Near-linear Time Decoding of Ta-Shma's Codes via Splittable Regularity

Fernando Granha Jeronimo* Shashank Srivastava† Madhur Tulsiani‡

Abstract

The Gilbert–Varshamov bound non-constructively establishes the existence of binary codes of distance $1/2 - \varepsilon/2$ and rate $\Omega(\varepsilon^2)$. In a breakthrough result, Ta-Shma [STOC 2017] constructed the first *explicit* family of nearly optimal binary codes with distance $1/2 - \varepsilon/2$ and rate $\Omega(\varepsilon^{2+\alpha})$, where $\alpha \to 0$ as $\varepsilon \to 0$. Moreover, the codes in Ta-Shma's construction are ε-balanced, where the distance between distinct codewords is not only bounded from below by $1/2 - \varepsilon/2$, but also from above by $1/2 + \varepsilon/2$.

Polynomial time decoding algorithms for (a slight modification of) Ta-Shma's codes appeared in [FOCS 2020], and were based on the Sum-of-Squares (SoS) semidefinite programming hierarchy. The running times for these algorithms were of the form $N^{O_\alpha(1)}$ for unique decoding, and $N^{O_{\varepsilon,\alpha}(1)}$ for the setting of "gentle list decoding", with large exponents of N even when α is a fixed constant. We derive new algorithms for both these tasks, running in time $\tilde{O}_\varepsilon(N)$. Our algorithms also apply to the general setting of decoding direct-sum codes.

Our algorithms follow from new structural and algorithmic results for collections of k-tuples (ordered hypergraphs) possesing a "structured expansion" property, which we call *splittability*. This property was previously identified and used in the analysis of SoS-based decoding and constraint satisfaction algorithms, and is also known to be satisfied by Ta-Shma's code construction. We obtain a new weak regularity decompomposition for (possibly sparse) splittable collections $W \subseteq [n]^k$, similar to the regularity decomposition for dense structures by Frieze and Kannan [FOCS 1996]. These decompositions are also computable in near-linear time $\tilde{O}(|W|)$, and form a key component of our algorithmic results.

*`University of Chicago`. `granha@uchicago.edu`. Supported in part by NSF grant CCF-1816372.
†`TTIC`. `shashanks@ttic.edu`. Supported in part by NSF grant CCF-1816372.
‡`TTIC`. `madhurt@ttic.edu`. Supported by NSF grant CCF-1816372.

Contents

1 Introduction

A binary code $\mathcal{C} \subseteq \mathbb{F}_2^N$ is said to be ε-balanced if any two distinct codewords $x, y \in \mathcal{C}$ satisfy $\Delta(x,y) \in [(1-\varepsilon)/2, (1+\varepsilon)/2]$, where $\Delta(x,y)$ denotes the relative distance between the two codewords. Finding explicit and optimal constructions of such codes, and indeed of codes where the distances are at least $(1-\varepsilon)/2$ is a central problem in coding theory [Gur10, Gur09], with many applications to the theory of pseudorandomness [Vad12]. Recently, Ta-Shma [TS17] gave a breakthrough construction of (a family of) explicit ε-balanced codes, with near-optimal rates, for arbitrarily small $\varepsilon > 0$. For the case of codes with distance at least $(1-\varepsilon)/2$, the existential rate-distance tradeoffs established by Gilbert [Gil52] and Varshamov [Var57], prove the existence of codes with rate $\Omega(\varepsilon^2)$, while McEliece et al. [MRRW77] prove an upper bound of $O(\varepsilon^2 \log(1/\varepsilon))$ on the rate. On the other hand, Ta-Shma's result yields an *explicit* family of codes with rate $\Omega(\varepsilon^{2+o(1)})$.

Decoding algorithms. The near-optimal ε-balanced codes of Ta-Shma [TS17] (which we will refer as Ta-Shma codes) were not known to be efficiently decodable at the time of their discovery. In later work, polynomial-time unique decoding algorithms for (a slight modification of) these codes were developed in [JQST20] (building on [AJQ+20]) using the Sum-of-Squares (SoS) hierarchy of semidefinite programming (SDP) relaxations. For unique decoding of codes with rates $\Omega(\varepsilon^{2+\alpha})$ (when $\alpha > 0$ is an arbitrarily small constant) these results yield algorithms running in time $N^{O_\alpha(1)}$. These algorithms also extend to the case when α is a vanishing function of ε, and to the problem of list decoding within an error radius of $1/2 - \varepsilon'$ (for ε' larger than a suitable function of ε) with running time $N^{O_{\varepsilon,\varepsilon',\alpha}(1)}$. However, the $O_\alpha(1)$ exponent of N obtained in the unique decoding case is quite large even for a fixed constant α (say $\alpha = 0.1$), and the exponent in the list decoding case grows with the parameter ε.

In this work, we use a different approach based on new weak regularity lemmas (for structures identified by the SoS algorithms), resulting in near-linear time algorithms for both the above tasks. The algorithms below work in time $\tilde{O}_\varepsilon(N)$ for ε-balanced Ta-Shma codes with rates $\Omega(\varepsilon^{2+\alpha})$, even when α is a (suitable) vanishing function of ε.

Theorem 1.1 (Near-linear Time Unique Decoding)**.** *For every $\varepsilon > 0$ sufficiently small, there are explicit binary linear Ta-Shma codes $\mathcal{C}_{N,\varepsilon,\alpha} \subseteq \mathbb{F}_2^N$ for infinitely many values $N \in \mathbb{N}$ with*

(i) distance at least $1/2 - \varepsilon/2$ (actually ε-balanced),

(ii) rate $\Omega(\varepsilon^{2+\alpha})$ where $\alpha = O(1/(\log_2(1/\varepsilon))^{1/6})$, and

(iii) an $r(\varepsilon) \cdot \tilde{O}(N)$ time unique decoding algorithm that that decodes within radius $1/4 - \varepsilon/4$ and works with high probability,

where $r(\varepsilon) = \exp(\exp(\mathrm{polylog}(1/\varepsilon)))$.

We can also obtain list decoding results as in [JQST20], but now in near-linear time.

Theorem 1.2 (Near-linear Time Gentle List Decoding)**.** *For every $\varepsilon > 0$ sufficiently small, there are explicit binary linear Ta-Shma codes $\mathcal{C}_{N,\varepsilon,\alpha} \subseteq \mathbb{F}_2^N$ for infinitely many values $N \in \mathbb{N}$ with*

(i) distance at least $1/2 - \varepsilon/2$ (actually ε-balanced),

(ii) rate $\Omega(\varepsilon^{2+\alpha})$ where $\alpha = O(1/(\log_2(1/\varepsilon))^{1/6})$, and

(iii) an $r(\varepsilon) \cdot \tilde{O}(N)$ time list decoding algorithm that decodes within radius $1/2 - 2^{-\Theta((\log_2(1/\varepsilon))^{1/6})}$ and works with high probability,

where $r(\varepsilon) = \exp(\exp(\mathrm{poly}(1/\varepsilon)))$.

While Theorem 1.2 yields a list decoding radius close to $1/2$, we remark that the above tradeoff between the list decoding radius and rate, is far from the state-of-the-art of $1/2 - \varepsilon$ radius with rate $\Omega(\varepsilon^3)$ of Guruswami and Rudra [GR06]. Considering a three way tradeoff involving distance, rate, and list-decoding radius, Theorem 1.2 can be seen as close to optimal with respect to the first two parameters, and quite far off with respect to the third one. Finding an algorithm for codes with optimal tradeoffs in all three parameters, is a very interesting open problem. Another interesting problem is understanding the optimal dependence of the "constant" factors $r(\varepsilon)$ in the running times. We have not tried to optimize these factors in our work.

Direct-Sum Codes and "Structured Pseudorandomness". Ta-Shma's code construction can be viewed as a special case of "distance amplification via direct-sum", an operation with several applications in coding and complexity theory [ABN+92, IW97, GI01, IKW09, DS14, DDG+15, Cha16, DK17, Aro02]. Given a (say) linear code $\mathcal{C}_0 \subseteq \mathbb{F}_2^n$ and a collection of tuples $W \subseteq [n]^k$, we define it's "direct-sum lifting" as $\mathcal{C} = \mathrm{dsum}_W(\mathcal{C}_0) \subseteq \mathbb{F}_2^{|W|}$ where

$$\mathrm{dsum}_W(\mathcal{C}_0) \;:=\; \left\{ (z_{i_1} + \cdots + z_{i_k})_{(i_1,\ldots,i_k) \in W} \;\mid\; z \in \mathcal{C}_0 \right\} .$$

It is easy to see that if $\mathcal{C}_0$ is ε_0-balanced for a constant ε_0, then taking $W = [n]^k$ results in $\mathrm{dsum}_W(\mathcal{C}_0)$ being ε-balanced with $\varepsilon = \varepsilon_0^k$ (though with vanishing rate). A standard sampling argument shows that a random $W \subseteq [n]^k$ with $|W| = O(n/\varepsilon^2)$ also suffices, while yielding rate $\Omega(\varepsilon^2)$. Rozenman and Wigderson [Bog12] suggested a derandomization of this argument using a "pseudorandom" W constructed from the collection of all length-$(k-1)$ walks on a suitable expander graph. While this result can be shown to achieve a rate of $\Omega(\varepsilon^{4+o(1)})$, Ta-Shma achieves a rate of $\Omega(\varepsilon^{2+o(1)})$ using a carefully constructed *sub-collection* of walks on an expander with a special form.

The above results show that pseudorandomness can be used to amplify distance, since the collections W above behave *like* a random W. However, finding decoding algorithms for such codes requires understanding properties of these collections which are *unlike* a random W, since random collections yield codes with (essentially) random generator matrices, where we do not expect efficient algorithms.

Our results can be viewed as showing that when the collection W satisfies a form of "structured multi-scale pseudorandomness" property [1] called *splittability* (identified in previous work), it can be exploited for algorithm design. One can think of splittability as capturing properties of the complete set $[n]^k$, which are not present in a (sparse) random $W \subseteq [n]^k$. For the case of $k = 4$, when $W = [n]^4$, if we consider a graph between pairs (i_1, i_2) and (i_3, i_4), which are connected when $(i_1, \ldots, i_4) \in W$, then this defines an expanding (complete) graph when $W = [n]^4$. On the other hand, for a random W of size $O(n)$,

[1] As discussed later, there are several notions of "structured pseudorandom" for (ordered and unordered) hypergraphs. We describe splittability here, since this is the one directly relevant for our algorithmic applications.

such a graph is a matching with high probability. Splittability requires various such graphs defined in terms of W to be expanders.

Definition 1.3 (Splittability, informal)**.** *Given $W \subseteq [n]^k$ and $a, b \in [k]$, let $W[a,b] \subseteq [n]^{b-a+1}$ denote the tuples obtained by considering $(i_a, \dots, i_b)$ for every $(i_1, \dots, i_k) \in W$. We say W can be τ-split at position t, if the bipartite graph with vertex sets $W[1,t]$ and $W[t+1,k]$, edge-set W, and (normalized) biadjacency matrix $\mathsf{S}_t \in \mathbb{R}^{W[1,t] \times W[t+1,k]}$, is an expander satisfying $\sigma_2(\mathsf{S}_t) \leq \tau$. We say that W is τ-splittable if for all $1 \leq a \leq t < b \leq k$, $W[a,b]$ can be τ-split at position t.*

Note that when $k = 2$, this coincides with the definition of (bipartite) graph expansion. It is also easy to show that collections of length-$(k-1)$ walks on a graph with second singular value λ, satisfy the above property with $\tau = \lambda$. The sub-collections used by Ta-Shma can also be shown to splittable (after a a slight modification) and we recall this proof from [JQST20] in Appendix A.

The key algorithmic component in our decoding results, is a general *list decoding* result for codes constructed via direct-sum operations, which reduces the task of list decoding for $\mathrm{dsum}_W(\mathcal{C}_0)$ to that of unique decoding for the code $\mathcal{C}_0$, when W is τ-splittable for an appropriate τ. The splittability property was identified and used in previous work [AJQ+20, JQST20], for the analysis of SoS based algorithms, which obtained the above reduction in $N^{O_\varepsilon(1)}$ time. Regularity based methods also allow for near-linear time algorithms in this general setting of direct-sum codes, with a simpler and more transparent proof (and improved dependence of the list decoding radius on τ and k).

Theorem 1.4 (List Decoding Direct Sum (informal version of Theorem 5.1))**.** *Let $\mathcal{C}_0 \subseteq \mathbb{F}_2^n$ be an ε_0-balanced linear code, which is unique-decodable to distance $(1-\varepsilon_0)/4$ in time $\mathcal{T}_0$. Let $W \subseteq [n]^k$ be a τ-splittable collection of tuples. Let $\mathcal{C} = \mathrm{dsum}_W(\mathcal{C}_0)$ be ε-balanced, and let β be such that*

$$\beta \gg \max\left\{\sqrt{\varepsilon},\ \left(\tau \cdot k^3\right)^{1/2},\ \left(\frac{1}{2} + 2\varepsilon_0\right)^{k/2}\right\}.$$

Then, there exists a randomized algorithm, which given $\tilde{y} \in \mathbb{F}_2^W$, recovers the list

$$\mathcal{L}_\beta(\tilde{y}) := \{y \in \mathcal{C} \mid \Delta(\tilde{y}, y) \leq 1/2 - \beta\},$$

with probability at least $1 - o(1)$, in time $\tilde{O}(C_{\beta,k,\varepsilon_0} \cdot (|W| + \mathcal{T}_0))$, where $C_{k,\beta,\varepsilon_0}$ only depends on k, β and ε_0.

Splittable Regularity. The technical component of our results is a novel understanding of splittable structures, via weak regularity lemmas. This provides a different way of exploiting "structured pseudorandomness" properties in hypergraphs, which may be of interest beyond applications considered here.

For the case of graphs (i.e., $k = 2$), several weak regularity lemmas are known which can be applied to (say) dense subgraphs of an expanding graph [RTTV08, TTV09, COCF09, BV20]. As in the Frieze-Kannan [FK96] weak regularity lemma for dense graphs, these lemmas decompose the adjacency matrix $A_{W'}$ of a subgraph $W' \subseteq W$, as a weighted sum of a small number of cut matrices ($\mathbf{1}_{S_\ell}\mathbf{1}_{T_\ell}^\mathsf{T}$ for $S_\ell, T_\ell \subseteq [n]$), such that one can use this decomposition to count the number of edges between any subsets $S, T \subseteq [n]$ i.e.,

$$\left|\mathbf{1}_S^\mathsf{T}\left(A_{W'} - \sum_\ell c_\ell \cdot \mathbf{1}_{S_\ell}\mathbf{1}_{T_\ell}^\mathsf{T}\right)\mathbf{1}_T\right| \leq \varepsilon \cdot |W|.$$

This can be thought of as computing an "approximation" of $A_{W'}$ using a small number of cut matrices $\mathbf{1}_{S_j}\mathbf{1}_{T_j}^{\mathsf{T}}$, which is "indistinguishable" by any cut matrix $\mathbf{1}_S\mathbf{1}_T^{\mathsf{T}}$.

More generally, one can think of the above results as approximating any function $g : W \to [-1,1]$ (with $g = \mathbf{1}_{W'}$ in the example above) with respect to a family of "split" functions $\mathcal{F} \subseteq \{f : [n] \to [-1,1]\}^{\otimes 2}$, where the approximation itself is a sum of a small number of of functions from $\mathcal{F}$ i.e., for all $f_1, f_2 \in \mathcal{F}$

$$\left| \left\langle g - \sum_{\ell} c_\ell \cdot f_{\ell,1} \otimes f_{\ell,2} \, , f_1 \otimes f_2 \right\rangle \right| \;\leq\; \varepsilon \cdot |W| \; .$$

Our regularity lemma for splittable $W \subseteq [n]^k$, extends the above notion of approximation, using k-wise split functions of the form $f_1 \otimes \cdots \otimes f_k$. We obtain near-linear time weak regularity decompositions for classes of k-wise cut functions of the form

$$\mathrm{CUT}^{\otimes k} := \{\pm \mathbf{1}_{S_1} \otimes \cdots \otimes \mathbf{1}_{S_k} \mid S_1, \ldots, S_k \subseteq [n]\},$$

and also for signed version of these k-wise cut functions

$$\mathrm{CUT}_{\pm}^{\otimes k} := \{\pm \chi_{S_1} \otimes \cdots \otimes \chi_{S_k} \mid S_1, \ldots, S_k \subseteq [n]\},$$

where $\chi_S = (-1)^{\mathbf{1}_S}$. For our decoding results, we will use $\mathrm{CUT}_{\pm}^{\otimes k}$. Our near-linear time weak regularity decomposition result is given next.

Theorem 1.5 (Efficient Weak Regularity (informal version of Theorem 4.11))**.** *Let $W \subseteq [n]^k$ and let $\mathcal{F}$ be either $\mathrm{CUT}^{\otimes k}$ or $\mathrm{CUT}_{\pm}^{\otimes k}$. Suppose $g \in \mathbb{R}^{[n]^k}$ is supported on W and has bounded norm. For every $\delta > 0$, if W is τ-splittable with $\tau = O(\delta^2/k^3)$, then we can find $h = \sum_{\ell=1}^{p} c_\ell \cdot f_\ell$ in $\widetilde{O}_{k,\delta}(|W|)$ time, where $p = O(k^2/\delta^2)$, $f_\ell \in \mathcal{F}$ and $c_\ell \in \mathbb{R}$, such that h is a good approximator to g in the following sense*

$$\max_{f \in \mathcal{F}} \quad \langle g - h, f \rangle \;\leq\; \delta \cdot |W| \,,$$

where the inner product is over the counting measure on $[n]^k$.

We note that an existential version of the above theorem follows known abstract versions of the Frieze-Kannan regularity lemma [TTV09, BV20], via a relatively simple use of splittability. However, making a black-box application of known regularity lemmas algorithmic, requires computing a form of "tensor cut-norm", which is believed to be hard to even approximate in general[2] (unlike the matrix case). The nontrivial component of the result above, is obtaining a regularity lemma which allows for a *near-linear time computation*, while still achieving parameters close to the existential version.

Related Work. As discussed above, the decoding results in this paper, were derived earlier using algorithms based on the SoS hierarchy [AJQ+20, JQST20], though with significantly larger running times (and somewhat worse dependence on parameters). A common thread in the SoS algorithms is to relate the task of decoding, to that of solving instances

[2] Strictly speaking, we only need to approximate this for "splittable" tensors. It is possible that one could use existing regularity lemmas black box, and use splittability to design a fast algorithm for tensor cut-norm. In our proof, we instead choose to use the matrix cut-norm algorithms as black-box, and use splittability to modify the proof of the regularity lemma.

of constraint satisfaction problems with k variables in each constraint (k-CSPs). The original weak regularity lemma of Frieze and Kannan [FK96] was indeed motivated by the question of approximately solving k-CSPs on dense structures (see also [KV09]). Several extensions of the Frieze-Kannan lemma are known, particularly for various families of sparse pseudorandom graphs [KR02, RTTV08, TTV09, OGT15, BV20]. Oveis-Gharan and Trevisan [OGT15] also proved a new weak regularity lemma for "low threshold-rank" graphs, which was used to obtain approximation algorithms for some 2-CSPs, where the previously known algorithms used the SoS hierarchy [BRS11, GS11]. Our work can be viewed as an extension of these ideas to the case of k-CSPs.

Ideas based on regularity lemmas, were also employed in the context of list decoding of Reed-Muller codes, by Bhowmick and Lovett [BL18]. They use analogues of the abstract weak regularity lemma [TTV09] and the Szemerédi regularity lemma over finite fields, but these are only used to prove bounds on the list size, rather than in the algorithm itself. On the other hand, our decoding algorithm crucially uses the decomposition obtained via our weak regularity lemma for (real-valued functions on) splittable structures.

In general, expansion phenomena have a rich history of interaction with coding theory (e.g., [GI01, Gur04, GI05, RWZ20]) including to the study of linear (or near-linear) time decoding backing to the seminal work of Sipser and Spielman [SS96]. The codes in [SS96] were good codes, though not near optimal in terms of distance-rate trade-off. Several other notions of "structured pseudorandomness" for hypergraphs (referred to as high-dimensional expansion) have also been considered in literature, which also have connections to the decoding of good codes. In particular, the notion of "double sampler" was used to obtain algorithms for the list decoding for direct-product codes [DHK$^+$19]. The notions of local spectral expansion [DK17], cosystolic expansion [EK16], and multilayer agreement samplers [DDHRZ20], are also used to connect structured pseudorandomness to the design of locally testable codes. The notion of splittability was also studied for unordered hypergraphs in terms of "complement walks" by Dinur and Dikstein [DD19], and in terms of "swap walks" in [AJT19], for high-dimensional expanders defined via local spectral expansion.

2 A Technical Overview

We now give a more detailed overview of some of the technical components of our proof.

Splittability. The key structural property used for our algorithmic and structural results, is the "structured pseudorandomness" of ordered hypergraphs $W \subseteq [n]^k$, which we call *splittability*. The canonical example one can think of for this case, is a collection of all length-$(k-1)$ walks on a (say) d-regular expander graph G on n vertices. Note that this satisfies $|W[a,b]| = d^{b-a} \cdot n$, where $W[a,b]$ represents the collection of sub-tuples with coordinates between indices a and b i.e., portions of the walks between the a^{th} and b^{th} step. We will restrict our discussion in this paper only to *d-regular collections* $W \subseteq [n]^k$ satisfying $|W[a,b]| = d^{b-a} \cdot n$.

We briefly sketch why the collection of length-3 walks (i.e., the case $k=4$) is splittable. Recall that splittability requires various graphs with sub-tuples to be expanding, and in particular consider the graph between $W[1,2]$ and $W[3,4]$, with edge-set $W[1,4]$. If $E(G)$

is the set of edges in G included with both orientations, then note that $W[1,2] = W[3,4] = E(G)$, and $(i_1, i_2), (i_3, i_4)$ are connected iff $(i_2, i_3) \in E(G)$. If $M \in \mathbb{R}^{W[1,2] \times W[3,4]}$ denotes the biadjacency matrix of the bipartite graph H on $W[1,2] \times W[3,4]$, then up to permutations of rows and columns, we can write M as $\mathsf{A}_G \otimes \mathsf{J}_d / d$, where J_d denotes the $d \times d$ all-1s matrix and A_G is the normalized adjacency matrix of G, since each tuple $(i_2, i_3) \in E(G)$ contributes d^2 edges in H (for choices of i_1 and i_4). Thus $\sigma_2(M) = \sigma_2(\mathsf{A}_G)$, which is small if G is an expander. A similar argument also works for splits in other positions, and for longer walks.

The above argument can also be extended to show that the sub-collections of walks considered by Ta-Shma (after a slight modification) are splittable, though the structure and the corresponding matrices are more involved there (see Appendix A).

Regularity for graphs and functions. We first consider an analytic form of the Frieze-Kannan regularity lemma (based on [TTV09]). Let $g : \mathcal{X} \to [-1,1]$ be any function on a finite space $\mathcal{X}$ with an associated probability measure μ, and let $\mathcal{F} \subseteq \{f : \mathcal{X} \to [-1,1]\}$ be any class of functions closed under negation. Say we want to construct a "simple approximation/decomposition" h, which is indistinguishable from g, for all functions in f i.e.,

$$\langle g - h, f \rangle_\mu \;=\; \mathop{\mathbb{E}}_{x \sim \mu}\left[(g(x) - h(x)) \cdot f(x)\right] \;\leq\; \delta \qquad \forall f \in \mathcal{F}\,.$$

We can view the regularity lemma as saying that such an h can always be constructed as a sum of $1/\delta^2$ functions from $\mathcal{F}$. Indeed, we can start with $h^{(0)} = 0$, and while there exists f_ℓ violating the above condition, we update $h^{(\ell+1)} = h^{(\ell)} + \delta \cdot f_\ell$. The process must stop in $1/\delta^2$ steps, since $\|g - h^{(\ell)}\|^2$ can be shown to decrease by δ^2 in every step.

$$\left\| g - h^{(\ell)} \right\|_\mu^2 - \left\| g - h^{(\ell+1)} \right\|_\mu^2 \;=\; 2\delta \cdot \left\langle g - h^{(\ell)}, f_\ell \right\rangle_\mu - \delta \cdot \|f_\ell\|_\mu^2 \;\geq\; \delta^2\,.$$

In fact, the above can be seen as gradient descent for minimizing the convex function $F(h) = \sup_{f \in \mathcal{F}} \langle g - h, f \rangle_\mu$. Taking $\mathcal{X} = [n]^2$ with μ as uniform on $[n]^2$, $g = \mathbf{1}_{E(G)}$ for a (dense) graph G, and $\mathcal{F}$ as all functions (cut matrices) of the form $\pm \mathbf{1}_S \mathbf{1}_T^\mathsf{T}$ yields the weak regularity lemma for graphs, since we get $h = \sum_\ell c_\ell \cdot f_\ell = \sum_\ell c_\ell \cdot \mathbf{1}_{S_\ell} \mathbf{1}_{T_\ell}^\mathsf{T}$ such that

$$\langle g - h, f \rangle_\mu \leq \delta \quad \forall f \in \mathcal{F} \quad \Leftrightarrow \quad \frac{1}{n^2} \cdot \left| E_G(S,T) - \sum_\ell c_\ell \, |S_\ell \cap S| \, |T_\ell \cap T| \right| \leq \delta \quad \forall S, T \subseteq [n]\,.$$

Note that the inner product in the above analytic argument can be chosen to be according to any measure on $\mathcal{X}$, and not just the uniform measure. In particular, taking $W \subseteq [n]^2$ to be the edge-set of a (sparse) d-regular expander with second singular value (say) λ, and $\mu = \mu_2$ to be uniform over W, we obtain the regularity lemma for subgraphs of expanders. In this case, after obtaining the approximation with respect to μ, one shows using the expander mixing lemma that if $\langle g - h, f \rangle_{\mu_2} \leq \delta$, then $\langle g - ({}^d/_n) \cdot h, f \rangle_{\mu_1 \otimes \mu_1} \leq ({}^d/_n) \cdot \delta'$, where μ_1 denotes the uniform measure on $[n]$ and $\delta' = \delta + \lambda$. This gives a sparse regularity lemma, since for $G \subseteq W$ and $g = \mathbf{1}_G$,

$$\left\langle g - \left(\frac{d}{n}\right) h, f \right\rangle_{\mu_1^{\otimes 2}} \leq \frac{d}{n} \cdot \delta' \quad \forall f \in \mathcal{F} \quad \Leftrightarrow \quad \left| E_G(S,T) - \sum_\ell c_\ell \cdot \frac{d}{n} \, |S_\ell \cap S| \, |T_\ell \cap T| \right| \leq \delta' \cdot nd \quad \forall S, T\,.$$

The *algorithmic step* in the above proofs, is finding an f_ℓ such that $\langle g - h, f_\ell \rangle > \delta$. For the function class $\mathcal{F}$ corresponding to cut matrices, this corresponds to solving a problem of the form $\max_{S,T} \left|\mathbf{1}_S^{\mathsf{T}} M \mathbf{1}_T\right|$ for an appropriate matrix M at each step. This equals the cut-norm and can be (approximately) computed using the SDP approximation algorithm of Alon and Naor [AN04]. Moreover, this can be implemented in near-linear time in the sparsity of M, using known fast, approximate SDP solvers of Lee and Padmanabhan [LP20] or of Arora and Kale [AK07] (see Section 4.5 for details).

Splittable regularity. For our regularity lemma, the class $\mathcal{F}$ comprises of "k-split functions" of the form $f_1 \otimes \cdots \otimes f_k$, where for each f_t can be thought of as $\mathbf{1}_{S_t}$ (or $(-1)^{\mathbf{1}_{S_t}}$) for some $S_t \subseteq [n]$. An argument similar to the one above, with the measure μ_k uniform on $W \subseteq [n]^k$, can yield an *existential version* of the splittable regularity lemma, similar to the one for expander graphs (we now transition from μ_k to $\mu_1^{\otimes k}$ using a simple generalization of the expander mixing lemma to splittable collections). However, the algorithmic step in the above procedure, requires computing

$$\max_{f_1,\ldots,f_k \in \mathcal{F}} \langle g - h, f_1 \otimes \cdots \otimes f_k \rangle$$

Unfortunately, such an algorithmic problem is hard to even approximate in general, as opposed to the 2-split case for graphs. Another approach is to first compute an approximation of a given $g : W \to [-1,1]$, in terms of 2-split functions of the form $f_1 \otimes f_2$, where $f_1 : W[1,t] \to [-1,1]$ and $f_2 : W[t+1,k] \to [-1,1]$, and then inductively approximate f_1 and f_2 in terms of 2-split functions, and so on. Such an induction does yield an algorithmic regularity lemma, though naively approximating the component functions f_1 and f_2 at each step, leads to a significantly lossy dependence between the final error, the splittability parameter τ, and k.

We follow a hybrid of the two approaches above. We give an inductive argument, which at step t, approximates g via h_t which is a sum of t-split functions. However, instead of simply applying another 2-split to each term in the decomposition h_t to compute h_{t+1}, we build an approximation for *all of* h_t using the regularity argument above from scratch. We rely on the special structure of h_t to solve the algorithmic problem $\max_{f_1,\ldots,f_{t+1}} \langle h_t - h_{t+1}, f_1 \otimes \cdots \otimes f_{t+1} \rangle$, reducing it to a matrix cut-norm computation[3]. This yields near-optimal dependence of the error on τ and k, needed for our coding applications.

Decoding direct-sum codes using regularity. We now consider the problem of decoding, from a received, possibly corrupted, $\tilde{y} \in \mathbb{F}_2^W$, to obtain the closest $y \in \mathrm{dsum}_W(\mathcal{C}_0)$ (or a list) i.e., finding $\mathrm{argmin}_{z_0 \in \mathcal{C}_0} \Delta(\tilde{y}, \mathrm{dsum}_W(z_0))$. Let $g : [n]^k \to \{-1,1\}$ be defined as $g(i_1,\ldots,i_k) = (-1)^{\tilde{y}_{(i_1,\ldots,i_k)}}$ if $(i_1,\ldots,i_k) \in W$ and 0 otherwise. Also, for any $z \in \mathbb{F}_2^n$, define the function χ_z as $\chi_z(i) = (-1)^{z_i}$. As before, let μ_1 denote the uniform measure on $[n]$.

[3]Strictly speaking, we also need to be careful about the bit-complexity of our matrix entries, to allow for near-linear time computation. However, all the entries in matrices we consider will have bit-complexity $O_{k,\delta}(\log n)$.

Using that g is 0 outside W, and that $|W| = d^{k-1} \cdot n$, we get

$$\begin{aligned} 1 - 2 \cdot \Delta(\tilde{y}, \operatorname{dsum}_W(z)) &= \mathop{\mathbb{E}}_{(i_1,\ldots,i_k) \in W} [g(i_1, \ldots, i_k) \cdot \chi_z(i_1) \cdots \chi_z(i_k)] \\ &= \left(\frac{n}{d}\right)^{k-1} \cdot \mathop{\mathbb{E}}_{(i_1,\ldots,i_k) \in [n]^k} [g(i_1, \ldots, i_k) \cdot \chi_z(i_1) \cdots \chi_z(i_k)] \\ &= \left(\frac{n}{d}\right)^{k-1} \cdot \left\langle g, \chi_z^{\otimes k} \right\rangle_{\mu_1^{\otimes k}} . \end{aligned}$$

At this point, we modify the problem in three ways. First, instead of restricting the optimization to $z_0 \in \mathcal{C}_0$, we widen the search to all $z \in \mathbb{F}_2^n$. We will be able to show that because of the pseudorandom (distance amplification) properties of W, a good (random) solution z found by our algorithm, will be within the unique decoding radius of $\mathcal{C}_0$ (with high probability). Secondly, using the fact that for splittable W, the function g has an approximation $h = \sum_{\ell=1}^{p} c_\ell \cdot f_{\ell,1} \otimes \cdots \otimes f_{\ell,k}$ given by the regularity lemma, we can restrict our search to z which (approximately) maximize the objective

$$\left\langle h, \chi_z^{\otimes k} \right\rangle_{\mu_1^{\otimes k}} = \sum_{\ell=1}^{p} c_\ell \cdot \prod_{t \in [k]} \langle f_{\ell,t}, \chi_z \rangle_{\mu_1}$$

Finally, instead of searching for $\chi_z : [n] \to \{-1, 1\}$, we further widen the search to $\overline{f} : [n] \to [-1, 1]$. A random "rounding" choosing each $\chi_z(i)$ independently so that $\mathbb{E}[\chi_z] = \overline{f}$ should preserve the objective value with high probability. We now claim that the resulting search for functions $\overline{f}$ maximizing $\left\langle h, \overline{f}^{\otimes k} \right\rangle_{\mu_1^{\otimes k}}$, can be solved via a simple brute-force search. Note that the objective only depends on the inner products with a finite number of functions $\{f_{\ell,t}\}_{\ell \in [p], t \in [k]}$ with range $\{-1, 1\}$. Partitioning the space $[n]$ in 2^{pk} "atoms" based on the values of these functions, we can check that it suffices to search over $\overline{f}$, which are constant on each atom. Moreover, it suffices to search the values in each atom, up to an appropriate discretization η, which can be done in time $O\left((1/\eta)^{2^{pk}}\right)$.

For the problem of list decoding $\tilde{y}$ up to radius $1/2 - \beta$, we show that each $z_0 \in \mathcal{C}_0$, such that $\operatorname{dsum}_W(z_0)$ is in the list, there must be an $\overline{f}$ achieving a large value of $\left\langle h, \overline{f}^{\otimes k} \right\rangle_{\mu_1^{\otimes k}}$ which then yields a z within the unique decoding radius of z_0. Since we enumerate over all $\overline{f}$, this recovers the entire list. Details of the decoding algorithm are given in Section 5.

3 Preliminaries

We now introduce some notation. The asymptotic notation $\widetilde{O}(r(n))$ hides polylogarithmic factors in $r(n)$.

3.1 Codes

We briefly recall some standard code terminology. Given $z, z' \in \mathbb{F}_2^n$, recall that the relative Hamming distance between z and z' is $\Delta(z, z') := |\{i \mid z_i \neq z'_i\}| / n$. A binary code is any subset $\mathcal{C} \subseteq \mathbb{F}_2^n$. The distance of $\mathcal{C}$ is defined as $\Delta(\mathcal{C}) := \min_{z \neq z'} \Delta(z, z')$ where $z, z' \in \mathcal{C}$. We say that $\mathcal{C}$ is a linear code if $\mathcal{C}$ is a linear subspace of $\mathbb{F}_2^n$. The rate of $\mathcal{C}$ is $\log_2(|\mathcal{C}|)/n$, or equivalently $\dim(\mathcal{C})/n$ if $\mathcal{C}$ is linear.

Definition 3.1 (Bias). *The* bias *of a word $z \in \mathbb{F}_2^n$ is defined as* $\operatorname{bias}(z) := \left|\mathbb{E}_{i\in[n]}(-1)^{z_i}\right|$. *The bias of a code $\mathcal{C}$ is the maximum bias of any non-zero codeword in $\mathcal{C}$.*

Definition 3.2 (ε-balanced Code). *A binary code $\mathcal{C}$ is ε-*balanced *if* $\operatorname{bias}(z+z') \le \varepsilon$ *for every pair of distinct $z, z' \in \mathcal{C}$.*

Remark 3.3. *For linear binary code $\mathcal{C}$, the condition* $\operatorname{bias}(\mathcal{C}) \le \varepsilon$ *is equivalent to $\mathcal{C}$ being an ε-balanced code.*

3.2 Direct Sum Lifts

Starting from a code $\mathcal{C} \subseteq \mathbb{F}_2^n$, we amplify its distance by considering the *direct sum lifting* operation based on a collection $W(k) \subseteq [n]^k$. The direct sum lifting maps each codeword of $\mathcal{C}$ to a new word in $\mathbb{F}_2^{|W(k)|}$ by taking the k-XOR of its entries on each element of $W(k)$.

Definition 3.4 (Direct Sum Lifting). *Let $W(k) \subseteq [n]^k$. For $z \in \mathbb{F}_2^n$, we define the* direct sum lifting *as* $\operatorname{dsum}_{W(k)}(z) = y$ *such that* $y_{(i_1,\dots,i_k)} = \sum_{j=1}^k z_{i_j}$ *for all $(i_1, \dots, i_k) \in W(k)$. The direct sum lifting of a code $\mathcal{C} \subseteq \mathbb{F}_2^n$ is*

$$\operatorname{dsum}_{W(k)}(\mathcal{C}) = \{\operatorname{dsum}_{W(k)}(z) \mid z \in \mathcal{C}\}.$$

We will omit $W(k)$ from this notation when it is clear from context.

Remark 3.5. *We will be concerned with collections $W(k) \subseteq [n]^k$ arising from length-$(k-1)$ walks on expanding structures (mostly arising from Ta-Shma's direct sum construction [TS17]).*

We will be interested in cases where the direct sum lifting reduces the bias of the base code; in [TS17], structures with such a property are called *parity samplers*, as they emulate the reduction in bias that occurs by taking the parity of random samples.

Definition 3.6 (Parity Sampler). *A collection $W(k) \subseteq [n]^k$ is called an $(\varepsilon_0, \varepsilon)$-*parity sampler *if for all $z \in \mathbb{F}_2^n$ with* $\operatorname{bias}(z) \le \varepsilon_0$, *we have* $\operatorname{bias}(\operatorname{dsum}_{W(k)}(z)) \le \varepsilon$.

3.3 Splittable Tuples

We now formally define the *splittability* property for a collection of tuples $W(k) \subseteq [n]^k$. For $1 \le a \le b \le k$, we define $W[a,b] \subseteq [n]^{(b-a+1)}$ as

$$W[a,b] := \{(i_a, i_{a+1}, \dots, i_b) \mid (i_1, i_2, \dots, i_k) \in W(k)\}.$$

We will work with d-regular tuples in the following sense.

Definition 3.7 (Regular tuple collection). *We say that $W(k) \subseteq [n]^k$ is d-regular if for every $1 \le a \le b \le k$, we have*

- $|W[a,b]| = d^{b-a} \cdot n$,
- $W[a] = [n]$.

A collection $W(k)$ being d-regular is analogous to a graph being d-regular.

Example 3.8. *The collection $W(k)$ of all length-$(k-1)$ walks on a d-regular connected graph $G = ([n], E)$ is a d-regular collection of tuples.*

The space of functions $\mathbb{R}^{W[a,b]}$ is endowed with an inner product associated to the uniform measure $\mu_{[a,b]}$ on $W[a,b]$. We use the shorthand μ_b for $\mu_{[1,b]}$.

Definition 3.9 (Splitable tuple collection)**.** *Let $\tau > 0$. We say that a collection $W(k) \subseteq [n]^k$ is* τ-splittable *if it is d-regular and either $k = 1$ or for every $1 \le a \le t < b \le k$ we have*

- *the* split *operator* $\mathsf{S}_{W[a,s],W[t+1,b]} \in \mathbb{R}^{W[a,t] \times W[t+1,b]}$ *defined as*

$$\left(\mathsf{S}_{W[a,t],W[t+1,b]}\right)_{(i_a,\ldots,i_t),(i_{t+1},\ldots,i_k)} := \frac{\mathbf{1}\left[(i_a,\ldots,i_t,i_{t+1},\ldots i_k) \in W[a,b]\right]}{d^{k-t}}$$

satisfy $\sigma_2(\mathsf{S}_{W[a,t],W[t+1,b]}) \le \tau$, where σ_2 denotes the second largest singular value.

Example 3.10. *The collection $W(k)$ of all length-$(k-1)$ walks on a d-regular a graph $G = ([n], E)$ whose normalized adjacency matrix has second largest singular value at most τ is a collection of τ-splittable tuples as shown in [AJQ+20].*

Example 3.11. *The collection $W(k)$ of tuples arising (from a slight modification) of the direct sum construction of Ta-Shma [TS17] is a τ-splittable as shown in [JQST20]. Precise parameters are recalled later as Theorem A.1 of Appendix A.*

3.4 Factors

It will be convenient to use the language of factors, to search the decompositions identified by regularity lemmas, for relevant codewords. This concept (from ergodic theory) takes a rather simple form in our finite settings: it is just a partition of base set $\mathcal{X}$, with an associated operation of averaging functions defined on $\mathcal{X}$, separately over each piece.

Definition 3.12 (Factors and measurable functions)**.** *Let $\mathcal{X}$ be a finite set. A* factor *$\mathcal{B}$ is a partition of the set $\mathcal{X}$, and the subsets of the partition are referred to as* atoms of the factor. A function $f : \mathcal{X} \to \mathcal{R}$ is said to measurable with respect to $\mathcal{B}$ ($\mathcal{B}$-measurable) if f is constant on each atom of $\mathcal{B}$.

Definition 3.13 (Conditional averages)**.** *If $f : \mathcal{X} \to \mathbb{R}$ is a function, μ is a measure on the space $\mathcal{X}$, and $\mathcal{B}$ is a factor, then we define the* conditional average function $\mathbb{E}[f|\mathcal{B}]$ *as*

$$\mathbb{E}\left[f|\mathcal{B}\right](x) \; := \; \mathop{\mathbb{E}}_{y \sim \mu|\mathcal{B}(x)}\left[f(y)\right] ,$$

where $\mathcal{B}(x)$ denotes the atom containing x. Note that the function $\mathbb{E}[f|\mathcal{B}]$ is measurable with respect to $\mathcal{B}$.

We will need the following simple observation regarding conditional averages.

Proposition 3.14. *Let $h : \mathcal{X} \to \mathbb{R}$ be a $\mathcal{B}$-measurable function, and let $f : \mathcal{X} \to \mathbb{R}$ be any function. Then, for any measure μ over $\mathcal{X}$, we have*

$$\langle h, f \rangle_\mu \; = \; \langle h, \mathbb{E}\left[f|\mathcal{B}\right] \rangle_\mu \, .$$

Proof. By definition of the $\mathcal{B}$-measurability, h is constant on each atom, and thus we can write $h(x)$ as $h(\mathcal{B}(x))$.

$$\begin{aligned}\langle h, f\rangle_\mu = \mathop{\mathbb{E}}_{x\sim\mu}[h(x)\cdot f(x)] &= \mathop{\mathbb{E}}_{x\sim\mu}\mathop{\mathbb{E}}_{y\sim\mu|\mathcal{B}(x)}[h(y)\cdot f(y)] \\ &= \mathop{\mathbb{E}}_{x\sim\mu}\left[h(\mathcal{B}(x))\cdot \mathop{\mathbb{E}}_{y\sim\mu|\mathcal{B}(x)}[f(y)]\right] \\ &= \mathop{\mathbb{E}}_{x\sim\mu}[h(x)\cdot \mathbb{E}[f|\mathcal{B}](x)] = \langle h, \mathbb{E}[f|B]\rangle_\mu . \qquad \blacksquare\end{aligned}$$

The factors we will consider will be defined by a finite collection of functions appearing in a regularity decomposition.

Definition 3.15 (Function factors)**.** *Let $\mathcal{X}$ and $\mathcal{R}$ be finite sets, and let $\mathcal{F}_0 = \{f_1, \ldots, f_r : \mathcal{X} \to \mathcal{R}\}$ be a finite collection of functions. We consider the factor $\mathcal{B}_{\mathcal{F}_0}$ defined by the functions in $\mathcal{F}_0$, as the factor with atoms $\{x \mid f_1(x) = c_1, \ldots, f_r(x) = c_r\}$ for all $(c_1, \ldots, c_r) \in \mathcal{R}^r$.*

Remark 3.16. *Note that when the above function are indicators for sets i.e., each $f_j = \mathbf{1}_{S_j}$ for some $S_j \subseteq \mathcal{X}$, then the function factor $\mathcal{B}_{\mathcal{F}_0}$ is the same as the σ-algebra generated by these sets. Also, given the functions $f_1, \ldots, f_r$ as above, the function factor $\mathcal{B}_{\mathcal{F}_0}$ can be computed in time $O(|\mathcal{X}| \cdot |\mathcal{R}|^r)$.*

3.5 Functions and Measures

We describe below some classes of functions, and spaces with associated measures, arising in our proof. The measures we consider are either uniform on the relevant space, or are products of measures on its component spaces.

Function classes. Let $S \subseteq [n]$. We define $\chi_S \colon [n] \to \{\pm 1\}$ as $\chi_S(i) := (-1)^{\mathbf{1}_{i\in S}}$ (we observe that as defined χ_S is not a character[4]). We need the following two collection of functions for which algorithmic results will be obtained.

Definition 3.17 (CUT functions)**.** *We define the set of* $0/1$ CUT *cut functions as*

$$\mathrm{CUT}^{\otimes k} := \{\pm \mathbf{1}_{S_1} \otimes \cdots \otimes \mathbf{1}_{S_k} \mid S_1, \ldots, S_k \subseteq [n]\},$$

and defined the set of ± 1 CUT *functions as*

$$\mathrm{CUT}_{\pm}^{\otimes k} := \{\pm \chi_{S_1} \otimes \cdots \otimes \chi_{S_k} \mid S_1, \ldots, S_k \subseteq [n]\}.$$

We will use a higher-order version of cut norm.

Definition 3.18. *Let $g \in \mathbb{R}^{[n]^k}$, the k-tensor cut norm is*

$$\|g\|_{\square^{\otimes k}} := \max_{f \in \mathrm{CUT}^{\otimes k}} \langle g, f\rangle ,$$

where the inner product is over the counting measure on $[n]^k$.

[4]Strictly speaking χ_S is not a character but by identifying the elements of $[n]$ with those of a canonical basis of $\mathbb{F}_2^n$ it becomes a character for $\mathbb{F}_2^n$.

Some of our results hold for more general class of functions.

Definition 3.19 (t-split functions)**.** *Suppose $W(k)$ is a regular collection of k-tuples. For $t \in \{0, \dots, k-1\}$, we define a generic class of tensor product functions $\mathcal{F}_t$ as*

$$\mathcal{F}_t \subseteq \left\{ \pm f_1 \otimes \cdots \otimes f_t \otimes f_{t+1} \mid f_j \subseteq \mathbb{R}^{W[1]} \text{ for } i \leq t, f_{t+1} \subseteq \mathbb{R}^{W[t+1,k]}, \left\|f_j\right\|_\infty \leq 1 \right\}.$$

To avoid technical issues, we assume that each $\mathcal{F}_t$ is finite.

Fixing some $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$, we define the set of functions that are linear combinations of function from $\mathcal{F}$ with coefficients of bounded support size and bounded ℓ_1-norm as follows

$$\mathcal{H}(R_0, R_1, \mathcal{F}) := \left\{ \sum_{\ell=1}^{p} c_\ell \cdot f_\ell \mid p \leq R_0, \sum |c_\ell| \leq R_1, f_\ell \in \mathcal{F} \right\}.$$

Measures and inner products. Recall that $\mu_1 := \mu_{[1,1]}$ is the uniform measure on $W[1]$ (equivalently uniform measure on $W[i]$ since $W(k)$ is regular) and $\mu_{[t+1,k]}$ is the uniform measure on $W[t+1,k]$. We define following measure ν_t as

$$\nu_t := (\mu_1)^{\otimes t} \otimes \left(\mu_{[t+1,k]}\right).$$

Note that ν_0 is the equal to μ_k and ν_{k-1} is equal to $\mu_1^{\otimes k}$. We will need to consider inner products of functions according to various measures defined above, which we will denote as $\langle \cdot, \cdot \rangle_\mu$ for the measure μ. When a measure is not indicated, we take the inner product $\langle f, g \rangle$ to be according to the counting measure on the domains of the functions f and g.

4 Weak Regularity for Splittable Tuples

We will show how functions supported on a (possibly) sparse splittable collection of tuples $W(k) \subseteq [n]^k$ admit weak regular decompositions in the style of Frieze and Kannan [FK96]. In Section 4.1, we start by showing an abstract regularity lemma for functions that holds in some generality and does not require splittability. Next, in Section 4.2, we show that splittable collections of tuples satisfy suitable (simple) generalizations of the expander mixing lemma for graphs which we call splittable mixing lemma. By combining this abstract weak regularity decomposition with splittable mixing lemmas, we obtain *existential* decomposition results for splittable tuples in Section 4.3. Then, we proceed to make these existential results not only algorithmic but near-linear time computable in Section 4.4. These algorithmic results will rely on fast cut norm like approximation algorithms tailored to our settings and this is done in Section 4.5. As mentioned previously, this last step borrows heavily from known results [AN04, AK07, LP20].

4.1 Abstract Weak Regularity Lemma

We now show a weak regularity decomposition lemma for functions that works in some generality and does not require splittability. We now fix some notation for this section. Let $\mathcal{X}$ be a finite set endowed with a probability measure μ. Let $\mathbb{R}^{\mathcal{X}}$ be a Hilbert space endowed with inner product $\langle f, g \rangle_\mu = \mathbb{E}_\mu [f \cdot g]$ and associated norm $\|\cdot\|_\mu = \sqrt{\langle \cdot, \cdot \rangle_\mu}$. Let $\mathcal{F} \subseteq \{f \colon \mathcal{X} \to \mathbb{R} \mid \|f\|_\mu \le 1\}$ be a finite collection of functions such that $\mathcal{F} = -\mathcal{F}$.

In a nutshell, given any $g \in \mathbb{R}^{\mathcal{X}}$, the abstract weak regularity lemma will allow us to find an approximator h, with respect to the semi-norm $g - h \mapsto \max_{f \in \mathcal{F}} \langle g - h, f \rangle$, which is a linear combinations of a certain *small* number of functions from $\mathcal{F}$ (where this number depends only on the approximation accuracy and the norm $\|g\|_\mu$). This means that g and h have approximately the same correlations with functions from $\mathcal{F}$. We will produce h in an iterative procedure, where at each step an oracle of the following kind (*cf.*, Definition 4.1) is invoked.

Definition 4.1 (Correlation Oracle)**.** *Let $1 \ge \delta \ge \delta' > 0$ be accuracy parameters and $B > 0$. We say that $\mathcal{O}_{\mu,B}$ is a (δ, δ')-correlation oracle for $\mathcal{F}$ if given $h \in \mathbb{R}^{\mathcal{X}}$ with $\|h\|_\mu^2 = O(B)$ if there exists $f \in \mathcal{F}$ with $\langle h, f \rangle \ge \delta$, then $\mathcal{O}_{\mu,B}$ returns some $f' \in \mathcal{F}$ with $\langle h, f' \rangle \ge \delta'$.*

More precisely, our abstract weak regularity decomposition is as follows.

Lemma 4.2 (Abstract Weak Regularity)**.** *Let $\mathcal{O}_{\mu,B}$ be a (δ, δ')-correlation oracle for $\mathcal{F}$ with $\delta \ge \delta' > 0$. Let $g \colon \mathcal{X} \to \mathbb{R}$ satisfy $\|g\|_\mu^2 \le B$. Then, we can find $h = \sum_{\ell=1}^{p} c_\ell \cdot f_\ell \in \mathcal{H}(B/(\delta')^2, B/\delta', \mathcal{F})$ with $f_\ell \in \mathcal{F}$, $c_\ell \in [\delta'/(1 + \delta'/\sqrt{B})^p, \delta']$ and $\|h\|_\mu^2 \le B$ such that*

$$\max_{f \in \mathcal{F}} \quad \langle g - h, f \rangle_\mu \ \le \ \delta.$$

Furthermore, if $\mathcal{O}_{\mu,B}$ runs in time $\mathcal{T}_{\mathcal{O}_{\mu,B}}$, then h can be computed in

$$\widetilde{O}\left(\mathrm{poly}(B, 1/\delta') \cdot (\mathcal{T}_{\mathcal{O}_{\mu,B}} + |\mathrm{Supp}(\mu)|)\right)$$

time, where $\mathrm{Supp}(\mu)$ is the support of μ. The function h is constructed in Algorithm 4.3 as the final function in a sequence of approximating functions $h^{(\ell)} \in \mathcal{H}(B/(\delta')^2, B/\delta', \mathcal{F})$.

The proof is based on the following algorithm.

Algorithm 4.3 (Regularity Decomposition Algorithm)**.**

Input $g \colon \mathcal{X} \to \mathbb{R}$

Output $h = \sum_{\ell=1}^{p} c_\ell \cdot f_\ell$

- *Let Π be the projector onto the convex ball $\{g' \in \mathbb{R}^{\mathcal{X}} \mid \|g'\|_\mu^2 \le B\}$.*
- *Let $\ell = 0$ and $h^{(\ell)} = 0$*
- *While $\max_{f \in \mathcal{F}} \left\langle g - h^{(\ell)}, f \right\rangle_\mu \ge \delta$:*
  - $\ell = \ell + 1$
  - *Let $f_\ell \in \mathcal{F}$ be such that $\left\langle g - h^{(\ell-1)}, f_\ell \right\rangle_\mu \ge \delta'$* (Correlation Oracle $\mathcal{O}_{\mu,B}$ Step)
  - *Let $c_\ell = \delta'$*
  - $h^{(\ell)} = \Pi(h^{(\ell-1)} + c_\ell \cdot f_\ell)$
- *Let $p = \ell$*
- *return $h = \sum_{\ell=1}^{p} c_\ell \cdot f_\ell$*

We will need the following general fact about projections onto a convex body.

Fact 4.4 (Implicit in Lemma 3.1 of [Bub15]). *Let $\mathcal{Y}$ be a compact convex body in a finite dimensional Hilbert space $\mathcal{V}$ equipped with inner product $\langle \cdot, \cdot \rangle_\nu$ and associated norm $\|\cdot\|_\nu$. Let $\Pi_\mathcal{Y}$ be projector onto $\mathcal{Y}$. Then, for $y \in \mathcal{Y}$ and $x \in \mathcal{V}$, we have*

$$\|y - x\|_\nu^2 \;\geq\; \|y - \Pi_\mathcal{Y}(x)\|_\nu^2 \;+\; \|\Pi_\mathcal{Y}(x) - x\|_\nu^2 .$$

Proof of Lemma 4.2. We will show that the norm of $\left\|g - h^{(\ell)}\right\|_\mu$ strictly decreases as the algorithm progresses. Computing we obtain

$$\begin{aligned}
\left\|g - h^{(\ell)}\right\|_\mu^2 &= \left\|g - \Pi(h^{(\ell-1)} + c_\ell \cdot f_\ell)\right\|_\mu^2 \\
&\leq \left\|g - (h^{(\ell-1)} + c_\ell \cdot f_\ell)\right\|_\mu^2 - \left\|(h^{(\ell-1)} + c_\ell \cdot f_\ell) - \Pi(h^{(\ell-1)} + c_\ell \cdot f_\ell)\right\|_\mu^2 \quad \text{(By Fact 4.4)} \\
&\leq \left\|g - (h^{(\ell-1)} + c_\ell \cdot f_\ell)\right\|_\mu^2 \\
&= \left\|g - h^{(\ell-1)}\right\|_\mu^2 + c_\ell^2 \cdot \|f_\ell\|_\mu^2 - 2c_\ell \cdot \left\langle g - h^{(\ell-1)}, f_\ell \right\rangle_\mu \\
&\leq \left\|g - h^{(\ell-1)}\right\|_\mu^2 - (\delta')^2
\end{aligned}$$

where the inequality follows from $c_\ell = \delta'$, the bound $\|f_\ell\|_\mu \leq 1$ and

$$\left\langle g - h^{(\ell-1)}, f_\ell \right\rangle_\mu \geq \delta'.$$

Since $\|g\|_\mu^2 \leq B$ and $\left\|g - h^{(\ell)}\right\|_\mu^2$ decreases by at least $(\delta')^2$ in each iteration, we conclude that the algorithm halts in at most $p \leq B/(\delta')^2$ steps.

By construction each c_ℓ is initialized to δ' and can not increase (it can only decrease due to projections). Thus, we obtain $\sum_{\ell=1}^p |c_\ell| \leq p \cdot \delta' \leq B/\delta'$. Also by construction at termination $\|h\|_\mu^2 \leq B$. It remains to show that $c_\ell \geq \delta'/(1 + \delta'/\sqrt{B})^p$. Note that the projection $\Pi(h^{(\ell-1)} + c_\ell \cdot f_\ell)$ at each iteration either does nothing to the coefficients c_ℓ's or scales them by a factor of at most $(1 + \delta'/\sqrt{B})$ since $\left\|h^{(\ell-1)}\right\|_\mu + \|c_\ell \cdot f_\ell\|_\mu \leq \sqrt{B}(1 + \delta'/\sqrt{B})$. This readily implies the claimed lower bound on the coefficients c_ℓ's at termination. Moreover, we have $h^{(\ell)} \in \mathcal{H}(B/(\delta')^2, B/\delta', \mathcal{F})$ also by construction.

Running Time: The decomposition algorithm calls the correlation oracle at most $p + 1$ times. Since the coefficients c_ℓ always lie in $[\delta'/(1 + \delta'/\sqrt{B})^p, \delta'] \subseteq [\delta'/\exp(p\delta'/\sqrt{B}), \delta']$, the bit complexity is $C = O(p\delta'/\sqrt{B})$ and computing the projection (which amounts to computing $h^{(\ell)}/\left\|h^{(\ell)}\right\|_\mu$ if $\left\|h^{(\ell)}\right\|_\mu^2 > B$) takes at most $\widetilde{O}(p^2 \cdot \mathrm{poly}(C) \cdot |\mathrm{Supp}(\mu)|)$. Then the total running time is at most

$$\widetilde{O}(p(\mathcal{T}_{\mathcal{O}_{\mu,B}} + p^2 \cdot \mathrm{poly}(C) \cdot |\mathrm{Supp}(\mu)|)) \;=\; \widetilde{O}\left(\mathrm{poly}(B, 1/\delta') \cdot (\mathcal{T}_{\mathcal{O}_{\mu,B}} + |\mathrm{Supp}(\mu)|)\right),$$

concluding the proof. ∎

Remark 4.5. *If we are only interested in an existential version of Lemma 4.2, we can always use a trivial existential (δ, δ)-correlation oracle. However, to obtain weak regularity decompositions efficiently in our settings, we will later use efficient (δ, δ')-correlation oracle with $\delta' = \Omega(\delta)$.*

4.2 Splittable Mixing Lemma

A splittable collection of tuples gives rise to several expanding split operators (see Definition 3.9). This allows us to show that a splittable collection satisfies some higher-order analogues of the well known expander mixing lemmas for graphs (*cf.*,[HLW06][Section 2.4]) as we make precise next.

Lemma 4.6 (Splittable Mixing Lemma)**.** *Suppose $W(k) \subseteq [n]^k$ is a τ-splittable collection of tuples. For every $t \in \{0, \ldots, k-2\}$ and every $f, f' \in \mathcal{F}_{t+1}$, we have*

$$\left| \langle f', f \rangle_{\nu_{t+1}} \; - \; \langle f', f \rangle_{\nu_t} \right| \; \leq \; \tau.$$

Proof. Let $f = f_1 \otimes \cdots \otimes f_t \otimes f_{t+1} \otimes f_{t+2}$ and $f' = f'_1 \otimes \cdots \otimes f'_t \otimes f'_{t+1} \otimes f'_{t+2}$. We have

$$\begin{aligned}\left| \langle f', f \rangle_{\nu_{t+1}} - \langle f', f \rangle_{\nu_t} \right| &= \left| \prod_{i=1}^{t} \mathop{\mathbb{E}}_{\mu_1} f_i f'_i \right| \cdot \left| \mathop{\mathbb{E}}_{\mu_1 \otimes \mu_{[t+2,k]}} f_{t+1} f'_{t+1} \otimes f_{t+2} f'_{t+2} - \mathop{\mathbb{E}}_{\mu_{[t+1,k]}} f_{t+1} f'_{t+1} \otimes f_{t+2} f'_{t+2} \right| \\ &\leq \left| \mathop{\mathbb{E}}_{\mu_1 \otimes \mu_{[t+2,k]}} f_{t+1} f'_{t+1} \otimes f_{t+2} f'_{t+2} - \mathop{\mathbb{E}}_{\mu_{[t+1,k]}} f_{t+1} f'_{t+1} \otimes f_{t+2} f'_{t+2} \right| .\end{aligned}$$

Let $f''_{t+1} = f_{t+1} f'_{t+1}$ and $f''_{t+2} = f_{t+2} f'_{t+2}$. Note that

$$\mathop{\mathbb{E}}_{\mu_1 \otimes \mu_{[t+2,k]}} f''_{t+1} \otimes f''_{t+2} - \mathop{\mathbb{E}}_{\mu_{[t+1,k]}} f''_{t+1} \otimes f''_{t+2} = \left\langle f''_{t+1}, \left(\frac{\mathsf{J}_{\mathrm{rec}}}{|W[t+2,k]|} - \mathsf{S}_{W[t+1],W[t+2,k]} \right) f''_{t+2} \right\rangle_{\mu_1},$$

where $\mathsf{J}_{\mathrm{rec}}$ is the (rectangular) $|W[t+1]| \times |W[t+2,k]|$ all ones matrix. Using the τ-splittability assumption, we have the following bound on the largest singular value

$$\sigma \left(\frac{\mathsf{J}_{\mathrm{rec}}}{|W[t+2,k]|} - \mathsf{S}_{W[t+1],W[t+2,k]} \right) \leq \sigma_2 \left(\mathsf{S}_{W[t+1],W[t+2,k]} \right) \leq \tau.$$

Then

$$\left| \mathop{\mathbb{E}}_{\mu_1 \otimes \mu_{[t+2,k]}} f_{t+1} f'_{t+1} \otimes f_{t+2} f'_{t+2} - \mathop{\mathbb{E}}_{\mu_{[t+1,k]}} f_{t+1} f'_{t+1} \otimes f_{t+2} f'_{t+2} \right| \leq \tau,$$

concluding the proof. ∎

We can iterate the preceding lemma to obtain the following.

Lemma 4.7 (Splittable Mixing Lemma Iterated)**.** *Suppose $W(k) \subseteq [n]^k$ is a τ-splittable collection of tuples. For every $f = f_1 \otimes \cdots \otimes f_k \in \mathcal{F}_{k-1}$, we have*

$$\left| \mathop{\mathbb{E}}_{\nu_0} f \; - \; \mathop{\mathbb{E}}_{\nu_{k-1}} f \right| \; \leq \; (k-1) \cdot \tau.$$

Proof. Let $1 \in \mathcal{F}_{k-1}$ be the constant 1 function. Note that for any $t \in \{0, \ldots, k-1\}$ the restriction of any $f' \in \mathcal{F}_{k-1}$ to the support of ν_t which we denote by $f'|_t$ belongs to $\mathcal{F}_t$. It is immediate that $\langle f, 1\rangle_{\nu_t} = \langle f|_t, 1\rangle_{\nu_t}$. Computing we obtain

$$\begin{aligned}
\left|\mathop{\mathbb{E}}_{\nu_0} f - \mathop{\mathbb{E}}_{\nu_{k-1}} f\right| = \left|\langle f, 1\rangle_{\nu_0} - \langle f, 1\rangle_{\nu_{k-1}}\right| &\le \sum_{i=0}^{k-2} \left|\langle f, 1\rangle_{\nu_i} - \langle f, 1\rangle_{\nu_{i+1}}\right| \\
&= \sum_{i=0}^{k-2} \left|\langle f|_t, 1|_t\rangle_{\nu_i} - \langle f|_{t+1}, 1|_{t+1}\rangle_{\nu_{i+1}}\right| \\
&\le \sum_{i=0}^{k-2} \tau, && \text{(By Lemma 4.6)}
\end{aligned}$$

finishing the proof. ■

In Section 4.4, we will need two corollaries of the splittable mixing lemma which we prove now.

Claim 4.8. *Let $W(k) \subseteq [n]^k$ be a τ-splittable collection of tuples. Let $t \in \{0, \ldots, k-2\}$ and $h_{t+1} \in \mathcal{H}(R_0, R_1, \mathcal{F}_{t+1})$. For every $f \in \mathcal{F}_{t+1}$, we have*

$$\left|\langle h_{t+1}, f\rangle_{\nu_{t+1}} - \langle h_{t+1}, f\rangle_{\nu_t}\right| \le \tau \cdot R_1.$$

Proof. Since $h_{t+1} \in \mathcal{H}(R_0, R_1, \mathcal{F}_{t+1})$, we can write $h_{t+1} = \sum_\ell c_\ell \cdot f_\ell$, where $f_\ell \in \mathcal{F}_{t+1}$ and $\sum_\ell |c_\ell| \le R_1$. By the splittable mixing lemma, *cf.*, Lemma 4.6, we have

$$\left|\langle h_{t+1}, f\rangle_{\nu_{t+1}} - \langle h_{t+1}, f\rangle_{\nu_t}\right| \le \sum_\ell |c_\ell| \cdot \left|\langle f_\ell, f\rangle_{\nu_{t+1}} - \langle f_\ell, f\rangle_{\nu_t}\right| \le \tau \cdot R_1. \qquad ■$$

Claim 4.9. *Let $W(k) \subseteq [n]^k$ be a τ-splittable collection of tuples. Let $t \in \{0, \ldots, k-2\}$ and $h_{t+1} \in \mathcal{H}(R_0, R_1, \mathcal{F}_{t+1})$. Then*

$$\left|\|h_{t+1}\|^2_{\nu_{t+1}} - \|h_{t+1}\|^2_{\nu_t}\right| \le \tau \cdot R_1^2.$$

Proof. Since $h_{t+1} \in \mathcal{H}(R_0, R_1, \mathcal{F}_{t+1})$, we can write $h_{t+1} = \sum_\ell c_\ell \cdot f_\ell$, where $f_\ell \in \mathcal{F}_{t+1}$ and $\sum_\ell |c_\ell| \le R_1$. By the splittable mixing lemma, *cf.*, Lemma 4.6, we have

$$\left|\langle h_{t+1}, h_{t+1}\rangle_{\nu_{t+1}} - \langle h_{t+1}, h_{t+1}\rangle_{\nu_t}\right| \le \sum_{\ell,\ell'} |c_\ell| \cdot |c_{\ell'}| \cdot \left|\langle f_\ell, f_{\ell'}\rangle_{\nu_{t+1}} - \langle f_\ell, f_{\ell'}\rangle_{\nu_t}\right| \le \tau \cdot R_1^2. \qquad ■$$

4.3 Existential Weak Regularity Decomposition

Using the abstract weak regularity lemma, Lemma 4.2, together splittable mixing lemmas of Section 4.2, we can obtain (non-constructive) existential weak regularity decompositions for splittable structures.

Lemma 4.10 (Existential Weak Regularity for Splittable Tuples)**.** *Let $W(k) \subseteq [n]^k$ be a τ-splittable structure. Let $g \in \mathbb{R}^{W[1]^k}$ be supported on $W(k)$ with $\|g\|_{\mu_k} \le 1$. Let $\mathcal{F} = \mathcal{F}_{k-1}$ (cf., Definition 3.19) be arbitrary. For every $\delta > 0$, if $\tau \le O(\delta^2/(k-1))$, then there exists $h \in \mathbb{R}^{W[1]^k}$ supported on $O(1/\delta^2)$ functions in $\mathcal{F}$ such that*

$$\max_{f \in \mathcal{F}} \langle g - h, f\rangle \le \delta \cdot |W(k)|,$$

where the inner product is over the counting measure on $W[1]^k$.

Proof. Apply the weak regularity Lemma 4.2, with parameters δ and δ' equal to δ, collection $\mathcal{F}$, input function g, measure $\mu = \mu_k$ (i.e., uniform measure on $W(k)$) and a non-explicit correlation oracle based on the existential guarantee. This yields $h = \sum_{\ell=1}^{p} c_\ell \cdot f_\ell \in \mathcal{H}(1/\delta^2, 1/\delta, \mathcal{F})$ where

$$\max_{f \in \mathcal{F}} \quad \langle g - h, f \rangle_{\mu_k} \leq \delta.$$

Let $f \in \mathcal{F}$. We claim that $h' = h \cdot |W(k)| \,/\, |W[1]|^k$ satisfies the conclusion of the current lemma. For this, we bound

$$\left| |W(k)| \langle g - h, f \rangle_{\mu_k} - \langle g - h', f \rangle \right| \leq \left| |W(k)| \langle g, f \rangle_{\mu_k} - \langle g, f \rangle \right| + \sum_{\ell=1}^{p} |c_\ell| \cdot \left| |W(k)| \langle f_\ell, f \rangle_{\mu_k} - \frac{|W(k)|}{|W[1]|^k} \langle f_\ell, f \rangle \right|.$$

The first term in the RHS above is zero since

$$|W(k)| \langle g, f \rangle_{\mu_k} = \sum_{\mathfrak{s} \in W(k)} g(\mathfrak{s}) \cdot f(\mathfrak{s}) = \langle g, f \rangle,$$

where in the second equality we used that g is supported on $W(k)$. Suppose that $f = f_1 \otimes \cdots \otimes f_k$ and $f_\ell = f_{\ell,1} \otimes \cdots \otimes f_{\ell,k}$. Set $f'_\ell = (f_1 \cdot f_{\ell,1}) \otimes \cdots \otimes (f_k \cdot f_{k,1})$ where $(f_j \cdot f_{j,1})$ is the pointwise product of f_j and $f_{j,1}$. Note that

$$\langle f_\ell, f \rangle_{\mu_k} = \mathop{\mathbb{E}}_{\nu_0} [f'_\ell] \quad \text{and} \quad \frac{\langle f_\ell, f \rangle}{|W[1]|^k} = \mathop{\mathbb{E}}_{\nu_{k-1}} [f'_\ell],$$

where we recall that μ_k is equal to ν_0 and $\mu_1^{\otimes k}$ is equal to ν_{k-1}. Moreover, f'_ℓ is the tensor product of k functions in $\mathbb{R}^{X[1]}$ of ℓ_∞-norm at most 1. By the splittable mixing lemma (*cf.*, Lemma 4.7), we have

$$\left| \mathop{\mathbb{E}}_{\nu_0} [f'_\ell] - \mathop{\mathbb{E}}_{\nu_{k-1}} [f'_\ell] \right| \leq (k-1) \cdot \tau.$$

Hence, we obtain

$$\begin{aligned} \left| |W(k)| \langle g - h, f \rangle_{\mu_k} - \langle g - h', f \rangle \right| &\leq \sum_{\ell=1}^{p} |c_\ell| \cdot |W(k)| \cdot \left| \mathop{\mathbb{E}}_{\nu_0} [f'_\ell] - \mathop{\mathbb{E}}_{\nu_{k-1}} [f'_\ell] \right| \\ &\leq \sum_{\ell=1}^{p} |c_\ell| \cdot (k-1) \cdot \tau \cdot |W(k)| \leq \delta \cdot |W(k)|, \end{aligned}$$

from which the lemma readily follows. ∎

4.4 Efficient Weak Regularity Decomposition

The goal of this section is to prove an efficient version of weak regularity that can be computed in near-linear time. We obtain parameters somewhat comparable to those parameters of the existential weak regularity in Lemma 4.10 above with a mild polynomial factor loss of $\Theta(1/k^2)$ on the splittability requirement.

Theorem 4.11. *[Efficient Weak Regularity] Let $W(k) \subseteq [n]^k$ be a τ-splittable collection of tuples. Let $g \in \mathbb{R}^{W[1]^k}$ be supported on $W(k)$ with $\|g\|_{\mu_k} \leq 1$. Suppose $\mathcal{F}$ is either $\mathrm{CUT}^{\otimes k}$ or $\mathrm{CUT}^{\otimes k}_{\pm}$. For every $\delta > 0$, if $\tau \leq \delta^2/(k^3 \cdot 2^{20})$, then we can find $h = \sum_{\ell=1}^{p} c_\ell \cdot f_\ell$ with $p = O(k^2/\delta^2)$, $c_1, \ldots, c_p \in \mathbb{R}$ and functions $f_1, \ldots, f_p \in \mathcal{F}$, such that $\|h\|_{\mu_1^{\otimes k}} \leq 2$ and h is a good approximator to g in the following sense*

$$\max_{f \in \mathcal{F}} \quad \left\langle g - \left(\frac{d}{n}\right)^{k-1} h, f \right\rangle \;\leq\; \delta \cdot |W(k)| \,,$$

where the inner product is over the counting measure on $W[1]^k$. Furthermore, h can be found in $\widetilde{O}(2^{2^{\widetilde{O}(k^2/\delta^2)}} \cdot |W(k)|)$ time.

Warm-up: We first sketch a simpler naive algorithmic weak regularity decompoistion for $\mathrm{CUT}^{\otimes k}$ whose parameters are much worse than the existential parameters of Lemma 4.10, but it can be computed in near-linear time. The fast accumulation of errors will explain our motivation in designing the efficient algorithm underlying Theorem 4.11. The reader only interested in the latter is welcome to skip ahead.

Lemma 4.12 (Naive Efficient Weak Regularity)**.** *Let $W' \subseteq W(k)$ where $W(k)$ is τ-splittable. Let $\mathcal{F}$ be either $\mathrm{CUT}^{\otimes k}$ or $\mathrm{CUT}^{\otimes k}_{\pm}$. For every $\delta > 0$, if $\tau \leq (O(\delta))^{2^k}$, then we can find h supported on $(O(1/\delta))^{2^k}$ functions of $\mathcal{F}$ such that*

$$\max_{f \in \mathcal{F}} \quad \langle \mathbf{1}_{W'} - h, f \rangle \;\leq\; (k-1) \cdot \delta \cdot |W(k)| \,,$$

where the inner product is over the counting measure on $W[1]^k$. Furthermore, this can be done in time $\tilde{O}_\delta(|W(k)|)$.

Proof Sketch: In this sketch, our goal is to show the fast accumulation of errors when applying the weak regularity decomposition for matrices. For simplicity, we assume that this can be done in near-linear time on the number of non-zero entries of the matrix. Precise details and much better parameters are given in the proof of Theorem 4.11.

Applying the matrix regularity decomposition to $\mathbf{1}_{W'}$, viewed a matrix in $\mathbb{R}^{W[1,k-1] \times W[k]}$ supported on $W[1,k]$, with accuracy parameter $\delta_1 > 0$, we get in $\tilde{O}_{\delta_1}(|W[1,k]|)$ time

$$\left\| \mathbf{1}_{W'} \;-\; \frac{d}{n} \sum_{\ell_1=1}^{p_1} c_{\ell_1} \cdot \mathbf{1}_{S_{\ell_1}} \otimes \mathbf{1}_{T_{\ell_1}} \right\|_{\square} \;\leq\; \delta_1 \cdot |W[1,k]| \,,$$

where $p_1 = O(1/\delta_1^2)$ and $\sum_{\ell_1} |c_{\ell_1}| \leq O(1/\delta_1)$.

In turn, for each $\mathbf{1}_{S_{\ell_1}}$ viewed a matrix in $\mathbb{R}^{W[1,k-2] \times W[k-1]}$ supported on $W[1,k-1]$, we apply the matrix regularity decomposition with accuracy parameter $\delta_2 > 0$ getting in $\tilde{O}_{\delta_2}(|W[1,k-1]|)$ time

$$\left\| \mathbf{1}_{S_{\ell_1}} \;-\; \frac{d}{n} \sum_{\ell_2=1}^{p_2} c_{\ell_2,\ell_1} \cdot \mathbf{1}_{S_{\ell_2,\ell_1}} \otimes \mathbf{1}_{T_{\ell_2,\ell_1}} \right\|_{\square} \;\leq\; \delta_2 \cdot |W[1,k-1]| \,,$$

where $p_2 = O(1/\delta_2^2)$ and $\sum_{\ell_2} |c_{\ell_2,\ell_1}| \le O(1/\delta_2)$. Continuing this process inductively with accuracy parameters $\delta_3, \ldots, \delta_{k-1}$, we obtain

$$h := \left(\frac{d}{n}\right)^{k-1} \sum_{\ell_1}^{p_1} \cdots \sum_{\ell_{k-1}=1}^{p_{k-1}} c_{\ell_1} \ldots c_{\ell_1,\ldots,\ell_{k-1}} \cdot \mathbf{1}_{T_{\ell_1,\ldots,\ell_{k-1}}} \otimes \cdots \otimes \mathbf{1}_{T_{\ell_1}},$$

in time $\widetilde{O}_{\delta_1,\ldots,\delta_{k-1}}(|W(k)|)$. We show that h is close in k-tensor cut norm (*cf.*, Definition 3.18) to $\mathbf{1}_{W'}$. Computing we have

$$\begin{aligned}
&\left\|\mathbf{1}_{W'} - h\right\|_{\square^{\otimes k}} \le \\
&\sum_{j=0}^{k-2} \sum_{\ell_1=1}^{p_1} \cdots \sum_{\ell_j=1}^{p_j} \left|c_{\ell_1} \ldots c_{\ell_1,\ldots,\ell_j}\right| \cdot \\
&\qquad \left\| \mathbf{1}_{S_{\ell_1,\ldots,\ell_j}} - \left(\frac{d}{n}\right)^{k-j-1} \sum_{\ell_{j+1}=1}^{p_{j+1}} c_{\ell_1,\ldots,\ell_{j+1}} \cdot \mathbf{1}_{S_{\ell_1,\ldots,\ell_{j+1}}} \otimes \mathbf{1}_{T_{\ell_1,\ldots,\ell_{j+1}}} \right\|_{\square^{\otimes k-j}} \cdot \\
&\qquad \left(\frac{d}{n}\right)^{j} \cdot \left\| \mathbf{1}_{T_{\ell_1,\ldots,\ell_j}} \otimes \cdots \otimes \mathbf{1}_{T_{\ell_1}} \right\|_{\square^{\otimes j}} \\
\le\ & \sum_{j=0}^{k-2} \sum_{\ell_1=1}^{p_1} \cdots \sum_{\ell_j=1}^{p_j} d^j \cdot \left|c_{\ell_1} \ldots c_{\ell_1,\ldots,\ell_j}\right| \cdot \\
&\qquad \left\| \mathbf{1}_{S_{\ell_1,\ldots,\ell_j}} - \left(\frac{d}{n}\right)^{k-j-1} \sum_{\ell_{j+1}=1}^{p} c_{\ell_1,\ldots,\ell_{j+1}} \cdot \mathbf{1}_{S_{\ell_1,\ldots,\ell_{j+1}}} \otimes \mathbf{1}_{T_{\ell_1,\ldots,\ell_{j+1}}} \right\|_{\square} \\
\le\ & \sum_{j=0}^{k-2} \sum_{\ell_1=1}^{p_1} \cdots \sum_{\ell_j=1}^{p_j} d^j \cdot \left|c_{\ell_1} \ldots c_{\ell_1,\ldots,\ell_j}\right| \cdot \delta_{j+1} \cdot |W[1, k-j]| \\
\le\ & |W(k)| \sum_{j=0}^{k-2} \delta_{j+1} \prod_{\ell=1}^{j} O(1/\delta_\ell).
\end{aligned}$$

By setting $\delta_j = \Theta(\delta^{2^j})$, the LHS becomes at most $(k-1) \cdot \delta \cdot |W(k)|$. $\square$

We now proceed to prove our main result in this section, namely Theorem 4.11. Fist, we establish some extra notation now. Let $W(k)$ be a d-regular collection of tuples. Most of our derivations which are existential hold for a generic $\mathcal{F}_t$ (*cf.*, Definition 3.19). However, we only derive near-linear time algorithmic results when $\mathcal{F}_t$ is either the CUT functions

$$\mathcal{F}_t^{0/1} := \left\{ \pm \mathbf{1}_{S_1} \otimes \cdots \otimes \mathbf{1}_{S_t} \otimes \mathbf{1}_T \mid S_j \subseteq W[1], T \subseteq W[t+1, k] \right\},$$

or "signed" CUT functions

$$\mathcal{F}_t^{\pm 1} := \left\{ \pm \chi_{S_1} \otimes \cdots \otimes \chi_{S_t} \otimes \chi_T \mid S_j \subseteq W[1], T \subseteq W[t+1, k] \right\},$$

where above we recall that for $S \subseteq [n]$, we have $\chi_S(i) = (-1)^{\mathbf{1}_{i \in S}}$ for $i \in [n]$. Observe that the condition $S_j \subseteq W[1]$ is equivalent to $S_j \subseteq W[i]$ since $W(k)$ is d-regular.

For quick reference, we collect the notation needed in our algorithmic weak regularity decomposition in the following table.

$$\mathcal{F}_t := \left\{ \pm f_1 \otimes \cdots \otimes f_t \otimes f_{t+1} \mid f_j \subseteq \mathbb{R}^{W[1]} \text{ for } i \leq t, f_{t+1} \subseteq \mathbb{R}^{W[t+1,k]}, \|f_j\|_\infty \leq 1 \right\}$$

$$\mathcal{F}_t^{0/1} := \left\{ \pm \mathbf{1}_{S_1} \otimes \cdots \otimes \mathbf{1}_{S_t} \otimes \mathbf{1}_T \mid S_j \subseteq W[1], T \subseteq W[t+1,k] \right\} \subseteq \mathcal{F}_t$$

$$\mathcal{F}_t^{\pm 1} := \left\{ \pm \chi_{S_1} \otimes \cdots \otimes \chi_{S_t} \otimes \chi_T \mid S_j \subseteq W[1], T \subseteq W[t+1,k] \right\} \subseteq \mathcal{F}_t$$

$$\mathcal{H}(R_0, R_1, \mathcal{F}) := \left\{ \sum_{\ell=1}^p c_\ell \cdot f_\ell \mid p \leq R_0, \sum |c_\ell| \leq R_1, f_\ell \in \mathcal{F} \right\}$$

μ_1 is the uniform distribution on $W[1]$ and $\mu_{[t+1,k]}$ is the uniform distribution on $W[t+1,k]$

$$\nu_t := (\mu_1)^{\otimes t} \otimes \left(\mu_{[t+1,k]} \right)$$

Our main result of this section, namely, the near-linear time weak regularity decomposition Theorem 4.11, can be readily deduced from Lemma 4.13 below.

Lemma 4.13 (Efficient Weak Regularity Induction). *Let $W(k) \subseteq [n]^k$ be a τ-splittable d-regular collection of tuples. Let $g \in \mathcal{F}_0$ and $t \in \{0, \ldots, k-1\}$ with $\|g\|_{\mu_k} \leq 1$. For every $\delta > 0$, if $\tau \leq \delta^2/(k \cdot 2^{18})$, then there exists $h_t \in \mathcal{H}(O(1/\delta^2), 2^8(1+1/k)^t/\delta, \mathcal{F}_t)$ with $\|h_t\|_{\nu_t}^2 \leq (1+1/k)^t$ such that*

$$\max_{f \in \mathcal{F}_t} \quad \left\langle g - \left(\frac{d}{n}\right)^t h_t, f \right\rangle_{\nu_t} \leq 2 \cdot \left(\frac{d}{n}\right)^t \cdot t \cdot \delta.$$

Furthermore, the function h_t can be found in $\widetilde{O}((2t)^{2^{O(1/\delta^2)}} \cdot |W(k)|)$ time.

We restate Theorem 4.11 below and then prove it assuming Lemma 4.13.

Theorem 4.11. *[Efficient Weak Regularity] Let $W(k) \subseteq [n]^k$ be a τ-splittable collection of tuples. Let $g \in \mathbb{R}^{W[1]^k}$ be supported on $W(k)$ with $\|g\|_{\mu_k} \leq 1$. Suppose $\mathcal{F}$ is either $\mathrm{CUT}^{\otimes k}$ or $\mathrm{CUT}_{\pm}^{\otimes k}$. For every $\delta > 0$, if $\tau \leq \delta^2/(k^3 \cdot 2^{20})$, then we can find $h = \sum_{\ell=1}^p c_\ell \cdot f_\ell$ with $p = O(k^2/\delta^2)$, $c_1, \ldots, c_p \in \mathbb{R}$ and functions $f_1, \ldots, f_p \in \mathcal{F}$, such that $\|h\|_{\mu_1^{\otimes k}} \leq 2$ and h is a good approximator to g in the following sense*

$$\max_{f \in \mathcal{F}} \quad \left\langle g - \left(\frac{d}{n}\right)^{k-1} h, f \right\rangle \leq \delta \cdot |W(k)|,$$

where the inner product is over the counting measure on $W[1]^k$. Furthermore, h can be found in $\widetilde{O}(2^{2^{\widetilde{O}(k^2/\delta^2)}} \cdot |W(k)|)$ time.

Proof. Set $\mathcal{F}_t = \mathcal{F}_t^{0/1}$ if $\mathcal{F} = \mathrm{CUT}^{\otimes k}$ or set $\mathcal{F}_t = \mathcal{F}_t^{\pm 1}$ if $\mathcal{F} = \mathrm{CUT}_{\pm}^{\otimes k}$. We apply Lemma 4.13 with $t = k-1$, accuracy δ as $\delta/(2k)$ and input function g. This gives $h_t = \sum_{\ell=1}^p c'_\ell \cdot f_\ell \in \mathcal{H}(O(k^2/\delta^2), O(k/\delta), \mathcal{F}_t)$ such that

$$\max_{f \in \mathcal{F}_t} \quad \left\langle g - \left(\frac{d}{n}\right)^t h_t, f \right\rangle_{\nu_t} \leq 2 \cdot \left(\frac{d}{n}\right)^t \cdot t \cdot \delta. \tag{1}$$

Note that $\nu_t = \nu_{k-1} = \mu_1^{\otimes k}$ is the uniform measure on $W[1]^k$. Since $W(k)$ is d-regular, $|W(k)| = |W[1]|^k \cdot (d/n)^{k-1}$. Set $h = \cdot h_t$. Then the guarantee in Eq. (1) becomes

$$\max_{f \in \mathcal{F}} \left\langle g - \left(\frac{d}{n}\right)^{k-1} h, f \right\rangle \leq \delta \cdot |W(k)|,$$

where the inner product is under the counting measure. By Lemma 4.13, we have $\|h_t\|_{\nu_t}^2 \leq (1+1/k)^t \leq e$, so $\|h_t\|_{\nu_t} \leq 2$. Then $\|h\|_{\mu_1^{\otimes k}} \leq 2$. The running time follows from Lemma 4.13 completing the proof. ∎

We now prove Lemma 4.13 above assuming the following algorithmic result which we prove later.

Lemma 4.14. *[Algorithmic Weak Regularity Step] Let $\delta > 0$ and $t \in \{0, \ldots, k-2\}$. Let $h_t \in \mathcal{H}(O(B/\delta^2), O(B/\delta), \mathcal{F}_t)$ with $\|h_t\|_{\nu_t}^2 \leq B$. Then there exists $h_{t+1} \in \mathcal{H}(O(B/\delta^2), 2^8 B/\delta, \mathcal{F}_{t+1})$ with $\|h_{t+1}\|_{\nu_t}^2 \leq B$ such that*

$$\max_{f \in \mathcal{F}_{t+1}} \langle h_t - h_{t+1}, f \rangle_{\nu_t} \leq \delta.$$

Furthermore, each h_{t+1} can be found in time $\widetilde{O}((2t)^{2^{O(1/\delta^2)}} \cdot |W(k)|)$.

Proof of Lemma 4.13. We will prove the lemma with the following simple equivalent conclusion

$$\left\langle g - \left(\frac{d}{n}\right)^t h_t, f \right\rangle_{\nu_t} \leq 2 \cdot \left(\frac{d}{n}\right)^t \cdot t \cdot \delta \qquad \Leftrightarrow \qquad \left\langle \left(\frac{n}{d}\right)^t g - h_t, f \right\rangle_{\nu_t} \leq 2 \cdot t \cdot \delta,$$

which we will prove holds for every $f \in \mathcal{F}_t$. The base case $t = 0$ follows immediately by setting $h_0 = g$. Let $t \in \{0, \ldots, k-2\}$. Since $h_t \in \mathcal{H}(O(1/\delta^2), 2^8(1+1/k)^t/\delta, \mathcal{F}_t)$, invoking Lemma 4.14 with accuracy parameter δ and input function h_t, we obtain $h_{t+1} \in \mathcal{H}(O(1/\delta^2), 2^8(1+1/k)^{t+1}/\delta, \mathcal{F}_{t+1})$ satisfying

$$\max_{f \in \mathcal{F}_{t+1}} \langle h_t - h_{t+1}, f \rangle_{\nu_t} \leq \delta. \tag{2}$$

Let $f \in \mathcal{F}_{t+1}$. We will show that h_{t+1} satisfies the conclusion of the lemma. Expanding we have

$$\begin{aligned} \left\langle \left(\frac{n}{d}\right)^{t+1} g - h_{t+1}, f \right\rangle_{\nu_{t+1}} &= \underbrace{\left\langle \left(\frac{n}{d}\right)^t g - h_t, f \right\rangle_{\nu_t}}_{(i)} + \left(\frac{n}{d}\right)^t \cdot \underbrace{\left(\frac{n}{d} \langle g, f \rangle_{\nu_{t+1}} - \langle g, f \rangle_{\nu_t}\right)}_{(ii)} \\ &\quad + \underbrace{\langle h_t - h_{t+1}, f \rangle_{\nu_t}}_{(iii)} + \underbrace{\langle h_{t+1}, f \rangle_{\nu_t} - \langle h_{t+1}, f \rangle_{\nu_{t+1}}}_{(iv)}. \end{aligned}$$

We will bound each of the terms in RHS above.

Term (i): Suppose $f = f_1 \otimes \cdots \otimes f_{t+1} \otimes f_{t+2} \in \mathcal{F}_{t+1}$. Let $f' = f_1 \otimes \cdots \otimes f_t \otimes f'_{t+1}$, where $f'_{t+1} = (f_{t+1} \otimes f_{t+2})|_{W[t+2,k]}$, so that $f' \in \mathcal{F}_t$. Using the induction hypothesis, we have

$$\left\langle \left(\frac{n}{d}\right)^t g - h_t, f \right\rangle_{\nu_t} = \left\langle \left(\frac{n}{d}\right)^t g - h_t, f' \right\rangle_{\nu_t} \leq 2 \cdot t \cdot \delta.$$

Term (ii): Since $g \in \mathcal{F}_0$, it is supported on $W(k)$ and so we have

$$\begin{aligned}\langle g, f\rangle_{\nu_t} &= \frac{1}{|W[1]|^t\,|W[t+1,k]|}\sum_{\mathfrak{s}\in W(k)} g(\mathfrak{s})\cdot f(\mathfrak{s})\\ &= \frac{n}{d}\cdot\frac{1}{|W[1]|^{t+1}\,|W[t+2,k]|}\sum_{\mathfrak{s}\in W(k)} g(\mathfrak{s})\cdot f(\mathfrak{s}) \;=\; \frac{n}{d}\cdot\langle g, f\rangle_{\nu_{t+1}}.\end{aligned}$$

where the second equality follows from $|W[t+1,k]| = d\cdot|W[t+2,k]|$ by the d-regular assumption.

Term (iii): By Eq. (2), we have $\langle h_t - h_{t+1}, f\rangle_{\nu_t} \le \delta$.

Term (iv): For notional convenience, set $R_1 = 2^8(1+1/k)^{t+1}/\delta$. Since $h_{t+1}\in\mathcal{H}(\infty, R_1, \mathcal{F}_{t+1})$ and the splittability parameter τ satisfies $\tau \le \delta^2/(k\cdot 2^{18})$, from Claim 4.8 we obtain

$$\langle h_{t+1}, f\rangle_{\nu_t} - \langle h_{t+1}, f\rangle_{\nu_{t+1}} \;\le\; \tau\cdot R_1 \;\le\; \delta.$$

Putting everything together yields

$$\left\langle \left(\frac{n}{d}\right)^{t+1} g - h_t, f\right\rangle_{\nu_{t+1}} \;\le\; \underbrace{2\cdot t\cdot\delta}_{(i)} + \left(\frac{n}{d}\right)^t\cdot\underbrace{0}_{(ii)} + \underbrace{\delta}_{(iii)} + \underbrace{\delta}_{(iv)} \;\le\; 2\cdot(t+1)\cdot\delta,$$

concluding the claimed inequality.

Now we use the bound $\|h_{t+1}\|_{\nu_t}^2 \le \|h_t\|_{\nu_t}^2$ from Lemma 4.14 together with the splittability assumption $\tau \le \delta^2/(k\cdot 2^{18})$ to bound the norm $\|h_{t+1}\|_{\nu_{t+1}}^2$ under the new measure ν_{t+1}. Under these assumptions and using Claim 4.9 we get

$$\begin{aligned}\left|\|h_{t+1}\|_{\nu_{t+1}}^2 - \|h_{t+1}\|_{\nu_t}^2\right| &\le \tau\cdot R_1^2 \le \frac{\delta^2}{k\cdot 2^{18}}\cdot\frac{2^{16}(1+1/k)^{2(t+1)}}{\delta^2}\\ &\le \frac{(1+1/k)^t}{k}.\end{aligned}$$

where we used the bounds on τ, R_1 and $(1+1/k)^{(t+2)} \le 4$ for $0 \le t \le k-2$. From the previous inequality and the induction hypothesis $\|h_t\|_{\nu_t}^2 \le (1+1/k)^t$, we finally get $\|h_{t+1}\|_{\nu_{t+1}}^2 \le (1+1/k)^{t+1}$ as desired. ■

We now show a near-linear time weak regularity decomposition for special functions of the form $h_t \in \mathcal{H}(O(1/\delta^2), O(1/\delta), \mathcal{F}_t)$ that admit a tensor product structure. The goal is to design a correlation oracle that exploits the special tensor product structure of the function $(h_t - h_{t+1}^{(\ell)})$, where $h_{t+1}^{(\ell)}$ is the ℓth approximator of h_t in the abstract weak regularity algorithm (*cf.*, Algorithm 4.3).

Lemma 4.14. *[Algorithmic Weak Regularity Step] Let $\delta > 0$ and $t \in \{0,\ldots,k-2\}$. Let $h_t \in \mathcal{H}(O(B/\delta^2), O(B/\delta), \mathcal{F}_t)$ with $\|h_t\|_{\nu_t}^2 \le B$. Then there exists $h_{t+1} \in \mathcal{H}(O(B/\delta^2), 2^8 B/\delta, \mathcal{F}_{t+1})$ with $\|h_{t+1}\|_{\nu_t}^2 \le B$ such that*

$$\max_{f\in\mathcal{F}_{t+1}} \quad \langle h_t - h_{t+1}, f\rangle_{\nu_t} \;\le\; \delta.$$

Furthermore, each h_{t+1} can be found in time $\widetilde{O}((2t)^{2^{O(1/\delta^2)}}\cdot|W(k)|)$.

Our correlation oracle for higher-order tensors will make calls to a correlation oracle for matrices Theorem 4.15 (i.e., 2-tensors) stated below. This matrix oracle is presented in Section 4.5 and it follows from a simple combination of a matrix cut norm approximation algorithm by Alon and Naor [AN04] with known fast SDP solvers for sparse matrices such as those by Lee and Padmanabhan [LP20] and Arora and Kale [AK07].

Theorem 4.15. *[Alon–Naor Correlation Oracle] Let $\mathcal{F}$ be either $\mathrm{CUT}^{\otimes 2}$ or $\mathrm{CUT}_{\pm}^{\otimes 2}$ and μ be the uniform measure supported on at most m elements of $[n'] \times [n']$. There exists an algorithmic $(\delta, \alpha_{\mathbf{AN}} \cdot \delta)$-correlation oracle $\mathcal{O}_{\mu,B}$ running in time $\mathcal{T}_{\mathcal{O}_{\mu,B}} = \tilde{O}\left(\mathrm{poly}(B/\delta) \cdot (m + n')\right)$, where $\alpha_{\mathbf{AN}} \geq 1/2^4$ is an approximation ratio constant.*

Proof. We will apply the abstract weak regularity lemma, *cf.*,Lemma 4.2, with $\mathcal{F} = \mathcal{F}_{t+1}$, δ, $\delta' = \delta/2^8$ and $\mu = \nu_t$. This will result in a function from $\mathcal{H}(O(B/\delta^2), 2^8 B/\delta, \mathcal{F}_{t+1})$.

Correlation oracle task: To make this application take near-linear time, we need to specify a correlation oracle $\mathcal{O}_{\nu_t} = \mathcal{O}_{\nu_t, O(1)}$ and now we take advantage of the special tensor structure in our setting. We want an oracle that given

$$h_t = \sum_{\ell=1}^{p} c_\ell \cdot g_\ell, \quad g_\ell \in \mathcal{F}_t, \qquad g_\ell = g_{\ell,1} \otimes \cdots \otimes g_{\ell,t} \otimes \underbrace{g_{\ell,t+1}}_{\in \mathbb{R}^{W[t+1,k]}} \text{ and}$$

$$h_{t+1} = \sum_{\ell=1}^{p} c'_\ell \cdot g'_\ell, \quad g'_\ell \in \mathcal{F}_{t+1}, \quad g'_\ell = g'_{\ell,1} \otimes \cdots \otimes g'_{\ell,t} \otimes \underbrace{g'_{\ell,t+1}}_{\in \mathbb{R}^{W[1]}} \otimes \underbrace{g'_{\ell,t+2}}_{\in \mathbb{R}^{W[t+2,k]}},$$

if there exists

$$f = f_1 \otimes \cdots \otimes f_t \otimes \underbrace{f_{t+1}}_{\in \mathbb{R}^{W[1]}} \otimes \underbrace{f_{t+2}}_{\in \mathbb{R}^{W[t+2,k]}} \in \mathcal{F}_{t+1}$$

satisfying

$$\langle h_t - h_{t+1}, f \rangle_{\nu_t} \geq \delta,$$

for some $f \in \mathcal{F}_{t+1}$, finds $f' \in \mathcal{F}_{t+1}$ in near-linear time such that

$$\left\langle h_t - h_{t+1}, f' \right\rangle_{\nu_t} \geq \delta' = \frac{\delta}{2^8}.$$

Here, h_{t+1} is the current approximator of h_t in the abstract weak regularity algorithm and, by Lemma 4.2, $h_{t+1} \in \mathcal{H}(O(1/\delta^2), 2^8(1+1/k)^{t+1}/\delta, \mathcal{F}_{t+1})$. Expanding $\langle h_t - h_{t+1}, f \rangle_{\nu_t}$ we get

$$\begin{aligned} \langle h_t - h_{t+1}, f \rangle_{\nu_t} \;=\; & \sum_{\ell=1}^{p} c_\ell \underbrace{\prod_{j=1}^{t} \langle g_{\ell,j}, f_j \rangle_{\mu_1}}_{\gamma_\ell} \cdot \langle g_{\ell,t+1}, f_{t+1} \otimes f_{t+2} \rangle_{\mu_{[t+1,k]}} \;- \\ & \sum_{\ell=1}^{p} c'_\ell \underbrace{\prod_{j=1}^{t} \left\langle g'_{\ell,j}, f_j \right\rangle_{\mu_1}}_{\gamma'_\ell} \cdot \langle g'_{\ell,t+1} \otimes g'_{\ell,t+2}, f_{t+1} \otimes f_{t+2} \rangle_{\mu_{[t+1,k]}}, \end{aligned}$$

where we define $\gamma_\ell := \prod_{j=1}^{t} \langle g_{\ell,j}, f_j \rangle_{\mu_1}$ and $\gamma'_\ell := \prod_{j=1}^{t} \left\langle g'_{\ell,j}, f_j \right\rangle_{\mu_1}$ for $\ell \in [p]$, $j \in [t]$. Suppose $g_{\ell,j} = f_{S_{\ell,j}}$ and $g'_{\ell,j} = f_{S'_{\ell,j}}$ for $\ell \in [p]$, $j \in [t]$, where $f_{S_{\ell,j}}, f_{S'_{\ell,j}}$ are either $\mathbf{1}_{S_{\ell,j}}, \mathbf{1}_{S'_{\ell,j}}$ or $\chi_{S_{\ell,j}}, \chi_{S'_{\ell,j}}$ depending on $\mathcal{F}_t$ being $\mathcal{F}_t^{0/1}$ or $\mathcal{F}_t^{\pm 1}$, respectively.

Sigma-algebra brute force: Now for each $j \in [t]$, we form the σ-algebra Σ_j generated by $\{S_{\ell,j}, S'_{\ell,j}\}_{\ell\in[p]}$ which can be done in $2^p \cdot \widetilde{O}(|W[1]|)$ time by Remark 3.16 and yields at most 2^p atoms. Hence, the generation of all these σ-algebras takes at most $t \cdot 2^p \cdot \widetilde{O}(|W[1]|)$ time. Suppose $f_j = f_{S_j}$ for some $S_j \subseteq W[1]$. Let $\eta > 0$ be an approximation parameter to be specified shortly. For each atom $\sigma_{j'} \in \Sigma_j$, we enumerate over all possible values for the ratio $\left|\sigma_{j'} \cap S_j\right| / \left|\sigma_{j'}\right|$ up to accuracy η. More precisely, if $\left|\sigma_{j'}\right| \geq 1/\eta$, we consider the values

$$0, 1\cdot\eta, 2\cdot\eta, \ldots, \lfloor 1/\eta \rfloor \cdot \eta,$$

and we consider $0, 1/\left|\sigma_{j'}\right|, 2/\left|\sigma_{j'}\right|, \ldots, \left|\sigma_{j'}\right| / \left|\sigma_{j'}\right|$ otherwise. Let $\left|\Sigma_j\right|$ denote the number of atoms in Σ_j. This enumeration results in $\prod_{j=1}^{t}(1/\eta)^{\left|\Sigma_j\right|}$ configurations which allows us to approximate any realizable values for $\left\langle g_{\ell,j}, f_j \right\rangle_{\mu_1}$ within additive error at most $4\cdot\eta$ since either

$$\left\langle g_{\ell,j}, f_j \right\rangle_{\mu_1} \; = \; \mathbb{E}_{\mu_1}\left[\mathbf{1}_{S_{\ell,j}} \cdot \mathbf{1}_{S_j}\right] = \frac{\left|S_{\ell,j} \cap S_j\right|}{|W[1]|} = \frac{1}{|W[1]|}\sum_{\sigma_{j'} \subseteq S_{\ell,j}} \left|\sigma_{j'} \cap S_j\right| \quad \text{or}$$

$$\begin{aligned} \left\langle g_{\ell,j}, f_j \right\rangle_{\mu_1} \; &= \; \mathbb{E}_{\mu_1}\left[\chi_{S_{\ell,j}} \cdot \chi_{S_j}\right] \; = \; \frac{|W[1]| - 2(\left|S_{\ell,j}\right| + \left|S_j\right| - 2\left|S_{\ell,j} \cap S_j\right|)}{|W[1]|} \\ &= \; \frac{|W[1]| - 2(\left|S_{\ell,j}\right| + \sum_{\sigma_{j'}}\left|\sigma_{j'} \cap S_j\right| - 2\sum_{\sigma_{j'} \subseteq S_{\ell,j}}\left|\sigma_{j'} \cap S_j\right|)}{|W[1]|}, \end{aligned}$$

according to $\mathcal{F}_{t+1}$. We can approximate $\left\langle g'_{\ell,j}, f_j \right\rangle_{\mu_1}$ similarly. In turn, we can approximate each of the realizable values in $\{\gamma_\ell, \gamma'_\ell\}_{\ell\in[p]}$ within additive error $4\cdot t\cdot\eta$ by some configuration of fractional value assignment to the atoms of each σ-algebra.

Invoking the matrix correlation oracle: Let $\mathsf{A} := \sum_\ell \left(c_\ell \cdot \gamma_\ell \cdot g_{\ell,t+1} + c'_\ell \cdot \gamma'_\ell \cdot g'_{\ell,t+1} \otimes g'_{\ell,t+2}\right)$. We conveniently view A as a *sparse* matrix of dimension $|W[t+1]| \times |W[t+2,k]|$ with at most $|W[t+1,k]|$ non-zeros entries. Define $\varphi_{\mathsf{A}}(f_{t+1}, f_{t+2}) := \langle \mathsf{A}, f_{t+1} \otimes f_{t+2}\rangle_{\mu_{[t+1,k]}}$. Define

$$\mathsf{OPT}(\mathsf{A}) := \max_{f_{t+1}, f_{t+2}} \quad \varphi_{\mathsf{A}}(f_{t+1}, f_{t+2}), \tag{3}$$

where f_{t+1}, f_{t+2} range over valid $f_{S_{t+1}}, f_{S_{t+2}}$ (again according to kind of $\mathcal{F}_{t+1}$ we have). In the computation of $\mathsf{OPT}(\mathsf{A})$, we have incurred so far an additive error of at most

$$4\cdot t\cdot\eta\cdot\sum_\ell(|c_\ell| + |c'_\ell|).$$

Let $\widetilde{\mathsf{A}}$ be obtained from A by zeroing out all entries of absolute value smaller than $\delta/8$. Note that $\mathsf{OPT}(\widetilde{\mathsf{A}}) \geq \mathsf{OPT}(\mathsf{A}) - \delta/8$ and the absolute value of the entries of $\widetilde{\mathsf{A}}$ lie $[\delta/8, O(1/\delta)]$. For each entry of A, we compute a rational approximation $\pm P/Q$ where $Q = \Theta(1/\delta)$ and $P \in [1, O(1/\delta)]$ obtaining $\widetilde{\mathsf{A}}'$ such that

$$\mathsf{OPT}(\widetilde{\mathsf{A}}') \;\geq\; \mathsf{OPT}(\widetilde{\mathsf{A}}) - \delta/8 \;\geq\; \mathsf{OPT}(\widetilde{\mathsf{A}}) \;\geq\; \mathsf{OPT}(\mathsf{A}) - \delta/4.$$

Using Theorem 4.15 with accuracy parameter $\delta/4$ and input matrix $\widetilde{\mathsf{A}}'$, we obtain in $\mathcal{T}_{\mathsf{A}} := \widetilde{O}(\mathrm{poly}(1/\delta)\cdot|W[t+1,k]|)$ time, with an extra additive error of $\delta/4$ and a multiplicative guarantee of $\alpha_{\mathbf{AN}}$, a 2-tensor $\tilde{f}_{t+1} \otimes \tilde{f}_{t+2}$ satisfying

$$\varphi_{\tilde{\mathsf{A}}}(\tilde{f}_{t+1}, \tilde{f}_{t+2}) \;\geq\; \alpha_{\mathbf{AN}} \cdot \left(\mathsf{OPT}(\mathsf{A}) \;-\; 2 \cdot \frac{\delta}{4} \;-\; 4 \cdot t \cdot \eta \cdot \sum_{\ell} (|c_\ell| + |c'_\ell|) \right).$$

Since $h_t \in \mathcal{H}(O(1/\delta^2), 2^8 \cdot (1 + 1/k)^t / \delta, \mathcal{F}_t)$ and $h_{t+1} \in \mathcal{H}(O(1/\delta^2), 2^8 \cdot (1 + 1/k)^{t+1} / \delta, \mathcal{F}_{t+1})$, we have $\sum_\ell (|c_\ell| + |c'_\ell|) \leq 2^{10}/\delta$ and $p = O(1/\delta^2)$. By choosing $\eta \leq O(\delta^2/t)$ appropriately, we can bound

$$4 \cdot t \cdot \eta \cdot \sum_{\ell} (|c_\ell| + |c'_\ell|) \;\leq\; 4 \cdot t \cdot \frac{2^{10}}{\delta} \cdot \eta \;\leq\; \frac{\delta}{4}.$$

Hence, $\varphi_{\tilde{\mathsf{A}}}(\tilde{f}_{t+1}, \tilde{f}_{t+2}) \geq \alpha_{\mathbf{AN}} \cdot \delta/4$ since we are under the assumption that $\mathsf{OPT}(\mathsf{A}) \geq \delta$.

Running Time: First, observe that with our choices of parameters the total number of configurations m_{config} is at most

$$m_{\text{config}} \;\leq\; \prod_{j=1}^{t} (1/\eta)^{|\Sigma_j|} \;\leq\; \left(\frac{t}{\delta^2} \right)^{2^p} \;\leq\; (2t)^{2^{O(1/\delta^2)}},$$

so that the correlation oracle $\mathcal{O}_{\nu_t}$ takes time at most

$$m_{\text{config}} \cdot \mathcal{T}_{\mathsf{A}} \;\leq\; (2t)^{2^{O(1/\delta^2)}} \cdot \widetilde{O}(\text{poly}(1/\delta) \cdot |W[t+1, k]|) \;=\; \widetilde{O}((2t)^{2^{O(1/\delta^2)}} \cdot |W[t+1, k]|).$$

Using the running time of the oracle $\mathcal{O}_{\nu_t}$, the total running time of the weak regularity decomposition follows from Lemma 4.2 which concludes the proof. ■

4.5 Near-linear Time Matrix Correlation Oracles

The main result of this section, Theorem 4.15 below, is a near-linear time correlation oracle for $\text{CUT}^{\otimes 2}$ and $\text{CUT}_{\pm}^{\otimes 2}$. We combine the constant factor approximation algorithms of Alon–Naor [AN04] for $\|\mathsf{A}\|_{\infty \to 1}$ and $\|\mathsf{A}\|_{\square}$ based on semi-definite programming (SDP) with the faster SDP solvers for sparse matrices such as those by Lee and Padmanabhan [LP20] and by Arora and Kale [AK07]. We point out that these SDP solvers provide additive approximation guarantees which are sufficient for approximating several CSPs, e.g., MaxCut, but they do not seem to provide non-trivial multiplicative approximation guarantees for $\|\mathsf{A}\|_{\infty \to 1}$ or $\|\mathsf{A}\|_{\square}$ in general. Since in our applications of computing regularity decomposition we are only interested in additive approximations, those solvers provide non-trivial sufficient approximation guarantees for $\|\mathsf{A}\|_{\infty \to 1}$ or $\|\mathsf{A}\|_{\square}$ in our settings.

Theorem 4.15. *[Alon–Naor Correlation Oracle] Let $\mathcal{F}$ be either $\text{CUT}^{\otimes 2}$ or $\text{CUT}_{\pm}^{\otimes 2}$ and μ be the uniform measure supported on at most m elements of $[n'] \times [n']$. There exists an algorithmic $(\delta, \alpha_{\mathbf{AN}} \cdot \delta)$-correlation oracle $\mathcal{O}_{\mu,B}$ running in time $\mathcal{T}_{\mathcal{O}_{\mu,B}} = \tilde{O}(\text{poly}(B/\delta) \cdot (m + n'))$, where $\alpha_{\mathbf{AN}} \geq 1/2^4$ is an approximation ratio constant.*

Theorem 4.15 is a simple consequence of the following theorem.

Theorem 4.16. *Let $\mathsf{A} \in \mathbb{R}^{n \times n}$ be a matrix of integers with at most m non-zero entries. Let $\delta \in (0, 2^{-5}]$ be an accuracy parameter. Suppose that*

$$\mathsf{OPT} := \max_{x_i, y_i \in \{\pm 1\}} \sum_{i,j=1}^{n} \mathsf{A}_{i,j} x_i y_j \;\geq\; \delta \cdot m.$$

Then, with high probability,i.e., $o_n(1)$, we we can find in $\tilde{O}\left(\mathrm{poly}(\|\mathsf{A}\|_\infty/\delta)\cdot(m+n)\right)$ time vectors $\tilde{x},\tilde{y}\in\{\pm 1\}^n$ such that

$$\sum_{i,j=1}^{n} \mathsf{A}_{i,j}\tilde{x}_i\tilde{y}_j \;\geq\; \frac{1}{4}\cdot\mathsf{OPT},$$

and find sets $\tilde{S},\tilde{T}\subseteq[n]$ such that

$$\left|\sum_{i\in\tilde{S},j\in\tilde{T}} \mathsf{A}_{i,j}\right| \;\geq\; \frac{1}{2^4}\cdot\|\mathsf{A}\|_{\square},$$

where $\|\mathsf{A}\|_{\square}$ is the cut norm of A.

The proof of the preceding theorem will rely on the following result which encapsulates the known sparse SDP solvers [AK07, LP20]. For concreteness, we will rely on [LP20] although the guarantee from [AK07] also suffice for us.

Lemma 4.17. *[Sparse SDP Solver Wrapper based on [LP20] and partially on [AK07]] Let $\mathsf{C}\in\mathbb{R}^{n\times n}$ be a matrix with at most m non-zero entries. For every accuracy $\gamma>0$, with high probability we can find in time $\widetilde{O}((m+n)/\mathrm{poly}(\gamma))$ vectors $u_1,\ldots,u_n\in\mathbb{R}^n$ in the unit ball (i.e., $\|u_i\|\leq 1$) such that that the matrix $\widetilde{\mathsf{X}}_{i,j}:=\langle u_i,u_j\rangle$ satisfies*

$$\mathrm{Tr}\left(\mathsf{C}\cdot\widetilde{\mathsf{X}}\right) \;\geq\; \max_{\mathsf{X}\succeq 0, X_{i,i}\leq 1}\mathrm{Tr}\left(\mathsf{C}\cdot\mathsf{X}\right)-\gamma\sum_{i,j}\left|\mathsf{C}_{i,j}\right|.$$

Proof of Theorem 4.16. We now implement the strategy mentioned above of combing the approximation algorithms of Alon–Naor [AN04] with the near-linear time sparse SDP solvers. We still need to argue that this indeed leads to the claimed approximation guarantees while being computable in near-linear time overall. We point out that Alon–Naor actually give a constant factor SDP based approximation algorithm for $\|\mathsf{A}\|_{\infty\to 1}$ from which a constant factor approximation algorithm for $\|\mathsf{A}\|_{\square}$ can be readily deduced from in near-linear time incurring an extra $1/4$ factor approximation loss[5]. Using the matrix A, we set

$$\mathsf{C}:=\frac{1}{2}\begin{pmatrix}0 & \mathsf{A}\\ \mathsf{A}^\dagger & 0\end{pmatrix}.$$

The SDP relaxation of Alon–Naor for $\|\mathsf{A}\|_{\infty\to 1}$ becomes

$$\begin{aligned}\max\quad & \mathrm{Tr}(\mathsf{C}\cdot\mathsf{X}) && =: \mathsf{SDP}^*\\ \text{s.t.}\quad & \mathsf{X}_{i,i}\;\leq\;1 && \forall i\in[2n]\\ & \mathsf{X}\;\succeq\;0,\end{aligned}$$

except for the constraints $X_{i,i}\leq 1$ which they instead take to be $X_{i,i}=1$. This technical difference will play a (small) role in the rounding of this SDP since Alon–Naor analysis relies on Gram vectors of X being on the unit sphere. Moreover, we will be solving this

[5] In Section 5.4 of Alon–Naor [AN04], there is a transformation avoiding any loss in the approximation ratio. Since constant factors are not asymptotically important for us, we rely on the simpler transformation which loses a factor of $1/4$. It simply consists in choosing $\widetilde{S}\in\{\{i\mid\widetilde{x}_i=1\},\{i\mid\widetilde{x}_i=-1\}\}$ and $\widetilde{T}\in\{\{j\mid\widetilde{y}_j=1\},\{j\mid\widetilde{y}_j=-1\}\}$ maximizing $\mathbf{1}_{\widetilde{S}}^t\mathsf{A}\mathbf{1}_{\widetilde{T}}$, which can be done in near-linear time given as input $\widetilde{x},\widetilde{y}$.

SDP within only a weak additive approximation guarantee[6]. Although these technical differences need to be handled, this will be simple to do.

Applying the solver of Lemma 4.17 with accuracy parameter $\gamma = \delta^2 / \|\mathsf{A}\|_\infty$ to the above SDP, we obtain in $\widetilde{O}(\mathrm{poly}(\|\mathsf{A}\|_\infty / \delta) \cdot (m+n))$ time vectors $u_1, \ldots, u_{2n} \in \mathbb{R}^{2n}$ in the unit ball so that the matrix $\widetilde{\mathsf{X}}_{i,j} := \langle u_i, u_j \rangle$ satisfy

$$\mathrm{Tr}\left(\mathsf{C} \cdot \widetilde{\mathsf{X}}\right) \;\geq\; \max_{\mathsf{X} \succeq 0, X_{i,i} \leq 1} \mathrm{Tr}\left(\mathsf{C} \cdot \mathsf{X}\right) \;-\; \delta^2 \cdot m.$$

By assumption, we have $\mathsf{SDP}^* := \max_{\mathsf{X} \succeq 0, X_{i,i} \leq 1} \mathrm{Tr}\left(\mathsf{C} \cdot \mathsf{X}\right) \geq \mathsf{OPT} \geq \delta \cdot m$, in which case the above guarantee becomes

$$\mathrm{Tr}\left(\mathsf{C} \cdot \widetilde{\mathsf{X}}\right) \;\geq\; (1-\delta) \cdot \mathsf{SDP}^*.$$

To obtain diagonal entries equal to 1 in our SDP solution we simply consider the new SDP solution $\widetilde{\mathsf{X}}' = \widetilde{\mathsf{X}} + \Lambda$, where Λ is the diagonal matrix defined as $\Lambda_{i,i} := 1 - \widetilde{\mathsf{X}}_{i,i}$. Gram vectors $u'_1, \ldots, u'_{2n}$ of $\widetilde{\mathsf{X}}'$ can be obtained in near-linear time from $u_1, \ldots, u_{2n}$ and Λ by setting

$$u'_i := u_i \oplus \sqrt{\Lambda_{i,i}} \cdot e_i \;\in\; \mathbb{R}^{2m} \oplus \mathbb{R}^{2m},$$

where $e_i \in \mathbb{R}^{2m}$ has a one at the ith position and zero everywhere else. Observe that for our particular C, we have

$$\mathrm{Tr}\left(\mathsf{C} \cdot \widetilde{\mathsf{X}}'\right) \;=\; \mathrm{Tr}\left(\mathsf{C} \cdot \widetilde{\mathsf{X}}\right).$$

We now proceed to round $\widetilde{\mathsf{X}}'$ according to the rounding scheme of Alon–Naor [AN04] (*cf.*,Section 5.1) which was chosen because it is simple enough to easily afford a near-linear time computation while providing a $\approx 0.27 \geq 1/4$ approximation guarantee [7] This rounding consists in sampling a Gaussian vector $g \sim N(0, \mathsf{I}_d)$ and setting $\widetilde{x}_i := \mathrm{sgn}\langle u'_i, g\rangle$ and $\widetilde{y}_{i+n} := \mathrm{sgn}\langle u'_{i+n}, g\rangle$ for $i \in [n]$. To analyze the approximation guarantee, the following identity is used.

Fact 4.18 (Alon–Naor [AN04], *cf.*,Eq. 5)**.** *Let $u, w \in \mathbb{R}^d$ be unit vectors in ℓ_2-norm. Then*

$$\frac{\pi}{2} \cdot \mathbb{E}\left[\mathrm{sgn}\langle u, g\rangle \,\mathrm{sgn}\langle w, g\rangle\right] = \langle u, w\rangle + \mathbb{E}\left[\left(\langle u, g\rangle - \sqrt{\frac{\pi}{2}}\,\mathrm{sgn}\langle u, g\rangle\right)\left(\langle w, g\rangle - \sqrt{\frac{\pi}{2}}\,\mathrm{sgn}\langle w, g\rangle\right)\right],$$

where the expectations are taken with respect to a random Gaussian vector $g \sim N(0, \mathsf{I}_d)$.

Using Fact 4.18, the expected value of the rounding, i.e.,

$$\mathbb{E}\left[\sum_{i,j} \mathsf{A}_{i,j}\,\mathrm{sgn}\langle u'_i, g\rangle\,\mathrm{sgn}\left\langle u'_{j+n}, g\right\rangle\right],$$

becomes

$$\frac{2}{\pi} \cdot \sum_{i,j} \mathsf{A}_{i,j}\left\langle u'_i, u'_{j+n}\right\rangle + \frac{2}{\pi} \cdot \sum_{i,j} \mathsf{A}_{i,j}\mathbb{E}\left[\left(\langle u'_i, g\rangle - \sqrt{\frac{\pi}{2}}\,\mathrm{sgn}\langle u'_i, g\rangle\right)\left(\left\langle u'_{j+n}, g\right\rangle - \sqrt{\frac{\pi}{2}}\,\mathrm{sgn}\left\langle u'_{j+n}, g\right\rangle\right)\right],$$

[6] This may not be sufficient to obtain $X_{i,i} \approx 1$ by an extremality argument

[7] Alon–Naor [AN04] have a more sophisticated rounding scheme that achieves $0.56 \geq 1/2$ approximation. In our applications, it is important to have a constant factor approximation, but the distinction between $1/2$ and the weaker $1/4$ factor approximation guarantee is not asymptotically relevant.

As in Alon–Naor [AN04], we will use the fact that $\langle u_i', g\rangle - \sqrt{\frac{\pi}{2}}\operatorname{sgn}\langle u_i', g\rangle$ and $\left\langle u_{j+n}', g\right\rangle - \sqrt{\frac{\pi}{2}}\operatorname{sgn}\left\langle u_{j+n}', g\right\rangle$ are themselves vectors on a Hilbert space with norm squared $\pi/2-1$. Then, in our setting we obtain

$$\begin{aligned}
\mathbb{E}\left[\sum_{i,j}\mathsf{A}_{i,j}\operatorname{sgn}\langle u_i', g\rangle\operatorname{sgn}\left\langle u_{j+n}', g\right\rangle\right] &\geq \frac{2}{\pi}(1-\delta)\cdot\mathsf{SDP}^* - \left(1-\frac{2}{\pi}\right)\cdot\mathsf{SDP}^* \\
&\geq \frac{2}{\pi}\left(2-\frac{\pi}{2}-\delta\right)\cdot\mathsf{SDP}^* \\
&\geq \left(\frac{1}{4}+\Omega(1)\right)\cdot\mathsf{SDP}^* && \text{(Since } \delta\leq 2^{-5}\text{)} \\
&\geq \left(\frac{1}{4}+\Omega(1)\right)\cdot\mathsf{OPT},
\end{aligned}$$

as claimed. By standard techniques, this guarantee on the expected value of the rounded solution can be used to give with high probability a guarantee of $1/4\cdot\mathsf{OPT}$ (namely, by repeating this rounding scheme $O(\operatorname{poly}(1/\gamma)\cdot\log(n))$ times). ■

We now proceed to establish the sparse SDP solver wrapper claimed in Lemma 4.17. For concreteness, we will use the following sparse SDP solver result of Lee–Padmanabhan [LP20]. The analogous result of Arora–Kale [AK07] with slightly worse parameters also suffices for our purposes, but the main result of [LP20] is stated in more convenient form.

Theorem 4.19 (Adapted from Theorem 1.1 of [LP20])**.** *Given a matrix $\mathsf{C}\in\mathbb{R}^{n\times n}$ with m non-zero entries, parameter $\gamma\in(0,1/2]$, with high probability, in time $\widetilde{O}((m+n)/\gamma^{3.5})$, it is possible to find a symmetric matrix $\mathsf{Y}\in\mathbb{R}^{n\times n}$ with $O(m)$ non-zero entries and diagonal matrix $\mathsf{S}\in\mathbb{R}^{n\times n}$ so that $\widetilde{X}=\mathsf{S}\cdot\exp\mathsf{Y}\cdot\mathsf{S}$ satisfies*

- $\widetilde{X}\succeq 0$,
- $\widetilde{X}_{i,i}\leq 1$ *for every* $1\leq i\leq n$, *and*
- $\operatorname{Tr}(\mathsf{C}\cdot\widetilde{X}) \;\geq\; \max_{\mathsf{X}\succeq 0, X_{i,i}\leq 1}\operatorname{Tr}(\mathsf{C}\cdot\mathsf{X}) - \gamma\sum_{i,j}\left|\mathsf{C}_{i,j}\right|$.

Furthermore, we have $\|\mathsf{Y}\|_{\mathrm{op}}\leq O(\log(n)/\gamma)$ (cf.,Lemma C.2.3 of [LP20]).

Remark 4.20. *We observe that Theorem 4.19 differs from Theorem 1.1 of [LP20] only by an additional bound on $\|\mathsf{Y}\|_{\mathrm{op}}$. This bound is important in analyzing the error when approximating (matrix) exponential of Y.*

We now show how we can approximate the Gram vectors of the SDP solution of Theorem 4.19. We rely on part of the analysis in Arora–Kale [AK07].

Claim 4.21. *Let $\mathsf{C}\in\mathbb{R}^{n\times n}$ be a matrix with at most m non-zero entries and $\gamma\in(0,1/2]$. Suppose $\widetilde{X}=\mathsf{S}\cdot\exp\mathsf{Y}\cdot\mathsf{S}$ satisfy the conclusions of Theorem 4.19 given $\mathsf{C}\in\mathbb{R}^{n\times n}$ and accuracy γ. Then with high probability we can find in $\widetilde{O}(\operatorname{poly}(1/\gamma)\cdot(m+n))$ time approximate Gram vectors $u_1,\ldots,u_n\in\mathbb{R}^n$ such that $\widetilde{X}'_{i,j}:=\langle u_i,u_j\rangle$ satisfy*

- $\widetilde{X}'_{i,i}\leq 1$ *for every* $1\leq i\leq n$, *and*

- $\mathrm{Tr}(\mathsf{C} \cdot \widetilde{X}') \geq \mathrm{Tr}\left(\mathsf{C} \cdot \widetilde{\mathsf{X}}\right) - \gamma \sum_{i,j} |\mathsf{C}_{i,j}|$.

Proof. Since $\widetilde{X} = (\mathsf{S} \cdot \exp(\mathsf{Y}/2))(\mathsf{S} \cdot \exp(\mathsf{Y}/2))^t$, the rows of $\mathsf{S} \cdot \exp(\mathsf{Y}/2)$ can be taken as Gram vectors $u_1, \ldots, u_n \in \mathbb{R}^n$ of $\widetilde{X}$. If we knew the rows of $\exp(\mathsf{Y}/2)$, we could readily recover these Gram vectors since S is diagonal. As observed in Arora–Kale [AK07], computing $\exp(\mathsf{Y}/2)$ may be computationally expensive, so instead one can approximate the matrix-vector product $\exp(\mathsf{Y}/2)u$ using $d = O(\log(n)/\gamma^2)$ random Gaussian vectors $u \sim N(0, I_n)$. By the Johnson–Lindenstrauss Lemma and scaling by $\sqrt{n/d}$, with high probability we obtain vectors $\widetilde{u}_1, \ldots, \widetilde{u}_n$ satisfying for every $i, j \in [n]$ say

$$\left|\langle u_i, u_j \rangle - \langle \widetilde{u}_i, \widetilde{u}_j \rangle\right| \leq \frac{\gamma}{6}.$$

In particular, whp $\|\widetilde{u}_i\|_2^2 \leq 1 + \gamma/6$. Thus, by normalizing the vectors $\widetilde{u}_i$ with $\|\widetilde{u}_i\|_2 > 1$ to have ℓ_2-norm one the preceding approximation deteriorates to

$$\left|\langle u_i, u_j \rangle - \langle \widetilde{u}_i, \widetilde{u}_j \rangle\right| \leq \gamma/2.$$

To compute each the matrix-vector product $\exp(\mathsf{Y}/2)u$ in $\widetilde{O}(\mathrm{poly}(1/\gamma) \cdot (m+n))$, we rely on the following lemma.

Lemma 4.22 (Arora–Kale [AK07], *cf.*,Lemma 6)**.** *Let $\mathcal{T}_{\mathsf{Y}}$ be the time needed to compute the matrix-vector product $\mathsf{Y}u$. Then the vector $v := \sum_{i=0}^{k} \mathsf{Y}^i u/(i!)$ can be computed in $O(k \cdot \mathcal{T}_{\mathsf{Y}})$ time and if $k \geq \max\{e^2 \cdot \|\mathsf{Y}\|_{\mathrm{op}}, \ln(1/\delta)\}$, then $\|\exp(\mathsf{Y})u - v\|_2 \leq \delta$.*

By noting that $\|\mathsf{Y}\|_{\mathrm{op}} \leq O(\log(n)/\gamma)$ and the time $\mathcal{T}_{\mathsf{Y}}$ (*cf.*, Lemma 4.22) $\mathsf{Y}u$ is $\widetilde{O}((m+n)/\gamma)$, applying Lemma 4.22 with say $\delta \leq \mathrm{poly}(\gamma/n)$ we can approximate each $\exp(\mathsf{Y}/2)u$ in time $\widetilde{O}((m+n)/\gamma)$. Therefore, the total running is $\widetilde{O}(\mathrm{poly}(1/\gamma) \cdot (m+n))$ as claimed. Then the actual Gram vectors still satisfy

$$\left|\langle u_i, u_j \rangle - \langle \widetilde{u}_i, \widetilde{u}_j \rangle\right| \leq \gamma.$$

Hence, we get

$$\mathrm{Tr}(\mathsf{C} \cdot \widetilde{X}') \geq \mathrm{Tr}\left(\mathsf{C} \cdot \widetilde{\mathsf{X}}\right) - \gamma \sum_{i,j} \left|\mathsf{C}_{i,j}\right|,$$

concluding the proof. ∎

We are ready to prove Lemma 4.17 which is restated below for convenience.

Lemma 4.17. *[Sparse SDP Solver Wrapper based on [LP20] and partially on [AK07]] Let $\mathsf{C} \in \mathbb{R}^{n \times n}$ be a matrix with at most m non-zero entries. For every accuracy $\gamma > 0$, with high probability we can find in time $\widetilde{O}((m+n)/\mathrm{poly}(\gamma))$ vectors $u_1, \ldots, u_n \in \mathbb{R}^n$ in the unit ball (i.e., $\|u_i\| \leq 1$) such that that the matrix $\widetilde{\mathsf{X}}_{i,j} := \langle u_i, u_j \rangle$ satisfies*

$$\mathrm{Tr}\left(\mathsf{C} \cdot \widetilde{\mathsf{X}}\right) \geq \max_{\mathsf{X} \succeq 0, X_{i,i} \leq 1} \mathrm{Tr}\left(\mathsf{C} \cdot \mathsf{X}\right) - \gamma \sum_{i,j} \left|\mathsf{C}_{i,j}\right|.$$

Proof of Lemma 4.17. Follows by combining the SDP solution $\widetilde{X}$ of Theorem 4.19 with the fast approximate Gram vector computation of Claim 4.21, the latter yielding another approximated SDP solution $\widetilde{X}'$. In both of these computations, we use accuracy parameter $\gamma/2$ so that

$$\begin{aligned}\mathrm{Tr}(\mathsf{C}\cdot\widetilde{X}') &\geq \mathrm{Tr}\left(\mathsf{C}\cdot\widetilde{\mathsf{X}}\right)-\frac{\gamma}{2}\sum_{i,j}\left|\mathsf{C}_{i,j}\right| \\ &\geq \max_{\mathsf{X}\succeq 0, X_{i,i}\leq 1}\mathrm{Tr}\left(\mathsf{C}\cdot\mathsf{X}\right)-\frac{\gamma}{2}\sum_{i,j}\left|\mathsf{C}_{i,j}\right|-\frac{\gamma}{2}\sum_{i,j}\left|\mathsf{C}_{i,j}\right|.\end{aligned}$$

Moreover, each step takes $\widetilde{O}(\mathrm{poly}(1/\gamma)\cdot(m+n))$ which concludes the proof. ∎

5 Regularity Based Decoding

5.1 List Decoding of Direct-Sum Codes

We now develop list-decoding algorithms for direct-sum codes, using the regularity lemmas obtained in the previous section. We will prove the following theorem.

Theorem 5.1. *Let $\mathcal{C}_0 \subset \mathbb{F}_2^n$ be a code with $\mathrm{bias}(\mathcal{C}_0) \leq \varepsilon_0$, which is unique-decodable to distance $(1-\varepsilon_0)/4$ in time $\mathcal{T}_0$. Let $W \subseteq [n]^k$ be a d-regular, τ-splittable collection of tuples, and let $\mathcal{C} = \mathrm{dsum}_W(\mathcal{C}_0)$ be the corresponding direct-sum lifting of $\mathcal{C}_0$ with $\mathrm{bias}(\mathcal{C}) \leq \varepsilon$. Let β be such that*

$$\beta \;\geq\; \max\left\{\sqrt{\varepsilon},\; \left(2^{20}\cdot\tau\cdot k^3\right)^{1/2},\; 2\cdot\left(\frac{1}{2}+2\varepsilon_0\right)^{k/2}\right\}.$$

Then, there exists a randomized algorithm, which given $\tilde{y} \in \mathbb{F}_2^W$, recovers the list $\mathcal{L}_\beta(\tilde{y}) := \{y \in \mathcal{C} \mid \Delta(\tilde{y}, y) \leq 1/2 - \beta\}$ with probability $1 - o(1)$, in time $\tilde{O}(C_{\beta,k,\varepsilon_0}\cdot(|W| + \mathcal{T}_0))$, where $C_{k,\beta,\varepsilon_0} = (6/\varepsilon_0)^{2^{O(k^3/\beta^2)}}$.

To obtain the decoding algorithm, we first define a function $g : [n]^k \to \{-1, 1\}$ supported on W as

$$g(i_1,\ldots,i_k) \;:=\; \begin{cases}(-1)^{\tilde{y}_{(i_1,\ldots,i_k)}} & \text{if } (i_1,\ldots,i_k)\in W\\ 0 & \text{otherwise}\end{cases}$$

For each $z \in \mathbb{F}_2^n$, we also consider the similar function $\chi_z : [n] \to \{-1, 1\}$ defined as $\chi_z(i) = (-1)^{z_i}$. We first re-state the decoding problem in terms of the functions g and χ_z.

Claim 5.2. *Let $z \in \mathbb{F}_2^n$, and let the functions g and χ_z be as above. Then,*

$$\Delta(\tilde{y}, \mathrm{dsum}_W(z)) \leq \frac{1}{2}-\beta \quad\Leftrightarrow\quad \left\langle g, \chi_z^{\otimes k}\right\rangle_{\mu_k} = \left(\frac{n}{d}\right)^{k-1}\cdot\left\langle g, \chi_z^{\otimes k}\right\rangle_{\mu_1^{\otimes k}} \geq 2\beta.$$

Proof. We have

$$\begin{aligned}\Delta(\tilde{y}, \mathrm{dsum}_W(z)) &= \mathop{\mathbb{E}}_{(i_1,\ldots,i_k)\sim W}\left[\mathbb{1}_{\{\tilde{y}_{(i_1,\ldots,i_k)} \neq z_{i_1}+\cdots+z_{i_k} \mod 2\}}\right] \\ &= \mathop{\mathbb{E}}_{(i_1,\ldots,i_k)\sim\mu_k}\left[\frac{1-g(i_1,\ldots,i_k)\cdot\prod_{t\in[k]}\chi_z(i_t)}{2}\right] = \frac{1}{2}-\frac{1}{2}\cdot\left\langle g,\chi_z^{\otimes k}\right\rangle_{\mu_k}.\end{aligned}$$

Finally, using the fact that g is only supported on W, and $|W| = d^{k-1} \cdot n$ by d-regularity, we have $\langle g, f \rangle_{\mu_k} = (n/d)^{k-1} \cdot \langle g, f \rangle_{\mu_1^{\otimes k}}$ for any function $f : [n]^k \to \mathbb{R}$. ■

Note that each element of the list $\mathcal{L}_\beta(\tilde{y})$ must be equal to $\mathrm{dsum}_W(z)$ for some $z \in \mathcal{C}_0$. Thus, to search for all such z, we will consider the decomposition h of the function g, given by Theorem 4.11 with respect to the class of functions $\mathcal{F} = \mathrm{CUT}_\pm^{\otimes k}$. Since the functions $\chi_z^{\otimes k}$ belong to $\mathcal{F}$, it will suffice to only consider the inner product $\langle h, \chi_z^{\otimes k} \rangle_{\mu_1^{\otimes k}}$.

Also, since the approximating function h is determined by a small number of functions, say $\{f_1, \ldots, f_r : [n] \to \{-1, 1\}\}$, it will suffice to (essentially) consider only the functions measurable in the factor $\mathcal{B}$ determined by $f_1, \ldots, f_r$. Recall that the factor $\mathcal{B}$ is simply a partition of $[n]$ in 2^r pieces according to the values of $f_1, \ldots, f_r$. Also, since any $\mathcal{B}$-measurable function is constant on each piece, it is completely specified by $|\mathcal{B}|$ real values. We will only consider functions taking values in $[-1, 1]$, and discretize this space to an appropriate accuracy η, to identify all relevant $\mathcal{B}$-measurable functions with the set $\{0, \pm\eta, \pm 2\eta, \ldots, \pm 1\}^{|\mathcal{B}|}$. The decoding procedure is described in the following algorithm.

Algorithm 5.3 (List Decoding)**.**

Input $\tilde{y} \in \mathbb{F}_2^W$

Output *List* $\mathcal{L} \subseteq \mathcal{C}$

- *Obtain the approximator h given by Theorem 4.11 for $\mathcal{F} = \mathrm{CUT}_\pm^{\otimes k}$, $\delta = \beta$, and the function $g : [n]^k \to \{-1, 1\}$ defined as*

$$g(i_1, \ldots, i_k) \;:=\; \begin{cases} (-1)^{\tilde{y}(i_1, \ldots, i_k)} & \text{if } (i_1, \ldots, i_k) \in W \\ 0 & \text{otherwise} \end{cases}$$

- *Let h be of the form $h = \sum_{j=1}^{p} c_j \cdot f_{j_1} \otimes \cdots \otimes f_{j_k}$, with each $f_{j_t} : [n] \to \{-1, 1\}$. Let $\mathcal{B}$ be the factor determined by the functions $\{f_{j_t}\}_{j \in [p], t \in [k]}$.*

- *Let $\mathcal{L} = \emptyset$ and let $\eta = 1/\lceil (2/\varepsilon_0) \rceil$. For each $\mathcal{B}$-measurable function $\overline{f}$ given by a value in $D_\eta := \{0, \pm\eta, \pm 2\eta, \ldots, \pm 1\}$ for every atom of $\mathcal{B}$:*

  - *Sample a random function $\chi : [n] \to \{-1, 1\}$ by independently sampling $\chi(i) \in \{-1, 1\}$ for each i, such that $\mathbb{E}[\chi(i)] = \overline{f}(i)$. Take $\tilde{z} \in \mathbb{F}_2^n$ to be such that $\chi = \chi_{\tilde{z}}$.*
  - *If there exists $z \in \mathcal{C}_0$ such that*

$$\Delta(\tilde{z}, z) \;\le\; \frac{(1 - \varepsilon_0)}{4} \quad \text{and} \quad \Delta(\tilde{y}, \mathrm{dsum}_W(z)) \;\le\; \frac{1}{2} - \beta\,,$$

    then $\mathcal{L} \leftarrow \mathcal{L} \cup \{\mathrm{dsum}_W(z)\}$.

- *Return $\mathcal{L}$.*

Note that by our choice of the β in Theorem 5.1, we have that $\tau \le \beta^2/(2^{20}k^3)$. Thus, we can indeed apply Theorem 4.11 to obtain the function h as required by the algorithm. To show that the algorithm can recover the list, we will need to show that for each z such that $\mathrm{dsum}_W(z) \in \mathcal{L}_\beta$, the sampling procedure finds a $\tilde{z}$ close to z with significant probability. To analyze this probability, we first prove the following claim.

Claim 5.4. *Let $z \in \mathbb{F}_2^n$ and let $\overline{f} : [n] \to D_\eta$ be a minimizer of $\|\mathbb{E}[\chi_z|\mathcal{B}] - \overline{f}\|_\infty$ among all $\mathcal{B}$-measurable functions in $D_\eta^{|\mathcal{B}|}$. Then, over the random choice of χ such that $\mathbb{E}[\chi] = \overline{f}$, we have*

$$\mathop{\mathbb{E}}_{\chi}\left[\langle \chi, \chi_z \rangle_{\mu_1}\right] \;=\; \left\langle \overline{f}, \chi_z \right\rangle_{\mu_1} \;\geq\; \|\mathbb{E}\left[\chi_z|\mathcal{B}\right]\|_{\mu_1}^2 - \eta\,.$$

Proof. By linearity of the inner product, we have

$$\mathop{\mathbb{E}}_{\chi}\left[\langle \chi, \chi_z \rangle_{\mu_1}\right] \;=\; \langle \mathbb{E}\left[\chi\right], \chi_z \rangle_{\mu_1} \;=\; \left\langle \overline{f}, \chi_z \right\rangle_{\mu_1} \;=\; \left\langle \overline{f}, \mathbb{E}\left[\chi_z|\mathcal{B}\right] \right\rangle_{\mu_1},$$

where the last equality used Proposition 3.14 and the fact that $\overline{f}$ is $\mathcal{B}$-measurable. Since $\mathbb{E}[\chi_z|\mathcal{B}]$ takes values in $[-1,1]$ and $\overline{f}$ is the minimizer over all functions in $D_\eta^{|\mathcal{B}|}$, we must have $\|\mathbb{E}[\chi_z|\mathcal{B}] - \overline{f}\|_\infty \leq \eta$. Using this pointwise bound, we get

$$\begin{aligned}\left\langle \overline{f}, \mathbb{E}\left[\chi_z|\mathcal{B}\right] \right\rangle_{\mu_1} &= \mathop{\mathbb{E}}_{i\sim\mu_1}\left[\overline{f}(i)\cdot \mathbb{E}\left[\chi_z|\mathcal{B}\right](i)\right] \\ &\geq \mathop{\mathbb{E}}_{i\sim\mu_1}\left[\left(\mathbb{E}\left[\chi_z|\mathcal{B}\right](i)\right)^2 - \eta\cdot\left|\mathbb{E}\left[\chi_z|\mathcal{B}\right](i)\right|\right] \;\geq\; \|\mathbb{E}\left[\chi_z|\mathcal{B}\right]\|_{\mu_1}^2 - \eta\,. \qquad \blacksquare\end{aligned}$$

We next show that when $z \in \mathbb{F}_2^n$ is such that $\langle g, \chi_z^{\otimes k}\rangle$ is large, then the norm of the conditional expectation $\mathbb{E}[\chi_z|\mathcal{B}]$ is also large, and hence the sampling procedure finds a $\tilde{z}$ close to z. When we have a $z \in \mathcal{C}_0$ with such a property, we can use $\tilde{z}$ to recover z using the unique decoding algorithm for $\mathcal{C}_0$.

Lemma 5.5. *Let $z \in \mathbb{F}_2^n$ be such that*

$$\left\langle g, \chi_z^{\otimes k}\right\rangle_{\mu_k} = \left(\frac{n}{d}\right)^{k-1}\cdot\left\langle g, \chi_z^{\otimes k}\right\rangle_{\mu_1^{\otimes k}} \geq\; 2\beta\,.$$

Then, we have $\|\mathbb{E}[\chi_z|\mathcal{B}]\|_{\mu_1}^2 \geq (\beta/2)^{2/k}$.

Proof. Let h be the approximating function obtained by applying Theorem 4.11 to g with approximation error $\delta = \beta$. Note that we have $\|h\|_{\mu_1^{\otimes k}} \leq 2$, and for any $f \in \mathrm{CUT}_\pm^{\otimes k}$,

$$\left(\frac{n}{d}\right)^{k-1}\cdot\left\langle g - \left(\frac{d}{n}\right)^{k-1}\cdot h\,, f\right\rangle_{\mu_1^{\otimes k}} \;\leq\; \delta\,.$$

Using $f = \chi_z^{\otimes k}$ and $\delta = \beta$, we get

$$\left\langle h, \chi_z^{\otimes k}\right\rangle_{\mu_1^{\otimes k}} \;\geq\; 2\beta - \delta \;\geq\; \beta\,.$$

Using Proposition 3.14, and the fact that $\mathcal{B}$ is defined so that all functions in the decomposition of h are (by definition) $\mathcal{B}$-measurable, we have

$$\left\langle h, \chi_z^{\otimes k}\right\rangle_{\mu_1^{\otimes k}} \;=\; \sum_{j=1}^{p} c_j \prod_{t=1}^{k}\langle f_{j_t}, \chi_z\rangle_{\mu_1} \;=\; \sum_{j=1}^{p} c_j \prod_{t=1}^{k}\langle f_{j_t}, \mathbb{E}\left[\chi_z|\mathcal{B}\right]\rangle_{\mu_1} \;=\; \left\langle h, \left(\mathbb{E}\left[\chi_z|\mathcal{B}\right]\right)^{\otimes k}\right\rangle_{\mu_1^{\otimes k}}.$$

Combining the above with Cauchy-Schwarz, we get

$$\beta \;\leq\; \left\langle h, \chi_z^{\otimes k}\right\rangle_{\mu_1^{\otimes k}} \;\leq\; \|h\|_{\mu_1^{\otimes k}}\cdot\left\|\left(\mathbb{E}\left[\chi_z|\mathcal{B}\right]\right)^{\otimes k}\right\|_{\mu_1^{\otimes k}} \;=\; \|h\|_{\mu_1^{\otimes k}}\cdot\|\mathbb{E}\left[\chi_z|\mathcal{B}\right]\|_{\mu_1}^k\,.$$

Using $\|h\|_{\mu_1^{\otimes k}} \leq 2$ then gives $\|\mathbb{E}[\chi_z|\mathcal{B}]\|_{\mu_1}^2 \geq (\beta/2)^{2/k}$. $\blacksquare$

Using the above results, we can now complete the analysis of the algorithm.

Proof of Theorem 5.1. We first argue that for any codeword $z \in \mathcal{C}_0$ such that $\mathrm{dsum}_W(z) \in \mathcal{L}_\beta$, sampling a random function χ (with $\mathbb{E}[\chi] = \overline{f}$ for an appropriate $\overline{f}$) finds a $\tilde{z}$ close to z with significant probability. Let $\overline{f} \in D_\eta^{\mathcal{B}}$ be the minimizer of $\|\chi_z - \overline{f}\|_\infty$, for such a $z \in \mathcal{C}_0$. We have by Claim 5.4 that $\mathbb{E}_\chi[\langle \chi, \chi_z \rangle_{\mu_1}] \geq \|\mathbb{E}[\chi_z|\mathcal{B}]\|_{\mu_1}^2 - \eta$. Since $\Delta(\tilde{y}, \mathrm{dsum}_W(z)) \leq 1/2 - \beta$, we have by Claim 5.2 that $\langle g, \chi_z^{\otimes k} \rangle_{\mu_k} \geq 2\beta$. Thus, by Lemma 5.5, we have that $\|\mathbb{E}[\chi_z|\mathcal{B}]\|_{\mu_1}^2 \geq (\beta/2)^{2/k}$. Combining these, and using the lower bound on β, we get that

$$\mathop{\mathbb{E}}_{\chi}\left[\langle \chi, \chi_z \rangle_{\mu_1}\right] \;\geq\; \left(\frac{\beta}{2}\right)^{2/k} - \eta \;\geq\; \frac{1}{2} + 2\varepsilon_0 - \eta \;\geq\; \frac{1}{2} + \frac{3\varepsilon_0}{2}.$$

Since $\langle \chi, \chi_z \rangle_{\mu_1}$ is the average of n independent (not necessarily identical) random variables $\{\chi(i) \cdot \chi_z(i)\}_{i \in [n]}$ in the range $[-1, 1]$, we get by Hoeffding's inequality that

$$\mathop{\mathbb{P}}_{\chi}\left[\langle \chi, \chi_z \rangle_{\mu_1} \leq \frac{1}{2} + \varepsilon_0\right] \;\leq\; \mathop{\mathbb{P}}_{\chi}\left[\left|\langle \chi, \chi_z \rangle_{\mu_1} - \mathop{\mathbb{E}}_{\chi}\left[\langle \chi, \chi_z \rangle_{\mu_1}\right]\right| \geq \frac{\varepsilon_0}{2}\right] \;\leq\; \exp\left(-\varepsilon_0^2 \cdot n/8\right).$$

Thus, given a good sample χ satisfying $\langle \chi, \chi_z \rangle_{\mu_1} \geq 1/2 + \varepsilon_0$, we can recover the above $z \in \mathcal{C}_0$ such that $\mathrm{dsum}_W(z) \in \mathcal{L}_\beta$, via the unique decoding algorithm for $\mathcal{C}_0$. Also, given the right $\overline{f}$, we sample a good χ with probability at least $1 - \exp(-\varepsilon_0^2 \cdot n/8)$. A union bound then gives

$$\mathbb{P}\left[\mathcal{L} = \mathcal{L}_\beta\right] \;\geq\; 1 \;-\; \left|\mathcal{L}_\beta\right| \cdot \exp(-\varepsilon_0^2 \cdot n/8).$$

Using $\beta \geq \sqrt{\varepsilon}$, we get that $\left|\mathcal{L}_\beta\right| \leq (1/\varepsilon)$ by the Johnson bound, which yields the desired probability bound.

Running time. Using Theorem 4.11, the decomposition h can be computed in time $\tilde{O}(C_{\beta,k} \cdot |W|)$. Given the functions $f_1, \ldots, f_r$ forming the decomposition h, the factor $\mathcal{B}$ can be computed in time $O(2^r \cdot n)$. For a chosen $\overline{f}$ in the sampling step, a sample χ can be computed in time $O(n)$, and the decoding problem for the corresponding $\tilde{z}$ can be solved in time $\mathcal{T}_0$. Also, the distance $\Delta(\tilde{y}, \mathrm{dsum}_W(z))$ can be computed in time $O(|W|)$. Since the total number of sampling steps is at most $(3/\eta)^{|B|}$ and the number of functions in the decomposition h is $O(k^3/\beta^2)$ from Theorem 4.11, we get that the total number of sampling steps is $(6/\varepsilon_0)^{2^{O(k^3/\beta^2)}}$. Thus, the total running time is bounded by $\tilde{O}(C_{k,\beta,\varepsilon_0} \cdot (|W| + \mathcal{T}_0))$, where $C_{k,\beta,\varepsilon_0} = (6/\varepsilon_0)^{2^{O(k^3/\beta^2)}}$. ∎

6 Near-linear Time Decoding of Ta-Shma's Codes

We now proceed to prove our main result, namely Theorem 1.1, which establishes a near-linear time *unique* decoding algorithm for Ta-Shma's codes [TS17]. It will follow from the regularity based list decoding algorithm for direct sum codes, Theorem 5.1, applied to the decoding of a slight modification of Ta-Shma's construction from [JQST20] that yields a splittable collection of tuples for the direct sum.

Theorem 1.1 (Near-linear Time Unique Decoding). *For every $\varepsilon > 0$ sufficiently small, there are explicit binary linear Ta-Shma codes $\mathcal{C}_{N,\varepsilon,\alpha} \subseteq \mathbb{F}_2^N$ for infinitely many values $N \in \mathbb{N}$ with*

(i) distance at least $1/2 - \varepsilon/2$ (actually ε-balanced),

(ii) rate $\Omega(\varepsilon^{2+\alpha})$ where $\alpha = O(1/(\log_2(1/\varepsilon))^{1/6})$, and

(iii) an $r(\varepsilon) \cdot \tilde{O}(N)$ time unique decoding algorithm that that decodes within radius $1/4 - \varepsilon/4$ and works with high probability,

where $r(\varepsilon) = \exp(\exp(\operatorname{polylog}(1/\varepsilon)))$.

We now state the properties and guarantees needed in our work of this slightly modified version of Ta-Shma's direct sum construction of near optimal ε-balanced codes. To make the decoding task more transparent, we will additionally require the base code in Ta-Shma's construction have the following technical property.

Definition 6.1. *We say that a code has symbol multiplicity $m \in \mathbb{N}$ if it can be obtained from another code by repeating each symbol of its codeword m times.*

Theorem A.1. *[Ta-Shma's Codes (implicit in [TS17])] Let $c > 0$ be an universal constant. For every $\varepsilon > 0$ sufficiently small, there exists $k = k(\varepsilon)$ satisfying $\Omega(\log(1/\varepsilon)^{1/3}) \leq k \leq O(\log(1/\varepsilon))$, $\varepsilon_0 = \varepsilon_0(\varepsilon) > 0$, and positive integer $m = m(\varepsilon) \leq (1/\varepsilon)^{o(1)}$ such that Ta-Shma's construction yields a collection of τ-splittable tuples $W = W(k) \subseteq [n]^k$ satisfying:*

(i) For every linear ε_0-balanced code $\mathcal{C}_0 \subseteq \mathbb{F}_2^n$ with symbol multiplicity m, the direct sum code $\operatorname{dsum}_W(\mathcal{C}_0)$ is:

  (i.1) ε-balanced (parity sampling).

  (i.2) if $\mathcal{C}_0$ has rate $\Omega(\varepsilon_0^c/m)$, then $\operatorname{dsum}_W(\mathcal{C}_0)$ has rate $\Omega(\varepsilon^{2+o(1)})$ (near optimal rate)

(ii) $\tau \leq \exp(-\Theta(\log(1/\varepsilon)^{1/6}))$ (splittability).

(iii) W is constructible in $\operatorname{poly}(|W|)$ time (explicit construction).

Ta-Shma's construction is based on a generalization of the zig-zag product of Reingold, Vadhan and Wigderson [RVW00]. To make the exposition more self-contained, we recall the slight modification from [JQST20] in Appendix A, but it is not exhaustive exposition. The interested reader is referred to Ta-Shma [TS17] for the original construction for aspects not covered here.

Ta-Shma's code construction requires an ε_0-balanced base code $\mathcal{C}_0 \subseteq \mathbb{F}_2^n$ whose distance will be amplified by taking the direct sum with a carefully chosen collection of tuples W yielding an ε-balanced code $\mathcal{C} = \operatorname{dsum}_W(\mathcal{C}_0)$. Since we our goal is to achieve near-linear time encoding and decoding of $\mathcal{C}$, we take an "off-the-shelf" base code $\mathcal{C}_0$ that is linear time encodable and decodable (near-linear time also suffices). A convenient choice is the linear binary code family of Guruswami–Indyk [GI05] that can be encoded and decoded in linear time. The rate versus distance trade-off is at the so-called Zyablov bound. In particular, it yields codes of distance $1/2 - \varepsilon_0$ with rate $\Omega(\varepsilon_0^3)$, but for our applications rate $\operatorname{poly}(\varepsilon_0)$ suffices (or with some extra steps even any rate depending only on ε_0 suffices, see Remark 6.5). We will use Corollary 6.2 implicit in [GI05].

Corollary 6.2. *[Implicit in Guruswami–Indyk [GI05]] For every $\varepsilon_0 > 0$, there exists a family of ε_0-balanced binary linear codes $\mathcal{C}_0 \subseteq \mathbb{F}_2^n$ of rate $\Omega(\varepsilon_0^3)$ which can be encoded in $O_{\varepsilon_0}(n)$ time and can be decoded in $O(\exp(\mathrm{poly}(1/\varepsilon_0)) \cdot n)$ time from up to a fraction $1/4 - \varepsilon_0$ of errors. Furthermore, every code in the family is explicitly specified given a binary linear code of blocklength $\mathrm{poly}(1/\varepsilon_0)$ which can be constructed in probabilistic $O(\mathrm{poly}(1/\varepsilon_0))$ or deterministic $2^{O(\mathrm{poly}(1/\varepsilon_0))}$ time.*

We first prove the (gentle) *list* decoding result of Ta-Shma's codes.

Theorem 1.2 (Near-linear Time Gentle List Decoding)**.** *For every $\varepsilon > 0$ sufficiently small, there are explicit binary linear Ta-Shma codes $\mathcal{C}_{N,\varepsilon,\alpha} \subseteq \mathbb{F}_2^N$ for infinitely many values $N \in \mathbb{N}$ with*

(i) distance at least $1/2 - \varepsilon/2$ (actually ε-balanced),

(ii) rate $\Omega(\varepsilon^{2+\alpha})$ where $\alpha = O(1/(\log_2(1/\varepsilon))^{1/6})$, and

(iii) an $r(\varepsilon) \cdot \tilde{O}(N)$ time list decoding algorithm that decodes within radius $1/2 - 2^{-\Theta((\log_2(1/\varepsilon))^{1/6})}$ and works with high probability,

where $r(\varepsilon) = \exp(\exp(\mathrm{poly}(1/\varepsilon)))$.

Proof. We start by dealing with a simple technical issue of making the base code in Ta-Shma's construction have the required symbol multiplicity. Let $\mathcal{C}_0' \subseteq \mathbb{F}_2^{n'}$ be an ε_0-balanced code from Corollary 6.2 which we will use to obtain a base code in Ta-Shma's construction where $\varepsilon_0 > 0$ is a suitable value prescribed by this construction.

Ta-Shma's construction then takes $\mathcal{C}_0' \subseteq \mathbb{F}_2^{n'}$ and forms a new code $\mathcal{C}_0 \subseteq \mathbb{F}_2^n$ by repeating each codeword symbol $m \leq (1/\varepsilon)^{o(1)}$ times. By Claim 6.6, $\mathcal{C}_0$ is an ε_0-balanced code that can be unique decoded within the same (fractional) radius of $\mathcal{C}_0'$ in time $\mathcal{T}_0(n) = r \cdot \mathcal{T}_0'(n') + \widetilde{O}(r^2 \cdot n')$, where $\mathcal{T}_0(n)'$ is the running time of an unique decoder for $\mathcal{C}_0'$. Since by Corollary 6.2 $\mathcal{T}_0(n') = O(\exp(\mathrm{poly}(1/\varepsilon_0)) \cdot n')$ and $\varepsilon_0 \gg \varepsilon$, the decoding time of $\mathcal{C}_0$ becomes $\mathcal{T}_0(n) = O(\exp(\mathrm{poly}(1/\varepsilon)) \cdot n)$.

Let $W = W(k)$ be a collection of tuples from Ta-Shma's construction Theorem A.1 so that $\mathcal{C} = \mathrm{dsum}_W(\mathcal{C}_0)$ is ε-balanced, $\tau \leq \exp(-\Theta(\log(1/\varepsilon)^{1/6}))$ and $k = \Omega(\log(1/\varepsilon)^{1/3})$. We will invoke our list decoding algorithm Theorem 5.1 whose list decoding radius $1/2 - \beta$ has to satisfy

$$\beta \;\geq\; \max\left\{\sqrt{\varepsilon},\;\left(2^{20} \cdot \tau \cdot k^3\right)^{1/2},\, 2 \cdot \left(\frac{1}{2} + 2\varepsilon_0\right)^{k/2}\right\}.$$

Using our values of τ and k together with the fact that $\varepsilon_0 < 1$ is bounded away form 1 by a constant amount gives

$$\beta \;\geq\; \max\left\{\sqrt{\varepsilon},\,,\;\exp(-\Theta((\log(1/\varepsilon))^{1/6})), \exp(-\Theta((\log(1/\varepsilon))^{1/3}))\right\}.$$

Hence, we can take $\beta = \exp(-\Theta(\log(1/\varepsilon)^{1/6}))$. Now, we compute the list decoding running proving a (crude) upper bound on its dependence on ε. By Theorem 5.1, the list decoding time

$$\tilde{O}(C_{\beta,k,\varepsilon_0} \cdot (|W| + \mathcal{T}_0(n))),$$

where $C_{k,\beta,\varepsilon_0} = (6/\varepsilon_0)^{2^{O(k^3/\beta^2)}}$. For our choices of parameters, this decoding time can be (crudely) bounded by $\tilde{O}(\exp(\exp(\mathrm{poly}(1/\varepsilon))) \cdot N)$. ∎

The gentle *list* decoding theorem above readily implies our main result for *unique* decoding if we are only interested in $\widetilde{O}_\varepsilon(N)$ decoding time without a more precise dependence on ε. We prove our main result, Theorem 1.1, for *unique* decoding making more precise the dependence of the running time on ε.

Proof. Proof of Theorem 1.1 We proceed as in the proof of Theorem 1.2 expect that we take $\beta = 1/4$ in the list decoding radius $1/2 - \beta$ so that by performing list decoding we can recover all codewords in the unique decoding radius of the corrupted codeword regardless of the bias of the code $\mathcal{C}_{N,\varepsilon,\alpha}$.

We now recompute the running time. By Theorem 5.1, the list decoding time

$$\tilde{O}(C_{\beta,k,\varepsilon_0} \cdot (|W| + \mathcal{T}_0(n))),$$

where $C_{k,\beta,\varepsilon_0} = ({}^{6}/_{\varepsilon_0})^{2^{O(k^3/\beta^2)}}$. For our choices of parameters, this decoding time can be (crudely) bounded by $\tilde{O}(\exp(\exp(\text{polylog}(1/\varepsilon))) \cdot N)$. ■

6.1 Choosing the Base Code

We now describe the (essentially) "off-the-shelf" base codes from Guruswami and Indyk [GI05] which we use in Ta-Shma's construction. We will need to prove that balanced codes can be easily obtained from [GI05]. The argument is quite simple and borrows from standard considerations related to the Zyablov and Gilbert–Varshamov bounds.

Corollary 6.2. *[Implicit in Guruswami–Indyk [GI05]] For every $\varepsilon_0 > 0$, there exists a family of ε_0-balanced binary linear codes $\mathcal{C}_0 \subseteq \mathbb{F}_2^n$ of rate $\Omega(\varepsilon_0^3)$ which can be encoded in $O_{\varepsilon_0}(n)$ time and can be decoded in $O(\exp(\text{poly}(1/\varepsilon_0)) \cdot n)$ time from up to a fraction $1/4 - \varepsilon_0$ of errors. Furthermore, every code in the family is explicitly specified given a binary linear code of blocklength $\text{poly}(1/\varepsilon_0)$ which can be constructed in probabilistic $O(\text{poly}(1/\varepsilon_0))$ or deterministic $2^{O(\text{poly}(1/\varepsilon_0))}$ time.*

Theorem 6.3 (Guruswami–Indyk [GI05], *cf.*,Theorem 5)**.** *For every $\gamma > 0$ and for every $0 < R < 1$, there exists a family of binary linear* concatenated *codes of rate R, which can be encoded in linear time and can be decoded in linear time from up to a fraction e of errors, where*

$$e \;\geq\; \max_{R<r<1} \frac{(1 - r - \gamma) \cdot H_2^{-1}(1 - R/r)}{2}. \tag{4}$$

$H_2^{-1}(x)$ is defined as the unique ρ in the range $0 \leq \rho \leq 1/2$ satisfying $H_2(\rho) = x$. Every code in the family is explicitly specified given a constant sized binary linear code which can be constructed in probabilistic $O(\log(1/\gamma)R^{-1}/\gamma^4)$ or deterministic $2^{O(\log(1/\gamma)R^{-1}/\gamma^4)}$ time [8].

As stated the codes in Theorem 6.3 are not necessarily balanced. We will see shortly that this can be easily achieved by choosing balanced inner codes in the concatenated code construction of Guruswami–Indyk [GI05]. To compute bounds on the parameters, we will use the following property about binary entropy.

[8] Note that dependence $\log(1/\gamma)R^{-1}/\gamma^4$ is slightly worse than that claimed in [GI05], but not qualitatively relevant here nor in [GI05].

Fact 6.4 ([GRS19],*cf.*,Lemma 3.3.7 abridged)**.** *Let H_2^{-1} be the inverse of the restriction of H_2 to $[0,1/2]$ (where H_2 is bijective). For every small enough $\varepsilon > 0$,*

$$H_2^{-1}(x - \varepsilon^2/C_2) \;\geq\; H_2^{-1}(x) - \varepsilon,$$

where C_2 is a constant.

Proof of Corollary 6.2. To achieve a final binary code of rate R, Guruswami and Indyk [GI05] concatenate an outer code of rate $r > R$ and distance $1 - r - \gamma$ (over a non-binary alphabet of size $O_\gamma(1)$) with an inner binary linear code of rate R/r at the GV bound whose distance $\rho \in [0,1/2]$ satisfy $R/r = 1 - H_2(\rho)$ (since it is at the GV bound), or equivalently $\rho = H_2^{-1}(1 - R/r)$. By choosing $\gamma = \Theta(\varepsilon_0)$ and $R = \Theta(\varepsilon_0^3)$ in Theorem 6.3, the decoding error e can be lower bounded by letting $r = \Theta(\varepsilon_0)$ so that Fact 6.4 implies that Eq. (4) becomes

$$e \;\geq\; \max_{R<r<1} \frac{(1-r-\gamma)\cdot H_2^{-1}(1-R/r)}{2} \;\geq\; \frac{1}{4} - \varepsilon_0.$$

To obtain codes that are ε_0-balanced, we require that the inner codes used in this code concatenation not only lie on the Gilbert–Varshamov bound but are also balanced. It is well known that with high probability a random binary linear code at the GV bound designed to have minimum distance $1/2 - \gamma/2$ also has maximum distance at most $1/2 + \gamma/2$, i.e., the code is γ-balanced. Therefore, we assume that our inner codes are balanced.

For our concrete choices of parameters, $\rho = 1/2 - \Theta(\varepsilon_0)$ and we also require the inner code to be $\Theta(\varepsilon_0)$-balanced. Note that any non-zero codeword of the concatenated is obtained as follows: each of the $\geq (1 - r - \gamma)$ non-zero symbols of the outer codeword is replaced by an inner codeword of bias bias $\Theta(\varepsilon_0)$ and the remaining $\leq r + \gamma$ zero symbols are mapped to zero (since the inner code is linear). Hence, the bias of the concatenated codeword is at most

$$(1-r-\gamma)\cdot\Theta(\varepsilon_0) \;+\; 1\cdot(r+\gamma),$$

which can be taken to be ε_0 by suitable choices of hidden constants. ∎

Remark 6.5. *Guruswami–Indyk [GI05] codes have several nice properties making them a convenient choice for base codes in Ta-Shma's construction, but they are not crucial here. We observe that for our purposes we could have started with any family of good binary linear codes admitting near-linear time encoding and decoding. From this family, we could boost its distance using a simpler version of Ta-Shma's construction (rounds I and II of [JQST20][Section 8]) and our near-linear time decoder Theorem 5.1 for direct sum. This would result in an alternative family of linear binary ε_0-balanced codes of rate $\Omega(\varepsilon_0^{2+\alpha})$, for some arbitrarily small constant $\alpha > 0$, that can be encoded and decoded in near-linear time. We also point out that for these base codes any rate* $\mathrm{poly}(\varepsilon_0)$ *suffices our purposes.*

To handle the technical requirement of a base code in Ta-Shma's construction having a symbol multiplicity property (*cf.*, Definition 6.1), we use the following observation.

Claim 6.6. *Let $\mathcal{C}_0 \subseteq \mathbb{F}_2^n$ be an ε_0-balanced linear code of dimension D_0. Suppose that $\mathcal{C}_0$ is uniquely decodable within (fractional) radius $\delta_0 \in (0,1]$ in time $\mathcal{T}_0(n)$. Let $m \in \mathbb{N}$ and $\mathcal{C} \subseteq \mathbb{F}_2^{m\cdot n}$ be the code formed by replicating m times each codeword from $\mathcal{C}_0$, i.e.,*

$$\mathcal{C} := \{z_1 \cdots z_m \in \mathbb{F}_2^{m\cdot n} \mid z_1 = \cdots = z_m \in \mathcal{C}_0\}.$$

Then, $\mathcal{C}$ is an ε_0-balanced linear code of dimension D_0 that can be uniquely decoded within (fractional) radius δ_0 in time $m \cdot \mathcal{T}_0(n) + \widetilde{O}(m^2 \cdot n)$.

Proof. The only non-immediate property is the unique decoding guarantees of $\mathcal{C}$. Given $\tilde{y} \in \mathbb{F}_2^{m \cdot n}$ within δ_0 (relative) distance of $\mathcal{C}$. Let β_i be the fraction of errors in the ith $\mathbb{F}_2^n$ component $\tilde{y}$. By assumption $\mathbb{E}_{i \in [m]} \beta_i \le \delta_0$, so there is at least one of such component that can be correctly uniquely decoded. We issue unique decoding calls for $\mathcal{C}_o$ on each component $i \in [m]$. For each successful decoding say $z \in \mathcal{C}_0$, we let $y = z \ldots z \in \mathbb{F}_2^{m \cdot n}$ and check whether $\Delta(\tilde{y}, y) \le \delta_0$ returning y if this succeeds. Finally, observe that this procedure indeed takes at most the claimed running time. ■

Acknowledgement

We thank Dylan Quintana for stimulating discussions during the initial phases of this project.

References

[ABN+92] N. Alon, J. Bruck, J. Naor, M. Naor, and R. Roth. Construction of asymptotically good, low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on Information Theory*, 28:509–516, 1992. 2

[AGHP92] N. Alon, O. Goldreich, J. Håstad, and R. Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures and Algorithms*, 3(3):289–304, 1992. 49

[AJQ+20] Vedat Levi Alev, Fernando Granha Jeronimo, Dylan Quintana, Shashank Srivastava, and Madhur Tulsiani. List decoding of direct sum codes. In *Proceedings of the 31st ACM-SIAM Symposium on Discrete Algorithms*, pages 1412–1425. SIAM, 2020. 1, 3, 4, 10

[AJT19] Vedat Levi Alev, Fernando Granha Jeronimo, and Madhur Tulsiani. Approximating constraint satisfaction problems on high-dimensional expanders. In *Proceedings of the 60th IEEE Symposium on Foundations of Computer Science*, pages 180–201, 2019. 5

[AK07] Sanjeev Arora and Satyen Kale. A combinatorial, primal-dual approach to semidefinite programs. In *Proceedings of the 39th ACM Symposium on Theory of Computing*, STOC '07, pages 227–236, 2007. 7, 12, 23, 25, 26, 28, 29

[AN04] Noga Alon and Assaf Naor. Approximating the cut-norm via grothendieck's inequality. In *Proceedings of the 36th ACM Symposium on Theory of Computing*, pages 72–80, 2004. 7, 12, 23, 25, 26, 27, 28

[Aro02] Sanjeev Arora. How NP got a new definition: a survey of probabilistically checkable proofs. In *Proceedings of the International Congress of Mathematicians*, pages 637–648, 2002. Volume 3. 2

[BL18] A. Bhowmick and S. Lovett. The list decoding radius for Reed–Muller codes over small fields. *IEEE Transactions on Information Theory*, 64(6):4382–4391, 2018. 5

[Bog12] Andrej Bogdanov. A different way to improve the bias via expanders. Lecture notes, April 2012. URL: `http://www.cse.cuhk.edu.hk/~andrejb/csc5060/notes/12L12.pdf`. 2

[BRS11] Boaz Barak, Prasad Raghavendra, and David Steurer. Rounding semidefinite programming hierarchies via global correlation. In *Proceedings of the 52nd IEEE Symposium on Foundations of Computer Science*, pages 472–481, 2011. 5

[Bub15] Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Found. Trends Mach. Learn.*, 8(3-4):231–357, November 2015. 14

[BV20] Greg Bodwin and Santosh Vempala. A unified view of graph regularity via matrix decompositions, 2020. `arXiv:1911.11868`. 3, 4, 5

[Cha16] Siu On Chan. Approximation resistance from pairwise-independent subgroups. *J. ACM*, 63(3), August 2016. 2

[COCF09] Amin Coja-Oghlan, Colin Cooper, and Alan Frieze. An efficient sparse regularity concept. In *Proceedings of the 20th ACM-SIAM Symposium on Discrete Algorithms*, SODA '09, page 207–216, 2009. 3

[DD19] Yotam Dikstein and Irit Dinur. Agreement testing theorems on layered set systems. In *Proceedings of the 60th IEEE Symposium on Foundations of Computer Science*, 2019. 5

[DDG$^+$15] Roee David, Irit Dinur, Elazar Goldenberg, Guy Kindler, and Igor Shinkar. Direct sum testing. ITCS '15, pages 327–336, New York, NY, USA, 2015. ACM. 2

[DDHRZ20] Yotam Dikstein, Irit Dinur, Prahladh Harsha, and Noga Ron-Zewi. Locally testable codes via high-dimensional expanders. *arXiv preprint arXiv:2005.01045*, 2020. 5

[DHK$^+$19] Irit Dinur, Prahladh Harsha, Tali Kaufman, Inbal Livni Navon, and Amnon Ta-Shma. List decoding with double samplers. In *Proceedings of the 30th ACM-SIAM Symposium on Discrete Algorithms*, pages 2134–2153, 2019. 5

[DK17] Irit Dinur and Tali Kaufman. High dimensional expanders imply agreement expanders. In *Proceedings of the 58th IEEE Symposium on Foundations of Computer Science*, pages 974–985, 2017. 2, 5

[DS14] Irit Dinur and David Steurer. Direct product testing. In *Proceedings of the 29th IEEE Conference on Computational Complexity*, CCC '14, pages 188–196, 2014. 2

[EK16] Shai Evra and Tali Kaufman. Bounded degree cosystolic expanders of every dimension. In *Proceedings of the 48th ACM Symposium on Theory of Computing*, pages 36–48. ACM, 2016. 5

[FK96] A. Frieze and R. Kannan. The regularity lemma and approximation schemes for dense problems. In *Proceedings of the 37th IEEE Symposium on Foundations of Computer Science*, 1996. 3, 5, 12

[GI01] Venkatesan Guruswami and Piotr Indyk. Expander-based constructions of efficiently decodable codes. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, pages 658–667, 2001. 2, 5

[GI05] V. Guruswami and P. Indyk. Linear-time encodable/decodable codes with near-optimal rate. *IEEE Transactions on Information Theory*, 51(10):3393–3400, 2005. 5, 34, 35, 36, 37

[Gil52] E.N. Gilbert. A comparison of signalling alphabets. *Bell System Technical Journal*, 31:504–522, 1952. 1

[GR06] Venkatesan Guruswami and Atri Rudra. Explicit capacity-achieving list-decodable codes. In *Proceedings of the 38th ACM Symposium on Theory of Computing*, pages 1–10, 2006. 2

[GRS19] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. Essential coding theory. Available at `https://cse.buffalo.edu/faculty/atri/courses/coding-theory/book/index.html`, 2019. 37

[GS11] Venkatesan Guruswami and Ali Kemal Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for graph partitioning and quadratic integer programming with psd objectives. In *FOCS*, pages 482–491, 2011. 5

[Gur04] Venkatesan Guruswami. Guest column: Error-correcting codes and expander graphs. *SIGACT News*, 35(3):25–41, September 2004. 5

[Gur09] Venkatesan Guruswami. List decoding of binary codes–a brief survey of some recent results. In *Coding and Cryptology*, pages 97–106. Springer Berlin Heidelberg, 2009. 1

[Gur10] Venkatesan Guruswami. Bridging Shannon and Hamming: List error-correction with optimal rate. In *ICM*, 2010. 1

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43(04):439–562, August 2006. 15

[IKW09] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson. New direct-product testers and 2-query PCPs. In *Proceedings of the 41st ACM Symposium on Theory of Computing*, STOC '09, pages 131–140, 2009. 2

[IW97] Russell Impagliazzo and Avi Wigderson. $P = BPP$ unless E has sub-exponential circuits. In *Proceedings of the 29th ACM Symposium on Theory of Computing*, pages 220–229, 1997. 2

[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 IEEE Symposium on Foundations of Computer Science*, 2020. 1, 3, 4, 10, 33, 34, 37, 42

[KR02] Y. Kohayakawa and V. Rödl. Szemerédi's regularity lemma and quasi-randomness. In *Recent advances in algorithms and combinatorics*. Springer, Berlin, 2002. URL: citeseer.ist.psu.edu/kohayakawa02szemeredis.html. 5

[KV09] Ravindran Kannan and Santosh Vempala. *Spectral algorithms*. Now Publishers Inc, 2009. 5

[LP20] Yin Tat Lee and Swati Padmanabhan. An $\widetilde{O}(m/\varepsilon^{3.5})$-cost algorithm for semidefinite programs with diagonal constraints. In *Conference on Learning Theory, COLT 2020, 9-12 July 2020, Virtual Event [Graz, Austria]*, volume 125, pages 3069–3119, 2020. 7, 12, 23, 25, 26, 28, 29

[MRRW77] R. McEliece, E. Rodemich, H. Rumsey, and L. Welch. New upper bounds on the rate of a code via the Delsarte-MacWilliams inequalities. *IEEE Transactions on Information Theory*, 23(2):157–166, 1977. 1

[OGT15] Shayan Oveis Gharan and Luca Trevisan. A new regularity lemma and faster approximation algorithms for low threshold rank graphs. *Theory of Computing*, 11(9):241–256, 2015. URL: http://www.theoryofcomputing.org/articles/v011a009, doi:10.4086/toc.2015.v011a009. 5

[RTTV08] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Dense subsets of pseudorandom sets. In *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*, 2008. 3, 5

[RVW00] O. Reingold, S. Vadhan, and A. Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *Proceedings of the 41st IEEE Symposium on Foundations of Computer Science*, 2000. 34, 43, 45

[RWZ20] N. Ron-Zewi, M. Wootters, and G. Zémor. Linear-time erasure list-decoding of expander codes. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 379–383, 2020. 5

[SS96] M. Sipser and D. Spielman. Expander codes. *IEEE Transactions on Information Theory*, 42(6):1710–1722, 1996. Preliminary version in *Proc. of FOCS'94*. 5

[TS17] Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *Proceedings of the 49th ACM Symposium on Theory of Computing*, STOC 2017, pages 238–251, New York, NY, USA, 2017. ACM. 1, 9, 10, 33, 34, 42, 46, 49

[TTV09] L. Trevisan, M. Tulsiani, and S. Vadhan. Boosting, regularity and efficiently simulating every high-entropy distribution. In *Proceedings of the 24th IEEE Conference on Computational Complexity*, 2009. 3, 4, 5, 6

[Vad12] Salil P. Vadhan. *Pseudorandomness*. Now Publishers Inc., 2012. 1

[Var57] R.R. Varshamov. Estimate of the number of signals in error correcting codes. *Doklady Akademii Nauk SSSR*, 117:739–741, 1957. 1

A Properties of Ta-Shma's Construction

The goal of this section is to provide a reasonably self-contained compilation of the properties of the slightly modified version of Ta-Shma code construction [TS17] from [JQST20]. The properties we need are collected in Theorem A.1.

Theorem A.1. *[Ta-Shma's Codes (implicit in [TS17])] Let $c > 0$ be an universal constant. For every $\varepsilon > 0$ sufficiently small, there exists $k = k(\varepsilon)$ satisfying $\Omega(\log(1/\varepsilon)^{1/3}) \le k \le O(\log(1/\varepsilon))$, $\varepsilon_0 = \varepsilon_0(\varepsilon) > 0$, and positive integer $m = m(\varepsilon) \le (1/\varepsilon)^{o(1)}$ such that Ta-Shma's construction yields a collection of τ-splittable tuples $W = W(k) \subseteq [n]^k$ satisfying:*

- *(i) For every linear ε_0-balanced code $\mathcal{C}_0 \subseteq \mathbb{F}_2^n$ with symbol multiplicity m, the direct sum code $\mathrm{dsum}_W(\mathcal{C}_0)$ is:*
  - *(i.1) ε-balanced (parity sampling).*
  - *(i.2) if $\mathcal{C}_0$ has rate $\Omega(\varepsilon_0^c/m)$, then $\mathrm{dsum}_W(\mathcal{C}_0)$ has rate $\Omega(\varepsilon^{2+o(1)})$ (near optimal rate)*
- *(ii) $\tau \le \exp(-\Theta(\log(1/\varepsilon)^{1/6}))$ (splittability).*
- *(iii) W is constructible in $\mathrm{poly}(|W|)$ time (explicit construction).*

We first recall the s-wide replacement product in Appendix A.1, then describe Ta-Shma's original construction based on it in Appendix A.2, describe our modification to obtain splittability in Appendix A.3, derive the splittability property in Appendix A.4, and finally choose parameters in terms of desired bias ε of the code we construct in Appendix A.5. We refer the reader to [TS17] for formal details beyond those we actually need here.

A.1 The s-wide Replacement Product

Ta-Shma's code construction is based on the so-called s-wide replacement product [TS17]. This is a derandomization of random walks on a graph G that will be defined via a product operation of G with another graph H (see Definition A.3 for a formal definition). We will refer to G as the *outer* graph and H as the *inner* graph in this construction.

Let G be a d_1-regular graph on vertex set $[n]$ and H be a d_2-regular graph on vertex set $[d_1]^s$, where s is any positive integer. Suppose the neighbors of each vertex of G are labeled 1, 2, ..., d_1. For $v \in V(G)$, let $v_G[j]$ be the j-th neighbor of v. The s-wide replacement product is defined by replacing each vertex of G with a copy of H, called a "cloud". While the edges within each cloud are determined by H, the edges between clouds are based on the edges of G, which we will define via operators $\mathsf{G}_0, \mathsf{G}_1, \ldots, \mathsf{G}_{s-1}$. The i-th operator G_i specifies one inter-cloud edge for each vertex $(v, (a_0, \ldots, a_{s-1})) \in V(G) \times V(H)$, which goes to the cloud whose G component is $v_G[a_i]$, the neighbor of v in G indexed by the i-th coordinate of the H component. (We will resolve the question of what happens to the H component after taking such a step momentarily.)

Walks on the s-wide replacement product consist of steps with two different parts: an intra-cloud part followed by an inter-cloud part. All of the intra-cloud substeps simply move to a random neighbor in the current cloud, which corresponds to applying the operator $\mathsf{I} \otimes \mathsf{A}_H$, where A_H is the normalized adjacency matrix of H. The inter-cloud substeps are all deterministic, with the first moving according to G_0, the second according to G_1, and so on, returning to G_0 for step number $s+1$. The operator for such a walk taking $k-1$ steps on the s-wide replacement product is

$$\prod_{i=0}^{k-2} \mathsf{G}_{i \bmod s}(\mathsf{I} \otimes \mathsf{A}_H).$$

Observe that a walk on the s-wide replacement product yields a walk on the outer graph G by recording the G component after each step of the walk. The number of $(k-1)$-step walks on the s-wide replacement product is

$$|V(G)| \cdot |V(H)| \cdot d_2^{k-1} = n \cdot d_1^s \cdot d_2^{k-1},$$

since a walk is completely determined by its intra-cloud steps. If d_2 is much smaller than d_1 and k is large compared to s, this is less than nd_1^{k-1}, the number of $(k-1)$-step walks on G itself. Thus the s-wide replacement product will be used to simulate random walks on G while requiring a reduced amount of randomness (of course this simulation is only possible under special conditions, namely, when we are uniformly distributed on each cloud).

To formally define the s-wide replacement product, we must consider the labeling of neighbors in G more carefully.

Definition A.2 (Rotation Map)**.** *Suppose G is a d_1-regular graph on $[n]$. For each $v \in [n]$ and $j \in [d_1]$, let $v_G[j]$ be the j-th neighbor of v in G. Based on the indexing of the neighbors of each vertex, we define the rotation map* [9] $\mathrm{rot}_G\colon [n] \times [d_1] \to [n] \times [d_1]$ *such that for every $(v, j) \in [n] \times [d_1]$,*

$$\mathrm{rot}_G((v, j)) = (v', j') \Leftrightarrow v_G[j] = v' \text{ and } v'_G[j'] = v.$$

Furthermore, if there exists a bijection $\varphi\colon [d_1] \to [d_1]$ such that for every $(v, j) \in [n] \times [d_1]$,

$$\mathrm{rot}_G((v, j)) = (v_G[j], \varphi(j)),$$

then we call rot_G locally invertible*.*

If G has a locally invertible rotation map, the cloud label after applying the rotation map only depends on the current cloud label, not the vertex of G. In the s-wide replacement product, this corresponds to the H component of the rotation map only depending on a vertex's H component, not its G component. We define the s-wide replacement product as described before, with the inter-cloud operator G_i using the i-th coordinate of the H component, which is a value in $[d_1]$, to determine the inter-cloud step.

Definition A.3 (s-wide replacement product)**.** *Suppose we are given the following:*

- *A d_1-regular graph $G = ([n'], E)$ together with a locally invertible rotation map* $\mathrm{rot}_G\colon [n'] \times [d_1] \to [n'] \times [d_1]$.
- *A d_2-regular graph $H = ([d_1]^s, E')$.*

[9]This kind of map is denoted rotation map in the zig-zag terminology [RVW00].

And we define:

- *For* $i \in \{0, 1, \ldots, s-1\}$, *we define* $\mathrm{Rot}_i \colon [n'] \times [d_1]^s \to [n'] \times [d_1]^s$ *as, for every* $v \in [n']$ *and* $(a_0, \ldots, a_{s-1}) \in [d_1]^s$,

$$\mathrm{Rot}_i((v, (a_0, \ldots, a_{s-1}))) := (v', (a_0, \ldots, a_{i-1}, a_i', a_{i+1}, \ldots, a_{s-1})),$$

  where $(v', a_i') = \mathrm{rot}_G(v, a_i)$.

- *Denote by* G_i *the operator realizing* Rot_i *and let* A_H *be the normalized random walk operator of* H*. Note that* G_i *is a permutation operator corresponding to a product of transpositions.*

Then $k-1$ *steps of the* s*-wide replacement product are given by the operator*

$$\prod_{i=0}^{k-2} \mathsf{G}_{i \bmod s}(\mathsf{I} \otimes \mathsf{A}_H).$$

Ta-Shma instantiates the s-wide replacement product with an outer graph G that is a Cayley graph, for which locally invertible rotation maps exist generically.

Remark A.4. *Let* R *be a group and* $A \subseteq R$ *where the set* A *is closed under inversion. For every Cayley graph* $\mathrm{Cay}(R, A)$*, the map* $\varphi \colon A \to A$ *defined as* $\varphi(g) = g^{-1}$ *gives rise to the locally invertible rotation map*

$$\mathrm{rot}_{\mathrm{Cay}(R,A)}((r, a)) = (r \cdot a, a^{-1}),$$

for every $r \in R$, $a \in A$.

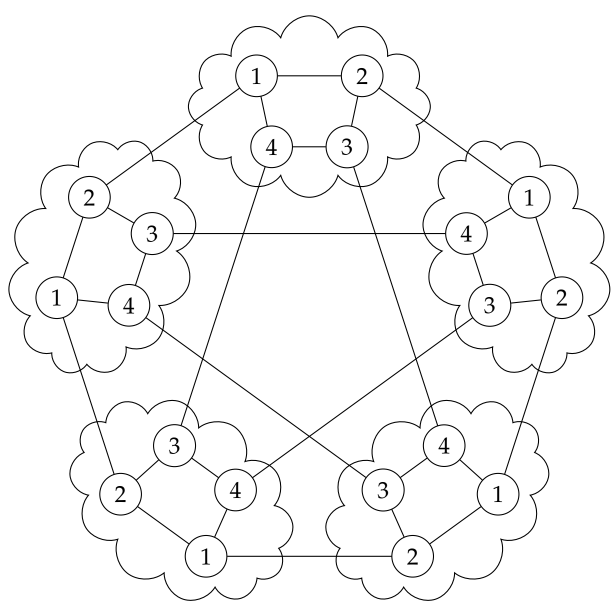

Figure 1: An example of the 1-wide replacement product with outer graph $G = K_5$ and inner graph $H = C_4$. Vertices are labeled by their H components. Note that the rotation map is locally invertible, with $\varphi(1) = 2$, $\varphi(2) = 1$, $\varphi(3) = 4$, and $\varphi(4) = 3$.

A.2 The Construction

Let $n' = |V(G)|, m = d_1^s = |V(H)|$ and $n = n' \cdot m = |V(G) \times V(H)|$. Ta-Shma's code construction works by starting with a constant bias code $\mathcal{C}_0'$ in $\mathbb{F}_2^{n'}$, repeating each codeword $m = d_1^s$ times to get a new ε_0-biased code $\mathcal{C}_0$ in $\mathbb{F}_2^n$, and boosting $\mathcal{C}_0$ to arbitrarily small bias using direct sum liftings. Recall that the direct sum lifting is based on a collection $W(k) \subseteq [n]^k$, which Ta-Shma obtains using $k-1$ steps of random walk on the s-wide replacement product of two regular expander graphs G and H. The graph G is on n' vertices and other parameters like degrees d_1 and d_2 of G and H respectively are chosen based on target code parameters.

To elaborate, every $k-1$ length walk on the replacement product gives a sequence of k vertices in the replacement product graph, which can be seen as an element of $[n]^k$. This gives the collection $W(k)$ with $|W(k)| = n' \cdot d_1^s \cdot d_2^{k-1}$ which means the rate of lifted code is smaller than the rate of $\mathcal{C}_0'$ by a factor of $d_1^s d_2^{k-1}$. However, the collection $W(k)$ is a parity sampler and this means that the bias decreases (or the distance increases) from that of $\mathcal{C}_0$. The relationship between this decrease in bias and decrease in rate with some careful parameter choices allows Ta-Shma to obtain nearly optimal ε-balanced codes.

A.3 Tweaking the Construction

Recall the first s steps in Ta-Shma's construction are given by the operator

$$\mathsf{G}_{s-1}(\mathsf{I} \otimes \mathsf{A}_H)\mathsf{G}_{s-2} \cdots G_1(\mathsf{I} \otimes \mathsf{A}_H)\mathsf{G}_0(\mathsf{I} \otimes \mathsf{A}_H).$$

Naively decomposing the above operator into the product of operators $\prod_{i=0}^{s-1} \mathsf{G}_i(\mathsf{I} \otimes \mathsf{A}_H)$ is not good enough to obtain the *splittability* property which would hold provided $\sigma_2(\mathsf{G}_i(\mathsf{I} \otimes \mathsf{A}_H))$ was small for every i in $\{0, \ldots, s-1\}$. However, each $\mathsf{G}_i(\mathsf{I} \otimes \mathsf{A}_H)$ has $|V(G)|$ singular values equal to 1 since G_i is an orthogonal operator and $(\mathsf{I} \otimes \mathsf{A}_H)$ has $|V(G)|$ singular values equal to 1. To avoid this issue we will tweak the construction to be the following product

$$\prod_{i=0}^{s-1}(\mathsf{I} \otimes \mathsf{A}_H)\mathsf{G}_i(\mathsf{I} \otimes \mathsf{A}_H).$$

The operator $(\mathsf{I} \otimes \mathsf{A}_H)\mathsf{G}_i(\mathsf{I} \otimes \mathsf{A}_H)$ is exactly the walk operator of the zig-zag product $G \textcircled{z} H$ of G and H with a rotation map given by the (rotation map) operator G_i. This tweaked construction is slightly simpler in the sense that $G \textcircled{z} H$ is an undirected graph. We know by the zig-zag analysis that $(\mathsf{I} \otimes \mathsf{A}_H)\mathsf{G}_i(\mathsf{I} \otimes \mathsf{A}_H)$ is expanding as long G and H are themselves expanders. More precisely, we have a bound that follows from [RVW00].

Fact A.5. *Let G be an outer graph and H be an inner graph used in the s-wide replacement product. For any integer $0 \leq i \leq s-1$,*

$$\sigma_2((I \otimes \mathsf{A}_H)G_i(I \otimes \mathsf{A}_H)) \leq \sigma_2(G) + 2 \cdot \sigma_2(H) + \sigma_2(H)^2.$$

This bound will imply *splittability* as shown in Appendix A.4. We will need to argue that this modification still preserves the correctness of the parity sampling and that it can be achieved with similar parameter trade-offs.

The formal definition of a length-t walk on this slightly modified construction is given below.

Definition A.6. *Let $k \in \mathbb{N}$, G be a d_1-regular graph and H be a d_2-regular graph on d_1^s vertices. Given a starting vertex $(v, u) \in V(G) \times V(H)$, a $(k-1)$-step walk on the tweaked s-wide replacement product of G and H is a tuple $((v_1, u_1), \dots, (v_k, u_k)) \in (V(G) \times V(H))^k$ such that*

- *$(v_1, u_1) = (v, u)$, and*
- *for every $1 \le i < k$, we have (v_i, u_i) adjacent to (v_{i+1}, u_{i+1}) in $(\mathsf{I} \otimes \mathsf{A}_H)\mathsf{G}_{(i-1) \bmod s}(\mathsf{I} \otimes \mathsf{A}_H)$.*

Note that each $(\mathsf{I} \otimes \mathsf{A}_H)\mathsf{G}_{(i-1) \bmod s}(\mathsf{I} \otimes \mathsf{A}_H)$ is a walk operator of a d_2^2-regular graph. Therefore, the starting vertex (v, u) together with a degree sequence $(m_1, \dots, m_k) \in [d_2^2]^{k-1}$ uniquely defines a $(k-1)$-step walk.

A.3.1 Parity Sampling

We argue informally why parity sampling still holds with similar parameter trade-offs. In particular, we formalize a key result underlying parity sampling and, in Appendix A.5, we compute the new trade-off between bias and rate in some regimes. In Appendix A.1, the definition of the original s-wide replacement product as a purely graph theoretic operation was given. Now, we explain how Ta-Shma used this construction for parity sampling obtaining codes near the GV bound.

For a word $z \in \mathbb{F}_2^{V(G)}$ in the base code, let P_z be the diagonal matrix, whose rows and columns are indexed by $V(G) \times V(H)$, with $(\mathsf{P}_z)_{(v,u),(v,u)} = (-1)^{z_v}$. Proving parity sampling requires analyzing the operator norm of the following product

$$\mathsf{P}_z \prod_{i=0}^{s-1} (\mathsf{I} \otimes \mathsf{A}_H)\mathsf{G}_i \mathsf{P}_z (\mathsf{I} \otimes \mathsf{A}_H), \tag{5}$$

when $\operatorname{bias}(z) \le \varepsilon_0$. Let $\mathbf{1} \in \mathbb{R}^{V(G) \times V(H)}$ be the all-ones vector, scaled to be of unit length under the ℓ_2 norm, and W be the collection of all $(t-1)$-step walks on the tweaked s-wide replacement product. Ta-Shma showed (and it is not difficult to verify) that

$$\operatorname{bias}\left(\operatorname{dsum}_W(z)\right) = \left| \left\langle \mathbf{1}, \mathsf{P}_z \prod_{i=0}^{k-2} (\mathsf{I} \otimes \mathsf{A}_H)\mathsf{G}_{i \bmod s} \mathsf{P}_z (\mathsf{I} \otimes \mathsf{A}_H) \mathbf{1} \right\rangle \right|.$$

The measure used in this inner product is the usual counting measure over $\mathbb{R}^{V(G) \times V(H)}$. From the previous equation, one readily deduces that

$$\operatorname{bias}\left(\operatorname{dsum}_W(z)\right) \le \sigma_1 \left(\mathsf{P}_z \prod_{i=0}^{s-1} (\mathsf{I} \otimes \mathsf{A}_H)\mathsf{G}_i \mathsf{P}_z (\mathsf{I} \otimes \mathsf{A}_H) \right)^{\lfloor (k-1)/s \rfloor}.$$

The key technical result obtained by Ta-Shma is the following, which is used to analyze the bias reduction as a function of the total number walk steps $k-1$. Here θ is a parameter used in obtaining explicit Ramanujan graphs.

Fact A.7 (Theorem 24 abridged [TS17]). *If H is a Cayley graph on $\mathbb{F}_2^{s \log d_1}$ and $\varepsilon_0 + 2 \cdot \theta + 2 \cdot \sigma_2(G) \le \sigma_2(H)^2$, then*

$$\left\| \prod_{i=0}^{s-1} \mathsf{P}_z \mathsf{G}_i (\mathsf{I} \otimes \mathsf{A}_H) \right\|_{\mathrm{op}} \le \sigma_2(H)^s + s \cdot \sigma_2(H)^{s-1} + s^2 \cdot \sigma_2(H)^{s-3},$$

where $\mathsf{P}_z \in \mathbb{R}^{(V(G)\times V(H))\times(V(G)\times V(H))}$ *is the* sign operator *of a* ε_0 *biased word* $z \in \mathbb{F}_2^{V(G)}$ *defined as a diagonal matrix with* $(P_z)_{(v,u),(v,u)} = (-1)^{z_v}$ *for every* $(v,u) \in V(G) \times V(H)$.

We reduce the analysis of Ta-Shma's tweaked construction to an analog of Fact A.7. In doing so, we only lose one extra step as shown below.

Corollary A.8. *If* H^2 *is a Cayley graph on* $\mathbb{F}_2^{s \log d_1}$ *and* $\varepsilon_0 + 2\cdot\theta + 2\cdot\sigma_2(G) \le \sigma_2(H)^4$*, then*

$$\left\|\prod_{i=0}^{s-1}(\mathsf{I}\otimes \mathsf{A}_H)\mathsf{P}_z\mathsf{G}_i(\mathsf{I}\otimes \mathsf{A}_H)\right\|_{\mathrm{op}} \le \sigma_2(H^2)^{s-1} + (s-1)\cdot\sigma_2(H^2)^{s-2} + (s-1)^2\cdot\sigma_2(H^2)^{s-4},$$

where P_z *is the* sign operator *of an* ε_0*-biased word* $z \in \mathbb{F}_2^{V(G)}$ *as in Fact A.7.*

Proof. We have

$$\begin{aligned}\left\|\prod_{i=0}^{s-1}(\mathsf{I}\otimes \mathsf{A}_H)\mathsf{P}_z\mathsf{G}_i(\mathsf{I}\otimes \mathsf{A}_H)\right\|_{\mathrm{op}} &\le \|(\mathsf{I}\otimes\mathsf{A}_H)\|_{\mathrm{op}}\left\|\prod_{i=1}^{s-1}\mathsf{P}_z\mathsf{G}_i(\mathsf{I}\otimes\mathsf{A}_H^2)\right\|_{\mathrm{op}}\|\mathsf{P}_z\mathsf{G}_0(\mathsf{I}\otimes\mathsf{A}_H)\|_{\mathrm{op}}\\ &\le \left\|\prod_{i=1}^{s-1}\mathsf{P}_z\mathsf{G}_i(\mathsf{I}\otimes\mathsf{A}_H^2)\right\|_{\mathrm{op}}\\ &\le \sigma_2(H^2)^{s-1} + (s-1)\cdot\sigma_2(H^2)^{s-2} + (s-1)^2\cdot\sigma_2(H^2)^{s-4},\end{aligned}$$

where the last inequality follows from Fact A.7. ∎

Remark A.9. *We know that in the modified construction* H^2 *is a Cayley graph since* H *is a Cayley graph.*

A.4 Splittability

In this subsection, we focus on the splittability parameters arising out of the construction described above. The collection $W(k) \subseteq [n]^k$ is obtained from taking $k-1$ step walks on s-wide replacement as described above, which is d_2^2-regular. Recall from Definition 3.9 that we need to show $\sigma_2(\mathsf{S}_{W[a,t],W[t+1,b]}) \le \tau$ for all $1 \le a < t < b \le k$, where,

$$\left(\mathsf{S}_{W[a,t],W[t+1,b]}\right)_{(i_a,\cdots,i_t),(i_{t+1},\cdots,i_b)} := \frac{\mathbf{1}[(i_a,\cdots,i_t,i_{t+1},\cdots,i_b)\in W[a,b]]}{d_2^{2(b-s)}}$$

Lemma A.10. *Let* $1 \le a < t < b \le k$*. Suppose* G *is a* d_1*-regular outer graph on vertex set* $[n]$ *with walk operator* G_t *used at step* s *of a walk on the* s*-wide replacement product and* H *is a* d_2*-regular inner graph on vertex set* $[m]$ *with normalized random walk operator* A_H*. Then there are orderings of the rows and columns of the representations of* $\mathsf{S}_{W[a,t],W[t+1,b]}$ *and* A_H *as matrices such that*

$$\mathsf{S}_{W[a,t],W[t+1,b]} = ((I\otimes\mathsf{A}_H)G_t(I\otimes\mathsf{A}_H))\otimes\mathsf{J}/d_2^{2(b-t-1)},$$

where $\mathsf{J} \in \mathbb{R}^{[d_2]^{2(t-a)}\times[d_2]^{2(b-t-1)}}$ *is the all ones matrix.*

Proof. Partition the set of walks $W[a,t]$ into the sets $W_{1,1},\dots,W_{n',m}$, where $w \in W_{i,j}$ if the last vertex of the walk $i_t = (v_t,u_t)$ satisfies $v_t = i$ and $u_t = j$. Similarly, partition $W[t+1,b]$ into the sets $W'_{1,1},\dots,W'_{n',m}$, where $(i_{t+1},\cdots,i_b) \in W'_{i,j}$ if the first vertex of the walk $i_{t+1} = (v_{t+1},u_{t+1})$ satisfies $v_{t+1} = i$ and $u_{t+1} = j$. Note that $\left|W_{i,j}\right| = d_2^{2(t-a)}$ and $\left|W'_{i,j}\right| = d_2^{2(b-t-1)}$ for all $(i,j) \in [n'] \times [m]$, since there are d_2^2 choices for each step of the walk.

Now order the rows of the matrix $\mathsf{S}_{W[a,t],W[t+1,b]}$ so that all of the rows corresponding to walks in $W_{1,1}$ appear first, followed by those for walks in $W_{1,2}$, and so on in lexicographic order of the indices (i,j) of $W_{i,j}$, with an arbitrary order within each set. Do a similar re-ordering of the columns for the sets $W'_{1,1},\dots,W'_{n',m}$. Observe that

$$\begin{aligned}\left(\mathsf{S}_{W[a,t],W[t+1,b]}\right)_{(i_a,\cdots,i_t),(i_{t+1},\cdots,i_b)} &= \frac{\mathbf{1}_{(i_a,\cdots,i_t,i_{t+1},\cdots,i_b)\in W[a,b]}}{d_2^{2(b-t)}}\\ &= \frac{d_2^2 \cdot (\text{weight of transition from } i_t \text{ to } i_{t+1} \text{ in } (I\otimes \mathsf{A}_H)G_t(I\otimes \mathsf{A}_H))}{d_2^{2(b-t)}},\end{aligned}$$

which only depends on the adjacency of the last vertex of $(i_a,\cdots,i_t)$ and the first vertex of $(i_{t+1},\cdots,i_b)$. If the vertices $i_t = (v_t,u_t)$ and $i_{t+1} = (v_{t+1},u_{t+1})$ are adjacent, then

$$\left(\mathsf{S}_{W[a,t],W[t+1,b]}\right)_{(i_a,\cdots,i_t),(i_{t+1},\cdots,i_b)} = ((I\otimes \mathsf{A}_H)G_t(I\otimes \mathsf{A}_H))_{(v_t,u_t),(v_{t+1},u_{t+1})} / d_2^{2(b-t-1)},$$

for every $(i_a,\cdots,i_t) \in W[a,t]$ and $(i_{t+1},\cdots,i_b) \in W[t+1,b]$; and otherwise $\left(\mathsf{S}_{W[a,t],W[t+1,b]}\right)_{(i_a,\cdots,i_t),(i_{t+1},\cdots,i_b)} = 0$. Since the walks in the rows and columns are sorted according to their last and first vertices, respectively, the matrix $\mathsf{S}_{W[a,t],W[t+1,b]}$ exactly matches the tensor product $((I\otimes \mathsf{A}_H)G_t(I\otimes \mathsf{A}_H)) \otimes \mathsf{J}/d_2^{2(b-t-1)}$. ■

Corollary A.11. *Let $1 \le a < t < b \le k$. Suppose G is a d_1-regular outer graph with walk operator G_t used at step t of a walk on the s-wide replacement product and H is a d_2-regular inner graph with normalized random walk operator A_H. Then*

$$\sigma_2(\mathsf{S}_{W[a,t],W[t+1,b]}) = \sigma_2((I\otimes \mathsf{A}_H)G_t(I\otimes \mathsf{A}_H)).$$

Proof. Using Lemma A.10 and the fact that

$$\sigma_2(((I\otimes \mathsf{A}_H)G_t(I\otimes \mathsf{A}_H)) \otimes \mathsf{J}/d_2^{2(b-t-1)}) = \sigma_2((I\otimes \mathsf{A}_H)G_t(I\otimes \mathsf{A}_H)),$$

the result follows. ■

Remark A.12. *Corollary A.11 is what causes the splittability argument to break down for Ta-Shma's original construction, as $\sigma_2(\mathsf{G}_t(\mathsf{I}\otimes \mathsf{A}_H)) = 1$.*

A.5 Parameter Choices

In this section, we choose parameters to finally obtain Theorem A.1, for which we must argue about bias, rate and splittability.

A graph is said to be an (n,d,λ)-graph provided it has n vertices, is d-regular, and has second largest singular value of its normalized adjacency matrix at most λ.

Notation A.13. *We use the following notation for the graphs G and H used in the s-wide replacement product.*

- *The outer graph G will be an (n'', d_1, λ_1)-graph.*

- *The inner graph H will be a (d_1^s, d_2, λ_2)-graph.*

The parameters $n'', d_1, d_2, \lambda_1, \lambda_2$ and s are yet to be chosen.

We are given the dimension D of the desired code and its bias $\varepsilon \in (0, 1/2)$. We set a parameter $\alpha \leq 1/128$ such that (for convenience) $1/\alpha$ is a power of 2 and

$$\frac{\alpha^5}{4\log_2(1/\alpha)} \geq \frac{1}{\log_2(1/\varepsilon)}. \tag{6}$$

By replacing $\log_2(1/\alpha)$ with its upper bound $1/\alpha$, we observe that $\alpha = \Theta(1/\log_2(1/\varepsilon)^{1/6})$ satisfies this bound, and so we choose $s = \Theta(\log_2(1/\varepsilon)^{1/6})$.

The inner graph H. The choice of H is same as Ta-Shma's choice. More precisely, we set $s = 1/\alpha$ and $d_2 = s^{4s}$. We obtain a Cayley graph $H = \mathsf{Cay}(\mathbb{F}_2^{4s\log_2(d_2)}, A)$ such that H is an $(n_2 = d_2^{4s}, d_2, \lambda_2)$ graph where $\lambda_2 = b_2/\sqrt{d_2}$ and $b_2 = 4s\log_2(d_2)$. (The set of generators, A, comes from a small bias code derived from a construction of Alon et al. [AGHP92].)

The base code $\mathcal{C}_0$. This is dealt with in detail in Section 5. We choose $\varepsilon_0 = 1/d_2^2$ and use Corollary 6.2 to obtain a code $\mathcal{C}_0'$ in $\mathbb{F}_2^{n'}$ that is ε_0-biased and has a blocklength $\Omega(D/\varepsilon_0^c)$ for some constant c. Call this blocklength of $\mathcal{C}_0'$ to be n'. Next we replicate the codewords $m = d_1^s$ times to get code $\mathcal{C}_0$ in $\mathbb{F}_2^n$ with the same bias but a rate that is worse by a factor of m. In the proofs below, we only use properties of $\mathcal{C}_0$ that is of multiplicity m, has rate $\Omega(\varepsilon_0^c)/m$ and has bias ε_0, as specified in Theorem A.1.

The outer graph G. Set $d_1 = d_2^4$ so that $n_2 = d_1^s$ as required by the s-wide replacement product. We apply Ta-Shma's explicit Ramanujan graph lemma (Lemma 2.10 in [TS17]) with parameters n', d_1 and θ to obtain an (n'', d_1, λ_1) Ramanujan graph G with $\lambda_1 \leq 2\sqrt{2}/\sqrt{d_1}$ and $n'' \in [(1-\theta)n', n']$ or $n'' \in [(1-\theta)2n', 2n']$. Here, θ is an error parameter that we set as $\theta = \lambda_2^4/6$ (this choice of θ differs from Ta-Shma). Because we can construct words with block length $2n'$ (if needed) by duplicating each codeword, we may assume w.l.o.g. that n'' is close to n' and $(n' - n'') \leq \theta n' \leq 2\theta n''$. See [TS17] for a more formal description of this graph.

Note that $\lambda_1 \leq \lambda_2^4/6$ since $\lambda_1 \leq 3/\sqrt{d_1} = 3/d_2^2 = 3 \cdot \lambda_2^4/b_2^4 \leq \lambda_2^4/6$. Hence, $\varepsilon_0 + 2\theta + 2\lambda_1 \leq \lambda_2^4$, as needed to apply Corollary A.8.

The walk length. Set the walk length $k-1$ to be the smallest integer such that

$$(\lambda_2^2)^{(1-5\alpha)(1-\alpha)(k-1)} \leq \varepsilon.$$

This will imply using Ta-Shma's analysis that the bias of the final code is at most ε as shown later.

$s = 1/\alpha, \quad s = \Theta(\log(1/\varepsilon)^{1/6})$, so that $\frac{\alpha^3}{4\log_2(1/\alpha)} \geq \frac{1}{\log_2(1/\varepsilon)}$
$H : (n_2, d_2, \lambda_2), \quad n_2 = d_1^s, \quad d_2 = s^{4s}, \quad \lambda_2 = \frac{b_2}{\sqrt{d_2}}, \quad b_2 = 4s \log d_2$
$\mathcal{C}_0'$: bias $\varepsilon_0 = 1/d_2^2$, blocklength $n' = O(D/\varepsilon_0^c)$
$\mathcal{C}_0$: bias $\varepsilon_0 = 1/d_2^2$, multiplicity $m = d_1^s$, blocklength $n = O(mD/\varepsilon_0^c)$
$G : (n'', d_1, \lambda_1), \quad n'' \approx n' = O(D/\varepsilon_0^c), \quad d_1 = d_2^4, \quad \lambda_1 \leq \frac{2\sqrt{2}}{d_1}$
k : smallest integer such that $(\lambda_2^2)^{(1-5\alpha)(1-\alpha)(k-1)} \leq \varepsilon$

Proof of Theorem A.1. We will prove it in the following claims. We denote by $W(k) \subseteq [n]^k$ the collection of walks on the s-wide replacement product obtained above, and we denote by $\mathcal{C}$ the final code obtained by doing the direct sum operation on $\mathcal{C}_0$ using the collection of tuples $W(k)$. The explicitness of $W(k)$ follows from Ta-Shma's construction since all the objects used in the construction have explicit constructions.

Next, the multiplicity $m = d_1^s = d_2^{4s} = s^{16s^2} = 2^{16s^2 \log s} \leq (2^{s^6})^{o(1)} = (1/\varepsilon)^{o(1)}$.

Claim A.14. *We have $k - 1 \geq s/\alpha = s^2$, and that $k - 1 \leq 2s^5$, so that*

$$\Theta(\log(1/\varepsilon)^{1/3}) \leq k \leq \Theta(\log(1/\varepsilon))$$

Proof. Using $d_2 = s^{4s}$ and Eq. (6), we have

$$\begin{aligned}
\left(\frac{1}{\lambda_2^2}\right)^{(1-5\alpha)(1-\alpha)s/\alpha} &\leq \left(\frac{1}{\lambda_2^2}\right)^{s/\alpha} = \left(\frac{d_2}{b_2^2}\right)^{s/\alpha} \leq (d_2)^{s/\alpha} = s^{4s^2/\alpha} \\
&= 2^{4s^2\log_2(s)/\alpha} = 2^{4\log_2(1/\alpha)/\alpha^3} \leq 2^{\log_2(1/\varepsilon)} = \frac{1}{\varepsilon}.
\end{aligned}$$

Hence, $\varepsilon \geq (\lambda_2^2)^{(1-5\alpha)(1-\alpha)s/\alpha}$ and thus $k-1$ must be at least s/α.

In the other direction, we show that $(\lambda_2^2)^{(1-5\alpha)(1-\alpha)2s^5} \leq \varepsilon$, which will imply $k \leq \Theta(s^5) \Rightarrow k \leq \Theta(s^6) = \Theta(\log(1/\varepsilon))$.

$$(\lambda_2^2)^{(1-5\alpha)(1-\alpha)2s^5} \leq \left(\frac{b_2^2}{d_2}\right)^{s^5} \leq \left(\frac{1}{s^{3s}}\right)^{s^5} = 2^{-\Theta(s^6 \log s)} \leq 2^{-\Theta(s^6)} = 2^{-\log(1/\varepsilon)} \leq \varepsilon$$

■

Remark A.15. *By the minimality of k, we have $(\lambda_2^2)^{(1-5\alpha)(1-\alpha)(k-2)} \geq \varepsilon$. Since $1/(k-1) \leq \alpha$, we get $(\lambda_2^2)^{(1-5\alpha)(1-\alpha)^2(k-1)} \geq \varepsilon$. This will be useful in rate computation.*

Claim A.16. *The code $\mathcal{C}$ is ε-balanced.*

Proof. Using Corollary A.8, we have that the final bias

$$b := \left(\sigma_2(H^2)^{s-1} + (s-1)\cdot\sigma_2(H^2)^{s-2} + (s-1)^2\cdot\sigma_2(H^2)^{s-4}\right)^{\lfloor (k-1)/s \rfloor}$$

is bounded by

$$\begin{aligned}
b &\leq (3(s-1)^2\sigma_2(H^2)^{s-4})^{((k-1)/s)-1} && \text{(Using } \sigma_2(H^2) \leq 1/3s^2\text{)}\\
&\leq ((\sigma_2(H^2)^{s-5})^{(k-1-s)/s}\\
&= \sigma_2(H^2)^{(1-5/s)(1-s/(k-1))(k-1)}\\
&\leq \sigma_2(H^2)^{(1-5\alpha)(1-\alpha)(k-1)}\\
&= \left(\lambda_2^2\right)^{(1-5\alpha)(1-\alpha)(k-1)} \leq \varepsilon,
\end{aligned}$$

where the last inequality follows from $s = 1/\alpha$ and $k-1 \geq s/\alpha$, the latter from Claim A.14. ■

Claim A.17. *$\mathcal{C}$ has rate $\Omega(\varepsilon^{2+28\cdot\alpha})$.*

Proof. The support size is the number of walks of length k on the s-wide replacement product of G and H (each step of the walk has d_2^2 options), which is

$$\begin{aligned}
|V(G)||V(H)|d_2^{2(k-1)} = n'' \cdot d_1^s \cdot d_2^{2(k-1)} &= n'' \cdot d_2^{2(k-1)+4s} \leq n' \cdot d_2^{2(k-1)+4s}\\
&= \Theta\left(\frac{D}{\varepsilon_0^c} \cdot d_2^{2(k-1)+4s}\right)\\
&= \Theta\left(D \cdot (d_2^2)^{k-1+2s+c}\right)\\
&= O\left(D \cdot (d_2^2)^{(1+3\alpha)(k-1)}\right),
\end{aligned}$$

where the penultimate equality follows from the assumption that ε_0 is a constant.

Note that $d_2^\alpha = d_2^{1/s} = s^4 \geq b_2$ since $b_2 = 4s\log_2(d_2) = 16s^2\log_2(s) \leq s^4$. Thus,

$$d_2^{1-2\alpha} = \frac{d_2}{d_2^{2\alpha}} \leq \frac{d_2}{b_2^2} = \frac{1}{\sigma_2(H^2)}.$$

We obtain

$$\begin{aligned}
(d_2^2)^{(k-1)} &\leq \left(\frac{1}{\sigma_2(H^2)}\right)^{\frac{2(k-1)}{1-2\alpha}}\\
&\leq \left(\frac{1}{\varepsilon}\right)^{\frac{2}{(1-2\alpha)(1-5\alpha)(1-\alpha)^2}} && \text{(Using Remark A.15)}\\
&\leq \left(\frac{1}{\varepsilon}\right)^{2(1+10\alpha)},
\end{aligned}$$

which implies a block length of

$$O\left(D \cdot (d_2^2)^{(1+3\alpha)(k-1)}\right) = O\left(D\left(\frac{1}{\varepsilon}\right)^{2(1+10\alpha)(1+3\alpha)}\right) = O\left(D\left(\frac{1}{\varepsilon}\right)^{2(1+14\alpha)}\right).$$

■

Claim A.18. *$W(k)$ is τ-splittable for $\tau \leq 2^{-\Theta(\log(1/\varepsilon)^{1/6})}$.*

Proof. As we saw in Corollary Corollary A.11, the splittability τ can be upper bounded by $\sigma_2((I \otimes \mathsf{A}_H)G_t(I \otimes \mathsf{A}_H))$, which is at most $\sigma_2(G) + 2 \cdot \sigma_2(H) + \sigma_2(H)^2$ by Fact A.5. So, the collection $W(k)$ is τ-splittable for

$$\begin{aligned}
\tau \leq \sigma_2(G) + 2 \cdot \sigma_2(H) + \sigma_2(H)^2 \leq 4\lambda_2 &= 4b_2/d_2^{1/2} \\
&= 64s^2 \log s/s^{2s} \\
&= 2^{-\Theta(s \log s)} \\
&\leq 2^{-\Theta(s)} \\
&= 2^{-\Theta(\log(1/\varepsilon)^{1/6})}
\end{aligned}$$

■

■