SIAM J. Comput.
Vol. 44, No. 5, pp. 1287–1324

MAKING THE LONG CODE SHORTER*

BOAZ BARAK[†], PARIKSHIT GOPALAN[‡], JOHAN HÅSTAD[§], RAGHU MEKA[¶], PRASAD RAGHAVENDRA[‖], AND DAVID STEURER[#]

Abstract. The *long code* is a central tool in hardness of approximation, especially in questions related to the Unique Games Conjecture. We construct a new code that is exponentially more efficient, but can still be used in many of these applications. Using the new code we obtain exponential improvements over several known results, including the following: (1) For any $\varepsilon > 0$, we show the existence of an n-vertex graph G where every set of $o(n)$ vertices has expansion $1-\varepsilon$, but G's adjacency matrix has more than $\exp(\log^{\delta} n)$ eigenvalues larger than $1-\varepsilon$, where δ depends only on ε. This answers an open question of Arora, Barak, and Steurer [*Proceedings of the* 2010 *IEEE* 51*st Annual Symposium on Foundations of Computer Science*, 2010, pp. 563–572], who asked whether one can improve over the noise graph on the Boolean hypercube that has $\mathrm{poly}(\log n)$ such eigenvalues. (2) A gadget that reduces Unique Games instances with linear constraints modulo K into instances with alphabet k with a blowup of $k^{\mathrm{polylog}(K)}$, improving over the previously known gadget with blowup of $k^{\Omega(K)}$. (3) An n-variable integrality gap for Unique Games that survives $\exp(\mathrm{poly}(\log\log n))$ rounds of the semidefinite programming version of the Sherali–Adams hierarchy, improving on the previously known bound of $\mathrm{poly}(\log\log n)$. We show a connection between the local testability of linear codes and Small-Set Expansion in certain related Cayley graphs and use this connection to derandomize the noise graph on the Boolean hypercube.

Key words. hardness of approximation, Cayley graphs, locally testable codes, expanders

AMS subject classification. 68Q17

DOI. 10.1137/130929394

1. Introduction.

Khot's *Unique Games Conjecture (UGC)* [Kho02] has been the focus of intense research effort in the last few years. The conjecture posits the hardness of approximation for a certain constraint satisfaction problem and shows promise for settling many open questions in theory of approximation algorithms. Specifically, an instance Γ of the Unique Games problem with n variables and alphabet Σ is described by a collection of constraints of the form (x, y, π), where π is a permutation over Σ. An *assignment* to Γ is a mapping f from $[n]$ to Σ, and f's value is the fraction of constraints (x, y, π) such that $f(y) = \pi(f(x))$. The UGC is that for any $\varepsilon > 0$, there is some finite Σ such that it is **NP** hard to distinguish between the case that a Unique Games instance Γ with alphabet Σ has an assignment satisfying a $1-\varepsilon$ fraction of the constraints, and the case that every assignment satisfies at most a ε fraction of Γ's constraints.

Many works have been devoted to studying the plausibility of the UGC, as well as exploring its implications and obtaining unconditional results motivated by this

*Received by the editors July 16, 2013; accepted for publication (in revised form) June 29, 2015; published electronically October 27, 2015. An extended abstract of this paper appeared in *Proceedings of the* 2012 *IEEE* 53*rd Annual Symposium on Foundations of Computer Science*.
http://www.siam.org/journals/sicomp/44-5/92939.html

[†] Microsoft Research New England, Cambridge, MA 02142 (b@boazbarak.org).

[‡] Microsoft Research, Mountain View, CA 94043 (parik@microsoft.com).

[§] Royal Institute of Technology, Stockholm S-100 44, Sweden (johanh@kth.se).

[¶] University of California, Los Angeles, Los Angeles, CA 94043 (raghuvardhan@gmail.com). The work of the fourth author was done in part while visiting Microsoft Research, Silicon Valley.

[‖] University of California, Berkeley, Berkeley, CA 94720 (prasad@cs.berkeley.edu). The work of the fifth author was done while at the Georgia Institute of Technology, Atlanta, GA.

[#] Cornell University, Ithaca, NY 14853 (dsteurer@cs.cornell.edu). The work of the sixth author was done while at Microsoft Research New England, Cambridge, MA.

effort. Tantalizingly, at the moment we have very little evidence for the truth of this conjecture. One obvious reason to believe the UGC is that no algorithm is known to contradict it, though that of course may have more to do with our proof techniques for algorithm analysis than actual computational difficulty. Thus perhaps the strongest evidence for the conjecture comes from results showing particular instances on which certain natural algorithms will fail to solve the problem. However, even those integrality gaps are quantitatively rather weak. For example, while Arora, Barak, and Steurer [ABS10] showed a subexponential upper bound on an algorithm for Unique Games and the related Small-Set Expansion problem, the hardest known instances for their algorithm only required quasi-polynomial time [Kol10]. Similarly (and related to this), known integrality gaps for Unique Games and related problems do not rule out their solution by an $O(\log n)$-round semidefinite hierarchy, an algorithm that can be implemented in quasi-polynomial (or perhaps even polynomial [BRS11]) time.

The *long code* has been a central tool in many of these works. This is the set of "dictator" functions mapping $\mathbb{F}_2^N$ to $\mathbb{F}_2$ that have the form $(x_1, \ldots, x_N) \mapsto x_i$ for some i. Many hardness reductions (especially from Unique Games) and constructions of integrality gap instances use the long code as a tool. However, this is also the source of their inefficiency, as the long code is indeed quite long. Specifically, it has only N codewords but dimension 2^N, which leads to exponential blowup in many of these applications. In this work, we introduce a different code, which we call the "short code," that is exponentially more efficient and can be used in the long code's place in many of these applications, leading to significant quantitative improvements. In particular, we use our code to show instances on which the [ABS10] algorithm, as well as certain semidefinite hierarchies, requires almost subexponential time, thus considerably strengthening the known evidence in support of the UGC.

1.1. Our results.

At the heart of the long code's applications lies its connection with the *noisy hypercube*. This is the weighted graph $H_{N,\varepsilon}$ whose vertices are elements in $\mathbb{F}_2^N$, where a random neighbor of $x \in \mathbb{F}_2^N$ is obtained by flipping each bit of x independently with probability ε.[1] It is not too hard to show that the codewords of the long code correspond to the top eigenvectors of the noisy hypercube, which also give the minimal bisections of the graph, cutting only a ε fraction of edges. In addition, several converse results are known, showing that bisections (and more general functions) cutting few edges are close to these top eigenvectors (or *dictatorships*) in some sense. (One such result is the "Majority Is Stablest" theorem of [MOO05].) The inefficiency of the long code is manifested in the fact that the number of vertices of the noisy cube is exponential in the number N of its top eigenvectors.

The short code. Another way to describe the long code is that it encodes $x \in \mathbb{F}_2^n$ by a binary vector v_x of length 2^{2^n}, where $v_x(f) = f(x)$ for every function $f : \mathbb{F}_2^n \to \mathbb{F}_2$. This view also accounts for the name "long code," since one can see that this is the longest possible encoding of x without having repeated coordinates. For every subset $\mathcal{D}$ of functions mapping $\mathbb{F}_2^n$ to $\mathbb{F}_2$, we define the *$\mathcal{D}$-short code* to be the code that encodes x by a vector v_x of length $|\mathcal{D}|$, where $v_x(f) = f(x)$ for every $f \in \mathcal{D}$. Note that this is a very general definition that encapsulates any code without repeated coordinates. For $d \in \mathbb{N}$, we define the *d-short code* to be the $\mathcal{D}$-short code where $\mathcal{D}$ is the set of all polynomials over $\mathbb{F}_2^n$ of degree at most d. Note that the 1-short code is the Hadamard code, while the n-short code is the long code. We use the

[1] This graph is closely related and has similar properties to the unweighted graph where we connect x and y if their Hamming distance is at most εN.

name "short code" to denote the d-short code for $d = O(1)$. Note that the short code has 2^n codewords and dimension roughly 2^{n^d}, and hence only quasi-polynomial blowup, as opposed to the exponential blowup of the long code. Our main contribution is a construction of a "derandomized" noisy cube, which is a small subgraph of the noisy cube that enjoys the same relations to the short code (including a Majority Is Stablest theorem) as the original noisy cube has to the long code. As a result, in many applications one can use the short code and the derandomized cube in place of the long code and the noisy cube, obtaining an exponential advantage. Using this approach we obtain the following results.

Small-set expanders with many large eigenvalues. Our first application, and the motivation to this work, is a question of Arora, Barak, and Steurer [ABS10]: How many eigenvectors with eigenvalue at least $1 - \varepsilon$ can an n-vertex *small-set expander* graph have? We say a graph is a small-set expander if all sufficiently small subsets of vertices have, say, at least a 0.9 fraction of their neighbors outside the set. [ABS10] showed an upper bound of $n^{O(\varepsilon)}$ on the number of large (i.e., greater than $1 - \varepsilon$) eigenvalues of a small-set expander. Arora, Barak, and Steurer then observed that the subspace enumeration algorithm of [KT07, Kol10] for approximating SMALL-SET EXPANSION in an input graph takes time at most exponential in this number, which they then use to give an algorithm with similar running time for the UNIQUE GAMES problem. Up to this work, the best lower bound was polylog(n), with the example being the noisy cube, and hence as far as we knew, the algorithm of [ABS10] could solve the SMALL-SET EXPANSION problem in quasi-polynomial time, which in turn might have had significant implications for the UNIQUE GAMES problem as well. Our derandomized noisy cube yields an example with roughly exponentially more eigenvalues than the noisy cube.

THEOREM 1.1. *For every $\varepsilon > 0$, there is an n-vertex small-set expander graph with $2^{(\log n)^{\Omega(1/\log(1/\varepsilon))}}$ eigenvectors with corresponding eigenvalues at least $1 - \varepsilon$.*

Theorem 1.1 actually follows from a more general result connecting locally testable codes to small-set expanders, which we instantiate with the Reed–Muller code. See section 2 for details.

Efficient integrality gaps. There is a standard semidefinite programming (SDP) relaxation for the UNIQUE GAMES problem, known as the "basic SDP" [FL92, KV05, RS09]. Several works have shown upper and lower bounds on the approximation guarantees of this relaxation, and for constant alphabet size, the relation between the alphabet size and approximation guarantee is completely understood [CMM06]. However, for an unbounded alphabet, there was still a big gap in our understanding of the relation between the approximation guarantee and the number of variables. Gupta and Talwar [GT06] showed that if the relaxation's value is $1 - \varepsilon$, there is an assignment satisfying a $1 - O(\varepsilon \log n)$ fraction of constraints. On the other hand, Khot and Vishnoi [KV05] gave an integrality gap instance where the relaxation's value was $1 - 1/\operatorname{poly}(\log\log n)$,[2] but the objective value (maximum fraction of constraints satisfied by any assignment) was $o(1)$. It was a natural question whether this could be improved (e.g., see [Lee11]), and indeed our short code allows us to obtain an almost exponential improvement.

[2] Throughout, for any function f, poly$(f(n))$ denotes a function g satisfying $g(n) = f(n)^{\Omega(1)}$.

Theorem 1.2. *There is an n-variable instance of* Unique Games *with objective value $o(1)$ but for which the standard SDP relaxation has value at least $1 - 1/\operatorname{qpolylog}(n)$.*[3]

Integrality gaps for SDP hierarchies. Our best evidence for the hardness of the UGC comes from integrality gap instances for SDP *hierarchies*. These are strengthened versions of the basic SDP where one obtains tighter relaxations by augmenting them with additional constraints; we refer the reader to [CT10] for a good overview of SDP hierarchies. These hierarchies are generally parametrized by a number r (often called the *number of rounds*), where the first round corresponds to the basic SDP, and the nth round (where n is the instance size) corresponds to the exponential brute force algorithm that always computes an optimal answer. Generally, the rth round of each such hierarchy can be evaluated in $n^{O(r)}$ time (though in some cases $n^{O(1)}2^{O(r)}$ time suffices [BRS11]). In this paper we consider two versions of these hierarchies—the SA-SDP hierarchy and the weaker LH hierarchy defined in the work of Raghavendra and Steurer [RS09]. Loosely speaking, the rth round of the SA-SDP hierarchy adds the constraints of the rth round of the Sherali–Adams linear programming hierarchy (see [SA90]) to the basic SDP; the rth round of the LH hierarchy (here LH stands for "local constraints hierarchy") augments the basic SDP with the constraint that every subset of r vectors from the vector solution embeds isometrically into the ℓ_1 metric. (See Appendix B and [RS09] for more details.)

Barak, Raghavendra, and Steurer [BRS11] (see also [GS11]) showed that for every $\varepsilon > 0$, n^{ε} rounds of the SA-SDP hierarchy yield a nontrivial improvement over the basic SDP. The UGC predicts that this is optimal, in the sense that $n^{o(1)}$ rounds of any hierarchy should not improve the worst-case approximation ratio above the basic SDP.[4] However, this prediction is far from being verified, with the best lower bounds given by [RS09] (see also [KS09]), which showed instances that require $\log^{\Omega(1)} n$ rounds for the LH hierarchy, and $(\log\log n)^{\Omega(1)}$ rounds for the SA-SDP hierarchy. Moreover, these instances are *known* to be solvable in quasi-polynomial time [Kol10] and, in fact, via $\operatorname{polylog}(n)$ rounds of the SA-SDP hierarchy [BRS11] . Thus prior work gave no evidence that the Unique Games problem cannot be solved in quasi-polynomial time. In this work we obtain almost exponentially more efficient integrality gaps, resisting $\operatorname{qpoly}(\log n)$ rounds of the SA-SDP hierarchy and $\operatorname{qqpoly}(n)$ rounds of the LH hierarchy. The latter is the first superlogarithmic SDP hierarchy lower bound for Unique Games for any SDP hierarchy considered in the literature.

Theorem 1.3. *For every $\varepsilon > 0$ there is some $k = k(\varepsilon)$ such that for every n there is an n-variable instance Γ of* Unique Games *with alphabet size k such that the objective value of Γ is at most ε, but the value on Γ of both $\operatorname{qpoly}(\log n)$ rounds of the $SA - SDP$ hierarchy and $\operatorname{qqpoly}(n)$ rounds of the LH hierarchy is at least $1-\varepsilon$.*

A corollary of the above theorem is a construction of an n-point metric of negative type such that all sets of size up to some $k = \operatorname{qqpoly}(n)$ embed isometrically into ℓ_1, but the whole metric requires $\operatorname{qpolylog}(n)$ distortion to embed into ℓ_1. We remark that Theorem 1.3 actually yields a stronger result than stated here—as a function of k, our results (as was the case with the previous ones) obtain a close to optimal gap between the objective value and the SDP value of these hierarchies; in particular we

[3] For functions $f, g : \mathbb{N} \to [0, \infty)$ we write $f = \operatorname{qpoly}(g)$ if $f = \exp(\operatorname{polylog}(g))$, that is, if there are constants $C > c > 0$ such that for all sufficiently large n, $\exp((\log g(n))^c) \leqslant f(n) \leqslant \exp((\log g(n))^C)$. (Note that we allow $c < 1$, and so $f = \operatorname{qpoly}(g)$ does not imply that $f > g$.) Similarly, we define $\operatorname{qpolylog}(g) = \operatorname{qpoly}(\log g)$ and write $f = \operatorname{qqpoly}(g)$ if $f = \exp(\exp(\operatorname{poly}(\log\log g)))$.

[4] This is under the widely believed assumption that $\mathbf{NP} \not\subseteq \mathbf{Dtime}(\exp(n^{o(1)})$.

show that in the above number of rounds one cannot improve on the approximation factor of the Goemans–Williamson algorithm for Max-Cut. It is a fascinating open question whether these results can be extended to the stronger *Lasserre* hierarchy. Some very recent results of Barak et al. [BHK$^+$11] (obtained subsequent to this work) indicate that new ideas may be needed to do this, since the UNIQUE GAMES instances constructed here and in prior works are not integrality gaps for some absolute constant number of rounds of the Lasserre hierarchy.

Alphabet reduction gadget. Khot et al. [KKMO07] used the long code to show an "alphabet reduction" gadget for UNIQUE GAMES. They showed how to reduce a UNIQUE GAMES instance with some large alphabet K to an instance with an arbitrarily small alphabet. (In particular, they showed how one can reduce arbitrary UNIQUE GAMES instances into binary alphabet instances, which turns out to be equivalent to the Max-Cut problem.) However, quantitatively their result was rather inefficient, incurring an exponential in K blowup of the instance. By replacing the long code with our "short code," we obtain a more efficient gadget, incurring only a *quasi-polynomial* blowup. One caveat is that, because the short code doesn't support arbitrary permutations, this reduction works only for UNIQUE GAMES instances whose constraints are affine functions over $\mathbb{F}_2^k$, where $k = \log K$; however, this class of UNIQUE GAMES seems sufficiently rich for many applications.[5]

THEOREM 1.4. *For every ε there are k, δ, and a reduction that for every ℓ maps any n-variable* UNIQUE GAMES *instance Γ whose constraints are affine permutations over alphabet $\mathbb{F}_2^\ell$ into an $n \cdot \exp(\mathrm{poly}(\ell, k))$-variable* UNIQUE GAMES *instance Γ' of alphabet k such that if the objective value of Γ is larger than $1 - \delta$, then the objective value of Γ' is larger than $1 - \varepsilon$, and if the objective value of Γ is smaller than δ, then the objective value of Γ' is smaller than ε.*

Once again, our quantitative results are stronger than stated, and as in [KKMO07], we obtain a nearly optimal relation between the alphabet size k and the soundness and completeness thresholds. In particular, for $k = 2$ our results match the parameters of the Max-Cut algorithm of Goemans and Williamson. Our alphabet reduction gadget suggests a new approach to proving the UGC by using it as an "inner PCP." For example, one could first show hardness of UNIQUE GAMES with very large alphabet (polynomial or even subexponential in the number of variables) and then applying alphabet reduction. At the very least, coming up with plausible hard instances for UNIQUE GAMES should be easier with a large alphabet.

Remark 1.1. The long code is also used as a tool in applications that do not involve the UGC. On a high level, there are two properties that make the long code useful in hardness of approximation: (i) It has a 2-query test obtained from the noisy hypercube, and (ii) it has many symmetries, and in particular one can read off any function of x from the xth codeword. Our short code preserves property (i) but (as is necessary for a more efficient code) does not preserve property (ii), as one can read off only low degree polynomials of x (also it is symmetric only under affine transformations). We note that if one does not care about property (i) and is happy with a 3-query test, then it's often possible to use the Hadamard code, which is more efficient than the short code (indeed it's essentially equal to the d-short code for $d = 1$). Thus, at least in the context of hardness of approximation, it seems that the applications in which the short code will be most useful are those where property (i) is the crucial one.

[5] For example, because the multiplicative group of the field $\mathbb{F}_{2^n}$ is cyclic, one can represent constraints of the form $x_i - x_j = c_{i,j} \pmod{2^n - 1}$ as linear constraints over $\mathbb{F}_2^n$ (i.e., constraints of the form $x_i = C_{i,j} x_j$, where $C_{i,j}$ is an invertible linear map over $\mathbb{F}_2^n$).

Despite the name “short code,” our code is not the shortest possible code. While in our applications, dimension linear in the number of codewords is necessary (e.g., one can’t have a graph with more eigenvalues than vertices), it’s not clear that the dimension needs to be polynomial. It is a very interesting open question to find shorter codes that can still be used in the above applications.

2. Our techniques.

To explain our techniques we focus on our first application—the construction of a small-set expander with many eigenvalues close to 1. The best way to view this construction is as a derandomization of the noisy hypercube, and so it will be useful to recall why the noisy hypercube itself is a small-set expander.

Recall that the ε-noisy hypercube is the graph $H_{N,\varepsilon}$ whose vertex set is $\{\pm 1\}^N$, where we sample a neighbor of x by flipping each bit independently with probability ε. The eigenvectors in $H_{N,\varepsilon}$ are given by the parity functions $\chi_\alpha(x) = \prod_{i\in\alpha} x_i$ for subsets $\alpha \subseteq [N]$, and the corresponding eigenvalues are $\lambda_\alpha = (1-2\varepsilon)^{|\alpha|}$. Thus λ_α depends only on the degree $|\alpha|$ of χ_α. In particular, the “dictator” functions $\chi_{\{i\}}(x) = x_i$ have eigenvalue $1-2\varepsilon$, and they correspond to balanced cuts (where vertices are partitioned based on the value of x_i) with edge expansion ε. As α increases, λ_α decreases, becoming smaller than a universal constant < 1 at around $|\alpha| = O(1/\varepsilon)$.

Given $f : \{\pm 1\}^N \to \{0,1\}$ which is the indicator of a set S, its Fourier expansion $f(x) = \sum_\alpha \hat{f}(\alpha)\chi_\alpha(x)$ can be viewed as expressing the vector f in the eigenvector basis. The edge expansion of S is determined by the distribution of its Fourier mass; sets where most of the Fourier mass is on large sets will expand well. Given this connection, Small-Set Expansion follows from the fact that the indicator functions of small sets have most of their mass concentrated on large Fourier coefficients. More precisely, a set S of measure μ has most of its Fourier mass on coefficients of degree $\Omega(\log(1/\mu))$. This follows from the so-called (2,4)-hypercontractive inequality for low-degree polynomials—that for every degree-d polynomial f,

$$\mathop{\mathbb{E}}_{x\in\{\pm 1\}^N}[f(x)^4] \leqslant C \mathop{\mathbb{E}}_{x\in\{\pm 1\}^N}[f(x)^2]^2 \tag{2.1}$$

for some C depending only on d. (See section 4.1 for the proof, though some intuition can be obtained by noting that if f is a characteristic function of a set S of measure $\mu = o(1)$, then $\mathbb{E}[f^2]^2 = \mu^2$ and $\mathbb{E}[f^4] = \mu$, and hence (2.1) shows that f cannot be an $O(1)$-degree polynomial.)

By a “derandomized hypercube” we mean a graph on much fewer vertices that still (approximately) preserves the above properties of the noisy hypercube. Specifically, we want to find a very small subset $\mathcal{D}$ of $\{\pm 1\}^N$ and a subgraph G of $H_{N,\varepsilon}$ whose vertex set is $\mathcal{D}$ such that (i) G will have an eigenvalue profile similar to $H_{N,\varepsilon}$ and, in particular, have N eigenvalues close to 1, and (ii) G will be a a small-set expander. To get the parameters we are looking for, we’ll need to have the size of $\mathcal{D}$ be at most $\mathrm{qpoly}(N)$.

A natural candidate is to take $\mathcal{D}$ to be a random set, but it is not hard to show that this will not work.[6] A better candidate might be a linear subspace $\mathcal{D} \subseteq \mathbb{F}_2^N$ that looks suitably pseudorandom. We show that in fact it suffices to choose a subspace $\mathcal{D}$ whose dual $\mathcal{C} = \mathcal{D}^\perp$ is a sufficiently good locally testable code. (We identify $\mathbb{F}_2^N$ with $\{\pm 1\}^N$ via the usual map $(b_1,\ldots,b_N) \mapsto ((-1)^{b_1},\ldots,(-1)^{b_N})$.)

[6] The noisy hypercube places most of its edge weight on pairs with Hamming distance $\varepsilon \cdot N$, which form an exponentially small fraction of all pairs. Therefore, a random subset of subexponential size will essentially contains no edges with high probability.

Our construction requires an asymptotic family of $[N, K, D]_2$ linear codes $\mathcal{C} \subseteq \mathbb{F}_2^N$ with a sufficiently large constant distance D. The codes must be equipped with an εN-query local tester which when given a received word $\alpha \in \mathbb{F}_2^N$ samples a codeword q of weight at most εN from a distribution $\mathcal{T}$ on $\mathcal{C}^\perp$ and accepts if $\langle \alpha, q\rangle = 0$. The test clearly accepts codewords in $\mathcal{C}$; we also require it to reject words that are distance at least $D/10$ from every codeword in $\mathcal{C}$ with probability 0.49. Given such a locally testable code $\mathcal{C}$, we consider the Cayley graph[7] G whose vertices are the codewords of the dual code $\mathcal{D} = \mathcal{C}^\perp$, while the (appropriately weighted) edges correspond to the distribution $\mathcal{T}$. That is, a vertex of G is a codeword $x \in \mathcal{D}$, while a random neighbor of x is obtained by picking a random q from $\mathcal{T}$ and moving to $x + q$.

Because $\mathcal{D}$ is a subspace, it is easy to show that the eigenvectors of G are linear functions of the form $\chi_\alpha(x)$ for $x, \alpha \in \mathbb{F}_2^N$ (where if $\alpha \oplus \alpha' \in \mathcal{C}$, then χ_α and $\chi_{\alpha'}$ are identical on G's vertices). Moreover, from the way we designed the graph, for every $\alpha \in \mathbb{F}_2^n$, the corresponding eigenvalue λ_α is equal to $\mathbb{E}_{q\in\mathcal{T}}[(-1)^{\langle\alpha,q\rangle}] = 1 - 2\,\mathbb{P}_{\mathcal{T}}[\text{Test rejects } \alpha]$. This connection between the spectrum of G and the local testability of $\mathcal{C}$ allows us to invoke machinery from coding theory in our analysis.

From this one can deduce that the eigenvalue spectrum of G does indeed resemble the hypercube in the range close to 1. In particular, for at least half the coordinates i, $\chi_{\{i\}}(x) = x_i$ is a distinct eigenvector with eigenvalue at most $1 - 4\varepsilon$ and gives a bad cut in G (where vertices are partitioned based on the value of x_i). On the other hand, for any eigenvector χ of G, choose α of minimal weight such that $\chi = \chi_\alpha$. Now if $|\alpha| > D/10$, this means that the distance of α from $\mathcal{C}$ is at least $D/10$, which using the testing property implies that $\lambda_\alpha \leqslant 1 - 2 \cdot 0.49 = 0.02$.

If we can show that indicator functions of small sets have most of their Fourier mass on such eigenvectors (with small eigenvalue), this will imply that small sets have good expansion. For small subsets of the hypercube, recall that this is proved using (2,4)-hypercontractivity for low-degree polynomials. The key observation is that the inequality

$$\mathop{\mathbb{E}}_{x\in\mathcal{D}}[f(x)^4] \leqslant C \mathop{\mathbb{E}}_{x\in\mathcal{D}}[f(x)^2]^2 \tag{2.2}$$

still holds for all polynomials f of degree $d < D/4$. This is because the distance of $\mathcal{C}$ is D; hence the distribution of a random x in $\mathcal{D}$ is D-wise independent, which means that the expectation of any polynomial of degree at most D is equal over such x and over a uniform x in $\{\pm 1\}^N$. Thus (2.2) follows from (2.1), completing our proof.

We instantiate this approach by using for $\mathcal{C}$ the Reed–Muller code consisting of polynomials in n variables over $\mathbb{F}_2$ of degree $n - d - 1$. This is a code of distance $D = 2^{d-1}$. We note that the degree $n-d-1$ and hence the rate of the code $\mathcal{C}$ are very high. The graph is over the codewords of $\mathcal{D} = \mathcal{C}^\perp$ that is itself the Reed–Muller code of polynomials over $\mathbb{F}_2^n$ of degree d. Our basic tester consists of selecting a random minimum weight codeword of $\mathcal{D}$.[8] Thus our graph $\mathcal{G}$ has as its vertices the d-degree polynomials over $\mathbb{F}_2^n$ with an edge between all polynomials p, q such that $p - q$ is a product of d linearly independent affine functions (as those are the minimal weight codewords in the Reed–Muller code). We use the optimal analysis of Bhattacharyya et al. [BKS$^+$10] to argue about the local testability of $\mathcal{C}$ which is a high-degree Reed–Muller code. We should note that this test is very closely related to the Gowers

[7] Cayley graphs are usually defined to be unweighted graphs. However, the definition can be generalized straightforwardly to weighted graphs.

[8] For many applications we amplify the success of this tester by selecting a sum of t random such words; this corresponds to taking some power of the basic graph $\mathcal{G}$ described.

uniformity test that was first analyzed in the work of Alon et al. [AKK$^+$05], but our application requires the stronger result from [BKS$^+$10].

2.1. Other applications. We now briefly outline how we use the above tools to obtain more efficient versions of several other constructions such as alphabet reduction gadgets and integrality gaps for Unique Games and other problems.

Efficient integrality gaps for Unique Games. To begin with, the graph we construct can be used to prove Theorem 1.2. Fix a constant $\delta > 0$ (say 1/100). The goal is to construct an M-variable instance Γ of Unique Games where every assignment can satisfy at most a δ fraction of the constraints, but for which the standard SDP relaxation has value of at least $1 - 1/\operatorname{qpoly}(\log M)$. The basic idea is to simply take the graph $\mathcal{G}$ we constructed above and turn it into an instance of Unique Games by considering it to be the *label extended graph* of some Unique Games instance. We now elaborate a bit below, leaving the full details to section 6. Recall that a Unique Games instance Γ with M variables and alphabet Σ is described by a collection of constraints of the form (x, y, π), where π is a permutation over Σ. An *assignment* to Γ is a mapping f from $[M]$ to Σ, and f's value is the fraction of constraints (x, y, π) such that $f(y) = \pi(f(x))$. The *label extended graph* corresponding to Γ is the graph G_Γ over vertices $[M] \times \Sigma$, where for every constraint of the form (x, y, π) and $\sigma \in \Sigma$ we add an edge between (x, σ) and $(y, \pi(\sigma))$. It is not hard to see that an assignment of value $1 - \varepsilon$ corresponds to a subset S containing exactly M of G_Γ's vertices with small expansion (i.e., a ε fraction of the edges from S leave the set). Thus if G_Γ is an expander for sets of measure $1/|\Sigma|$ in G_Γ, then there is no nearly satisfying assignment for the Unique Games instance Γ. In our case, our graph $\mathcal{G}$ has the degree-d polynomials over $\mathbb{F}_2^n$ as its vertices, and we transform it into a Unique Games instance whose variables correspond to degree-d polynomials *without linear terms*. The alphabet Σ consists of all linear functions over $\mathbb{F}_2^n$. We ensure that the graph $\mathcal{G}$ is the label extended graph of Γ by setting the permutations accordingly: Given a polynomial p without a linear term, and a function q that is a product of d affine functions,[9] if we write $q = q' + q''$, where q'' is the linear part of q, then we add a constraint of the form $(p, p + q', \pi)$, where π is the permutation that maps a linear function r into $r + q''$. Some not too difficult calculations show that the top eigenvectors of our graph $\mathcal{G}$ yield a solution for the semidefinite program for Γ (if the top eigenvectors are $f^1, \dots, f^K$, our vector solution will associate with each vertex x the vector $(f^1(x), \dots, f^K(x))$). By choosing carefully the parameters of the graph $\mathcal{G}$, the instance Γ will have SDP value $1 - 1/\operatorname{qpoly}(\log M)$, where M is the number of variables.

Derandomized invariance principle. While hypercontractivity of low-degree polynomials suffices for some applications of the long code, other applications require other theorems, and in particular the *invariance principle*, shown for the hypercube by Mossel, O'Donnell, and Oleszkiewicz [MOO05]. Roughly speaking, their invariance principle says that for "nice" functions f on the vertices of the N-dimensional noisy hypercube, the distribution of $f(x)$, where x is a random vertex, is close to the distribution of $f(y)$, where y consists of N independent standard Gaussian random variables (appropriately extending f to act on $\mathbb{R}^N$). To obtain a more efficient version of these applications, we first show that the same holds even when x is a random vertex in our smaller subset of N-dimensional strings—the Reed–Muller codewords. Our central tool is a recent result by Meka and Zuckerman [MZ13] which derandomizes the

[9] Actually, to get better parameters, we take some power t of $\mathcal{G}$, meaning that we consider q that is a sum of t functions that are products of d affine functions.

invariance principle of Mossel, O'Donnell, and Oleszkiewicz. Our key insight is that taking a random Reed–Muller codeword can, in fact, be viewed as an instantiation of the Meka–Zuckerman generator, which involves splitting the input into blocks via a pairwise independent hash function, and using independent k-wise independent distributions in each block. This allows us to obtain a version of the Majority Is Stablest theorem for our graph, which is the main corollary of the invariance principle that is used in applications of the long code. See section 5 for more details

Efficient alphabet reduction. With the Majority Is Stablest theorem in hand, proving Theorem 1.4 (efficient alphabet reduction for UNIQUE GAMES), is fairly straightforward. The idea is to simply replace the noisy hypercube gadget used by [KKMO07] with our derandomized hypercube. This is essentially immediate in the case of alphabet reduction to a binary alphabet (i.e., reduction to Max-Cut) but requires a bit more work when reducing to a larger alphabet. See Appendix A for more details.

Efficient hierarchy integrality gaps. Our proof of Theorem 1.3 again works by plugging in our short code/derandomized noisy hypercube in place of the long code in the previous integrality gap constructions [KV05, KS09, RS09]. Specifically, these constructions worked by starting with an integrality gap for UNIQUE GAMES where the basic SDP yields $1-1/r$, and then composing it with an alphabet reduction gadget to obtain a new instance; Raghavendra and Steurer [RS09] showed that the composed instances resist poly(r) rounds of the SA-SDP hierarchy and exp(poly(r)) rounds of the LH hierarchy. These constructions used the noisy cube twice—both to obtain the basic UNIQUE GAMES gap instance and to obtain the alphabet reduction gadget. We simply plug in our short code in both usages—using for the basic UNIQUE GAMES instance the efficient version obtained in Theorem 1.2, and for the alphabet reduction gadget the efficient version obtained in Theorem 1.4. (Luckily, our UNIQUE GAMES instance has affine constraints and so is compatible with our alphabet reduction gadget.) The result essentially follows in a blackbox way from the analysis of [RS09]. See Appendix B for details.

3. Preliminaries.

Let G be a regular graph with vertex set V. For a subset $S \subseteq V$ we define the *volume* of S, denoted $\mu(S)$, to be $|S|/|V|$. We define the *expansion* of S, denoted $\Phi(S)$, to be the probability over a random edge (u, v), conditioned on $u \in S$ such that $v \notin S$. Equivalently (since G is regular), $\Phi(S) = G(S, V \setminus S)/(\deg_G \cdot|S|)$, where $\deg_G$ is the degree of the graph G and $G(S, V \setminus S)$ is the number of edges going from S to $V \setminus S$. Throughout, we denote the normalized adjacency matrix of a graph G also by G, and refer to the spectrum of the adjacency matrix as the spectrum of the graph G. Note that by definition, every regular graph has maximum eigenvalue 1. In this paper, we use *expectation norms* for real-valued functions. That is, for a function $f\colon S \to \mathbb{R}$ and $p \geqslant 1$, we let $\|f\|_p := (\mathbb{E}_{x\in S}\,|f(x)|^p)^{1/p}$.

Many of the UNIQUE GAMES instances that appear in this work belong to a special subclass of UNIQUE GAMES, namely $\mathbb{F}_2^n$-MAX-2LIN instances defined below.

DEFINITION 3.1. *Given a group $\mathcal{H}$, an $\mathcal{H}$-MAX-2LIN instance consists of a system of linear equations over the group $\mathcal{H}$, where each equation is of the form $x_i - x_j = c_{ij}$ for some $c_{ij} \in \mathcal{H}$.*

Locally testable codes. Let $\mathcal{C}$ be an $[N, K, D]_2$ code; that is, $\mathcal{C}$ is a K-dimensional linear subspace of $\mathbb{F}_2^N$ with minimum distance D $(= \min\{\mathsf{wt}(x) : x \in \mathcal{C}\})$. (In this paper, we are mostly interested in the extremely high rate regime when $H = N - K$ is very small compared to N and are happy with D being a sufficiently large constant.) Let $\Delta(x, y) \in \{0, \ldots, N\}$ denote Hamming distance between $x, y \in \mathbb{F}_2^N$. For $\alpha \in \mathbb{F}_2^N$

and a code $\mathcal{C}$, we define

$$\Delta(\alpha, \mathcal{C}) \overset{\text{def}}{=} \min_{c \in \mathcal{C}} \Delta(\alpha, c).$$

DEFINITION 3.2. *We say a distribution $\mathcal{T}$ over $\mathbb{F}_2^N$ is a* canonical tester *for $\mathcal{C}$ if every vector in the support of the distribution $\mathcal{T}$ is a codeword $q \in C^{\perp}$. The* query complexity *of $\mathcal{T}$ is the maximum weight of a vector in its support. The tester's* soundness curve $s_{\mathcal{T}} \colon \mathbb{N} \to [0,1]$ *is defined as*

$$s_{\mathcal{T}}(k) \overset{\text{def}}{=} \min_{\substack{\alpha \in \mathbb{F}_2^N \\ \Delta(\alpha, \mathcal{C}) \geqslant k}} \mathop{\mathbb{P}}_{q \sim \mathcal{T}} \{\langle \alpha, q \rangle = 1\}.$$

Similarly, we denote the rejection probability *of $\mathcal{T}$ for a vector $\alpha \in \mathbb{F}_2^N$ by $s_{\mathcal{T}}(\alpha) = \mathbb{P}_{q \sim \mathcal{T}} \{\langle \alpha, q \rangle = 1\}$. We let the* query probability $\tau \in [0,1]$ *of a tester be the expected fraction of queried coordinates, that is, $\tau = \mathbb{E}_{q \sim \mathcal{T}} \mathsf{wt}(q)/N$. We say that a tester $\mathcal{T}$ with query probability τ is* smooth *if for any coordinate $i \in [N]$, $\mathbb{P}_{q \sim \mathcal{T}} \{q_i = 1\} = \tau$, and we say it is* 2-smooth *if, in addition, for any two distinct coordinates $i \neq j$, $\mathbb{P}_{q \sim \mathcal{T}} \{q_i = q_j = 1\} = \tau^2$.*

If the tester $\mathcal{T}$ is clear from the context, we will sometimes drop the subscript of the soundness curve/rejection probability $s_{\mathcal{T}}$. In the setting of this paper, we will consider testers with query probability slowly going to 0 (with N). Further, given a canonical tester $\mathcal{T}$, it is easy to amplify the probability of rejection by repeating the test and taking the XOR of the results.

Finally, the following simple lemma gives some estimates for rejection probabilities of vectors for smooth testers.

LEMMA 3.3. *If $\mathcal{T}$ is a smooth canonical tester with query probability τ, then $s_{\mathcal{T}}(\alpha) \leqslant \Delta(\alpha, \mathcal{C}) \cdot \tau$ for every vector $\alpha \in \mathbb{F}_2^N$. Furthermore, if $\mathcal{T}$ is 2-smooth, then $s_{\mathcal{T}}(\alpha) \geqslant (1-\gamma) \cdot \Delta(\alpha, \mathcal{C}) \cdot \tau$ for every vector $\alpha \in \mathbb{F}_2^N$ with $\Delta(\alpha, \mathcal{C})\tau \leqslant \gamma$.*

Proof. Fix $\alpha \in \mathbb{F}_2^N$ and let $k = \Delta(\alpha, \mathcal{C})$. Without loss of generality, we may assume $\mathsf{wt}(\alpha) = k$. By renaming coordinates, we may assume $\alpha_1 = \cdots = \alpha_k = 1$ and $\alpha_{k+1} = \cdots = \alpha_N = 0$. Then, $s_{\mathcal{T}}(\alpha) \leqslant \mathbb{P}_{q \sim \mathcal{T}} \{q_1 = 1\} + \cdots + \mathbb{P}_{q \sim \mathcal{T}} \{q_k = 1\} = k \cdot \tau$. On the other hand,

$$s_{\mathcal{T}}(\alpha) \geqslant \sum_{i=1}^{k} \mathop{\mathbb{P}}_{q \sim \mathcal{T}} \{q_i = 1\} - \sum_{0 \leqslant i < j \leqslant k} \mathop{\mathbb{P}}_{q \sim \mathcal{T}} \{q_i = q_j = 1\} \geqslant k\tau - k^2\tau^2 \geqslant (1-\gamma) \cdot k\tau. \qquad \square$$

We review the prerequisites for Majority Is Stablest and UNIQUE GAMES related results in the corresponding sections.

4. Small-set expanders from locally testable codes.

In this section we first use some known properties of hypercontractive norms to give a sufficient condition for graphs to be small-set expanders. We then describe a generic way to construct graphs satisfying this condition from locally testable codes, proving Theorem 1.1.

4.1. Subspace hypercontractivity and SMALL-SET EXPANSION.

Let $\mathcal{V}$ be a subspace of the set of functions from V to $\mathbb{R}$ for some finite set V. We denote by $P_{\mathcal{V}}$ the projection operator to the space $\mathcal{V}$. For $p, q \geqslant 1$, we define

$$\|\mathcal{V}\|_{p \to q} \overset{\text{def}}{=} \max_{f \colon V \to \mathbb{R}} \frac{\|P_{\mathcal{V}} f\|_q}{\|f\|_p}.$$

We now relate this notion to SMALL-SET EXPANSION. We first show that a subspace $\mathcal{V}$ with bounded $(4/3) \to 2$ norm cannot contain the characteristic function of a small set.

Lemma 4.1. *Let $f : V \to \{0,1\}$ such that $\mu = \mathbb{E}_{x\in V}[f(x)]$; then $\|P_{\mathcal{V}} f\|_2^2 \leqslant \|\mathcal{V}\|_{4/3\to 2}^2 \mu^{3/2}$.*

Proof. The proof is by direct calculation

$$\|P_{\mathcal{V}} f\|_2^2 \leqslant \|\mathcal{V}\|_{4/3\to 2}^2 \|f\|_{4/3}^2 = \|\mathcal{V}\|_{4/3\to 2}^2 \mu^{(3/4)\cdot 2}. \qquad \square$$

Note that if $\|\mathcal{V}\|_{4/3\to 2} = O(1)$ and $\mu = o(1)$, then $\|P_{\mathcal{V}} f\|_2^2 = o(\|f\|_2^2)$, meaning the projection of f onto V is small. It is often easier to work with the $2 \to 4$ norm instead of the $4/3 \to 2$ norm. The following lemma allows us to use a bound on the former to bound the latter.

Lemma 4.2.

$$\|\mathcal{V}\|_{4/3\to 2} \leqslant \|\mathcal{V}\|_{2\to 4}.$$

Proof. Let $f : V \to \mathbb{R}$ and let $f' = P_{\mathcal{V}} f$. We know that

$$\begin{aligned}
\mathbb{E}[f'^2] &= \mathbb{E}[f' \cdot f] \quad \text{(since } f' \text{ is the projection of } f\text{)}\\
&\leqslant \mathbb{E}[f'^4]^{1/4}\, \mathbb{E}[f^{4/3}]^{3/4} \quad \text{(by Hölder's inequality)}\\
&= \mathbb{E}[(P_{\mathcal{V}} f')^4]^{1/4}\, \mathbb{E}[f^{4/3}]^{3/4} \quad \text{(projection is idempotent)}\\
&\leqslant \|\mathcal{V}\|_{2\to 4}\, \mathbb{E}[(f')^2]^{1/2}\, \mathbb{E}[f^{4/3}]^{3/4}.
\end{aligned}$$

Dividing by $\|f\|_2 = \mathbb{E}[f^2]^{1/2}$ yields the result. $\square$

We now conclude that graphs for which the top eigenspace has bounded $2 \to 4$ norm are small-set expanders. The lemma can be viewed qualitatively as a generalization of one direction of the classical Cheeger's inequality relating combinatorial expansion to eigenvalue gap [Che70].

Lemma 4.3. *Let $G = (V, E)$ be a regular graph, and let $\mathcal{V}$ be the span of the eigenvectors of G with eigenvalue larger than λ. Then, for every $S \subseteq V$,*

$$\Phi(S) \geqslant 1 - \lambda - \|\mathcal{V}\|_{2\to 4}^2 \sqrt{\mu(S)}.$$

Proof. Let f be the characteristic function of S, and write $f = f' + f''$, where $f' = P_{\mathcal{V}} f$ (and so $f'' = f - f'$ is the projection to the eigenvectors with value at most λ). Let $\mu = \mu(S)$. We know that

$$\Phi(S) = 1 - \langle f, Gf\rangle / \|f\|_2^2 = 1 - \langle f, Gf\rangle/\mu. \tag{4.1}$$

By Lemmas 4.1 and 4.2,

$$\begin{aligned}
\langle f, Gf\rangle = \langle f', Gf'\rangle + \langle f'', Gf''\rangle &\leqslant \|f'\|_2^2 + \lambda\|f''\|_2^2 \leqslant \|\mathcal{V}\|_{4/3\to 2}^2 \mu^{3/2} + \lambda\mu\\
&\leqslant \|\mathcal{V}\|_{2\to 4}^2 \mu^{3/2} + \lambda\mu\,.
\end{aligned}$$

Plugging this into (4.1) yields the result. $\square$

4.2. Cayley graphs on codes. Motivated by the previous section, we now construct a graph for which the projection operator onto the top eigenspace is hypercontractive, i.e., has small $2 \to 4$ norm, while also having high rank.

Let $\mathcal{C} \subseteq \mathbb{F}_2^N$ be an $[N, K, D]_2$ code. The graph we construct will be a Cayley graph with vertices indexed by $\mathcal{C}^\perp$ and edges drawn according to a canonical local tester $\mathcal{T}$ for $\mathcal{C}$. Let $\mathrm{Cay}(\mathcal{C}^\perp, \mathcal{T})$ denote the (weighted) Cayley graph with vertex set $\mathcal{C}^\perp$ and edges generated by $\mathcal{T}$. We describe the graph more precisely by specifying the neighbor distribution for a random walk on the graph. For a vertex $p \in \mathcal{C}^\perp$, a random neighbor has the form $p + q$ with q sampled from the tester $\mathcal{T}$. (Since the group $\mathcal{C}^\perp$ has characteristic 2, the graph $\mathrm{Cay}(\mathcal{C}^\perp, \mathcal{T})$ is symmetric for every tester $\mathcal{T}$.)

We will argue that if the tester $\mathcal{T}$ has small query complexity and good soundness, then the graph $\mathrm{Cay}(\mathcal{C}^\perp, \mathcal{T})$ has many large eigenvalues while being a small-set expander.

Theorem 4.4. *Let $\mathcal{C}$ be an $[N, K, D]_2$ linear code that has a canonical tester $\mathcal{T}$ with query complexity εN and soundness curve $s()$, and let $k < D/5$. The graph $\mathrm{Cay}(\mathcal{C}^\perp, \mathcal{T})$ has $2^{N-K} = 2^H$ vertices with at least $N/2$ eigenvalues larger than $1 - 4\varepsilon$. All subsets S of $\mathcal{C}^\perp$ have expansion at least*

$$\Phi(S) \geqslant 2s(k) - 3^k \sqrt{\mu(S)}.$$

By XORing the results of multiple tests, one can let the soundness $s(k)$ tend to $1/2$. Hence, if $s(k)$ is significantly larger than ε (for appropriate k), one can obtain a graph with many large eigenvalues such that small enough sets have near-perfect expansion.

Eigenfunctions and eigenvalues. We identify the graph $G = \mathrm{Cay}(\mathcal{C}^\perp, \mathcal{T})$ by its normalized adjacency matrix. For every vector $\alpha \in \mathbb{F}_2^N$, the character $\chi_\alpha \colon \mathcal{C}^\perp \to \{\pm 1\}$ with $\chi_\alpha(p) = (-1)^{\langle \alpha, p\rangle}$ is an eigenfunction of G. If two vectors $\alpha, \beta \in \mathbb{F}_2^N$ belong to the same coset of $\mathcal{C}$, they define the same character over $C^\perp$ since $\langle \alpha + \beta, p\rangle = 0$ for all $p \in \mathcal{C}^\perp$, while if $\alpha + \beta \notin \mathcal{C}$, then $\langle \chi_\alpha, \chi_\beta\rangle = 0$. Thus, the set of characters of $\mathcal{C}^\perp$ corresponds canonically to the quotient space $\mathbb{F}_2^N/\mathcal{C}$. If we fix a single representative α for every coset in $\mathbb{F}_2^N/\mathcal{C}$, we have exactly $2^{N-K} = 2^H$ distinct, mutually orthogonal characters. We define the degree of a character as follows:

$$\deg(\chi_\alpha) = \min_{c \in \mathcal{C}} \mathsf{wt}(\alpha + c) = \Delta(\alpha, \mathcal{C})\,. \tag{4.2}$$

Note that if $\deg(\chi_\alpha) < D/2$, then the minimum weight representative in $\alpha + \mathcal{C}$ is unique. (This uniqueness will allow us later to define low-degree influences of functions; see section 5.)

We let λ_α denote the eigenvalue corresponding to character χ_α. The following observation connects the soundness of the canonical tester to the spectrum of G.

Lemma 4.5. *For any $\alpha \in \mathbb{F}_2^N$, $\lambda_\alpha = 1 - 2s(\alpha)$.*

Proof. From standard facts about Cayley graphs, it follows that

$$\lambda_\alpha = \mathop{\mathbb{E}}_{q \in \mathcal{T}}[\chi_\alpha(q)] = \mathop{\mathbb{E}}_{q \in \mathcal{T}}[(-1)^{\alpha \cdot q}] = 1 - 2 \mathop{\mathbb{P}}_{q \in \mathcal{T}}[\alpha \cdot q = 1] = 1 - 2s(\alpha). \tag{4.3}$$ □

We use this to show that many *dictator cuts* in G which correspond to characters with degree 1 have eigenvalues close to 1. We let λ_i, χ_i denote $\lambda_{\{i\}}, \chi_{\{i\}}$. As noted before, for $D > 2$ these are distinct characters.

Corollary 4.6. *We have $\lambda_i \geqslant 1 - 4\varepsilon$ for at least $N/2$ coordinates $[i] \in N$.*

Proof. We have $\lambda_i = 1 - 2\,\mathbb{P}_{q \in \mathcal{T}}[q_i = 1]$. Since $\mathsf{wt}(q) \leqslant \varepsilon N$ for every $q \in \mathcal{T}$,

$$\sum_{i=1}^{N} \mathop{\mathbb{P}}_{q \in \mathcal{T}}[q_i = 1] \leqslant \varepsilon N.$$

So we can have $\mathbb{P}_{q \in \mathcal{T}}[q_i = 1] \geqslant 2\varepsilon$ for at most $N/2$ coordinates. □

Another immediate consequence of Lemma 4.5 is that large-degree characters have small eigenvalues.

Corollary 4.7. *If $\deg(\chi_\alpha) \geqslant k$, then $\lambda_\alpha \leqslant 1 - 2s(k)$.*

Subspace hypercontractivity. Given a function $f\colon \mathcal{C}^\perp \to \mathbb{R}$, we can write it (uniquely) as a linear combination of the characters $\{\chi_\alpha\}_{\alpha\in\mathbb{F}_2^N/\mathcal{C}}$:

$$f(p) = \sum_{\alpha\in\mathbb{F}_2^N/\mathcal{C}} \hat{f}(\alpha)\chi_\alpha(p)\,,$$

where $\hat{f}(\alpha) = \langle \chi_\alpha, f\rangle$ is the *Fourier transform* of f (over the abelian group $\mathcal{C}^\perp$).

We define the *degree* of f, denoted $\deg(f)$, to be $\max_{\alpha:\hat{f}(\alpha)\neq 0}\deg(\chi_\alpha)$. Note that $\deg(f+g) \leqslant \max\{\deg(f),\deg(g)\}$ and $\deg(fg) \leqslant \deg(f)+\deg(g)$. The following crucial observation follows immediately from the fact that $\mathcal{C}$ has minimum distance$D$.

FACT 4.8. *The uniform distribution on $\mathcal{C}^\perp$ is $(D-1)$-wise independent. That is, for any $\alpha\in\mathbb{F}_2^N$ such that $1\leqslant \mathsf{wt}(\alpha) < D$, we have $\mathbb{E}_{p\in\mathcal{C}^\perp}[\chi_\alpha(p)] = 0$.*

This fact has the following corollary.

LEMMA 4.9. *Let $\ell < (D-1)/4$, and let $\mathcal{V}$ be the subspace of functions with degree at most ℓ. Then $\|\mathcal{V}\|_{2\to 4} \leqslant 3^{\ell/2}$.*

Proof. The proof follows from the following two facts:

1. This bound on the 2 → 4 norm is known to hold for true low-degree polynomials under the uniform distribution on the hypercube by the Bonami–Beckner–Gross inequality [O'D08].
2. The expectation of polynomials of degree up to $4\ell < D-1$ is the same under the uniform distribution and a $(D-1)$-wise independent distribution.

Given $f : \mathbb{R}^n \to \mathbb{R}$, let f^ℓ denote its projection onto the space $\mathcal{V}$ spanned by characters where $\deg(\chi_\alpha)\leqslant \ell$. We have

$$\|f^\ell\|_4^4 = \mathop{\mathbb{E}}_{p\in\mathcal{C}^\perp}[f^\ell(p)^4] = \mathop{\mathbb{E}}_{p\in\{0,1\}^N}[f^\ell(p)^4]\,,$$

$$\|f\|_2^2 \geqslant \|f^\ell\|_2^2 = \mathop{\mathbb{E}}_{p\in\mathcal{C}^\perp}[f^\ell(p)^2] = \mathop{\mathbb{E}}_{p\in\{0,1\}^N}[f^\ell(p)^2]\,.$$

By the 2 → 4 hypercontractivity for degree-ℓ polynomials over $\{0,1\}^N$,

$$\mathop{\mathbb{E}}_{p\in\{0,1\}^N}[f^\ell(p)^4] \leqslant 9^\ell \mathop{\mathbb{E}}_{p\in\{0,1\}^N}[f^\ell(p)^2]^2\,.$$

So we conclude that

$$\mathop{\mathbb{E}}_{p\in\mathcal{C}^\perp}[f^\ell(p)^4] \leqslant 9^\ell \mathop{\mathbb{E}}_{p\in\mathcal{C}^\perp}[f^\ell(p)^2]^2 \leqslant 9^\ell \mathop{\mathbb{E}}_{p\in\mathcal{C}^\perp}[f(p)^2]^2\,,$$

which implies that $\|\mathcal{V}\|_{2\to 4}\leqslant 3^{\ell/2}$. ∎

Combining the above bound with Lemma 4.3, we get that, if the local tester rejects sufficiently far codewords with high probability, then the resulting graph is a small-set expander.

COROLLARY 4.10. *For every vertex subset S in the graph $\mathrm{Cay}(\mathcal{C}^\perp,\mathcal{T})$ and every $k < D/5$, we have*

$$\Phi(S) \geqslant 2s(k) - 3^k\mu(S)^{\frac{1}{2}}.$$

In particular, as $s(k)$ tends to 1/2, the expansion of small sets tends to 1. This corollary together with Corollary 4.6 completes the proof of Theorem 4.4.

4.3. A canonical tester for Reed–Muller codes. We instantiate the construction from the previous section for the Reed–Muller code. Let $\mathcal{C} = \mathsf{RM}(n, n-d-1)$ be the Reed–Muller code on n variables of degree $n-d-1$, which has $N = 2^n$, $H = \sum_{j \leqslant d} \binom{n}{j}$, and $D = 2^{d+1}$. Specifically, $\mathsf{RM}(n, n-d-1)$ consists of tables of evaluations of n-variate $(n-d-1)$-degree polynomials over $\{0,1\}^n$. Bhattacharyya et al. [BKS$^+$10] analyze the canonical tester $\mathcal{T}_{\mathsf{RM}}$ which samples a random minimum weight codeword from $\mathcal{C}^{\perp}$. It is well known that the dual of $\mathsf{RM}(n, n-d-1)$ is exactly $\mathsf{RM}(n,d)$ and that the minimum weight codewords in $\mathsf{RM}(n,d)$ are products of d linearly independent affine forms. They have weight $2^{n-d} = \varepsilon N$, where $\varepsilon = 2^{-d}$. Thus, our graph $\mathrm{Cay}_{\mathsf{RM}} = \mathrm{Cay}(\mathsf{RM}_{n,d}, \mathcal{T}_{\mathsf{RM}})$ has as its vertices the d-degree polynomials over $\mathbb{F}_2^n$ with an edge between every pair of polynomials P, Q such that $P - Q$ is equal to a minimum weight codeword, which are known to be products of d linearly independent affine forms.

Theorem 4.11 (see [BKS$^+$10]). *There exists a constant $\eta_0 > 0$ such that for all n, d, and $k < \eta_0 2^d$ the tester $\mathcal{T}_{\mathsf{RM}}$ described above has soundness $s(k) \geqslant (k/2) \cdot 2^{-d}$.*

Theorem 4.11 allows us to estimate the eigenvalue profile of $\mathrm{Cay}_{\mathsf{RM}}$ and shows that small sets have expansion close to $O(\eta_0)$. From here, we can get near-perfect expansion by taking short random walks. To avoid cumbersome discretization issues, we work with continuous-time random walks on graphs instead of the usual discrete random walks.

Definition 4.12. *For a graph G the continuous-time random walk on G with parameter t is described by the (stochastic) matrix $G(t) = e^{-t(I-G)}$. $G(t)$ and G have the same eigenvectors, and the eigenvalues of $G(t)$ are $\{e^{-t(1-\mu_i)}\}$, where $\{\mu_i\}$ is the spectrum of G.*

We will view $\mathrm{Cay}_{\mathsf{RM}}(t)$ as a weighted graph. We show that its eigenvalue profile is close to that of the noisy cube; this stronger statement will be useful later.

Lemma 4.13. *Let $t = \varepsilon 2^{d+1}$ for $\varepsilon > 0$ and $\rho = e^{-\varepsilon}$. Let $\{\lambda_\alpha\}$ denote the eigenvalues of $\mathrm{Cay}_{\mathsf{RM}}(t)$.*

- *If $\deg(\chi_\alpha) = k$, then $\lambda_\alpha \leqslant \max(\rho^{k/2}, \rho^{\mu_0 2^d})$, where μ_0 is an absolute constant.*
- *For all $\delta < \delta_0$ for some constant δ_0, if $\deg(\chi_\alpha) = k < \delta^2 2^{d+1}$, then $|\lambda_\alpha - \rho^k| \leqslant \delta$.*

Proof. Let $\{\mu_\alpha\}$ be the eigenvalues of $\mathrm{Cay}_{\mathsf{RM}}$ corresponding to the character χ_α so that $\lambda_\alpha = e^{-t(1-\mu_\alpha)}$. Let $\tau = 2^{-(d+1)}$. Since the canonical tester $\mathcal{T}_{\mathsf{RM}}$ for $\mathcal{C}$ is 2-smooth, by Lemma 3.3, $\mu_\alpha = 1 - k\tau \pm k^2\tau^2$. Hence, $\lambda_\alpha = e^{-t(1-\mu_\alpha)} = e^{-\varepsilon k(1 \pm k\tau)} = \rho^k e^{-\varepsilon k^2 \tau}$.

For $k \leqslant 2^d$, $1 - \mu_\alpha = k\tau \pm k^2\tau^2 \geqslant k\tau/2$. Therefore, if $\deg(\chi_\alpha) \leqslant 2^d$, $\lambda_\alpha = e^{-\varepsilon 2^{d+1}(1-\mu_\alpha)} \leqslant e^{-\varepsilon 2^{d+1} k\tau/2} = \rho^{k/2}$. For $k > 2^d$, by Corollary 4.7, $\mu_\alpha < 1 - 2s(k) < C_0$ for a universal constant $C_0 < 1$. Therefore, $|\lambda_\alpha| < e^{-\varepsilon 2^{d+1}(1-C_0)} = \rho^{(1-C_0)2^{d+1}} < \rho^{\mu_0 2^d}$ for $\mu_0 < (1-C_0)/2$.

We now prove the second bound. If $\varepsilon k^2 \tau < \delta/10$, we have $\lambda_\alpha = \rho^{-k}(1 \pm \delta)$ which implies $|\lambda_\alpha - \rho^k| \leqslant \delta$. Otherwise, if $\varepsilon k^2\tau \geqslant \delta/10$, our assumption $k < \delta^2 2^{d+1}$ implies $\varepsilon k > 1/(10\delta)$; hence $\varepsilon^{-\varepsilon k} \leqslant e^{-\frac{1}{10\delta}} \leqslant \delta/4$ for all $\delta < \delta_0$. For $k \leqslant 2^d$, $(1-\mu_\alpha) = k\tau \pm k^2\tau^2 \geqslant k\tau/2$. Hence $\lambda_\alpha \leqslant e^{-tk\tau/2} \leqslant e^{-\frac{1}{20\delta}} \leqslant \delta/4$ for all $\delta < \delta_0$. In this case, we get $|\lambda_\alpha - \rho^k| \leqslant |\lambda_\alpha| + |\rho^k| \leqslant \delta/2$. ∎

Since the eigenvectors stay the same, $\mathrm{Cay}_{\mathsf{RM}}(t)$ inherits the hypercontractive properties of $\mathrm{Cay}_{\mathsf{RM}}$. In particular, by Lemma 4.9, $\|\mathcal{V}\|_{2\to 4} \leqslant 3^{\ell/2}$, where $\mathcal{V}$ denotes polynomials of degree $\ell \leqslant \frac{D-1}{4}$. Combining Lemmas 4.3 and 4.13, we obtain a graph with Small-Set Expansion and many large eigenvalues.

Theorem 4.14. *For any $\varepsilon, \eta > 0$, there exists a graph G with $2^{(\log|G|)^{\frac{1}{d}}}$ eigenvalues larger than $1-\varepsilon$ for $d = \log(1/\varepsilon) + \log\log(1/\eta) + O(1)$ and where every set $S \subseteq G$ has expansion*

$$\Phi(S) \geqslant 1 - \eta - 3^{\frac{c_1}{\varepsilon}\log(1/\eta)}\sqrt{\mu(S)}$$

for some constant c_1.

Proof. Let μ_0, δ_0 be constants from previous lemma. Fix $\ell = \frac{c_1}{\varepsilon}\log(\frac{1}{\eta})$ so that $e^{-\varepsilon\ell/2} = \eta$ and $d = \log(\ell) + c_2$ so that $\ell \leqslant \min(\mu_0 2^{d+1}, 2^d/5)$. Consider the graph $\mathrm{Cay}_{\mathsf{RM}}(t)$ of the continuous random walk on $\mathrm{Cay}_{\mathsf{RM}}$ where $t = \varepsilon 2^{d+1}$ as in Lemma 4.13. Note that the graph has $|G| = \sum_{j\leqslant d}\binom{N}{i}$ vertices. Let $\{\mu_\alpha\}$ be the spectrum of $\mathrm{Cay}_{\mathsf{RM}}$, and let λ_α be the spectrum of $\mathrm{Cay}_{\mathsf{RM}}(t)$.

Then, for every $\alpha \in \mathbb{F}_2^N/\mathcal{C}$, $\deg(\chi_\alpha) = 1$, we have $s_\tau(\alpha) \geqslant 2^{-d}$. Hence $\mu_\alpha \geqslant 1 - 2^{-d+1}$, $\lambda_\alpha \geqslant e^{-t2^{-d+1}} = e^{-4\varepsilon}$. Therefore, there are at least $N = 2^{(\log|G|)^{1/d}}$ eigenvalues which are larger than $1-4\varepsilon$.

Since $\ell < \mu_0 2^{d+1}$, by Lemma 4.13, if $\deg(\chi_\alpha) > \ell$, $\lambda_\alpha \leqslant \eta$. Let $\mathcal{V}$ be the subspace spanned by characters of degree at most ℓ. Since $\ell < 2^d/5$ by Lemma 4.9, $\|\mathcal{V}\|_{2\to4} \leqslant 3^{\ell/2}$. Therefore, by Lemma 4.3, for any set $S \subseteq G$ with $\mu(S) \leqslant \delta$,

$$\Phi(S) \geqslant 1 - \eta - 3^{\frac{c_1}{\varepsilon}\log(1/\eta)}\sqrt{\mu(S)}. \qquad \square$$

Remark 4.15 (coding application). The fact that our graph is a Cayley graph over $\mathbb{F}_2^n$ has a potentially interesting implication for coding theory. By looking at the set of edge labels as the rows of a generating matrix for a code, we know that large Fourier coefficients correspond to low weight codewords, and hence we get a code of dimension $m = \binom{n}{d}$ that has an almost exponential (i.e., 2^n) number of codewords of low weight, but yet has small *generalized Hamming distance* in the sense that every subspace of codimension $\omega(1)$ contains a codeword of fractional Hamming weight $1 - o(1)$. In particular, by setting d to be a function slowly tending to infinity, we can get a linear code for which correcting from a $o(1)$ fraction of *corruption* errors requires an almost exponential list size, but for which one can correct a fraction approaching 1 of *erasure* errors using a list of constant size. (The code obtained by taking all edges of our graph has an almost exponential blowup, but this can be reduced by subsampling the edges.)

5. Majority Is Stablest over codes.

In this section we show an analogue of the Majority Is Stablest result of Mossel et al. for the Reed–Muller graph we constructed in the previous section; this will help us replace the noisy cube with the Reed–Muller graph in various Unique Games gadgets.

We first review some definitions. For a function $f : \{\pm1\}^N \to \mathbb{R}$ and $\ell > 0$, define

$$\mathrm{Inf}_i^{\leqslant\ell}(f) = \sum_{\alpha\in\{0,1\}^N, |\alpha|\leqslant\ell, \alpha_i=1} |\hat{f}(\alpha)|^2.$$

For $\rho > 0$, let $\Gamma_\rho : [0,1] \to [0,1]$ be the Gaussian noise stability curve defined as follows. For $\mu \in [0,1]$, let $t \in \mathbb{R}$ be such that $\mathbb{P}_{g\leftarrow\mathcal{N}(0,1)}[g < t] = \mu$. Then, $\Gamma_\rho(\mu) = \mathbb{P}_{X,Y}[X \leqslant t, Y \leqslant t]$, where $(X,Y) \in \mathbb{R}^2$ is a two-dimensional mean zero Gaussian random vector with covariance matrix $\left(\begin{smallmatrix}1 & \rho\\ \rho & 1\end{smallmatrix}\right)$. We refer the reader to Appendix B in Mossel, O'Donnell, and Oleszkiewicz [MOO05] for a more detailed discussion on Γ_ρ.

Let $P(x) = \sum_{I\subseteq[N]} a_I \prod_{i\in I} x_i$ be an N-variate multilinear polynomial $P : \mathbb{R}^N \to \mathbb{R}$. We define $\|P\|^2 = \sum_I a_I^2$ and say P is ε-regular if for every $i \in [N]$, $\sum_{i\ni I} a_I^2 \leqslant \varepsilon^2 \cdot \|P\|^2$.

Throughout this section, we let $T_{\rho,N}$ (we omit N when the dimension is clear) denote the noisy hypercube graph with second largest eigenvalue ρ.

5.1. Majority Is Stablest and invariance. The following theorem shows that, in the context of noise stability, a *regular* function on the hypercube behaves like a function on Gaussian space.

Theorem 5.1 (Majority Is Stablest [MOO05]). *Let $f\colon \{\pm1\}^N \to [0,1]$ be a function with $\mathbb{E} f = \mu$. Suppose $\mathrm{Inf}_i^{\leqslant 10\log(1/\tau)}(f) \leqslant \tau$ for all $i \in [N]$. Then,*

$$\langle f, T_\rho f\rangle \leqslant \Gamma_\rho(\mu) + \tfrac{10\log\log(1/\tau)}{(1-\rho)\log(1/\tau)}\,,$$

where T_ρ is the Boolean noise graph with second largest eigenvalue ρ and Γ_ρ is the Gaussian noise stability curve.

We will need the following ingredient of the proof of Theorem 5.1 from [MOO05]. For $a, b \in \mathbb{R}$, let $\zeta_{[a,b]}\colon \mathbb{R} \to \mathbb{R}_+$ be the functional $\zeta_{[a,b]}(x) = \max\{a - x, x - b, 0\}^2$. For a real-valued random variable X, the expectation $\mathbb{E}\,\zeta(X)$ is the L_2^2-distance of X to the set of $[a, b]$-valued random variables (over the same probability space as X). We will be interested in the case $a = 0$ and $b = 1$. For this case, we abbreviate $\zeta = \zeta_{[0,1]}$.

Theorem 5.2 (invariance principle [MOO05, Theorem 3.19]). *Let P be a τ-regular N-variate real multilinear polynomial with degree at most ℓ and $\|P\|^2 \leqslant 1$. Then,*

$$\left| \mathop{\mathbb{E}}_{x\in\{\pm1\}^N} \zeta \circ P(x) - \mathop{\mathbb{E}}_{y\sim\mathcal{N}(0,1)^N} \zeta \circ P(y) \right| \leqslant 2^{O(\ell)}\sqrt{\tau}\,.$$

We will need the following corollary that can handle functions that are not $[0,1]$-valued as in the theorem but just close to $[0,1]$-valued functions.

Corollary 5.3. *Let $f\colon \{\pm1\}^N \to \mathbb{R}$ be a function with $\mathbb{E} f = \mu$ and $\mathbb{E}\,\zeta \circ f \leqslant \tau$. Suppose $\mathrm{Inf}_i\, f^{\leqslant 30\log(1/\tau)} \leqslant \tau$ for all $i \in [N]$. Then,*

$$\langle f, T_\rho f\rangle \leqslant \Gamma_\rho(\mu) + \tfrac{40\log\log(1/\tau)}{(1-\rho)\log(1/\tau)}\,,$$

where T_ρ is the Boolean noise graph with second largest eigenvalue ρ and Γ_ρ is the Gaussian noise stability curve. (Here, we assume that τ is small enough.)

Proof. Let f' be the closest $[0,1]$-valued function to f. Since $\|f - f'\| \leqslant \sqrt{\tau}$, it follows that $\mathrm{Inf}_i^{\leqslant 20\log(1/\tau)} f' \leqslant \tau + O(\sqrt{\tau}) \ll \tau^{1/3}$ and $\mathbb{E} f' \leqslant \mathbb{E} f + \sqrt{\tau}$. Since $\langle f, T_\rho f\rangle \leqslant \langle f', T_\rho f'\rangle + O(\sqrt{\tau})$, the corollary follows by applying Theorem 5.1 to the function f'. (Here, we also use that fact that $\Gamma_\rho(\mu + \sqrt{\tau}) \leqslant \Gamma_\rho(\mu) + 2\sqrt{\tau}$. See Lemma B.3 in [MOO05].) □

We remark that although we specialize to Reed–Muller codes in this section, most of the arguments generalize appropriately to arbitrary codes with good canonical testers modulo a conjecture about bounded independence distributions fooling low-degree polynomial threshold functions. We briefly discuss this in section 5.3.

To state our version of Majority Is Stablest we first extend the notion of influences to functions over Reed–Muller codes. For $n, d \in \mathbb{N}$, $N = 2^n$, let $\mathcal{C} \subseteq \mathbb{F}_2^N$ be the Reed–Muller code $\mathsf{RM}(n, n-d-1)$, and let $\mathcal{C}^\perp \subseteq \mathbb{F}_2^N$ be its dual $\mathsf{RM}(n, d)$. For the rest of this section we assume that a set of representatives corresponding to the minimum weight codeword in each coset is chosen for the coset space $\mathbb{F}_2^N/\mathcal{C}$.

DEFINITION 5.4. *For a function $f : \mathcal{C}^\perp \to \mathbb{R}$ and $i \in [N]$, $\ell > 0$, the ℓ-degree influence of coordinate i in f is defined by*

$$\mathrm{Inf}_i^{\leqslant \ell}(f) = \sum_{\substack{\alpha \in \mathbb{F}_2^N/\mathcal{C},\\ |\alpha| \leqslant \ell, \alpha_i = 1}} \hat{f}(\alpha)^2 .$$

(Recall that the Fourier coefficient $\hat{f}(\alpha) = \mathbb{E}_{x \in \mathcal{C}^\perp}[\chi_\alpha(x)]$.) As all α's with weight less than half the distance of $\mathcal{C}$ fall into different cosets of $\mathcal{C}$, for $\ell < D/2$, the above expression simplifies to

$$\mathrm{Inf}_i^{\leqslant \ell}(f) = \sum_{\alpha \in \mathbb{F}_2^N,\ |\alpha| \leqslant \ell,\ \alpha_i = 1} \hat{f}(\alpha)^2 .$$

The sum of ℓ-degree influences of a function f can be bounded as below.

LEMMA 5.5. *For a function $f : \mathcal{C}^\perp \to \mathbb{R}$ and $\ell < D/2$,*

$$\sum_{i \in [N]} \mathrm{Inf}_i^{\leqslant \ell}(f) \leqslant \ell\, \mathbb{V}[f],$$

where $\mathbb{V}[f] = \mathbb{E}[f^2] - (\mathbb{E}[f])^2$ denotes the variance of f.

Proof. The lemma is an easy consequence from the definition of $\mathrm{Inf}^{\leqslant \ell}(f)$ and the fact that $\mathbb{V}[f] = \sum_{\alpha \neq 0} \hat{f}(\alpha)^2$. We include the proof for the sake of completeness:

$$\begin{aligned}\sum_{i \in [N]} \mathrm{Inf}_i^{\leqslant \ell}(f) &= \sum_{i \in [N]} \sum_{\alpha \in \mathbb{F}_2^N,\ |\alpha| \leqslant \ell,\ \alpha_i = 1} \hat{f}(\alpha)^2 \\ &= \sum_{\alpha \in \mathbb{F}_2^N,\ |\alpha| \leqslant \ell,\ \alpha \neq 0} |\alpha| \hat{f}(\alpha)^2 \\ &\leqslant \ell \sum_{\alpha \in \mathbb{F}_2^N,\ |\alpha| \leqslant \ell,\ \alpha \neq 0} \hat{f}(\alpha)^2 \leqslant \ell\, \mathbb{V}[f] \qquad \square\end{aligned}$$

We are now ready to state the main result of this section generalizing the Majority Is Stablest result to Reed–Muller codes. Let $\mathcal{T}_{\mathsf{RM}}$ be the canonical tester for $\mathcal{C}$ as defined in section 4.3.

THEOREM 5.6. *There exist universal constants c, C such that the following holds. Let G be a continuous-time random walk on the* Reed–Muller *graph* $\mathrm{Cay}(\mathcal{C}^\perp, \mathcal{T}_{\mathsf{RM}})$ *with parameter $t = \varepsilon 2^{d+1}$. Let $f\colon \mathcal{C}^\perp \to [0,1]$ be a function on $\mathcal{C}^\perp$ with $\mathbb{E}_{x \sim \mathcal{C}^\perp}[f(x)] = \mu$ and $\max_{i \in [N]} \mathrm{Inf}_i^{\leqslant 30 \log(1/\tau)}(f) < \tau$. Then, for $d > C \log(1/\tau)$,*

$$\langle f, Gf \rangle \leqslant \Gamma_\rho(\mu) + \frac{c \log\log(1/\tau)}{(1-\rho)\log(1/\tau)}, \tag{5.1}$$

where $\rho = e^{-\varepsilon}$ and $\Gamma_\rho\colon \mathbb{R} \to \mathbb{R}$ is the noise stability curve of Gaussian space.

The proof of the theorem proceeds in three steps. We first show that the eigenvalue profile of the graph G is close to the eigenvalue profile of the Boolean noise graph (see Lemma 4.13). We then show an invariance principle for low-degree polynomials (and as a corollary for *smoothed functions*), showing that they have similar behavior under the uniform distribution over the hypercube and the uniform distribution over the appropriate Reed–Muller code. Finally, we use the invariance principle to translate the Majority Is Stablest result in the hypercube setting to the Reed–Muller code.

The above approach is similar to that of Mossel, O'Donnell, and Oleszkiewicz, who translate a Majority Is Stablest result in the Gaussian space to the hypercube using a similar invariance principle.

We first state the invariance principle that we use below (see the next subsection for the proof). Recall the definition of the functional $\zeta\colon \mathbb{R} \to \mathbb{R}$ from section 5.1.

THEOREM 5.7. *Let $N = 2^n$ and $d \geqslant 4\log(1/\tau)$. Let $P : \mathbb{R}^N \to \mathbb{R}$ be a τ-regular polynomial of degree at most ℓ. Then, for $x \in_u \{\pm 1\}^N$, $z \in_u \mathsf{RM}(n,d)$,*

$$|\mathbb{E}[\zeta \circ P(x)] - \mathbb{E}[\zeta \circ P(z)]| \leqslant 2^{c_1 \ell}\sqrt{\tau},$$

for a universal constant $c_1 > 0$.

The (somewhat technical) proof of Theorem 5.6 from the above invariance principle closely follows the argument of Mossel, O'Donnell, and Oleszkiewicz and is deferred to Appendix C.

5.2. Invariance principles over Reed–Muller codes. The various invariance principles of Mossel, O'Donnell, and Oleszkiewicz [MOO05] are essentially equivalent (up to some polynomial loss in error estimates) to saying that for any low-degree regular polynomial P, the polynomial threshold function (PTF) sign($P(\ )$) cannot distinguish between the uniform distribution over the hypercube and the standard multivariate Gaussian distribution $\mathcal{N}(0,1)^N$.

THEOREM 5.8. *Let $P : \mathbb{R}^N \to \mathbb{R}$ be an ε-regular polynomial of degree at most ℓ. Then, for any $x \sim \{\pm 1\}^N$, $y \leftarrow \mathcal{N}(0,1)^N$,*

$$|\mathbb{E}[\mathrm{sign}(P(x))] - \mathbb{E}[\mathrm{sign}(P(y))]| \leqslant O(\ell\varepsilon^{1/(2\ell+1)}).$$

Ideally, we would like a similar invariance principle to hold even when x is chosen uniformly from the codes of the earlier sections instead of being uniform over the hypercube. Such an invariance principle will allow us to analyze alphabet reductions and integrality gaps based on graphs considered in earlier sections (e.g., the Reed–Muller graph). We obtain such generic invariance principles applicable to all codes modulo certain plausible conjectures on low-degree polynomials being fooled by bounded independence.

For the explicit example of the Reed–Muller code we bypass the conjectures and directly show an invariance principle by proving that the uniform distribution over the Reed–Muller code fools low-degree PTFs. To do so, we will use the specific structure of the Reed–Muller code along with the pseudorandom generator (PRG) for PTFs of Meka and Zuckerman [MZ13]. Specifically, we show that the uniform distribution over the Reed–Muller code can be seen as an instantiation of the PRG of [MZ13] and then use the latter's analysis as a blackbox. Call a smooth function $\psi : \mathbb{R} \to \mathbb{R}$ B-nice if $|\psi^{(4)}(t)| \leqslant B$ for every $t \in \mathbb{R}$.

THEOREM 5.9. *Let $N = 2^n$ and $d \geqslant \log \ell + 2\log(1/\varepsilon) + 2$. Let $P : \mathbb{R}^N \to \mathbb{R}$ be an ε-regular multilinear polynomial of degree at most ℓ. Let $x \leftarrow \mathcal{N}(0,1)^N$, $z \sim \mathsf{RM}(n,d)$. Then, for every 1-nice function $\psi : \mathbb{R} \to \mathbb{R}$,*

$$|\mathbb{E}[\psi(P(x))] - \mathbb{E}[\psi(P(z))]| \leqslant \ell^2 9^\ell \varepsilon^2 .$$

To prove the theorem, we first discuss the PRG construction of [MZ13]. Let $t = 1/\varepsilon^2$ and $M = N/t$. Let $\mathcal{H} : [N] \to [t]$ be a family of almost pairwise independent hash functions[10] and let $\mathcal{D} \equiv \mathcal{D}_{4\ell}$ be a (4ℓ)-wise independent distribution over $\{\pm 1\}^m$. The PRG of [MZ13], $G_{\mathcal{H},\mathcal{D}}$, can now be defined by the following algorithm:

[10] A hash family $\mathcal{H}$ is almost pairwise independent if for every $i \neq j \in [N]$, $a, b \in [t]$, $\mathbb{P}_{h \in_u \mathcal{H}}[h(i) = a \wedge h(j) = b] \leqslant (1+\alpha)/t^2$ for $\alpha = O(1)$.

1. Choose a random $h \in \mathcal{H}$ and partition $[N]$ into t blocks $B_1, \dots, B_t$, with $B_j = \{i : h(i) = j\}$.
2. Choose independent samples $x_1, \dots, x_t \leftarrow \mathcal{D}$ and let $y \in \{\pm 1\}^N$ be chosen according to an arbitrary distribution independent of $x_1, \dots, x_t$.
3. Output[11]

$$(5.2) \quad z' \in \{\pm 1\}^N, \text{ with } z'_i = z_i \cdot y_i \text{ for } i \in [N], \text{ where } z_{|B_j} = x_j \text{ for } j \in [t].$$

Meka and Zuckerman show that $G_{\mathcal{H},\mathcal{D}}$ as above fool (arbitrary) low-degree polynomials. Below we state their result for regular PTFs, which suffices for our purposes and gives better quantitative bounds.

THEOREM 5.10 (Lemma 5.10 in [MZ13]). *Let $P : \mathbb{R}^N \to \mathbb{R}$ be an ε-regular multilinear polynomial of degree at most ℓ. Then, for $x \in_u \{\pm 1\}^N$, and $y \in \{\pm 1\}^N$ generated according to $G_{\mathcal{H},\mathcal{D}}$,*

$$|\mathbb{E}[\psi(P(x))] - \mathbb{E}[\psi(P(y))]| \leqslant \frac{1}{3}\ell^2 9^\ell \varepsilon^2 .$$

We next show that the uniform distribution over $\mathsf{RM}(n,d)$ for a sufficiently high d is equivalent to $G_{\mathcal{H},\mathcal{D}}$ as above for an appropriately chosen hash family $\mathcal{H}$ and (4ℓ)-wise independent distribution $\mathcal{D}$. Below we identify $[N]$ with $\mathbb{F}_2^n$ and $[t]$ with $\mathbb{F}_2^c$ for $c = 2\log(1/\varepsilon)$.

Proof of Theorem 5.9. For simplicity, in the following discussion we view $\mathsf{RM}(n,d)$ as generating a vector in $\mathbb{F}_2^N$ and show that the uniform distribution over $\mathsf{RM}(n,d)$ has the appropriate independence structure as required by Theorem 5.10, albeit with $\{\pm 1\}$ replaced with $\{0,1\}$. This does not affect the analysis of the generator.

At the outset, the idea is to view random low-degree polynomials as an instantiation of the Meka and Zuckerman PRG. Fix $\mathcal{D}$ to be the uniform distribution over polynomials of degree $d-c$ over $n-c$ variables for $c = 2\log(1/\varepsilon)$. Let $\mathcal{H}$ denote the family of hash functions from $\mathbb{F}_2^n \to \mathbb{F}_2^c$ given by affine maps. The family $\mathcal{H}$ is almost pairwise independent. The Meka–Zuckerman construction will partition the $[N] = \mathbb{F}_2^n$ coordinates into 2^c buckets using the hash family $\mathcal{H}$, and within each bucket assign a sample from distribution $\mathcal{D}$, namely an $(n-c)$·variate, degree-$(d-c)$ polynomial. The key observation is that the resulting function on $\mathbb{F}_2^n$ (over all buckets) is an n-variate polynomial of degree at most d. Therefore, uniform distribution over $\mathsf{RM}(n,d)$ can be expressed as a mixture of Meka–Zuckerman PRG constructions. The details of the argument are presented below.

Let $c = 2\log(1/\varepsilon)$, and let $\mathcal{S}$ be the subspace of polynomials of the form

$$Q_1(x_1, \dots, x_n) = \sum_{a \in \{0,1\}^c} \mathbb{1}(x_{|[c]} = a) \cdot P_a(x_{c+1}, \dots, x_n),$$

where the polynomials P_a each have degree at most $d-c$. Note that we can sample a uniformly random element $Q_1 \in \mathcal{S}$ by choosing independent, uniformly random degree at most $d-c$ polynomials $P_a : \mathbb{F}_2^{n-c} \to \mathbb{F}_2$ for $a \in \{0,1\}^c$ and setting Q_1 as above. This is because each collection $(P_a)_{a \in \{0,1\}^c}$ leads to a unique element of $\mathcal{S}$, and together they cover all elements of $\mathcal{S}$.

Let $\mathcal{S}'$ be a subspace of degree-d, n-variate polynomials such that $\mathcal{S} \cap \mathcal{S}' = \{0\}$ and $\mathcal{S}, \mathcal{S}'$ together span all degree-d polynomials. Let $\mathcal{A} : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be the space of all

[11] The description we give here is slightly different from that of [MZ13] due to the presence of the string y. However, the analysis of [MZ13] works without any changes for this case as well.

affine transformations. For $A \in \mathcal{A}$, let $h_A : [N] \to [t]$ be defined by $h_A(x) = A(x)_{|[c]}$, and let $\mathcal{H} \equiv \{h_A : A \in \mathcal{A}\}$. It is easy to see that for $A \in_u \mathcal{A}$, the hash functions h_A are almost pairwise independent. Observe that for $Q_1 \in_u \mathcal{S}, Q_2 \in_u \mathcal{S}'$, and $A \in_u \mathcal{A}$, the polynomial $Q(\ ) = (Q_1 + Q_2)(A(\ ))$ is uniformly distributed over all n-variate degree-d polynomials.

Now, fix a polynomial $Q_2 \in \mathcal{S}'$. Then, for a random $Q_1 \in_u \mathcal{S}$, we have

$$Q(x) = \sum_{a \in \{0,1\}^c} 1(h_A(x) = a) \cdot P_a(u_{a+1}, \dots, u_n) + Q_2(u),$$

where $u = Ax$ and the polynomials $(P_a)_{a \in \{0,1\}^c}$ are independent uniformly random polynomials of degree at most $d - c$ in $n - c$ variables. Let $\mathcal{D}$ denote the distribution of $(P'(u))_{u \in \mathbb{F}_2^{n-c}}$ for P' a uniformly random polynomial of degree at most $d - c$ in $n - c$ variables. Then, for every fixed $A \in \mathcal{A}$ and $Q_2 \in \mathcal{S}'$, the distribution of the evaluations of Q restricted to different *buckets* $B_a = \{x : h_A(x) = a\}$ are independent of one another. Moreover, within each bucket B_a, the evaluations vector $(Q_1(x))_{x \in B_a}$ is distributed as $\mathcal{D}$, which is $(2^{d-c} - 1)$-wise independent.

Therefore, for every fixed $Q_2 \in \mathcal{S}'$, the distribution of $z = (Q(x))_{x \in \mathbb{F}_2^n}$ is the same as the output of $G_{\mathcal{H},\mathcal{D}}$ as defined in (5.2), where $y = Q_2(A(x))$. The theorem now follows from Theorem 5.10. ∎

The invariance principle of Theorem 5.9 combined with the appropriate choice of the smooth function ψ gives us the following corollaries.

Proof of Theorem 5.7. The proof follows from using Theorem 5.9 and an argument as in Theorem 3.19 of [MOO05], where the authors get a similar conclusion for the hypercube starting from an invariance principle for the hypercube to the Gaussian space. ∎

COROLLARY 5.11. *Let $N = 2^n$ and $d \geqslant \log \ell + 2\log(1/\varepsilon) + 2$. Let $P : \mathbb{R}^N \to \mathbb{R}$ be an ε-regular polynomial of degree at most ℓ. Then, for $x \in_u \{\pm 1\}^N$, $z \in_u \mathsf{RM}(n,d)$,*

$$|\mathbb{E}[\text{sign}(P(x))] - \mathbb{E}[\text{sign}(P(z))]| \leqslant O(\ell \varepsilon^{1/(2\ell+1)}).$$

Proof. The proof follows from Theorem 5.9 and Lemma 5.8 in [MZ13]. ∎

Finally a similar argument in the proof Theorem 5.9, using a minor modification of the full analysis of the PRG from [MZ13, Theorem 5.17], shows that Reed–Muller codes with $d = \Omega(\ell \log(1/\varepsilon))$ fool all degree-ℓ PTFs. We exclude the proof in this work as we do not need the more general statement in our applications

THEOREM 5.12. *There exists a constant $C > 0$ such that the following holds. Let $N = 2^n$ and $d = C\ell \log(1/\varepsilon)$. Let $P : \mathbb{R}^N \to \mathbb{R}$ be a multilinear polynomial of degree at most ℓ. Then, for $x \in_u \{\pm 1\}^N$, $z \in_u \mathsf{RM}(n,d)$,*

$$|\mathbb{E}[\text{sign}(P(x))] - \mathbb{E}[\text{sign}(P(z))]| \leqslant \varepsilon.$$

5.3. Invariance principles over codes. Our main tool for proving the Majority Is Stablest result over Reed–Muller codes, Theorem 5.6, was the invariance principle, Theorem 5.7. We conjecture that similar results should hold for any linear code with sufficiently large dual distance so that the codewords have bounded independence. In particular, we conjecture that bounded independence fools arbitrary low-degree PTFs over $\{\pm 1\}^n$.

The conjecture is known to be true for half-spaces [DGJ+09], degree-2 PTFs [DKN10], and for Gaussians with bounded independence [Kan11].

CONJECTURE 5.13. *For all $d \in \mathbb{N}$ and $\varepsilon > 0$, there exists $k = k(d, \varepsilon)$ such that the following holds: Let Q be an n-variate multilinear real polynomial with degree d.*

Let X be a k-wise independent distribution over $\{\pm 1\}^n$, and let Y be the uniform distribution over $\{\pm 1\}^n$. Then, $|\mathbb{E} \operatorname{sign} \circ Q(X) - \mathbb{E} \operatorname{sign} \circ Q(Y)| \leqslant \varepsilon$.

Finally, we remark that for the application to Majority Is Stablest it suffices to show a weaker invariance principle applicable to the ζ functional.

Conjecture 5.14. *For all $d \in \mathbb{N}$ and $\varepsilon > 0$, there exist $k = k(d, \varepsilon)$ and $\eta = \eta(\varepsilon)$ such that the following holds: Let Q be an n-variate multilinear real polynomial with degree d. Let X be a k-wise independent distribution over $\{\pm 1\}^n$, and let Y be the uniform distribution over $\{\pm 1\}^n$. Suppose that $\mathbb{E}\, Q(X)^2 \leqslant 1$ and $\mathbb{E}\, \zeta \circ Q(X) \leqslant \eta$. Then, $\mathbb{E}\, \zeta \circ Q(Y) \leqslant \varepsilon$.*

We show in the appendix that Conjecture 5.13 implies Conjecture 5.14.

Lemma 5.15. *Let X be a 20ℓ-wise independent distribution over $\{\pm 1\}^N$ that ε-fools every τ-regular degree-ℓ PTF. Then, for every τ-regular N-variate multilinear real polynomial Q with degree at most ℓ and $\mathbb{E}\, Q(X) \leqslant 1$, we have for the uniform distribution Y over $\{\pm 1\}^N$,*

$$\mathbb{E}\, \zeta \circ Q(Y) \leqslant \mathbb{E}\, \zeta \circ Q(X) + 2^{O(\ell)} \varepsilon^{0.9}\,.$$

6. Efficient integrality gaps for Unique Games.

In this section, we present constructions of SDP integrality gap instances starting from a code $\mathcal{C}$ along with a local tester. To this end, we make an additional assumption on the code $\mathcal{C}$. Specifically, let us suppose there exists a subcode $\mathcal{H}$ of $\mathcal{D} = \mathcal{C}^\perp$ with distance $\frac{1}{2}$. Formally, we show the following result.

Theorem 6.1. *Let $\mathcal{C}$ be an $[N, K, D]_2$ linear code with a canonical tester $\mathcal{T}$ as described in Definition 3.2. Furthermore, let $\mathcal{H}$ be a subcode of $\mathcal{D} = \mathcal{C}^\perp$ with distance $\frac{1}{2}$. Then, there exists an instance of* Unique Games, *more specifically an $\mathcal{H}$-*Max-2Lin *instance, whose vertices are $\mathcal{D}$ ($|\mathcal{D}| = 2^{N-K}$) and alphabet $\mathcal{H}$ such that*

- *The optimum value of the natural SDP relaxation for* Unique Games *is at least $\left(1 - \frac{2t}{N}\right)^2$, where t is the number of queries made by the canonical tester $\mathcal{T}$.*
- *No labelling satisfies more than a*

$$\min_{k \in [0, D/5]} \left(1 - 2s(k) + \frac{3^k}{|\mathcal{H}|^{\frac{1}{2}}}\right)$$

fraction of constraints.

Instantiating the above theorem with the Reed–Muller code and its canonical tester, we obtain the following explicit SDP integrality gap instance.

Corollary 6.2. *For every integer n, $\delta > 0$, there exists an $\mathbb{F}_2^n$-*Max-2Lin *instance Γ on $M = 2^{2^{\log^2 n}}$ vertices such that the optimum value of the SDP relaxation on Γ is $1 - O(\frac{\log(1/\delta)}{n}) = 1 - O(\frac{\log(1/\delta)}{2^{(\log\log M)^{1/2}}})$, while every labelling of Γ satisfies at most a $O(\delta)$ fraction of edges.*

Proof. Fix the code $\mathcal{C}$ to be the Reed–Muller code $\mathsf{RM}(n, n - \log n)$ of degree $d = n - \log n$ over n variables. Its dual $\mathcal{D} = \mathcal{C}^\perp = \mathsf{RM}(n, \log n)$ consists of polynomials of degree $d = \log n$. The block length of the code $\mathcal{D}$ is $N = 2^n$, while the rate is $K = 2^n - \sum_{i \leqslant d} \binom{n}{i} \leqslant 2^n - O(2^{\log^2 n})$. This code contains the Hadamard code $\mathcal{H}$ which is of relative distance $\frac{1}{2}$.

Let $\mathcal{T}_{\mathsf{RM}}$ denote the canonical Reed–Muller tester for $\mathsf{RM}(n, n - \log n)$, and let $\mathcal{T}_{\mathsf{RM}}^{\oplus r}$ denote the XOR of r-independent tests. Let us fix $r = 100 \log(1/\delta)$, thus yielding a canonical tester making $t = \log(1/\delta) \cdot 2^{n-d}$ queries. By the work of [BKS+10], this tester has a soundness of at least $s(k) = \frac{1}{2} - (1 - k/2^{d+1})^r/2$. With $k = 2^d/10$, the above soundness is at least $s(k) \geqslant 1/2 - \delta/2$. Using Theorem 6.1, the optimum value of the resulting $\mathbb{F}_2^n$-Max-2Lin instance is at most δ. On the other hand, the SDP value is at least

$$(1-2t/N)^2 = 1 - 100\log(1/\delta)2^{n-d}/2^n = 1 - O\left(\frac{\log 1/\delta}{2^d}\right) = 1 - \frac{\log(1/\delta)}{n}\,. \qquad \square$$

Starting from $\mathcal{C}$, we construct an SDP integrality gap instance $\Gamma(\mathcal{C},\mathcal{T})$ for Unique Games as described below. The integrality gap instance can be thought of as a derandomization of the Unique Games integrality gap instance constructed by Khot and Vishnoi [KV05].

In the construction of Khot and Vishnoi [KV05], the code $\mathcal{D}$ consists of all strings over $\mathbb{F}_2^n$, which clearly contains the Hadamard code $\mathcal{H}$. The vertices of the Unique Games instance are elements of $\mathcal{D} = \mathbb{F}_2^n$, and the edges are given by the noisy hypercube graph. The set of labels are the Hadamard codewords $\mathcal{H}$. Hence, a labelling $\ell : \mathcal{D} \to \mathcal{H}$ assigns a Hadamard codeword for each point in $\mathbb{F}_2^n$. The constraints of the unique game are set up to ensure that *nearby points receive the same label*. Specifically, if (c, c') is an edge in the noisy hypercube graph, then it is natural to include the constraint $\ell(c) = \ell(c')$. Equality constraints alone would result in a trivial Unique Games instance which is satisfied by labelling all vertices with the same label. Hence, we will also enforce the condition that the labelling *obeys the linear structure of cosets of Hadamard code*. Specifically, if $c \in \mathbb{F}_2^n$ and $h \in \mathcal{H}$, we will want $\ell(c+h) = \ell(c) + h$. The constraints of the Khot–Vishnoi integrality gap instance are obtained by incorporating both of the above-mentioned tests into the constraints of the instance. The instance admits a very natural SDP solution with value close to 1, arising out of the natural embedding of $\mathbb{F}_2^n$ as $\{-1,1\}^n$. On the other hand, Small-Set Expansion of the noisy hypercube can be used to show that the instance does not admit good labellings.

We will construct derandomizations of the Khot–Vishnoi instance by using a derandomization of the noisy hypercube. In particular, we will replace $\mathcal{D} = \mathbb{F}_2^n$ by a different code of much smaller size, but which retains the Small-Set Expansion property of the noisy hypercube. The SDP solution is again an immediate consequence of the natural embedding of $\mathcal{D}$ into $\{-1,1\}^n$. Analogous to the Khot–Vishnoi example, the soundness argument is again a direct consequence of the Small-Set Expansion of the underlying graph. The formal description of the integrality gap instance is presented below.

The vertices of $\Gamma(\mathcal{C},\mathcal{T})$ are the codewords of $\mathcal{D}$. The alphabet of the Unique Games instance $\Gamma(\mathcal{C},\mathcal{T})$ are the codewords in $\mathcal{H}$. The constraints of Unique Games instance $\Gamma(\mathcal{C},\mathcal{T})$ are given by the tests of the following verifier.
The input to the verifier is a labelling $\ell : \mathcal{D} \to \mathcal{H}$. Let us denote by $R = |\mathcal{H}|$. The verifier proceeds as follows:

- Sample codewords $c \in \mathcal{D}$ and $h, h' \in \mathcal{H}$ uniformly at random.
- Sample a codeword $q \in \mathcal{D}$ from the tester $\mathcal{T}$.
- Test if

$$\ell(c+q+h) - \ell(c+h') = h - h'.$$

SDP solution. Here we construct SDP vectors that form a feasible solution to a natural SDP relaxation of Unique Games [KV05].

(6.1)

$$\text{Maximize} \mathop{\mathbb{E}}_{c\in\mathcal{D},h,h'\in\mathcal{H}} \mathop{\mathbb{E}}_{q\in\mathcal{T}} \left[\frac{1}{R}\sum_{\ell\in\mathcal{H}} \langle \boldsymbol{b}_{c+h',\ell+h}, \boldsymbol{b}_{c+q+h,\ell+h'}\rangle\right]$$

(6.2)

$$\text{subject to } \langle \boldsymbol{b}_{c,h}, \boldsymbol{b}_{c,h'}\rangle = 0 \qquad \forall c\in\mathcal{D}, h\neq h'\in\mathcal{H},$$

$$(6.3) \qquad \langle \boldsymbol{b}_{c,h}, \boldsymbol{b}_{c',h'}\rangle \geqslant 0 \qquad \forall c,c'\in\mathcal{D}, h,h'\in\mathcal{H},$$

$$(6.4) \qquad \sum_{\ell\in\mathcal{H}} \langle \boldsymbol{b}_{c,\ell}, \boldsymbol{b}_{c,\ell}\rangle = R \qquad \forall c\in\mathcal{D}.$$

For a vector $c \in \mathbb{F}_2^m$, we will use $(-1)^c \in \mathbb{R}^m$ to denote the vector whose coordinates are given by $(-1)^c_i = (-1)^{c_i}$. For a pair of vectors c, c', we have

$$\langle (-1)^c, (-1)^{c'}\rangle = 1 - 2\Delta(c,c')\,.$$

For each vertex $c \in \mathcal{D}$ associate vectors $\{\boldsymbol{b}_{c,h} = (-1)^{c+h} \otimes (-1)^{c+h} | h \in \mathcal{H}\}$. Notice that for a pair of vectors $\boldsymbol{b}_{c,h}, \boldsymbol{b}_{c',h'}$, we have

$$\langle \boldsymbol{b}_{c,h}, \boldsymbol{b}_{c',h'}\rangle = \langle (-1)^{c+h}, (-1)^{c'+h'}\rangle^2 = (1 - 2\Delta(c+h, c'+h'))^2\,.$$

Since the distance of the code $\mathcal{H}$ is $\frac{1}{2}$, we have

$$(6.5) \qquad \langle \boldsymbol{b}_{c,h}, \boldsymbol{b}_{c,h'}\rangle = (1-2\Delta(h,h'))^2 = \begin{cases} 1 & \text{if } h = h', \\ 0 & \text{if } h \neq h'. \end{cases}$$

In other words, for every vertex c, the corresponding SDP vectors are orthonormal. The objective value of the SDP solution is given by

$$\begin{aligned} \mathsf{OBJ} &= \mathop{\mathbb{E}}_{c\in\mathcal{D},h,h'\in\mathcal{H}} \mathop{\mathbb{E}}_{q\in\mathcal{T}} \left[\frac{1}{R}\sum_{\ell\in\mathcal{H}} \langle \boldsymbol{b}_{c+h',\ell+h}, \boldsymbol{b}_{c+q+h,\ell+h'}\rangle\right] \\ &= \mathop{\mathbb{E}}_{c\in\mathcal{D},h\in\mathcal{H}} \mathop{\mathbb{E}}_{q\in\mathcal{T}} \left[\frac{1}{R}\sum_{\ell\in\mathcal{H}} (1-2\Delta(c+h'+\ell+h, c+q+h+\ell+h'))^2\right] \\ &= \mathop{\mathbb{E}}_{c\in\mathcal{D},h\in\mathcal{H}} \mathop{\mathbb{E}}_{q\in\mathcal{T}} \left[(1-2\Delta(0,q))^2\right] \\ &\geqslant \left(1 - \frac{2t}{N}\right)^2, \end{aligned}$$

where t is the number of queries made by the canonical tester $\mathcal{T}$ for $\mathcal{C}$.

Soundness. Let $\ell : \mathcal{D} \to \mathcal{H}$ be an arbitrary labelling of the UNIQUE GAMES instance $\Gamma(\mathcal{C},\mathcal{T})$. For each $p \in \mathcal{H}$, define a function $f_p \colon \mathcal{D} \to [0,1]$ as follows:

$$f_p(c) = \mathop{\mathbb{E}}_{h\in\mathcal{H}} \left[\mathbb{I}[\ell(c+h) = p+h]\right].$$

The fraction of constraints satisfied by the labelling ℓ is given by

$$\mathsf{OBJ} = \mathop{\mathbb{E}}_{c\in\mathcal{D},h,h'\in\mathcal{H}} \mathop{\mathbb{E}}_{q\in\mathcal{T}} \left[\sum_{p\in\mathcal{H}} \mathbb{I}[\ell(c+h') = p+h'] \cdot \mathbb{I}[\ell(c+q+h) = p+h]\right]$$

$$= \mathop{\mathbb{E}}_{c\in\mathcal{D}} \mathop{\mathbb{E}}_{q\in\mathcal{T}} \left[\sum_{p\in\mathcal{H}} \mathop{\mathbb{E}}_{h'\in\mathcal{H}} \mathbb{I}[\ell(c+h') = p+h'] \cdot \mathop{\mathbb{E}}_{h\in\mathcal{H}} \mathbb{I}[\ell(c+q+h) = p+h] \right]$$

$$= \mathop{\mathbb{E}}_{c\in\mathcal{D}} \mathop{\mathbb{E}}_{q\in\mathcal{T}} \left[\sum_{p\in\mathcal{H}} f_p(c) f_p(c+q) \right] \tag{6.6}$$

$$= \sum_{p\in\mathcal{H}} \langle f_p, G f_p \rangle, \tag{6.7}$$

where $G = \mathrm{Cay}(\mathcal{C}^\perp, \mathcal{T})$ is the graph associated with the code $\mathcal{C}^\perp$ and tester $\mathcal{T}$.

The expectation of the function f_p is given by

$$\begin{aligned}\mathop{\mathbb{E}}_{c\in\mathcal{D}} f_p(c) &= \mathop{\mathbb{P}}_{c\in\mathcal{D}, h\in\mathcal{H}} [\ell(c+h) = p+h] \\ &= \mathop{\mathbb{P}}_{c\in\mathcal{D}, h\in\mathcal{H}} [\ell(c) = p+h] \text{ because } (c+h, h) \sim (c, h) \\ &= \frac{1}{|\mathcal{H}|} = \frac{1}{R}.\end{aligned}$$

Since f_p is bounded in the range $[0, 1]$ we have

$$\langle f_p, f_p \rangle = \mathop{\mathbb{E}}_{c\in\mathcal{D}} [f_p(c)^2] \leqslant \mathop{\mathbb{E}}_{c\in\mathcal{D}} [f_p(c)] = \frac{1}{R}.$$

Applying Corollary 4.10, we get that for each p,

$$\langle f_p, G f_p \rangle \leqslant \frac{1}{R} \cdot \min_{k\in[0,\frac{\mathcal{D}}{5}]} \left(1 - 2s(k) + \frac{3^k}{R^{1/2}} \right).$$

Substituting the previous equation into (6.7), we get that the fraction of constraints satisfied by ℓ is at most

$$\min_{k\in[0,\frac{\mathcal{D}}{5}]} \left(1 - 2s(k) + \frac{3^k}{R^{1/2}} \right).$$

Appendix A. Efficient alphabet reduction. The *long code* over a (nonbinary) alphabet Q consists of the set of dictator functions $\{f_1, \ldots, f_N : Q^N \to Q\}$, where $f_i(x) = x_i$ for all $x \in Q^N$.

A natural 2-query test for this code was proposed by Khot et al. [KKMO07] and analyzed in Mossel, O'Donnell, and Oleszkiewicz [MOO05]. The queries of the test are associated with the edges of the ε*-noise graph* on Q^N. In this graph, the weight of an edge (x, y) is its probability in the following sampling procedure: Sample $x \in Q^N$ uniformly at random and resample each coordinate of $x \in Q^N$ independently with probability ε to generate $y \in Q^N$.

In this section, we present a more efficient code that serves as an analogue for the long code over a nonbinary alphabet. For $n, d \in \mathbb{N}$, let $N = 2^n$, and let $\mathcal{C} \subseteq \mathbb{F}_2^N$ be the Reed–Muller code $\mathsf{RM}(n, n-d-1)$ and let $\mathcal{D} = \mathcal{C}^\perp \in \mathbb{F}_2^N$ be its dual $\mathsf{RM}(n, d)$. Let $\mathcal{T} \subseteq \mathcal{D}$ denote the canonical test set for the code $\mathcal{C}$ as in section 4.3.

Let $t \in \mathbb{N}$ and let $Q = \mathbb{F}_2^t$. We define the following distribution $\mathcal{T}_t$ over $\mathcal{D}^t$ (the t-fold direct sum of $\mathcal{D}$, a subspace of $\mathbb{F}_2^{t\cdot N}$):

- Sample c from the test set $\mathcal{T} \subseteq \mathcal{D}$.
- Sample $w = (w^{(1)}, \ldots, w^{(t)})$ from $\mathbb{F}_2^t$ at random.
- Sample $z = (z^{(1)}, \ldots, z^{(t)}) \in \mathcal{D}^t$ by setting

$$z^{(i)} = \begin{cases} c & \text{if } w^{(i)} = 1\,, \\ 0 & \text{if } w^{(i)} = 0\,. \end{cases}$$

Consider the continuous-time random walk on the graph $\mathrm{Cay}(\mathcal{D}^t, \mathcal{T}_t)$ with parameter $\varepsilon \cdot 2^d$ (starting at point $0 \in \mathcal{D}^t$). Let $\mathcal{T}_{\varepsilon,t}$ be the distribution over $\mathcal{D}^t$ corresponding to this random walk. The Cayley graph $\mathrm{Cay}(\mathcal{D}^t, \mathcal{T}_{\varepsilon,t})$ will serve us as an analogue of the ε-noise graph on Q^N.

Spectrum. In the following we will demonstrate that (part of) the spectrum of the Cayley graph $\mathrm{Cay}(\mathcal{D}^t, \mathcal{T}_{t,\varepsilon})$ corresponds to the spectrum of the ε-noise graph on Q^N. To this end, we recall the spectrum of the ε-noise graph on Q^N. First, we define a convenient basis for the functions on $Q = \mathbb{F}_2^t$. We will denote the coordinates of a vector $\alpha \in Q = \mathbb{F}_2^t$ by $\alpha = (\alpha^{(1)}, \ldots, \alpha^{(t)})$. The set of characters of $\mathbb{F}_2^t$ is $\{\chi_\alpha \colon \mathbb{F}_2^t \to \{\pm 1\} \mid \alpha \in \mathbb{F}_2^t\}$, where

$$\chi_\alpha(x) = (-1)^{\sum_j \alpha^{(j)} x^{(j)}}\,.$$

Since the noise graph on Q^N is a Cayley graph over the abelian group $\mathbb{F}_2^{tN}$, the characters of this group form a basis of eigenfunctions. For $\beta = (\beta_1, \ldots, \beta_N) \in Q^N$, let $\chi_\beta \colon Q^N \to \{\pm 1\}$ denote the character

$$\chi_\beta(x_1, \ldots, x_N) = \prod_{i \in [N]} \chi_{\beta_i}(x_i)\,.$$

The eigenvalue of χ_β in the ε-noise graph on Q^N $(1-\varepsilon)^{\mathsf{hwt}(\beta)}$, where $\mathsf{hwt}(\beta) = |\{i \mid \beta_i \neq 0^t\}|$, is the Hamming weight of β as a length-N string over alphabet Q. (In this section $\mathsf{hwt}(\beta)$ will always refer to the Hamming weight of string β over alphabet Q.)

The canonical eigenfunctions of $\mathrm{Cay}(\mathcal{D}^t, \mathcal{T}_t)$ and $\mathrm{Cay}(\mathcal{D}^t, \mathcal{T}_{t,\varepsilon})$ are indexed by $\beta \in Q^N/\mathcal{C}^t$. (Note that $\mathcal{C}^t$ is the orthogonal complement of $\mathcal{D}^t$.) Analogously to the definition in section 4.2, we define the degree of a character $\chi_\beta \colon \mathcal{D}^t \to \{\pm 1\}$ for $\beta \in Q^N/\mathcal{C}^t$ as

$$\deg(\chi_\beta) = \mathsf{wt}(\beta) = \min_{\beta' \in \beta} \mathsf{hwt}(\beta')\,,$$

where $\mathsf{hwt}(\beta') = |\{i \in [N] \mid \beta'_i \neq 0^t\}|$ is the Hamming weight of β' seen as a length-N string over alphabet Q. (Here, the minimum is over all $\beta' \in Q^N$ that lie in the same coset as β in $Q^N/\mathcal{C}^t$.)

The following lemma is an analogue of Lemmas 3.3 and 4.13 and shows that the eigenvalues of $\mathrm{Cay}(\mathcal{D}^t, \mathcal{T}_t)$ are similar to the eigenvalues of the ε-noise graph.

Lemma A.1. *Let $\beta \in Q^N/\mathcal{C}^t$. The eigenvalue λ_β of the character χ_β in the graph* $\mathrm{Cay}(\mathcal{D}^t, \mathcal{T}_t)$ *satisfies $\lambda_\beta = 1 - \mathsf{wt}(\beta)/2^d \pm O(\mathsf{wt}(\beta)/2^d)^2$ and $\lambda_\beta \leqslant 1 - \Omega(1/t) \cdot \min\{\mathsf{wt}(\beta) \cdot 2^{-d}, 1\}$.*

Proof. We will first prove an upper bound on λ_β for the case that $\mathsf{wt}(\beta) \gg 2^d$. We write $\beta = (\beta^{(1)}, \ldots, \beta^{(t)})$ with $\beta^{(i)} \in \mathbb{F}_2^N$. Let $z = (z^{(1)}, \ldots, z^{(t)}) \in \mathcal{D}^t$ be a string drawn from the distribution $\mathcal{T}_t$. Note that $z^{(i)} = w^{(i)} \cdot c$, where $w = (w^{(1)}, \ldots, w^{(t)})$ and c are sampled as in the definition of $\mathcal{T}_t$. Since w is a random vector in $\mathbb{F}_2^t$, we can upper bound λ_β:

$$\begin{aligned}\lambda_\beta &= \mathop{\mathbb{E}}_{z}(-1)^{\langle\beta,z\rangle}\\ &= 1-2\mathop{\mathbb{P}}_{w\in\mathbb{F}_2^t,\ c\in\mathcal{T}}\left\{\sum_{i=1}^t w^{(i)}\langle\beta^{(i)},c\rangle=1\right\}\\ &= 1-\mathop{\mathbb{P}}_{c\in\mathcal{T}}\left\{\exists i,\ \langle\beta^{(i)},c\rangle=1\right\}\\ &\leqslant 1-\max_{i\in[t]}\mathop{\mathbb{P}}_{c\in\mathcal{T}}\left\{\langle\beta^{(i)},c\rangle=1\right\}.\end{aligned}$$

Without loss of generality, we may assume that $\beta^{(t)}$ has Hamming weight (as a binary string) at least $\mathsf{wt}(\beta)/t$. By Theorem 4.11, if $\mathsf{wt}(\beta)>\eta 2^{-d}$ for sufficiently small $\eta>0$, we can upper bound $\lambda_\beta\leqslant 1-\Omega(\eta/t)$.

Next, we will estimate λ_β (from below and above) for $\mathsf{wt}(\beta)\ll 2^{-d}$. Let $I\subseteq[N]$ be the set of coordinates $i\in[N]$ with $\beta_i\neq 0^t$. We claim

$$\lambda_\beta = 1-\mathop{\mathbb{P}}_{c}\{|I\cap\operatorname{supp}(c)|=1\}\pm O(1)\cdot\mathop{\mathbb{P}}_{c}\{|I\cap\operatorname{supp}(c)|\geqslant 2\}.$$

We write $\beta=(\beta_1,\dots,\beta_N)$ with $\beta_i\in\mathbb{F}_2^t$. Then, $\langle\beta,z\rangle=\sum_{i\in[N]}c_i\langle w,\beta_i\rangle$. We refine the event $\langle\beta,z\rangle=1$ according to the cardinality of $I\cap\operatorname{supp}(c)$. If $I\cap\operatorname{supp}(c)=\emptyset$, then $\langle\beta,z\rangle=0$. On the other hand, conditioned on $|I\cap\operatorname{supp}(c)|$, the event $\langle\beta,z\rangle=1$ is equivalent to the event $\langle w,\beta_{i_0}\rangle=1$ with $\{i_0\}=I\cap\operatorname{supp}(c)$. Since $\beta_{i_0}\neq 0^t$, this event has (conditional) probability $1/2$. Hence,

$$\mathop{\mathbb{P}}_{z}\left\{\langle\beta,z\rangle=1\right\}=\tfrac12\mathop{\mathbb{P}}_{c\in\mathcal{T}}\left\{|I\cap\operatorname{supp}(c)|=1\right\}\pm\mathop{\mathbb{P}}_{c\in\mathcal{T}}\left\{|I\cap\operatorname{supp}(c)|\geqslant 2\right\},$$

which implies the claimed estimate for λ_β.

It remains to estimate the distribution of $|I\cap\operatorname{supp}(c)|$. The argument is similar to the proof of Lemma 3.3. For every coordinate $i\in[N]$, we have $\mathbb{P}_{c\in\mathcal{T}}\{c_i=1\}=2^{-d}$. Thus, $\mathbb{P}\{|I\cap\operatorname{supp}(c)|=1\}\leqslant|I|\cdot 2^{-d}=\mathsf{wt}(\beta)/2^d$. On the other hand, for any two distinct coordinates $i\neq j\in[N]$, we have $\mathbb{P}_{c\in\mathcal{T}}\{c_i=c_j=1\}=2^{-2d}$. Therefore,

$$\mathbb{P}\{|I\cap\operatorname{supp}(c)|=1\}\geqslant\sum_{i\in I}\mathbb{P}\{c_i=1\}-\sum_{i<j\in I}\mathbb{P}\{c_i=c_j=1\}\geqslant\mathsf{wt}(\beta)/2^d-(\mathsf{wt}(\beta)/2^d)^2.$$

Similarly, $\mathbb{P}\{|I\cap\operatorname{supp}(c)|\geqslant 2\}\leqslant(\mathsf{wt}(\beta)/2^d)^2$. We conclude that

$$\lambda_\beta=1-\mathsf{wt}(\beta)/2^d\pm O(\mathsf{wt}(\beta)/2^d)^2.$$

(Note that the estimate is meaningful only when $\mathsf{wt}(\beta)\ll 2^d$.) □

If the character χ_β has eigenvalue λ_β in the graph $\operatorname{Cay}(\mathcal{D}^t,\mathcal{T}_t)$, then it has eigenvalue $e^{-\varepsilon(1-\lambda_\beta)/2^d}$ in $\operatorname{Cay}(\mathcal{D}^t,\mathcal{T}_{t,\varepsilon})$. Similarly to Lemma 4.13, the eigenvalue of a character χ_β is close to $e^{-\varepsilon\mathsf{wt}(\beta)}$ in the graph $\operatorname{Cay}(\mathcal{D}^t,\mathcal{T}_{t,\varepsilon})$.

Lemma A.2.

– *If* $\mathsf{wt}(\beta)\leqslant\delta^2 2^d$ *for sufficiently small* δ*, then the character* χ_β *has eigenvalue* $e^{-\varepsilon\cdot\mathsf{wt}(\beta)}\pm\delta$ *in the graph* $\operatorname{Cay}(\mathcal{D}^t,\mathcal{T}_{t,\varepsilon})$.

– *For an absolute constant* c_0 *and all* $\beta\in Q^N/\mathcal{D}^t$, $\lambda_\beta\leqslant\max(\rho^{\mathsf{wt}(\beta)/c_0t},\rho^{2^d/c_0t})$.

Influences. Let $\beta\in Q^N/\mathcal{C}^t$. Suppose $\mathsf{wt}(\beta)<\mathsf{wt}(\mathcal{C}^t)/2$. (Note that $\mathcal{C}^t\subseteq Q^N$ has the same minimum distance as $\mathcal{C}\subseteq\mathbb{F}_2^N$.) In this case, we will identify β with the (unique) codeword of minimum weight in the equivalence class $\beta\in Q^N/\mathcal{C}^t$.

Definition A.3. *For a function* $f\colon\mathcal{D}^t\to\mathbb{R}$*, a coordinate* $i\in[N]$*, and a degree bound* $\ell<\operatorname{dist}(\mathcal{C}^t)/2$*, we define the* ℓ-degree influence of coordinate i on f *as*

$$\operatorname{Inf}_i^{\leqslant\ell}(f)=\sum_{\beta\in Q^N/\mathcal{C}^t,\ \beta_i\neq 0^t,\ \mathsf{wt}(\beta)\leqslant\ell}\hat f(\beta)^2.$$

(Here, β_i refers to the ith coordinate of the unique minimum weight representative of the equivalence class β.)

A.1. Majority Is Stablest. In this section, we show an analogue of the Majority Is Stablest theorem of [MOO05] on the ε-noise graph on Q^N just as Theorem 5.6 showed an analogue of the Majority Is Stablest theorem over the Boolean noise graph.

THEOREM A.4. *For every $\varepsilon, \delta, t > 0$, there exist L, d, τ such that if G denotes the graph $\mathrm{Cay}(\mathcal{D}^t, \mathcal{T}_{t,\varepsilon})$ constructed using Reed–Muller codes of degree d, then for every function $f\colon \mathcal{D}^t \to [0,1]$ with $\max_{i\in[N]} \mathrm{Inf}_i^{\leqslant L}(f) < \tau$,*

$$\langle f, Gf \rangle \leqslant \Gamma_\rho(\mu) + \delta, \tag{A.1}$$

where $\rho = e^{-\varepsilon}$, $\mu = \mathbb{E}_{x\sim\mathcal{D}^t}[f(x)]$, and $\Gamma_\rho\colon \mathbb{R} \to \mathbb{R}$ is the noise stability curve over Gaussian space.

Given the characterization of the spectrum of $\mathrm{Cay}(\mathcal{D}^t, \mathcal{T}_{t,\varepsilon})$ (Lemma A.2), the proof of Theorem A.4 is similar to that of Theorem 5.6. For the sake of completeness, we include a proof sketch in Appendix C.1.

A.2. 2-query test. We will now describe a dictatorship test for functions on $\mathcal{D}^t$, analogous to the 2-query dictatorship test on the ε-noise graph.

We are interested in functions $f\colon \mathcal{D}^t \to Q$ where $Q = \mathbb{F}_2^t$. Note that $v \in \mathcal{D}^t$ can also be thought of as $v \in Q^N$. For all $\beta \in \mathbb{F}_2^n$, the β^{th} *dictator function* χ_β from $\mathcal{D}^t \subseteq Q^N$ to Q is given by

$$\chi_\beta(c) = c_\beta.$$

Clearly, the dictator functions are linear functions over $\mathcal{D}^t$, i.e., $\chi_\beta(c + c') = \chi_\beta(c) + \chi_\beta(c')$. This linearity is used to perform the 2-query test via *folding*. Note that for each $\alpha \in Q$, the constant function $\alpha(x) = \alpha$ for all $x \in \mathbb{F}_2^n$ belongs to the code $\mathcal{D}^t$. We will *fold* the function by enforcing that for all $\alpha \in Q$, $f(c + \alpha) = f(c) + \alpha$ for all $\alpha \in Q$.

The details of the 2-query dictatorship test are described below.

DICT
Input: $f\colon \mathcal{D}^t \to Q$

Folding. The function is assumed to satisfy $f(c+r) = f(c) + r$ for every $c \in \mathcal{D}^t$ and $r \in Q$. This is enforced by *folding* the table of the function f.

- Sample a vertex $c \in \mathcal{D}^t$.
- Sample a neighbor $c' \in \mathcal{D}^t$ of the vertex c in the Cayley graph $\mathrm{Cay}(\mathcal{D}^t, \mathcal{T}_{t,\varepsilon})$.
- Sample $r \in Q$ uniformly at random.
- Accept if $f(c+r) - r = f(c')$.

Given a function $f\colon \mathcal{D}^t \to Q$, we can arithmetize the value of the test in terms of Q functions $\{f_\alpha\}_{\alpha\in Q}$ that are defined as

$$f_\alpha(x) = \mathbb{I}[f(x) = \alpha]\,.$$

Due to folding, we have $f_\alpha(x) = f_{\alpha+r}(x + r)$ for all $r \in Q$. For each $\alpha \in Q$, the expectation of f_α is given by

$$\mathop{\mathbb{E}}_{c\in\mathcal{D}^t} f_\alpha(c) = \mathop{\mathbb{P}}_{c\in\mathcal{D}^t, r\in Q}[f(c+r) = \alpha] = \frac{1}{Q},$$

where we used the fact that f is folded. The probability of acceptance of the 2-query test can be written in terms of the functions f_α as follows:

$$\begin{aligned}\mathbb{P}[\text{Test accepts } f] &= \sum_{\alpha\in Q} \mathop{\mathbb{E}}_{(c,c')\sim\mathrm{Cay}(\mathcal{D}^t,\mathcal{T}_{t,\varepsilon})} [f_{\alpha+r}(c+r)f_\alpha(c')] \\ &= \sum_{\alpha\in Q} \mathop{\mathbb{E}}_{(c,c')\sim\mathrm{Cay}(\mathcal{D}^t,\mathcal{T}_{t,\varepsilon})} [f_\alpha(c)f_\alpha(c')]\,,\end{aligned}$$

where $(c,c') \sim \mathrm{Cay}(\mathcal{D}^t,\mathcal{T}_{t,\varepsilon})$ denotes a uniformly random edge in the graph $\mathrm{Cay}(\mathcal{D}^t,\mathcal{T}_{t,\varepsilon})$.

Theorem A.5. *The 2-query dictatorship test* DICT *described above satisfies the following completeness and soundness:*

- *(Completeness) Every dictator function $\chi_\beta(x) = x_i$ is accepted by the test with probability at least $1-\varepsilon$.*
- *(Soundness) For every $\delta > 0$, there exist τ, L such that if f satisfies the condition $\max_{i\in[N]} \mathrm{Inf}_i^{\leqslant L}(f_\alpha) \leqslant \tau$ for all $\alpha \in Q$, then f is accepted with probability at most*

$$Q\cdot\Gamma_\rho\left(\tfrac{1}{Q}\right)+\delta\,,$$

where $\rho = e^{-\varepsilon}$.

Completeness. Recall that for a $c \in \mathcal{C}^\perp$ generated from distribution $\mathcal{T}_\varepsilon$, for each $x \in \mathbb{F}_2^n$ (see Lemma 4.5),

$$\mathop{\mathbb{P}}_{c\sim\mathcal{T}_\varepsilon}[c(x)=0] \geqslant 1-O(\varepsilon)\,.$$

It is easy to see that by construction, this property holds for the distribution $\mathcal{T}_{t,\varepsilon}$ also, namely,

$$\mathop{\mathbb{P}}_{c\sim\mathcal{T}_{t,\varepsilon}}[c(x)=0] \geqslant 1-O(\varepsilon)\,.$$

Hence for a random edge (c,c') in the Cayley graph $\mathrm{Cay}(\mathcal{D}^t,\mathcal{T}_{t,\varepsilon})$ and a $\beta \in \mathbb{F}_2^n$, $c(\beta) = c'(\beta)$ with probability $1-\varepsilon$. Therefore, for each $\beta \in \mathbb{F}_2^n$, the βth dictator function satisfies the test with probability $1-O(\varepsilon)$.

Soundness. The probability of acceptance of the 2-query test is given by

$$Pr[\text{Test accepts } f] = \sum_{\alpha\in Q} \mathop{\mathbb{E}}_{(c,c')\sim\mathrm{Cay}(\mathcal{D}^t,\mathcal{T}_{t,\varepsilon})} [f_\alpha(c)f_\alpha(c')]\,.$$

By applying Theorem 5.6, there is an appropriate choice of L, τ such that if $\max_{i\in[N]} \mathrm{Inf}_i^{\leqslant L}(f_\alpha) \leqslant \tau$ for all α, then the probability of acceptance can be bounded by

$$\mathbb{P}[\text{Test accepts } f] = \sum_{\alpha\in Q}\langle f_\alpha, Gf_\alpha\rangle \leqslant Q\cdot\Gamma_\rho\left(\tfrac{1}{Q}\right)+\delta\,,$$

where $\rho = e^{-\varepsilon}$ and $G = \mathrm{Cay}(\mathcal{D}^t,\mathcal{T}_{t,\varepsilon})$. The conclusion follows.

Appendix B. Hierarchy integrality gaps for Unique Games and related problems. This section is devoted to the construction of an integrality gap instance for a hierarchy of SDP relaxations to Unique Games. More specifically, we consider the LH_r and SA_r SDP hierarchies described in [RS09]. For these SDP hierarchies, we will demonstrate the following integrality gap constructions.

Theorem B.1. *For every $\varepsilon, \delta > 0$, there exists an $\mathbb{F}_2^t$-Max-2Lin instance I for some positive integer t such that no labelling satisfies more than a δ fraction of edges of Γ, while there exists an SDP solution such that*

– *the SDP solution is feasible for LH_R with $R = \exp(\exp(\Omega(\log\log^{1/2} N)))$;*

– *the SDP solution is feasible for $SA\text{-}SDP_R$ with $R = \exp(\Omega(\log\log^{1/2} N))$;*

– *the SDP solution has value $1 - O(\varepsilon)$,*

where N is the number of vertices in the instance I.

Remark B.2. Composing the above SDP integrality gap with Unique Games–based hardness reductions yields corresponding gap instances for several classes of problems such as constraint satisfaction problems (CSPs) and ordering CSPs such as maximum acyclic subgraph. Specifically, up to $\exp(\exp(\Omega(\log\log^{1/2} N)))$ rounds of the LH hierarchy or the $\exp(\Omega(\log\log^{1/2} N))$ rounds of the SA-SDP hierarchy can be shown to have the same SDP integrality gap as the simple SDP relaxation for every CSP. For the sake of brevity, we omit a formal statement of this result here.

Towards showing Theorem B.1, we follow the approach outlined in [RS09]. At a high level, the idea is to start with an integrality gap instance Γ for a simple SDP relaxation for Unique Games over a large alphabet. The instance Γ is reduced to an instance $\Psi_{\varepsilon,Q,d}(\Gamma)$ of Unique Games over a smaller alphabet using a reduction similar to that of Khot et al. [KKMO07]. Moreover, the SDP solution to the simple SDP relaxation of Γ can be translated to a solution for several rounds of the SDP hierarchy for $\Psi_{\varepsilon,Q,d}(\Gamma)$.

Let Γ be an instance of $\mathbb{F}_2^n$-Max-2Lin over a set of vertices $V(\Gamma)$ and edges $E(\Gamma)$. On every edge $(u, v) \in E(\Gamma)$, there is a constraint of the form $u - v = \alpha_{uv}$ for some $\alpha \in \mathbb{F}_2^n$. We will reduce Γ to an instance of Q-Max-2Lin using the 2-query test described in Appendix A.

Translations. Notice that the Reed–Muller code is invariant under translation of its coordinates. Therefore, the code $\mathcal{D}^t$ and the test distributions $\mathcal{T}_{t,\varepsilon}$ are both invariant under translation. Formally, for an $\alpha \in \mathbb{F}_2^n$, the translation operator $T_\alpha \colon Q^N \to Q^N$ is defined by

$$(T_\alpha \circ c)_\beta = c_{\beta+\alpha} \qquad \forall c \in Q^N,\ \beta \in \mathbb{F}_2^n\,.$$

Given a codeword $c \in \mathcal{D}^t$, we have $T_\alpha \circ c \in \mathcal{D}^t$.

The vertices of $\Psi_{\varepsilon,Q,d}(\Gamma)$ are $V(\Gamma) \times \mathcal{D}^t$. Let $\ell : V(\Gamma) \times \mathcal{D}^t \to Q$ be a labelling of the instance $\Psi_{\varepsilon,Q,d}(\Gamma)$.

Folding. The labelling ℓ is assumed to satisfy $\ell(v, c + r) = \ell(v, c) + r$ for every vertex $v \in V(\Gamma)$, $c \in \mathcal{D}^t$, and $r \in Q$. This is enforced by “folding.”

The constraints of $\Psi_{\varepsilon,Q,d}(\Gamma)$ are given by the queries of the following verifier:

– Sample a vertex $u \in V(\Gamma)$ uniformly at random. Sample two neighbors $v_1, v_2 \in N(u)$ of u uniformly at random. Let the constraint on the edge (u, v_i) be $v_i - u = \alpha_i$ for $i \in \{1, 2\}$.

– Sample an element $c_1 \in \mathcal{D}^t$ uniformly at random, and sample a neighbor $c_2 \in \mathcal{D}^t$ of c_1 in the graph $\mathrm{Cay}(\mathcal{D}^t, \mathcal{T}_{t,\varepsilon})$.

– Sample an element $r \in Q$ uniformly at random.

– Test if $\ell(v_1, (T_{\alpha_1} \circ c_1) + r) - r = \ell(v_2, T_{\alpha_2} \circ c_2)$.

We are now ready to describe the reduction from Γ to an instance of $\mathbb{F}_2^n$-Max-2Lin.

Soundness.

Lemma B.3. *For all sufficiently small constants $\varepsilon, \delta > 0$ and all choices of $Q = 2^t$, there exist γ, d such that if no labelling of Γ satisfies more than a γ fraction of edges, then every labelling of $\Psi_{\varepsilon,Q,d}(\Gamma)$ satisfies at most a $Q\Gamma_\rho\left({}^1\!/_Q\right) + \delta$ fraction of constraints, where $\rho = e^{-\varepsilon}$.*

Proof. Let $\ell : V \times \mathcal{D}^t \to Q$ be a labelling of the instance $\Psi_{\varepsilon,Q,d}(\Gamma)$. For each vertex $v \in V(\Gamma)$, let $F^v : \mathcal{D}^t \to Q$ denote the labelling ℓ restricted to the vertex v, i.e., $F^v(c) \overset{\text{def}}{=} \ell(v,c)$. For each vertex $v \in V(\Gamma)$ and $q \in Q$, define $f_q^v : \mathcal{D}^t \to [0,1]$ as

$$f_q^v(c) \overset{\text{def}}{=} \mathbb{I}[F^v(c) = q]\,.$$

Due to folding we have $f_q^v(c) = f_{q+r}^v(c+r)$ for all $r \in Q$. Moreover, this implies that $\mathbb{E}_{c\in\mathcal{D}^t} f_q^v = \frac{1}{Q}$. Finally, for a vertex $u \in V(\Gamma)$ and $r \in Q$, define

$$h_r^u(p) \overset{\text{def}}{=} \mathop{\mathbb{E}}_{v\in N(u)} f_r^v(T_{\alpha_{uv}} \circ p)\,.$$

Clearly, for the functions h_r^u also we have

$$\mathop{\mathbb{E}}_{p} h_r^u = \frac{1}{Q} \qquad \forall u \in V(\Gamma),\ r \in Q. \tag{B.1}$$

The probability of acceptance of the verifier can be arithmetized in terms of the functions h_r^u:

$$\begin{aligned}
&\mathbb{P}[\text{verifier accepts}]\\
&= \mathop{\mathbb{E}}_{u\in V(\Gamma)} \mathop{\mathbb{E}}_{v_1,v_2\in N(u)} \mathop{\mathbb{E}}_{c_1,c_2\in \mathrm{Cay}(\mathcal{D}^t,\mathcal{T}_{t,\varepsilon})} \mathop{\mathbb{E}}_{r\in Q} \left[\sum_{q\in Q} f_{q+r}^{v_1}(T_{\alpha_1}\circ c_1 + r) f_q^{v_2}(T_{\alpha_2}\circ c_2)\right]\\
&= \mathop{\mathbb{E}}_{u\in V(\Gamma)} \mathop{\mathbb{E}}_{v_1,v_2\in N(u)} \mathop{\mathbb{E}}_{c_1,c_2\in \mathrm{Cay}(\mathcal{D}^t,\mathcal{T}_{t,\varepsilon})} \left[\sum_{q\in Q} f_{q}^{v_1}(T_{\alpha_1}\circ c_1) f_q^{v_2}(T_{\alpha_2}\circ c_2)\right] \qquad \text{(folding)}\\
&= \mathop{\mathbb{E}}_{u\in V(\Gamma)} \mathop{\mathbb{E}}_{c_1,c_2\in \mathrm{Cay}(\mathcal{D}^t,\mathcal{T}_{t,\varepsilon})} \left[\sum_{q\in Q} \mathop{\mathbb{E}}_{v_1\in N(u)} f_{q}^{v_1}(T_{\alpha_1}\circ c_1) \cdot \mathop{\mathbb{E}}_{v_2\in N(u)} f_q^{v_2}(T_{\alpha_2}\circ c_2)\right]\\
&= \mathop{\mathbb{E}}_{u\in V(\Gamma)} \mathop{\mathbb{E}}_{c_1,c_2\in \mathrm{Cay}(\mathcal{D}^t,\mathcal{T}_{t,\varepsilon})} \left[\sum_{q\in Q} h_q^u(c_1) h_q^u(c_2)\right]\\
&= \mathop{\mathbb{E}}_{u\in V(\Gamma)} \left[\sum_{q\in Q} \langle h_q^u, H h_q^u\rangle\right] \qquad (\text{where } H = \mathrm{Cay}(\mathcal{D}^t,\mathcal{T}_{t,\varepsilon})).
\end{aligned}$$

Suppose the probability of acceptance of the verifier is at least $Q \cdot \Gamma_\rho(1/Q) + \delta$. By simple averaging, for at least a $\delta/2$ fraction of the vertices $u \in V(\Gamma)$, we have

$$\sum_{q\in Q} \langle h_q^u, H h_q^u\rangle \geqslant Q\Gamma_\rho(1/Q) + \frac{\delta}{2}\,.$$

Let us refer to such a vertex u as being *good*.

Fix the parameters τ, L, d to those obtained by applying Theorem A.4 with parameters $\varepsilon, \delta/2Q$. Recall that by (B.1), we have $\mathbb{E}_{\mathcal{D}^t}[h_q^u] = \frac{1}{Q}$. Applying Theorem A.4, if for each $q \in Q$, $\max_{\alpha\in\mathbb{F}_2^n} \mathrm{Inf}_\alpha^{\leqslant l}(h_q^u) \leqslant \tau$, then

$$\sum_{q\in Q} \langle h_q^u, G h_q^u\rangle \leqslant Q\Gamma_\rho(1/Q) + Q \cdot \frac{\delta}{2Q}.$$

This implies that for each *good* vertex u there exist q, α such that $\mathrm{Inf}_\alpha^{\leqslant L}(h_q^u) \geqslant \tau$. We will use these influential coordinates to decode a labelling for the $\mathbb{F}_2^n$-Max-2Lin instance Γ.

For each vertex $v \in V(\Gamma)$ define the set of influential coordinates S_v as

$$S_v = \{\alpha \in \mathbb{F}_2^n \,|\, \mathrm{Inf}_\alpha^{\leqslant L}(h_q^v) \geqslant \tau/2 \text{ for some } q \in Q\} \cup \{\alpha \in \mathbb{F}_2^n \,|\, \mathrm{Inf}_\alpha^{\leqslant L}(f_q^v) \geqslant \tau/2 \text{ for some } q \in Q\}. \tag{B.2}$$

Using Lemma C.1, for each of the functions h_q^v or f_q^v, there are at most $2L/\tau$ coordinates with influence greater than $\tau/2$. Therefore, for each vertex v the set S_v is of size at most $2 \cdot Q \cdot 2L/\tau = 4QL/\tau$.

Define an assignment of labels $A : V(\Gamma) \to \mathbb{F}_2^n$ as follows. For each vertex v, sample a random $\alpha \in S_v$ and assign $A(v) = \alpha$.

Fix one good vertex u and a corresponding q, α such that $\mathrm{Inf}_\alpha^{\leqslant L}(h_q^u) \geqslant \tau$. By definition of h_q^u this implies that

$$\mathrm{Inf}_\alpha^{\leqslant L}\left(\mathop{\mathbb{E}}_{v \in N(u)} T_{\alpha_{uv}} \circ f_q^v\right) \geqslant \tau\,,$$

which by convexity of influences yields

$$\tau \leqslant \mathop{\mathbb{E}}_{v \in N(u)}[\mathrm{Inf}_\alpha^{\leqslant L}(T_{\alpha_{uv}} \circ f_q^v)] = \mathop{\mathbb{E}}_{v \in N(u)}[\mathrm{Inf}_{\alpha - \alpha_{uv}}^{\leqslant L}(f_q^v)]\,.$$

Hence, for at least a $\tau/2$ fraction of the neighbors $v \in N(u)$, the coordinate $\alpha - \alpha_{uv}$ has influence at least $\tau/2$ on f_q^v. Therefore, for every good vertex u, for at least a $\tau/2$ fraction of its neighbors $v \in N(u)$, the edge (u, v) is satisfied by the labelling A with probability at least $\frac{1}{|S_u|}\frac{1}{|S_v|} \geqslant \tau^2/16Q^2L^2$. Since there is at least a $\delta/2$ fraction of good vertices u, the expected fraction of edges satisfied by the labelling A is at least $\frac{\delta}{2} \cdot \frac{\tau}{2} \cdot \frac{\tau^2}{16Q^2L^2} = \frac{\delta\tau^3}{64Q^2L^2}$.

Choosing the soundness γ of the outer unique game Γ to be lower than $\frac{\delta\tau^3}{64Q^2L^2}$ yields a contradiction. This shows that the value of any labelling ℓ to $\Psi_{\varepsilon,Q,d}(\Gamma)$ is less than $Q\Gamma_\rho(1/Q) + \delta$. ∎

SDP solution. We will construct feasible solutions to certain strong SDP relaxations of $\Psi_{\varepsilon,Q,d}(\Gamma)$ by appealing to the work of [RS09]. The SDP hierarchies that we consider are referred to as the LH and SA-SDP hierarchies. Informally, the rth-level LH relaxation (LH$_r$) consists of the simple SDP relaxation for Unique Games augmented by local distributions μ_S over integral assignments for every set S of at most r vertices. The local distribution μ_S is required to be consistent with the inner products of the SDP vectors. Alternately, this SDP hierarchy can be thought of as the simple SDP relaxation augmented by every valid constraint on at most r vertices.

The SA-SDP hierarchy is a somewhat stronger hierarchy that requires the local distributions μ_S to be consistent with each other, namely, μ_S and μ_T agree on $S \cap T$. Alternately, the SA-SDP hierarchy corresponds to the simple SDP relaxation augmented with r rounds of SA linear program variables. We refer the reader to [RS09] for formal definitions of the SA-SDP and LH hierarchies.

Lemma B.4. *Suppose Γ has an SDP solution that is of value $1 - \eta$; then there exists an SDP solution to the instance $\Psi_{\varepsilon,Q,d}(\Gamma)$ such that*

- *the SDP solution is feasible for LH$_R$ with $R = 2^{\Omega(\varepsilon/\eta^{1/4})}$;*
- *the SDP solution is feasible for SA-SDP$_R$ with $R = \Omega(\varepsilon/\eta^{1/4})$;*
- *the SDP solution has value $1 - O(\varepsilon) - o_\eta(1)$ on $\Psi_{\varepsilon,Q,d}(\Gamma)$.*

Proof. This lemma is a direct consequence of Theorem 9 from [RS09].

In [RS09], the authors start with an integrality gap instance Γ for the simple SDP for Unique Games, and then perform a traditional long code–based reduction to obtain an instance $\Phi_{\varepsilon,Q}(\Gamma)$.

The crucial observation is the following.

Observation B.5. *The vertices of $\Psi_{\varepsilon,Q,d}(\Gamma)$ are a subset of vertices of $\Phi_{\varepsilon,Q}(\Gamma)$—the instance obtained by the traditional Q-ary long code reduction on Γ.*

Proof. The vertices of $\Psi_{\varepsilon,Q,d}(\Gamma)$ are pairs of the form (v, c), where $v \in V(\Gamma)$ and $c \in \mathcal{D}^t$. The codeword $c \in \mathcal{D}^t$ can be thought of as a string of length $N = 2^n$ over the alphabet $Q = \mathbb{F}_2^t$, namely, $c \in Q^{2^n}$. The vertices of the instance $\Phi_{\varepsilon,Q}(\Gamma)$ obtained via a traditional Q-ary long code reduction are $V(\Gamma) \times Q^{2^n}$. Hence the observation follows. □

In [RS09], the authors construct an SDP solution for the instance $\Phi_{\varepsilon,Q}(\Gamma)$ that is feasible for LH_R relaxation with $R = 2^{\Omega(\varepsilon/\eta^{1/4})}$ and for SA-SDP$_R$ relaxation with $R = \Omega(\varepsilon/\eta^{1/4})$. As noted in Observation B.5, the vertices of $\Psi_{\varepsilon,Q}(\Gamma)$ are a subset of the vertices of $\Phi_{\varepsilon,Q}(\Gamma)$. Therefore, the same SDP solution constructed in [RS09] when restricted to the instance $\Psi_{\varepsilon,Q,d}(\Gamma)$ yields a feasible solution for the corresponding LH_R and SA-SDP$_R$ relaxations.

To finish the proof, we need to show that the value of the SDP solution from [RS09] is $1 - 2\varepsilon - o_\eta(1)$.

The traditional long code–based reduction to get $\Phi_{\varepsilon,Q}(\Gamma)$ uses the noise stability test as the inner gadget. Namely, to test whether a function $f : \mathbb{F}_Q^{2^n} \to \mathbb{F}_Q$ is a dictator function, the verifier picks $x \in \mathbb{F}_Q^{2^n}$ uniformly at random, rerandomizes each coordinate of x independently with probability ε, and then tests whether $f(x) = f(y)$. Composing this noise stability test with the outer unique game Γ yields the instance $\Phi_{\varepsilon,Q}(\Gamma)$. The value of the SDP solution constructed for $\Phi_{\varepsilon,Q}(\Gamma)$ in [RS09] depends only on the expected Hamming distance between the queries x, y. More precisely, in Claim 2 of [RS09], the authors show that if the distribution on the queries $(x, y) \in \mathbb{F}_Q^{2^n} \times \mathbb{F}_Q^{2^n}$ is chosen to be an arbitrary distribution NS over $\mathbb{F}_Q^{2^n} \times \mathbb{F}_Q^{2^n}$, the SDP objective value of the solution is given by

$$\mathop{\mathbb{P}}_{\{x,y\}\sim \mathsf{NS}, \ell \in [2^n]}[x_\ell = y_\ell] - \varepsilon\,.$$

The instance $\Psi_{\varepsilon,Q,d}$ is obtained by using the following distribution of x, y over $\mathbb{F}_Q^{2^n} \times \mathbb{F}_Q^{2^n}$: Sample (c_1, c_2), an edge in $\mathrm{Cay}(\mathcal{D}^t, \mathcal{T}_{t,\varepsilon})$.

By construction, for any coordinate $\ell \in [2^n]$, $\mathbb{P}[x_\ell = y_\ell = 1 - O(\varepsilon)$. Therefore, using Claim 2 of [RS09], the SDP objective value on the instance $\Psi_{\varepsilon,Q,d}(\Gamma)$ is at least $1 - O(2\varepsilon) - o_\eta(1)$. □

Proof of Theorem B.1. Fix $t = \lceil 10/\varepsilon \log(1/\delta)\rceil$ and $Q = 2^t$. By our choice of Q, we have $Q\Gamma_{e^{-\varepsilon}}(1/Q) \leqslant \delta$ (see Appendix B in [MOO05] for such asymptotic bounds on Γ).

Fix γ, d depending on ε, δ, and Q as dictated by Lemma B.3. Let Γ be the Unique Games instance obtained by Corollary 6.2 with the optimal integral value set to γ. In particular, Γ is an $\mathbb{F}_2^n$-Max-2Lin instance that has $M = 2^{2^{\log^2 n}}$ vertices. Its SDP optimum for the simple Unique Games SDP relaxation is at least $1 - O(C(\varepsilon,\delta)/n)$ $(\eta = O(C(\varepsilon,\delta)/n))$ for some constant $C(\varepsilon,\delta)$ depending on ε, δ.

Now we apply the reduction to $\mathbb{F}_2^t$-Max-2Lin outlined below to obtain an instance $\Psi_{\varepsilon,Q,d}(\Gamma)$. The number of vertices of the instance $\Psi_{\varepsilon,Q,d}(\Gamma)$ is $|V(\Gamma)| \times |\mathcal{D}^t|$. Note that the choice of the degree d is a constant (say $d(\varepsilon,\delta)$) depending on ε, δ. Hence, the number of points in $\mathcal{D}^t$ is given by $|\mathcal{D}^t| = 2^{O(n^{d(\varepsilon,\delta)})}$. Therefore, the number of vertices of $\Psi_{\varepsilon,Q,d}(\Gamma)$ is $N = 2^{2^{\log^2 n}} \cdot 2^{O(n^{d(\varepsilon,\delta)})} = 2^{2^{O(\log^2 n)}}$. Equivalently, we have $n = 2^{\Omega(\log\log^{1/2} N)}$.

- By Lemma B.3, the optimal labelling to $\Psi_{\varepsilon,Q,d}(\Gamma)$ satisfies at most a $Q\Gamma_{e^{-\varepsilon}}$ $(1/Q) + \delta = O(\delta)$ fraction of constraints.
- By Lemma B.4, there exists an SDP solution to the instance $\Psi_{\varepsilon,Q,d}(\Gamma)$ with value $1 - O(\varepsilon) - o_\eta(1)$. Since $\eta = O(C(\varepsilon,\delta)/n)$, for large enough choice of n, the SDP value is at least $1 - O(\varepsilon)$.
- The SDP solution is feasible for LH_R for $R = 2^{\Omega(\varepsilon/\eta^{1/4})} = 2^{c(\varepsilon,\delta)n^{1/4}} = \exp(\exp(\Omega(\log\log^{1/2} N)))$ rounds, where $c(\varepsilon,\delta)$ is a constant depending on ε and δ. Furthermore, the SDP solution is also feasible for SA-SDP$_R$ for $R = \Omega(\varepsilon/\eta^{1/4}) = c(\varepsilon,\delta)n^{1/4} = \exp(\Omega(\log\log^{1/2} N))$. □

Appendix C. Missing proofs.

Proof of Theorem 5.6. Suppose $d \geqslant C\log(1/\tau)$ for C to be chosen later, and fix $\gamma < 1/8$ for γ to be chosen later. Let $\ell = \log(1/\tau)/4c_1 < \tau^2 2^{d+1}$, where c_1 is the constant from Theorem 5.7. For $\alpha \in \mathbb{F}_2^N/\mathcal{C}$, let λ_α be the eigenvalues of G. Then, by Lemma 4.13,

$$|\lambda_\alpha - \rho^k| < \tau \text{ for } k \leqslant \ell \quad |\lambda_\alpha| < \rho^{\ell/2} \text{ for } k > \ell. \tag{C.1}$$

Let $g = G^\gamma f$ and $G' = G^{1-2\gamma}$. Then, the graph G' has the same eigenfunctions as G—χ_α for $\alpha \in \mathbb{F}_2^N/\mathcal{C}$ with eigenvalues $\lambda'_\alpha = \lambda_\alpha^{1-2\gamma}$. From the above equation, it is easy to check that, for $\rho' = \rho^{1-2\gamma}$,

$$|\lambda'_\alpha - (\rho')^k| < \sqrt{\tau} \text{ for } k \leqslant \ell, \quad |\lambda'_\alpha| < (\rho')^{\ell/2} \text{ for } k > \ell. \tag{C.2}$$

Further, as the eigenvalues of G are each at most 1, the coordinate influences of g are no larger than those of f.

Now, decompose $g = g^{\leqslant\ell} + g^{>\ell}$ into a low-degree part $g^{\leqslant\ell} = \sum_{\alpha\in\mathbb{F}_2^n,\ \mathsf{wt}(\alpha)\leqslant\ell} \hat{g}(\alpha)\chi_\alpha$ and a high-degree part $g^{>\ell} = \sum_{\alpha\in\mathbb{F}_2^n/\mathcal{C},\ \Delta(\alpha,\mathcal{C})>\ell} \hat{g}(\alpha)\chi_\alpha$. Then,

$$\langle f, Gf\rangle = \langle g, G'g\rangle = \langle g^{\leqslant\ell}, G'g^{\leqslant\ell}\rangle + \langle g^{>\ell}, G'g^{>\ell}\rangle \leqslant \langle g^{\leqslant\ell}, G'g^{\leqslant\ell}\rangle + \mu \cdot \max_{\alpha\in\mathbb{F}_2^N/\mathcal{C},\ \Delta(\alpha,\mathcal{C})>\ell} \lambda'_\alpha\,.$$

Hence, using (C.2) (and the crude bound $\mu \leqslant 1$),

$$\langle f, Gf\rangle = \sum_{\alpha\in\mathbb{F}_2^N,\ \mathsf{wt}(\alpha)\leqslant\ell} (\rho')^{\mathsf{wt}(\alpha)} \hat{g}(\alpha)^2 + (\rho')^\ell + \sqrt{\tau}\,. \tag{C.3}$$

Observe that $g^{\leqslant\ell}$ is a multilinear polynomial of degree at most ℓ and, as the ℓ-degree influences of g are at most τ, $g^{\leqslant\ell}$ is τ-regular.

Let $S \subseteq \{\pm 1\}^N$ be the set of $\{\pm 1\}$-vectors corresponding to the Reed–Muller code $\mathcal{C}^\perp = \mathsf{RM}(n,d)$; that is, for every codeword $c \in \mathcal{C}^\perp$, the set S contains the vector $((-1)^{c_1}, \ldots, (-1)^{c_N})$. Then, as g is $[0,1]$-valued on $\mathcal{C}^\perp$ and ζ measures distance to bounded random variables, by (C.1),

$$\begin{aligned}\mathop{\mathbb{E}}_{z\sim S}[\zeta \circ g^{\leqslant\ell}(z)] \leqslant \mathop{\mathbb{E}}_{z\sim S}[(g(z) - g^{\leqslant\ell}(z))^2] = \mathop{\mathbb{E}}_{z\sim S}[(g^{>\ell}(z))^2] = \mathop{\mathbb{E}}_{z\sim S}[(G^\gamma f^{>\ell}(z))^2]\\ \leqslant \max_{\alpha:|\alpha|>\ell} (\lambda_\alpha^\gamma)^2 \leqslant \rho^{\gamma\ell}.\end{aligned}$$

Hence, by Theorem 5.7 (recall that $\ell = \log(1/\tau)/4c_1$),

$$\mathop{\mathbb{E}}_{x\sim\{\pm1\}^N}[\zeta \circ g^{\leqslant \ell}(x)] \leqslant \mathop{\mathbb{E}}_{z\sim S}[\zeta \circ g^{\leqslant \ell}(z)] + 2^{O(\ell)}\sqrt{\tau} \leqslant \underbrace{\rho^{\gamma\ell} + \tau^{1/4}}_{\eta :=} .$$

Now, as $\mathcal{C}^{\perp}$ is ℓ-wise independent ($\ell < 2^{d+1}$),

$$\mathop{\mathbb{E}}_{x\sim\{\pm1\}^N}[g^{\leqslant \ell}(x)] = \mathop{\mathbb{E}}_{z\sim S}[g^{\leqslant \ell}(z)] = \mathop{\mathbb{E}}_{z\sim S}[g(z)] \pm \mathop{\mathbb{E}}_{z\sim S}[(g^{>\ell}(z))^2]^{1/2} \leqslant \mu + \sqrt{\eta}.$$

Therefore, by Corollary 5.3,

$$\langle g^{\leqslant \ell}, T_{\rho'} g^{\leqslant \ell}\rangle = \sum_{\alpha:\mathsf{wt}(\alpha)\leqslant \ell} (\rho')^{\mathsf{wt}(\alpha)} \hat{g}(\alpha)^2 \leqslant \Gamma_{\rho'}(\mu + \sqrt{\eta}) + \frac{O(\log\log(1/\eta))}{(1-\rho')\log(1/\eta)}. \tag{C.4}$$

Since $\Gamma_{\rho'}(\mu+\sqrt{\eta}) \leqslant \Gamma_{\rho'}(\mu) + 2\sqrt{\eta}$ and $\Gamma_{\rho}(\mu) \leqslant \Gamma_{\rho'}(\mu) + |\rho-\rho'|/(1-\rho)$ (cf. Lemma B.3, Corollary B.5 in [MOO05]), it follows from (C.3) and (C.4) that

$$\begin{aligned}\langle f, Gf\rangle = \langle g, G'g\rangle &\leqslant \Gamma_\rho(\mu) + O\left(\frac{|\rho-\rho'|}{1-\rho}\right) + O(\sqrt{\eta}) + \frac{O(\log\log(1/\eta))}{(1-\rho)\log(1/\eta)} \\ &\quad + \rho^{(1-2\gamma)\ell} + \sqrt{\tau} \\ &= \Gamma_\rho(\mu) + O\left(\frac{\gamma\log(1/\rho)}{1-\rho} + \rho^{\gamma\ell/2} + \tau^{1/8} + \frac{\log\log(1/\eta)}{(1-\rho)\log(1/\eta)}\right).\end{aligned}$$

(Here we used the estimate $|\rho-\rho'| = |\rho-\rho^{1-2\gamma}| = O(\gamma\log(1/\rho))$.) By choosing $d \geqslant C\log(1/\tau)$ and $\gamma = CK\log\log(1/\tau)/(\log(1/\tau)\log(1/\rho))$ for an appropriately large constant C, the above expression simplifies to

$$\langle f, Gf\rangle \leqslant \Gamma_\rho(\mu) + \frac{O(\log\log(1/\tau))}{(1-\rho)\log(1/\tau)}. \qquad \square$$

Proof of Lemma 5.15. Since X fools τ-regular degree-ℓ PTFs, we have for all $u \geqslant 0$,

$$\Big|\mathbb{P}\{\zeta\circ Q(X) > u\} - \mathbb{P}\{\zeta\circ Q(Y) > u\}\Big| \leqslant O(\varepsilon).$$

By hypercontractivity and 20ℓ-wise independence of X,

$$\mathbb{P}\{\zeta\circ Q(X) > u\} \leqslant \mathbb{P}\left\{|Q(X)| > \sqrt{u}\right\} \leqslant u^{-10}\,\mathbb{E}\,Q(X)^{20} \leqslant u^{-10}2^{O(\ell)}.$$

Since $\zeta\circ Q(X)$ is a nonnegative random variable,

$$\mathbb{E}\,\zeta\circ Q(X) = \int \mathbb{P}\{\zeta\circ Q(X) > u\}\,\mathrm{d}u.$$

Hence, we can bound its expectation:

$$\begin{aligned}\mathbb{E}\,\zeta\circ Q(X) &= \int_{u\geqslant 0} \mathbb{P}\{\zeta\circ Q(X) > u\}\,\mathrm{d}u \\ &= \int_{0\leqslant u\leqslant M} \mathbb{P}\{\zeta\circ Q(X) > u\}\,\mathrm{d}u \pm 2^{O(\ell)}\int_{u\geqslant M} u^{-10}\,\mathrm{d}u \\ &= \int_{0\leqslant u\leqslant M} \mathbb{P}\{\zeta\circ Q(Y) > u\}\,\mathrm{d}u \pm O\left(\varepsilon M + 2^{O(\ell)}/M^9\right) \\ &= \mathbb{E}\,\zeta\circ Q(Y) \pm O\left(\varepsilon M + \ell^{O(\ell d)}/M^9\right).\end{aligned}$$

(In the last step, we used that $\mathbb{P}\{\zeta\circ Q(Y) > u\} \leqslant u^{-10}2^{O(\ell)}$, a consequence of hypercontractivity.) Choosing $M = 2^{O(\ell)}/\varepsilon^{0.1}$ (so that $\varepsilon M = 2^{O(\ell)}/M^9$), we conclude that $\mathbb{E}\,\zeta\circ Q = \mathbb{E}\,\zeta\circ Q \pm \varepsilon^{0.9}2^{O(\ell)}$. $\square$

C.1. Proofs from section A.1. The following lemma shows a bound on the sum of influences.

LEMMA C.1. *For a function $f\colon \mathcal{D}^t \to \mathbb{R}$ and $\ell < \operatorname{dist}(\mathcal{C}^t)/2$, the sum of ℓ-degree influences of f is at most $\sum_{i\in[N]} \operatorname{Inf}_i^{\leqslant \ell}(f) \leqslant \ell\, \mathbb{V}[f]$.*

Proof. The usual identity for the total (low-degree) influence holds:

$$\sum_{i\in[N]} \operatorname{Inf}_i^{\leqslant\ell}(f) = \sum_{\beta\in Q^N/\mathcal{C}^t,\ \mathsf{wt}(\beta)\leqslant\ell} \mathsf{wt}(\beta)\hat{f}(\beta)^2 \leqslant \ell\, \mathbb{V} f\,. \qquad \square$$

Analogously to Theorem 5.7, the following invariance principle can be shown for regular multilinear polynomials.

THEOREM C.2. *Let $N = 2^n$ and t be an integer. For every $\tau, \ell > 0$, there exists C such that for $d > C\log(1/\tau)$ the following holds: If $P : \mathbb{R}^{Nt} \to \mathbb{R}$ is a τ-regular polynomial of degree at most ℓ, then, for $x \in_u \{\pm 1\}^{Nt}$, $z \in_u \mathsf{RM}(n,d)^t$,*

$$|\mathbb{E}[\zeta \circ P(x)] - \mathbb{E}[\zeta\circ P(z)]| \leqslant 2^{c_1 \ell}\sqrt{\tau}$$

for a universal constant $c_1 > 0$.

The proof follows easily from the proofs of Theorems 5.9 and 5.7 and the fact that if $\mathsf{RM}(n,d)$ satisfies the properties of the PRG in [MZ13], then so does $\mathsf{RM}(n,d)^t$. We omit the proof.

The work of Mossel, O'Donnell, and Oleszkiewicz [MOO05] also obtains bounds on noise stability of functions over product spaces of large alphabets, namely Q^N. The following corollary is a consequence of Theorem 4.4 in [MOO05]. The proof is analgous to that of Corollary 5.3, the corollary to Theorem 5.1.

COROLLARY C.3. *Let $f\colon Q^N \to \mathbb{R}$ be a function with $\mathbb{E} f = \mu$ and $\mathbb{E}\,\zeta\circ f \leqslant \tau$. Suppose $\operatorname{Inf}_i f^{\leqslant 30\log(1/\tau)/\log Q} \leqslant \tau$ for all $i \in [N]$. Then,*

$$\langle f, T_\rho f\rangle \leqslant \Gamma_\rho(\mu) + O\left(\tfrac{\log Q \log\log(1/\tau)}{(1-\rho)\log(1/\tau)}\right),$$

where T_ρ is the noise graph on Q^N with second largest eigenvalue ρ and Γ_ρ is the Gaussian noise stability curve. (Here, we assume that τ is small enough.)

Now we are ready to present the proof of the Majority Is Stablest theorem over $\mathcal{D}^t$ (Theorem A.4) using Theorem C.2 and Corollary C.3.

Proof of Theorem A.4. Let $Q = 2^t$. Fix $d \geqslant C\log(1/\tau)$ for a sufficiently large constant C to be chosen later. Let $\gamma < 1/8$ be a constant depending on ε, δ whose value will be chosen later. Let $\ell = \log(1/\tau)/4c_1 < \tau^2 2^{d+1}$, where c_1 is the constant from Theorem C.2. For $\alpha \in Q^N/\mathcal{C}$, let λ_α be the eigenvalues of G. Then, by Lemma A.2,

$$\text{(C.5)} \qquad |\lambda_\alpha - \rho^{\mathsf{wt}(\alpha)}| < \tau \text{ for } \mathsf{wt}(\alpha)\leqslant \ell, \qquad |\lambda_\alpha| < \rho^{\Omega(\ell/t)} \text{ for } \mathsf{wt}(\alpha) > \ell.$$

Let $g = G^\gamma f$ and $G' = G^{1-2\gamma}$. Then, the graph G' has the same eigenfunctions as G—χ_α for $\alpha \in Q^N/\mathcal{C}$ with eigenvalues $\lambda'_\alpha = \lambda_\alpha^{1-2\gamma}$. From the above equation, it is easy to check that, for $\rho' = \rho^{1-2\gamma}$,

$$\text{(C.6)} \qquad |\lambda'_\alpha - (\rho')^{\mathsf{wt}(\alpha)}| < \sqrt{\tau} \text{ for } \mathsf{wt}(\alpha)\leqslant\ell, \qquad |\lambda'_\alpha| < (\rho')^{\Omega(\ell/t)} \text{ for } \mathsf{wt}(\alpha) > \ell.$$

Further, as the eigenvalues of G are each at most 1, the coordinate influences of g are no larger than those of f. Now, decompose $g = g^{\leqslant\ell} + g^{>\ell}$ into a low-degree part $g^{\leqslant\ell} = \sum_{\alpha\in Q^N,\ \mathsf{wt}(\alpha)\leqslant\ell} \hat{g}(\alpha)\chi_\alpha$ and a high-degree part $g^{>\ell} = \sum_{\alpha\in Q^N/\mathcal{C},\ \mathsf{wt}(\alpha)>\ell}\hat{g}(\alpha)\chi_\alpha$. Then,

$$\langle f, Gf\rangle = \langle g, G'g\rangle = \langle g^{\leqslant\ell}, G'g^{\leqslant\ell}\rangle + \langle g^{>\ell}, G'g^{>\ell}\rangle \leqslant \langle g^{\leqslant\ell}, G'g^{\leqslant\ell}\rangle + \mu\cdot \max_{\alpha\in Q^N/\mathcal{C},\ \deg(\alpha)>\ell} \lambda'_\alpha\,.$$

Hence, using (C.6) (and the crude bound $\mu \leqslant 1$),

$$\langle f, Gf \rangle = \sum_{\alpha \in Q^N/\mathcal{C},\, \mathsf{wt}(\alpha) \leqslant \ell} (\rho')^{\mathsf{wt}(\alpha)} \hat{g}(\alpha)^2 + (\rho')^{\Omega(\ell/t)} + \sqrt{\tau}\,. \tag{C.7}$$

Observe that $g^{\leqslant \ell}$ is a multilinear polynomial of degree at most $\ell \cdot t$. Since the ℓ-degree influences of g are at most τ, it implies that the multilinear polynomial $g^{\leqslant \ell}$ is τ-regular.

Let $S \subseteq \{\pm 1\}^{Nt}$ be the set of $\{\pm 1\}$-vectors corresponding to the Reed–Muller code $\mathcal{D}^t$; that is, for every codeword $c = (c^{(1)}, c^{(2)}, \ldots, c^{(t)}) \in \mathcal{D}^t$, the set S contains the vector $((-1)^{c_1^{(1)}}, \ldots (-1)^{c_j^{(i)}}, (-1)^{c_N^{(t)}})$. Then, as g is $[0,1]$-valued on $\mathcal{D}^t$ and ζ measures distance to bounded random variables, by (C.5),

$$\begin{aligned} \mathop{\mathbb{E}}_{z \sim S}[\zeta \circ g^{\leqslant \ell}(z)] \leqslant \mathop{\mathbb{E}}_{z \sim S}[(g(z) - g^{\leqslant \ell}(z))^2] = \mathop{\mathbb{E}}_{z \sim S}[(g^{>\ell}(z))^2] &= \mathop{\mathbb{E}}_{z \sim S}[(G^\gamma f^{>\ell}(z))^2] \\ &\leqslant \max_{\alpha : \mathsf{wt}(\alpha) > \ell} (\lambda_\alpha^\gamma)^2 \leqslant \rho^{\Omega(\gamma \ell / t)}. \end{aligned}$$

Hence, by Theorem C.2 (recall that $\ell = \log(1/\tau)/4c_1$),

$$\mathop{\mathbb{E}}_{x \sim \{\pm 1\}^N}[\zeta \circ g^{\leqslant \ell}(x)] \leqslant \mathop{\mathbb{E}}_{z \sim S}[\zeta \circ g^{\leqslant \ell}(z)] + 2^{O(\ell)} \sqrt{\tau} \leqslant \underbrace{\rho^{\Omega(\gamma \ell)} + \tau^{1/4}}_{\eta :=}\,.$$

Now, as $\mathcal{D}^t$ is ℓ-wise independent ($\ell < 2^{d+1}$),

$$\mathop{\mathbb{E}}_{x \sim \{\pm 1\}^N}[g^{\leqslant \ell}(x)] = \mathop{\mathbb{E}}_{z \sim S}[g^{\leqslant \ell}(z)] = \mathop{\mathbb{E}}_{z \sim S}[g(z)] \pm \mathop{\mathbb{E}}_{z \sim S}[(g^{>\ell}(z))^2]^{1/2} \leqslant \mu + \sqrt{\eta}.$$

Therefore, by Corollary C.3,

$$\langle g^{\leqslant \ell}, T_{\rho'} g^{\leqslant \ell} \rangle = \sum_{\alpha : \mathsf{wt}(\alpha) \leqslant \ell} (\rho')^{\mathsf{wt}(\alpha)} \hat{g}(\alpha)^2 \leqslant \Gamma_{\rho'}(\mu + \sqrt{\eta}) + \frac{O(t \log\log(1/\tau))}{(1 - \rho') \log(1/\tau)}. \tag{C.8}$$

Since $\Gamma_{\rho'}(\mu + \sqrt{\eta}) \leqslant \Gamma_{\rho'}(\mu) + 2\sqrt{\eta}$ and $\Gamma_\rho(\mu) \leqslant \Gamma_{\rho'}(\mu) + |\rho - \rho'|/(1 - \rho)$ (cf. Lemma B.3, Corollary B.5 in [MOO05]), it follows from (C.7) and (C.8) that

$$\begin{aligned} \langle f, Gf \rangle = \langle g, G' g \rangle \leqslant \Gamma_\rho(\mu) + O\left(\frac{|\rho - \rho'|}{1 - \rho}\right) + O(\sqrt{\eta}) + \frac{O(t \log\log(1/\tau))}{(1 - \rho) \log(1/\tau)} \\ + \rho^{\Omega((1 - 2\gamma)\ell/t)} + \sqrt{\tau}. \end{aligned}$$

By a sufficiently small choice of τ, and fixing $\ell = \log(1/\tau)/4c_1$ and

$$\gamma = 100 t c_1 \log\log(1/\tau) / (\log(1/\tau) \log(1/\rho))$$

(so that $\rho^{\Omega(\gamma \ell / t)} < 1/\log(1/\tau)$ and $|\rho - \rho'| = O(\frac{t}{\log 1/\rho \log 1/\tau})$), the error term in the above expression can be made smaller than δ. $\square$

Acknowledgments. We have had many beneficial conversations on this question with a number of researchers, including Venkatesan Guruswami, Ran Raz, Alex Samorodnitsky, Amir Shpilka, Daniel Spielman, Ryan O'Donnell, Subhash Khot, and Madhu Sudan. We thank Avi Wigderson for suggesting the name "short code."

REFERENCES

[ABS10] S. Arora, B. Barak, and D. Steurer, *Subexponential algorithms for unique games and related problems*, in Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science (FOCS), 2010, pp. 563–572.

[AKK$^+$05] N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn, and D. Ron, *Testing Reed-Muller codes*, IEEE Trans. Inform. Theory, 51 (2005), pp. 4032–4039.

[BHK$^+$11] B. Barak, F. G. S. L. Brandao, A. W. Harrow, J. Kelner, D. Steurer, and Y. Zhou, *Hypercontractivity, sum-of-squares proofs, and their applications*, in Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC), New York, 2012, pp. 307–326.

[BKS$^+$10] A. Bhattacharyya, S. Kopparty, G. Schoenebeck, M. Sudan, and D. Zuckerman, *Optimal testing of Reed-Muller codes*, in Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science (FOCS), 2010, pp. 488–497.

[BRS11] B. Barak, P. Raghavendra, and D. Steurer, *Rounding semidefinite programming hierarchies via global correlation*, in Proceedings of the 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science (FOCS), 2011, pp. 472–481.

[Che70] J. Cheeger, *A lower bound for the smallest eigenvalue of the Laplacian*, Problems in Analysis (Papers dedicated to Salomon Bochner, 1969), Princeton University Press, Princeton, NJ, 1970, pp. 195–199.

[CMM06] M. Charikar, K. Makarychev, and Y. Makarychev, *Near-optimal algorithms for unique games*, in Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC), 2006, pp. 205–214.

[CT10] E. Chlamtac and M. Tulsiani, *Convex Relaxations and Integrality Gaps*, in Handbook on Semidefinite, Conic and Polynomial Optimization, Springer, New York, 2012, pp. 139–169.

[DGJ$^+$09] I. Diakonikolas, P. Gopalan, R. Jaiswal, R. A. Servedio, and E. Viola, *Bounded independence fools halfspaces*, SIAM J. Comput., 39 (2010), pp. 3441–3462.

[DKN10] I. Diakonikolas, D. M. Kane, and J. Nelson, *Bounded independence fools degree-*2 *threshold functions*, in Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science (FOCS), 2010, pp. 11–20.

[FL92] U. Feige and L. Lovasz, *Two-prover one-round proof systems: Their power and their problems*, in Proceedings of the 24th Annual ACM Symposium on Theory of Computing (Victoria, BC, Canada), N. Alon, ed., 1992, pp. 733–744.

[GS11] V. Guruswami and A. K. Sinop, *The complexity of finding independent sets in bounded degree (hyper)graphs of low chromatic number*, in Proceedings of the Twenty-Second Annual ACH-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, ACM, New York, (SODA), 2011.

[GT06] A. Gupta and K. Talwar, *Approximating unique games*, in Proceedings of the Seventeenth Annual ACH-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, ACM, New York, (SODA), 2006, pp. 99–106.

[Kan11] D. M. Kane, *k-independent Gaussians fool polynomial threshold functions*, in Proceedings of the 26th Annual IEEE Conference on Computational Complexity, 2011, pp. 252–261.

[Kho02] S. Khot, *On the power of unique* 2*-prover* 1*-round games*, in Proceedings of the 34th Annual ACH-SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, ACM, New York, (STOC), 2002, pp. 767–775.

[KKMO07] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell, *Optimal inapproximability results for MAX-CUT and other* 2*-variable CSPs?*, SIAM J. Comput., 37 (2007), pp. 319–357.

[Kol10] A. Kolla, *Spectral algorithms for unique games*, Comput. Complexity, 20 (2011), pp. 177–206.

[KS09] S. Khot and Rishi Saket, *SDP integrality gaps with local ℓ_1-embeddability*, in Proceedings of the 2009 IEEE 50th Annual Symposium on Foundations of Computer Science (FOCS), 2009, pp. 565–574.

[KT07] A. Kolla and M. Tulsiani, *Playing Random and Expanding Unique Games*, manuscript, 2007.

[KV05] S. Khot and N. K. Vishnoi, *The Unique Games Conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1*, in Proceedings of the 2005 IEEE 46th Annual Symposium on Foundations of Computer Science (FOCS), 2005, pp. 53–62.

[Lee11] J. Lee, *PSD Lifting and Unique Games Integrality Gaps*, tcs math blog, http://tcsmath.org/2011/02/23/psd-lifiting-and-unique-games-integrality-gaps/ (23 February 2011).

[MOO05] E. Mossel, R. O'Donnell, and K. Oleszkiewicz, *Noise stability of functions with low influences: invariance and optimality*, in Proceedings of the 2005 IEEE 46th Annual Symposium on Foundations of Computer Science (FOCS), 2005, pp. 21–30.

[MZ13] R. Meka and D. Zuckerman, *Pseudorandom generators for polynomial threshold functions*, SIAM J. Comput., 42 (2013), pp. 1275–1301.

[O'D08] R. O'Donnell, *Some topics in analysis of Boolean functions,* in Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC), 2008, pp. 569–578.

[RS09] P. Raghavendra and D. Steurer, *Integrality gaps for strong SDP relaxations of Unique Games*, in Proceedings of the 2009 IEEE 50th Annual Symposium on Foundations of Computer Science (FOCS), 2009, pp. 575–585.

[SA90] H. D. Sherali and W. P. Adams, *A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems*, SIAM J. Discrete Math., 3 (1990), pp. 411–430.