Electronic Colloquium on Computational Complexity, Report No. 93 (2025)

Eigenvalue Bounds for Symmetric Markov Chains on Multislices With Applications

Prashanth Amireddy* Amik Raj Behera† Srikanth Srinivasan ‡ Madhu Sudan§

June 23, 2025

Abstract

We consider random walks on "balanced multislices" of any "grid" that respects the "symmetries" of the grid, and show that a broad class of such walks are good spectral expanders. (A grid is a set of points of the form $\mathcal{S}^n$ for finite $\mathcal{S}$, and a balanced multi-slice is the subset that contains an equal number of coordinates taking every value in $\mathcal{S}$. A walk respects symmetries if the probability of going from $u = (u_1, \ldots, u_n)$ to $v = (v_1, \ldots, v_n)$ is invariant under simultaneous permutations of the coordinates of u and v.) Our main theorem shows that, under some technical conditions, every such walk where a single step leads to an almost $\mathcal{O}(1)$-wise independent distribution on the next state, conditioned on the previous state, satisfies a non-trivially small singular value bound.

We give two applications of our theorem to error-correcting codes: (1) We give an analog of the Ore-DeMillo-Lipton-Schwartz-Zippel lemma for polynomials, and junta-sums, over balanced multislices. (2) We also give a local list-correction algorithm for d-junta-sums mapping an arbitrary grid $\mathcal{S}^n$ to an Abelian group, correcting from a near-optimal $(\frac{1}{|\mathcal{S}|^d} - \varepsilon)$ fraction of errors for every $\varepsilon > 0$, where a d-junta-sum is a sum of (arbitrarily many) d-juntas (and a d-junta is a function that depends on only d of the n variables).

Our proofs are obtained by exploring the representation theory of the symmetric group and merging it with some careful spectral analysis.

*School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA. Supported in part by a Simons Investigator Award and NSF Award CCF 2152413 to Madhu Sudan and a Simons Investigator Award to Salil Vadhan. Email: pamireddy@g.harvard.edu

†Department of Computer Science, University of Copenhagen, Denmark. Supported by Srikanth Srinivasan's start-up grant from the University of Copenhagen. Email: ambe@di.ku.dk

‡Department of Computer Science, University of Copenhagen, Denmark. Supported by the European Research Council (ERC) under grant agreement no. 101125652 (ALBA). Email: srsr@di.ku.dk

§School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA. Supported in part by a Simons Investigator Award, NSF Award CCF 2152413 and AFOSR award FA9550-25-1-0112. Email: madhu@cs.harvard.edu

ISSN 1433-8092

Contents

1 Introduction

Consider the following natural random walk whose states are the balanced vectors of $\{0,1\}^n$, i.e., the balanced Boolean slice with an equal number of 0s and 1s, where a single step of the random walk takes a state u to a state v at Hamming distance exactly $n/2$ from it. One would expect this random walk to mix extremely rapidly, and indeed this is known. The underlying graph here is a special case of a Johnson graph whose entire eigenspectrum is well known [Del78] and, in particular, implies that the second eigenvalue of this graph is $o_n(1)$.

Now consider the following variant of the above random walk: The states now are elements of the 'balanced multislice' in $\{-1,0,1\}^n$, i.e. vectors with exactly 1/3rd fraction of the letters -1, 0 and 1, and in a single step from a balanced vector u to a random balanced vector v obtained by flipping exactly 1/3 fraction of each of the letters of u to -1, 0 and 1. (So for a single coordinate i, v_i is uniform in $\{-1,0,1\}$ conditioned on u_i.) It is intuitive to believe that such a random walk should also converge to the uniform distribution over all balanced vectors extremely fast, but, as far as we know, it was not even known that the second-eigenvalue of this random walk (or its transition probability matrix) has value $o_n(1)$.

The gap in the understanding between the Boolean and non-Boolean cases in such problems can be significant for fundamental reasons. For example, for the alternate version of the random walk where the transition is defined by a uniformly random transposition of coordinates, it took a decade after optimal bounds on the mixing time were proved in the Boolean case [DS87] to prove similar results in the non-Boolean setting [Sca97]. We refer the reader to the work of Filmus, O'Donnell, and Wu [FOW22] for a nice overview of the challenges posed by the non-Boolean setting in such problems. Some of these obstructions have to do with associated representations of the symmetric group that play a role in the corresponding proofs; these representations are simpler ('multiplicity-free') in the Boolean setting than in the non-Boolean setting. This also creates difficulties in resolving the questions we consider.

The main contribution of this work is to address some of the challenges alluded to above. In particular, we show that the variant random walk described in the second paragraph also has fast mixing, specifically by giving a $o_n(1)$ bound on its second eigenvalue. Indeed, we study this question in more generality for balanced multislices, with "nearly balanced moves". We believe the questions carry intrinsic interest and should find broad applications in the field. We justify this belief partially by describing two applications in coding theory:

- The first gives a near-tight distance bound on codes obtained by evaluations of polynomials on balanced multislices.
- The second gives a local list-correction algorithm for subclasses of polynomials evaluated on *grids*. (Note that the second application does not refer to balanced multislices in the problem definition — the multislices arise naturally in the design and analysis of the local correction algorithm!)

We elaborate on our setting and results, the applications, and the technical challenges below.

1.1 Multislices and Random Walks

By a *grid*, we refer to sets of the form $\mathcal{S}^n$ for some finite set $\mathcal{S}$ and positive integer n. (Usually we think of $s := |\mathcal{S}|$ as a constant and study the growth of relevant parameters as a function of n). The *balanced multislice* of a grid $\mathcal{S}^n$ is the set

$$\mathcal{S}^n_\mu := \left\{ \mathbf{a} \in \mathcal{S}^n \;\middle|\; \forall \sigma \in \mathcal{S}, |\{i \in [n] | a_i = \sigma\}| = \frac{n}{s} \right\}.$$

(Note that a multislice is non-empty if and only if s divides n. We will drop the term "balanced" in the future and simply refer to multislices to keep the term short.)

The random walks we consider have the multislice of some grid as their state space. Recall that such a random walk can be described by a $\mathcal{S}^n_\mu \times \mathcal{S}^n_\mu$ matrix W with $W(\mathbf{a}, \mathbf{b})$ denoting the probability of transition from state $\mathbf{a}$ to $\mathbf{b}$. We consider walks where every step of the walk makes a "nearly balanced move". To elaborate, let us define the *generalized Hamming distance*[1] $\Delta(\mathbf{a}, \mathbf{b})$, for vectors $\mathbf{a}, \mathbf{b} \in \mathcal{S}^n$, to be the $\mathcal{S} \times \mathcal{S}$ matrix given by $\Delta_{\sigma,\tau}(\mathbf{a}, \mathbf{b}) = |\{i \in [n] | a_i = \sigma, b_i = \tau\}|$. We say that a generalized Hamming distance parameter $\Delta \in \mathbb{Z}^{\mathcal{S}\times\mathcal{S}}$ *determines* a random walk matrix, denoted W_Δ, if for each vertex $\mathbf{a}$, the random step corresponding to W_Δ is obtained by picking, uniformly at random, a vertex $\mathbf{b}$ on the multislice such that $\Delta(\mathbf{a}, \mathbf{b}) = \Delta$.

For constant $C < \infty$, we say that a generalized Hamming distance parameter $\Delta \in \mathbb{Z}^{\mathcal{S}\times\mathcal{S}}$ is C-*balanced* if each entry of Δ is $\frac{m}{s} \pm \left(C\sqrt{m \log m}\right)$ where $m = n/s$. In other words, all the entries of Δ are equal up to a difference of at most $2C\sqrt{m \log m}$. Informally, when considering $n \to \infty$ we refer to Δ, as also a random walk matrix W_Δ determined by Δ, as "nearly balanced" if Δ is C-balanced for some constant C. Here, we note that W_Δ is a well-defined random walk matrix over the multislice only if Δ/m is a doubly-stochastic matrix (i.e., every row and every column of Δ sums to m).

Note that the mixing time of a random walk matrix W is closely tied to the second largest singular value, which we denote by $\sigma_2(W)$. (In particular, the singular values satisfy $1 = \sigma_1 \geqslant \sigma_2 \geqslant \cdots \sigma_N \geqslant 0$ and we let $\sigma_2(W) = \sigma_2$. If the walk is symmetric, then this captures the second eigenvalue. Specifically, if the eigenvalues are $1 = \lambda_1 \geqslant \lambda_2 \geqslant \cdots \lambda_N \geqslant -1$ where $N = |\mathcal{S}^n_\mu|$, then $\sigma_2(W) = \max\{|\lambda_2|, |\lambda_N|\}$.) Our main goal is to bound the value of $\sigma_2(W)$ by some function $o_n(1)$ that tends to 0 with growing n for a broad class of random walk matrices W over the multislice S^n_μ. In general, it is desirable to have such singular value bounds, and such random walk matrices are said to have good "spectral expansion" or "fast mixing".

The following theorem gives such a fast mixing result on the balanced multislice for all nearly balanced walks that "respect symmetries". More formally, for a permutation $\pi \in \mathrm{Sym}_n$ and $\mathbf{a} \in \mathcal{S}^n$, let $\pi(\mathbf{a})$ denote the action of π on $\mathcal{S}^n_\mu$, i.e., $\pi(\mathbf{a}) := (a_{\pi^{-1}(1)}, \ldots, a_{\pi^{-1}(n)})$. For a stochastic matrix $M \in \mathbb{R}^{\mathcal{S}^n_\mu \times \mathcal{S}^n_\mu}$, we say M *respects symmetries* if for all permutations $\pi \in \mathrm{Sym}_n$ and for all $\mathbf{a}, \mathbf{b} \in \mathcal{S}^n_\mu$ we have $M(\pi(\mathbf{a}), \pi(\mathbf{b})) = M(\mathbf{a}, \mathbf{b})$. Our main theorem shows that walks that respect symmetries and are nearly balanced have fast mixing.

[1] Sometimes, this is also called as a "meet table" in algebraic combinatorics.

Theorem 1.1 (Singular value bound for nearly balanced walks). *For every $s \geqslant 2$ and $C < \infty$, there exists $\tau > 0$ such that for every finite set $\mathcal{S}$ of size s and sufficiently large $n \in \mathbb{N}$, the following holds:*
If W is a stochastic matrix over the multislice $\mathcal{S}_\mu^n$ that respects symmetries*, and satisfies the condition that*

$$W(\mathbf{a}, \mathbf{b}) > 0 \quad \Rightarrow \quad \Delta(\mathbf{a}, \mathbf{b}) \text{ is } C\text{-balanced} \quad \forall\, \mathbf{a}, \mathbf{b} \in \mathcal{S}_\mu^n,$$

then $\sigma_2(W) \leqslant 1/n^\tau$.

The above result implies that the random walk on the balanced multislice mentioned earlier in this section (which corresponds to $s = 3$ and C-balanced generalized distance parameter with $C = 0$) has its second largest eigenvalue polynomially bounded. In fact, Theorem 1.1 is more general and covers multislices over any grid $\mathcal{S}$ of constant size (i.e., for every $|\mathcal{S}| = \mathcal{O}(1)$). Additionally, it is robust to perturbations of transition probabilities as long as the transition probabilities are nearly balanced.

Indeed Theorem 1.1 follows from our main technical theorem, stated as Theorem 1.3 below, which abstracts the main properties that suffice to prove the bound on the second largest singular value. Specifically Theorem 1.3 shows that, in addition to the symmetries respected by the matrix W, the important features that suffice to prove fast mixing are:

1. The next state of the random walk is "almost $\mathcal{O}(1)$-wise independent" conditioned on the current state
2. The Frobenius norm of W is polynomially bounded in n.

We elaborate on these conditions below before stating our main technical result Theorem 1.3.

For a distribution D supported on $\mathcal{S}^n$ and set $T \subseteq [n]$, we let D_T denote the marginal distribution supported on $\mathcal{S}^T$ induced by projecting a random variable $x \sim D$ to its coordinates in T. Recall that a distribution D is k-wise independent if for every set $T \subseteq [n]$ with $|T| \leqslant k$ we have D_T is the uniform distribution on $\mathcal{S}^T$. Recall further that D is ε-almost k-wise independent if for every set $T \subseteq [n]$ with $|T| \leqslant k$ we have D_T is ε-close in total variation distance to the uniform distribution on $\mathcal{S}^T$.

In the following definition we view the rows of a stochastic matrix $M \in \mathbb{R}^{\mathcal{S}_\mu^n \times \mathcal{S}_\mu^n}$, denoted $M(\mathbf{a}) := (M(\mathbf{a}, \mathbf{b}) | \mathbf{b} \in \mathcal{S}_\mu^n)$ for $\mathbf{a} \in \mathcal{S}_\mu^n$, as distributions supported on $\mathcal{S}^n$ (which have zero support outside $\mathcal{S}_\mu^n$).

Definition 1.2 (ε-almost k-wise independent matrix). *For parameter $k \in \mathbb{N}$ and $\epsilon > 0$ we say that a stochastic matrix $M \in \mathbb{R}^{\mathcal{S}_\mu^n \times \mathcal{S}_\mu^n}$ is* ε-almost k-wise independent *if for every row $\mathbf{a} \in \mathcal{S}_\mu^n$, the distribution $M(\mathbf{a})$ is ε-almost k-wise independent.*

Finally we recall that for a matrix $M \in \mathbb{R}^{N \times N}$, its Frobenius norm, denoted $\|M\|_F$, is the quantity $\sqrt{\sum_{(i,j) \in N \times N} M(i,j)^2}$.

We now state the main theorem of our work.

Theorem 1.3 (Singular Value Bound for Markov Chains on Balanced Multislice). *For every $\kappa > 0$, and $s \in \mathbb{N}$ with $\kappa \geqslant s$ there exists $c_1.c_2, c_3 < \infty$ such that for every $\varepsilon > 0$ and every sufficiently large $n \in \mathbb{N}$ that is divisible by s, the following holds:*
Suppose $\mathcal{S}$ is a set of size s, and $M \in \mathbb{R}^{\mathcal{S}_\mu^n \times \mathcal{S}_\mu^n}$ is a stochastic matrix that satisfies the following three conditions:

1. *The matrix M respects symmetries.*
2. *$\|M\|_F \leqslant c_1 \cdot n^\kappa$.*
3. *The matrix M is ε-almost k-wise independent for $k = 10s\kappa$.*

Then we have $\sigma_2(M) \leqslant \max\{c_2/n, c_3 \cdot \varepsilon\}$.

If the Markov chain is symmetric, then the singular values correspond to the eigenvalues, and hence Theorem 1.3 yields eigenvalue bounds for symmetric Markov chains satisfying the properties mentioned above. Theorem 1.3 is proved in Section 3. The proof involves many standard and some new elements of representation theory for the symmetric group. We elaborate on this in Section 1.2.2. We also note that Theorem 1.1 immediately follows from Theorem 1.3, modulo some calculations that verify that Condition (2) above applies to C-balanced matrices. For more details, see Section 3.5.

To illustrate the applicability of Theorem 1.1, we give two examples, both related to coding theoretic aspects of polynomials and other polynomial-like functions that we refer to as *junta-sums*. These results extend corresponding works in the Boolean setting [ABPSS25; ABSS25], obtaining natural generalizations to non-Boolean settings.

Distance of polynomials and junta-sums on multislices. A function $f : \mathcal{S}^n \to G$ is called a d-junta if it depends on only d of the n variables, i.e, there exists a set $I \subseteq [n]$, $|I| \leqslant d$ and a function $g : \mathcal{S}^I \to G$ such that for all $\mathbf{a} \in \mathcal{S}^n$, $f(\mathbf{a}) = g(\mathbf{a}|_I)$ where $\mathbf{a}|_I$ is the projection of $\mathbf{a}$ to the coordinates in I. Here we could allow G to be any set, though in this work G will denote an Abelian group. We say f is a *degree d junta-sum* (or simply a *d-junta-sum*) if there exists d-junta's $f_1, \ldots, f_k : \mathcal{S}^n \to G$ such that $f = f_1 + \cdots + f_k$.
When $G = \mathbb{F}$ is a field and $\mathcal{S} \subseteq \mathbb{F}$, then degree d junta-sums are closely related to the notion of degree d polynomials. In particular, every degree-d polynomial is also a degree-d junta-sum, and degree-d junta-sums are polynomials of degree at most $(s-1)d$ where $s = |\mathcal{S}|$. Junta-sums come up naturally when studying questions related to testing direct sums and low-degree polynomials [DG19; BP21; ASS23].

A well-studied question about degree-d polynomials is: How often can a non-zero polynomial be zero on a grid? The well-known and oft-discovered Ore-DeMillo-Lipton-Schwartz-Zippel lemma [Ore22; DL78; Zip79; Sch80] (henceforth ODLSZ lemma) asserts that a non-zero degree-d polynomial over a field $\mathbb{F}$ is non-zero with probability at least $s^{-d/(s-1)}$ over the uniform distribution over $\mathcal{S}^n$. When $\mathcal{S} = \mathbb{F}$, the precise bound is $\delta(q, d) = (1 - \beta/q)q^{-\alpha}$, where α and β are the quotient and

remainder respectively when d is divided by $q-1$. The former bound immediately implies that a degree-d junta-sum is non-zero with probability at least s^{-d} over $\mathcal{S}^n$ (and the claim even extends to arbitrary $\mathcal{S}$ and Abelian groups G).

A natural related question then becomes — *how do these bounds change when considering natural subsets T that are not grids (or more generally product sets)?* Recent work has begun to address such questions [ABSS25; KKS24]. In this work, we consider the case of the balanced multislice i.e., $T = \mathcal{S}^n_\mu$. Despite the simple nature of these questions, the answer does not seem to have been pinned down before, with the exception of the Boolean case that was resolved recently [ABSS25]. We are able to show a clean connection between $\mathcal{S}^n$ and $\mathcal{S}^n_\mu$ that allows us to show that these probabilities (in the worst case) differ by at most $\lambda_2(W)$ for some nearly balanced walk over the multislice. This allows us to prove the following theorem, which generalizes the work of [ABSS25] beyond the Boolean case.

Theorem 1.4 (Polynomial distance over multislice). *For every finite field $\mathbb{F} = \mathbb{F}_q$, if a degree d polynomial $P(\mathbf{x})$ is such that $P(\mathbf{a}) \neq 0$ for some $\mathbf{a} \in \mathbb{F}^n_\mu$ on the balanced multislice, then*

$$\Pr_{\mathbf{b}\sim\mathbb{F}^n_\mu}[P(\mathbf{b}) \neq 0] \geqslant \delta(q,d) - \frac{1}{n^{\Omega_q(1)}},$$

where $\delta(q,d) = (1-\beta/q)q^{-\alpha}$, where α and β are the quotient and remainder respectively when d is divided by $q-1$.

We prove this theorem in Section 4.

Note that $\delta(q,d)$ is exactly the distance of the space of degree-d polynomials on the field $\mathbb{F}_q$ and hence the above theorem says that the distance of the space of polynomials on the balanced multislice is nearly exactly what it is in the grid $\mathbb{F}^n_q$.[2] An analogous statement can also be made for junta-sums, getting a bound that almost matches the bound over grids, i.e., $1/s^d$ (see Theorem 4.2).

Following the proof idea of [ABSS25], both cases are handled by a similar proof technique that first proves a quantitatively weak bound on the probability that f is non-zero[3](see Corollary 5.11), and then randomly identifies a small grid inside the multislice. On each such grid, we can apply the ODLSZ lemma as a black-box to assert that if f is non-zero within the randomly identified small grid, then it is non-zero with the 'correct' probability (either $\delta(q,d)$ or s^{-d}). It suffices, therefore, to prove that f is non-zero on most of the grids, which is where the main technical theorem regarding the expansion of the walk on the balanced multislice comes into play. We use our eigenvalue bounds along with the quantitatively weak bound already obtained to establish that all but an $n^{-\Omega(1)}$-fraction of the grids satisfy this property. See Section 1.2.3 for the proof overview

[2] An important subtlety here is that there are polynomials that are non-zero in the grid $\mathbb{F}^n_q$ but are zero at all points on the multislice. That is the reason this theorem is only stated for polynomials that are non-zero as functions on the multislice. This is analogous to similar restrictions we place on polynomials in the setting of grids (e.g., in the setting of the Boolean cube, we only consider non-zero *multilinear* polynomials).

[3] We prove these bounds by an adaptation of the standard inductive strategy used to prove the standard ODLSZ lemma. Unfortunately, we are unable to use this strategy to prove a tight bound.

and Section 4 for a formal proof.

Local Correction of Junta-Sums. Our next application considers the local correction problem for junta-sums over *grids*. Here a (possibly randomized) corrector is given oracle access to a function f that is known to be δ-close (in normalized Hamming distance) to some degree-d junta-sum P, and also given a point $\mathbf{a} \in \mathcal{S}^n$ and needs to output $P(\mathbf{a})$ (with high probability) while making few oracle queries to f.
In the list-correction setting, the amount of error δ may be too high for P to be defined uniquely by f and δ, but it may be known a priori that the list size is bounded. In the local list-correction problem, the goal for the corrector is to make a few queries to f to produce several "oracle" algorithms, such that for every degree d junta-sum P that is δ-close to f, there is an algorithm with oracle access to f that computes P. We refer the reader to Section 2 for more formal definitions.

Local correction algorithms for low-degree polynomials have played a central role in complexity theory, for example [GL89; STV01]. While most of the early works like [GKZ08; BL18] considered the setting where $\mathcal{S} = G = \mathbb{F}$, some recent works have considered the setting of $\mathcal{S} = \{0, 1\}$ and general abelian G such as [ABPSS25] (Note that when $|\mathcal{S}| = 2$, then degree-d polynomials are the same as degree-d junta-sums.)

For general $\mathcal{S}$ and Abelian group G, even the list-decoding radius was not completely understood till this work. We prove that for $\delta = |\mathcal{S}|^{-d} - \epsilon$ there are most $\mathcal{O}_\varepsilon(1)$ degree d junta-sums P that are δ-close to any given function f. (This bound is tight in that for $\delta = |\mathcal{S}|^{-d}$ the number of junta-sums grows with n.) This motivates the corresponding local list-correction problem, which we solve tightly in this work. We state an informal version below and point to Theorem 5.1 for the more precise version.

Theorem 1.5 (Local list-correction of junta-sums (Informal)). *For every set $\mathcal{S}$, every Abelian group G, every integer d and $\varepsilon > 0$, there exists an $L = L(\varepsilon, d, \mathcal{S})$ such that the following holds.*

There exists an algorithm $\mathcal{A}$ that on oracle access to a function $f : \mathcal{S}^n \to G$, outputs L oracle algorithms $\psi_1, \ldots, \psi_L$ such that for every degree d junta-sum $P : \mathcal{S}^n \to G$ that is $(1/s^d - \varepsilon)$-close to f, there exists $i \in [L]$ such that $\psi_i^f(\cdot)$ computes P (with high probability for every input).
The query complexity of $\mathcal{A}$ and $\psi_1, \ldots, \psi_L$ is $\mathrm{poly}(\log n)$.

This theorem is formalized as Theorem 5.1 with more explicit bounds on the error probability and query complexity, and is proved in Section 5.

This theorem generalizes a theorem of Amireddy, Behera, Paraashar, Srinivasan, and Sudan ([ABPSS25, Theorem 1.3.4]) who solved the corresponding problem over the Boolean cube $\{0, 1\}^n$. (Note that in the Boolean setting, junta-degree is the same as algebraic degree, and their result is thus expressed in terms of the latter phrase.) Our extension follows the same sequence of steps as employed in [ABPSS25]. Their work ultimately ends up using the expansion properties of Boolean multislices, which, as we've noted earlier, is well-understood. Extending their work to general grids requires a

number of changes that we elaborate on in Section 1.2.4, with the most significant change being the use of Theorem 1.1 instead of the expansion results on the Boolean slice.

1.2 Techniques and Proof Overview

In this section, we first review known methods for bounding the singular values of walks that respect symmetries and explain where there is a gap in knowledge. We then show how we overcome these challenges by overviewing the proof of our main theorem Theorem 1.3 in Section 1.2.2. Next, we give an overview of the proof of the ODLSZ theorem for multislices, Theorem 1.4, in Section 1.2.3. Finally, we discuss the proof of the local correction theorem for grids, Theorem 1.5 in Section 1.2.4.

1.2.1 Prior Approaches and Obstructions

We describe some prior cases where random walk matrices respecting symmetries (i.e., the first condition of Theorem 1.3) have been studied and explain the special properties in play there.

Boolean Hypercube and Cayley graphs A broad class of examples bounding eigenvalues of highly symmetric graphs are the bounds on the eigenvalues of Cayley graphs over abelian groups - this captures random walks on the Boolean hypercube and many more general settings. Here it is well known that the random walk matrix is *diagonalizable*[4] and the eigenvectors of the random walk matrix depend only on the group (and not the set of generators). This makes it possible to determine the entire eigenspectrum for many basic groups using Fourier analysis. We note that Cayley graphs over some non-abelian groups have been studied, but general results are mostly lacking. In these cases, the random walk matrix is typically not diagonalizable, but can be made block diagonal, using the representation theory of the underlying group. This is a complex tool, and many basic questions are unanswered as we elaborate below.

Boolean slices One well-studied setting that happens to be the special case of $s = 2$ of our problem is the setting of Boolean slices. Here $\mathcal{S} = \{0, 1\}$ and $\mathcal{S}_\mu^n$ is the balanced Boolean slice (all points in $\{0, 1\}^n$ of Hamming weight exactly $n/2$). This setting has particular relevance to the analysis of Boolean functions and combinatorics; see, e.g. [Del78; Fil16; Fil23]. The random walk matrices in this setting lie in the *Johnson scheme*, which is an algebra of symmetric matrices that commute with one another. This implies that all such matrices can be diagonalized *simultaneously*, i.e., there exists one unitary matrix U such that for every random walk matrix M on the Boolean slice that respects symmetries, we have that UMU^T is diagonal. This implies that all such matrices M have the same eigenvectors. The works [Fil16; Sri11] gave explicit descriptions of the common eigenspaces. This can be quite useful when analyzing the spectrum of such matrices and in a recent example [ABSS25] used this description to show that a particular random walk matrix on the balanced Boolean slice is a good spectral expander (see [ABSS25, Lemma 3.2]).

[4] A matrix $M \in \mathbb{R}^{N \times N}$ is said to be diagonalizable if there is a unitary matrix $U \in \mathbb{R}^{N \times N}$ such that UMU^T is a diagonal matrix.

1.2.2 Spectral Expansion of Multislice Walks

Turning to our setting, our matrix M is not *diagonalizable* and so the techniques from the analysis of Cayley graphs on abelian groups as well as the random walk on the Boolean slice, do not work in this setting. We have to resort to the use of representation theory, but here as we alluded to earlier, our understanding is not as complete. In what follows, we explain what representation theory implies for our setting and how we build on it.

Summary of known facts from representation theory The fact that our matrix M respects symmetries allows us to invoke results from the representation theory of the symmetric group Sym_n. We cover these results in detail in Proposition 3.1 and Theorem 3.3 (Parts (1) and (2)). Essentially, we can use representation theory to show that our matrix M can be block-diagonalized with relatively few blocks.

Specifically[5] there is an orthonormal matrix $U = U_{\mathcal{S},n}$ such that for every M respecting symmetries the matrix UMU^T is block diagonal with blocks $M_0, M_1, \ldots, M_t$ where t and the "shape" of the blocks is known from standard representation theory. In particular, $M_0 = [1]$ is just a 1×1 matrix that contributes the top singular value (which is 1), and $M_1, \ldots, M_t$ determine $\sigma_2(M) < 1$.

Now let us understand the structure of M_i's in more detail. Each block M_i is a Kronecker/tensor product of a "small" matrix $A_i \in \mathbb{R}^{m(i)\times m(i)}$ with a somewhat larger identity matrix i.e.,

$$M_i = A_i \otimes \mathrm{Id}_{k(i)}, \qquad \text{where } \mathrm{Id}_k \text{ is the } k \times k \text{ identity matrix.}$$

See Figure 1 for an informal pictorial description. Both the quantities $m(i)$ and $k(i)$ are known from the representation theory of Sym_n (and in particular only depend on $\mathcal{S}$ and n and are independent of the particular matrix M). However, the small matrices A_i's do depend on M, and more importantly, the matrix U is not too well-understood. (In particular, we need to understand the effect of U on A_i and this is not clear.) In particular, if we were to arrange i such that $m(i)$'s are non-decreasing, then $k(i)$'s are also non-decreasing. Intuition from Fourier analysis in the Abelian world would suggest that λ_2 comes from $M_1 = A_1 \otimes \mathrm{Id}_{k(1)}$ but, as far as we are aware, even this is *not known*.

Our analysis Given that the A_i's are not determined by only $\mathcal{S}$ and n, and U is not explicitly understood, we need to find some crude ways to bound the singular values of M. We give such an analysis in Section 3 and summarize the essence here. We start with the following informal observation.

Observation 1.6. *If for some $i \in [t]$, the quantity $k(i)$ is a polynomial larger than the Frobenius norm of M, then every singular value corresponding to the block M_i must necessarily be small.*

The observation follows from the fact that the Frobenius norm is lower bounded by the sum of the squares of the singular values of M_i, each of which repeats at least $k(i)$ times, so each singular value must be small.

[5] This conclusion only requires that M respects symmetries (see Theorem 1.3).

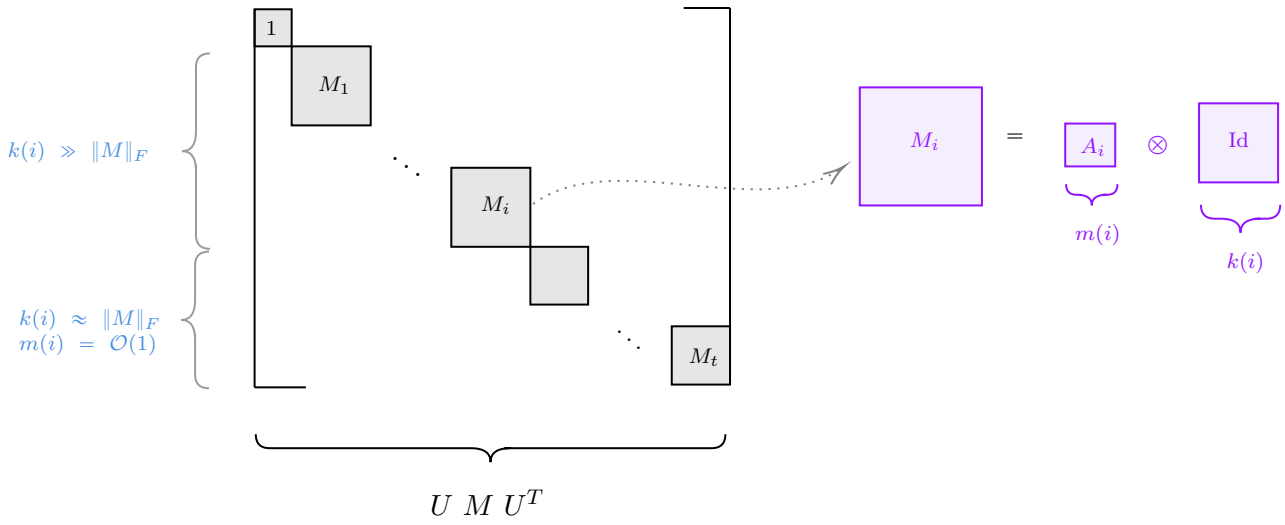

Figure 1: An informal visualization for the block diagonalization of our matrix M

In the context of our goal, the observation above immediately allows us to eliminate all the large blocks in the block diagonalization of M and turns the focus to the small blocks — where $k(i)$ is bounded by some polynomial in n. Here, standard facts of representations of Sym_n imply that the corresponding matrix A_i is of constant size (dependent on $\mathcal{S}$ and the exponent of n in $k(i)$ but not on n). However, this does not immediately translate to bounds on the singular values, since these depend on the actual matrix A_i and its entries. Ideally, we would have liked to get our hands on the singular vectors of M corresponding to singular vectors of the A_i's (after a change of basis according to U), but such vectors were not known (and we do not get such vectors either). Fortunately, a collection of vectors that span the space corresponding to the singular vectors of A_i's is known. In particular, we use a description given by Dafni, Filmus, Lifshitz, Lindzey, and Vinyals [DFLLV21] — see Definition 3.8. Our main contribution adds two observations about this collection of vectors, called "special vectors" below.

Our main contribution here does get something almost as good, for our purposes (i.e., to show that each M_i has top singular value $o_n(1)$):

> *We observe that the special vectors given in Definition 3.8 are "weakly orthogonal" in the sense that they have $\Omega(1)$ volume (in the space they span). We further observe that these functions are junta-like and so shrink significantly when acted on by ε-almost k-wise independent matrices for sufficiently large constant k. (See Theorem 3.3, Parts 3(c) and 3(d)).*

More specifically, the work of [DFLLV21] yields $\dim(A_i) = m(i)$ many vectors (see Section 3.3) of M that are supported on coordinates of one of the blocks of M_i after transforming the basis according to U. We show that while these vectors do not form an orthogonal basis, they are sufficiently divergent to ensure their determinantal volume is large (see Lemma 3.12). Thus, bounding the length of the vectors obtained by applying the linear map A_i by $o_n(1)$ suffices to bound the spectral norm of A_i and hence also M_i. Details of this part may be found in Section 3.3. We give some more insight in the next paragraph.

To show our singular value bounds, we use the fact that the vectors in the basis correspond to $\mathcal{O}(1)$-junta's. Specifically, note that a vector that M acts on can be viewed as a function f from $\mathcal{S}_\mu^n$ to $\mathbb{R}$, which can in turn be viewed as a partial function from $\mathcal{S}^n$ to $\mathbb{R}$. We show that this function depends on only $\mathcal{O}(1)$-coordinates of the input vector. (See Section 3.3.) This property now combines nicely with our third condition in Theorem 1.3 which asserts that after multiplication by M any vector f looks essentially random when projected on $\mathcal{O}(1)$-coordinates and so has little correlation left with f — this immediately translates into an upper bound on the singular value of M corresponding to coordinates in M_i, and yields a proof of Theorem 1.3.

Why does our Theorem 1.1 hold only for nearly-balanced walks? The reason is related to Observation 1.6. We note that the Frobenius norm condition is not completely natural, and indeed the natural matrices in our applications do not satisfy this condition (and we have to find workarounds). The Frobenius norm restriction is satisfied by nearly-balanced walks as considered in Theorem 1.1, and indeed it is one of the reasons why Theorem 1.1 is restricted to such nearly balanced walks.

1.2.3 Distance Lemma over Balanced Multislice

In this subsection, we discuss the proof overview for Theorem 1.4. The strategy is a generalization of the proof for [ABSS25, Lemma 3.2]. The idea is to find a random copy of $\mathcal{S}^{n/s}$ inside the balanced multislice $\mathcal{S}_\mu^n$ such that it is a good *sampler* for $\mathcal{S}_\mu^n$, i.e. if we choose points from this subgrid $\mathcal{S}^{n/s}$ at random, then the corresponding points in $\mathcal{S}_\mu^n$ behave like random samples. As we explain now, this guarantee essentially allows us to move from balanced multislice to subgrid, where we have a complete understanding of distance. For every non-zero d-junta-sum $P : \mathcal{S}_\mu^n \to G$, we choose a random copy of $\mathcal{S}^{n/s}$ inside $\mathcal{S}_\mu^n$ and restrict P to this copy. With the sampling guarantee, we can argue that the restricted d-junta-sum is also non-zero on the subgrid $\mathcal{S}^{n/s}$, and we get the claimed bound by applying Claim 2.6 on this restricted polynomial. Next, we explain the process of finding a random copy of the n/s-dimensional subgrid inside the balanced multislice.

The key step in our proof is to show that we can find a random copy of $\mathcal{S}^{n/s}$ inside $\mathcal{S}_\mu^n$, which is a sampler for $\mathcal{S}_\mu^n$. We do it by randomly grouping the coordinates $x_1, \ldots, x_n$ into n/s buckets of size s and in each bucket, we randomly assign distinct values to the s coordinates. We prove that for two random points in $\mathcal{S}^{n/s}$, their corresponding points in the balanced multislice $\mathcal{S}_\mu^n$ are almost pairwise independent. We show this via the second moment method and the expander mixing lemma. We use our main theorem Theorem 1.3 to show that the random walk on $\mathcal{S}_\mu^n$ arising from the above random process has good spectral expansion, making the expander mixing lemma applicable in this context.

1.2.4 Local List Correction for Junta-Sums

Our local list corrector (see Theorem 5.1) is a generalization of [ABPSS25, Theorem 1.3.4] to d-junta-sums and arbitrary grids $\mathcal{S}^n$ (instead of degree-d *polynomials* and *Boolean* cube). We do not dwell on the algorithm here, but only highlight and discuss the key technical difference in our work and the previous work of [ABPSS25]. We request the reader to please refer to Section 5 for more

details on the algorithm.

An important step of our local list corrector involves a random restriction of $\mathcal{S}^{sk}$ to a subgrid $\mathcal{S}^k$, as follows: We randomly group the sk coordinates into k groups of size s, and identify all the s coordinates in a group together by a single new coordinate. To show that our local list corrector has small error probability, we need the following guarantee from the above random restriction: If a d-junta-sum $P \in \mathcal{J}_d(\mathcal{S}^{sk})$ is non-zero on the balanced multislice, then with high probability, it continues to be a non-zero junta-sum on $\mathcal{S}^k$ after the random restriction. For this, we show that the above-mentioned random process can be interpreted as finding a random copy of the balanced multislice in $\mathcal{S}^k$ inside the balanced multislice in $\mathcal{S}^{sk}$. Similar to the distance lemma for multislices (Theorem 4.2), we show that we get a good sampler using Theorem 1.3.

We now briefly touch upon some of the additional challenges in going from the Boolean case of [ABPSS25] to junta-sums over grids $\mathcal{S}^n$, in the context of local list-correction. For the local correction algorithm in the unique decoding regime, the main idea is to reduce the problem to the Boolean case but over a biased distribution instead of the uniform one; the proof then proceeds by a mostly straightforward generalization of the local corrector from [ABPSS25] for the uniform distribution. The overall template for proving the combinatorial bound is also similar to that over the Boolean cube, except now we will need more general anti-concentration lemmas and distance lemmas for junta-sums. As already described in the above paragraph, going from the combinatorial bound for list-decodability to the local list-corrector is the main technical challenge we overcome in this work by making use of the fact that a certain random embedding of the multislice of $\mathcal{S}^k$ inside the multislice of $\mathcal{S}^n$ is a good sampler.

1.3 Organization

In Section 2, we give some definitions that we are going to use throughout the article. In Section 3, we prove the main theorem of our work (Theorem 1.3), which itself is organized as follows: we start with giving some necessary background on representation theory for the symmetric group, then use it to prove Theorem 1.3, and finally show that "typical" random-walk matrices are good spectral expanders. In the subsequent sections, we give applications of our main theorem. In Section 4, we prove a near-optimal distance lemma for junta-sums and polynomials over balanced multislice (see Theorem 4.2 and Theorem 5.10). In Section 5, we give a local list corrector for d-junta-sums over $\mathcal{S}^n$ (see Theorem 5.1).

2 Preliminaries

We begin by describing some standard notation and terminology we will use throughout the paper.

For a set of parameters $\alpha_1, \ldots, \alpha_t$, the notation $\mathcal{O}_{\alpha_1,\ldots,\alpha_t}(\cdot)$ hides factors depending on $\alpha_1, \ldots, \alpha_t$. Similarly for $\Theta_{\alpha_1,\ldots,\alpha_t}(\cdot), \Omega_{\alpha_1,\ldots,\alpha_t}(\cdot)$ and so on. Although this is not standard, we will use the notation $\widetilde{\mathcal{O}}(\cdot)$ to hide $(\log\log n)^{\mathcal{O}(1)}$ factors (generally this notation is used to hide $(\log n)^{\mathcal{O}(1)}$) factors). We use $|\mathbf{x}|$ to denote the Hamming weight of $\mathbf{x}$, i.e., the number of non-zero coordinates. Let $\mathrm{Bern}(p)^n$ denote the distribution over $\{0,1\}^n$ where each bit is chosen from the Bernoulli distribution $\mathrm{Bern}(p)$ independently. For two distributions X, Y over the same finite domain, we let

$\mathrm{SD}(X, Y)$ denote the statistical distance between the distributions. We let $\|\mathbf{v}\|_2$ denote the ℓ_2 norm of a vector $\mathbf{v} \in \mathbb{R}^N$.

For any $s \in \mathbb{N}$, we use $\mathbb{Z}_s$ to denote the cyclic group $\mathbb{Z}/s\mathbb{Z}$, and not to be confused by the p-adic field $\mathbb{Z}_s$. We say that a group is a *torsion group* if all its elements have finite order. The *exponent* of a torsion group is the least common multiple of the orders of all its elements.

Let n and s be two natural numbers where n is divisible by s and let $\mathcal{S}_\mu^n$ denote the balanced multislice over a finite set $\mathcal{S}$ of size s, i.e.,

$$\mathcal{S}_\mu^n := \left\{ \mathbf{a} \in \mathcal{S}^n \;\middle|\; \forall \sigma \in \mathcal{S}, |\{i \in [n] | a_i = \sigma\}| = \frac{n}{s} \right\}.$$

Similarly, for any $\lambda = (\lambda_0, \ldots, \lambda_{s-1})$ with $\lambda_0 + \ldots + \lambda_{s-1} = n$, we define the multislice $\mathcal{S}_\lambda^n$ as follows:

$$\mathcal{S}_\lambda^n := \{ \mathbf{a} \in \mathcal{S}^n \mid \forall \sigma \in \mathcal{S}, |\{i \in [n] | a_i = \sigma\}| = \lambda_i \}.$$

Then, we define the generalized Hamming distance between points in the (balanced) multislice as follows:

Definition 2.1 (**Generalized Hamming distance).** *We define the generalized Hamming distance $\Delta(\mathbf{a}, \mathbf{b})$ between two points $\mathbf{a}, \mathbf{b} \in \mathcal{S}_\mu^n$ to be the $\mathcal{S} \times \mathcal{S}$ matrix where the (σ, τ)-th entry is given by $|\{i \in [n] : a_i = \sigma \text{ and } b_i = \tau\}|$.*

Example 2.1.1 (A generalized Hamming distance matrix for $n = 9$ and $s = 3$.)**.** *Let* $\mathbf{u} = 000111222$ *and* $\mathbf{v} = 110201022$*. Then,*

$$\Delta(\mathbf{u}, \mathbf{v}) \;=\; \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix}.$$

We now define the notion of nearly balanced generalized Hamming distance (or *C-balanced* profiles to be more precise).

Definition 2.2 (**C-balanced generalized Hamming distance).** *For $C \geqslant 0$, we say that a generalized Hamming distance parameter $P \in \mathbb{Z}^{\mathcal{S} \times \mathcal{S}}$ w.r.t a multislice $\mathcal{S}_\mu^n$ is* C-balanced *if every entry of P is in the range $\frac{m}{s} \pm \sqrt{Cm \log m}$ where $m = n/s$.*

We now use generalized Hamming distance matrices to define a random walk on the balanced multislice $\mathcal{S}_\mu^n$.

Definition 2.3 (**Random walk determined by a generalized Hamming distance matrix).** *We say that a generalized Hamming distance matrix $P \in \mathbb{Z}^{\mathcal{S} \times \mathcal{S}}$* determines *a random walk matrix, denoted W_P, if for each vertex $\mathbf{a} \in \mathcal{S}_\mu^n$ in the multislice, the random step corresponding to W_P is obtained by picking, uniformly at random, a vertex $\mathbf{b} \in \mathcal{S}_\mu^n$ such that $\Delta(\mathbf{a}, \mathbf{b}) = P$. That is, the $\mathbf{a}$-th row of W_P (denoted $W_P(\mathbf{a})$) is the uniform distribution over $\{\mathbf{b} \in \mathcal{S}_\mu^n : \Delta(\mathbf{a}, \mathbf{b}) = P\}$.*

We now give some necessary background for random walk matrices more generally. We refer the reader to the survey [Vad12] for more discussion.

We say that a matrix $W \in \mathbb{R}^{N \times N}$ is a *random walk matrix* if for every $i \in [N]$ (which may be referred to as *vertices*), the i-th row of the matrix, denoted $W(i)$, is a probability distribution over $[N]$. It is clear that every random walk matrix has an eigenvector of $\mathbf{1}$ with eigenvalue 1. If W is symmetric, then it has real eigenvalues $1 = \lambda_1 \geqslant \cdots \geqslant \lambda_N \geqslant -1$; and we define $\lambda_2(W) = \max\{|\lambda_2|, |\lambda_n|\}$ to be the second largest eigenvalue of W in absolute value. Equivalently, one can show that

$$\lambda_2(W) = \max_{\mathbf{v} \in \mathbb{R}^N : \mathbf{v}^\top \mathbf{1} = 0} \frac{\|W\mathbf{v}\|_2}{\|\mathbf{v}\|_2}.$$

We will also deal with random walk matrices that are not necessarily symmetric. We say a square matrix is *stochastic* if all its entries are non-negative and each row element sums to 1. We say that a matrix is *doubly stochastic* if both the matrix and its transpose are stochastic. We observe that doubly stochastic matrices have $\mathbf{1}$ as both a left eigenvector and right eigenvector. Furthermore, it has singular values $1 = \sigma_1 \geqslant \sigma_2 \geqslant \ldots \sigma_N \geqslant 0$, where N is the order of the matrix; and we use $\sigma_2(W)$ to mean σ_2. Similar to the case of symmetric matrices, we have for every doubly stochastic matrix $W \in \mathbb{R}^{N \times N}$:

$$\sigma_2(W) = \max_{\mathbf{v} \in \mathbb{R}^N : \mathbf{v}^\top \mathbf{1} = 0} \frac{\|W\mathbf{v}\|_2}{\|\mathbf{v}\|_2}.$$

For symmetric matrices, singular values are simply the absolute values of the eigenvalues. Hence, if W is a symmetric random walk matrix with eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_N$ and singular values $\sigma_1 \geqslant \cdots \geqslant \sigma_N$, then $\lambda_1 = \sigma_1 = 1$ and $\lambda_2(W) = \sigma_2(W)$.

We observe the following property of the random walks determined by generalized Hamming distance between points on the multislice (Definition 2.3).

Observation 2.4. *For every generalized Hamming distance matrix $P \in \mathbb{Z}^{s \times s}$ defined with respect to a multislice $\mathcal{S}^n_\mu$, we have that $W_P^\top = W_{P^\top}$ is a random walk matrix. In particular, W_P is doubly stochastic.*

We will now show how to bound the eigenvalues of a convex combination of random walk matrices.

Lemma 2.5 (Singular value bound for convex combinations). *Suppose $W = \sum_{i \in [t]} \alpha_i W_i$, where $W_i \in \mathbb{R}^{N \times N}$ are doubly stochastic matrices and $\alpha_1 \ldots, \alpha_t \in [0, 1]$ are such that $\sum_{i \in [t]} \alpha_i = 1$. Let $S \subseteq [t]$ be arbitrary. Then W is also a doubly stochastic matrix with*

$$\sigma_2(W) \leqslant \max_{i \in S}\{\sigma_2(W_i)\} + \sum_{i \notin S} \alpha_i.$$

Proof. We observe that each row (similarly column) of W is a convex combination of probability distributions, so is also a probability distribution; hence W is indeed doubly stochastic. In other words, $\mathbf{1}$ is both a left eigenvector and right eigenvector. Hence, we have that

$$\sigma_2(W) = \max_{\mathbf{u} \in \mathbb{R}^N : \mathbf{u}^\top \mathbf{1} = 0 \text{ and } \|\mathbf{u}\|_2 = 1} \|W\mathbf{u}\|_2.$$

Now letting $\mathbf{u} \in \mathbb{R}^N$ be an arbitrary vector such that $\|\mathbf{u}\|_2 = 1$ and $\mathbf{u}^\top \mathbf{1} = 0$, we will bound $\|W\mathbf{u}\|_2$. We have

$$\|W\mathbf{u}\|_2 = \left\| \sum_{i\in[t]} \alpha_i W_i \mathbf{u} \right\|_2 \leqslant \sum_{i\in[t]} \alpha_i \|W_i\mathbf{u}\|_2 = \sum_{i\in S} \alpha_i \|W_i\mathbf{u}\|_2 + \sum_{i\notin S} \alpha_i \|W_i\mathbf{u}\|_2$$
$$\leqslant \left(\sum_{i\in S} \alpha_i\right)\left(\max_{i\in S} \|W_i\mathbf{u}\|_2\right) + \left(\sum_{i\notin S} \alpha_i\right)(1) \leqslant \max_{i\in S}\{\sigma_2(W_i)\} + \sum_{i\notin S} \alpha_i,$$

where we are using the triangle inequality for the first inequality, and that each W_i is a random walk matrix for the second inequality. ■

We now move on to the definitions needed for our local list-correction application in Section 5.

Local Correction and Junta-Sums

We say that a family of functions $\mathcal{F}$ from a finite domain D to a (finite or infinite) co-domain G, is (q, ε)-*locally correctable* if there exists a q-query algorithm $\mathcal{A}$, which when given query access to a function $f : D \to G$ such that $\delta(f, P) \leqslant \varepsilon$ for some $P \in \mathcal{F}$, and an input index $i \in D$, outputs $P(i)$ with probability at least $3/4$. In words, the algorithm $\mathcal{A}$ is able to "correct" any given index of the received word f by making only a few queries. Since P has to be unique for such an algorithm to exist, we are always in the regime when the fraction of errors is less than half the distance of the code i.e., $\varepsilon < \delta(\mathcal{F})/2$.

We say that $\mathcal{F}$ is (ε, L)-*list-decodable* if if for every function $f : D \to \mathcal{F}$, there exists at most L functions $P \in \mathcal{F}$ such that $\delta(f, P) \leqslant \varepsilon$. While this is a purely combinatorial guarantee for the code, the notion of local list-correction makes it more "algorithmic".

We say that $\mathcal{F}$ is $(\varepsilon, q_1, q_2, L)$ *locally list-correctable* if there exists a q_1-query algorithm $\mathcal{A}$, which when given query access to f, outputs at most L many q_2-query local correction algorithms $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_L$ such that for every $P \in \mathcal{F}$ such that $\delta(f, P) \leqslant \varepsilon$, there exists at least one index $i \in [L]$ such that $\mathcal{A}_i$ is a local correction algorithm for P i.e., on input $i \in D$, it makes q_2 queries to f and outputs $P(i)$ with probability at least $3/4$.

For an Abelian group G, let $\mathcal{J}_d(\mathcal{S}^n, G)$ (or simply $\mathcal{J}_d$ when $\mathcal{S}$ and G are clear from context) denote the family of functions from $\mathcal{S}^n \to G$ that can be expressed as a sum of d-juntas (i.e., a *d-junta-sum*). We may sometimes also refer to d-junta-sums as functions of *junta-degree* d. We then have the following observation regarding d-junta-sums.

Claim 2.6 (*Distance of junta-sums*, see e.g. [ASS23], Claim 2.7)**.** *For every two distinct junta-sums $P \neq Q \in \mathcal{J}_d(\mathcal{S}^n, G)$ where $|\mathcal{S}| = s$, we have*

$$\Pr_{\mathbf{a}\sim\mathcal{S}^n}[P(\mathbf{a}) \neq Q(\mathbf{a})] \geqslant \frac{1}{s^d}.$$

That is, junta-sums form a code of distance $\delta_{\mathcal{J}} = 1/s^d$ where $s = |\mathcal{S}|$. Indeed, the local correction and list-decodability properties of this family only depends on the size of $\mathcal{S}$, so we will often assume that $\mathcal{S} = \mathbb{Z}_s$ or $\mathcal{S} = [s]$ without loss of generality. We also use the following claim, where for $a \in \mathbb{Z}_s$,

the function $\delta_a : \mathbb{Z}_s \to \mathbb{Z}$ is defined as: $\delta_0(x) = 1$, and for $a \neq 0$, we define $\delta_a(x) = 1$ if $x = a$, and $\delta_a(x) = 0$ otherwise.

Claim 2.7 (**Junta-polynomial representation**, [ASS23] Claim 2.5)**.** *Every $P \in \mathcal{J}_d(\mathbb{Z}_s^n, G)$ can be uniquely expressed as:*

$$P(\mathbf{x}) = \sum_{\mathbf{a} \in \mathbb{Z}_s^n : |\mathbf{a}| \leqslant d} g_{\mathbf{a}} \cdot \prod_{i \in [n] : a_i \neq 0} \delta_{a_i}(x_i), \qquad \text{where each } g_{\mathbf{a}} \in G.$$

We call the above representation $\sum_{\mathbf{a} \in \mathbb{Z}_s^n} g_{\mathbf{a}} \cdot \prod_{i \in [n] : a_i \neq 0} \delta_{a_i}(x_i)$ as a *junta-polynomial* and its *junta-degree* is the size of the largest $|\mathbf{a}|$ such that the coefficient $g_{\mathbf{a}} \neq 0$; in particular, we call the terms being added as *monomials*. Generalizing Claim 2.7 one can show that every function $f : \mathbb{Z}_s^n \to G$ has a unique junta-polynomial representing it and f is a d-junta-sum if and only if the degree of that junta-polynomial is at most d. In turn, this immediately implies that f *depends* on the i-th coordinate if and only if the variable x_i appears (as $\delta_a(x_i)$ for some $a \in \mathbb{Z}_s \backslash \{0\}$) in a non-zero monomial in the junta-polynomial representation.

Partitions and Tableaux

We end this section with some more terminology about integer partitions and multislices, which will be needed in our proofs. All the definitions in this subsection are standard and can be found in any standard text on algebraic combinatorics or representation theory for the symmetric group. For example, see [Sag13, Chapter 2] or [Sta12; Sta24].

Partitions For every natural number $n \in \mathbb{N}$, let $\mathcal{P}(n)$ denote the set of partitions of n. We will frequently use Ferrers diagram to represent partitions. Let $\lambda^* \in \mathcal{P}(n)$ denote the dual partition of λ.

SYT and SSYT For a partition $\lambda \in \mathcal{P}(n)$, a *standard Young tableau* is a tableau of shape λ in which the entries in each row and each column are *strictly* increasing. A *semi-standard Young tableau* is a tableau of shape λ in which the entries in each row are *weakly* increasing and entries in each column are *strictly* increasing. For a pair of partitions $\lambda, \mu \in \mathcal{P}(n)$, the set $\mathrm{SSYT}(\lambda, \mu)$ denotes the set of semi-standard Young tableaux of shape λ and type μ. Similarly, $\mathrm{SYT}(\lambda)$ denotes the set of standard Young tableaux of shape λ.

For any $\lambda, \mu \in \mathcal{P}(n)$, we associate two quantities:

- f_λ denotes the number of Standard Young Tableaux of shape λ with content $[n]$, i.e. $f_\lambda = |\mathrm{SYT}(\lambda)|$.
- $K_{\lambda\mu}$ denotes the number of distinct Semi-Standard Young Tableaux of shape λ and type μ, i.e. $K_{\lambda\mu} = |\mathrm{SSYT}(\lambda, \mu)|$. This is also known as the Kostka number of the pair (λ, μ).

Dominance Order For two partitions $\lambda, \mu \in \mathcal{P}(n)$, *dominance order* is a partial order on partitions, defined as follows: Suppose $\lambda = (\lambda_1, \ldots, \lambda_\ell)$ and $\mu = (\mu_1, \ldots, \mu_m)$, then $\lambda \trianglerighteq \mu$ if for every

$1 \leqslant i \leqslant \min\{\ell, m\}$,

$$\lambda_1 + \ldots + \lambda_i \;\geqslant\; \mu_1 + \ldots + \mu_i.$$

For every positive integer n, Sym_n denotes the group of permutations on n elements and $\mathrm{Sym}[\mathcal{S}]$ denotes the group of permutations on $\mathcal{S}$.

Linear Algebra

Singular Value Decompositions For a matrix $M \in \mathbb{R}^{n\times n}$, the *singular value decomposition* (SVD) of M is given by orthonormal matrices $U, V \in \mathbb{R}^{n\times n}$ such that:

$$M \;=\; UDV^{-1} \;\Leftrightarrow\; M \;=\; UDV^T, \qquad\qquad (V^{-1} = V^T \text{ for orthonormal } V),$$

where D is a diagonal entries and the diagonal entries of D are the *singular values* of M. We will denote the singular values of M by $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n$.
In particular, if M has all real eigenvalues, then $V = U$ and the singular values correspond to the absolute values of the eigenvalues.

We now explain how SVDs behave under the tensor product. For two matrices M_1 and M_2 with SVDs

$$M_1 \;=\; U_1 D_1 V_1^T \quad \text{and} \quad M_2 \;=\; U_2 D_2 V_2^T,$$

then the SVD of $M_1 \otimes M_2$ is:

$$M_1 \otimes M_2 \;=\; (U_1 \otimes U_2)\cdot(D_1 \otimes D_2)\cdot(V_1 \otimes V_2)^T. \tag{1}$$

Definition 2.8 (Volume of a parallelepiped)**.** *Suppose $\{v_1, \ldots, v_r\} \in \mathbb{R}^r$ is a set of linearly independent vectors. Fix an arbitrary total order '$\prec$' on the vectors i.e., there exists a $\pi \in \mathrm{Sym}_r$ such that*

$$v_{\pi(1)} \prec v_{\pi(2)} \prec \cdots \prec v_{\pi(n)}.$$

Let $\widetilde{v}_{\pi(1)} = v_{\pi(1)}$ and for every $2 \leqslant i \leqslant r$, let $\widetilde{v}_{\pi(i)}$ denote the vector orthogonal to $\mathrm{span}\left\{v_{\pi(j)} \;\middle|\; j < i\right\}$. *Then the* volume *of the parallelepiped spanned by $v_1, \ldots, v_r$, denoted by* $\mathrm{Vol}(\{v_1, \ldots, v_r\})$, *is defined to be*

$$\prod_{j=1}^{r} \|\widetilde{v}_i\|,$$

where $\|\cdot\|$ is the norm with respect to the standard inner product on $\mathbb{R}$.
It also turns out that the volume is equal to $|\det(\Lambda)|$ where the columns of Λ are $v_1, \ldots, v_r$.

For any matrix $A \in \mathbb{R}^{r\times r}$, we will denote by $\|A\|_2$ the spectral norm of A i.e.

$$\|A\|_2 \;=\; \sup_{\mathbf{x}\neq\mathbf{0}} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \;=\; \max_{\sigma \text{ is a singular value of } A} \sigma. \tag{2}$$

Subgrids of $\mathcal{S}^n$

It will be very useful in our algorithms to be able to restrict the given function to a smaller subgrid and analyze this restriction. We construct such subgrids by first permuting a subset of the variables and then identifying them into a smaller set of variables. More precisely, we have the following definition.

Definition 2.9 (Embedding a smaller grid into $\mathcal{S}^n$)**.** *Fix any $k \in \mathbb{N}$ and $k \leqslant n$. Let $h : [n] \to [k]$ be a hash function. For each $i \in [n]$, let $\Pi_i \in \mathrm{Sym}[\mathcal{S}]$ (i.e. a permutation on elements of $\mathcal{S}$) and $\Pi = (\Pi_1, \ldots, \Pi_n)$. For every $\mathbf{y} \in \mathcal{S}^k$, define $x_{h,\Pi}(\mathbf{y}) \in \mathcal{S}^n$ as follows:*

$$x_{h,\Pi}(\mathbf{y})_i = \Pi_i(y_{h(i)}), \qquad \text{for all } \ i \in [n]$$

and the subset $C_{h,\Pi} \subset \mathcal{S}^n$ is defined as:

$$C_{h,\Pi} \ = \ \left\{ x_{h,\Pi}(\mathbf{y}) \,\middle|\, \mathbf{y} \in \mathcal{S}^k \right\}$$

Further, a random subgrid $C_{h,\Pi}$ is obtained by sampling a uniformly random permutation $\Pi_i \sim \mathsf{Sym}[\mathcal{S}]$ independently for all $i \in [n]$ and sampling a uniformly random hash function $h : [n] \to [k]$.

In simple words, the above definition gives us a way to embed a k-dimensional grid $\mathcal{S}^k$ into a n-dimensional grid $\mathcal{S}^n$, where the hash function h governs how the k-coordinates are mapped into n-coordinates and Π governs which value the i^{th} coordinate takes.

The following sampling lemma (proved in Appendix B) will be useful for local (list) correction of junta-sums.

Lemma 2.10 (**Sampling lemma for random subgrids**)**.** *Let $C_{h,\Pi} \subset \mathcal{S}^n$ be a subgrid sampled randomly as per Definition 2.9. Fix any $T \subseteq \mathcal{S}^n$ and let $\mu := |T|/s^n$. Then, for any $\varepsilon, \eta > 0$*

$$\Pr_{h,\Pi}\left[\left| \frac{|T \cap C_{h,\Pi}|}{s^k} - \frac{|T|}{s^n} \right| \geqslant \varepsilon \right] \ < \ \eta$$

as long as $k \geqslant \max\left\{ \frac{A}{\varepsilon^8 \eta^4} \cdot \log\left(\frac{1}{\varepsilon\eta}\right), B \cdot s^4 \log s \right\}$ for a large enough absolute constants $A, B > 0$.

3 Singular Value Bounds for Random Walks on Balanced Multi-slices

Organization of this section. In this section, we will prove Theorem 1.3. At a high level, the proof proceeds as follows:

1. We give the necessary background on representation theory for finite groups in Proposition 3.1. We then instantiate it for Sym_n and state the exact requirements we need to prove for our purpose in Theorem 3.3. These two steps can be found in Section 3.1.
2. We then in Section 3.2 argue that Theorem 3.3 is sufficient to prove Theorem 1.3.
3. We devote Section 3.3 and Section 3.4 to prove Theorem 3.3. In particular, we start with describing some "special vectors" which we use to prove Theorem 3.3. The description of these vectors is combinatorial in nature, and we prove certain properties about them. Finally, we prove Theorem 3.3 in Section 3.4.

Notation For any two natural numbers n and s with n divisible by s, μ denotes the s-tuple $(n/s, \ldots, n/s)$ and $\mathcal{S}^n_\mu$ is the set of all points in $\mathcal{S}^n$ which are on the balanced multi-slice μ. Let $N := |\mathcal{S}^n_\mu| = \binom{n}{n/s,\ldots,n/s}$. For any n, $\mathcal{P}(n)$ denotes the set of partitions of n. Throughout this section, we will assume that s is an absolute constant.

3.1 Representation Theory Primer

In this and the following subsections, whenever we mention a representation, we refer to a complex finite-dimensional representation of finite groups. For interested readers, we refer to [Sag13, Chapter 1] for the relevant background on the representation theory of finite groups.

Let G be a finite group and V be a $\mathbb{C}$-vector space with $\dim(V) < \infty$. Let $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ be an inner product that is preserved under the representation ρ, i.e. for every $g \in G$, for every $u, v \in V$,

$$\langle \rho(g)u,\ \rho(g)v \rangle \ = \ \langle u, v \rangle.$$

Basic results of representation theory imply the following.

Proposition 3.1 (Standard facts on representations for finite groups)**.** *Suppose $\rho : G \to \mathrm{GL}(V)$ is a representation of G and $W \in \mathrm{End}(V, V)$ commutes with the representation ρ, i.e. for every $g \in G$,*

$$\rho(g) \circ W \ \equiv \ W \circ \rho(g) \qquad \textit{(equivalent as linear operators).}$$

In other words, W is an intertwining operator from (ρ, V) to itself. Then,

1. *There exists sub-representations $V_1, \ldots, V_r$ such that*

$$V \ \cong \ \bigoplus_{i=1}^{r} V_i$$

where $\{V_1, \ldots, V_r\}$ are orthogonal subspaces (with respect to the inner product mentioned above).
Moreover, for every $i \in [r]$, the following holds. There exists an irreducible representation U_i and an integer $m_i \geqslant 1$ such that

$$V \ \cong \ \bigoplus_{j=1}^{m_i} V_{i,j} \qquad \textit{and} \qquad V_{i,1} \cong \cdots \cong V_{i,m_i} \cong U_i.$$

The subrepresentation V_i is called as the isotypic component *of (ρ, V) corresponding to the irreducible representation U_i. This means*

$$\dim(V_i) = m_i \cdot \dim(U_i) \qquad \textit{and} \qquad \dim(V) = \sum_{i=1}^{r} m_i \cdot \dim(U_i).$$

2. *Fix any $i \in [r]$. Let $\mathcal{B}_{i,1}$ be an ordered basis for $V_{i,1}$. For every $2 \leqslant j \leqslant m_i$, there exists an unique isomorphism $\mathcal{L}_{i,j} : V_{i,1} \to V_{i,j}$. Let $\mathcal{B}_{i,j}$ denote the image of $\mathcal{B}_{i,1}$ under $\mathcal{L}_{i,j}$, and $\mathcal{B}_{i,j}$ is a basis for $V_{i,j}$. Let $\mathcal{B}_i$ be an ordered basis for V_i obtained by concatenating $\mathcal{B}_{i,1}, \ldots, \mathcal{B}_{i,m_i}$ in that order. Similarly, $\mathcal{B}$ obtained by concatenating $\mathcal{B}_1, \ldots, \mathcal{B}_r$ is an ordered basis for V.*

3. *The linear map W preserves the isotypic components, i.e. for each $i \in [r]$, $W|_{V_i} \in \mathrm{End}(V_i, V_i)$. In particular, under the ordered basis $\mathcal{B}$, the map W (when viewed as a $\dim(V) \times \dim(V)$ matrix) has the following structure:*

$$W \;=\; \bigoplus_{i=1}^{r} W_i, \qquad\qquad (\textit{direct sum of matrices})$$

where for each $i \in [r]$, W_i is a $\dim(V_i) \times \dim(V_i)$ dimensional matrix.
Furthermore, for every $i \in [r]$, there exists a unique $m_i \times m_i$ dimensional matrix A_i such that

$$W_i \;=\; A_i \otimes \mathrm{Id}_{\dim(U_i)}, \qquad \textit{where } \mathrm{Id_k} \textit{ is the } k \times k \textit{ dimensional identity matrix.}$$

4. *Fix any $i \in [r]$. For every non-zero $v \in V_{i,1}$, define a $\mathbb{C}$-space $Y_{i,v} := \mathrm{span}\{v, \mathcal{L}_{i,2}(v), \ldots, \mathcal{L}_{i,m_i}(v)\}$. Then $W_i|_{Y_{i,v}} \in \mathrm{End}(Y_{i,v}, Y_{i,v})$ and $W_i|_{Y_{i,v}} = A_i$ when we represent the linear map in the ordered basis $(v, \mathcal{L}_{i,2}(v), \ldots, \mathcal{L}_{i,m_i}(v))$.*

The following corollary is immediate from the third item of Proposition 3.1.

Corollary 3.2. *We follow the same notation from Proposition 3.1. Suppose $\{\beta_1^i, \ldots, \beta_{m_i}^i\}$ is the multi-set of singular values of A_i where each $\beta_j^i \in \mathbb{C}$. Then using Equation (1), we get that in the multi-set of singular values of W_i, the frequency of β_j^i is equal to the frequency of β_j^i in the multi-set $\{\beta_1^i, \ldots, \beta_{m_i}^i\}$ times $\dim(U_i)$.*

Now we turn to the representation theory for Sym_n. In particular, we will be considering the representation of Sym_n on the space of functions on a slice of $\mathcal{S}^n$.

Space of functions on a slice For any partition $\lambda \in \mathcal{P}(n)$, let $\mathbb{M}^\lambda$ denote the $\mathbb{C}$-vector space of functions over the slice $\mathcal{S}_\lambda^n$ i.e.

$$\mathbb{M}^\lambda \;=\; \{f : \mathcal{S}_\lambda^n \to \mathbb{C}\}.$$

It is easy to see that $\dim(\mathbb{M}^\lambda) = |\mathcal{S}_\lambda^n|$. There is a natural action of the symmetric group Sym_n on $\mathbb{M}^\lambda$: For all $\pi \in \mathrm{Sym}_n$ and for all $f \in \mathbb{M}^\lambda$,

$$(\pi f)(\mathbf{x}) \;=\; f(\pi^{-1}\mathbf{x}), \qquad\qquad \text{where } \; \pi^{-1}\mathbf{x} = (x_{\pi^{-1}(1)}, \ldots, x_{\pi^{-1}(n)})$$

Representation of Sym_n Let $(\rho, \mathbb{M}^\mu)$ denote the following $\mathbb{C}$ representation of Sym_n:

$$\rho : \mathrm{Sym}_n \to \mathrm{GL}(\mathbb{M}^\mu)$$

$$(\rho(\pi)\; f)(\mathbf{x}) \;=\; f(\pi^{-1}\mathbf{x}), \qquad \forall\, \pi \in \mathrm{Sym}_n, \quad \forall\, f \in \mathbb{M}^\mu. \tag{3}$$

Invariant inner product Next we mention an inner product $\langle \cdot, \cdot \rangle$ on the space $\mathbb{M}^\mu \times \mathbb{M}^\mu$ which will be invariant under the representation ρ. The inner product is defined as follows:

$$\langle \cdot, \cdot \rangle : \mathbb{M}^\mu \times \mathbb{M}^\mu \to \mathbb{R}_{\geqslant 0}$$

$$\langle f, g\rangle \;=\; \mathbb{E}_{\mathbf{x}\sim\mathcal{S}_\mu^n}[f(\mathbf{x})\cdot g(\mathbf{x})] \;=\; \frac{1}{N}\sum_{\mathbf{x}\in\mathcal{S}_\mu^n}[f(\mathbf{x})\cdot g(\mathbf{x})]. \tag{4}$$

It is not hard to see that the above inner product is invariant under the representation ρ, i.e., for every $f, g \in \mathbb{M}^\mu$, the following holds:

$$\langle \rho(\pi)f,\ \rho(\pi)g\rangle \;=\; \langle f,\ g\rangle.$$

This representation is quite well-studied in the representation theory for finite groups. There is a complete understanding of the decomposition of $\mathbb{M}^\lambda$ into its irreducible representations. In particular, the irreducible representations of $(\rho, \mathbb{M}^\mu)$ are given by the *Specht modules* $\mathbb{S}^\lambda \subset \mathbb{M}^\lambda$. See [Sag13, Chapter 2] for an excellent exposition on the irreducible decompositions. Next, we state Theorem 3.3, which we will use to prove Theorem 1.3. We do not require specific details of the irreducible representation, so we only state what is sufficient for our purpose.

Theorem 3.3. *Fix any $n, s \in \mathbb{N}$ where n is divisible by s and let $\mu = (n/s, \dots, n/s) \in \mathcal{P}(n)$. The following holds:*

1. *The subrepresentations of $\mathbb{M}^\mu$ are indexed by $\lambda \in \mathcal{P}(n)$ and in particular, there exists subrepresentations $V_{\lambda,1}, \dots, V_{\lambda,m_\lambda}$ for an integer $m_\lambda \in \mathbb{N}$ such that:*

$$\mathbb{M}^\mu \;\cong\; \bigoplus_{\lambda\trianglerighteq\mu}\ \bigoplus_{j=1}^{m_\lambda} V_{\lambda,j},$$

   where $V_{\lambda,1} \cong \cdots \cong V_{\lambda,m_\lambda}$.

   (a) *For $\lambda = (n)$, $m_\lambda = 1$ and $\dim(V_{\lambda,1}) = 1$. This corresponds to the trivial subrepresentation spanned by the function that takes the value 1 at each point of $\mathcal{S}_\mu^n$.*

2. *Let $c \in \mathbb{N}$ denote an absolute constant > 1. For every partition $\lambda = (\lambda_1, \dots, \lambda_\ell) \in \mathcal{P}(n)$, we have*

   (a) *If $\lambda_2 > c$, then $\dim(V_{\lambda,1}) = \cdots = \dim(V_{\lambda,m_\lambda}) \geqslant n^{0.8c}$.*

   (b) *If $\lambda_2 \leqslant c$, then $m_\lambda \leqslant s^{cs}$. As s and c are constants, $m_\lambda = \mathcal{O}_{s,c}(1)$.*

3. *For every constant $c \in \mathbb{N}$, the following holds. Fix any $\lambda = (\lambda_1, \dots, \lambda_\ell) \in \mathcal{P}(n)$ such that $\lambda_2 \leqslant c$. Then there exists vectors $u_1^\lambda, \dots, u_{m_\lambda}^\lambda \in \mathbb{M}^\mu$ where for every $j \in [m_\lambda]$, the vector $u_j^\lambda \in V_{\lambda,j}$, satisfying the following conditions:*

   (a) *For every $j > 1$, u_j^λ is the image of u_1^λ under the unique isomorphism between representations $V_{\lambda,1}$ and $V_{\lambda,j}$.*

   (b) *For every $j \in [m_\lambda]$, $\|u_j^\lambda\|_2 = \Theta_{s,c}(1)$. Here, the norm is with respect to the invariant inner product stated in Equation (4).*

   (c) *If $\mathcal{D}$ is a probability distribution on the balanced multislice $\mathcal{S}_\mu^n$ such that $\mathcal{D}$ is an ε-almost k-wise independent and uniform distribution for some $k \geqslant cs$, then,*

$$\left|\mathbb{E}_{\mathbf{x}\sim\mathcal{D}}[u_j^\lambda(\mathbf{x})]\right| \;\leqslant\; \mathcal{O}_{s,c}(\varepsilon), \qquad\qquad \text{for all}\ \ j \in [m_\lambda].$$

(d) *The volume of the parallelepiped (see Definition 2.8) formed by* $\left\{u_j^\lambda\right\}_{j=1}^{m_\lambda}$ *is at least a constant, i.e.,* $\mathrm{Vol}(u_1^\lambda, \ldots, u_{m_\lambda}^\lambda) = \Omega_{s,c}(1)$.

Remark 3.4. *We note that items* 1 *and* 2 *in the Theorem 3.3 are standard results in the representation theory of* Sym_n*, or simple consequences thereof. Also see Theorem 3.13 for more details on item* 1*. The new technical observations we make are in proving items 3.(c) and 3.(d) of Theorem 3.3. As we will elaborate later in Section 3.3, our proof for item* 3 *of Theorem 3.3 uses [DFLLV21]. In this work, we analyze a set of functions described already in [DFLLV21] and show that they satisfy additional properties, which allows us to prove our main theorem (Theorem 1.3).*

We will first show how Theorem 3.3 implies Theorem 1.3. We defer the proof of Theorem 3.3 to Section 3.4. We will also require the following lemma on estimating the singular values of small matrices. It says that if we have a set of linearly independent vectors whose parallelepiped has a significant volume, then they are "useful" in estimating the singular values.

Lemma 3.5 (Estimating singular values using special vectors)**.** *As a special case, in this lemma, we will work with the standard inner product in Euclidean space* $\mathbb{R}^r$*. The lengths, volumes etc. below are defined using this standard inner product.*

Let $Q \in \mathbb{R}^{r\times r}$ *be a matrix and* $\{v_1, \ldots, v_r\}$ *be a set of linearly independent vectors satisfying the following conditions:*

1. *For each* $j \in [r]$, $\|v_j\|_2 \leqslant m$ *for some* $m \in \mathbb{R}_{>0}$.
2. *For each* $j \in [r]$, $\|Qv_j\|_2 \leqslant q$ *for some* $q \in \mathbb{R}_{>0}$.
3. *The volume (recall Definition 2.8)* $\mathrm{Vol}(v_1, \ldots, v_r) \geqslant \tau$.

Then[6] $\|Q\|_2 \leqslant r \cdot \dfrac{\max\{m^r, 1\} \cdot r!}{\tau} \cdot q$.

Proof of Lemma 3.5. By definition of spectral norm,

$$\|Q\|_2 = \sup_{\mathbf{x}:\|\mathbf{x}\|_2=1} \|Q\mathbf{x}\|_2, \qquad \text{where } \mathbf{x} \in \mathrm{span}(v_1, \ldots, v_r).$$

Choose an arbitrary $\mathbf{x} \in \mathrm{span}(v_1, \ldots, v_r)$ with $\|\mathbf{x}\|_2 = 1$. We know there exists coefficients $\alpha_1, \ldots, \alpha_r \in \mathbb{C}$ such that $\mathbf{x} = \sum_{i=1}^r \alpha_i v_i$. We first upper bound $|\alpha_i|$ for all $i \in [r]$.

Let $\Lambda \in \mathbb{R}^{r\times r}$ denote the matrix whose columns are $v_1, \ldots, v_r$. Let $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_r)$. Then $\Lambda\boldsymbol{\alpha} = \mathbf{x}$. In other words,

$$\alpha_i = \frac{\det(\Lambda_i)}{\det(\Lambda)},$$

where Λ_i is the matrix whose i^{th} column v_i is replaced by $\mathbf{x}$. We have $\|\mathbf{x}\|_\infty \leqslant \|\mathbf{x}\|_2 = 1$ and for every $j \in [r]$, $\|v_j\|_\infty \leqslant \|v_j\|_2 \leqslant m$. Combining these two, we have $\|\Lambda_i\|_\infty \leqslant \max\{m, 1\}$. This implies

[6] Recall the notation in Equation (2).

that $\det(\Lambda_i) \leqslant r! \cdot \|\Lambda_i\|_\infty^r \leqslant \max\{m^r, 1\} \cdot r!$. Recall also that the volume of the parallelepiped spanned by $v_1, \ldots, v_r$ is given by $|\det(\Lambda)|$. Thus we have,

$$|\alpha_i| \leqslant \frac{\max\{m^r, 1\} \cdot r!}{\tau}, \qquad \text{for all } i \in [r].$$

Now let us consider $\|Q\mathbf{x}\|_2$:

$$\|Q\mathbf{x}\|_2 \leqslant \sum_{i=1}^{r} |\alpha_i| \cdot \|Qv_i\|_2 \leqslant r \cdot \frac{\max\{m^r, 1\} \cdot r!}{\tau} \cdot q.$$

This finishes the proof of Lemma 3.5. ■

3.2 Proof of Our Main Technical Theorem (Theorem 1.3)

In this subsection, we give the proof of our main technical theorem (Theorem 1.3), assuming Theorem 3.3. We defer the proof of Theorem 3.3 to Section 3.4.

Proof of Theorem 1.3 (using Theorem 3.3). The first condition on the matrix M implies that M commutes with the representation $(\rho, \mathbb{M}^\mu)$ (see Equation (3)). Using the third item of Proposition 3.1 and Corollary 3.2, we know that the singular values of M can be divided into groups indexed by partitions $\lambda \in \mathcal{P}(n)$. We can classify the singular values of M into three categories:

- Singular values corresponding to the partition $\lambda = (n)$. As stated in 1.(a) of Theorem 3.3, it corresponds to the 1-dimensional vector space and thus has singular value 1.
- Singular values indexed by partitions $\lambda = (\lambda_1, \ldots, \lambda_\ell) \in \mathcal{P}(n)$ with $\lambda_2 > c > 1$ (here c is the constant from Theorem 3.3).
- Singular values indexed by partitions $\lambda = (\lambda_1, \ldots, \lambda_\ell) \in \mathcal{P}(n)$ with $\lambda_2 \leqslant c$ and $\lambda \neq (n)$ (here c is the constant from Theorem 3.3).

To bound $\sigma_2(M)$, we only need to upper bound the singular values in the second and third categories. Before proceeding, we set some notation for convenience. Applying the third item of Proposition 3.1 on the matrix M, we know the following: For every partition $\lambda \in \mathcal{P}(n)$ with $\lambda \trianglerighteq \mu$, there exists a square matrix $\widetilde{M}_\lambda$ of dimensions $m_\lambda \times m_\lambda$ such that

$$M = \bigoplus_{\lambda \trianglerighteq \mu} \left(\widetilde{M}_\lambda \otimes \mathrm{Id}_{\dim(V_\lambda)} \right), \qquad \text{where } \dim(V_\lambda) = m_\lambda \cdot \dim(V_{\lambda,1}).$$

Corollary 3.2 tells us that the multiset of singular values of M is essentially governed by the multiset of singular values of $\widetilde{M}_\lambda$'s for different λ's. For every $\lambda \trianglerighteq \mu$, let $\{\beta_1^\lambda, \ldots, \beta_{m_\lambda}^\lambda\}$ denote the multiset of singular values of $\widetilde{M}_\lambda$ (i.e. we account for repetitions too).

To upper bound the singular values of the second category, we use the upper bound on the **Frobenius norm** of M. More precisely, we will show the following lemma.

Lemma 3.6. *Let $\kappa, c_1 > 0$ be constants such that the Frobenius norm $\|M\|_F \leqslant c_1 \cdot n^{\kappa}$ (see Theorem 1.3) and let $c = 4\kappa$. Let $\lambda = (\lambda_1, \ldots, \lambda_\ell) \in \mathcal{P}(n)$ with $\lambda_2 > c$. Then[a],*

$$\|\widetilde{M}_\lambda\|_2 \;\leqslant\; \frac{1}{c_1' \cdot n^{\kappa}},$$

where c_1' is a constant depending on c_1 and κ.

[a] Recall the notation in Equation (2).

Proof of Lemma 3.6. We know that the square of the Frobenius norm equals the sum of singular values squared, i.e. if $\left\{\beta_1^\gamma, \ldots, \beta_{m_\gamma}^\gamma\right\}$ is the multiset of singular values of $\widetilde{M}_\gamma$, then

$$\sum_{\gamma \trianglerighteq \mu} \dim(V_{\gamma,1}) \cdot \sum_{i=1}^{m_\gamma} |\beta_i^\gamma|^2 \;\leqslant\; \|M\|_F^2.$$

Fix an arbitrary $\lambda \in \mathcal{P}(n)$ with $\lambda_2 > c$. As every term on the left side of the above inequality is a non-negative number, we have the following inequality for any $i \in [m_\lambda]$:

$$\dim(V_{\lambda,1}) \cdot |\beta_i^\lambda|^2 \;\leqslant\; \|M\|_F^2$$

$$\Rightarrow \quad \Omega_\kappa(n^{4\kappa}) \cdot |\beta_i^\lambda|^2 \;\leqslant\; c_1^2 \cdot n^{2\kappa} \qquad \text{(Using 2.(a) of Theorem 3.3 and 2 of Theorem 1.3)}$$

$$\Rightarrow |\beta_i^\lambda| \;\leqslant\; \frac{1}{c_1 \cdot n^{\kappa}}.$$

Since the above upper bound holds for every $i \in [m_\lambda]$, we get the desired bound on $\|\widetilde{M}_\lambda\|_2$. This finishes the proof of Lemma 3.6. ■

Next, we have to upper bound the singular values of the third category, and this requires more steps in comparison to the previous lemma. We use the **ε-almost k-wise independence** of M in this step. We start by stating the bound.

Lemma 3.7. *Let $c = 4\kappa$ (the same constant from Lemma 3.6) and $\lambda \in \mathcal{P}(n)$ be a partition with $\lambda \trianglerighteq \mu$, $\lambda_2 \leqslant c$, and $\lambda \neq (n)$. Then,*

$$\|\widetilde{M}_\lambda\|_2 \;\leqslant\; \mathcal{O}_{s,\kappa}(\varepsilon),$$

where ε is the distance parameter in the third item of Theorem 1.3.

Proof of Lemma 3.7. Fix a partition $\lambda \trianglerighteq \mu$ with $\lambda_2 \leqslant c$ and $\lambda \neq (1, 1, \ldots, 1)$ for rest of the proof. Let $u_1^\lambda, \ldots, u_{m_\lambda}^\lambda$ be the vectors guaranteed from the third item of Theorem 3.3. The idea is to use Lemma 3.5 on the vectors u_j^λ's and the matrix $\widetilde{M}_\lambda$, but we need to be careful, as we explain below.

From the third and fourth items of Proposition 3.1, we have,

$$Mu_j^\lambda \;=\; \widetilde{M}_\lambda u_j^\lambda, \qquad\qquad \text{where } \widetilde{M}_\lambda \text{ is the operator } M|_{Y_\lambda}$$

$$\Rightarrow\; \|Mu_j^\lambda\|_2 \;=\; \|\widetilde{M}_\lambda u_j^\lambda\|_2,$$

where both the norms are with respect to the invariant inner product defined in Equation (4).

Fixing an orthonormal basis. Let

$$Y_\lambda := \mathrm{span}(u_1^\lambda, \ldots, u_{m_\lambda}^\lambda)$$

and let $(w_1, \ldots, w_{m_\lambda})$ be an ordered *orthonormal* (with respect to the *invariant* inner product defined in Equation (4)) basis for the space Y_λ.
Let $\widetilde{A}_\lambda$ denote the $m_\lambda \times m_\lambda$ matrix representing the operator $\widetilde{M}_\lambda$ under the orthonormal basis $(w_1, \ldots, w_{m_\lambda})$.

The singular values remain invariant under the choice of basis[7], thus it is enough to bound $\|\widetilde{A}_\lambda\|_2$. The idea is to use Lemma 3.5 on $\widetilde{A}_\lambda$ and vectors u_j^λ's to bound $\|\widetilde{A}_\lambda\|_2$. There is some subtelty regarding norms in using Lemma 3.5, which one needs to be careful about.

Expressing the u_j^λ's in the orthonormal basis. For every $j \in [m_\lambda]$, let

$$u_j^\lambda = \alpha_{j,1} w_1 + \ldots + \alpha_{j,m_\lambda} w_{m_\lambda} \quad \text{and} \quad \boldsymbol{\alpha}_j := (\alpha_{j,1}, \ldots, \alpha_{j,m_\lambda}).$$

Then,

$$\|u_j^\lambda\|_2 \;=\; \|\boldsymbol{\alpha}_j\|_2,$$

where the left norm is with respect to the *invariant* inner product defined in Equation (4) and the right norm is the *standard* inner product on $\mathbb{R}^{m_\lambda}$. Using 3.(b) of Theorem 3.3, we get that for every $j \in [m_\lambda]$, $\|\boldsymbol{\alpha}_j\|_2 = \Theta_{s,c}(1)$.

Norm after applying the operator $\widetilde{M}_\lambda$. Now we have the following equality:

$$\widetilde{M}_\lambda u_j^\lambda \;=\; \widetilde{A}_\lambda \boldsymbol{\alpha}_j \quad \Rightarrow \quad \|\widetilde{M}_\lambda u_j^\lambda\|_2 \;=\; \|\widetilde{A}_\lambda \boldsymbol{\alpha}_j\|_2,$$

where the left norm is with respect to the invariant inner product defined in Equation (4) and the right norm is with respect to the standard inner product. Hence, we get

$$\|Mu_j^\lambda\|_2 \;=\; \|\widetilde{A}_\lambda \boldsymbol{\alpha}_j\|_2, \qquad\qquad \text{for every } \; j \in [m_\lambda].$$

[7] To see this quickly, note that singular values of a matrix A are the positive square roots of the eigenvalues of AA^T and eigenvalues are independent of the choice of basis.

Upper bounding the norm after applying the operator We now show that for every $j \in [m_\lambda]$,

$$\|\widetilde{A}_\lambda \boldsymbol{\alpha}_j\|_2 \leqslant \mathcal{O}_{s,\kappa}(\varepsilon),$$

where the norm is with respect to the standard inner product. From the previous paragraph, it is enough to show that for every $j \in [m_\lambda]$, the norm $\|Mu_j^\lambda\|_2 \leqslant \mathcal{O}_{s,c}(\varepsilon)$, where the norm is with respect to the invariant inner product.

Using the definition of the invariant inner product from Equation (4), we get,

$$\|Mu_j^\lambda\|_2^2 \;=\; \mathbb{E}_{\mathbf{x}\sim\mathcal{S}_\mu^n}\left(\mathbb{E}_{\mathbf{y}\sim M(\mathbf{x})}[u_j^\lambda(\mathbf{y})]\right)^2$$

For every $\mathbf{x} \in \mathcal{S}_\mu^n$, the third item of Theorem 1.3 says that $M(\mathbf{x})$ is ε-almost k-wise independent for $k = 10s\kappa \geqslant cs$. Applying item 3.(c) of Theorem 3.3 on $M(\mathbf{x})$ for an arbitrary $\mathbf{x} \in \mathcal{S}_\mu^n$, we get

$$\mathbb{E}_{\mathbf{y}\sim M(\mathbf{x})}[u_j^\lambda(\mathbf{y})] \;=\; \mathcal{O}_{s,\kappa}(\varepsilon).$$

As this holds for every $\mathbf{x} \in \mathcal{S}_\mu^n$, we get,

$$\mathbb{E}_{\mathbf{x}\sim\mathcal{S}_\mu^n}\left(\mathbb{E}_{\mathbf{y}\sim M(\mathbf{x})}[u_j^\lambda(\mathbf{y})]\right)^2 \;=\; \mathbb{E}_{\mathbf{x}\sim\mathcal{S}_\mu^n}\left(\mathcal{O}_{s,c}(\varepsilon^2)\right) \;=\; \mathcal{O}_{s,\kappa}(\varepsilon^2).$$

Hence we have shown that for every $j \in [m_\lambda]$, we get

$$\|\widetilde{A}_\lambda \boldsymbol{\alpha}_j\|_2 \;=\; \|Mu_j^\lambda\|_2 \;=\; \mathcal{O}_{s,\kappa}(\varepsilon).$$

Volume of the parallelepiped. Fix an arbitrary order on the u_j's and recall from Definition 2.8 that

$$\mathrm{Vol}(u_1^\lambda, \ldots, u_{m_\lambda}^\lambda) = \|\tilde{u}_1^\lambda\|_2 \cdots \|\tilde{u}_{m_\lambda}^\lambda\|_2,$$

where $\tilde{u}_j$ is as defined in Definition 2.8 and the above norms are with respect to the invariant inner product defined in Equation (4). Similarly, we have $\mathrm{Vol}(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{m_\lambda})$, in which the norm is with respect to the standard inner product.

Observe that for every $j \in [m_\lambda]$, $\mathrm{span}(\tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_{j-1}) \cong \mathrm{span}(\tilde{u}_1^\lambda, \ldots, \tilde{u}_{j-1}^\lambda)$, i.e. they are isometric as inner product spaces. Now the component of u_j^λ orthogonal to the $(j-1)$ dimensional subspace has the same norm (in the invariant inner product) as the component of $\boldsymbol{\alpha}_j$ has norm under the standard inner product. Thus,

$$\mathrm{Vol}(u_1^\lambda, \ldots, u_{m_\lambda}^\lambda) \;=\; \mathrm{Vol}(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{m_\lambda}) = \Omega_{s,\kappa}(1),$$

where the final lower bound is from item 3.(d) of Theorem 3.3.

Now we apply Lemma 3.5 on $\widetilde{A}_\lambda$ and vectors $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{m_\lambda}$. Using 2.(b) of Theorem 3.3, we know that $m_\lambda = \mathcal{O}_{s,c}(1)$. This gives us the desired bound and finishes the proof of Lemma 3.7. ■

Thus, we have proved Lemma 3.6 and Lemma 3.7, which gives an upper bound on the singular values of the second and third categories, respectively. Combining Lemma 3.6 and Lemma 3.7, we finish the proof of Theorem 1.3. ■

3.3 Construction of Special Vectors

In this section, we describe the vectors specified in the third item of Theorem 3.3. The construction combines standard literature on the representation theory of the symmetric group [Sag13] with the recent work of [DFLLV21]. We need to recall the definition here to show that they satisfy the properties claimed in Theorem 3.3.

Throughout this section, fix a partition $\lambda = (\lambda_1, \dots, \lambda_\ell) \in \mathcal{P}(n)$ and assume that $2 \leqslant \ell \leqslant s$. We will consider (Young) tableaux T of shape λ, which contain cells $T[i,j]$ where $1 \leqslant i \leqslant \ell$ and for $1 \leqslant j \leqslant \lambda_i$. Further, we will also consider permutations of such tableaux by permutations that rearrange the elements in each column of T. Let $C_\lambda := \mathrm{Sym}_{\lambda_1^*} \times \cdots \times \mathrm{Sym}_{\lambda_{\lambda_1}^*}$ and given a permutation $\sigma \in C_\lambda$, we denote by T^σ the tableau obtained by rearranging the contents of the cells of T according to σ. For every $\sigma = (\sigma^{(1)}, \dots, \sigma^{(\lambda_1)}) \in C_\lambda$, $\mathrm{sgn}(\sigma) := \mathrm{sgn}(\sigma^{(1)}) \cdots \mathrm{sgn}(\sigma^{(\lambda_1)})$.

We define T_0 to be the canonical tableau of shape λ where the cells are labelled as follows:

$$T_0[i,j] \;:=\; \sum_{p<i} \lambda_p + j, \qquad i \in [\ell],\ j \in [\lambda_i]. \tag{5}$$

The following is a diagram for the canonical tableau T_0 for some partition $\lambda \in \mathcal{P}(n)$.

$T_0 \;=$

1	2	$\dots$	$\lambda_1 - 1$	λ_1
$\lambda_1 + 1$	$\lambda_1 + 2$	$\dots$	λ_2	
$\vdots$	$\vdots$			
$n-2$	$n-1$			
n				

Next, we define a polynomial on $\mathcal{S}_\mu^n$ for every semi-standard Young tableau of shape λ and content[8] μ. Recall that $\mu \in \mathcal{P}(n)$ is the partition of n given by $(\frac{n}{s}, \dots, \frac{n}{s})$.

Definition 3.8. *[DFLLV21, Section 5.2]. Given a tableau T' of shape λ with distinct labels from $[n]$ and another tableau T of shape λ with content μ, we define a corresponding $\mathbb{R}$-valued function $e_{T',T} : \mathcal{S}_\mu^n \to \{0,1\}$ by*

$$e_{T',T}(\boldsymbol{x}) = \begin{cases} 1, & \text{if } \{x_{T'[i,1]}, \dots, x_{T'[i,\lambda_i]}\} = \{T[i,1], \dots, T[i,\lambda_i]\} \text{ as multisets for each } 1 \leqslant i \leqslant \ell, \text{ (*)} \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

[8] i.e. a tableau with μ_1 many 0s, μ_2 many 1s, and so on until μ_s many $(s-1)$s

Finally, given a $T \in \mathrm{SSYT}(\lambda, \mu)$*, define the function* $\chi_T : \mathcal{S}_\mu^n \to \mathbb{Z}$ *by*

$$\chi_T(\boldsymbol{x}) = \sum_{\sigma \in C_\lambda} \mathrm{sgn}(\sigma) \cdot e_{T_0^\sigma, T}(\boldsymbol{x}), \tag{7}$$

where T_0^σ *is the tableau obtained after* σ *acts on the canonical tableau* T_0*.*

Observation 3.9. *Note that condition (*) in Equation (6) could equivalently have been stated in terms of rows* $i \in \{2, \ldots, \ell\}$ *as the condition for* $i = 1$ *is implied by the others (since the input* $\boldsymbol{x}$ *is a point in* $\mathcal{S}_\mu^n$*). Overall, this implies that* $e_{T_0^\sigma, T}$ *(and hence* χ_T*) depends only on variables whose index appears in one of the first* λ_2 *columns of* T*. In particular, if* $\lambda_2 \leqslant c$*, this implies that* χ_T *is a* γ*-junta for* $\gamma \leqslant c\ell \leqslant cs$*.*

Next we define a total order on the set $\mathrm{SSYT}(\lambda, \mu)$.

Definition 3.10 (Total order on SSYTs)**.** *Let* $\lambda \in \mathcal{P}(n)$ *and* $\mu = (n/s, \ldots, n/s)$*. Given two distinct SSYTs* $S, T \in \mathrm{SSYT}(\lambda, \mu)$*, we say that* $S < T$ *if there exists* $2 \leqslant i \leqslant \ell$ *and* $j \in [\lambda_i]$ *such that the following holds:*

1. *For every* $k > i$ *and for every* $j' \in [\lambda_k]$*, we have* $S[k, j'] = T[k, j']$*, i.e. the* k^{th} *rows of* S *and* T *are equal.*
2. *For every* $\lambda_i \geqslant j' > j$ *such that* $S[i, j'] = T[i, j']$*.*
3. *Finally,* $S[i, j] < T[i, j]$*.*

We leave it to the reader to check that this defines a total order on $\mathrm{SSYT}(\lambda, \mu)$*.*

We will need the following claim regarding the aforementioned ordering.

Claim 3.11. *Assume that* $S, T \in \mathrm{SSYT}(\lambda, \mu)$ *and* $\sigma \in C_\lambda$ *are such that*

- *either* $S < T$
- *or* $S = T$ *and* σ *is not the identity permutation.*

Then, for any $\sigma \in C_\lambda$*, there exists an* $i \in \{2, \ldots, \ell\}$ *such that the multisets* $\{S^\sigma[i, 1], \ldots, S^\sigma[i, \lambda_i]\}$ *and* $\{T[i, 1], \ldots, T[i, \lambda_i]\}$ *are distinct.*

Proof. Choose i to be the largest number such that σ moves the contents of some cell in the ith row of S, assuming σ is not the identity; otherwise, set $i = 0$. Assuming that $i \neq 0$, for each cell in the i^{th} row moved by σ, we note that the contents of this row can only decrease, since the columns of S are strictly increasing and σ does not change the contents of any row $i' > i$. In particular, this implies the claim in the case that $S = T$. We therefore assume that $S \neq T$ and $S < T$ for the rest of the proof.

Let i_0 be the largest number such that the i_0^{th} rows of S and T differ. Note that $i_0 \in \{2, \ldots, \ell\}$. Further, fix j_0 to be the rightmost cell on this row where S and T differ. Note that $S[i_0, j_0] < T[i_0, j_0]$.

We note that we are immediately done if $i < i_0$ since in this case

$$\{S[i_0, 1], \ldots, S[i_0, \lambda_{i_0}]\} = \{S^\sigma[i_0, 1], \ldots, S^\sigma[i_0, \lambda_{i_0}]\} \neq \{T[i_0, 1], \ldots, T[i_0, \lambda_{i_0}]\}$$

So we may assume that $i \geqslant i_0$, and in particular that σ is not the identity.

Now, we consider two cases.

- If $i > i_0$, then we have

$$\sum_{j \in [\lambda_i]} S^\sigma[i, j] < \sum_{j \in [\lambda_i]} S[i, j] = \sum_{j \in [\lambda_i]} T[i, j]$$

  implying the claim in this case.

- If $i = i_0$, consider the rightmost cell (numbered j, say) where S^σ and T differ on this row. Note that $S^\sigma[i, j_0] \leqslant S[i, j_0] < T[i, j_0]$ and hence $j \geqslant j_0$.

  Consider the multiplicities of the element $t := T[i, j]$ in the ith rows of S, S^σ and T, which we denote m_S, m_{S^σ} and m_T respectively. Note that $m_S \leqslant m_T$ because $j \geqslant j_0$ and $S < T$. We also know that $S^\sigma[i, j] \neq t$ by definition of j. Finally, note that for any $j' < j$ we have either $S^\sigma[i, j'] = S[i, j']$ or $S^\sigma[i, j'] < S[i, j'] \leqslant S[i, j] \leqslant T[i, j]$ with the latter two inequalities following from the fact that S is an SSYT and the fact that $j \geqslant j_0$. This implies that $m_{S^\sigma} < m_T$, the multisets defined by the ith row in the two tableaux S^σ and T cannot be equal.

This finishes the proof of the claim. ■

The main result of this subsection is the following lemma, which shows the existence of the special vectors as stated in Theorem 3.3.

Lemma 3.12. *Let $\lambda \in \mathcal{P}(n)$ and $c \in \mathbb{N}$. Assume that $\lambda_2 \leqslant c$. For every $T \in \mathrm{SSYT}(\lambda, \mu)$, the function χ_T satisfy the following properties:*

1. *For each $T \in \mathrm{SSYT}(\lambda, \mu)$, we have $\|\chi_T\|_2 = \mathcal{O}_{s,c}(1)$.*
2. *Let $\varepsilon > 0$ be arbitrary and assume k is an integer such that $k \geqslant cs$. For any ε-almost k-wise independent distribution $\mathcal{D}$ supported on $\mathcal{S}_\mu^n$, we have*

$$|\mathbb{E}_{\mathbf{x} \sim \mathcal{D}}[\chi_T(\mathbf{x})]| \leqslant \mathcal{O}_{s,c}(\varepsilon). \tag{8}$$

3. *We have,*

$$\mathrm{Vol}(\{\chi_T \mid T \in \mathrm{SSYT}(\lambda, \mu)\}) \;=\; \Omega_{s,c}(1).$$

Proof. The first item follows almost immediately from the definition of χ_T in Equation (7) above. From this definition, we get

$$\|\chi_T\|_2 \leqslant \|\chi_T\|_\infty = \max_{\mathbf{x} \in \mathcal{S}_\mu^n} |\chi_T(\boldsymbol{x})| \leqslant |C_\lambda| \cdot \max_{\sigma, \mathbf{x}} |e_{T_0^\sigma, T}(\mathbf{x})| \leqslant |C_\lambda| \leqslant (s!)^c \tag{9}$$

where the first inequality is trivial, the second is the triangle inequality applied to Equation (7), the third follows from the fact $|e_{T_0^\sigma,T}(\boldsymbol{x})| \leqslant 1$ for each $\mathbf{x}$, and the last follows from the fact that λ is c-good.

For the second item, we note that by Observation 3.9 and the ε-almost k-wise independence of $\mathcal{D}$, we have

$$\mathbb{E}_{\mathbf{x}\sim\mathcal{D}}[\chi_T(\mathbf{x})] \quad \leqslant \quad \varepsilon \cdot \max_{\mathbf{x}\in\mathcal{S}^n} |\chi_T(\mathbf{x})| \; + \; \mathbb{E}_{\mathbf{a}\sim\mathcal{S}_\mu^n}[\chi_T(\mathbf{a})]. \leqslant \mathcal{O}_{s,c}(\varepsilon) + \mathbb{E}_{\mathbf{x}\sim\mathcal{S}_\mu^n}[\chi_T(\mathbf{x})]$$

where the second inequality uses the bound on $|\chi_T(\mathbf{x})|$ proved above. To bound the latter term, we note that for by symmetry, for each $\sigma \in C_\lambda$, the quantity $\mathbb{E}_{\mathbf{x}\sim\mathcal{S}_\mu^n}[e_{T_0^\sigma,T}(\mathbf{x})]$ is exactly the same. Since the signed sum defining χ_T has the same number of positive and negative signs, the sum of the expectations is 0. This proves the second item of the claim.

The third item needs a definition. Given a $S \in \mathrm{SSYT}(\lambda,\mu)$, define subset $A_S \subset \mathcal{S}_\mu^n$ as follows:

$$A_S \;:=\; \left\{\mathbf{x} \in \mathcal{S}_\mu^n \;\middle|\; x_{T_0[i,j]} = S[i,j],\; i \in [\ell],\; j \leqslant \min\{\lambda_2, \lambda_i\}\right\},$$

i.e. we define A_S using the first λ_2 columns of S.

Note that for each $\mathbf{x} \in A_S$, we have the following:

- $e_{T_0,S}(\mathbf{x}) = 1$. This follows immediately from the definition of $e_{T_0,S}$ above.
- Now fix $T \in \mathrm{SSTY}(\lambda,\mu)$ and $\sigma \in C_\lambda$ such that either $T > S$ or $S = T$ and σ is not the identity permutation (here the identity permutation in C_λ refers to $\mathrm{id} \times \cdots \times \mathrm{id}$). We claim that $e_{T_0^\sigma,T}(\mathbf{x}) = 0$. To see this, start by labelling each cell of T_0 with the value of the corresponding variable, which leads to a tableau S' which agrees with S on all cells in the first λ_2 columns. Since $e_{T_0^\sigma,T}$ depends only the variables in these columns, we may change $\mathbf{x}$ in the other coordinates to ensure that $S' = S$.

  Now, we observe that for any $\sigma \in C_\lambda$, the multiset $\{x_{T_0^\sigma[i,1]}, \ldots, x_{T_0^\sigma[i,\lambda_i]}\}$ is equal to the multiset $\{S^\sigma[i,1], \ldots, S^\sigma[i,\lambda_i]\}$. In particular, by Claim 3.11, there exists an $i \in [\ell]$ so that the multiset $\{x_{T_0^\sigma[i,1]}, \ldots, x_{T_0^\sigma[i,\lambda_i]}\}$ is not equal to $\{T[i,1], \ldots, T[i,\lambda_i]\}$, implying that $e_{T_0^\sigma,T}(\mathbf{x}) = 0$.

The above implies that for each $\mathbf{x} \in A_S$, we have

- $\chi_S(\mathbf{x}) = 1$ and
- $\chi_T(\mathbf{x}) = 0$ for each $T > S$.

For each $S \in \mathrm{SSYT}(\lambda,\mu)$, let $\tilde{\chi}_S$ denote the projection of χ_S to the vector space orthogonal to the span of $\{\chi_T \mid S < T\}$.
To bound $\|\tilde{\chi}_S\|_2$, we recall that $\tilde{\chi}_S = \chi_S - \chi$ for some χ in the span of $\{\chi_T \mid S < T\}$. By the above argument, we know that $\chi(\mathbf{x}) = 0$ and hence that $\tilde{\chi}_S(\boldsymbol{x}) = 1$ for each $\mathbf{x} \in A_S$. Hence, we get

$$\|\tilde{\chi}_S\|_2^2 = \mathbb{E}_{\mathbf{x}\sim\mathcal{S}_\mu^n} \tilde{\chi}_S(\mathbf{x})^2 \geqslant \frac{|A_S|}{N}$$

where $N = |\mathcal{S}_\mu^n|$. So to prove the claim, it suffices to show that the latter quantity is $\Omega_{s,c}(1)$.

For each $i \in \{0, \ldots, s-1\}$, let γ_i denote the number of cells in the first λ_2 columns of S that are i

Define $\gamma := \gamma_0 + \ldots + \gamma_{s-1} \leqslant cs$. Using Stirling's approximation and $\gamma \leqslant cs \leqslant n/2$, we get,

$$|A_S| \;=\; \binom{n-\gamma}{\frac{n}{s}-\gamma_0,\ldots,\frac{n}{s}-\gamma_{s-1}} \;\geqslant\; \frac{(n-\gamma)^{n-\gamma}\cdot s^{n-\gamma}}{(n-s)^{n-\gamma}} \cdot \Omega\left(\frac{\sqrt{\pi n}}{(2\pi(\frac{n}{s}))^{s/2}}\right)$$

Again using Stirling's approximation for $N = \binom{n}{n/s,\ldots,n/s}$, we get,

$$\frac{|A_S|}{|\mathcal{S}^n_\mu|} \;\geqslant\; \frac{(n-\gamma)^{n-\gamma}}{(n-s)^{n-\gamma}} \cdot \frac{1}{s^\gamma} \cdot \Omega\left(\frac{\sqrt{\pi n}}{(2\pi(\frac{n}{s}))^{s/2}} \cdot \frac{(2\pi(\frac{n}{s}))^{s/2}}{\sqrt{2\pi n}}\right)$$

$$\geqslant\; \left(1-\frac{\gamma-s}{n-s}\right)^{n-\gamma} \cdot \frac{1}{s^\gamma} \cdot \Omega(1)$$

$$\geqslant\; \Omega\left(\left(1-\frac{2\gamma}{n}\right)^{n/2\gamma\cdot 2\gamma} \cdot \frac{1}{s^\gamma}\right) \;\geqslant\; \Omega\left(\left(\frac{1}{e^2 s}\right)^\gamma\right) \;=\; \Omega_{s,\gamma}(1).$$

As $\gamma \leqslant cs$, we get that $|A_S|/N = \Omega_{s,c}(1)$.

The volume of the parallelepiped is equal to the product of $\|\tilde{\chi}_T\|'s$. As we showed above, $\tilde{\chi}_T = \Omega_{s,c}(1)$, and thus we get,

$$\mathrm{Vol}(\{\chi_T \mid T \in \mathrm{SSYT}(\lambda,\mu)\}) \;\geqslant\; \prod_{T\in\mathrm{SSYT}(\lambda,\mu)} \|\tilde{\chi}_T\|_2 \;=\; \Omega_{s,c}(1).$$

This finishes the proof of Lemma 3.12. ■

3.4 Putting Everything Together

Now we are ready to combine everything and prove Theorem 3.3. To do so, we will use the following standard result on the representation of Sym_n. The proof can be found in standard texts on representation theory for the symmetric group or [Sag13].

Theorem 3.13 (Young's Rule)**.** *(See for e.g. [Sag13, Corollary 2.11.2]). Fix any $n, s \in \mathbb{N}$ where n is divisible by s and let $\mu = (n/s,\ldots,n/s) \in \mathcal{P}(n)$. For every $\lambda \in \mathcal{P}(n)$ with $\lambda \trianglerighteq \mu$, let $V_{\lambda,j}$ and m_λ be as defined in Theorem 3.3. Then,*

$$\dim(V_{\lambda,1}) = \cdots = \dim(V_{\lambda,m_\lambda}) = f_\lambda \quad \text{and} \quad m_\lambda = K_{\lambda\mu},$$

where f_λ and $K_{\lambda\mu}$ are defined in Section 2.

Next we prove two claims regarding f_λ and m_λ for certain partitions $\lambda \in \mathcal{P}(n)$. These two claims will be used to prove the item 2 of Theorem 3.3.

Claim 3.14 (Lower bound on the algebraic multiplicity of certain eigenvalues)**.** *[EFP11, Lemma 2][9]. Let $c \in \mathbb{N}$ be a constant with $c > 10s$. Then for any partition $\lambda \in \mathcal{P}(n)$ with $\lambda_2 > c$,*

$$f_\lambda > \Omega_c(n^c).$$

[9] There is a minor typo in the statement of Lemma 2 in [EFP11]. It should be "of length **at most**..." instead of "of length greater than..."

Claim 3.15 (Multiplicity for c-good partitions)**.** *Let $c \in \mathbb{N}$ be a constant and $\lambda \in \mathcal{P}(n)$ with $\lambda_2 \leqslant c$ and $\lambda \neq (n)$. Let m_λ be as defined in the statement of Theorem 3.3. Then, $m_\lambda \leqslant s^{sc} = \mathcal{O}_{s,c}(1)$.*

Proof of Claim 3.15. From Theorem 3.13, we know that $m_\lambda = K_{\lambda\mu}$, i.e. the Kostka numbers for shape λ and type μ. We are interested in upper bounding $K_{\lambda\mu}$ for a c-good partition λ. We have $\lambda_2 + \ldots + \lambda_\ell \leqslant cs$. For each cell in the second row till the last row, there are at most s many choices. As there are $\leqslant cs$ such cells, we get that $K_{\lambda\mu} \leqslant s^{cs}$. This finishes the proof of Claim 3.15. ■

Now we are ready to put all the claims and lemmas together to finish the proof of Theorem 3.3.

Proof of Theorem 3.3. The first item follows by combining the first item of Proposition 3.1 and Theorem 3.13. Item 2.(a) follows from Claim 3.14 and item 2.(b) follows from Claim 3.15.
Finally, we show that the vectors χ_T's meet the conditions stated in the third item.

1. For 3.(a), we note that the literature on the representation theory of Sym_n (see [Sag13, Section 2.9 & Section 2.10][10]) identifies for each $\lambda \in \mathcal{P}(n)$ exactly m_λ many linearly independent ways of embedding the irreducible representation $\mathbb{S}^\lambda$ ($\mathbb{S}^\lambda$ is the unique irreducible representation, or Specht module, corresponding to partition λ) into the representation $\mathbb{M}^\mu$. These embeddings are indexed by elements of $\mathrm{SSYT}(\lambda, \mu)$ and denoted by $\Theta_T : \mathbb{S}^\lambda \to \mathbb{M}^\mu$. Given $T \in \mathrm{SSYT}(\lambda, \mu)$, let $V_{\lambda,T}$ denote the image of $\mathbb{S}^\lambda$ under the corresponding embedding Θ_T.

   It can be checked that the various χ_T are the images of the same element $v \in \mathbb{S}^\lambda$ under Θ_T (see also [DFLLV21]). This implies that for $S, T \in \mathrm{SSYT}(\lambda, \mu)$ χ_T is the image of χ_S under the unique isomorphism from $V_{\lambda,S}$ to $V_{\lambda,T}$. This proves 3.(a).

2. Item 1 of Lemma 3.12 shows that they satisfy 3.(b).

3. Item 2 of Lemma 3.12 shows that they satisfy 3.(c).

4. Item 3 of Lemma 3.12 shows that they satisfy 3.(d).

This finishes the proof of Theorem 3.3. ■

3.5 Singular Value Bound for Nearly Balanced Random Walks

We now use the statement of Theorem 1.3 to derive the singular value bounds for nearly balanced random walks on the multislice, as stated in Theorem 1.1.

For this, we will need the following lemma.

Lemma 3.16. *For every $s \geqslant 2$ and $C < \infty$, there exists $\tau > 0$ such that for every finite set $\mathcal{S}$ of size s and sufficiently large $n \in \mathbb{N}$, if a generalized Hamming distance parameter $\Delta \in \mathbb{Z}^{\mathcal{S}\times\mathcal{S}}$ over the multislice $\mathcal{S}_\mu^n$ is C-balanced, we have that $\sigma_2(W_\Delta) \leqslant 1/n^\tau$, where W_Δ is the random walk matrix*

[10] In the literature of representation theory, these isomorphisms are stated in the language of **tabloids** and **polytabloids**. In Appendix A, we provide a translation between the language of tabloids/polytabloids and points/functions.

determined by Δ.

The above statement implies the claimed general result (i.e., Theorem 1.1) for random walk matrices that are not necessarily given by a single generalized Hamming distance parameter, but as long as they are *supported* on balanced generalized Hamming distance parameters.

We first prove Lemma 3.16.

Proof of Lemma 3.16. Here we directly apply our main result Theorem 1.3, bounding the singular values of matrices satisfying certain properties. For this, we show that W_Δ satisfies the three properties needed to apply Theorem 1.3.

- **Permutation invariance:** For every permutation π of $[n]$, W_Δ is unchanged if the rows and columns are changed according to the permutation induced by π (denoted $\pi()$) on the balanced multislice (denoted V in this proof). This is because the value of the entry $W_\Delta(\mathbf{a},\mathbf{b})$ only depends on $\Delta(\mathbf{a},\mathbf{b})$, which doesn't get altered by π, i.e., we have $\Delta(\mathbf{a},\mathbf{b}) = \Delta(\pi(\mathbf{a}),\pi(\mathbf{b}))$.
- **Bounded Frobenius norm:** We will show that $\|W_\Delta\|_F \leqslant n^{O_s(1)}$. Denoting $m := n/s$ and the rows of Δ by $\mathbf{p}(0),\dots,\mathbf{p}(s-1)$, we note that for each $\mathbf{a} \in V$, there are exactly $D := \binom{m}{\mathbf{p}(0)}\cdots\binom{m}{\mathbf{p}(s-1)}$ [11] points $\mathbf{b} \in V$ such that $\Delta(\mathbf{a},\mathbf{b}) = \Delta$. Hence, we have that

$$W_\Delta(\mathbf{a},\mathbf{b}) \;=\; \begin{cases} 1/D, & \text{if } \Delta(\mathbf{a},\mathbf{b}) = P \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, we have

$$\begin{aligned} \|W_\Delta\|_F^2 &= \sum_{\mathbf{a},\mathbf{b}\in V} W_\Delta(\mathbf{a},\mathbf{b})^2 \\ &= |V|D/D^2 \\ &= \binom{sm}{m,\dots,m} \Big/ \left(\binom{m}{\mathbf{p}(0)}\cdots\binom{m}{\mathbf{p}(s-1)}\right) \end{aligned} \tag{10}$$

In order to bound the above quantity, let $\mathbf{q} = (q_0,\dots,q_{s-1}) \in \mathbb{Z}^{\mathbb{Z}_s}$ be such that $\sum_{j\in\mathbb{Z}_s} q_j = m$ and $|q_j - q_{j'}| \leqslant 1$ for all $j,j' \in \mathbb{Z}_s$ (such a $\mathbf{q}$ always exists; indeed each $\lfloor m/s \rfloor \leqslant q_j \leqslant \lceil m/s \rceil$). We will first show that $\frac{p_0!\dots p_{s-1}!}{q_0!\dots q_{s-1}!}$ is upper bounded by $m^{\mathcal{O}_s(1)}$, where $\mathbf{p} := \mathbf{p}(\alpha) = (p_0,\dots,p_{s-1})$ for an arbitrary $\alpha \in \mathbb{Z}_s$.

Claim 3.17. $\frac{p_0!\dots p_{s-1}!}{q_0!\dots q_{s-1}!} \leqslant m^{\mathcal{O}_s(1)}$.

Proof. We consider the following sequence of vectors: $\mathbf{p} = \mathbf{p}^{(0)},\mathbf{p}^{(1)},\dots,\mathbf{p}^{(t)} = \mathbf{q}$ (for some finite t), where two adjacent $\mathbf{p}^{(i-1)}$ and $\mathbf{p}^{(i)}$ differ in exactly two coordinates (say $c_i \neq c_i' \in \mathbb{Z}_s$) such that $p^{(i)}_{c_i} = p^{(i-1)}_{c_i} + 1$ and $p^{(i)}_{c_i'} = p^{(i-1)}_{c_i'} - 1$, for all $i \in [t]$. We note that since each $p_j \in \frac{m}{s} \pm \sqrt{Cm\log m}$ (as P is a *balanced* generalized Hamming distance matrix), such a sequence can be realized with $t \leqslant s\sqrt{Cm\log m}$ by repeatedly picking the smallest and largest elements

[11] Here, for a vector of integers $\mathbf{p} = (p_1,\dots,p_s)$, $\binom{m}{\mathbf{p}}$ denotes $\binom{m}{p_1,\dots,p_s}$.

of $\mathbf{p}$ and adding one to the smallest element and subtracting one from the largest one. This will also ensure that for each intermediate $i \in [t]$, we have the invariant $p_j^{(i)} \in \frac{m}{s} \pm \sqrt{Cm \log m}$ for all $j \in \mathbb{Z}_s$.

Now, we note that for all $i \in [t]$,

$$\frac{\prod_j p_j^{(i-1)}!}{\prod_j p_j^{(i)}!} = \frac{p_{c_i}^{(i-1)}! p_{c_i'}^{(i-1)}!}{p_{c_i}^{(i)}! p_{c_i'}^{(i)}!} = \frac{p_{c_i'}^{(i-1)}}{p_{c_i}^{(i)}} \leqslant 1 + \mathcal{O}\left(\sqrt{\frac{Cs^2 \log m}{m}}\right).$$

Using the above bound for all $i \in [t]$ and multiplying them, we get

$$\frac{p_0! \dots p_{s-1}!}{q_0! \dots q_{s-1}!} \leqslant \left(1 + \mathcal{O}\left(\sqrt{\frac{Cs^2 \log m}{m}}\right)\right)^t \leqslant \left(1 + \mathcal{O}\left(\sqrt{\frac{Cs^2 \log m}{m}}\right)\right)^{s\sqrt{Cm \log m}} \leqslant m^{\mathcal{O}(Cs^2)}, \tag{11}$$

where for the last inequality, we are using the inequality $1 + x \leqslant e^x$.

■

Now, continuing the computation of (10), we have

$$\begin{aligned}
\|W_\Delta\|_F^2 &= \binom{sm}{m, \dots, m} \Big/ \left(\binom{m}{\mathbf{p}(0)} \cdots \binom{m}{\mathbf{p}(s-1)}\right) \\
&\leqslant m^{O(Cs^3)} \cdot \frac{(sm)! q_0!^s \dots q_{s-1}!^s}{m!^{2s}} && \text{(using Equation (11))} \\
&\leqslant m^{\mathcal{O}(Cs^3)} \cdot \frac{(sm)!}{m!^s} \cdot \left(\frac{\lceil m/s \rceil!^s}{m!}\right)^s && \text{(as } q_j \leqslant \lceil m/s \rceil\text{)} \\
&\leqslant m^{\mathcal{O}(Cs^3)} \cdot s^{sm} \cdot \left(\frac{(m/(es))^m}{(m/e)^m}\right)^s && \text{(using Stirling's inequality)} \\
&\leqslant m^{\mathcal{O}(Cs^3)}.
\end{aligned}$$

Hence $\|W_\Delta\|_F \leqslant n^{\mathcal{O}_{s,C}(1)}$.

- **ε-almost k-wise independence:** We will show that for every $k \leqslant \mathcal{O}_s(1)$, W_Δ is ε-almost k-wise independent (see Definition 1.2) for some $\varepsilon = 1/n^{\Omega_s(1)}$. That is, for every $\mathbf{a} \in V$ and $T \in \binom{[n]}{k}$, we will show that

$$\mathrm{SD}(W_\Delta(\mathbf{a})|_T, U_T) \leqslant \varepsilon,$$

where U_T denotes the uniform distribution over the coordinates given by T.

Let $T = T^{(0)} \cup \cdots \cup T^{(s-1)}$ be a partition of T, where $T^{(i)} = \mathbf{a}^{-1}(i) \cap T$ for $i \in \mathbb{Z}_s$. We will fix an arbitrary $\mathbf{b} \in \mathbb{Z}_s^T$ and upper bound the difference $\left|\Pr[W_\Delta(\mathbf{a})|_T = \mathbf{b}] - \frac{1}{s^k}\right|$. For this, let $\mathbf{e} = (e_j)_{j\in\mathbb{Z}_s}$ denote the number of occurrences of $j \in \mathbb{Z}_s$ in $\mathbf{b}$. Furthermore, let $\mathbf{e}^{(i)} = (e_j^{(i)})_{j\in\mathbb{Z}_s}$ where $e_j^{(i)}$ denotes the number of occurrences of $j \in \mathbb{Z}_s$ in $\mathbf{b}$ when restricted to $T^{(i)}$. To make the notation cleaner, for the rest of the proof, we will use the notation $\mathbf{p}^{(i)}$ to mean $\mathbf{p}(i)$. We then have:

$$\Pr[W_\Delta(\mathbf{a})|_T = \mathbf{b}] = \binom{m - |T^{(0)}|}{\mathbf{p}^{(0)} - \mathbf{e}^{(0)}} \cdots \binom{m - |T^{(s-1)}|}{\mathbf{p}^{(s-1)} - \mathbf{e}^{(s-1)}} \Bigg/ \left(\binom{m}{\mathbf{p}^{(0)}} \cdots \binom{m}{\mathbf{p}^{(s-1)}}\right). \tag{12}$$

For each $i \in \mathbb{Z}_s$, we have

$$\begin{aligned}
\frac{\binom{m-|T^{(i)}|}{\mathbf{p}^{(i)}-\mathbf{e}^{(i)}}}{\binom{m}{\mathbf{p}^{(i)}}} &= \frac{(m - e_0^{(i)} - \cdots - e_{s-1}^{(i)})!}{(p_0^{(i)} - e_0^{(i)})! \ldots (p_{s-1}^{(i)} - e_{s-1}^{(i)})!} \cdot \frac{p_0^{(i)}! \ldots p_{s-1}^{(i)}!}{m!} \\
&= \frac{(p_0^{(i)} \ldots (p_0^{(i)} - e_0^{(i)} + 1)) \ldots (p_{s-1}^{(i)} \ldots (p_{s-1}^{(i)} - e_{s-1}^{(i)} + 1))}{m \ldots (m - e_0^i - \ldots e_{s-1}^{(i)} + 1)} \\
&\in \left[\left(\frac{\frac{m}{s} - \sqrt{Cm \log m} - |T^{(i)}|}{m} \right)^{|T^{(i)}|}, \left(\frac{\frac{m}{s} + \sqrt{Cm \log m}}{m - |T^{(i)}|} \right)^{|T^{(i)}|} \right] \\
&\qquad\qquad \text{(as each } p_j^{(i)} \in \tfrac{m}{s} \pm \sqrt{Cm \log m}\text{)} \\
&\subseteq \frac{1}{s^{|T^{(i)}|}} \left[1 \pm \frac{1}{m^{\Omega_{s,C}(1)}} \right]^{|T^{(i)}|}. \qquad \text{(using } |T^{(i)}| \leqslant k \leqslant O_s(1)\text{)}
\end{aligned}$$

Plugging the above bound into (12) and using $\sum_i |T^{(i)}| = |T| = k$ gives that

$$\Pr[W_\Delta(\mathbf{a})|_T = \mathbf{b}] \in \prod_{i\in\mathbb{Z}_s} \frac{1}{s^{|T^{(i)}|}} \left[1 \pm \frac{1}{m^{\Omega_s(1)}} \right]^{|T^{(i)}|} \subseteq \left[\frac{1}{s^k} \pm \frac{k}{m^{\Omega_s(1)}} \right].$$

Therefore, we have

$$\mathrm{SD}(W_\Delta(\mathbf{a})|_T, U_T) = \frac{1}{2} \sum_{\mathbf{b}\in\mathcal{S}^T} \left| \Pr[W_\Delta(\mathbf{a})|_T = \mathbf{b}] - \frac{1}{s^k} \right| \leqslant \frac{s^k \cdot k}{m^{\Omega_{s,C}(1)}} \leqslant \frac{1}{n^{\Omega_{s,C}(1)}}.$$

Thus we have shown that the three conditions needed to apply Theorem 1.3 hold for W_Δ. Therefore, we obtain $\sigma_2(W_\Delta) \leqslant 1/n^{\Omega_{s,C}(1)}$, finishing the proof of Lemma 3.16. ■

We now finish the proof of Theorem 1.1 using Lemma 3.16.

Theorem 1.1 (Singular value bound for nearly balanced walks). *For every $s \geqslant 2$ and $C < \infty$, there exists $\tau > 0$ such that for every finite set $\mathcal{S}$ of size s and sufficiently large $n \in \mathbb{N}$, the following holds:*
If W is a stochastic matrix over the multislice $\mathcal{S}_\mu^n$ that respects symmetries*, and satisfies the condition that*

$$W(\mathbf{a}, \mathbf{b}) > 0 \quad \Rightarrow \quad \Delta(\mathbf{a}, \mathbf{b}) \text{ is } C\text{-balanced} \quad \forall\, \mathbf{a}, \mathbf{b} \in \mathcal{S}_\mu^n,$$

then $\sigma_2(W) \leqslant 1/n^\tau$.

Proof of Theorem 1.1. The idea is to express W as a convex combination of W_Δ for Δ being C-balanced generalized Hamming distance parameters. We first show that for every $\mathbf{a} \in \mathcal{S}_\mu^n$, the $\mathbf{a}$-th row of W can be expressed as a convex combination of the $\mathbf{a}$-th rows of the random walk matrix determined by the individual generalized Hamming distance parameters (i.e, W_Δ). As W respects symmetries, for every $\mathbf{a}, \mathbf{b} \in \mathcal{S}_\mu^n$ and permutation $\pi \in \mathrm{Sym}_n$, we have that $W(\mathbf{a}, \mathbf{b}) = W(\pi(\mathbf{a}), \pi(\mathbf{b}))$. Now we note that if it holds that $\Delta(\mathbf{a}, \mathbf{b}) = \Delta(\mathbf{a}, \mathbf{c})$ for some $\mathbf{a}, \mathbf{b}, \mathbf{c}$, then there exists a permutation $\pi \in \mathrm{Sym}_n$ such that $\pi(\mathbf{a}) = \mathbf{a}$ and $\pi(\mathbf{b}) = \mathbf{c}$ (this can be obtained by permuting the coordinates of $\mathbf{a}$ that take identical values); hence we have $W(\mathbf{a}, \mathbf{b}) = W(\pi(\mathbf{a}), \pi(\mathbf{b})) = W(\mathbf{a}, \mathbf{c})$. Since W has positive entries only at cells corresponding to balanced generalized Hamming distance, we can thus express the $\mathbf{a}$-th row of W as the following convex combination:

$$W(\mathbf{a}) = \sum_{\Delta \text{ is } C\text{-balanced}} \alpha_{\mathbf{a},\Delta} W_\Delta(\mathbf{a}), \tag{13}$$

for some $\alpha_{\mathbf{a},\Delta} \geqslant 0$ such that $\sum_{\Delta \text{ is } C\text{-balanced}} \alpha_{\mathbf{a},\Delta} = 1$. We now show that $\alpha_{\mathbf{a},\Delta} = \alpha_{\mathbf{b},\Delta}$ for every $\mathbf{a}, \mathbf{b} \in \mathcal{S}_\mu^n$ and generalized Hamming distance parameter Δ. Let $\pi \in \mathrm{Sym}_n$ be a permutation such that $\pi(\mathbf{a}) = \mathbf{b}$ and let $\mathbf{c} \in \mathcal{S}_\mu^n$ be an arbitrary point such that $\Delta(\mathbf{a}, \mathbf{c}) = \Delta$. Further, let t_Δ denote the number of points $\mathbf{d} \in \mathcal{S}_\mu^n$ such that $\Delta(\mathbf{a}, \mathbf{d}) = \Delta$ (note that this does not depend on $\mathbf{a}$). Then using (13) we have the following.

$$W(\mathbf{a}, \mathbf{c}) = \alpha_{\mathbf{a},\Delta} \cdot \frac{1}{t_\Delta} \tag{14}$$

Since $W(\mathbf{a}, \mathbf{c}) = W(\pi(\mathbf{a}), \pi(\mathbf{c})) = W(\mathbf{b}, \pi(\mathbf{c}))$ and $\delta(\mathbf{b}, \pi(\mathbf{c})) = \Delta$, using (13) again, we have:

$$W(\mathbf{b}, \pi(\mathbf{c})) = \alpha_{\mathbf{b},\Delta} \cdot \frac{1}{t_\Delta} \tag{15}$$

From (14) and (15), we get that $\alpha_{\mathbf{a},\Delta} = \alpha_{\mathbf{b},\Delta} = \Delta$; hence we can simply denote $\alpha_{\mathbf{a},\Delta}$ by α_Δ. Now, using (13) for all the rows $\mathbf{a} \in \mathcal{S}_\mu^n$ of W, we obtain

$$W = \sum_{\Delta \text{ is } C\text{-balanced}} \alpha_\Delta W_\Delta,$$

where $\alpha_\Delta \geqslant 0$ and $\sum_{\Delta \text{ is } C\text{-balanced}} \alpha_\Delta = 1$.

Since $W_\Delta^\top = W_{\Delta^\top}$ is stochastic, we note that W_Δ is doubly stochastic. Thus, by applying Lemma 2.5 we conclude that $\lambda_2(W') \leqslant \max_{\Delta \text{ is } C\text{-balanced}}\{\lambda_2(W'_\Delta)\} \leqslant 1/n^\tau$, where $\tau > 0$ is a constant given by Lemma 3.16. This finishes the proof of Theorem 1.1.

■

4 Near-Optimal Distance Lemmas Over Balanced Multislices

In this section, we derive near-optimal polynomial lemmas for junta-sums and polynomials over the balanced multislice. More formally, for a finite set $\mathcal{S}$ of size $s \geqslant 2$, integer $d \geqslant 0$, positive integer n divisible by s and $\mu = (n/s, \ldots, n/s)$ (repeated s times), we recall that $\mathcal{S}^n_\mu \subseteq \mathcal{S}^n$ denotes the set of points in which each element $i \in \mathcal{S}$ appears n/s many times.

We also recall that $\mathcal{J}_d(\mathbb{Z}_s^n, G)$ denotes the family of d-junta sums from the domain $\mathbb{Z}_s^n$ to an Abelian group G. Similarly, we let $\mathcal{P}_d(\mathbb{F}_q^n)$ denote the family of polynomials of degree at most d over a finite field $\mathbb{F}_q$. The well-known ODLSZ lemma states that $\mathcal{P}_d(\mathbb{F}_q^n)$ forms a code of relative distance $\delta = \delta(q,d)$ independent of n. Stated more formally,

Lemma 4.1 (Polynomial distance lemma (ODLSZ lemma))**.** *(See e.g. [GRS23, Lemma 9.4.1]). For every finite field $\mathbb{F} = \mathbb{F}_q$, if a polynomial $P \in \mathcal{P}_d(\mathbb{F}^n)$ is such that $P(\mathbf{a}) \neq 0$ for some $\mathbf{a} \in \mathbb{F}^n$, then*

$$\Pr_{\mathbf{b}\sim\mathbb{F}^n}[P(\mathbf{b}) \neq 0] \geqslant \delta(q,d),$$

where $\delta(q,d) = (1-\beta/q)q^{-\alpha}$, where α and β are the quotient and remainder respectively when d is divided by $q-1$.

With this setup, we prove the following two main theorems in this section.

Theorem 4.2 (**Distance of junta-sums over multislice**)**.** *If a junta-sum $P \in \mathcal{J}_d(\mathcal{S}^n, G)$ is such that $P(\mathbf{a}) \neq 0$ for some $\mathbf{a} \in \mathcal{S}^n_\mu$, then*

$$\Pr_{\mathbf{b}\sim\mathcal{S}^n}[P(\mathbf{b}) \neq 0] \geqslant \frac{1}{s^d} - \frac{1}{n^{\Omega_s(1)}}.$$

As noted in Section 1 we also prove a similar theorem for algebraic degree as opposed to junta-degree. We recall the theorem statement below.

Theorem 1.4 (**Polynomial distance over multislice**)**.** *For every finite field $\mathbb{F} = \mathbb{F}_q$, if a degree d polynomial $P(\mathbf{x})$ is such that $P(\mathbf{a}) \neq 0$ for some $\mathbf{a} \in \mathbb{F}^n_\mu$ on the balanced multislice, then*

$$\Pr_{\mathbf{b}\sim\mathbb{F}^n_\mu}[P(\mathbf{b}) \neq 0] \geqslant \delta(q,d) - \frac{1}{n^{\Omega_q(1)}},$$

where $\delta(q,d) = (1-\beta/q)q^{-\alpha}$, where α and β are the quotient and remainder respectively when d is divided by $q-1$.

We first prove Theorem 4.2 below followed by the proof of Theorem 1.4, which is almost identical.

Proof of Theorem 4.2. Without loss of generality, we will assume that $\mathcal{S} = \mathbb{Z}_s$, so addition and subtraction of elements of $\mathcal{S}$ make sense. At a high level, the proof proceeds as follows. We consider a random walk matrix W over the multislice which we will describe below, and bound its eigenvalues. We will then use the "expander mixing lemma" to derive the required distance lower bound and finish the proof.

We shall define a random walk matrix W_{ODLSZ} over the points in the multislice, i.e., $V = \mathcal{S}^n_\mu$ and we let $N = |V| = \binom{sm}{m,\ldots,m}$ where $m = n/s$ is an integer. For each $\mathbf{a} \in V$, we define the distribution over the neighbors of $\mathbf{a}$ according to W_{ODLSZ} (or equivalently, the $\mathbf{a}$-th row of W_{ODLSZ}, denoted $W_{\mathrm{ODLSZ}}(\mathbf{a})$) as being the random variable output by the algorithm below (Algorithm 1).

Algorithm 1: The random walk matrix W_{ODLSZ}

Input: $\mathbf{a} \in V$

1 For $j \in \mathbb{Z}_s$, letting $\mathbf{a}^{-1}(j) \in \binom{[n]}{m}$ denote the coordinates of $\mathbf{a}$ with value j, sample uniformly random bijections $M_j : \mathbf{a}^{-1}(j) \to [m]$ independently for all $j \in \mathbb{Z}_s$.
2 Sample $\mathbf{y} = (y_1, \ldots, y_m) \sim \mathbb{Z}_s^m$ u.a.r.
3 Define $\mathbf{b} = (b_1, \ldots, b_n)$ as follows: For $i \in [n]$, we let $j := a_i$ and $b_i := y_{M_j(i)} + j$.
4 **return b**

We first note that $\mathbf{b}$ is always on the balanced multislice, i.e., $\mathbf{b} \in V$, so W_{ODLSZ} is a well-defined random walk matrix over the balanced multislice. We now argue that for every fixed $\mathbf{a} \in V$ such that $P(\mathbf{a}) \neq 0$ for a junta-sum $P \in \mathcal{J}_d(\mathcal{S}^n, G)$, it holds that

$$\Pr_{\mathbf{b} \sim W_{\mathrm{ODLSZ}}(\mathbf{a})}[P(\mathbf{b}) \neq 0] \geqslant 1/s^d.$$

To see this, we fix the bijections $M_j : a^{-1}(j) \to [m]$ in Step 1 of Algorithm 1 arbitrarily and get the probability bound over the uniformly random choice of $\mathbf{y}$ in Step 2. More precisely, letting

$$Q(\mathbf{y}) = P(\mathbf{b}) = P((y_{M_{a_i}(i)} + a_i)_{i \in [n]}),$$

we note that $Q : \mathcal{S}^m \to G$ is a d-junta-sum since P is a d-junta-sum. Moreover, $Q(\mathbf{0}) = P(\mathbf{a}) \neq 0$. Therefore, by applying Claim 2.6, we get that $\Pr_{\mathbf{y} \sim \mathcal{S}^m}[Q(\mathbf{y}) \neq 0] \geqslant 1/s^d$, and thus $\Pr_{\mathbf{b} \sim W_{\mathrm{ODLSZ}}(\mathbf{a})}[P(\mathbf{b}) \neq 0] \geqslant 1/s^d$.

Letting $U \subseteq V$ denote the set of points in V which evaluate P to a non-zero value, from the above discussion, we have that

$$\forall \mathbf{a} \in U, \quad \Pr_{\mathbf{b} \sim W_{\mathrm{ODLSZ}}(\mathbf{a})}[\mathbf{b} \in U] \geqslant 1/s^d. \tag{16}$$

We now use the expander mixing lemma.

Theorem 4.3 (**Expander mixing lemma** see e.g. [HLW06] Lemma 2.5)**.** *For every symmetric random walk matrix $W \in \mathbb{R}^{V \times V}$ over a finite vertex set V and $U \subset V$,*

$$\Pr_{\substack{\mathbf{a}\sim V \\ \mathbf{b}\sim W(\mathbf{u})}} [\mathbf{a} \in U \text{ and } \mathbf{b} \in U] \leqslant \left(\frac{|U|}{|V|}\right)^2 + \lambda_2(W)\left(\frac{|U|}{|V|}\right),$$

where $\lambda_2(W)$ denotes the second largest eigenvalue of W in absolute value.

In order to apply the above theorem, we will need to show that the random walk matrix W_{ODLSZ} we defined is symmetric and has a small $\lambda_2(W_{\text{ODLSZ}})$.

Lemma 4.4. *The random walk matrix W_{ODLSZ} as defined in Algorithm 1 is symmetric and satisfies $\lambda_2(W_{\text{ODLSZ}}) \leqslant 1/n^{\Omega_s(1)}$.*

We prove this lemma in Section 4.1.

We can now finish the proof of Theorem 4.2 assuming the above lemma. On the one hand, (16) implies that

$$\Pr_{\substack{\mathbf{a}\sim V \\ \mathbf{b}\sim W_{\text{ODLSZ}}(\mathbf{u})}} [\mathbf{a} \in U \text{ and } \mathbf{b} \in U] \geqslant \left(\frac{|U|}{|V|}\right)\frac{1}{s^d},$$

and on the other hand, Theorem 4.3 and Lemma 4.4 imply that

$$\Pr_{\substack{\mathbf{a}\sim V \\ \mathbf{b}\sim W_{\text{ODLSZ}}(\mathbf{u})}} [\mathbf{a} \in U \text{ and } \mathbf{b} \in U] \leqslant \left(\frac{|U|}{|V|}\right)\left(\frac{|U|}{|V|} + \frac{1}{n^{\Omega_s(1)}}\right).$$

Putting them together, we obtain that $\frac{|U|}{|V|} \geqslant \frac{1}{s^d} - \frac{1}{n^{\Omega_s(1)}}$, thus finishing the proof of Theorem 4.2. ■

We now prove the near-optimal distance lemma for algebraic degree (Theorem 1.4).

Proof of Theorem 1.4. The proof follows exactly the same approach as that of the distance lemma for junta-sums over the balanced multislice (i.e., Theorem 4.2). All the additions and subtraction of the domain elements are now instead done over the field $\mathbb{F}$ instead of the group $\mathbb{Z}_s$. The only other difference is in (16) where we now get a lower bound of $\delta(q, d)$ instead of $1/s^d$. This is because the restricted function $Q : \mathbb{F}^m \to \mathbb{F}$ is now a function of degree at most d, so we can apply the standard ODSLZ lemma (Lemma 4.1) instead of the junta-sum distance lemma (Claim 2.6) to get this bound. Due to its similarity with the proof of Theorem 4.2, we omit the rest of the details. ■

Hence, it only remains to prove the eigenvalue bounds for W_{ODLSZ}, i.e., Lemma 4.4, which we do in the next subsection.

4.1 Eigenvalue Bounds for W_{ODLSZ}

Before we proceed with the proof of Lemma 4.4, we remark that it doesn't immediately follow from our result for nearly *balanced* random walks (i.e, Theorem 1.1 from Section 3) since W_{ODLSZ} can potentially have non-zero weights even for edges whose generalized Hamming distance is far from being balanced. Moreover, it doesn't even immediately follow from our more general theorem (Theorem 1.3) from Section 3 as it requires a bounded Frobenius norm which isn't the case with W_{ODLSZ}. However, we are able to reduce it to a setting where Theorem 1.1 actually applies and use it to get the final bound.

Proof of Lemma 4.4. At a high level, we prove this in the following steps. We first provide an alternate description of the random walk matrix W_{ODLSZ} (defined in Algorithm 1) using generalized Hamming distance matrices. Then, we express W_{ODLSZ} as a convex combination $W_{\text{ODLSZ}} = \sum_{i\in[t]} \alpha_i W_i$ for some random walk matrices W_i where $\sum_i \alpha_i = 1$. Then, we use our expansion result for nearly balanced walks (Theorem 1.1) from Section 3 to bound $\lambda_2(W_i)$ for "most" $i \in [t]$, and use this to finally bound $\lambda_2(W_{\text{ODLSZ}})$. Before we go into the actual proof, we need to recall a few definitions.

For $\mathbf{a}, \mathbf{b} \in V$, we recall (from Definition 2.1) that $\Delta(\mathbf{a}, \mathbf{b}) \in \mathbb{Z}^{\mathbb{Z}_s \times \mathbb{Z}_s}$ denotes the *generalized Hamming distance matrix*, i.e., the (i, j)-th entry of the matrix equals the number of coordinates where $\mathbf{a}$ takes value i and $\mathbf{b}$ takes value j. We now recall the definition of W_{ODLSZ} (from Algorithm 1): For each $\mathbf{a} \in V$, its random neighbor $\mathbf{b} \sim W_{\text{ODLSZ}}(\mathbf{a})$ is obtained by setting $b_i = y_{M_j(i)} + j$, where $j = a_i$ and $M_j : \mathbf{a}^{-1}(j) \to [m]$ are bijections chosen u.i.a.r., and $\mathbf{y} \sim \mathbb{Z}_s^n$ is chosen independently. Hence, we see that

$$\Delta(\mathbf{a}, \mathbf{b})(i, j) = f_{j-i},$$

where f_j denotes the number of times $j \in \mathbb{Z}_s$ appears in $\mathbf{y}$. In fact, conditioned on $\Delta(\mathbf{a}, \mathbf{b}) = P$ for some fixed P, the conditional distribution of $\mathbf{b}$ is uniform over all points $\mathbf{b}$ such that $\Delta(\mathbf{a}, \mathbf{b}) = P$, since M_j's are uniform and independent bijections. In particular, this alternate description of W_{ODLSZ} shows that it is symmetric.

Now, for a "frequency vector" $\mathbf{f} = (f_0, \ldots, f_{s-1}) \in \mathbb{Z}^{\mathbb{Z}_s}$ where $\sum_j f_j = m$, we let $W_{\mathbf{f}}$ denote the random walk matrix where for each $\mathbf{a} \in V$, $W_{\mathbf{f}}(\mathbf{a})$ is the uniform distribution over

$$\left\{ \mathbf{b} \in V : \Delta(\mathbf{a}, \mathbf{b}) = (f_{j-i})_{(i,j)\in\mathbb{Z}_s^2} \right\}. \tag{17}$$

Then by our previous discussion, for each $\mathbf{a} \in V$, by conditioning on the choice of the frequency vectors resulting from $\mathbf{y} \sim \mathcal{S}^m$ and using the total probability law, we obtain

$$W_{\text{ODLSZ}}(\mathbf{a}) = \sum_{\substack{\mathbf{f}\in\mathbb{Z}^s \\ \sum_j f_j = m}} \alpha_{\mathbf{f}} W_{\mathbf{f}}(\mathbf{a}),$$

where $\alpha_{\mathbf{f}} = \binom{m}{\mathbf{f}}/s^m$ denotes the probability of getting the frequency vector $\mathbf{f}$ from a uniformly random $\mathbf{y} \in \mathcal{S}^m$.

The idea now is to apply our eigenvalue bound (Theorem 1.1) from Section 3 to the $W_{\mathbf{f}}$'s and then bound the eigenvalues of W_{ODLSZ}. However, there are two issues: First, the eigenvalue bound from Theorem 1.1 requires the matrix to be supported only on edges with nearly balanced

generalized Hamming distance, which isn't the case for *all* $W_{\mathbf{f}}$'s. Regardless, we show that this holds true for the "typical" $W_{\mathbf{f}}$'s and that this suffices. And secondly, we remark that each $W_{\mathbf{b}}$ need not be a symmetric matrix. However, since we already know that $W_{\text{ODLSZ}} = \sum_{\mathbf{f}} \alpha_{\mathbf{f}} W_{\mathbf{f}}$ is symmetric, we have that

$$W_{\text{ODLSZ}} = \sum_{\mathbf{f}} \alpha_{\mathbf{f}} \left(\frac{W_{\mathbf{f}} + W_{\mathbf{f}}^\top}{2} \right),$$

where now we see that the "components" $\frac{W_{\mathbf{f}}+W_{\mathbf{f}}^\top}{2}$ are symmetric.

We say that a frequency vector $\mathbf{f}$ is "bad" if there exists a $j \in \mathbb{Z}_s$ such that $f_j \notin \frac{m}{s} \pm \sqrt{\frac{10m \log m}{s}}$, and say that $\mathbf{f}$ is "good" otherwise. We have, by a Chernoff bound, that

$$\sum_{\mathbf{f} \text{ bad}} \alpha_{\mathbf{f}} \leqslant \frac{s}{m^{\Omega(1)}}. \tag{18}$$

Now we claim that for every good $\mathbf{f}$, it holds that $\lambda_2(W'_{\mathbf{f}})$ is small where $W'_{\mathbf{f}} := \frac{W_{\mathbf{f}}+W_{\mathbf{f}}^\top}{2}$:

Claim 4.5 (Eigenvalue bounds for $W'_{\mathbf{f}}$). *Suppose $\mathbf{f} = (f_0, \ldots, f_{s-1})$ is such that $f_j \in \frac{m}{s} \pm \sqrt{\frac{10m \log m}{s}}$ for all $j \in \mathbb{Z}_s$. Then, $\lambda_2(W'_{\mathbf{f}}) \leqslant \frac{1}{n^{\Omega_s(1)}}$, where $W'_{\mathbf{f}} := \frac{W_{\mathbf{f}}+W_{\mathbf{f}}^\top}{2}$.*

Proof. We note that the matrix $W'_{\mathbf{f}}$ respects symmetries and has non-zero entries only on entries corresponding to a balanced generalized Hamming distance of either Δ or $\Delta^\top$ (both of which are $(10/s)$-balanced). Hence the proof follows directly by applying Theorem 1.1 to the matrix $W'_{\mathbf{f}}$. ■

We now bound the eigenvalues of W_{ODLSZ} and finish the proof of Lemma 4.4. By applying Lemma 2.5 with S being the set of good $\mathbf{f}$, we have

$$\begin{aligned}
\lambda_2(W_{\text{ODLSZ}}) &= \lambda_2 \left(\sum_{\mathbf{f}} \alpha_{\mathbf{f}} W'_{\mathbf{f}} \right) \\
&\leqslant \max_{\mathbf{f} \text{ good}} \{\lambda_2(W'_{\mathbf{f}})\} + \sum_{\mathbf{f} \text{ bad}} \alpha_{\mathbf{f}} && \text{(using Lemma 2.5)} \\
&\leqslant \left(\frac{1}{n^{\Omega_s(1)}} \right) + \left(\sum_{\mathbf{f} \text{ bad}} \alpha_{\mathbf{f}} \right) && \text{(applying Claim 4.5 to the random walk matrix } W'_{\mathbf{f}}) \\
&\leqslant \frac{1}{n^{\Omega_s(1)}}. && \text{(using (18))}
\end{aligned}$$

The above bound shows that all the eigenvalues of W_{ODLSZ} except the largest one must be bounded above by $1/n^{\Omega_s(1)}$ in absolute value, i.e., $\lambda_2(W_{\text{ODLSZ}}) \leqslant 1/n^{\Omega_s(1)}$ proving Lemma 4.4. ■

5 Local List Correction of Junta-Sums

In this section, we will prove the following theorem which is a restatement of Theorem 1.5 with explicit bounds on the query complexity.

Theorem 5.1 (Local List Correction)**.** *For every Abelian group G and for every $\varepsilon > 0$, the space $\mathcal{J}_d(\mathcal{S}^n, G)$ is $(1/s^d - \varepsilon,\ \mathcal{O}_\varepsilon(1),\ \tilde{\mathcal{O}}_\varepsilon(\log n)^d,\ \mathcal{O}_\varepsilon(1))$-locally list correctable.*

In particular, there is a randomized algorithm $\mathcal{A}$ such that for a function $f : \mathcal{S}^n \to G$ and a parameter $\varepsilon > 0$, $A^f(\varepsilon)$ outputs with probability $\geqslant 3/4$ a list of randomized algorithms $\{\phi_i\}_{i=1}^L$ ($L = \mathcal{O}_\varepsilon(1)$) such that the following holds. For each junta degree-d function $P \in \mathcal{J}_d$ that is $(1/s^d - \varepsilon)$-close to f, there exists at least one randomized algorithm ϕ_i such that ϕ_i^f computes P correctly on every input in $\mathcal{S}^n$ with probability at least $3/4$.

The algorithm $\mathcal{A}$ makes $\mathcal{O}_\varepsilon(1)$ queries to f, while each ϕ_i makes $\tilde{\mathcal{O}}_\varepsilon(\log n)^d$ oracle queries to f.

We remark that for the Boolean case, [ABPSS25] proves that one can reduce the number of queries to a constant depending only on ε and the *torsion* (or *exponent*) of the group (see Section 2 for a definition). Similarly, in this case, we get a similar statement where the algorithm $\mathcal{A}$ makes $\mathcal{O}_\varepsilon(1)$ queries, and each ϕ_i makes $\mathcal{O}_{M,\varepsilon}(1)$ queries where M is the exponent of the torsion Abelian group G. More formally, we prove the following.

Theorem 5.2. *For every torsion Abelian group G of exponent $M > 0$ and every $\varepsilon > 0$, the family $\mathcal{J}_d(\mathcal{S}^n, G)$ is $(1/s^d - \varepsilon, \mathcal{O}_\varepsilon(1), \mathcal{O}_{M,\varepsilon}(1), \mathcal{O}_\varepsilon(1))$-locally list correctable.*

As stated earlier, [ABPSS25] gave a local list corrector for degree-d *polynomials* over $\mathcal{S} = \{0, 1\}$. We note that most of their proof can be extended to *junta sums* and general $\mathcal{S}$ with some extensions to their arguments. However, a key challenge was to show that certain random walk matrix has good spectral expansion. In particular, [ABPSS25, Lemma 5.1.1] is proved by analyzing the eigenvalues of matrices defined on Johnson graphs. To extend their argument to general grids, we have to analyze the eigenvalues of random walk matrices on the balanced multi-slice. In this section, we describe the random walk matrix arising from the analysis of our local list corrector and show that it has "large" spectral gap, using Theorem 1.1. We first give a quick overview of the algorithm, which is an extension of [ABPSS25, Algorithm 3 and Algorithm 4].

Overview of the local list corrector. Similar to the work of [ABPSS25], our local list corrector goes as follows:

- We design a local corrector $\mathcal{J}_d(\mathcal{S}^n, G)$ (see Theorem 5.3).
- We show a combinatorial list decoding bound for $\mathcal{J}_d(\mathcal{S}^n, G)$ (see Theorem 5.4).
- We design *approximating oracles* for $\mathcal{J}_d(\mathcal{S}^n, G)$ (see Theorem 5.5).
- Combining an approximating oracle with local corrector, we get a local list corrector. The bound on query complexity follows from the combinatorial list decoding bound.

Our key technical contribution is in analyzing the approximating oracles. We use a very similar algorithm for approximating oracles as in [ABPSS25], however the correctness is more involved. The first two steps are again analogous to [ABPSS25, Section 3 and Section 4]. Most of the arguments follow with a simple extension from $\{0, 1\}$ to $\mathcal{S}$, and few arguments require a bit more careful analysis. For the sake of completeness, we give a proof for local corrector and combinatorial

list decoding bound. We state the results for them below, and after that we will proceed with the local list corrector.

Theorem 5.3 (*Local correction of junta-sums*). *For every $\varepsilon > 0$, finite set $\mathcal{S}$ of size $s \geqslant 2$ and $d \geqslant 0$, Abelian group G, the family $\mathcal{J}_d(\mathcal{S}^n, G)$ is $(\widetilde{\mathcal{O}}_\varepsilon(\log n)^d, \delta_{\mathcal{J}}/2 - \varepsilon)$-locally correctable where $\delta_{\mathcal{J}} := 1/s^d$.*
Moreover, if G is a torsion Abelian group of exponent M, then the number of queries can be made $O_{M,\varepsilon}(1)$, i.e., $\mathcal{J}_d(\mathcal{S}^n, G)$ is $(\mathcal{O}_{M,\varepsilon}(1), \delta_J/2 - \varepsilon)$-locally correctable.

Theorem 5.4 (Combinatorial List Decoding Bound). *For every $\varepsilon > 0$, positive integers s, d, and Abelian group G, the family $\mathcal{J}_d(\mathcal{S}^n, G)$ is $(1/s^d - \varepsilon, \mathcal{O}_\varepsilon(1))$-list decodable.*

For every $f : \mathcal{S}^n \to G$ which is $(\frac{1}{s^d} - \varepsilon)$-close to $\mathcal{J}_d(\mathcal{S}^n, G)$, let $\mathsf{List}_\varepsilon(f)$ denote the set of d-junta-sums that have distance $\leqslant (1/s^d - \varepsilon)$ to f, i.e.

$$\mathsf{List}_\varepsilon(f) \;=\; \left\{ P \in \mathcal{J}_d(\mathcal{S}^n, G) \;\middle|\; \delta(f, P) \leqslant \frac{1}{s^d} - \varepsilon \right\}.$$

We give a proof for Theorem 5.3 and Theorem 5.4 later. We informally state a standard observation[12] in the literature of local list correctors which says that given local correctors, it is enough to design approximating oracles (see Theorem 5.5):

If there exists a local corrector, then it suffices to design an algorithm which outputs a list of algorithms with the guarantee - For every junta-sum P in the list, there exists an algorithm A in the list which computes P correctly on sufficiently large fraction of $\mathcal{S}^n$, and then we can run the local corrector on A.

So the focus in this is to design the approximating oracles. The following theorem is larger-grid analogue to [ABPSS25, Theorem 5.0.1].

Theorem 5.5 (Approximate oracles). *Fix $n \in \mathbb{N}$, $\varepsilon > 0$. Let $f : \mathcal{S}^n \to G$ be any function and $L(\varepsilon) := |\mathsf{List}_\varepsilon(f)|$. There exists a randomized algorithm $\mathcal{A}_1^f$ that makes at most $\mathcal{O}_\varepsilon(1)$ oracle queries and outputs deterministic algorithms $\Psi_1, \ldots, \Psi_{L'}$ satisfying the following property: With probability at least $3/4$, for every junta-sum $P \in \mathsf{List}_\varepsilon^f$, there exists a $j \in [L']$ such that*
1. *$\delta(\Psi_j, P) < 1/(10 \cdot 2^{d+1})$*
2. *For every $\mathbf{x} \in \mathcal{S}^n$, Ψ_j computes $P(\mathbf{x})$ by making at most $\mathcal{O}_\varepsilon(1)$ oracle queries to f.*

Here $L' = \mathcal{O}(L(\varepsilon/2) \log L(\varepsilon)) = \mathcal{O}_\varepsilon(1)$.

We first show that using Theorem 5.5 and Theorem 5.3, we can prove Theorem 5.1 (and Theorem 5.2).

[12] This is also used in [ABPSS24; ABPSS25]. See [ABPSS25, Section 5] for a more elaborate discussion on it.

Proof of Theorem 5.1 and Theorem 5.2. We first employ Algorithm 4 with oracle access to f and it outputs deterministic algorithms $\psi_1, \ldots, \psi_{L'}$ where $L' = \mathcal{O}(L(\varepsilon/2)\log L(\varepsilon))$. Next, we run the local corrector for $\mathcal{J}_d(\mathcal{S}^n, G)$ on each of ψ_j. This completes the description of our local list corrector. The correctness and query complexity now follow by combining Theorem 5.5 and Theorem 5.3. ■

Organization of the section We start by proving a sampling lemma for the *balanced slice* of the grid $\mathcal{S}^n$ in Section 5.1. The key tool to prove this sampling lemma will be to show that a certain random walk matrix (it arises from our sampling procedure) is a "good spectral expander" (see Theorem 5.9). We will prove by employing Theorem 1.1. After the sampling lemma, we then prove a sub-optimal distance lemma for d-juntas on multi-slices of $\mathcal{S}^n$ (see Theorem 5.10). Combining the sampling lemma (Lemma 5.7) and the distance lemma on slices (Theorem 5.10), we get Corollary 5.12. This corollary will be useful in showing that our local list correctors have a small error probability, i.e., Corollary 5.12 will bound the probability of our local list correctors making a certain type of error. Once we have these statements, we describe a subroutine in Section 5.3 and the local list correctors in Section 5.4. Finally, we analyze the algorithms in Section 5.5.

5.1 A Sampling Lemma for the Balanced Multislice

Definition 5.6. *Let $k, s \in \mathbb{N}$. For a s-to-1 map $\tau : [sk] \to [k]$, let $\mathsf{C}_\tau \subset \mathcal{S}^{sk}$ denote the k-dimensional subgrid obtained by identifying coordinates acccording to τ. More precisely, for every $\mathbf{y} \in \mathcal{S}^k$, let $x_\tau(\mathbf{y}) \in \mathcal{S}^{sk}$ be defined as follows:*

$$x_\tau(\mathbf{y})_i \;=\; y_{\tau(i)}, \qquad\qquad \text{for all } \; i \in [sk].$$

Define $\mathsf{C}_\tau := \left\{ x_\tau(\mathbf{y}) \;\middle|\; \mathbf{y} \in \mathcal{S}^k \right\}$.

The main lemma of this subsection is to show that if we sample a uniformly random s-to-1 map τ, then C_τ is a *good sampler* for the balanced slice of $\mathcal{S}^{sk}_{k,\ldots,k}$.

Lemma 5.7 (Sampler for the Balanced Slice)**.** *Let $k, s \in \mathbb{N}$. There exists an absolute constant $\eta = \eta(s) > 0$ such that for every subset $S \subseteq \mathcal{S}^{s^2k}_{sk,\ldots,sk}$, we have,*

$$\Pr_\tau \left[\left| \frac{|S|}{|\mathcal{S}^{s^2k}_{sk,\ldots,sk}|} - \frac{|S \cap \mathsf{C}_\tau|}{|\mathcal{S}^{sk}_{k,\ldots,k}|} \right| \geqslant \frac{1}{k^\eta} \right] \leqslant \mathcal{O}_s\left(\frac{1}{k^\eta}\right),$$

where the probability is over the choice of a random s-to-1 map $\tau : [s^2k] \to [sk]$.

Description of the matrix W: For this section, we will assume that n is divisible by s. We will use μ to denote the *balanced partition* of n into s rows, i.e. $\mu = (n/s, n/s, \ldots, n/s)$. Let N denote the number of points in the balanced slice $\mathcal{S}^n_\mu$, i.e. $N = |\mathcal{S}^n_\mu| = \binom{n}{n/s, n/s, \ldots, n/s}$.

Definition 5.8 (The matrix W)**.** *We will define the random walk matrix by the joint distribution over $\mathcal{S}^n_\mu \times \mathcal{S}^n_\mu$ represented by the matrix W/N. In particular, it is the joint probability distribution of $(\mathbf{u}, \mathbf{v})$ corresponding to picking a uniformly random vertex $\mathbf{u} \sim \mathcal{S}^n_\mu$ and $\mathbf{v} \sim W(\mathbf{a})$ is its random*

neighbor corresponding to taking a random step according to W. We then define W/N according to the distribution of the output of the following steps:

1. *Pick $\mathbf{a}, \mathbf{b} \sim \mathcal{S}^{sk}_{k,\dots,k}$ uniformly and independently at random.*
2. *Pick a s-to-1 map $\tau : [s^2k] \to [sk]$ uniformly at random.*
3. *Output $(\mathbf{u}, \mathbf{v}) = (x_\tau(\mathbf{a}), x_\tau(\mathbf{b}))$ (see Definition 5.6 for the definition of $x_\tau(\cdot)$).*

We note that W is symmetric since the joint probability distribution of $(\mathbf{u}, \mathbf{v})$ above is symmetric w.r.t. $\mathbf{u}$ and $\mathbf{v}$. We further claim below that W has good spectral expansion:

Theorem 5.9 (Spectral expansion of the random walk matrix)**.** *Let $W \in \mathbb{R}^{N \times N}$ be the symmetric random walk matrix as described previously. Denote the second largest eigenvalue of W (in terms of absolute value) by $\lambda_2(W)$. Then there exists $\nu = \nu(s) > 0$ such that*

$$\lambda_2(W) \leqslant \frac{1}{n^\nu}.$$

We first prove Lemma 5.7 assuming Theorem 5.9.

Proof of Lemma 5.7. Let $\sigma := |S|/|\mathcal{S}^{s^2k}_{sk,\dots,sk}|$. For every $\mathbf{y} \in \mathcal{S}^k$, define $Z(\mathbf{y})$ to be the indicator variable which is 1 if $x_\tau(\mathbf{y}) \in S$. For a uniformly random s-to-1 map τ, for every $\mathbf{y} \in \mathcal{S}^{sk}_{k,\dots,k}$, the random variable $x_\tau(\mathbf{y})$ is uniformly distributed in $\mathcal{S}^{s^2k}_{sk,\dots,sk}$. Thus for every $\mathbf{y} \in \mathcal{S}^{sk}_{k,\dots,k}$, the $\mathbb{E}_\tau[Z(\mathbf{y})] = \sigma$. Let $Z := |S \cap \mathsf{C}_\tau| = \sum_{\mathbf{y} \in \mathcal{S}^{sk}_{k,\dots,k}} Z_{\mathbf{y}}$ and by linearity of expectation, we have $\mathbb{E}_\tau[Z] = |\mathcal{S}^{sk}_{k,\dots,k}| \cdot \sigma$. We will now bound the variance of Z.

Using the linearity of expectation, we have,

$$\mathbb{E}_\tau[Z^2] = \sum_{\mathbf{a},\mathbf{b} \in \mathcal{S}^{sk}_{k,\dots,k}} \mathbb{E}_\tau[Z(\mathbf{a}) \cdot Z(\mathbf{b})] = \sum_{\mathbf{a},\mathbf{b} \in \mathcal{S}^{sk}_{k,\dots,k}} \Pr_\tau[x_\tau(\mathbf{a}) \in S \ \wedge \ x_\tau(\mathbf{b}) \in S]$$

If we sample $\mathbf{a}, \mathbf{b}$ uniformly and independently at random from $\mathcal{S}^{sk}_{k,\dots,k}$, then by the definition of the matrix W (Definition 5.8), we get the following equality:

$$\Pr_{\substack{\mathbf{a},\mathbf{b} \sim \mathcal{S}^{sk}_{k,\dots,k} \\ \tau}} [x_\tau(\mathbf{a}) \in S \ \wedge \ x_\tau(\mathbf{b}) \in S] = \Pr_{\substack{\mathbf{u} \sim \mathcal{S}^{s^2k}_{sk,\dots,sk} \\ \mathbf{v} \sim W(\mathbf{u})}} [\mathbf{u} \in S \ \wedge \ \mathbf{v} \in S]. \tag{19}$$

Using the Expander Mixing Lemma (see e.g. Theorem 4.3),

$$\begin{aligned} \Pr_{\substack{\mathbf{u} \sim \mathcal{S}^{s^2}_{sk,\dots,sk} \\ \mathbf{v} \sim W(\mathbf{u})}} [\mathbf{u} \in S \ \wedge \ \mathbf{v} \in S] &\leqslant \sigma^2 + \lambda_2(W) \\ \Rightarrow \mathbb{E}_\tau[Z^2] &\leqslant |\mathcal{S}^{sk}_{k,\dots,k}|^2 \cdot (\sigma^2 + \lambda_2(W)) \\ \Rightarrow \mathrm{Var}[Z] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 &\leqslant |\mathcal{S}^{sk}_{k,\dots,k}|^2 \cdot \lambda_2(W). \end{aligned}$$

Now using Chebyshev's inequality on Z, we get,

$$\Pr_{\tau}\left[\left|Z-\sigma\cdot|\mathcal{S}^{sk}_{k,\ldots,k}|\right|\geqslant\frac{1}{k^{\eta}}\cdot|\mathcal{S}^{sk}_{k,\ldots,k}|\right]\;\leqslant\;\mathrm{Var}(Z)\cdot\frac{k^{2\eta}}{|\mathcal{S}^{sk}_{k,\ldots,k}|^2}\;\leqslant\;\lambda_2(W)\cdot k^{2\eta}$$

From Theorem 5.9, we know that $\lambda_2(W)\leqslant 1/k^{\eta}$, which implies that

$$\Pr_{\tau}\left[\left|Z-\sigma\cdot|\mathcal{S}^{sk}_{k,\ldots,k}|\right|\geqslant\frac{1}{k^{\eta}}\cdot|\mathcal{S}^{sk}_{k,\ldots,k}|\right]\;\leqslant\;\frac{1}{k^{\eta}}.$$

This finishes the proof of Lemma 5.7. ■

We now prove Theorem 5.9 which bounds the eigenvalues of the random walk matrix W.

Proof of Theorem 5.9. We recall (from Definition 5.8) that the random walk W is over the balanced multislice $V:=\mathcal{S}^{s^2k}_{\mu}=\mathcal{S}^{s^2k}_{sk,\ldots,sk}$ and $N=|\mathcal{S}^{n}_{\mu}|$ denotes the number of vertices where $n=s^2k$. We will use the following equivalent description of W. We observe that W/N is the joint probability distribution of $(\mathbf{u},\mathbf{v})$ corresponding to picking a uniformly random vertex $\mathbf{u}\sim V$ and $\mathbf{v}\sim W(\mathbf{a})$ is its random neighbor corresponding to a taking a random step according to W. We now rephrase the description of W/N from Definition 5.8:

1. Pick $\mathbf{a},\mathbf{b}\sim\mathcal{S}^{sk}_{k,\ldots,k}$ uniformly and independently at random.
2. Let $P=\Delta(\mathbf{a},\mathbf{b})$ and $\widetilde{P}=sP$.
3. Output $(\mathbf{u},\mathbf{v})$ such that $\Delta(\mathbf{u},\mathbf{v})=\widetilde{P}$ uniformly at random.

The above output is indeed distributed according to W/N by noting that the map τ used in Definition 5.8 is chosen uniformly and independently from $\mathbf{a},\mathbf{b}\sim\mathcal{S}^{sk}_{k,\ldots,k}$ and for every such τ used in Definition 5.8, we have that $\Delta(\mathbf{u},\mathbf{v})=\Delta(x_\tau(\mathbf{a}),x_\tau(\mathbf{b}))=s\cdot\Delta(\mathbf{a},\mathbf{b})$. Now, applying a total probability rule over the choice of $\mathbf{a}$ and $\mathbf{b}$ in Step 1, we have:

$$\frac{W}{N}=\sum_{P}\alpha_P\frac{W_{sP}}{N},\tag{20}$$

where we use α_P to denote the probability that $\Delta(\mathbf{a},\mathbf{b})=P$ for $\mathbf{a},\mathbf{b}$ chosen uniformly and independently at random; and W_{sP} denotes the random walk over the multislice, determined by the generalized Hamming distance matrix sP (see Definition 2.3). We now say that a generalized Hamming distance matrix P is *good* if it is $(10/s)$-balanced (by Definition 2.2, this is equivalent to saying all entries of P are $\frac{k}{s}\pm\sqrt{\frac{10k\log k}{s}}$) and P is *bad* otherwise. It is easy to see that sP is 10-balanced w.r.t. the multislice $\mathcal{S}^{s^2k}_{\mu}$ if P is good. We first show that the mass of α_P on bad P is small: that is, we show that $\sum_{P\text{ bad}}\alpha_P$, which denotes the probability that $\Delta(\mathbf{a},\mathbf{b})$ is *not* $(10/s)$-balanced, is at most $1/k^{\Omega_s(1)}$. By fixing $\mathbf{a}$ and noting that $\mathbf{b}$ is still uniformly distributed over $\mathcal{S}^{sk}_{\mu}$, we see that each entry of P is distributed according to a hypergeometric distribution with a total of sk states and k success states, and we are picking k draws without replacement. By applying Hoeffding bound [Hoe94], we get that this probability is at most $1/2^{-\Omega_s(\sqrt{\log k/k})^2k}\leqslant 1/k^{-\Omega_s(1)}$.

Now by a union bound over all the k^2 entries of the matrix P, we get that the probability that P is not $(10/s)$-balanced is at most $1/n^{\Omega_s(1)}$ as claimed. That is,

$$\sum_{P \text{ bad}} \alpha_P \leqslant 1/k^{\Omega_s(1)}. \tag{21}$$

We now use our main eigenvalue bound from Section 3 to bound $\lambda_2(W'_{sP})$ when P is good, where $W'_{sP} := \frac{W_{sP}+W_{sP}^\top}{2}$. In particular, since we have that sP and $(sP)^\top$ are $(10/s)$-balanced for good P, by applying Theorem 1.1, we have for all good P that

$$\lambda_2(W'_{sP}) \leqslant 1/k^{\Omega_s(1)}. \tag{22}$$

We note that since we showed that W is symmetric, (20) implies that

$$W = \sum_P \alpha_P \left(\frac{W_{sP} + W_{sP}^\top}{2}\right) = \sum_P \alpha_P W'_{sP}.$$

We can therefore apply Lemma 2.5 with the set S being the set of good P to conclude that

$$\lambda_2(W) \leqslant \max_{P \text{ good}} \{\lambda_2(W'_{sP})\} + \sum_{P \text{ bad}} \alpha_P \leqslant 1/k^{\Omega_s(1)} \leqslant 1/n^{\Omega_s(1)},$$

by using (21) and (22). This finishes the proof of Theorem 5.9. ∎

5.2 Sub-optimal Distance Lemma Over Multislices

In this subsection, we prove that if a junta sum does *not* vanish on a multi-slice, then it does not vanish on at least a *constant fraction* of that multi-slice. It is a generalization of the distance lemma for junta-sums (Claim 2.6), generalized from grids to slices. In the case of $\mathcal{S} = \{0, 1\}$, such a statement was proved in [ABPSS25, Lemma 5.1.6]. They proved it by induction on the degree d. We observe that a similar induction also works for junta sums. We provide a proof below. For $\mathcal{S} = \{0, 1\}$, our lower bound matches [ABPSS25, Lemma 5.1.6].

For this, we will need the following notation. For integers $d \geqslant 0$ and $s \geqslant 2$ and $(n_i)_{i\in\mathcal{S}}$, let $n = \sum_{i\in\mathcal{S}} n_i$, $\mathbf{n} = (n_i)_{i\in\mathcal{S}} \in \mathcal{S}^n$. Let the multi-slice $\mathcal{S}^n_{\mathbf{n}} \subseteq \mathcal{S}^n$ denote the set of points which contain n_{i+1} many occurrences of the element i for all $i \in \mathcal{S}$. Let $\binom{n}{\mathbf{n}} = \binom{n}{n_0, n_1, \ldots, n_{s-1}}$ denote the size of $\mathcal{S}^n_{\mathbf{n}}$ (so it is zero if some n_i is negative). We also use the notation $\mathbf{n} - d$ to denote the tuple $((n_i - d))_{i\in\mathcal{S}}$.

Theorem 5.10 (Sub-optimal distance lemma for junta sums on multi-slices)**.** *For every $\mathbf{n} = (n_i)_{i\in\mathcal{S}}$ with $\sum_{i\in\mathcal{S}} n_i = n$, the following holds. If a junta-sum $P \in \mathcal{J}_d(\mathcal{S}^n, G)$ is non-zero on the multi-slice $\mathcal{S}^n_{\mathbf{n}}$ i.e. there exists a point $\mathbf{a} \in \mathcal{S}^n_{\mathbf{n}}$ such that $P(\mathbf{a}) \neq 0$, then*

$$|\{\mathbf{a} \in \mathcal{S}^n_{\mathbf{n}} \mid P(\mathbf{a}) \neq 0\}| \geqslant \binom{n - sd}{\mathbf{n} - d}.$$

Proof of Theorem 5.10. The proof of the theorem is by induction on d. The base case $d = 0$ is handled by noting that P is a constant function in this case. Now suppose $d \geqslant 1$. We shall assume that $n \geqslant sd + 1$ and $n_i \geqslant d$ for all $i \in \mathbb{Z}_s$, as the theorem statement is trivial otherwise.

We will assume that P is not a constant function over $\binom{[n]}{\mathbf{n}}$ as otherwise we are done. In particular, we can always find two points $\mathbf{a}, \mathbf{b} \in \binom{[n]}{\mathbf{n}}$ such that $P(\mathbf{a}) \neq P(\mathbf{b})$ and they differ in exactly two coordinates; this follows by noting that we can move from any point on the multislice to any other point by swapping elements a finite number of times. Without loss of generality, we can assume that $\mathbf{a}$ and $\mathbf{b}$ differ on the first and last coordinates; i.e., $a_1 = b_n = \alpha$ and $a_n = b_1 = \beta$ for some $\alpha \neq \beta \in \mathbb{Z}_s$. Let $\mathbf{n}' = (n'_i)_{i \in \mathbb{Z}_s}$ be defined by $n'_i = n_i$ for $i \notin \{\alpha, \beta\}$ and $n'_i = n_i - 1$ for $i \in \{\alpha, \beta\}$. We now consider the function $Q : \mathcal{S}^{n-2} \to G$ defined as:

$$Q(x_2, \ldots, x_{n-1}) = P(\alpha, x_2, \ldots, x_{n-1}, \beta) - P(\beta, x_2, \ldots, x_{n-1}, \alpha).$$

As $P(\mathbf{a}) - P(\mathbf{b}) \neq 0$, we see that Q is not identically zero over $\binom{[n-2]}{\mathbf{n}'}$. We also claim that Q is a $(d-1)$-junta-sum. Indeed, if

$$P(x_1, \ldots, x_n) = \sum_{\mathbf{c} \in \mathbb{Z}_s^n : |\mathbf{c}| \leqslant d} g_{\mathbf{c}} \cdot \prod_{i \in [n] : c_i \neq 0} \delta_{c_i}(x_i),$$

then in the junta-polynomial of Q, all the monomials that do not contain either x_1 or x_n will be canceled, while the monomials of degree d that contain either x_1 or x_n (or both) will reduce in degree. Hence, $Q \in \mathcal{J}_{d-1}(\mathcal{S}^{n-2}, d-1)$. Now, by induction hypothesis, we have that there are at least $\binom{n-2-s(d-1)}{\mathbf{n}'-(d-1)}$ choices for $\mathbf{d} \in \binom{[n-2]}{\mathbf{n}'}$ such that $Q(\mathbf{d}) \neq 0$. For each such $\mathbf{d}$, we have that $Q(\mathbf{d}) = P(\alpha, \mathbf{d}, \beta) - P(\beta, \mathbf{d}, \alpha) \neq 0$ so either $\mathbf{a}' = (\alpha, \mathbf{d}, \beta)$ or $\mathbf{b}' = (\beta, \mathbf{d}, \alpha)$ is a non-zero of P. Furthermore, we can verify that $\mathbf{a}', \mathbf{b}' \in \binom{[n]}{\mathbf{n}'}$. Let $\mathbf{e}$ denote the tuple which is 1 at all indices $i \notin \{\alpha, \beta\}$ and is 0 for $i \in \{\alpha, \beta\}$. Hence, the number of non-zeroes of P over $\binom{[n]}{\mathbf{n}}$ is at least the number of such $\mathbf{d}$ which is at least

$$\binom{n-2-s(d-1)}{\mathbf{n}'-(d-1)} = \binom{n-sd+(s-2)}{\mathbf{n}'-(d-1)} = \binom{n-sd+(s-2)}{(\mathbf{n}-d)+\mathbf{e}} \geqslant \binom{n-sd}{\mathbf{n}-d} \cdot \binom{s-2}{\mathbf{e}} \geqslant \binom{n-sd}{\mathbf{n}-d}.$$

■

Using Theorem 5.10, we immediately get the following corollary, which gives a lower bound on the fraction of non-zeroes on the balanced multislice.

Corollary 5.11. *Let $n, s, d \in \mathbb{N}$ with $n \geqslant sd$ divisible by s. Let $\mu = (n/s, \ldots, n/s)$. If a junta-sum $P \in \mathcal{J}(\mathcal{S}^n, d, G)$ is non-zero over $\mathcal{S}^n_\mu$, i.e. there exists $\mathbf{a} \in \mathcal{S}^n_\mu$ such that $P(\mathbf{a}) \neq 0$, then:*

$$\Pr_{\mathbf{x} \sim \mathcal{S}^n_\mu}[P(\mathbf{x}) \neq 0] \;\geqslant\; \frac{1}{(sd)^{sd}}.$$

Proof of Corollary 5.11. Let $n = ms$ for some $m \in \mathbb{Z}$. By Theorem 5.10, we have that the probability of a random point in $\mathcal{S}^n_\mu$ being non-zero for P is at least:

$$\frac{\binom{n-sd}{\mathbf{n}-d}}{\binom{n}{\mathbf{n}}} = \frac{(n-sd)!}{(m-d)!^s} \cdot \frac{m!^s}{n!}$$

$$
\begin{aligned}
&= \frac{m(m-1)\dots(m-d+1)}{n(n-1)\dots(n-d+1)} \cdots \frac{m(m-1)\dots(m-d+1)}{(n-(s-1)d)(n-(s-1)d-1)\dots(n-sd+1)} \\
&= \left(\frac{m}{n} \cdot \frac{m-1}{n-1} \cdot \dots \cdot \frac{m-d+1}{n-d+1}\right) \cdots \left(\frac{m}{n-(s-1)d} \cdot \frac{m-1}{n-(s-1)d-1} \cdot \dots \cdot \frac{m-d+1}{n-sd+1}\right) \\
&\geqslant \left(\frac{m-d+1}{n-d+1}\right)^d \cdots \left(\frac{m-d+1}{n-sd+1}\right)^d && \text{(using } \tfrac{a}{b} \geqslant \tfrac{a-i}{b-i} \text{ for } 0 < i < a < b\text{)} \\
&\geqslant \left(\frac{n-sd+s}{s(n-d+1)}\right)^{sd} \\
&\geqslant \frac{1}{(sd)^{sd}}. && \text{(using } \tfrac{n-sd+s}{n-d+1} \geqslant \tfrac{1}{d} \text{ since } n \geqslant sd-1\text{)}
\end{aligned}
$$

■

Using Lemma 5.7 and Corollary 5.11, we get the following corollary.

Corollary 5.12. *There exists an absolute constant $\eta > 0$ for which the following holds. Let $R \in \mathcal{J}_d(\mathcal{S}^{s^2k}, G)$ be a non-zero function and there exists a $\mathbf{w} \in \mathcal{S}^{s^2k}_{sk,\dots,sk}$ such that $R(\mathbf{w}) \neq 0$. Let $\tau : [s^2k] \to [sk]$ be a random s-to-1 map and C_τ be the subgrid as defined before. Then,*

$$\Pr_\tau[R|_{\mathsf{C}_\tau} \text{ vanishes on } \mathcal{S}^{sk}_{k,\dots,k}] \leqslant \frac{1}{k^\eta}.$$

Proof of Corollary 5.12. Let S denote the set of non-zeroes of R on the slice $\mathcal{S}^{s^2k}_{sk,\dots,sk}$, i.e. $S = \left\{\mathbf{a} \in \mathcal{S}^{s^2k}_{sk,\dots,sk} \;\middle|\; R(\mathbf{a}) \neq 0\right\}$. From Corollary 5.11, we know that

$$|S| \geqslant (1/(sd)^{sd}) \cdot |\mathcal{S}^{s^2k}_{sk,\dots,sk}| \;\Rightarrow\; \frac{|S|}{|\mathcal{S}^{s^2k}_{sk,\dots,sk}|} = \Omega(1).$$

$R|_{\mathsf{C}_\tau}$ does not vanish on $\mathcal{S}^{sk}_{k,\dots,k}$ if $S \cap \mathsf{C}_\tau \neq \varnothing$. Using Lemma 5.7, we know that the probability of $S \cap \mathsf{C}_\tau = \varnothing$ (over the randomness in choice of τ) is at most $1/k^\eta$. ■

5.3 Subroutine for Approximating Oracles

Definition 5.13 (Subgrid containing $\mathbf{b}$)**.** *Let $\mathsf{C} = C_{h,\Pi}$ be a k-dimensional subgrid of $\mathcal{S}^n$ as defined in Definition 2.9, where $h : [n] \to [k]$ is a hash function and $\Pi \in (\mathrm{Sym}[\mathcal{S}])^n$ is a tuple of permutations. For an arbitrary $\mathbf{b} \in \mathcal{S}^n$ and a permutation $\sigma \in \mathrm{Sym}_{sk}$ define a new hash function $h' : [n] \to [sk]$ as follows:*

$$h'(i) = \sigma(h(i) + k \cdot b_i), \qquad \text{for all } i \in [n]$$

For every $\mathbf{z} \in \mathcal{S}^{sk}$, define $x_{h',\Pi}(\mathbf{z}) := \Pi_i(z_{h'(i)})$. Define the subset $\mathsf{C}^{\mathbf{b}}_\sigma \subset \mathcal{S}^n$ as follows

$$\mathsf{C}^{\mathbf{b}}_\sigma := \left\{x_{h,\Pi}(\mathbf{z}) \;\middle|\; \mathbf{z} \in \mathcal{S}^{sk}\right\}.$$

We make a few observations from Definition 5.13. The first observation is that $\mathbf{b}$ is indeed in $\mathsf{C}^{\mathbf{b}}_{\sigma}$. The second observation is that for random $h, \Pi, \mathbf{b}$ and σ, the subgrid $\mathsf{C}^{\mathbf{b}}_{\sigma}$ is a random embedding of a sk-dimensional subgrid. The third observation is that C is a subgrid of $\mathsf{C}^{\mathbf{b}}_{\sigma}$ and is obtained by "randomly pairing" coordinates.

Observation 5.14. *The point $\mathbf{b} \in \mathcal{S}^n$ lies inside the subgrid $\mathsf{C}^{\mathbf{b}}_{\sigma}$, i.e. there exists a string $\mathbf{w} \in \mathcal{S}^{sk}_{k,\dots,k}$ such that $x_{h',\Pi}(\mathbf{w}) = \mathbf{b}$. More explicitly,*

$$w_{h(i)+k\cdot b_i} := \Pi_i^{-1}(b_i), \qquad \text{for all } i \in [n].$$

Also it is easy to see that the partition of $[n]$ induced by h' (as defined in Definition 5.13) is a refinement of the partition induced by h. This means $\mathsf{C} \subset \mathsf{C}^{\mathbf{b}}_{\sigma}$.

Observation 5.15. *Let h, Π, and $\mathbf{b}$ (as stated in Definition 5.13) be randomly chosen. Then $\mathsf{C}^{\mathbf{b}}$ is a random embedding of a sk-dimensional subgrid, i.e. there exists a random hash function $H : [n] \to [sk]$ and a random $\Pi' \in (\mathrm{Sym}[\mathcal{S}])^n$ such that $\mathsf{C}^{\mathbf{b}}_{\sigma}$ has the same distribution as $C_{H,\Pi'}$.*

Observation 5.16. *Let h, Π, and $\mathbf{b}$ (as stated in Definition 5.13) be randomly chosen. Conditioned on the grid $\mathsf{C}^{\mathbf{b}}$, the subgrid C has the following distribution:*
Sample a random s-to-1 map[13] $\tau : [sk] \to [k]$ and we identify s variables together.

5.4 The Algorithm

In this subsection, we give the description of the algorithms to prove Theorem 5.5. The algorithm proceeds in two steps, and this is similar to the algorithms in [ABPSS25, Section 5.2.2], barring a few changes to handle larger grids $\mathcal{S}$. We request the reader to refer to [ABPSS25, Section 5.2.2] for an overview and discussion on the algorithms.

In the following description, let $L(\varepsilon) = |\mathsf{List}_{\varepsilon}(f)|$, where recall that $\mathsf{List}_{\varepsilon}(f)$ is the set of d-junta-sums that are $(1/s^d - \varepsilon)$-close to f. Note that Algorithm 2 is a deterministic algorithm and all the randomness is in Algorithm 4,

[13] A map is s-to-1 if the pre-image of every element under the map has size exactly s, i.e., exactly s elements from the domain have the same image.

Algorithm 2: Approximating Algorithm $\Psi[\mathsf{C}, \sigma, Q]$

Input: Oracle access to the function f, a point $\mathbf{b} \in \mathcal{S}^n$

Let C' be a subgrid spanned by C and $\mathbf{b}$ using $\sigma \in \mathrm{Sym}_{sk}$ // see Definition 5.13

Let $\mathbf{w} \in \mathcal{S}^{sk}$ such that $x(\mathbf{w}) \in \mathsf{C}'$ and $x(\mathbf{w}) = \mathbf{b}$ // see Observation 5.14, $|\mathbf{w}| \in \mathcal{S}^{sk}_{k,\dots,k}$

Query f on the subgrid C' // Number of queries is s^{sk}

Find all degree-d junta-sums $R_1, \dots, R_{L''} \in \mathcal{J}_d(\mathcal{S}^{sk}, G)$ that are $\left(\frac{1}{s^d} - \frac{\varepsilon}{2}\right)$-close to $f|_{\mathsf{C}'}$

if *there exists an* $i \in [L'']$ *such that* $R_i|_{\mathsf{C}} = Q$ **then**

  pick any such i and **return** $R_i(\mathbf{w})$

else

  return 0 // An arbitrary value

Now we describe the randomized Algorithm 3 that returns the descriptions of the deterministic oracles.

Algorithm 3: Algorithm $\mathcal{A}_1$

Input: Oracle access to the function f

Choose $k \leftarrow B_d \left(\frac{L(\varepsilon/2)}{\varepsilon}\right)^c$ // B_d and c are constants, chosen later in the analysis

Set $\ell \leftarrow \log L(\varepsilon)$

$T \leftarrow \varnothing$

repeat

  Sample $\Pi \in (\mathrm{Sym}[\mathcal{S}])^n$ and a random hash function $h : [n] \to [k]$ // the first source of randomness

  Construct the subgrid $\mathsf{C} := C_{h,\Pi}$ // see Definition 2.9

  Query f on the subgrid C // Number of queries is 2^k

  Find all junta-sums $Q_1, \dots, Q_{L'} \in \mathcal{J}_d(\mathcal{S}^k, G)$ that are $\left(\frac{1}{s^d} - \frac{\varepsilon}{2}\right)$-close to $f|_{\mathsf{C}}$

  Pick a uniformly random permutation $\sigma \sim \mathrm{Sym}_{sk}$ // the second source of randomness

  $T \leftarrow T \cup \{(\mathsf{C}, \sigma, Q_1), \dots, (\mathsf{C}, \sigma, Q_{L'})\}$

until ℓ *times*

return $\Psi[\mathsf{C}, \sigma, Q]$ *for all* $(\mathsf{C}, \sigma, Q) \in T$ // Size of T is $\leqslant \ell L'$

5.5 Analysis of the Local List Corrector

In this subsection, we analyze Algorithm 4 and Algorithm 2 to prove Theorem 5.5. We recall the statement of Theorem 5.5.

Theorem 5.5 (Approximate oracles)**.** *Fix $n \in \mathbb{N}$, $\varepsilon > 0$. Let $f : \mathcal{S}^n \to G$ be any function and $L(\varepsilon) := |\mathsf{List}_\varepsilon(f)|$. There exists a randomized algorithm $\mathcal{A}_1^f$ that makes at most $\mathcal{O}_\varepsilon(1)$ oracle queries and outputs deterministic algorithms $\Psi_1, \ldots, \Psi_{L'}$ satisfying the following property:*
With probability at least $3/4$, for every junta-sum $P \in \mathsf{List}_\varepsilon^f$, there exists a $j \in [L']$ such that

1. *$\delta(\Psi_j, P) < 1/(10 \cdot 2^{d+1})$*
2. *For every $\mathbf{x} \in \mathcal{S}^n$, Ψ_j computes $P(\mathbf{x})$ by making at most $\mathcal{O}_\varepsilon(1)$ oracle queries to f.*

Here $L' = \mathcal{O}(L(\varepsilon/2) \log L(\varepsilon)) = \mathcal{O}_\varepsilon(1)$.

We start by show that in a single iteration of Algorithm 3, for every junta-sum $P \in \mathrm{List}_\varepsilon^f$, with probability at least $\geqslant 99/100$, there exists an approximating oracle $\Psi[\mathsf{C}, \sigma, Q]$ such that $\delta(P, \Psi[\mathsf{C}, \sigma, \mathsf{Q}])$ is at most $\leqslant 1/(10 \cdot s^{d+1})$.

Lemma 5.17 (Error w.r.t a fixed junta-sum in one iteration)**.** *Fix a junta-sum $P \in \mathrm{List}_\varepsilon^f$. Then for every iteration of Algorithm 3, the following holds:*
With probability $\geqslant 99/100$, over the randomness of Algorithm 3, there exists a tuple (C, σ, Q) such that

$$\delta(P, \Psi[\mathsf{C}, \sigma, Q]) \leqslant \frac{1}{10 \cdot s^{d+1}}.$$

Proof of Lemma 5.17. Fix a particular iteration of the main loop of Algorithm 3. In this iteration, there are three sources of errors:

1. Event $\mathcal{E}_{1,P}$ (depends on Π and h): There does not exist a junta-sum $Q \in \mathcal{J}_d(\mathcal{S}^k, G)$ such that $Q \equiv P|_{\mathsf{C}}$.
2. Event $\mathcal{E}_{2,P}$ (depends on $\Pi, h, \sigma, \mathbf{b}$): Consider a tuple $(\mathsf{C}, \sigma, Q_i) \in T$ added in this iteration. For the approximating algorithm $\Psi[\mathsf{C}, \sigma, Q_i]$ (Algorithm 2), there does not exist a junta-sum $R \in \mathcal{J}_d(\mathcal{S}^{sk}, G)$ such that $R \equiv P|_{\mathsf{C}'}$. Observe that this event is independent of Q_i and only depends on $\mathsf{C}, \mathbf{b}$, and σ.
3. Event $\mathcal{E}_{3,P}$ (depends on $\Pi, h, \sigma, \mathbf{b}$): Consider a tuple $(\mathsf{C}, \sigma, Q_i) \in T$ added in this iteration. For the approximating algorithm $\Psi[\mathsf{C}, \sigma, Q_i]$ (Algorithm 2), there exists two distinct junta-sums $R_i, R_j \in \mathcal{J}_d(\mathcal{S}^{sk}, G)$ such that $R_i|_{\mathsf{C}} \equiv R_j|_{\mathsf{C}}$ but $R_i(\mathbf{w}) \neq R_j(\mathbf{w})$. In this situation, Line 6 of Algorithm 2 is not a well-defined instruction. This event also only depends on $\mathsf{C}, \mathbf{b}$, and σ.

The probability of $\mathcal{E}_{1,P}$ and $\mathcal{E}_{2,P}$ can be upper bounded by using Lemma 2.10 on C and C' respectively. To upper bound, we use Corollary 5.12.

Claim 5.18 (Probabilities of the first two error events)**.** *Let $\mathcal{E}_{1,P}$ and $\mathcal{E}_{2,P}$ be as defined above. Then,*

$$\Pr_{\Pi,h}[\mathcal{E}_{1,P}] \leqslant \frac{1}{10000 \cdot s^{d+1}} \quad \text{and} \quad \Pr_{\Pi,h,\sigma,\mathbf{b}]}[\mathcal{E}_{2,P}] \leqslant \frac{1}{10000 \cdot s^{d+1}}.$$

Proof. Let us start with $\mathcal{E}_{1,P}$. Non-existence of a $Q \in \mathcal{J}_d(\mathcal{S}^k, G)$ such that $Q \equiv P|_{\mathsf{C}}$ is equivalent to $\delta(P|_{\mathsf{C}}, f|_{\mathsf{C}}) > (1/s^d - \varepsilon/2)$. Using Lemma 2.10, we get the desired bound.

For $\mathcal{E}_{2,P}$, we use Observation 5.15 and then proceed as in the case of $\mathcal{E}_{1,P}$. This finishes the proof of the claim. ■

The next claim is to upper bound the probability of the third error. Upper bounding this error uses the spectral expansion and is very different from the Boolean setting as in [ABPSS25].

Claim 5.19 (Probability of the third error event)**.** *Let $\mathcal{E}_{3,P}$ be as defined above. Then,*

$$\Pr_{\Pi,h,\sigma,\mathbf{b}}[\mathcal{E}_{3,P}] \leqslant \frac{1}{10000 \cdot s^{d+1}}.$$

Proof. Fix a subgrid C'. This fixes the junta sums $R_1, \ldots, R_{L''}$ in Line 4 of Algorithm 2. Consider any two distinct junta sums R_i and R_j such that they differ on at least one point in $\mathcal{S}^{sk}_{k,\ldots,k}$ (this includes the pairs which differ on $\mathbf{w}$). This means $R := R_i - R_j$ is non-zero on $\mathcal{S}^{sk}_{k,\ldots,k}$. We want to upper bound the probability that $R_i|_{\mathsf{C}} \equiv R_j|_{\mathsf{C}}$ i.e. $R|_{\mathsf{C}} \equiv 0$.

Using Observation 5.16 and Corollary 5.12, for appropriately chosen constants B_d and c, the probability of $R|_{\mathsf{C}}$ vanishing is $\leqslant 1/(10000 \cdot s^{d+1} \cdot L(\varepsilon/2)^2)$. We know that $L'' \leqslant L(\varepsilon/2)$. Doing an union bound on all possible pairs (R_i, R_j), we get the error probability is $\leqslant 1/(10000 \cdot s^{d+1})$. This finishes the proof of the claim. ■

Combining the above three claims to bound the final error probability is analogous to the proof in [ABPSS25, Lemma 5.3.1]. As the proof is quite similar, we skip it here.
This finishes the proof of Lemma 5.17. ■

The above lemma shows that for a fixed $P \in \mathsf{List}_\varepsilon(f)$, the algorithm returns an approximating oracle with high probability in a single iteration. We now use it to finish the proof of Theorem 5.5.

Proof of Theorem 5.5. We first show the correctness of Algorithm 4. Fix any $P \in \mathsf{List}_\varepsilon(f)$. From Lemma 5.17, we know that Algorithm 4 returns a tuple (C, σ, Q) for which $\Psi[\mathsf{C}, \sigma, Q]$ is $\leqslant 1/(10 \cdot s^{d+1})$-close with probability $\geqslant 0.99$. Algorithm 4 has $\ell = \log L(\varepsilon)$ many independent iterations. Thus at the end of ℓ iterations, the probability of the event that there is no tuple (C, σ, Q) added in T such that $\Psi[\mathsf{C}, \sigma, Q]$ is $\leqslant 1/(10 \cdot s^{d+1})$-close to P is $\leqslant 1/100^\ell$. By a union bound over all $P \in \mathsf{List}_\varepsilon(f)$, we get the desired correctness probability.

In Line 8 of Algorithm 4, $L' \leqslant L(\varepsilon/2)$. So in each iteration of Algorithm 4, at most $L(\varepsilon/2)$ tuples are added in T. Thus over ℓ iterations, at most $\mathcal{O}(L(\varepsilon/2) \log L(\varepsilon))$ tuples are added.

It remains to argue about the query complexity. In a single iteration of Algorithm 4, we make $s^k = s^{B_d(L(\varepsilon/2)/\varepsilon)^c}$ queries to f. There are $\ell = \log L(\varepsilon)$ iterations. From Theorem 5.4, we know that $L(\varepsilon/2) = \mathcal{O}_\varepsilon(1)$. Thus Algorithm 4 outputs the deterministic algorithms $\Psi_1, \ldots, \Psi_{L'}$ by making $\mathcal{O}_\varepsilon(1)$ queries to f.
For each deterministic algorithm $\Psi[\mathsf{C}, \sigma, Q]$, Algorithm 2 makes $s^{sk} = s^{sB_d(L(\varepsilon/2)/\varepsilon)^c}$ queries to f. From Theorem 5.4, we know that $L(\varepsilon) = \mathcal{O}_\varepsilon(1)$. Thus each Ψ_j makes $\mathcal{O}_\varepsilon(1)$ queries to f. This

shows the claimed query complexity.
This finishes the proof of Theorem 5.5. ■

Acknowledgments

We would like to thank the anonymous reviewers of RANDOM 2025 for many valuable comments, including pointers to crucial papers in the literature (specifically [DFLLV21]), that significantly improved some of our proofs.

References

[ABPSS24] Prashanth Amireddy, Amik Raj Behera, Manaswi Paraashar, Srikanth Srinivasan, and Madhu Sudan. "Local Correction of Linear Functions over the Boolean Cube". In: *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*. STOC 2024. Vancouver, BC, Canada: Association for Computing Machinery, 2024, pp. 764–775. ISBN: 9798400703836. DOI: 10.1145/3618260.3649746. URL: https://doi.org/10.1145/3618260.3649746 (cit. on pp. 45, 64).

[ABPSS25] Prashanth Amireddy, Amik Raj Behera, Manaswi Paraashar, Srikanth Srinivasan, and Madhu Sudan. "Low Degree Local Correction Over the Boolean Cube". In: *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2025, pp. 5504–5511. DOI: 10.1137/1.9781611978322.187. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611978322.187. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611978322.187 (cit. on pp. 7, 9, 13, 14, 44, 45, 49, 52, 55, 62, 63, 73, 75, 77, 80, 81, 83, 84, 86, 87, 89, 92).

[ABSS25] Prashanth Amireddy, Amik Raj Behera, Srikanth Srinivasan, and Madhu Sudan. "A Near-Optimal Polynomial Distance Lemma over Boolean Slices". In: *52nd International Colloquium on Automata, Languages, and Programming (ICALP 2025)*. Ed. by Keren Censor-Hillel, Fabrizio Grandoni, Joël Ouaknine, and Gabriele Puppis. Vol. 334. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2025, 11:1–11:17. ISBN: 978-3-95977-372-0. DOI: 10.4230/LIPIcs.ICALP.2025.11. URL: https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ICALP.2025.11 (cit. on pp. 7, 8, 10, 13).

[ASS23] Prashanth Amireddy, Srikanth Srinivasan, and Madhu Sudan. "Low-Degree Testing over Grids". In: *Approximation, Randomization, and Combinatorial Optimization (RANDOM)*. Vol. 275. 2023, 41:1–41:22. DOI: 10.4230/LIPICS.APPROX/RANDOM.2023.41. URL: https://doi.org/10.4230/LIPIcs.APPROX/RANDOM.2023.41 (cit. on pp. 7, 17, 18).

[BL18] Abhishek Bhowmick and Shachar Lovett. "The List Decoding Radius for Reed-Muller Codes Over Small Fields". In: *IEEE Trans. Inf. Theory* 64.6 (2018), pp. 4382–4391. DOI: 10.1109/TIT.2018.2822686. URL: https://doi.org/10.1109/TIT.2018.2822686 (cit. on p. 9).

[BP21] Andrej Bogdanov and Gautam Prakriya. "Direct Sum and Partitionability Testing over General Groups". In: *International Colloquium on Automata, Languages, and Programming, (ICALP)*. Vol. 198. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021, 33:1–33:19. DOI: 10.4230/LIPIcs.ICALP.2021.33 (cit. on p. 7).

[DFLLV21] Neta Dafni, Yuval Filmus, Noam Lifshitz, Nathan Lindzey, and Marc Vinyals. "Complexity Measures on the Symmetric Group and Beyond". In: *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*. Ed. by James R. Lee. Vol. 185. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021, 87:1–87:5. ISBN: 978-3-95977-177-1. DOI: 10.4230/LIPIcs.ITCS.2021.87. URL: https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2021.87 (cit. on pp. 12, 24, 29, 34, 56).

[Del78] Ph. Delsarte. "Hahn Polynomials, Discrete Harmonics, and t-Designs". In: *SIAM Journal on Applied Mathematics* 34.1 (1978), pp. 157–166. ISSN: 00361399. URL: http://www.jstor.org/stable/2100864 (visited on 06/04/2024) (cit. on pp. 4, 10).

[DL78] Richard A. DeMillo and Richard J. Lipton. "A probabilistic remark on algebraic program testing". In: *Information Processing Letters* 7.4 (1978), pp. 193–195. DOI: 10.1016/0020-0190(78)90067-4 (cit. on p. 7).

[DS87] Persi Diaconis and Mehrdad Shahshahani. "Time to Reach Stationarity in the Bernoulli–Laplace Diffusion Model". In: *SIAM Journal on Mathematical Analysis* 18.1 (1987), pp. 208–218. DOI: 10.1137/0518016. eprint: https://doi.org/10.1137/0518016. URL: https://doi.org/10.1137/0518016 (cit. on p. 4).

[DG19] Irit Dinur and Konstantin Golubev. "Direct Sum Testing: The General Case". In: *Approximation, Randomization, and Combinatorial Optimization (RANDOM)*. Vol. 145. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019, 40:1–40:11. DOI: 10.4230/LIPIcs.APPROX-RANDOM.2019.40 (cit. on p. 7).

[EFP11] David Ellis, Ehud Friedgut, and Haran Pilpel. "Intersecting Families of Permutations". In: *Journal of the American Mathematical Society* 24.3 (Jan. 2011), pp. 649–682. DOI: 10.1090/s0894-0347-2011-00690-5 (cit. on p. 33).

[ER60] Paul Erdös and Richard Rado. "Intersection theorems for systems of sets". In: *Journal of the London Mathematical Society* 1.1 (1960), pp. 85–90 (cit. on pp. 82, 90).

[Fil16] Yuval Filmus. "An Orthogonal Basis for Functions over a Slice of the Boolean Hypercube". In: *The Electronic Journal of Combinatorics* 23 (2016), p. 1. DOI: https://doi.org/10.37236/4567 (cit. on p. 10).

[Fil23] Yuval Filmus. "Junta Threshold for Low Degree Boolean Functions on the Slice". In: *The Electronic Journal of Combinatorics* 30 (2023). DOI: https://doi.org/10.37236/11115 (cit. on p. 10).

[FOW22] Yuval Filmus, Ryan O'Donnell, and Xinyu Wu. "Log-Sobolev inequality for the multislice, with applications". In: *Electron. J. Probab.* 27 (2022), Paper No. 33, 30. ISSN: 1083-6489. DOI: 10.1214/22-ejp749. URL: https://doi.org/10.1214/22-ejp749 (cit. on p. 4).

[GL89] Oded Goldreich and Leonid A. Levin. "A Hard-Core Predicate for all One-Way Functions". In: *ACM Symposium on Theory of Computing (STOC)*. 1989, pp. 25–32. DOI: 10.1145/73007.73010. URL: https://doi.org/10.1145/73007.73010 (cit. on p. 9).

[GKZ08] Parikshit Gopalan, Adam R. Klivans, and David Zuckerman. "List-Decoding Reed-Muller Codes over Small Fields". In: *ACM Symposium on Theory of Computing (STOC)*. 2008, pp. 265–274. DOI: 10.1145/1374376.1374417. URL: https://doi.org/10.1145/1374376.1374417 (cit. on p. 9).

[GRS23] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. *Essential Coding Theory (Book draft)*. 2023. URL: http://www.cse.buffalo.edu/atri/courses/coding-theory/book (cit. on p. 39).

[Hoe94] Wassily Hoeffding. "Probability inequalities for sums of bounded random variables". In: *The collected works of Wassily Hoeffding* (1994), pp. 409–426 (cit. on p. 48).

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. "Expander Graphs and their Applications". In: *Bulletin of the American Mathematical Society* 43 (2006), pp. 439–561 (cit. on p. 41).

[KKS24] Swastik Kopparty, Mrinal Kumar, and Harry Sha. "High Rate Multivariate Polynomial Evaluation Codes". In: *CoRR* abs/2410.13470 (2024). DOI: 10.48550/ARXIV.2410.13470. arXiv: 2410.13470. URL: https://doi.org/10.48550/arXiv.2410.13470 (cit. on p. 8).

[Kum52] Ernst Eduard Kummer. "Über die Ergänzungssätze zu den allgemeinen Reciprocitätsgesetzen." In: (1852) (cit. on p. 77).

[MNV16] Raghu Meka, Oanh Nguyen, and Van Vu. "Anti-concentration for Polynomials of Independent Random Variables". In: *Theory of Computing* 12.11 (2016), pp. 1–17. DOI:

10.4086/toc.2016.v012a011. URL: https://theoryofcomputing.org/articles/v012a011 (cit. on pp. 83, 84).

[ODo14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. DOI: 10.1017/CBO9781139814782 (cit. on p. 61).

[Ore22] Øystein Ore. "Über höhere kongruenzen". In: *Norsk Mat. Forenings Skrifter* 1.7 (1922), p. 15 (cit. on p. 7).

[PS97] Alessandro Panconesi and Aravind Srinivasan. "Randomized Distributed Edge Coloring via an Extension of the Chernoff–Hoeffding Bounds". In: *SIAM Journal on Computing* 26.2 (1997), pp. 350–368. DOI: 10.1137/S0097539793250767. eprint: https://doi.org/10.1137/S0097539793250767. URL: https://doi.org/10.1137/S0097539793250767 (cit. on p. 92).

[Sag13] Bruce E. Sagan. *The Symmetric Group - Representations, Combinatorial Algorithms, and Symmetric Functions*. Graduate Texts in Mathematics. Springer New York, NY, 2013. ISBN: 978-1-4757-6804-6. DOI: https://doi.org/10.1007/978-1-4757-6804-6 (cit. on pp. 18, 21, 23, 29, 33, 34, 60).

[Sca97] Fabio Scarabotti. "Time to reach stationarity in the Bernoulli-Laplace diffusion model with many urns". In: *Adv. in Appl. Math.* 18.3 (1997), pp. 351–371. ISSN: 0196-8858,1090-2074. DOI: 10.1006/aama.1996.0514. URL: https://doi.org/10.1006/aama.1996.0514 (cit. on p. 4).

[Sch80] Jacob T. Schwartz. "Fast Probabilistic Algorithms for Verification of Polynomial Identities". In: *J. ACM* 27.4 (1980), pp. 701–717. DOI: 10.1145/322217.322225 (cit. on p. 7).

[Sri11] Murali K. Srinivasan. "Symmetric chains, Gelfand–Tsetlin chains, and the Terwilliger algebra of the binary Hamming scheme". In: *Journal of Algebraic Combinatorics* 34 (2011). DOI: https://doi.org/10.1007/s10801-010-0272-2 (cit. on p. 10).

[Sta12] Richard P. Stanley. *Enumerative combinatorics. Volume 1*. Second. Vol. 49. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2012, pp. xiv+626 (cit. on p. 18).

[Sta24] Richard P. Stanley. *Enumerative combinatorics. Vol. 2*. Second. Vol. 208. Cambridge Studies in Advanced Mathematics. With an appendix by Sergey Fomin. Cambridge University Press, Cambridge, 2024, pp. xvi+783 (cit. on p. 18).

[STV01] Madhu Sudan, Luca Trevisan, and Salil P. Vadhan. "Pseudorandom Generators without the XOR Lemma". In: *J. Comput. Syst. Sci.* 62.2 (2001), pp. 236–266. DOI: 10.1006/JCSS.2000.1730. URL: https://doi.org/10.1006/jcss.2000.1730 (cit. on p. 9).

[Vad12] Salil P. Vadhan. "Pseudorandomness". In: *Foundations and Trends® in Theoretical Computer Science* 7.1–3 (2012), pp. 1–336. ISSN: 1551-305X. DOI: 10.1561/0400000010. URL: http://dx.doi.org/10.1561/0400000010 (cit. on pp. 16, 74).

[Zip79] Richard Zippel. "Probabilistic algorithms for sparse polynomials". In: *Symbolic and Algebraic Computation*. Springer Berlin Heidelberg, 1979, pp. 216–226. DOI: 10.1007/3-540-09519-5_73 (cit. on p. 7).

A Tabloids, Polytabloids, Multislices, and Functions

For a tableau t, tabloid of t, denoted by $\{\mathbf{t}\}$ is an equivalence class of tableaux (of the same shape) under the row equivalence relation. See [Sag13, Definition 2.1.4] for a formal definition. For a partition $\lambda \in \mathcal{P}(n)$, Tabloids($\lambda$) is a set of tabloids of shape λ. The symmetric group Sym_n acts naturally on tabloids as follows: For a permutation $\pi \in \mathrm{Sym}_n$, π acts on a $\{T\} \in \mathrm{Tabloids}(\lambda)$ by permuting the entries of $\{T\}$. For example if $\pi = (125)(46) \in S_6$, then

$$(125)(46)\ \begin{array}{|c|c|c|}\hline 1&2&3\\\hline 4&5\\\cline{1-2} 6\\\cline{1-1}\end{array}\ =\ \begin{array}{|c|c|c|}\hline 2&5&3\\\hline 6&1\\\cline{1-2} 4\\\cline{1-1}\end{array}$$

Tabloids and multislice In the remaining section, we will always use λ to denote a partition such that $\lambda \trianglerighteq \mu$, where $\mu = (n/s, \ldots, n/s)$. Note that $\ell(\lambda) \leqslant s$. We will use the convention that λ has exactly s many parts, where we append a λ with fewer than s parts with 0's.

We now observe that Tabloids(λ) and $\mathcal{S}^n_\lambda$ are in bijection, as follows. For any tabloid $\{\mathbf{t}\} \in \mathrm{Tabloids}(\lambda)$, it corresponds to the point $\mathbf{a} \in \mathcal{S}^n_\lambda$ where,

$$a_j = i \quad \text{if } j \in (i+1)^{th} \text{ row of } \{\mathbf{t}\}, \qquad \text{for all } j \in [n].$$

Similarly, for any point $\mathbf{a} \in \mathcal{S}^n_\lambda$, we get a corresponding tabloid $\{\mathbf{t}\} \in \mathrm{Tabloids}(\lambda)$ where for every $j \in [n]$, the $(i+1)^{th}$ row of $\{\mathbf{t}\}$ contains j if $a_i = j$. In simple words, the entries in the $(i+1)^{th}$ row of $\{T\}$ correspond to the coordinates which are i. Following is an example for $n = 9$ and $\lambda = (4, 3, 2)$:

$$001210201 \quad \leftrightarrow \quad \begin{array}{|c|c|c|c|}\hline 1&2&3&4\\\hline 5&6&7\\\cline{1-3} 8&9\\\cline{1-2}\end{array}$$

For a tableau $\mathbf{t}$, a *polytabloid* for $\mathbf{t}$, denoted by $e_{\mathbf{t}}$ is a linear combination of tabloids obtained by permuting the columns of $\mathbf{t}$. See [Sag13, Definition 2.3.2] for a formal definition. Using the above bijection, it is easy to see that for every tableau $\mathbf{t}$, the associated polytabloid $e_{\mathbf{t}}$ is a function on $\mathcal{S}^n_\lambda$.

B Subgrid Sampling Lemma

Here we give the proof of the subgrid sampling lemma from Section 2.

Proof of Lemma 2.10. The proof is an application of the second moment method with a consequence of the following hypercontractivity theorem (Theorem B.1) being used to bound the variance.

Theorem B.1 ([ODo14, Section 10.3]). *Let $E \subseteq \mathbb{Z}_s^n$ be a subset of density δ, i.e. $|E|/s^n = \delta$. Let $q \geqslant 2$. Then for any $0 \leqslant |\rho| \leqslant (1/(q-1)) \cdot (1/s)^{1-2/q}$,*

$$\Pr_{\substack{\mathbf{x}\sim\mathbb{Z}_s^n \\ \mathbf{y}\sim\mathcal{N}_\rho(\mathbf{x})}} [\mathbf{x} \in E \text{ and } \mathbf{y} \in E] \leqslant \delta^{2-2/q}.$$

More formally, for each $\mathbf{y} \in \mathbb{Z}_s^k$, let $Z_{\mathbf{y}} \in \{0,1\}$ be the indicator random variable that is 1 exactly when $x(\mathbf{y}) \in T$. Let Z denote the sum of all $Z_{\mathbf{y}}$ ($\mathbf{y} \in \mathbb{Z}_s^k$). The statement of the lemma is equivalently stated as

$$\Pr\left[\left|Z - \mu \cdot s^k\right| \geqslant \varepsilon \cdot s^k\right] < \eta \tag{23}$$

for k as specified above.

Since each $x(\mathbf{y})$ is uniformly distributed over $\mathbb{Z}_s^n$, it follows that each $Z_{\mathbf{y}}$ is a Bernoulli random variable that is 1 with probability μ. In particular, the mean of Z is $\mu \cdot s^k$.

We now bound the variance of Z. Let I_γ be the interval $[\frac{(1-\gamma)(s-1)}{s}, \frac{(1+\gamma)(s-1)}{s}]$ where $\gamma \leqslant 1/(s-1)$. We have

$$\begin{aligned}
\mathrm{Var}(Z) &= \sum_{\mathbf{y},\mathbf{y}'} \mathrm{Cov}(Z_{\mathbf{y}}, Z_{\mathbf{y}'}) \\
&= \sum_{\mathbf{y},\mathbf{y}':\delta(\mathbf{y},\mathbf{y}')\in I_\gamma} \mathrm{Cov}(Z_{\mathbf{y}}, Z_{\mathbf{y}'}) + \sum_{\mathbf{y},\mathbf{y}':\delta(\mathbf{y},\mathbf{y}')\notin I_\gamma} \mathrm{Cov}(Z_{\mathbf{y}}, Z_{\mathbf{y}'}) \\
&\leqslant \sum_{\mathbf{y},\mathbf{y}':\delta(\mathbf{y},\mathbf{y}')\in I_\gamma} \mathrm{Cov}(Z_{\mathbf{y}}, Z_{\mathbf{y}'}) + \sum_{\mathbf{y},\mathbf{y}':\delta(\mathbf{y},\mathbf{y}')\notin I_\gamma} 1 \\
&\leqslant \sum_{\mathbf{y},\mathbf{y}':\delta(\mathbf{y},\mathbf{y}')\notin I_\gamma} \mathrm{Cov}(Z_{\mathbf{y}}, Z_{\mathbf{y}'}) + s^{2k} \cdot \exp(-\Omega(\gamma^2 \cdot (k(s-1)/s))).
\end{aligned} \tag{24}$$

where the final inequality is an application of the Chernoff bound. On the other hand, for any $\mathbf{y}, \mathbf{y}'$ such that $\delta(\mathbf{y},\mathbf{y}') \in I_\gamma$, we have seen above that the pair $(x(\mathbf{y}), x(\mathbf{y}'))$ have the same distribution as a pair of random variables $(\mathbf{z}, \mathbf{z}')$ where $\mathbf{z}$ is chosen uniformly at random from $\mathbb{Z}_s^n$ and $\mathbf{z}'$ is sampled from the distribution $\mathcal{N}_\rho(\mathbf{z})$, where $\rho = 1 - \frac{s\delta(\mathbf{y},\mathbf{y}')}{s-1} \in [-\gamma, \gamma]$. Thus $|\rho| \leqslant \gamma$.

Choose γ such that $\gamma \leqslant 1/(s-1)$ and

$$C_1\sqrt{\frac{s\log k}{(s-1)k}} \leqslant \gamma \leqslant \min\left\{\frac{1}{4}, \frac{1}{(k/\log k)^{1/4}} \cdot \frac{1}{s}\right\},$$

for a large enough constant C_1. Such a γ exists since $k \geqslant B \cdot s^4 \log s$ for a large constant B.

Set $q = (k \log k)^{1/4}$. From Theorem B.1, and since $\gamma \leqslant 1/4$, for $(\mathbf{y}, \mathbf{y}')$ satisfying $\delta(\mathbf{y}, \mathbf{y}') \in I_\gamma$ we have

$$\begin{aligned}\operatorname{Cov}(Z_{\mathbf{y}}, Z_{\mathbf{y}'}) &= \Pr[x(\mathbf{y}) \in T \text{ and } x(\mathbf{y}') \in T] - \mu^2 \\ &\leqslant \mu^{2-2/q} - \mu^2 \\ &\leqslant \min\{\mu^{1.5}, \mu^2 \cdot (\exp(O((1/q) \cdot \log(1/\mu)) - 1)\}.\end{aligned}$$

Plugging into Equation (24) we get the following inequalities:

$$\operatorname{Var}(Z) \leqslant s^{2k} \cdot \mu^{1.5} + s^{2k} \cdot \frac{1}{k} \leqslant s^{2k} \cdot O\left(\frac{1}{k}\right) \quad \left(\text{if } \mu \leqslant \frac{1}{k}\right)$$

$$\operatorname{Var}(Z) \leqslant s^{2k} \cdot \mu^2 \cdot O\left(\left(\frac{\log k}{k}\right)^{1/4} \cdot \log(1/\mu)\right) + s^{2k} \cdot \frac{1}{k} \leqslant s^{2k} \cdot O\left(\left(\frac{\log k}{k}\right)^{1/4}\right) \quad \left(\text{if } \mu > \frac{1}{k}\right)$$

where we used the fact that $e^x \leqslant 1 + 2x$ for $|x| \leqslant 1/2$ for the first inequality and the fact that $\mu \leqslant 1$ for the second.

Finally, using Chebyshev's inequality, we get

$$\begin{aligned}\Pr\left[\left|Z - \mu \cdot s^k\right| \geqslant \varepsilon \cdot s^k\right] &= \Pr_{\mathbf{a},h}\left[|Z - \mathbb{E}[Z]| \geqslant \varepsilon \cdot s^k\right] \\ &\leqslant \frac{\operatorname{Var}(Z)}{\varepsilon^2 s^{2k}} \leqslant \frac{1}{\varepsilon^2} \cdot O\left(\left(\frac{\log k}{k}\right)^{1/4}\right) < \eta\end{aligned}$$

using the lower bound on k in the statement of the lemma. ■

C Local Correction

In this section, we show that the family of junta-sums can be locally corrected up to error approaching half the distance of the underlying code, i.e., we prove Theorem 5.3:

Theorem 5.3 (*Local correction of junta-sums*)**.** *For every $\varepsilon > 0$, finite set $\mathcal{S}$ of size $s \geqslant 2$ and $d \geqslant 0$, Abelian group G, the family $\mathcal{J}_d(\mathcal{S}^n, G)$ is $(\widetilde{\mathcal{O}}_\varepsilon(\log n)^d, \delta_{\mathcal{J}}/2 - \varepsilon)$-locally correctable where $\delta_{\mathcal{J}} := 1/s^d$.*

Moreover, if G is a torsion Abelian group of exponent M, then the number of queries can be made $O_{M,\varepsilon}(1)$, i.e., $\mathcal{J}_d(\mathcal{S}^n, G)$ is $(\mathcal{O}_{M,\varepsilon}(1), \delta_J/2 - \varepsilon)$-locally correctable.

Similar to the prior work on local correction of low-degree over the Boolean cube [ABPSS25], we divide the proof into two main steps:

- **Error reduction:** In this step, we give a way of reducing the error of the oracle $f : [s]^n \to G$ from $1/(2s^d) - \varepsilon$ to ε_1 for any given $\varepsilon_1 \leqslant 1/\Omega_{s,d}(\log n)^d$, by making $q_1 = \widetilde{O}_\varepsilon(1)$ queries to f. In particular, there exists a $\widetilde{O}_\varepsilon(1)$ query algorithm $\mathcal{A}$ such that, when given as oracle $f : [s]^n \to G$ such that $\delta(f, P) \leqslant 1/(2s^d) - \varepsilon$ for some $P \in \mathcal{J}_d([s]^n, G)$, it satisfies

  $$\Pr[\mathcal{A}^f(\mathbf{x}) \neq P(\mathbf{x})] \leqslant \varepsilon_1,$$

  where the above probability is both over the randomness of $\mathcal{A}$ and $\mathbf{x} \sim [s]^n$ is independently and uniformly chosen.

- **Correction in low-error regime:** Here, we now assume access to a *randomized* oracle $f' : [s]^n \to G$ such that $\Pr[f'(\mathbf{x}) \neq P(\mathbf{x})] \leqslant \varepsilon_1$ for $\mathbf{x} \sim [s]^n$ for some $P \in \mathcal{J}_d([s]^n, G)$, and design a $q_2 = O_{s,d}(1/\varepsilon_1)$ query algorithm $\mathcal{A}'$ such that for every $\mathbf{x} \in [s]^n$, we have

$$\Pr[\mathcal{A}'^{f'}(\mathbf{x}) \neq P(\mathbf{x})] \leqslant 1/4.$$

Hence, composing the algorithms $\mathcal{A}$ and $\mathcal{A}'$, we get a local corrector for f that uses at most $q_1 \cdot q_2 = \widetilde{O}_\varepsilon(\log n)^d$ queries. For the case of groups with small order, we follow the same line, except we change the threshold ε_1 to be at most $1/\Omega_{M,\varepsilon}(1)$, resulting in $q_1 = q_2 = O_{M,\varepsilon}(1)$. This would then finish the proof of Theorem 5.3.

While the error reduction procedure closely follows similar ideas as for the Boolean cube ($s = 2$) from prior work, the low-error regime needs some changes. We give the proofs for error reduction in Appendix C.1, and for the low-error local corrector in Appendix C.2. For the remainder of the section, we fix G to be an arbitrary Abelian group and assume that $s \geqslant 2$ (as $\mathcal{J}_d([s]^n, G)$ is a trivial family otherwise).

C.1 Error Reduction

The main goal of this subsection is to prove the following:

Lemma C.1 (Error reduction). *For every $\varepsilon_1 = 1/\Theta_{s,d}(\log n)^d$, there exists a $q_1 = \widetilde{O}_{s,d,\varepsilon}(1)$ query algorithm $\mathcal{A}$ such that for every $f : [s]^n \to G$ satisfying $\delta(f, P) \leqslant 1/(2s^d) - \varepsilon$ for some $P \in \mathcal{J}_d([s]^n, G)$, the following holds:*

$$\Pr[\mathcal{A}^f(\mathbf{x}) \neq P(\mathbf{x})] \leqslant \varepsilon_1,$$

where the probability is over a uniformly random $\mathbf{x} \sim [s]^n$, and an independent choice of the randomness of $\mathcal{A}$.

We will proceed in an almost identical way as done by [ABPSS25] with a natural extension of the notion of a *subcube* from $s = 2$ (i.e., Boolean cube) to general s. We show the following two key lemmas: the first one reduces the error from a small enough constant to "sub-constant" and the second one reduces it from $1/(2s^d) - \varepsilon$ to a small enough constant.

Lemma C.2 (Reduction from small constant to sub-constant error). *Fix any Abelian group G, any $s \geqslant 2$, and any positive integer d. The following holds for $\delta < 1/s^{\mathcal{O}(d)}$ and $K = s^{\mathcal{O}(d)}$ where the $\mathcal{O}(\cdot)$ hides a large enough absolute constant. For any η, δ, where $\eta < \delta$, there exists a randomized algorithm $\mathcal{A}$ with the following properties: Let $f : \mathbb{Z}_s^n \to G$ be a function and let $P : \mathbb{Z}_s^n \to G$ be a junta-degree-d function such that $\delta(f, P) \leqslant \delta$, and let $\mathcal{A}^f$ denote that $\mathcal{A}$ has oracle access to f. Then,*

$$\Pr[\delta(\mathcal{A}^f, P) > \eta] < 1/10,$$

where the above probability is over the internal randomness of $\mathcal{A}^f$. Further, for every $\mathbf{x} \in \{0,1\}^n$, $\mathcal{A}^f$ makes K^T queries to f and $T = \mathcal{O}\left(\log\left(\frac{\log(1/\eta)}{\log(1/\delta)}\right)\right)$.

We now state the second key error reduction lemma.

Lemma C.3 (Reduction to small constant error). *Fix any Abelian group G, any integer $s \geqslant 2$, and a positive integer d. For any η, δ, where $\eta < \delta$ and $\delta < 1/(2 \cdot s^d) - \varepsilon$ for $\varepsilon > 0$, there exists a randomized algorithm $\mathcal{A}$ with the following properties: Let $f : \mathbb{Z}_s^n \to G$ be a function and let $P : \mathbb{Z}_s^n \to G$ be a junta-degree d function such that $\delta(f, P) \leqslant \delta$, and let $\mathcal{A}^f$ denotes that $\mathcal{A}$ has oracle access to f, then*

$$\Pr[\delta(\mathcal{A}^f, P) > \eta] < 1/10,$$

where the above probability is over the internal randomness of $\mathcal{A}$, and for every $\mathbf{x} \in \mathbb{Z}_s^n$, $\mathcal{A}^f$ makes s^k queries to f, where $k = \mathrm{poly}(\frac{1}{\varepsilon}, \frac{1}{\eta}, s)$.

We prove the first lemma in Appendix C.1.1 and the second lemma in Appendix C.1.2. Below, we finish the proof of the main error reduction lemma of this section using the above two lemmas.

Proof of Lemma C.1. The proof proceeds in a similar way to [ABPSS24]: we apply the first step of error reduction (Lemma C.3) with $\eta = \eta_1 = O_{s,d}(1)$ being smaller than the value of δ needed to apply the second step (Lemma C.2), i.e., $\delta \leqslant O_{s,d}(1)$. This results in a number of queries which is the product of the number of queries from both the steps. Taking $\eta = \eta_2$ in the second error reduction step (i.e., Lemma C.2) to be equal to $\varepsilon_1 = 1/\Theta_{s,d}(\log n)^d$, we get that the total number of queries is $O_s\left(\frac{1}{\eta_1 \varepsilon}\right) \cdot O_{s,d}(1)^{\log\left(\log\left(\frac{\log(1/\eta_2)}{\log(1/\delta)}\right)\right)} \leqslant (\log\log n)^{O_{s,d,\varepsilon}(1)}$. ∎

C.1.1 Reduction from Small Constant to Sub-Constant Error

We will show that there is a randomized algorithm $\mathcal{A}^f$ that given oracle access to any function f that is δ-close to a junta-degree-d function P (think of δ as being a small enough constant depending on d), has the following property: with high probability over the internal randomness of $\mathcal{A}^f$, the function computed by $\mathcal{A}^f$ is η-close to P, where η can be much smaller than δ. We restate it formally below.

Lemma C.2 (Reduction from small constant to sub-constant error). *Fix any Abelian group G, any $s \geqslant 2$, and any positive integer d. The following holds for $\delta < 1/s^{\mathcal{O}(d)}$ and $K = s^{\mathcal{O}(d)}$ where the $\mathcal{O}(\cdot)$ hides a large enough absolute constant. For any η, δ, where $\eta < \delta$, there exists a randomized algorithm $\mathcal{A}$ with the following properties: Let $f : \mathbb{Z}_s^n \to G$ be a function and let $P : \mathbb{Z}_s^n \to G$ be a junta-degree-d function such that $\delta(f, P) \leqslant \delta$, and let $\mathcal{A}^f$ denote that $\mathcal{A}$ has oracle access to f. Then,*

$$\Pr[\delta(\mathcal{A}^f, P) > \eta] < 1/10,$$

where the above probability is over the internal randomness of $\mathcal{A}^f$. Further, for every $\mathbf{x} \in \{0,1\}^n$, $\mathcal{A}^f$ makes K^T queries to f and $T = \mathcal{O}\left(\log\left(\frac{\log(1/\eta)}{\log(1/\delta)}\right)\right)$.

In the rest of this subsection, we will prove Lemma C.2. The algorithm $\mathcal{A}^f$ in Lemma C.2 will be a recursive algorithm. Each recursive iteration of the algorithm $\mathcal{A}^f$ uses the same 'base algorithm' $\mathcal{B}$, which will be the core of our error reduction algorithm from small constant error. In the next lemma, we formally state the properties of the base algorithm.

Lemma C.4 (Base Error Reduction Algorithm)**.** *Fix any Abelian group G, any integer $s \geqslant 2$, and a positive integer d. The following holds for $K = s^{O(d)}$. For any $0 < \gamma < 1$, there exists a randomized algorithm $\mathcal{B}$ with the following properties: Let $g : \mathbb{Z}_s^n \to G$ be a function and let $P : \mathbb{Z}_s^n \to G$ be a junta-degree-d function such that $\delta(g, P) \leqslant \gamma$, and let $\mathcal{B}^g$ denote that $\mathcal{B}$ has oracle access to g, then*

$$\mathbb{E}[\delta(\mathcal{B}^g, P)] < O(K^2) \cdot \gamma^{1.5}$$

where the above expectation is over the internal randomness of $\mathcal{B}$. Further, for every $\mathbf{x} \in \mathbb{Z}_s^n$, $\mathcal{B}^g$ makes K queries to g.

We defer the construction of the base algorithm and proof of Lemma C.4 to later. For now, we assume Lemma C.4 and proceed to describe the recursive construction of $\mathcal{A}^f$ and prove Lemma C.2.

Proof of Lemma C.2. Let $\mathcal{B}$ be the algorithm given by Lemma C.4. We define a sequence of algorithms $\mathcal{A}_0^f, \mathcal{A}_1^f, \ldots,$ as follows.

The algorithm $\mathcal{A}_t^f$ computes a function mapping inputs in $\mathbb{Z}_s^n$ along with a uniformly random string from $\{0,1\}^{r_t}$ to a random group element in G.

- $\mathcal{A}_0^f$ just computes the function f. (In particular, $r_0 = 0$.)
- For each $t > 0$, we inductively define $r_t = r_{t-1} + r$, where r is the amount of randomness required by the base error reduction algorithm $\mathcal{B}$. On input $\mathbf{x} \in \mathbb{Z}_s^n$ and a uniformly random string σ_t, the algorithm $\mathcal{A}_t^f$ algorithm runs the algorithm $\mathcal{B}$ on $\mathbf{x}$ using the first r bits of σ_t as its source of randomness, and with oracle access to $\mathcal{A}_{t-1}^f$ using the remaining r_{t-1} bits of σ_t as randomness.

The algorithm $\mathcal{A}^f$ will be $\mathcal{A}_T^f$ for $T = C \cdot \log\left(\dfrac{\log(1/\eta)}{\log(1/\delta)}\right)$ where C is a large enough absolute constant chosen below.

Query complexity: An easy inductive argument shows that $\mathcal{A}^f$ makes at most K^T queries to f.

Error probability: We now analyze the error made by the above algorithms. We will argue inductively that for each $t \leqslant T$ and $\delta_t := \delta^{(1.1)^t}$, we have

$$\Pr_{\sigma_t}[\underbrace{\delta(\mathcal{A}_t^f(\cdot, \sigma_t), P) > \delta_t}_{:= \mathcal{E}_t}] \leqslant \sum_{j=1}^{t} \frac{1}{100^j} < \frac{1}{10}. \tag{25}$$

In the inductive proof, we will need that $\delta_0 = \delta < s^{-C_1 \cdot d}$ for a large enough absolute constant C_1.

We now proceed with the induction. The base case ($t = 0$) is trivial as $\delta(\mathcal{A}_t^f, P) = \delta_0$ by definition.

Now assume that $t > 1$. We decompose the random string σ_t into its first r bits, denoted σ, and its last r_{t-1} bits, denoted σ_{t-1}. We bound the probability in Equation (25) as follows. (Note that the event $\mathcal{E}_{t-1}$ below only depends on σ_{t-1}.)

$$\Pr_{\sigma_t}[\mathcal{E}_t] \leqslant \Pr_{\sigma_{t-1}}[\mathcal{E}_{t-1}] + \Pr_{\sigma_t}[\mathcal{E}_t \mid \neg\mathcal{E}_{t-1}] \leqslant \sum_{j=1}^{t-1} \frac{1}{100^j} + \Pr_{\sigma_t}[\mathcal{E}_t \mid \neg\mathcal{E}_{t-1}] \tag{26}$$

where we used the induction hypothesis for the second inequality. To bound $\Pr_{\sigma_t}[\mathcal{E}_t \mid \neg\mathcal{E}_{t-1}]$, fix any choice of σ_{t-1} so that $\neg\mathcal{E}_{t-1}$ holds, i.e. so that $\delta(\mathcal{A}^f_{t-1}, P) \leqslant \delta_{t-1}$. By the guarantee on $\mathcal{B}$, i.e. Lemma C.4, we know that

$$\mathbb{E}_\sigma[\delta(\mathcal{A}^f_t(\cdot, \sigma_t), P)] < \mathcal{O}(K^2) \cdot \gamma^{1.5},$$

where $\gamma = \delta(\mathcal{A}^f_{t-1}(\cdot, \sigma_{t-1}), P)$. Substituting it above, we get,

$$\mathbb{E}_\sigma[\delta(\mathcal{A}^f_t(\cdot, \sigma_t), P)] \leqslant \mathcal{O}(K^2) \cdot \delta^{1.5}_{t-1} \leqslant \delta^{1.25}_{t-1}$$

where for the final inequality, we use the fact that

$$\mathcal{O}(K^2) \cdot \delta^{0.25}_{t-1} \leqslant \mathcal{O}(K^2) \cdot \delta^{0.25}_0 \leqslant 1$$

as long as $\delta_0 = \delta \leqslant s^{-C_1 d}$ for a large enough constant C_1. Continuing the above computation, we see that by Markov's inequality

$$\Pr_\sigma[\mathcal{E}_t] \leqslant \frac{\delta^{1.25}_{t-1}}{\delta_t} = \delta^{\Omega((1.1)^t)} \leqslant \frac{1}{100^t}$$

where the final inequality holds for all t as long as $\delta \leqslant s^{-C_1 d}$ for a large enough constant C_1. Since this inequality holds for any choice of σ_{t-1} so that $\neg\mathcal{E}_{t-1}$ holds, we can plug this bound into Equation (26) to finish the inductive case of Equation (25).

Setting $T = C \cdot \log\left(\frac{\log(1/\eta)}{\log(1/\delta)}\right)$ for a large enough constant C, we see that $\delta_T < \eta$. In this case, Equation (25) implies the required bound on the error probability of $\mathcal{A}^f$. ■

Thus we have shown so far that given the base algorithm $\mathcal{B}$, we do get an error reduction algorithm from small constant error to error $\mathcal{O}(1/\log n)$. Now it remains to describe the base error reduction algorithm. In the next subsection, we describe the base algorithm $\mathcal{B}$ and prove Lemma C.4.

The base algorithm and its analysis. In the rest of this subsection, we prove Lemma C.4, which will then complete the proof of Lemma C.2. Before we describe $\mathcal{B}$, we will define an *error reduction gadget*.

Definition C.5 (Error-reduction Gadget for $\mathcal{J}_d$)**.** *For $\rho \in (0, 1/(s-1))$, an (ρ, q)-error reduction gadget for $\mathcal{J}_d$ is a distribution $\mathcal{D}$ over $(\mathbb{Z}^n_s)^q$ satisfying the following two properties:*

1. *There exists $c_1, \ldots, c_q \in \mathbb{Z}$ such that for any $(\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(q)}) \in \mathrm{supp}(\mathcal{D})$, the following holds true for each $P \in \mathcal{J}_d$ and each $\mathbf{a} \in \mathbb{Z}^n_s$*

$$P(\mathbf{a}) = c_1 P(\mathbf{a} + \mathbf{y}^{(1)}) + \ldots + c_q P(\mathbf{a} + \mathbf{y}^{(q)}) \tag{27}$$

   where the $\mathbf{a} + \mathbf{y}^{(i)} \in \mathbb{Z}^n_s$ is computed via a co-ordinate-wise sum modulo s.

2. *For any $i \in [q]$, the co-ordinates of $\mathbf{y}^{(i)}$ are i.i.d. random variables in $\mathbb{Z}_s$ that take the value 0 with probability p_i such that*

$$p_i \in \left[\frac{1}{s} - \rho \cdot \left(1 - \frac{1}{s}\right), \frac{1}{s} + \rho \cdot \left(1 - \frac{1}{s}\right)\right]$$

   *and each non-zero value in $\mathbb{Z}_s$ with probability $\frac{1-p_i}{s}$. We call such distributions ρ-*noisy distributions *over $\mathbb{Z}_s$.*

To prove Lemma C.4, we need an error-reduction gadget for $\mathcal{J}_d$, the space of junta-degree-d functions over a group G. This is given by the following lemma.

Lemma C.6 (Constructing an error-reduction gadget for $\mathcal{J}_d$)**.** *Fix any Abelian group G, $s \geqslant 2$ and any $\rho \in (0, 1/(s-1))$. Then $\mathcal{J}_d(\mathbb{Z}_s^n, G)$ has a (ρ, q)-error-reduction gadget where $q = ((1/\rho) + s)^{O(d)}$.*

Assuming the above lemma, we first finish the proof of Lemma C.4. For this, we will need the following technical claim.

Claim C.7. *Let y, z be independent random variables taking values in $\mathbb{Z}_s$ such that their distributions are ρ_1-noisy and ρ_2-noisy respectively. Then, $y - z$ is $(\rho_1 \cdot \rho_2)$-noisy.*

Proof. Let $\mathcal{D}_y$ and $\mathcal{D}_z$ denote the probability distributions of y and z respectively, which we think of as elements of $\mathbb{R}^s$.

We note that the condition that y is ρ_1-noisy can be restated as

$$\mathcal{D}_y = \varepsilon_1 \cdot \delta_0 + (1 - \varepsilon_1) \cdot \mathcal{U}$$

where $\mathcal{U}$ denotes the uniform distribution over $\mathbb{Z}_s$, δ_0 denotes the distribution that places all its mass on 0, and ε_1 is a (possibly negative) number satisfying $|\varepsilon_1| \leqslant \rho_1$.

A similar fact also holds for the random variable $-z \in \mathbb{Z}_s$, since z being ρ_2-noisy implies the same for $-z$.

Now, the distribution $\mathcal{D}$ of $y - z$ is the convolution $\mathcal{D}_y * \mathcal{D}_z$ giving us

$$\mathcal{D} = (\varepsilon_1 \cdot \delta_0 + (1 - \varepsilon_1) \cdot \mathcal{U}) * (\varepsilon_2 \cdot \delta_0 + (1 - \varepsilon_2) \cdot \mathcal{U}) = \varepsilon_1\varepsilon_2 \cdot \delta_0 + (1 - \varepsilon_1\varepsilon_2) \cdot \mathcal{U}$$

where the latter equality is by distributivity and the fact that the convolution of $\mathcal{U}$ with any distribution is $\mathcal{U}$.

Since $|\varepsilon_1| \leqslant \rho_1$ and $|\varepsilon_2| \leqslant \rho_2$, we have the claim. ■

In the algorithm, we use the error-reduction gadget to correct the junta-sum at a *random point* $\mathbf{a} \in \{0,1\}^n$. This process is likely to give the right answer except with probability $q\gamma$ since, after shifting, each query is now *uniformly* distributed and hence the chance that any of the queried points is an error point of g is at most γ. We reduce the error by repeating this process three times and taking a majority vote. To analyze this algorithm, we need to understand the probability that two iterations of this process both evaluate g at an error point. We do this using hypercontractivity (more specifically Theorem B.1).

Proof of Lemma C.4. Let $\mathcal{D}$ be a $(1/10s, q)$-error-reduction gadget as given by Lemma C.6. The algorithm $\mathcal{B}$, given oracle access to $g : \mathbb{Z}_s^n \to G$ and $\mathbf{a} \in \mathbb{Z}_s^n$, does the following.

- Repeat the following three times independently. Sample $(\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(q)})$ from $\mathcal{D}$ and compute

  $$c_1 g(\mathbf{a} + \mathbf{y}^{(1)}) + \cdots + c_q g(\mathbf{a} + \mathbf{y}^{(q)})$$

  where $c_1, \ldots, c_q$ are the coefficients corresponding to the error-reduction gadget, and the sums $\mathbf{a} + \mathbf{y}^{(i)}$ are computed in $\mathbb{Z}_s^n$.

- Output the plurality among the three group elements b_1, b_2, b_3 computed above.

The number of queries made by the algorithm is $K = O(q) = (10s + s)^{O(d)} = s^{O(d)}$ as claimed. So it only remains to analyze $\delta(\mathcal{B}^g, P)$. From now on, let $\mathbf{a}$ be a uniformly random input in $\{0,1\}^n$.

For $i \in \{1,2,3\}$, let $\mathcal{E}_i$ denote the event that $b_i \neq P(\mathbf{a})$. We have

$$\mathbb{E}[\delta(\mathcal{B}^g, P)] = \Pr[\mathcal{B}^g(\mathbf{a}) \neq P(\mathbf{a})] \leqslant \Pr[\mathcal{E}_1 \wedge \mathcal{E}_2] + \Pr[\mathcal{E}_2 \wedge \mathcal{E}_3] + \Pr[\mathcal{E}_1 \wedge \mathcal{E}_3].$$

It therefore suffices to show that each of the three terms in the final expression above is at most $O(q^2) \cdot \gamma^{1.5}$.

Without loss of generality, consider the event $\mathcal{E}_1 \wedge \mathcal{E}_2$. Let $(\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(q)})$ and $(\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(q)})$ be the two independent samples from $\mathcal{D}$ in the two corresponding iterations.

It follows from Equation (27) that the algorithm correctly computes $P(\mathbf{a})$ in the first iteration as long as none of the queried points lie in the set T of points where g and P differ. A similar statement also holds for the second iteration. This reasoning implies that

$$\Pr[\mathcal{E}_1 \wedge \mathcal{E}_2] \leqslant \sum_{i,j=1}^{q} \Pr[\underbrace{\mathbf{a} + \mathbf{y}^{(i)}}_{\mathbf{u}^{(i)}} \in T \wedge \underbrace{\mathbf{a} + \mathbf{z}^{(j)}}_{\mathbf{v}^{(j)}} \in T]. \tag{28}$$

We bound the latter expression using Theorem B.1.

Fix $i, j \in [q]$. Note that for every fixing of $\mathbf{y}^{(i)}$, the vector $\mathbf{u}^{(i)}$ is distributed uniformly over $\mathbb{Z}_s^n$ (because $\mathbf{a}$ is uniform over $\mathbb{Z}_s^n$). In particular, this implies that $\mathbf{u}^{(i)}$ is uniformly distributed and moreover that $\mathbf{u}^{(i)}$ and $\mathbf{y}^{(i)}$ are independent random variables.

Note, moreover, that $\mathbf{y}^{(i)}$ is independent of $\mathbf{z}^{(j)}$ and their entries are i.i.d. random variables over $\mathbb{Z}_s$ that are ρ-noisy. By Claim C.7 above, we see that the entries of $\mathbf{y}^{(i)} - \mathbf{z}^{(j)}$ are i.i.d. and $\rho^2 = (1/100s^2)$-noisy.

This means that $\mathbf{v}^{(j)} = \mathbf{u}^{(i)} + \mathbf{y}^{(i)} - \mathbf{z}^{(j)}$ is drawn from the noise distribution $\mathcal{N}_\sigma(\mathbf{u}^{(i)})$, where the parameter $\sigma \leqslant 1/100s^2$. Using Theorem B.1 with $q = 4$, we have

$$\Pr[\mathbf{u}^{(i)} \in T \wedge \mathbf{v}^{(j)} \in T] \leqslant \gamma^{1.5}.$$

Plugging this into Equation (28) implies the required bound on the probability of $\mathcal{E}_1 \wedge \mathcal{E}_2$. This concludes the analysis of $\mathcal{B}$. ■

We now show how to construct the error-reduction gadget and prove Lemma C.6. This requires the following claim (implied e.g. by Möbius inversion) that shows that any junta-degree-d function over $\{0,1\}^n$ (even with group coefficients) can be interpolated from its values on a Hamming ball of radius d. For completeness, we give a short proof.

Lemma C.8. *Fix $d \in \mathbb{N}$. For any natural number $m \geqslant d$ and any Hamming ball B of radius d,*

$$P(0^m) = \sum_{\mathbf{b} \in B} \alpha_{\mathbf{b}} P(\mathbf{b})$$

where the $\alpha_{\mathbf{b}}$ are integer coefficients.

Proof. Assume that

$$P(\mathbf{x}) = \sum_{\substack{\mathbf{a}\in\mathbb{Z}_s^n \\ \#\mathbf{a}\leqslant d}} g_{\mathbf{a}} \cdot \prod_{i\in[n]:\ a_i\neq 0} \delta_{a_i}(x_i).$$

By Möbius inversion, we know that

$$g_{\mathbf{a}} = \sum_{J\subseteq I} (-1)^{|I\setminus J|} P(1_J \circ \mathbf{a})$$

where $1_J \in \{0,1\}^m$ denotes the indicator vector of set J and $\circ$ denots co-ordinate-wise product. Putting the above equalities together gives us

$$P(\mathbf{x}) = \sum_{\#\mathbf{b}\leqslant d} \alpha'_{\mathbf{b},\mathbf{x}} P(\mathbf{b})$$

for suitable integer coefficients $\alpha'_{\mathbf{b},\mathbf{x}}$.

Now, assume B is the Hamming ball of radius d around the point $\mathbf{c} \in \mathbb{Z}_s^m$. Replacing $\mathbf{x}$ by $\mathbf{x}+\mathbf{c}$ in P does not increase the junta-degree of the function (since each co-ordinate of $\mathbf{x}+\mathbf{c}$ depends only on a single co-ordinate of $\mathbf{x}$). Applying this substitution above yields

$$P(\mathbf{x}+\mathbf{c}) = \sum_{\#\mathbf{b}\leqslant d} \alpha'_{\mathbf{b},\mathbf{x}} P(\mathbf{b}+\mathbf{c}) = \sum_{\mathbf{b}\in B} \alpha_{\mathbf{b},\mathbf{x}} P(\mathbf{b}).$$

Setting $\mathbf{x} = -\mathbf{c}$ yields the statement of the lemma. ■

We end this section by completing the proof of Lemma C.6.

Proof of Lemma C.6. The idea is to apply Lemma C.8 on a random subcube, as defined in Definition 2.9.

More precisely, let k, d be positive integers such that k is divisible by s and $k \geqslant s \cdot d$. Let $\mathbf{a} \in \mathbb{Z}_s^n$ be arbitrary. For each $i \in [n]$, let $\Pi_i \in \mathsf{Sym}[\mathbb{Z}_s]$ be chosen uniformly *from among bijections that map* 0 *to* a_i, and let $\mathbf{\Pi}$ denote $(\Pi_1, \ldots, \Pi_n)$. Also assume that $h : [n] \to [k]$ is chosen uniformly at random. Let $C = C_{\mathbf{\Pi},h}$ be the corresponding subcube of $\mathbb{Z}_s^n$ as defined in Definition 2.9. Let $Q(y_1, \ldots, y_k)$ denote $P|_C$, the restriction of P to this subcube.

Fix a Hamming ball B of radius d in $\mathbb{Z}_s^k$ centred at a point $\mathbf{c}$ with exactly k/s many occurrences of 0. Since Q is a function of junta-degree at most d, applying Lemma C.8 to Q and the ball B yields an equality

$$Q(0^k) = \sum_{\mathbf{b}\in\mathbf{B}} \alpha_{\mathbf{b}} Q(\mathbf{b}).$$

Since Q is a restriction of P, the above equality can be rephrased in terms of P as

$$P(x(0^k)) = \sum_{\mathbf{b}\in\mathbf{B}} \alpha_{\mathbf{b}} P(x(\mathbf{b})).$$

From the definition of the cube C, it follows that $x(0^k) = \mathbf{a}$ and thus the above gives us an equality of the type desired in an error-reduction gadget (Equation (27)). To finish the proof, we only need to argue that each $x(\mathbf{b})$ has the required distribution.

Note that for each $\mathbf{b} \in B$, we have

$$x(\mathbf{b}) = \mathbf{a} + \mathbf{b}'$$

where $\mathbf{b}'$ is the random vector in $\mathbb{Z}_s^n$ that at co-ordinate i takes the random value $\Pi_i(b_{h(i)})$. Since h is chosen uniformly at random and the Π_i's are independent and uniform subject to the constraint that $\Pi_i(0) = a_i$, it follows that the entries of $\mathbf{b}_h$ are independent and the ith co-ordinate is a $\mathbb{Z}_s$-valued random variable that takes the value 0 with probability equal to the proportion of 0's in $\mathbf{b}$ (which we denote σ) and each non-zero value in $\mathbb{Z}_s$ with the probability $(1-\sigma)/(s-1)$. In other words, the entries of $\mathbf{b}_h$ are ρ-noisy as long as

$$\sigma \in \left[\frac{1}{s} - \rho \cdot \left(1 - \frac{1}{s}\right), \frac{1}{s} + \rho \cdot \left(1 - \frac{1}{s}\right)\right].$$

To conclude the argument, note that $\mathbf{b}$ is at Hamming distance at most d from $\mathbf{c}$, implying that σ is in the range

$$\left[\frac{1}{s} - \frac{d}{k}, \frac{1}{s} + \frac{d}{k}\right].$$

Setting k to be the smallest multiple of s larger than $2d/\rho$ gives us the desired value for the parameter of the distribution of $\mathbf{b}$.

Finally, the number of queries q made by the error-reduction gadget is dictated by the size of a Hamming ball in $k = O(d/\rho)$ dimensions. This can be bounded by

$$\binom{k}{d} \cdot s^d \leqslant (k/d)^{O(d)} \cdot s^d \leqslant (1/\rho + s)^{O(d)} \cdot s^d = (1/\rho + s)^{O(d)}.$$

It follows that we have a $(\rho, ((1/\rho) + s)^{O(d)})$-error-reduction gadget. ■

C.1.2 Reduction to Small Constant Error

Now, we will show that there is a randomized algorithm $\mathcal{A}$ that given oracle access to any function f that is δ-close to a low junta-degree function P (think of δ to be very close to half the minimum distance, i.e. $1/(2 \cdot s^d) - \varepsilon$ for junta-degree d), has the following property: with high probability over the internal randomness of $\mathcal{A}$, $\mathcal{A}^f$ is η-close to P, where η is much smaller than δ. We recall it formally below.

Lemma C.3 (Reduction to small constant error). *Fix any Abelian group G, any integer $s \geqslant 2$, and a positive integer d. For any η, δ, where $\eta < \delta$ and $\delta < 1/(2 \cdot s^d) - \varepsilon$ for $\varepsilon > 0$, there exists a randomized algorithm $\mathcal{A}$ with the following properties: Let $f : \mathbb{Z}_s^n \to G$ be a function and let $P : \mathbb{Z}_s^n \to G$ be a junta-degree d function such that $\delta(f, P) \leqslant \delta$, and let $\mathcal{A}^f$ denotes that $\mathcal{A}$ has oracle access to f, then*

$$\Pr[\delta(\mathcal{A}^f, P) > \eta] < 1/10,$$

where the above probability is over the internal randomness of $\mathcal{A}$, and for every $\mathbf{x} \in \mathbb{Z}_s^n$, $\mathcal{A}^f$ makes s^k queries to f, where $k = \mathrm{poly}(\frac{1}{\varepsilon}, \frac{1}{\eta}, s)$.

Now we state an algorithm $\mathcal{A}^f$ below and use it to prove Lemma C.3..

Algorithm 4: Error Reduction Algorithm $\mathcal{A}^f$

Input: f and $\mathbf{a} \in \mathbb{Z}_s^n$

Choose $k = (s/(\varepsilon\eta))^{10}$

Sample a uniformly random $h : [n] \to [k]$ // h is the internal randomness of $\mathcal{A}^f$

Sample $\Pi_1, \ldots, \Pi_n \in \mathsf{Sym}[\mathbb{Z}_s]$ independently and uniformly at random subject to the condition that $\Pi_i(0) = a_i$ for each $i \in [n]$.

Construct the cube $\mathsf{C} := C_{\mathbf{\Pi},h}$ according to Definition 2.9

Let $\tilde{f} := f|_{\mathsf{C}}$ // $f|_{\mathsf{C}}$ is the restriction of f to the subcube C

Query $\tilde{f}$ on all inputs in $\mathbb{Z}_s^k$ to find the junta-sum $\tilde{P}$ on C such that $\delta(\tilde{f}, \tilde{P}) < 1/(2 \cdot s^d)$ // s^k queries to f

if *such a junta-sum* $\tilde{P}$ *is found* **then**

  return $\tilde{P}(0^k)$ // $x(0^k) = \mathbf{a}$

else

  return 0 // An arbitrary value

Proof of Lemma C.3. Let P be the (unique) junta-degree d function such that $\delta(f, P) < 1/(2 \cdot s^d)$. The junta-degree of P is at most d when P is restricted to $\mathsf{C} = C_{\mathbf{\Pi},h}$. If $\delta(P|_{\mathsf{C}}, \tilde{f}) < 1/(2 \cdot s^d)$, then $\tilde{P} = P|_{\mathsf{C}}$. In particular, $\tilde{P}(x(0^k)) = P(\mathbf{a})$, i.e. the output of the algorithm is correct.

Equivalently, $\mathcal{A}^f(\mathbf{a}) = P(\mathbf{a})$ unless $\delta(P|_{\mathsf{C}}, \tilde{f}) \geqslant 1/(2 \cdot s^d)$. In the next lemma, we will show that with high probability over random $\mathbf{a}$ and h, $\delta((P|_{\mathsf{C}}, \tilde{f}) < 1/(2 \cdot s^d)$.

Lemma C.9. *Sample* $\mathbf{a}$, $\mathbf{\Pi} = (\Pi_1, \ldots, \Pi_n)$ *and* h *as in the algorithm above. Let* $\mathsf{C} = C_{\mathbf{\Pi},h}$ *be the subcube of dimension* k *as described in Definition 2.9. Then,*

$$\Pr_{\mathbf{a},\mathbf{\Pi},h}[\delta(P|_{\mathsf{C}}, \tilde{f}) \geqslant 1/(2 \cdot s^d))] < \eta/10$$

We prove Lemma C.9 below. For now, let us assume Lemma C.9 and finish the proof of Lemma C.3. We have,

$$\Pr_{\mathbf{a},\mathbf{\Pi},h}[\delta(P|_{\mathsf{C}}, \tilde{f}) \geqslant 1/(2 \cdot s^d)] < \eta/10$$
$$\Rightarrow \mathbb{E}_{h,\mathbf{\Pi}}\left[\Pr_{\mathbf{a}}[\delta(P|_{\mathsf{C}}, \tilde{f}) \geqslant 1/(2 \cdot s^d)]\right] < \eta/10$$

Note that if we fix the internal randomness of $\mathcal{A}^f$ (i.e. the random bits used to choose $h, \mathbf{\Pi}$), then $\delta(\mathcal{A}^f, P)$ is at most $\Pr_{\mathbf{a}}[\delta(P|_{\mathsf{C}}, \tilde{f}) \geqslant 1/(2 \cdot s^d))]$, as the algorithm always outputs $P(\mathbf{a})$ correctly when $\delta(P|_{\mathsf{C}}, \tilde{f}) < 1/(2 \cdot s^d)$. Then from the above inequality, we have,

$$\mathbb{E}_{h,\mathbf{\Pi}}[\delta(\mathcal{A}^f, f)] < \eta/10$$
$$\Rightarrow \Pr_{h,\mathbf{\Pi}}[\delta(\mathcal{A}^f, f) > \eta] \leqslant 1/10 \qquad \text{(Markov's Inequality)}$$

As commented in Algorithm 4, for each $\mathbf{a} \in \mathbb{Z}_s^n$, $\mathcal{A}^f$ makes s^k queries to f. ■

Now we give the proof of Lemma C.9.

Proof of Lemma C.9. Let E denote the subset of points in $\mathbb{Z}_s^n$ where P and f disagree, i.e. $E := \{\mathbf{x} \in \mathbb{Z}_s^n \mid f(\mathbf{x}) \neq P(\mathbf{x})\}$. We know that $|E|/s^n \leqslant 1/(2 \cdot s^d) - \varepsilon$.

The fractional Hamming distance between $P|_{\mathsf{C}}, \tilde{f}$ is given by the relative size of the set $E \cap \mathsf{C}$ inside C. Note that since $\mathbf{a}$ is chosen at random and each Π_i is chosen at random satisfying $\Pi_i(0) = a_i$ (for each $i \in [n]$), we see that each Π_i is indeed a uniformly independent element of $\mathsf{Sym}[\mathbb{Z}_s]$. Hence, the subcube $\mathsf{C} = C_{\mathbf{\Pi},h}$ is a random subcube in the sense of Definition 2.9.

Applying the sampling lemma from Section 2 (i.e., Lemma 2.10), we get that for $k = (s/(\varepsilon\eta))^{10}$ (we assume without loss of generality that ε, η are small enough for k to satisfy the hypothesis of Lemma 2.10)

$$\Pr_{\mathbf{a},\mathbf{\Pi},h}[\delta(P|_{\mathsf{C}}, \tilde{f}) \geqslant 1/(2 \cdot s^d)] < \eta/10,$$

and this completes the proof of Lemma C.9. ∎

C.2 Correction in Low-Error Regime

Having just shown how to reduce the error, we will now prove that there is a local correction algorithm in this "low-error" regime.

Lemma C.10 (Local correction in low-error regime). *There exists $\varepsilon_1 = 1/\Theta_{s,d}(\log n)^d$ and a $q_2 = O_{s,d}(1/\varepsilon_1)$ query algorithm $\mathcal{A}$ such that for every randomized oracle $f : [s]^n \to G$ satisfying $\Pr_{\mathbf{x}\sim[s]^n}[f(\mathbf{x}) \neq P(\mathbf{x})] \leqslant \varepsilon_1$ for some $P \in \mathcal{J}_d([s]^n, G)$, it holds for every $\mathbf{x} \in [s]^n$ that:*

$$\Pr[\mathcal{A}^f(\mathbf{x}) \neq P(\mathbf{x})] \leqslant 1/4.$$

Using the above two lemmas, we can finish the proof of the first part of Theorem 5.3.

Proof of Theorem 5.3 for general Abelian groups. The proof follows by applying Lemma C.10 with the randomized oracle being $\mathcal{A}^f$ given by Lemma C.1. This yields a total number of queries of $q_1 \cdot q_2 = \widetilde{O}_{s,d,\varepsilon}(\log n)^d = \widetilde{O}_{\varepsilon}(\log n)^d$ as we can assume that $\varepsilon < \delta_{\mathcal{J}}/2 = 1/(2s^d)$. ∎

Before we prove Lemma C.10, we show the following claim which reduces the problem of local correction of junta-sums to local correction over the Boolean cube but with a biased distribution.

Lemma C.11 (Reduction to correction over biased cube). *Suppose there exists a q query algorithm $\mathcal{A}$ such that for every randomized oracle $f : \{0,1\}^n \to G$ satisfying $\Pr_{\mathbf{y}\sim\mathrm{Bern}(1/s)^n}[f(\mathbf{y}) \neq P(\mathbf{y})] \leqslant 10\varepsilon_1$ for some $P \in \mathcal{J}_d(\{0,1\}^n, G)$, it holds that $\Pr[\mathcal{A}^f(\mathbf{1}) \neq P(\mathbf{1})] \leqslant 1/4$.*

Then, there exists a $O(q/\varepsilon_1)$ query algorithm $\mathcal{A}'$ such that for every randomized oracle $f' : [s]^n \to G$ satisfying $\Pr_{\mathbf{x}\sim[s]^n}[f'(\mathbf{x}) \neq P'(\mathbf{x})] \leqslant \varepsilon_1$ for some $P' \in \mathcal{J}_d([s]^n, G)$, it holds for every $\mathbf{x} \in [s]^n$ that $\Pr[\mathcal{A}'^{f'}(\mathbf{x}) \neq P'(\mathbf{x})] \leqslant 1/4$.

Proof. We design $\mathcal{A}'$ using $\mathcal{A}$. Fix $\mathbf{x} \in [s]^n$ be arbitrarily and sample $\mathbf{x}' \in [s]^n$ but choosing $x'_i \in [s] \backslash \{x_i\}$ uniformly and independently at random. We then define $f : \{0,1\}^n \to G$ as follows: Given $\mathbf{y} \in \{0,1\}^n$, let $\mathbf{z} = \mathbf{z}(\mathbf{y}) \in [s]^n$ be defined by $z_i = x_i$ if $y_i = 1$ and $z_i = x'_i$ otherwise – then we define $f(\mathbf{y})$ to be equal to $f'(\mathbf{y}(\mathbf{z}))$; similarly we define $P : \{0,1\}^n \to G$ by $P(\mathbf{y}) = P'(\mathbf{y}(\mathbf{z}))$. Since P' is a d-junta-sum, so is P (for every choice of $\mathbf{x}'$), i.e., $P \in \mathcal{J}_d(\{0,1\}^n, G)$. Furthermore, we observe that for $\mathbf{y} \sim \mathrm{Bern}(1/s)^n$, the point $\mathbf{z}(\mathbf{y})$ is uniformly distributed over $[s]^n$ (over a random choice of $\mathbf{x}'$ and $\mathbf{y}$). In particular, we have

$$\mathop{\mathbb{E}}_{\mathbf{x}'}\left[\mathop{\mathbb{E}}_{\mathbf{y}\sim\mathrm{Bern}(1/s)^n}[\mathbb{1}[f(\mathbf{y}) \neq P(\mathbf{y})]]\right] = \Pr_{\mathbf{z}\sim[s]^n}[f'(\mathbf{z}) \neq P'(\mathbf{z})] \leqslant \varepsilon_1.$$

By Markov's inequality, therefore, $\Pr_{\mathbf{y}\sim\mathrm{Bern}(1/s)^n}[f(\mathbf{y}) \neq P(\mathbf{y})] \leqslant 10\varepsilon_1$ with probability at least 0.9 over the choice of $\mathbf{x}'$. Now using $\mathcal{A}$ and oracle access to f (which can be simulated using the oracle access to f'), we get a q query algorithm that outputs $P'(\mathbf{x})$ with probability at least $3/4 - 0.1$, which can be made at least $2/4$ by repeating this subroutine constant number of times. Finally, we have a $O(q)$ query algorithm $\mathcal{A}'$ such that $\Pr[\mathcal{A}'^{f'}(\mathbf{x}) \neq P'(\mathbf{x})] \leqslant 1/4$. ■

Proof of Lemma C.10. Using Lemma C.11, we have the ability to work with a biased distribution over the Boolean cube instead of a uniform distribution over $[s]^n$ (we note that the change of error from ε_1 to $10\varepsilon_1$ and the queries from q to $O(q/\varepsilon_1)$ are insignificant to the final asymptotic query complexity). Hence, it suffices to show that there exists $\varepsilon_1 = 1/\Theta_{s,d}(\log n)^d$ and a $O_{s,d}(\log n)^d$ query algorithm $\mathcal{A}$ such that for every randomized oracle $f : \{0,1\}^n \to G$ satisfying

$$\Pr_{\mathbf{x}\sim\mathrm{Bern}(1/s)^n}[f(\mathbf{x}) \neq P(\mathbf{x})] \leqslant \varepsilon_1$$

for some $P \in \mathcal{J}_d(\{0,1\}^n, G)$, it holds that $\Pr[\mathcal{A}^f(\mathbf{1}) \neq P(\mathbf{1})] \leqslant 1/4$. In other words, we want to locally correct low-junta-degree functions over the Boolean cube under a biased distribution. The high level idea is to adapt the construction for the unbiased distribution from [ABPSS25]. In particular, we prove the following key result, and the local corrector is then described in Algorithm 5.

Theorem C.12. *For a growing parameter k divisible by $10s^2d$, there exists $\mathcal{S} \subseteq \{0,1\}^k$ of size at most $O_{s,d}(k^d)$ such that the following conditions hold:*

- *$\mathcal{S}$ is* weight-balanced*: i.e., there exists a probability distribution $\mathcal{D}$ over $[k]$, such that for every $\mathbf{b} \in \mathcal{S}$: it holds that*

$$\left|\mathop{\mathbb{E}}_{i\sim\mathcal{D}}[b_i] - \frac{1}{s}\right| \leqslant \frac{1}{2^{\Omega_{s,d}(k)}}. \tag{29}$$

- *$\mathcal{S}$ is an* interpolating set*: i.e., for every Abelian group G and every $Q \in \mathcal{J}_d(\{0,1\}^k, G)$, there exist integers $(c_{\mathbf{b}})_{\mathbf{b}\in\mathcal{S}}$ such that*

$$Q(\mathbf{1}) = \sum_{\mathbf{b}\in\mathcal{S}} c_{\mathbf{b}} Q(\mathbf{b}).$$

Algorithm 5: Local corrector in low-error regime

Input: Oracle access to the function $f : \{0,1\}^n \to G$

1 Set $k = \Theta_{s,d}(\log n)$ so that the RHS term in (29) (i.e., $\frac{1}{2^{\Omega_{s,d}(k)}}$) is at most $\frac{1}{n^2}$ and let $\mathcal{S} \subseteq \{0,1\}^k$ be given by Theorem C.12.

2 Let $\mathcal{D}$ be the probability distribution over $[k]$ also given by Theorem C.12.

3 For $\mathbf{b} \in \mathcal{S}$, let $\mathbf{x} = \mathbf{x}(\mathbf{b}) \in \{0,1\}^n$ be the point obtained by setting $x_i = b_j$, where $j \sim \mathcal{D}$ is sampled independently for all $i \in [n]$.

4 Output $\sum_{\mathbf{b}\in\mathcal{S}} c_{\mathbf{b}} f(\mathbf{x}(\mathbf{b}))$, where $c_{\mathbf{b}}$ are integers given from Theorem C.12.

We first prove the correctness of Algorithm 5 before we provide a proof of Theorem C.12. We first note that the number of queries made by the local correction algorithm is equal to $|\mathcal{S}|$, which is $O_{s,d}(k^d) = O_{s,d}(\log n)^d$ as desired. It now remains to show that the probability of error $\Pr[\mathcal{A}^f(\mathbf{1}) \neq P(\mathbf{1})]$ is at most $1/4$, where $\mathcal{A}^f(\mathbf{1}) = \sum_{\mathbf{b}\in\mathcal{S}} c_{\mathbf{b}} f(\mathbf{x}(\mathbf{b}))$ is the output of Algorithm 5. Let $Q : \{0,1\}^k \to G$ be defined by $Q(\mathbf{y}) = P(\mathbf{x}(\mathbf{y}))$ i.e. it depends on the choice of randomness used in Step 3 of Algorithm 5. Since P is a d-junta-sum, so is Q, so by Theorem C.12, we know that

$$Q(\mathbf{1}) = \sum_{\mathbf{b}\in\mathcal{S}} c_{\mathbf{b}} Q(\mathbf{b}).$$

Equivalently, we thus get

$$P(\mathbf{1}) = \sum_{\mathbf{b}\in\mathcal{S}} c_{\mathbf{b}} P(\mathbf{x}(\mathbf{b})).$$

Hence, if all the queries to f by $\mathcal{A}$ output the value of P, then there is no error in the algorithm. However, there are two sources of error: firstly, $f(\mathbf{x}(\mathbf{b}))$ need not always be equal to $P(\mathbf{x}(\mathbf{b}))$. Indeed we are only guaranteed that they are equal with high probability for an input chosen from $\mathrm{Bern}(1/s)^n$ distribution. And secondly, the distribution of $\mathbf{x}(\mathbf{b})$ is not exactly identical to the $\mathrm{Bern}(1/s)^n$ distribution, but only statistically close to it. More precisely, we have

$$\Pr_{\mathbf{x}\sim\mathrm{Bern}(1/s)^n}[f(\mathbf{x}) \neq P(\mathbf{x})] \leqslant \varepsilon_1,$$

and the statistical distance between the distributions $\mathrm{Bern}(1/s)^n$ and $\mathbf{x}(\mathbf{b})$ is:

$$\mathrm{SD}(\mathrm{Bern}(1/s)^n, \mathbf{x}(\mathbf{b})) \leqslant \frac{1}{n},$$

for every $\mathbf{b} \in \mathcal{S}$ by the weight-balanced property of $\mathcal{S}$ as each bit of $\mathbf{x}(\mathbf{b})$ is $\frac{1}{n^2}$-close to $\mathrm{Bern}(1/s)$ and the n bits are all independent (see Step 3 of Algorithm 5); here are we using the property $\mathrm{SD}((X_1, X_2), (Y_1, Y_2)) \leqslant \mathrm{SD}(X_1, Y_1) + \mathrm{SD}(X_2, Y_2)$ if X_1, X_2 are independent and so are Y_1, Y_2 (see e.g. [Vad12] Lemma 6.3). Thus, we have for each $\mathbf{b} \in \mathcal{S}$, $\Pr[f(\mathbf{x}(\mathbf{b})) \neq P(\mathbf{x}(\mathbf{b}))] \leqslant \varepsilon_1 + \frac{1}{n}$. Now, applying a union bound over the queries made, we get

$$\Pr[\mathcal{A}^f(\mathbf{1}) \neq P(\mathbf{1})] \leqslant |\mathcal{S}| \cdot \left(\varepsilon_1 + \frac{1}{n}\right) \leqslant 1/4,$$

by taking $\varepsilon_1 = 1/\Theta_{s,d}(\log n)^d$ appropriately small.

This finishes the proof of Lemma C.10. ■

We now prove Theorem C.12.

Proof of Theorem C.12. We let $k = rm$, where $r = 10s^2d$ and identify $[k]$ with $[r] \times [m]$ arbitrarily and treat $\mathbf{y} \in \{0,1\}^k$ as a tuple of points in $\{0,1\}^r$, i.e., we let $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m)$ where each $\mathbf{y}_i \in \{0,1\}^r$ (equivalently we treat the point $\mathbf{y}$ as a Boolean $r \times m$ matrix with $\mathbf{y}_i$ being the column vectors). Then, we define the distribution $\mathcal{D} = \mathcal{D}(m)$ over $[k] \equiv [r] \times [m]$ so that the probability mass for (i,j) is proportional to $W_j^{(m)} = s^{m-j}$; in particular, we have $\Pr_{(i,j)\sim\mathcal{D}} = W_j^{(m)}/W^{(m)}$, where we denote $W = W^{(m)} = r\sum_{j=1}^{m} W_j^{(m)} = \frac{r(s^m-1)}{s-1}$.

There exists a subset $\mathcal{S} = \mathcal{S}_{m,d} \subseteq \{0,1\}^{r\times m}$ of size at most $(4rm)^d$ such that

- $\mathcal{S}$ is weight-balanced: i.e., for every $\mathbf{b} \in \mathcal{S}$, we have
$$\left| \sum_{(i,j)\in[r]\times[m]} \frac{W_j^{(m)}}{W^{(m)}} b_{j,i} - \frac{1}{s} \right| \leqslant \frac{d}{W^{(m)}}.$$
- $\mathcal{S}$ is a *hitting set*: i.e., for every Abelian group G and every non-zero $Q \in \mathcal{J}_d(\{0,1\}^k, G)$, there exists $\mathbf{b} \in \mathcal{S}$ such that $Q(\mathbf{b}) \neq 0$.

We note that the notion of a *hitting set* in the second item implies the interpolating set property in the statement of Theorem C.12 by using Claim 3.2.4 of [ABPSS25]. Moreover, we note that the RHS of the first item is at most $\frac{O(sd)}{2^m} = \frac{1}{2^{\Omega_{s,d}(k)}}$ as required. Thus, it remains to show the existence of the subset $\mathcal{S}_{m,d} \subseteq \{0,1\}^{r\times m}$ satisfying the above two conditions; we do this by induction on m.

Base case $m = 1$. We will make use of the following claim from [ABPSS25].

Claim C.13 ([ABPSS25] Claim 3.2.3)**.** *For every interval $I \subseteq \{0,1,\ldots,r\}$ of size at least $d+1$, there exists a subset $\mathcal{H}_{I,d} \subseteq \{0,1\}^r$ of size at most $(4r)^d$ such that*

- *$\mathcal{H}_{I,d}$ consists only of points $\mathbf{z}$ such that $|\mathbf{z}| \in I$, and*
- *For every non-zero $Q \in \mathcal{J}_d(\{0,1\}^k, G)$, there exists $\mathbf{z} \in \mathcal{H}_{I,d}$ such that $Q(\mathbf{z}) \neq 0$.*

Using the above with $I = [\frac{r}{s} - d, \frac{r}{s} + d]$, we directly get $\mathcal{S}_{1,d} = \mathcal{H}_{I,d}$ as the desired set – the weight-balanced property of $\mathcal{S}$ follows by the immediately as for every $\mathbf{b} \in \mathcal{S}_{1,d}$, we have $||\mathbf{b}| - \frac{r}{s}| \leqslant d$ by the first property of Claim C.13.

Induction step $m > 1$. Let $\mathcal{S}_{m-1,d'} \subseteq \{0,1\}^{r\times(m-1)}$ be given by the induction hypothesis, and similarly $\mathcal{H}_{I,d'} \subseteq \{0,1\}^r$ be given by Claim C.13 for $0 \leqslant d' \leqslant d$. Let $\mathbf{b} = (\mathbf{b}_1, \ldots, \mathbf{b}_{m-1}) \in \mathcal{S}_{m-1,d}$ be arbitrary. This gives us that

$$\left| \sum_{(i,j)\in[r]\times[m-1]} W_j^{(m-1)} b_{j,i} - \frac{W^{(m-1)}}{s} \right| \leqslant d. \tag{30}$$

We now show that there exists an interval $I_{\mathbf{b}} \subseteq \{0,1,\ldots,r\}$ of size at least $d+1$ such that for every $\mathbf{b}_m \in \{0,1\}^r$ with $|\mathbf{b}_m| \in I_{\mathbf{b}}$, it holds that $\mathbf{b}' = (\mathbf{b}_1, \ldots, \mathbf{b}_{m-1}, \mathbf{b}_m) \in \{0,1\}^{r\times m}$ satisfies the

weight-balanced property, i.e., by letting $\tau = \sum_{(i,j)\in[r]\times[m-1]} W_j^{(m-1)} b_{j,i} - \frac{W^{(m-1)}}{s}$ and $I_{\mathbf{b}}$ to be the interval $\frac{r}{s} - s\tau \pm d$ (which is well-defined as $|\tau| \leqslant d \leqslant \frac{r}{10s^2}$ from (30)), we have:

$$\left| \sum_{(i,j)\in[r]\times[m]} W_j^{(m)} b_{j,i} - \frac{W^{(m)}}{s} \right| = \left| |\mathbf{b}_m| - \frac{r}{s} + s\tau \right| \leqslant d. \tag{31}$$

Now, we are ready to describe $\mathcal{S} = \mathcal{S}_{m,d}$:

$$\mathcal{S} = \bigcup_{0\leqslant d'\leqslant d} \{\mathbf{b} \times \mathcal{H}_{I_{\mathbf{b}},d-d'} \ : \ \mathbf{b} \in \mathcal{S}_{m-1,d'}\}.$$

In particular, we show the following three properties for the above definition of $\mathcal{S}$.

- **Size.** We have that

$$\begin{aligned}
|\mathcal{S}| &\leqslant \sum_{d'=0}^{d} |\mathcal{S}_{m-1,d'}| \cdot |\mathcal{H}_{I_{\mathbf{b}},d-d'}| \\
&\leqslant \sum_{d'=0}^{d} (4(m-1)r)^{d'} \cdot (4r)^{d-d'} \\
&\quad \text{(using the induction hypothesis to upper bound } |\mathcal{S}_{m-1,d'}|) \\
&\leqslant (4r)^d \sum_{d'=0}^{d} (m-1)^{d'} \\
&\leqslant (4mr)^d.
\end{aligned}$$

- **Weight-balanced.** This follows from the discussion leading to (31).

- **Hitting set.** Let $Q \in \mathcal{J}_d(\{0,1\}^{r\times m}, G)$ be an arbitrary non-zero d-junta-sum. Treating it as a junta-polynomial in the last column of variables, we have for every $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_m) \in \{0,1\}^{r\times m}$:

$$Q(\mathbf{x}) = \sum_{A\subseteq[r]:|A|\leqslant d} Q_A(\mathbf{x}_1, \ldots, \mathbf{x}_{m-1}) \cdot \mathbf{x}_m^A.$$

  Since Q is non-zero, let $A \subseteq [r]$ be such that Q_A is a *non-zero* function of junta-degree $d' \leqslant d$. By induction hypothesis, we know there exists $\mathbf{b} \in \mathcal{S}_{m-1,d'}$ such that $Q_A(\mathbf{b}) \neq 0$. Letting $Q' : \{0,1\}^r \to G$ denote the restriction of Q obtained on setting $\mathbf{x}_i = \mathbf{b}_i$ for all $i \in [m-1]$, we note that Q' is a *non-zero* junta-polynomial of degree at most $d - d'$. Hence, there exists $\mathbf{b}_m \in \mathcal{H}_{I_{\mathbf{b}},d'}$ such that $Q'(\mathbf{b}) \neq 0$. Effectively, this shows that there exists $\mathbf{b}' \in \mathcal{S}$ such that $Q(\mathbf{b}') \neq 0$.

■

C.3 Correction for Torsion Groups

We now finish the proof for the "moreover" part of Theorem 5.3, i.e., we show a constant query local correction algorithm over torsion groups of constant exponent.

Proof of Theorem 5.3 for torsion Abelian groups. Similar to the case of general Abelian groups, we first apply the error reduction step from Appendix C.1 (but with a different threshold ε_1) and the reduce the local correction problem to that over the Boolean cube but with a biased distribution (i.e., Lemma C.11). Thus, it suffices to show that there exists a $q = O_{M,s,d}(1)$ query algorithm $\mathcal{A}$ such that for every randomized oracle $f : \{0,1\}^n \to G$ satisfying $\Pr_{\mathbf{x}\sim \mathrm{Bern}(1/s)^n}[f(\mathbf{x}) \neq P(\mathbf{x})] \leqslant \varepsilon_1$ for some $P \in \mathcal{J}(\{0,1\}^n, d)$, it holds that $\Pr[\mathcal{A}^f(\mathbf{1}) \neq P(\mathbf{1})] \leqslant 1/4$.

In particular, we set $\varepsilon_1 = \frac{1}{10\binom{sk}{k}}$ for a suitably large $k = \mathcal{O}_{M,s,d}(1)$ (so $\varepsilon_1 \geqslant \Omega_{M,s,d}(1)$). We state it as a lemma below:

Lemma C.14. *For every Abelian torsion group G of exponent M, there exists $k = \mathcal{O}_{M,s,d}(1)$ and a $q = O_{M,s,d}(1)$ query algorithm $\mathcal{A}$ such that for every randomized oracle $f : \{0,1\}^n \to G$ satisfying $\Pr_{\mathbf{x}\sim \mathrm{Bern}(1/s)^n}[f(\mathbf{x}) \neq P(\mathbf{x})] \leqslant \varepsilon_1$ for some $P \in \mathcal{J}(\{0,1\}^n, d)$ and $\varepsilon_1 = \frac{1}{10\binom{sk}{k}}$, it holds that $\Pr[\mathcal{A}^f(\mathbf{1}) \neq P(\mathbf{1})] \leqslant 1/4$.*

Now we note that by using the error-reduction lemma Lemma C.3 with $\eta = \varepsilon_1 \geqslant \Omega_{M,s,d}(1)$, we can convert a local corrector for error ε_1 to one with error up to $1/(2s^d) - \varepsilon$ with a $\mathcal{O}_\varepsilon(1)$ factor blow-up. Combining with the low-error local corrector of Lemma C.14, we obtain a local corrector over the biased distribution $\mathrm{Bern}(1/s)^n$ making $\mathcal{O}_{M,s,d}(1)$ queries. Therefore, by Lemma C.11, we also get a $\mathcal{O}_{M,\varepsilon}(1)$ query local corrector for d-junta-sums over $\mathcal{S}^n$ for error up to $1/(2s^d) - \varepsilon$.

■

We now prove Lemma C.14.

Proof of Lemma C.14. The proof proceeds in an identical manner to the analysis of [ABPSS25] by making use of Kummer's theorem which may be thought of as an analog of Lucas' theorem for prime powers. We state Kummer's theorem below, where the notation $S_p(n)$ denotes the sum of the digits of n when written in base p.

Theorem C.15 (Kummer's theorem [Kum52])**.** *Let $p \in \mathbb{N}$ be a prime. Then for any integers $a \geqslant b \geqslant 0$, the largest power of p that divides $\binom{a}{b}$ is equal to $\frac{S_p(b)+S_p(a-b)-S_p(a)}{p-1}$.*

Let $M = \prod_{j=1}^{\ell} p_j^{r_j}$ be the prime factorization of the exponent M of G (so $\ell \leqslant \log M$). For each $j \in [\ell]$, let $s_j \in \mathbb{N}$ be the smallest integer such that $p_j^{r_j s_j} > d$. Then, we choose $k = \prod_{j\in[\ell]} p_j^{3r_j s_j}$. Note that $p_j^{r_j(s_j-1)} \leqslant d$ and hence $k \leqslant \prod_{j\in[\ell]} (dp_j^{r_j})^3 \leqslant d^{3\ell} M^3 = \mathcal{O}_{M,d}(1)$ as needed. We then recall that $\varepsilon_1 = \frac{1}{10\binom{sk}{k}} = \Omega_{M,s,d}(1)$.

We claim that the algorithm below (Algorithm 6) is the desired local corrector. It queries f at a few inputs from some distribution and outputs $P(\mathbf{1})$ with probability at least $9/10$, where $P \in \mathcal{J}_d$ is the unique degree-d junta-sum such that $\delta(f, P) \leqslant \varepsilon_1$. We will need the following claim in order to describe the local corrector.

Claim C.16. *There exist integers $c_{\mathbf{b}} \in \mathbb{Z}$ for $\mathbf{b} \in \binom{[sk]}{k}$ such that for every d-junta-sum $Q(\mathbf{y}) \in \mathcal{J}_d(\{0,1\}^{sk}, G)$, we have that*

$$Q(\mathbf{1}) = \sum_{\mathbf{b} \in \binom{[sk]}{k}} c_{\mathbf{b}} \cdot Q(\mathbf{b}). \tag{32}$$

Algorithm 6: Local corrector for torsion groups

Input: Oracle access to a randomized function $f : \{0,1\}^n \to G$

1 Sample a uniformly random function $h : [n] \to [sk]$.

2 For $\mathbf{b} \in \binom{[sk]}{k}$, let $\mathbf{x} = \mathbf{x}_h(\mathbf{b}) \in \{0,1\}^n$ be the point obtained by setting $x_i = b_{h(i)}$ for $i \in [n]$.

3 Output $\sum_{\mathbf{b} \in \binom{[sk]}{k}} c_{\mathbf{b}} f(\mathbf{x}_h(\mathbf{b}))$, where $c_{\mathbf{b}}$ are integers given by Claim C.16.

The above algorithm is similar to Algorithm 5 with the main difference being the choice of the interpolating set in the last step from Claim C.16 (as opposed to the "weight balanced interpolating set" of Theorem C.12).

Assuming the correctness of Claim C.16, we shall now finish the proof of Lemma C.14.

Firstly, we note that the local corrector makes $\binom{sk}{k} = \mathcal{O}_{M,s,d}(1)$ queries as required. To prove correctness, for every $\mathbf{b} \in \binom{[sk]}{k}$, we note that the corresponding query point $x_h(\mathbf{b}) \in \{0,1\}^n$ is distributed according to $\mathrm{Bern}(1/s)^n$ since the map h used in Algorithm 6 is uniformly random and $\mathbf{b}$ has $1/s$ fraction of indices as ones. Thus, we have that $f(\mathbf{x}_h(\mathbf{b})) \neq P(\mathbf{x}_h(\mathbf{b}))$ with probability at most ε_1, so by a union bound over $\mathbf{b}$, we have that with probability at least $1 - \varepsilon_1 \cdot \binom{sk}{k} = 9/10$ (over the random choice of h and the randomness of f), that $f(\mathbf{x}_h(\mathbf{b})) = P(\mathbf{x}_h(\mathbf{b}))$ for all $\mathbf{b} \in \binom{[sk]}{k}$. Now, letting $Q \in \mathcal{J}_d(\{0,1\}^{sk}, G)$ denote the restriction of P defined as $Q(\mathbf{y}) = P(\mathbf{x}_h(\mathbf{y}))$, we see that the output of Algorithm 6 is equal to

$$\sum_{\mathbf{b} \in \binom{[sk]}{k}} c_{\mathbf{b}} P(\mathbf{x}_h(\mathbf{b})) = \sum_{\mathbf{b} \in \binom{[sk]}{k}} c_{\mathbf{b}} Q(\mathbf{b}) = Q(\mathbf{1}) = P(\mathbf{1}),$$

where we are using Claim C.16 for the second equality and $\mathbf{x}_h(\mathbf{1}) = \mathbf{1}$ for the last equality. Therefore, the output of the local correction algorithm (Algorithm 6) is indeed $P(\mathbf{1})$ with probability at least $9/10$. ■

It now remains to prove Claim C.16.

Proof of Claim C.16. By replacing the variables x_i with $1 - x_i$, we note that the claim is equivalent to proving that there exists $c_{\mathbf{b}} \in \mathbb{Z}$ for $\mathbf{b} \in \binom{[sk]}{(s-1)k}$ such that for every d-junta-sum $Q \in \mathcal{J}_d(\{0,1\}^{sk}, G)$, it holds that

$$Q(\mathbf{0}) = \sum_{\mathbf{b} \in \binom{[sk]}{(s-1)k}} c_{\mathbf{b}} \cdot Q(\mathbf{b}). \tag{33}$$

To show this, we proceed with the following assignments. For every $\mathbf{b} \in \binom{[sk]}{(s-1)k}$, we set $c_{\mathbf{b}} = 0$ if $\mathbf{b}$ contains a 1 in any of the last $k-d$ coordinates and we set $c_{\mathbf{b}} = A$ otherwise, where $A \in \mathbb{Z}$ will be decided later. Recall that $M = \prod_{j\in[\ell]} p_j^{r_j}$ and $k = \prod_{j\in[\ell]} p_j^{3r_j s_j}$, and we have that $p_j^{r_j s_j} > d \geqslant p_j^{r_j(s_j-1)}$ for all $j \in [\ell]$. By linearity, it suffices to show (33) for $Q(\mathbf{y})$ of the form $g \cdot \prod_{j\in I} y_j$ for all $I \in \binom{[sk]}{\leqslant d}$ and $g \in G$. According to our assignment of $c_{\mathbf{b}}$, it is clear that (33) holds true (with LHS = RHS = 0) if I contains any of the last $k-d$ coordinates. Otherwise, we have that $I \subseteq \binom{[(s-1)k+d]}{\leqslant d}$. If $I = \varnothing$, we have $Q(0^{sk}) = g$ and $\sum_{\mathbf{b}\in\binom{[sk]}{(s-1)k}} c_{\mathbf{b}} \cdot Q(\mathbf{b}) = \binom{(s-1)k+d}{(s-1)k} A \cdot g$. On the other hand, if $|I| = i \geqslant 1$, we have $Q(0^{sk}) = 0$ and $\sum_{\mathbf{b}\in\binom{[sk]}{k}} c_{\mathbf{b}} \cdot Q(\mathbf{b}) = \binom{(s-1)k+d-i}{(s-1)k-i} A \cdot g$ since every non-zero term must have $b_j = 1$ for all $j \in I$. Hence, it suffices to find an integer A satisfying the following two conditions:

$$g = \binom{(s-1)k+d}{(s-1)k} A \cdot g, \text{ for all } g \in G, \text{ and}$$
$$0 = \binom{(s-1)k+d-i}{(s-1)k-i} A \cdot g, \text{ for all } g \in G \text{ and } i \in [d].$$

Let $k' := (s-1)k$. Since the order of every element g divides the exponent M of the group, for the above two conditions to hold, it suffices if for all $j \in [\ell]$ and $i \in [d]$, p_j does not divide $\binom{k'+d}{k'}$ and that $p_j^{r_j}$ divides $\binom{k'+d-i}{k'-i}$ for all $i \in [d]$. Then we can take A to be any integer such that $A\binom{k'+d}{k'} + A'M = 1$ for some integer A' (such A and A' are guaranteed to exist as M and $\binom{k'+d}{k'}$ are coprime). The rest of the proof is dedicated to verifying these divisibility constraints hold.

- p_j **does not divide** $\binom{k'+d}{k'}$**:** We will represent all the numbers k', d, i etc. in base p_j. We note that the last $r_j s_j$ digits of k' are zeroes since $p_j^{r_j}$ divides k'. Furthermore, since $d < p_j^{r_j s_j}$, all the digits of d except the last $r_j s_j$ many are zeroes. Hence, the sum of digits of $k'+d$ is equal to the sum of the digits of k' and d combined. That is, $S_{p_j}(k') + S_{p_j}(d) - S_{p_j}(k'+d) = 0$. Applying Kummer's theorem (Theorem C.15) now finishes the proof.
- $p_j^{r_j}$ **divides** $\binom{k'+d-i}{k'-i}$**:** By Kummer's theorem (Theorem C.15), it suffices to show that

$$\frac{S_{p_j}(d) + S_{p_j}(k'-i) - S_{p_j}(k'+d-i)}{p_j - 1} \geqslant r_j. \tag{34}$$

We note that $S_{p_j}(k'+d-i) = S_{p_j}(k') + S_{p_j}(d-i)$ by the same argument as the above paragraph. In addition, we have the trivial bounds $S_{p_j}(d) \geqslant 1$ and $S_{p_j}(d-i) \leqslant (p_j-1)r_j s_j$. Finally, we give a lower bound for $S_{p_j}(k'-i)$. Since k' has at least $3r_j s_j$ trailing zeroes, we get that $S_{p_j}(k'-1) \geqslant S_{p_j}(k') + 3r_j s_j(p_j-1) - 1$. But we observe that $S_{p_j}(k'-i) = S_{p_j}((k'-1)-(i-1)) = S_{p_j}(k'-1) - S_{p_j}(i-1)$ since the number of trailing (p_j-1)'s of $k'-1$ exceeds the total number of (non-zero) digits of $(i-1)$. Therefore, we get

$$\begin{aligned} S_{p_j}(d) + S_{p_j}(k'-i) - S_{p_j}(k'+d-i) &\geqslant 1 + S_{p_j}(k'-1) - S_{p_j}(i-1) - S_{p_j}(k') - S_{p_j}(d-i) \\ &\geqslant 1 + (3r_j s_j(p_j-1) - 1) - (p_j-1)r_j s_j - (p_j-1)r_j s_j \\ &\geqslant r_j s_j(p_j-1) \\ &\geqslant r_j(p_j-1). \end{aligned}$$

This finishes the proof of (34), and hence Claim C.16 and Lemma C.14.

■

D Combinatorial List-Decodability

We prove the combinatorial list-decodability bound for junta-sums (i.e., Theorem 5.4).

Theorem 5.4 (Combinatorial List Decoding Bound)**.** *For every $\varepsilon > 0$, positive integers s, d, and Abelian group G, the family $\mathcal{J}_d(\mathcal{S}^n, G)$ is $(1/s^d - \varepsilon, \mathcal{O}_\varepsilon(1))$-list decodable.*

The proof can be broken into the following four steps:

- First, we reduce to the setting where G is finite.
- Second, we show the combinatorial bound for finite groups where every element has a sufficiently large order.
- Third, we show the combinatorial bound for p-primary groups where p is a sufficiently small prime.[14]
- Finally, we combine the above bounds to get a combinatorial bound for arbitrary Abelian groups.

In particular, we prove the following two theorems.

Theorem D.1 (Combinatorial bound for large order)**.** *For every $\varepsilon > 0$, positive integers s,d, there exists a $p = p(s, d, \varepsilon)$ such that for every Abelian group G which does not have any element of order at most p, the family $\mathcal{J}_d([s]^n, G)$ is $(1/s^d - \varepsilon, O_\varepsilon(1))$-list decodable.*

And we have:

Theorem D.2 (Combinatorial bound for p-primary groups)**.** *For every $\varepsilon > 0$, positive integers s, d, prime p, and finite p-primary group G, the family $\mathcal{J}_d([s]^n, G)$ is $(1/s^d - \varepsilon, O_{\varepsilon,p}(1)$-list decodable.*

Using the above two theorems, we finish the proof of Theorem 5.4.

Proof of Theorem 5.4. Given Theorem D.1 and Theorem D.2, the proof follows the same outline as the prior work [ABPSS25] on combinatorial bound for low-degree polynomials over the Boolean cube, so we defer the proof. ■

We now show the proof for the large order case in Appendix D.1 and the p-primary groups case in Appendix D.2.

D.1 Combinatorial Bound for Large Order

We prove Theorem D.1 in this subsection. Throughout this section, we assume that all the elements of G have order at least p (where $p = p(\varepsilon)$ is to be determined), $s \geqslant 2$, and $\varepsilon \in (0, 1/s^d)$ is arbitrary.

[14] An Abelian group is said to be *p-primary* if every element has order that is an exponent of p.

Following along the lines of the proof for the Boolean case ($s = 2$) from [ABPSS25], we prove (and use) an anti-concentration inequality for junta-sums depending on many variables. Once the right anti-concentration lemma is in place, the rest of the proof of the combinatorial bound is more or less identical to the Boolean case, except we are able to make some simplifications as we are only aiming for a bound of $O_\varepsilon(1)$ (as opposed to $\mathrm{poly}(1/\varepsilon)$ from [ABPSS25]). We state the anti-concentration inequality below and defer its proof to the end of this subsection. In order to state the lemma, we need a definition – we say that a function $f : [s]^n \to G$ *depends* on the i-th variable if there exists $\mathbf{x} \in [s]^n$ such that $f(\mathbf{x}) \neq f(\mathbf{x}')$ for some $\mathbf{x}' \in [s]^n$ that agrees with $\mathbf{x}$ on the coordinates $[n]\backslash\{i\}$.

Lemma D.3 (Anti-concentration lemma). *For integers $s \geqslant 2$ and $d \geqslant 1$, and every $\varepsilon > 0$, there exists $r = r(s, d, \varepsilon) > 0$ and $p = p(s, d, \varepsilon)$ such that for every Abelian group G which does not contain any element of order less than p, and every $P \in \mathcal{J}_d([s]^n, G)$ that depends on at least r variables, it holds that:*

$$\Pr_{\mathbf{a}\sim[s]^n}[P(\mathbf{a}) \neq 0] \geqslant 1/s^{d-1} - \varepsilon.$$

Note that this improves on the trivial bound of $1/s^d$ for general non-zero junta-sums. Given the above lemma, we now prove Theorem D.1. The proof proceeds in multiple stages – in each stage, we make the junta-sums in the list (of close-by junta-sums to a fixed function) more structured, thus pruning the list at each stage.

D.1.1 Pruning the List

For a function $f : [s]^n \to G$, let $L_\varepsilon(f) \subseteq \mathcal{J}_d([s]^n, G)$ denote the set (or rather "list") of junta-sums P such that $\delta(f, P) \leqslant 1/s^d - \varepsilon$. Our goal is to show that $|L_\varepsilon(f)| \leqslant O_\varepsilon(1)$. We first reduce the problem to counting the number of junta-sums in the list that depend only on a few variables.

Reducing to counting junta-sums depending on a few variables. If $P_1, P_2 \in L_\varepsilon(f)$, note that

$$\delta(P_1 - P_2, \mathbf{0}) = \delta(P_1, P_2) \leqslant \delta(f, P_1) + \delta(f, P_2) \leqslant 2/s^d - 2\varepsilon.$$

Now applying Lemma D.3 for $P = P_1 - P_2$ (which is also a d-junta-sum), we get that $P_1 - P_2$ depends on at most $r(\varepsilon)$ variables, as otherwise we get $1/s^{d-1} - \varepsilon \leqslant \delta(P_1 - P_2, \mathbf{0}) \leqslant 2/s^d - 2\varepsilon$ which would be a contradiction. Hence, if $L_\varepsilon(f) = \{P_1, P_2, \ldots, P_t\}$, we observe that $P_1 - P_t, P_2 - P_t, \ldots, P_{t-1} - P_t$ are distinct junta-sums that are in $L_\varepsilon(P_1 - f)$ and depend on at most $r = O_\varepsilon(1)$ variables. Therefore, it suffices to count such junta-sums to get a final combinatorial bound. In order to do this, we first count such junta-sums depending on the same set of variables.

Counting junta-sums depending on the same set of few variables. Without loss of generality, let the variable set on which the junta-sums depend on be $[r]$. That is, let $P_1, \ldots, P_t$ be the d-junta-sums that are at distance at most $1/s^d - \varepsilon$ from a function $f : [s]^n \to G$, and each P_i only depends on the first r variables. For $\mathbf{a} \in [s]^{n-r}$, let $f_\mathbf{a} : [s]^r \to G$ be the function obtained by setting the last $n - r$ variables of f to be uniformly random independently. Since $\delta(f, P_i) \leqslant 1/s^d - \varepsilon$ for every $i \in [t]$, we have $\mathbb{E}_\mathbf{a}[\delta(f_\mathbf{a}, P_i)] \leqslant 1/s^d - \varepsilon$, hence with probability at least $\varepsilon/2$ over the choice of $\mathbf{a}$, it holds that $\delta(f_\mathbf{a}, P_i) \leqslant 1/s^d - \varepsilon/2$ (where we are thinking of P_i as being a function from $[s]^r$ to G). By linearity of expectation, this means that the expected number of junta-sums P_i such that $P_i \in L_{\varepsilon/2}(f_\mathbf{a})$ is at least $\varepsilon t/2$. Hence, it suffices to show that $|L_{\varepsilon/2}(f')|$ for every

$f' : [s]^r \to G$ is $O_\varepsilon(1)$ to conclude that $t = O_\varepsilon(1)$. To do this, we note that $P_1 \neq P_2 \in L_{\varepsilon/2}(f')$ cannot agree on more than $1 - 1/s^d$ fraction of inputs, so for a given subset of $[s]^r$ of size $s^r - s^{r-d}$, there is at most one junta-sum in the list $L_{\varepsilon/2}(f')$ that agrees with f' on that subset. Therefore $|L_{\varepsilon/2}(f')| \leqslant \binom{s^r}{s^r - s^{r-d}} = O_\varepsilon(1)$ as $r = O_\varepsilon(1)$.

Reducing to the case where the variable sets form a sunflower. We recall that our goal is to prove that the number of d-junta-sums that are at distance at most $1/s^d - \varepsilon$ from a given $f : [s]^n \to G$ is $O_\varepsilon(1)$. However, from the above paragraph, we see that the number of such junta-sums depending on the same set of variables is $O_\varepsilon(1)$. Thus, it suffices to show the following:

Suppose $P_1, \ldots, P_t \in L_\varepsilon(f)$ are such that they depend on *distinct* subsets of variables. Then, $t = O_\varepsilon(1)$.

Now, consider the set system formed over the universe $[n]$ by the subsets of variables each P_i depends on. Applying the sunflower lemma (e.g. [ER60], Theorem 3) to this set system, we observe that if $t > r!(m-1)^r$, then there exists $P_{i_1}, \ldots, P_{i_m} \in L_\varepsilon(f)$ such that the subset of variables they depend on forms a *sunflower*: that is, if the subset of variables that P_i depends on is denoted by $V_i \subseteq [n]$, then there exists a *core* $C \subseteq [n]$ such that $V_{i_{j_1}} \cap V_{i_{j_2}} = C$ for every $j_1 \neq j_2 \in [m]$ and the *petals* $V_{i_j} \backslash C$ are non-empty. Hence, it suffices to show that $m = O_\varepsilon(1)$ to get that $t = O_\varepsilon(1)$. For the remainder of the proof, we shall assume that $i_j = j$ for $j \in [m]$, without loss of generality.

Reducing to the case where the variable sets are pairwise disjoint. While the application of the sunflower lemma in the above step results in a core C which can be non-empty, the goal of this step is to show that we can essentially assume that $C = \varnothing$ without loss of generality. We prove this by carefully setting the variables in C (which is assumed to be non-empty) to constants. We will switch the domain of the functions from $[s]^n$ to $\mathbb{Z}_s^n$ as we will be using junta-polynomial representations.

Let $\mathbf{x} = \mathbf{z} \cup (\mathbf{y}^{(1)} \cup \mathbf{y}^{(2)} \ldots \mathbf{y}^{(m)}) \cup \mathbf{w}$ be a partition of the variable set where $\mathbf{z}$ denotes the variables indexed by C, and $y^{(i)}$ denotes the variables that P_i depends on other than $\mathbf{z}$ (i.e., $y^{(i)}$ corresponds to the variables indexed by $V_i \backslash C$), and $\mathbf{w}$ are the remaining variables. We let $n_0 = |C| = |\mathbf{z}|$ and $n_i = |y^{(i)}|$. Then we note that we can express each P_i (for $i \in [m]$) as follows:

$$P_i(\mathbf{x}) = P_i(\mathbf{z}, \mathbf{y}^{(i)}) = \sum_{\mathbf{a} \in \mathbb{Z}_s^{n_i} : |\mathbf{a}| \leqslant d} \delta_{\mathbf{a}}(\mathbf{y}^{(i)}) \cdot P_{i,\mathbf{a}}(\mathbf{z}),$$

where we use the notation $\delta_{\mathbf{a}}(\mathbf{y}^{(i)}) = \prod_{j \in [n_i]} \delta_{a_j}(y_j^{(i)})$. Let $\mathbf{y}$-*degree* of P_i denote the maximum value of $|\mathbf{a}|$ for which $P_{i,\mathbf{a}}$ is non-zero; since P_i depends on $\mathbf{y^{(i)}}$ variables, the $\mathbf{y}$-degree must be in $[d]$. Moreover, since $|\mathbf{z}| \leqslant r = O_\varepsilon(1)$, the number of possible monomials (without considering coefficients) in $P_{i,\mathbf{a}}$ is $O_{s,d,\varepsilon}(1) = O_\varepsilon(1)$. Thus, assuming m is a large enough function of $1/\varepsilon$ (otherwise, we are done), using the pigeon-hole principle, we can assume without loss of generality that the $\mathbf{y}$-degree of the P_i's are all the same (say $d' \in [d]$) and that each $P_{i,\mathbf{a}}$ contains a non-zero coefficient for the monomial $\delta_{\mathbf{b}}(\mathbf{z})$ for some $\mathbf{b} \in \mathbb{Z}_s^{n_0}$, and that $\delta_{\mathbf{b}}(\mathbf{z})$ is a non-zero monomial with the maximal degree. Without loss of generality, let the first n_0' coordinates of $\mathbf{b}$ be zero and the remaining ones be non-zero, where $0 \leqslant n_0' \leqslant n_0$. We will first set the first n_0' variables in $\mathbf{z}$ (if $n_0' = 0$, we skip this step) uniformly at random: we note that setting these variables cannot cancel the monomial $\delta_{\mathbf{b}}(\mathbf{z})$ as by assumption, it is a monomial with maximal degree. Denoting the

restricted functions by $P'_1, \dots, P'_t$ and the restriction of f by f', we have that these are all distinct and each P'_i satisfies $\delta(f', P'_i) \leqslant 1/s^d - \varepsilon/2$ with probability at least $\varepsilon/2$. Thus, there exists a choice of assignments to the first n'_0 variables of $\mathbf{z}$ such that for at least $\varepsilon t/2$ many P'_i s, it holds that $\delta(f', P'_i) \leqslant 1/s^d - \varepsilon/2$. Without loss of generality, we assume that these are the initial $t' = \varepsilon t/2$ junta-sums. We now set the remaining variables of $\mathbf{z}$ uniformly at random. We note that for $i \in [t']$, since $P_{i,\mathbf{a}}$ is still non-zero even after setting some variables of $\mathbf{z}$ in the earlier step, with probability at least $1/s^{n_0 - n'_0}$, it holds that P_i is non-zero. However, since the junta-degree of P_i is at most d and the $\mathbf{y}$-degree of P_i is d', we must have that $n_0 - n'_0 \leqslant d - d'$. That is, denoting the final junta-sums after setting all the variables of $\mathbf{z}$ by P''_i respectively and the restricted function of f by f'', we have that P''_i is non-zero with $\Omega_\varepsilon(1)$ probability. Furthermore, each P''_i if non-zero has junta-degree at most d'. Since $\delta(f', P'_i) \leqslant 1/s^d - \varepsilon/2$ and we are only setting $n_0 - n'_0 \leqslant d - d'$ variables when going from P'_i to P''_i, we must have that $\delta(f'', P''_i) \leqslant 1/s^{d'} - \varepsilon/2$. Thus, we have reduced to the case where the junta-sums we want to count all depend on pairwise disjoint sets of variables (although the degree changes from d to d', we will use d for the rest of the proof for simplicity; similarly we use ε instead of $\varepsilon/2$).

Counting junta-sums depending on pairwise disjoint variables. To recap, we are now in the following setup: We have an arbitrary function $f : [s]^n \to G$ and distinct d-junta-sums $P_1, \dots, P_t$ depending on pairwise disjoint subsets of variables such that $\delta(f, P_i) \leqslant 1/s^d - \varepsilon$, and the goal is to show that $t = O_\varepsilon(1)$. The main idea is that the junta-sums behave "independently" as they depend on disjoint subsets of variables and so there cannot be many of them correlated with the same function f. More formally, we consider the following quantity:

$$\Pr_{\mathbf{x} \sim [s]^n} \left[\exists i \in [t] : \left| \{ j \in [t] : P_j(\mathbf{x}) = P_i(\mathbf{x}) \} \right| \geqslant (1 - 1/s^d + \varepsilon/2)t - 1 \right]. \tag{35}$$

On the one hand, since $\Pr_{\mathbf{x} \sim [s]^n, i \sim [t]}[f(\mathbf{x}) = P_i(\mathbf{x})] \geqslant 1 - 1/s^d + \varepsilon$, we have that (35) is at least $\varepsilon/2$ (i.e., with probability $\varepsilon/2$, at least $1 - 1/s^d + \varepsilon/2$ fraction of the junta-sums agree with f and so with each other). On the other hand, since any two distinct junta-sums agree on at most $1 - 1/s^d$ fraction of inputs and the events $P_j(\mathbf{x}) = P_i(\mathbf{x})$ are independent across different $j \neq i$, we have that (35) is at most $t/2^{\Omega(\varepsilon^2 t)}$. Combining both, we get $t = O_\varepsilon(1)$.

Proof of Theorem D.1. The above discussion finishes the proof of Theorem D.1. ■

D.1.2 Anti-concentration Lemma

We end with a proof of the anti-concentration lemma (Lemma D.3). For this, we will need the following claim about junta-sums that have a certain matching structure. This is analogous (and extends) the corresponding result of Meka, Nguyen and Vu [MNV16] used in the analysis for the Boolean case ($s = 2$) in [ABPSS25]. To state the claim, we say that two monomials of a junta-polynomial: $\delta_{\mathbf{a}}$ and $\delta_{\mathbf{b}}$ (where $\mathbf{a}, \mathbf{b} \in \mathbb{Z}_s^n$), are *disjoint*, if the non-zero indices of $\mathbf{a}$ and $\mathbf{b}$ are disjoint.

Claim D.4. *For integers $s \geqslant 2$ and $d \geqslant 1$ and every $\varepsilon > 0$, there exists $u = u(s, d, \varepsilon)$ and $p = p(d, \varepsilon)$ such that for every Abelian group G which does not contain any element of order less than p, and every d-junta-sum*

$$P(\mathbf{x}) = \sum_{\mathbf{a} \in \mathbb{Z}_s^n : |\mathbf{a}| \leqslant d} g_{\mathbf{a}} \cdot \delta_{\mathbf{a}}(\mathbf{x})$$

with at least u many pairwise disjoint non-zero monomials of degree d, it holds that:

$$\Pr_{\mathbf{a}\sim\mathbb{Z}_s^n}[P(\mathbf{a})=0]\leqslant\varepsilon.$$

Proof. The main idea is to reduce to the Boolean case and use the following result from [ABPSS25], which itself is derived using the anti-concentration result of Meka, Nguyen and Vu [MNV16] over the reals.

Lemma D.5 ([MNV16], [ABPSS25] Theorem 4.1.6 and Claim 4.1.5)**.** *For every positive integer d and $\varepsilon>0$, there exists $t=t(d,\varepsilon)$ and $p=p(d,\varepsilon)$ such that for every Abelian group G which does not contain any element of order less than p, and every junta-degree-d polynomial $P\in\mathcal{J}_d(\{0,1\}^n,G)$ with at least t many pairwise disjoint non-zero monomials, it holds that:*

$$\Pr_{\mathbf{a}\sim\{0,1\}^n}[P(\mathbf{a})=0]\leqslant\varepsilon.$$

We now show how to use the above lemma to deduce a similar inequality for general s i.e., we prove Claim D.4. Let u denote the number of pairwise disjoint non-zero monomials of degree d in the junta-polynomial representation of P. Assuming a sufficiently large lower bound on u, our goal is to show that

$$\Pr_{\mathbf{a}\sim\mathbb{Z}_s^n}[P(\mathbf{a})=0]\leqslant\varepsilon.$$

We choose a uniformly random $\mathbf{a}\in\mathbb{Z}_s^n$ as follows:

- Choose a random *subcube* $C\subseteq\mathbb{Z}_s^n$ by picking $\mathbf{u},\mathbf{v}\in\mathbb{Z}_s^n$, where $u_i\neq v_i\in\mathbb{Z}_s$ are chosen uniformly at random and independently over $i\in[n]$: more specifically, $C=\{u_1,v_1\}\times\cdots\times\{u_n,v_n\}$.
- Choose $\mathbf{a}\in C$ uniformly at random.

Let $t=t(d,\varepsilon/2)$ and $p=p(d,\varepsilon/2)$ be given by the functions $t(.,.)$ and $p(.,.)$ in Lemma D.5. Let $\mathbf{a}_1,\ldots,\mathbf{a}_r\in\mathbb{Z}_s^n$ (where $u=u(s,d,\varepsilon)$ will be decided later) be such that the monomials $\delta_{\mathbf{a}_i}(\mathbf{x})$ are pairwise disjoint monomials, with $|\mathbf{a}_i|=d$, and have non-zero coefficients in P, where $i\in[u]$. Let $S_i\subseteq[n]$ denote the indices where $\mathbf{a}_i$ is non-zero, so that S_i are pairwise disjoint for $i\in[u]$. Now, if $u_j=0$ and $v_j=a_j$ for all $j\in S_i$, we note that if we treat P restricted to C as function over the Boolean cube (with $u_j\mapsto 0$ and $v_j\mapsto 1$ for $j\in S_i$ and rest of the coordinates are mapped arbitrarily), the monomial $\prod_{j\in S_i}x_j$ has the same coefficient as that of $\delta_{\mathbf{a}_i}(\mathbf{x})$ in P (which is non-zero), since no other monomials can cancel this. Thus, if we can prove that there are at least t many of the $\mathbf{a}_i$'s for which it holds that $(u_j,v_j)=(0,a_j)$ for $j\in S_i$, then we have a multilinear polynomial over C (or equivalently over $\{0,1\}^n$) with at least t many non-zero disjoint monomials, in which case, we apply Lemma D.5 to conclude that for a random point in $\mathbf{a}\sim C$, the probability that $P(\mathbf{a})=0$ is at most $\varepsilon/2$. Therefore,

$$\Pr_{\mathbf{a}\sim\mathbb{Z}_s^n}[P(\mathbf{a})=0]\leqslant\varepsilon/2+\Pr[|\{i\in[r]:(u_j,v_j)=(0,a_j)\ \forall j\in S_i\}|\geqslant t].$$

Now, we observe that the events $(u_j,v_j)=(0,a_j)$ are independent across $j\in S_i$ and i. In particular, we have $\Pr[(u_j,v_j)=(0,a_j)\ \forall i\in S_i]=\left(\frac{1}{s(s-1)}\right)^d\geqslant\Omega_{s,d}(1)$. Hence, if u is a sufficiently large enough function of s,d,ε, by applying the Chernoff bound, we get that

$$\Pr[|\{i\in[u]:(u_j,v_j)=(0,a_j)\ \forall j\in S_i\}|\geqslant t]\leqslant\varepsilon/2,$$

which in turn implies that $\Pr_{\mathbf{a}\sim\mathbb{Z}_s^n}[P(\mathbf{a}) = 0] \leqslant \varepsilon$. ■

Finally, we finish the proof of the anti-concentration lemma.

Proof of Lemma D.3. The proof is by induction on d.

Base case $d = 1$. We take $r = u(s, 1, \varepsilon)$, where $u(.,.)$ is given by the function in Claim D.4, so that if P depends on r variables, we are guaranteed that there are at least r degree 1 pairwise disjoint monomials. Then, applying Claim D.4, we get $\Pr_{\mathbf{a}\in[s]^n}[P(\mathbf{a}) \neq 0] \geqslant 1 - \varepsilon$.

Induction step $d > 1$. The analysis is based on three cases.

- **Case 1:** There exists a variable (say x_1 w.l.o.g.) and an index $j \in [s-1]$ such that in the junta-polynomial representation

$$P(\mathbf{x}) = P_0(x_2, \ldots, x_n) + \sum_{j=1}^{s-1} \delta_j(x_1) P_j(x_2, \ldots, x_n),$$

  P_j depends on at least $r_1 = r(s, d-1, \varepsilon)$ variables. In this case, we note that for a random choice of $a_2, \ldots, a_n \in [s]$, by applying the induction hypothesis to P_j (which is a $(d-1)$-junta-sum), we have that $P_j(a_2, \ldots, a_n) \neq 0$ with probability at least $\frac{1}{s^{d-2}} - \varepsilon$. Thus, the restriction of P unto the variable x_1 is a non-constant function on setting $x_i = a_i$ for $i > 1$. Therefore, we have $\Pr_{\mathbf{a}\sim[s]^n}[P(\mathbf{a} \neq 0)] \geqslant \frac{1}{s} \cdot \left(\frac{1}{s^{d-2}} - \varepsilon\right) \geqslant \frac{1}{s^{d-1}} - \varepsilon$.

- **Case 2:** Suppose there exists $r_2 = u(s, d, 1/2)$ many pairwise disjoint non-zero monomials of degree d in P, where $u(.)$ is given by Claim D.4. Then, we immediately get

$$\Pr_{\mathbf{a}\sim[s]^n}[P(\mathbf{a}) = 0] \leqslant \frac{1}{2} \leqslant 1 - \frac{1}{s^{d-1}}.$$

- **Case 3:** Suppose neither Case 1 nor Case 2 occur. We now consider the set system Δ over $[n]$, where we include $S \in \Delta$ for $S \subseteq [n]$ if there exists $\mathbf{b} \in \mathbb{Z}_s^n$ such that the coefficient of $\delta_{\mathbf{b}}(\mathbf{x})$ is non-zero in P, and S is the set of non-zero indices of $\mathbf{b}$. Since there cannot be r_2 many $S \in \Delta$ that are pairwise disjoint and each $S \in \Delta$ is of size at most d, we can guarantee that there exists a small "cover"; i.e., there exists indices $i_1, \ldots, i_\ell \in [n]$ with $\ell \leqslant dr_2$ such that for every degree d non-zero monomial $\delta_{\mathbf{b}}(\mathbf{x})$ in P, there exists $j \in [\ell]$ such that $b_{i_j} \neq 0$ (i.e., x_{i_j} is contained in the corresponding monomial). We now count the number of monomials in P which contain the variable x_{i_j} for some $j \in [\ell]$. Since Case 1 does not occur, we can bound this by $(s-1) \cdot \binom{(s-1)r_1}{\leqslant d}$, where the $s-1$ factor accounts for the number of monomials where x_{i_j} appears as $\delta_{j'}(x_{i_j})$ for $j' \in [s-1]$, and the second factor $\binom{(s-1)r_1}{\leqslant d}$ bounds the number of non-zero monomials of a function depending only on at most r_1 variables. Now, we set the variables $\{x_{i_j} : j \in [\ell]\}$ arbitrarily and show that the restricted function of P is still non-zero. We note that once we set these variables, all the degree d monomials would reduce in degree as the variables begin set form a "cover", thus we can bound the probability of the restriction of P being non-zero as being at least $\frac{1}{s^{d-1}}$. Hence, it only remains to prove that the restriction is non-zero. To see this, we set $r = r(s, d, \varepsilon) = 1 + 2\ell(s-1)\binom{(s-1)r_1}{\leqslant d}$; this ensures that even

after setting all the variables $x_{i_j} : j \in [\ell]$, there is at least one non-zero monomial in the restricted function.

■

D.2 Combinatorial Bound for p-primary groups

In this section, we prove Theorem D.2. The high level proof approach again follows closely as that of the Boolean case $s = 2$ from [ABPSS25]. The proof consists of the following steps:

- The first step is to reduce the problem from general p-primary groups to the case of $\mathbb{Z}_p$. We prove a combinatorial bound for this case and "lift" it to the general case.
- Then we show that that we can instead count polynomials over a field $\mathbb{F}_q$ (for some $q = O(s, p)$) rather than junta-sums.
- In order to get the bound for the $\mathbb{F}_q$ case, we show that we can essentially assume without loss of generality that the polynomials in the list have pairwise disjoint leading monomials.
- Finally, we show a tail bound for the roots of polynomials with pairwise disjoint leading monomials, which results in a list size bound.

We divide the proof into two subsections; we prove the first two items above in Appendix D.2.1 and the next two items in Appendix D.2.2.

D.2.1 Reducing to the Case of Constant-sized Field $\mathbb{F}_q$

For a field $\mathbb{F}$ and a subset $S \subseteq \mathbb{F}$ of size $|S| = s \geqslant 2$, we note that $\mathcal{J}_d(S^n, \mathbb{F})$ is exactly the family of functions that can be uniquely expressed as a polynomial where each (non-zero) monomial has at most d variables and individual degree at most $(s-1)$ in each variable. For the remainder of this subsection, we will use this interpretation. We show next that we can always assume that $S \subseteq \mathbb{F}$ (and then use the polynomial interpretation) without much loss in parameters for the combinatorial bound.

Lemma D.6 (Reducing counting junta-sums to polynomials). *If $\mathcal{J}_d(S^n, \mathbb{F}_q)$ is $(1/s^d - \varepsilon, O_{q,\varepsilon}(1))$-list-decodable for every $\varepsilon > 0$ and every finite field $\mathbb{F}_q$ and subset $S \subseteq \mathbb{F}_q$ of size s, then $\mathcal{J}_d([s]^n, \mathbb{Z}_p)$ is also $(1/s^d - \varepsilon, O_{p,\varepsilon}(1))$-list-decodable for every prime p and every $\varepsilon > 0$.*

Proof. Fix an arbitrary prime p and let q be the smallest power of p that is at least s. Let $f : [s]^n \to \mathbb{Z}_p$ be arbitrary and $P_1, \ldots, P_t \in \mathcal{J}_d([s]^n, \mathbb{Z}_p)$ be distinct junta-sums such that $\delta(f, P_i) \leqslant \frac{1}{s^d} - \varepsilon$ for $i \in [t]$. We will prove that $t = O_{q,\varepsilon}(1)$ assuming that $\mathcal{J}_d(S^n, \mathbb{F}_q)$ is $(1/s^d - \varepsilon, O_{q,\varepsilon}(1))$-list-decodable. We shall identify $\mathbb{Z}_p$ with a subgroup of $\mathbb{F}_q$ of order p: in particular, let $H \subseteq \mathbb{F}_q$ be a subgroup of $\mathbb{F}_q$ that is homomorphic to $\mathbb{Z}_p$, via a group homomorphism $\sigma : \mathbb{Z}_p \to H$. Let $\phi : [s] \to S$ be an arbitrary bijection and let $g : S^n \to \mathbb{F}_q$ be defined by $g(\mathbf{x}) = \sigma(f(\phi^{-1}(\mathbf{x})))$. Similarly, for $i \in [t]$, let $Q_i : S^n \to \mathbb{F}_q$ be defined by $Q_i(\mathbf{x}) = \sigma(P_i(\phi^{-1}(\mathbf{x})))$. We claim that $Q_i \in \mathcal{J}_d(S^n, \mathbb{F}_q)$: indeed, if $P_i(\mathbf{x}) = \sum_{I \in \binom{[n]}{\leqslant d}} P_{i,I}(\mathbf{x}_I)$ for functions $P_{i,I} : [s]^I \to \mathbb{Z}_p$, then we have $Q_i(\mathbf{x}) = \sigma\left(\sum_{I \in \binom{[n]}{\leqslant d}} P_{i,I}(\phi^{-1}(\mathbf{x}_I))\right) = \sum_{I \in \binom{[n]}{\leqslant d}} \sigma(P_{i,I}(\phi^{-1}(\mathbf{x}_I)))$. Moreover, $\delta(f, P_i) = \delta(g, Q_i)$ and Q_i's are pairwise distinct functions since σ and ϕ are bijections. By our assumption that $\mathcal{J}_d(S^n, \mathbb{F}_q)$

is $(1/s^d - \varepsilon, O_{q,\varepsilon}(1))$-list-decodable, we get that $t \leqslant O_{q,\varepsilon}(1) = O_{p,\varepsilon}(1)$.

■

Lemma D.7 (Lifting the bound to general p-primary groups). *If $\mathcal{J}_d([s]^n, \mathbb{Z}_p)$ is $(1/s^d - \varepsilon, O_{p,\varepsilon}(1))$-list-decodable for every $\varepsilon > 0$ and every prime p, then $\mathcal{J}_d([s]^n, G)$ is also $(1/s^d - \varepsilon, O_{p,\varepsilon}(1))$-list-decodable for every finite p-primary group G and every $\varepsilon > 0$.*

Proof. The proof essentially follows the same outline as that from [ABPSS25] which handles $s = 2$. For arbitrary fixed $\varepsilon > 0$, let L be the list size; i,e., for every $g : [s]^n \to \mathbb{Z}_p$, there exists at most $L \leqslant O_{p,\varepsilon}(1)$ junta-sums $Q \in \mathcal{J}_d([s]^n, \mathbb{Z}_p)$ such that $\delta(g, Q) \leqslant 1/s^d - \varepsilon$. We will now show that for an arbitrary $f : [s]^n \to G$, that the number of junta-sums $P \in \mathcal{J}_d([s]^n, G)$ such that $\delta(f, P) \leqslant 1/s^d - \varepsilon$ is at most $L^{O(\log(1/\varepsilon))}$, thus giving the required bound. Using the notation $L_\varepsilon(f)$ to denote the set of junta-sums $P \in \mathcal{J}_d([s]^n, G)$ such that $\delta(f, P) \leqslant 1/s^d - \varepsilon$, our goal now is to prove an upper bound on $|L_\varepsilon(f)|$. In order to prove this, we will need the following setup. Since G is a finite p-primary group, there exists an element $h_0 \in G$ of order p; let $H_0 \subseteq G$ be the subgroup generated by h_0. We then note that the quotient group G/H_0 is again a p-primary group. By continuing this argument, we have a sequence of groups $G = G_0, G_1, \ldots, G_h$ for some $h \in \mathbb{N}$ such that $G_{i+1} = G_i/H_i$, where $H_i \subseteq G_i$ is a subgroup of order p (generated by some $h_i \in G_i$) and G_h is the trivial group containing just the identity element. Now, we let $f_0 = f$ and for $0 \leqslant i \leqslant h$, we define $f_i : [s]^n \to G_i$ by the recurrence

$$f_{i+1}(\mathbf{x}) = f_i(\mathbf{x}) \mod H_i.$$

We now define a rooted tree T as follows: there are $h + 1$ levels of the tree, with the root being level h and the leaves being level 0. We now describe the vertices and their labels bottom-up. The vertices in level 0 are in bijection with the junta-sums $L_\varepsilon(f)$ (we treat these junta-sums as the "labels" of the vertices). For a vertex with label $P_0 \in L_\varepsilon(f)$ in level 0, we let $P_1 \in \mathcal{J}_d([s]^n, G_1)$ defined by

$$P_1(\mathbf{x}) = P_0(\mathbf{x}) \mod H_0$$

be the label of the parent of this vertex: we note that $P_1 \in L_\varepsilon(f_1)$ since if f and P_0 agree, so do f_1 and P_1. Proceeding in a similar way, we construct all the above levels of the tree T and label its vertices. In particular, the parent of a vertex in level i labeled with $P_i \in L_\varepsilon(f_i)$ is set to be the junta-sum $P_{i+1} \in L_\varepsilon(f_{i+1})$ defined as:

$$P_{i+1}(\mathbf{x}) = P_i(\mathbf{x}) \mod H_i.$$

For a vertex v of T at level $i \in [0..h]$ and labeled with $P_i \in L_\varepsilon(f_i)$, we let

$$\rho(v) = 1/s^d - \delta(f_i, P_i).$$

Note that $\rho(v) \geqslant \varepsilon$ for all vertices v of T. We further show the following properties of $\rho(\cdot)$.

Claim D.8. *For the tree T and the function ρ defined over the vertices of T defined above, the following properties hold:*

- *Each non-leaf vertex of T has at most L children.*
- *If u is the parent of v, then $\rho(u) \geqslant \rho(v)$.*
- *If u has two distinct children v and w, then $\rho(u) \geqslant \rho(v) + \rho(w)$.*

We now finish the proof of Lemma D.7 using the above claim. We recall that the number of leaves in T is exactly $|L_\varepsilon(f)|$, which is what we want to upper bound. To do this, we argue that for any non-leaf node u of T with children $v_1, \dots, v_t$ (for some $1 \leqslant t \leqslant L$), it holds that

$$\rho(u)^\ell \geqslant \sum_{i \in [t]} \rho(v_i)^\ell, \tag{36}$$

where $\ell = \lceil \log L \rceil$. Then, applying this inequality for all the non-leaf vertices of the tree, we get that

$$\rho(\text{root})^\ell \geqslant \sum_{v \text{ is a leaf}} \rho(v)^\ell \geqslant (\#\text{leaves}) \cdot \varepsilon^\ell.$$

Using $\rho(\text{root}) \leqslant 1$, we thus get that $|L_\varepsilon(f)| = (\#\text{ leaves}) \leqslant (1/\varepsilon)^\ell = L^{O(\log(1/\varepsilon))} = O_{p,\varepsilon}(1)$ as desired. Hence, it only remains to show that (36) holds. For this, we will assume that t, the number of children of u is at least 2 as otherwise, we immediately have $\rho(u)^\ell \geqslant \rho(v_1)^\ell$ using Item 2 of Claim D.8. Further, let $\rho(v_1) \geqslant \rho(v_2) \geqslant \cdots \geqslant \rho(v_t)$ without loss of generality. Then using Item 3 of Claim D.8 and $t \leqslant L \leqslant 2^\ell$, we have

$$\begin{aligned}
\rho(u)^\ell &\geqslant (\rho(v_1) + \rho(v_2))^\ell \\
&\geqslant \rho(v_1)^\ell + (2^\ell - 1)\rho(v_2)^\ell \\
&\geqslant \rho(v_1)^\ell + (t-1)\rho(v_2)^\ell \\
&\geqslant \rho(v_1)^\ell + \rho(v_2)^\ell + \cdots + \rho(v_t)^\ell.
\end{aligned}$$

■

We now prove Claim D.8.

Proof of Claim D.8. Let u be an arbitrary non-leaf vertex of T at level $i+1$ (for some fixed $i \in [0..h-1]$), with children $v_1, \dots, v_t$. Suppose u is labeled by a junta-sum $P \in L_\varepsilon(f_{i+1})$ and v_j is labeled by a junta-sum $Q_j \in L_\varepsilon(f_i)$ for $j \in [t]$. Therefore, for all $j \in [t]$, we have that

$$P(\mathbf{x}) = Q_j(\mathbf{x}) \mod H_i. \tag{37}$$

Hence, if f_i and Q_j agree on some input, so do f_{i+1} and P; so $\delta(f_{i+1}, P) \leqslant \delta(f_i, Q_j)$ and $\rho(u) \geqslant \rho(v_j)$, thus proving Item 2. Our goal now is to show that $t \leqslant L$ and $\rho(u) \geqslant \rho(v_{j_1}) + \rho(v_{j_2})$ for $j_1 \neq j_2 \in [t]$. To do this, we let $c_1, c_2, \dots, c_M \in G_i$ be fixed coset representatives (where $M = |G_i|/p$ and the cosets are ordered arbitrarily) corresponding to the subgroup H_i of G_i. Then each element $g \in G_i$ can be uniquely written as $g = g' + \widehat{g}$ with $g' \in H_i$ and $\widehat{g} \in \{c_1, \dots, c_M\}$ being a coset representative.

Let

$$Q_j(\mathbf{x}) = \sum_{\mathbf{a} \in \mathbb{Z}_s^n : |\mathbf{a}| \leqslant d} g_{j,\mathbf{a}} \cdot \delta_{\mathbf{a}}(\mathbf{x}),$$

for some $g_{j,\mathbf{a}} \in G_i$. From (37), we see that $\widehat{g_{j,\mathbf{a}}} = \widehat{g_{1,\mathbf{a}}}$ for all $j \in [t]$. Now we define $\widehat{Q} : [s]^n \to G_i$ to be:

$$\widehat{Q}(\mathbf{x}) = \sum_{\mathbf{a} \in \mathbb{Z}_s^n : |\mathbf{a}| \leqslant d} \widehat{g_{1,\mathbf{a}}} \cdot \delta_{\mathbf{a}}(\mathbf{x}),$$

and d-junta-sums $\widetilde{Q}_j \in \mathcal{J}_d([s]^n, H_i)$ for $j \in [t]$, to be:

$$\widetilde{Q}_j(\mathbf{x}) = \sum_{\mathbf{a} \in \mathbb{Z}_s^n : |\mathbf{a}| \leqslant d} g'_{j,\mathbf{a}} \cdot \delta_{\mathbf{a}}(\mathbf{x}).$$

Since $Q_j(\mathbf{x}) = \widehat{Q}(\mathbf{x}) + \widetilde{Q}_j(\mathbf{x})$ and Q_j are pairwise distinct for $j \in [t]$, we have that $\widetilde{Q}_j$ are pairwise distinct for $j \in [t]$. Moreover for the function $\widetilde{f} : [s]^n \to G_i$ defined as $\widetilde{f}(\mathbf{x}) = f_i(\mathbf{x}) - \widehat{Q}(\mathbf{x})$, we have that $\delta(\widetilde{f}, \widetilde{Q}_j) = \delta(f_i, Q_j) \leqslant 1/s^d - \varepsilon$. Therefore, we get $t \leqslant L$ as H_i is isomorphic to $\mathbb{Z}_p$ and we have a list size bound of L for junta-sums over $\mathbb{Z}_p$. This proves Item 1 of the claim. To prove Item 3, let $j_1 \neq j_2 \in [t]$ be arbitrary and let $A_1, A_2 \subseteq [s]^n$ be the subset of points where f_i agrees with Q_{j_1} and Q_{j_2} respectively. Let $A \subseteq [s]^n$ be the subset of points where f_{i+1} agrees with P. From the proof of Item 2, we have that $A_1, A_2 \subseteq A$. Since two distinct d-junta-sums cannot agree on more than $1 - 1/s^d$ fraction of inputs (Claim 2.6), we have $|A_1 \cap A_2| \leqslant (1 - 1/s^d)s^n$. Hence, $|A| \geqslant |A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2| \geqslant s^n \left(2 - \delta(f_i, Q_{j_1}) - \delta(f_i, Q_{j_2}) - 1 + 1/s^d\right)$. Since $|A| = s^n \left(1 - \delta(f_{i+1}, P)\right)$, we get $\delta(f_{i+1}, P) \leqslant \delta(f_i, Q_{j_1}) + \delta(f_i, Q_{j_2}) - 1/s^d$, or equivalently $\rho(u) \geqslant \rho(v_{j_1}) + \rho(v_{j_2})$. ■

Having reduced the problem to showing combinatorial bound over $\mathbb{F}_q$, which we state below and prove in the next subsection.

Theorem D.9. *For every $\varepsilon > 0$, finite field $\mathbb{F}_q$ and subset $S \subseteq \mathbb{F}_q$ of size $s \geqslant 2$, the family $\mathcal{J}_d(S^n, \mathbb{F}_q)$ is $(1/s^d - \varepsilon, O_{q,\varepsilon}(1))$-list-decodable.*

With the above theorem, we can now finish the proof of the combinatorial bound for p-primary groups.

Proof of Theorem D.2. The proof follows by combining Lemma D.7, Lemma D.6 and Theorem D.9. ■

D.2.2 Combinatorial Bound for $\mathbb{F}_q$

Throughout this subsection, we fix a finite field $\mathbb{F}_q$ and a subset $S \subseteq \mathbb{F}_q$ of size s arbitrarily. We think of $q \geqslant 2, s \geqslant 2$ and $d \geqslant 1$ as constants. Furthermore, we fix a monomial ordering (denoted $\preceq$) over the monomials to be the *graded lexicographic order* (see [ABPSS25] for a definition) and denote by $\mathrm{LM}(P)$ the leading monomial of a polynomial P (assuming it is non-zero). We will use the notation $m_1 \succeq m_2$ to mean that $m_2 \preceq m_1$ and $m_1 \succnsim m_2$ to mean that $m_2 \preceq m_1$ and $m_1 \neq m_2$.

We show the following lemma which effectively reduces the list-decoding problem to bounding the number of polynomials in the list with pairwise distinct monomials.

Lemma D.10 (Distinct leading monomials). *If $P_1, \ldots, P_t \in \mathcal{J}_d(S^n, \mathbb{F}_q)$ are such that $\delta(f, P_i) \leqslant 1/s^d - \varepsilon$ for all $i \in [t]$ and some $f : S^n \to \mathbb{F}_q$, then there exists a function $f' : S^n \to \mathbb{F}_q$ such that there are at least $\ell \geqslant \Omega(\log_q t)$ many polynomials $Q_1, \ldots, Q_\ell \in \mathcal{J}_d(S^n, \mathbb{F}_q)$ such that $\delta(f', Q_i) \leqslant 1/s^d - \varepsilon$ and $\mathrm{LM}(Q_i)$ are pairwise distinct for $i \in [\ell]$.*

Then, we prove the following "tail bound" for polynomials with *pairwise disjoint* leading monomials.

Lemma D.11 (Tail bound for disjoint leading monomials). *Let $P_1, \ldots, P_t \in \mathcal{J}_d(S^n, \mathbb{F}_q)$ be such that $\mathrm{LM}(P_i)$ are pairwise disjoint for $i \in [t]$. Then:*

$$\Pr_{\mathbf{a} \sim S^n}\left[\left|\{i \in [t] : P_i(\mathbf{a}) = 0\}\right| \geqslant (1 - 1/s^d + \eta)t\right] \leqslant \exp(-\Omega(\eta^2 t)).$$

With the above lemmas in place, we are ready to finish the proof of the main result of this subsection i.e., Theorem D.9.

Proof of Theorem D.9. Using Lemma D.10, it suffices to show that if $Q_1, \ldots, Q_\ell \in \mathcal{J}_d(S^n, \mathbb{F}_q)$ are such that $\mathrm{LM}(Q_i)$ are pairwise distinct for $i \in [\ell]$ and $\delta(Q_i, f) \leqslant 1/s^d - \varepsilon$ for some $f : S^n \to \mathbb{F}_q$, then that $\ell = O_{q,\varepsilon}(1)$. Applying the sunflower lemma (see e.g. [ER60]) for the multisets determined by the leading monomials of Q_i's, we can find a subset of indices $i_1, \ldots, i_{\ell'}$ for some $\ell' \geqslant \Omega_d(\ell^{1/d})$ such that $\mathrm{LM}(Q_{i_j})$ form a sunflower for $j \in [\ell']$. That is, there exists variables $\{x_j : j \in C\}$ and $\{e_j \in \mathbb{Z}_s : j \in C\}$ where $C \in \binom{[n]}{\leqslant d}$ such that $\mathrm{LM}(Q_{i_j}) = \prod_{j' \in C} x_{j'}^{e_{j'}} \cdot m_j$ and m_j are monomials over the variables indexed by $[n] \backslash C$ and are pairwise disjoint for $j \in [\ell']$. Without loss of generality, we will assume that $i_j = j$ for all $j \in [\ell']$. We will now express each Q_i for $i \in [\ell']$ (uniquely) as follows:

$$Q_i(\mathbf{x}) = \prod_{j \in C} x_j^{e_j} \cdot Q_i^{(1)}(\mathbf{x}_{[n]\backslash C}) + Q_i^{(2)}(\mathbf{x}),$$

where $Q_i^{(1)}$ is a polynomial over variables indexed by $[n] \backslash C$ and $Q_i^{(2)}$ does not contain any monomial dividing $\prod_{j \in C} x_j^{e_j}$. We note that since the leading monomial of Q_i is $\prod_{j \in C} x_j^{e_j} \cdot m_i$, by our definition of monomial ordering, we must have that $\mathrm{LM}(Q_i^{(1)}) = m_i$. Let $\mathbf{a} \sim S^n$ be sampled by first choosing $\mathbf{a}' \sim S^{[n]\backslash C}$ uniformly at random and then $\mathbf{a}'' \sim S^C$ uniformly and independently. Letting $d' = d - |C|$, we may now apply the tail bound (Lemma D.11) to $Q_i^{(1)}$ to get that

$$\Pr_{\mathbf{a}'}\left[\left|\{i \in [t] : Q_i^{(1)}(\mathbf{a}') = 0\}\right| \geqslant (1 - 1/s^{d'} + \varepsilon/2)\ell'\right] \leqslant \exp(-\Omega(\varepsilon^2 \ell')).$$

In fact, by applying it to $Q_i^{(1)} - \alpha$ for $\alpha \in \mathbb{F}_q$ and by a union bound, we get that

$$\Pr_{\mathbf{a}'}\left[\exists \alpha \in \mathbb{F}_q \text{ such that } \left|\{i \in [t] : Q_i^{(1)}(\mathbf{a}') = \alpha\}\right| \geqslant (1 - 1/s^{d'} + \varepsilon/2)\ell'\right] \leqslant q \cdot \exp(-\Omega(\varepsilon^2 \ell')). \quad (38)$$

Now, let us use the notation

$$Q_i'(\mathbf{x}_C) = Q_i(\mathbf{x}_C, \mathbf{a}')$$

to denote the corresponding restricted functions obtained by setting the variables in $[n] \backslash C$ to $\mathbf{a}'$. Similarly, let $f' : S^C \to \mathbb{F}_q$ be the restricted function $f'(\mathbf{x}_C) = f(\mathbf{x}_C, \mathbf{a}')$. For a uniformly random choice of $\mathbf{a}'$, let $\mathcal{B}$ denote the "bad" event that the *multiset* of functions $\{Q_i' : i \in [\ell']\}$ has a function occurring at least $(1 - 1/s^{d'} + \varepsilon/2)\ell'$ many times. The bound from Equation (38) immediately implies that

$$\Pr_{\mathbf{a}'}[\mathcal{B}] \leqslant q \cdot \exp(-\Omega(\varepsilon^2 \ell')). \quad (39)$$

Conditioned on $\mathcal{B}$ not occurring, we note that there are at least $(1/s^{d'} - \varepsilon/2)\ell'$ indices $i \in [\ell']$ such that $Q'_i \neq f'$ as functions over S^C. More formally,

$$\Pr_{i\sim[\ell']}\left[Q'_i \neq f' \;\middle|\; \overline{\mathcal{B}}\right] \geqslant 1/s^{d'} - \varepsilon/2.$$

Since two different functions over $|C|$ variables must differ on a random input with probability at least $1/s^{|C|}$, we further get:

$$\Pr_{i\sim[\ell'],\mathbf{a}''}\left[Q'_i(\mathbf{a}'') \neq f'(\mathbf{a}'') \;\middle|\; \overline{\mathcal{B}}\right] \geqslant \left(\frac{1}{s^{d'}} - \frac{\varepsilon}{2}\right)\frac{1}{s^{|C|}} \geqslant \frac{1}{s^d} - \frac{\varepsilon}{2}. \tag{40}$$

Combining (39) and (40), we obtain

$$\Pr_{i\sim[\ell'],\mathbf{a}\sim S^n}\left[Q_i(\mathbf{a}) \neq f(\mathbf{a})\right] \geqslant \frac{1}{s^d} - \frac{\varepsilon}{2} - \frac{q}{2^{\Omega(\varepsilon^2\ell')}}.$$

However, we note that since $\delta(Q_i, f) \leqslant 1/s^d - \varepsilon$ for all $i \in [\ell']$, the left hand side of the above inequality must be at most $\frac{1}{s^d} - \varepsilon$. Put together, they give the required bound of $\ell' = O_{q,\varepsilon}(1)$, and thus $\ell = O_{q,\varepsilon}(1)$. ∎

We now give the proofs of the above two lemmas. First, we start with the reduction to counting polynomials with distinct leading monomials, i.e., Lemma D.10.

Proof of Lemma D.10. Let ℓ be an integer such that $t \in [q^\ell, q^{\ell+1})$ (so we have $\ell \geqslant \Omega(\log_q t)$). We will prove the following inductive claim. We recall that $L_\varepsilon(f)$ denotes the set of polynomials in $\mathcal{J}_d(S^n, \mathbb{F}_q)$ that are at distance at most $1/s^d - \varepsilon$ from the function f.

Inductive claim. For every $0 \leqslant i \leqslant \ell$, there exists a function $f_i : S^n \to \mathbb{F}_q$, polynomials $Q_1, Q_2, \ldots, Q_i \in \mathcal{J}_d(S^n, \mathbb{F}_q)$, and a set of polynomials $\mathcal{Q}_i \subseteq \mathcal{J}_d(S^n, \mathbb{F}_q)$ such that:

- $Q_1, \ldots, Q_i \in L_\varepsilon(f_i)$ and $Q \in L_\varepsilon(f_i)$ for all $Q \in \mathcal{Q}_i$,
- $\mathrm{LM}(Q_1) \gneqq \mathrm{LM}(Q_2) \gneqq \cdots \gneqq \mathrm{LM}(Q_i) \gneqq \mathrm{LM}(Q)$ for all $Q \in \mathcal{Q}_i$, and
- $|\mathcal{Q}_i| \geqslant q^{\ell-i}$.

We note that the base case $i = 0$ is true with $f_0 = f$ and $\mathcal{Q}_0 = \{P_1, P_2, \ldots, P_t\}$. And proving the inductive claim for $i = \ell$ finishes the proof of Lemma D.10. We now assume the inductive claim holds for a fixed $i < \ell$ and prove it for $i + 1$.

Let $P \in \mathcal{J}_d(S^n, \mathbb{F}_q)$ be the "plurality polynomial" of $\mathcal{Q}_i$, i.e., we determine each coefficient of P by taking a plurality vote of the corresponding coefficients from the polynomials in $\mathcal{Q}_i$ (by breaking ties arbitrarily). We then define $f_{i+1} : S^n \to \mathbb{F}_q$ to be

$$f_{i+1} = f_i - P.$$

We now define $Q'_1, \ldots, Q'_i, Q'_{i+1} \in \mathcal{J}_d(S^n, \mathbb{F}_q)$ and $\mathcal{Q}'_{i+1} \subseteq \mathcal{J}_d(S^n, \mathbb{F}_q)$ such that the three items of the inductive claim hold for them. We let $\mathcal{Q}'_i = \{Q - P : Q \in \mathcal{Q}_i\}$ and set $Q'_j = Q_j - P$ for $j \in [i]$. We now set Q'_{i+1} to be a polynomial from $\mathcal{Q}'_i$ with the greatest leading monomial (ignoring the

zero polynomial if it exists and breaking ties arbitrarily). Then we set $\mathcal{Q}'_{i+1}$ to be the subset of polynomials in $\mathcal{Q}'_i$ with leading monomial strictly smaller than that of Q'_{i+1}, i.e.:

$$\mathcal{Q}'_{i+1} = \{Q' \in \mathcal{Q}'_i : \mathrm{LM}(Q') \precnsim \mathrm{LM}(Q'_{i+1})\}.$$

It remains to prove that the three conditions of the inductive claim actually hold for the above definitions.

- We have that $\delta(Q'_j, f_{i+1}) = \delta(Q_j - P, f_i - P) = \delta(Q_j, f_i) \leqslant 1/s^d - \varepsilon$, therefore $Q'_j \in L_\varepsilon(f_{i+1})$ for all $j \in [i]$. Similarly, we have $Q'_{i+1} \in L_\varepsilon(f_{i+1})$ and $Q' \in L_\varepsilon(f_{i+1})$ for all $Q' \in \mathcal{Q}'_{i+1}$.
- We note that $\mathrm{LM}(P) \precnsim \mathrm{LM}(Q_i)$ since the coefficients of all monomials $m \succeq \mathrm{LM}(Q_i)$ in all $Q \in \mathcal{Q}'_i$ (and hence in P) are zero by the induction hypothesis. Therefore, $\mathrm{LM}(Q'_j) = \mathrm{LM}(Q_j)$ for $j \in [i]$ and we have $\mathrm{LM}(Q'_1) \succnsim \mathrm{LM}(Q'_2) \succnsim \cdots \succnsim \mathrm{LM}(Q'_i)$. It also follows that $\mathrm{LM}(Q'_i) \succnsim \mathrm{LM}(Q'_{i+1})$ and $\mathrm{LM}(Q'_{i+1}) \succnsim \mathrm{LM}(Q)$ for all $Q \in \mathcal{Q}'_{i+1}$ by the definitions of Q'_{i+1} and $\mathcal{Q}'_{i+1}$.
- We have that $|\mathcal{Q}'_i| = |\mathcal{Q}_i| \geqslant q^{\ell-i}$ by induction hypothesis. By the definition of Q'_{i+1}, we observe that $\mathrm{LM}(Q') \preceq \mathrm{LM}(Q'_{i+1})$ for all $Q' \in \mathcal{Q}'_i$ and we will show that at least $1/q$ fraction of Q''s have leading monomial *strictly* smaller than that of Q'_{i+1}. By the nature of the construction of P using the plurality vote, we observe that at least $1/q$ fraction of the polynomials $Q \in \mathcal{Q}'_i$ agree with P on the coefficient $\mathrm{LM}(Q'_{i+1})$. The corresponding polynomials $Q' = Q - P$ have coefficient of $\mathrm{LM}(Q_{i+1})$ as zero. In other words there are at least $|\mathcal{Q}_i|/q$ polynomials $Q' \in \mathcal{Q}'_i$ with leading coefficient strictly smaller than $\mathrm{LM}(Q'_{i+1})$, and hence by our definition of $\mathcal{Q}'_{i+1}$, it must be of size $|\mathcal{Q}'_{i+1}| \geqslant |\mathcal{Q}'_i|/q \geqslant q^{\ell-(i+1)}$.

This finishes the proof of the inductive claim. ■

We now prove the tail bound for polynomials with pairwise disjoint leading monomials (Lemma D.11).

Proof of Lemma D.11. We will use the following theorem of Panconesi and Srinivasan [PS97] which reduces the task of showing tail bounds to proving a certain "independence" relation among the events.

Theorem D.12 ([PS97] Theorem 3.4)**.** *Let $Z_1, \ldots, Z_t$ be Boolean random variables and $\alpha \in [0,1]$ be such that for every subset $S \subseteq [t]$, we have that $\Pr[\bigwedge_{i \in S} Z_i = 1] \leqslant \alpha^{|S|}$. Then, for every $\eta > 0$, we have*

$$\Pr\left[\sum_{i \in [t]} Z_i \geqslant (\alpha + \eta)t\right] \leqslant \exp(-\Omega(\eta^2 t)).$$

Because of the above theorem, it suffices to show the following: for every $t \in \mathbb{N}$ and non-zero polynomials $P_1, \ldots, P_t \in \mathcal{J}_d(S^n, \mathbb{F}_q)$ for which $\mathrm{LM}(P_i)$ are pairwise disjoint for $i \in [t]$, that:

$$\Pr_{\mathbf{a} \sim S^n}\left[P_i(\mathbf{a}) = 0 \ \text{ for all } i \in [t]\right] \leqslant \left(1 - \frac{1}{s^d}\right)^t. \tag{41}$$

The proof follows the same *footprint bound* technique as used in [ABPSS25]. In particular, letting

$$Z = \{\mathbf{a} \in S^n : P_i(\mathbf{a}) = 0 \text{ for all } i \in [t]\}$$

denote the set of common zeroes, we will prove an upper bound on the dimension of all functions from Z to $\mathbb{F}_q$ i.e., $|Z|$. Let $f : Z \to \mathbb{F}_q$ be an arbitrary function. We will show that it can be expressed as a polynomial of individual degree at most $s-1$ over $\mathbb{F}_q$, without using any monomial divisible by any of the $\mathrm{LM}(P_i)$ for $i \in [t]$. That is, these are the monomials $\mathbf{x}^{\mathbf{e}}$ for $\mathbf{e} \in E$, where $E \subseteq \mathbb{Z}_s^n$ is defined below, and $\mathrm{LM}(P_i) = \mathbf{x}^{\mathbf{m}_i}$ for some $\mathbf{m}_i \in \mathbb{Z}_s^n$[15]:

$$E = \{\mathbf{e} \in \mathbb{Z}_s^n : \forall i \in [t]\ \exists j \in [n]\ e_j < m_{i,j}\}.$$

Using the fact that the supports of $\mathbf{m}_i$ are pairwise disjoint (over $i \in [t]$) and are of size at most d, we get that $|E| \leqslant \left(1 - \frac{1}{s^d}\right)^t \cdot s^n$. We will show that there exists field elements $(c_{\mathbf{e}})_{\mathbf{e} \in E}$ such that $f(\mathbf{x}) = \sum_{\mathbf{e} \in E} c_{\mathbf{e}} \cdot \mathbf{x}^{\mathbf{e}}$ for all $\mathbf{x} \in Z$. This shall finish the proof of (41) and thus Lemma D.11 as it shows that $|Z| \leqslant |E| \leqslant \left(1 - \frac{1}{s^d}\right)^t \cdot s^n$. Hence, it remains to prove that f can be expressed as a linear combination of monomials in E. Since $Z \subseteq S^n$, we know that there exists a polynomial representation for f of individual degree at most $s-1$: suppose $Q(\mathbf{x}) = \sum_{\mathbf{e} \in \mathbb{Z}_s^n} c_{\mathbf{e}} \cdot \mathbf{x}^{\mathbf{e}}$ for some $c_{\mathbf{e}} \in \mathbb{F}_q$ is such that $f(\mathbf{x}) = Q(\mathbf{x})$ for all $\mathbf{x} \in Z$. If $c_{\mathbf{e}} = 0$ for all $\mathbf{e} \notin E$, we are done. Otherwise, there exists an $i \in [t]$ such that $\mathrm{LM}(P_i)$ divides $\mathbf{x}^{\mathbf{e}'}$ for some $\mathbf{e}'$ such that $c_{\mathbf{e}'} \neq 0$ (say that $\mathbf{x}^{\mathbf{e}'} = \mathrm{LM}(P_i) \cdot \mathbf{x}^{\mathbf{e}''}$); w.l.o.g. let $\mathbf{x}^{\mathbf{e}'}$ be the largest monomial in the monomial ordering such that this holds. Then, we note that we can replace the monomial $\mathbf{x}^{\mathbf{e}'}$ with the polynomial $\mathbf{x}^{\mathbf{e}''}(\mathrm{LM}(P_i) - P_i/c)$ in the polynomial $\sum_{\mathbf{e} \in \mathbb{Z}_s^n} c_{\mathbf{e}} \cdot \mathbf{x}^{\mathbf{e}}$ while still computing f, where $c \in \mathbb{F}_q^\times$ is the coefficient of $\mathrm{LM}(P_i)$ in P_i. This is due to the fact that P_i (and therefore P_i/c) evaluates to 0 over Z. Let Q' denote the polynomial obtained by such a transformation. We claim that $\mathrm{LM}(Q') \lneq \mathrm{LM}(Q)$. This is because all the new non-zero monomials introduced by the transformation are of the form $\mathbf{x}^{\mathbf{e}''} \cdot \mathbf{x}^{\mathbf{e}'''}$ for some $\mathbf{x}^{\mathbf{e}'''} \lneq \mathrm{LM}(P_i)$, and so $\mathbf{x}^{\mathbf{e}''} \cdot \mathbf{x}^{\mathbf{e}'''} \lneq \mathbf{x}^{\mathbf{e}''} \cdot \mathrm{LM}(P_i) = \mathbf{x}^{\mathbf{e}'} = \mathrm{LM}(Q)$ using the monomial ordering property. Hence, $\mathrm{LM}(Q') \lneq \mathrm{LM}(Q)$. While the leading monomial of the polynomial computing has decreased, it may be possible that Q' contains monomials with individual degree at least s. We now argue that we can design a new polynomial Q'' such that $\mathrm{LM}(Q'') \preceq \mathrm{LM}(Q')$ and $Q''(\mathbf{x}) = Q'(\mathbf{x}) = f(\mathbf{x})$ for all $\mathbf{x} \in Z$. The idea is to use the equation $\prod_{a \in S}(x_i - a) = 0$ to replace the powers of the variable x_i greater than $s-1$ with smaller powers — this only results in monomials that are smaller in the monomial order. Thus, by repeating the above two steps for a finite number of times, we will have a polynomial representing f only using monomials from E. ■

[15] Here we treat $\mathbb{Z}_s^n = \{0, 1, \ldots, s-1\}^n$ as a subset of $\mathbb{Z}^n$ and define the monomial $\mathbf{x}^{\mathbf{m}} = \prod_{i \in [n]} x_i^{m_i}$.

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il