

THE APPROXIMATE DEGREE OF DNF AND CNF FORMULAS*

ALEXANDER A. SHERSTOV

ABSTRACT. The *approximate degree* of a Boolean function $f: \{0, 1\}^n \rightarrow \{0, 1\}$ is the minimum degree of a real polynomial p that approximates f pointwise: $|f(x) - p(x)| \leq 1/3$ for all $x \in \{0, 1\}^n$. For every $\delta > 0$, we construct CNF and DNF formulas of polynomial size with approximate degree $\Omega(n^{1-\delta})$, essentially matching the trivial upper bound of n . This improves polynomially on previous lower bounds and fully resolves the approximate degree of constant-depth circuits (AC^0), a question that has seen extensive research over the past 10 years. Prior to our work, an $\Omega(n^{1-\delta})$ lower bound was known only for AC^0 circuits of depth that grows with $1/\delta$ (Bun and Thaler, FOCS 2017). Furthermore, the CNF and DNF formulas that we construct are the simplest possible in that they have *constant* width. Our result holds even for *one-sided* approximation: for any $\delta > 0$, we construct a polynomial-size constant-width CNF formula with one-sided approximate degree $\Omega(n^{1-\delta})$.

Our work has the following consequences.

- (i) We essentially settle the communication complexity of AC^0 circuits in the bounded-error quantum model, k -party number-on-the-forehead randomized model, and k -party number-on-the-forehead nondeterministic model: we prove that for every $\delta > 0$, these models require $\Omega(n^{1-\delta})$, $\Omega(n/4^k k^2)^{1-\delta}$, and $\Omega(n/4^k k^2)^{1-\delta}$, respectively, bits of communication even for polynomial-size constant-width CNF formulas.
- (ii) In particular, we show that the multiparty communication class coNP_k can be separated essentially optimally from NP_k and BPP_k by a particularly simple function, a polynomial-size constant-width CNF formula.
- (iii) We give an essentially tight separation, of $O(1)$ versus $\Omega(n^{1-\delta})$, for the one-sided versus two-sided approximate degree of a function; and $O(1)$ versus $\Omega(n^{1-\delta})$ for the one-sided approximate degree of a function f versus its negation $\neg f$.

Our proof departs significantly from previous approaches and contributes a novel, number-theoretic method for amplifying approximate degree.

* This manuscript is a much-expanded version of the STOC '22 paper, with several new results. Work supported by NSF grants CCF-1814947 and CCF-2220232.

Author affiliation: Computer Science Department, UCLA, Los Angeles, CA 90095. *Email:* sherstov@cs.ucla.edu.

CONTENTS

1. Introduction	3
1.1. Approximate degree of DNF and CNF formulas	4
1.2. Large-error approximation	5
1.3. One-sided approximation	5
1.4. Randomized multiparty communication	7
1.5. Nondeterministic and Merlin–Arthur multiparty communication	9
1.6. Multiparty communication classes	10
1.7. Quantum communication complexity	11
1.8. Previous approaches	12
1.9. Our proof	14
2. Preliminaries	17
2.1. General notation	17
2.2. Boolean strings and functions	18
2.3. Concentration of measure	19
2.4. Orthogonal content	21
2.5. Polynomial approximation	22
2.6. Dual polynomials	22
2.7. Symmetrization	25
2.8. Number theory	25
3. Balanced colorings	26
3.1. Existence of balanced colorings	26
3.2. Discrepancy defined	28
3.3. A low-discrepancy set	29
3.4. Discrepancy and balanced colorings	31
3.5. An explicit balanced coloring	34
4. Hardness amplification	35
4.1. Pseudodistributions from balanced colorings	35
4.2. Encoding via indistinguishable distributions	41
4.3. Hardness amplification for approximate degree	45
4.4. Hardness amplification for one-sided approximate degree	50
4.5. Specializing the parameters	56
5. Main results	57
5.1. Approximate degree of DNF and CNF formulas	58
5.2. Quantum communication complexity	60
5.3. Randomized multiparty communication	63
5.4. Nondeterministic and Merlin–Arthur multiparty communication	66
Acknowledgments	73
References	73
Appendix A. Constructing low-discrepancy integer sets	75

1. INTRODUCTION

Representations of Boolean functions by real polynomials play a central role in theoretical computer science. Our focus in this paper is on *approximate degree*, a particularly natural and useful complexity measure. Formally, the ϵ -approximate degree of a Boolean function $f: \{0, 1\}^n \rightarrow \{0, 1\}$ is denoted $\deg_\epsilon(f)$ and defined as the minimum degree of a real polynomial p that approximates f within ϵ pointwise: $|f(x) - p(x)| \leq \epsilon$ for all $x \in \{0, 1\}^n$. The standard choice of the error parameter is $\epsilon = 1/3$, which is a largely arbitrary setting that can be replaced by any other constant in $(0, 1/2)$ without affecting the approximate degree by more than a multiplicative constant. Since every function $f: \{0, 1\}^n \rightarrow \{0, 1\}$ can be computed with zero error by a polynomial of degree at most n , the ϵ -approximate degree is always at most n .

The notion of approximate degree originated three decades ago in the pioneering work of Nisan and Szegedy [31] and has since proved to be a powerful tool in theoretical computer science. Upper bounds on approximate degree have algorithmic applications, whereas lower bounds are a staple in complexity theory. On the algorithmic side, approximate degree underlies many of the strongest results obtained to date in computational learning, differentially private data release, and algorithm design in general. In complexity theory, the notion of approximate degree has produced breakthroughs in quantum query complexity, communication complexity, and circuit complexity. A detailed bibliographic overview of these applications can be found in [47, 17].

Approximate degree has been particularly prominent in the study of AC^0 , the class of polynomial-size constant-depth circuits with gates \vee, \wedge, \neg of unbounded fan-in. The simplest functions in AC^0 are conjunctions and disjunctions, which have depth 1, followed by polynomial-size CNF and DNF formulas, which have depth 2, followed in turn by higher-depth circuits. Lower bounds on the approximate degree of AC^0 functions have been used to settle the quantum query complexity of Grover search [8], element distinctness [1], and a host of other problems [14]; resolve the communication complexity of set disjointness in the two-party quantum model [33, 38] and number-on-the-forehead multiparty model [37, 38, 28, 20, 36, 11, 44, 43]; separate the communication complexity classes PP and UPP [13, 37]; and separate the polynomial hierarchy in communication complexity from the communication class UPP [34]. Despite this array of applications and decades of study, our understanding of the approximate degree of AC^0 has remained surprisingly fragmented and incomplete. In this paper, we set out to resolve this question in full.

In more detail, previous work on the approximate degree of AC^0 started with the seminal 1994 paper of Nisan and Szegedy [31], who proved that the OR function on n bits has approximate degree $\Theta(\sqrt{n})$. This was the best result until Aaronson and Shi's celebrated lower bound of $\Omega(n^{2/3})$ for the element distinctness problem [1]. In a beautiful paper from 2017, Bun and Thaler [17] showed that AC^0 contains functions in n variables with approximate degree $\Omega(n^{1-\delta})$, where the constant $\delta > 0$ can be made arbitrarily small at the expense of increasing the depth of the circuit. In follow-up work, Bun and Thaler [18] proved an $\Omega(n^{1-\delta})$ lower bound for approximating AC^0 circuits even with error exponentially close to $1/2$, where once again the circuit depth grows with $1/\delta$. A stronger yet result was obtained by Sherstov and Wu [49], who showed that AC^0 has essentially the maximum possible *threshold degree* (defined as the limit of ϵ -approximate degree as $\epsilon \nearrow 1/2$) and *sign-rank* (a generalization of threshold degree to arbitrary bases rather than just the basis of

monomials). Quantitatively, the authors of [49] proved a lower bound of $\Omega(n^{1-\delta})$ for threshold degree and $\exp(\Omega(n^{1-\delta}))$ for sign-rank, essentially matching the trivial upper bounds. As before, $\delta > 0$ can be made arbitrarily small at the expense of increasing the circuit depth. In particular, AC^0 requires a polynomial of degree $\Omega(n^{1-\delta})$ even for approximation to error doubly (triply, quadruply, quintuply...) exponentially close to $1/2$.

The lower bounds of [17, 18, 49] show that AC^0 functions have essentially the maximum possible complexity—but only if one is willing to look at circuits of *arbitrarily large* constant depth. What happens at *small* depths has been a wide open problem, with no techniques to address it. Bun and Thaler observe that their AC^0 circuit in [17] with approximate degree $\Omega(n^{1-\delta})$ can be flattened to produce a DNF formula of size $\exp(\log^{O(\log(1/\delta))} n)$, but this is superpolynomial and thus no longer in AC^0 . The only progress of which we are aware is an $\Omega(n^{3/4-\delta})$ lower bound obtained for polynomial-size DNF formulas in [14, 29]. This leaves a polynomial gap in the approximate degree for small depth versus arbitrary constant depth. Our main contribution is to definitively resolve the approximate degree of AC^0 by constructing, for any constant $\delta > 0$, a polynomial-size DNF formula with approximate degree $\Omega(n^{1-\delta})$. We now describe our main result and its generalizations and applications.

1.1. Approximate degree of DNF and CNF formulas. Recall that a *literal* is a Boolean variable x_1, x_2, \dots, x_n or its negation $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$. A conjunction of literals is called a *term*, and a disjunction of literals is called a *clause*. The *width* of a term or clause is the number of literals that it contains. A *DNF formula* is a disjunction of terms, and analogously a *CNF formula* is a conjunction of clauses. The *width* of a DNF or CNF formula is the maximum width of a term or clause in it, respectively. One often refers to DNF and CNF formulas of width k as k -DNF and k -CNF formulas. The *size* of a DNF or CNF formula is the total number of terms or clauses, respectively, that it contains. Thus, AC^0 circuits of depth 1 correspond precisely to clauses and terms, whereas AC^0 circuits of depth 2 correspond precisely to polynomial-size DNF and CNF formulas. Our main result on approximate degree is as follows.

THEOREM 1.1 (Main result). *Let $\delta > 0$ be any constant. Then for each $n \geq 1$, there is an (explicitly given) function $f: \{0, 1\}^n \rightarrow \{0, 1\}$ that has approximate degree*

$$\deg_{1/3}(f) = \Omega(n^{1-\delta})$$

and is computable by a DNF formula of size $n^{O(1)}$ and width $O(1)$.

Theorem 1.1 almost matches the trivial upper bound of n on the approximate degree of any function. Thus, the theorem shows that AC^0 circuits of depth 2 already achieve essentially the maximum possible approximate degree. This depth cannot be reduced further because AC^0 circuits of depth 1 have approximate degree $O(\sqrt{n})$. Finally, the DNF formulas constructed in Theorem 1.1 are the simplest possible in that they have *constant* width.

Recall that previously, a lower bound of $\Omega(n^{1-\delta})$ for AC^0 was known only for circuits of large constant depth that grows with $1/\delta$. The lack of progress on small-depth AC^0 prior to this paper had experts seriously entertaining [18] the possibility that AC^0 circuits of any given depth d have approximate degree $O(n^{1-\delta d})$, for

some constant $\delta_d = \delta_d(d) > 0$. Such an upper bound would have far-reaching consequences in computational learning and circuit complexity. Theorem 1.1 rules it out.

1.2. Large-error approximation. Any Boolean function can be approximated pointwise within $1/2$ in a trivial manner, by a constant polynomial. Approximation within $\frac{1}{2} - o(1)$, on the other hand, is a meaningful and extremely useful notion. We obtain the following strengthening of our main result, in which the approximation error is relaxed from $1/3$ to an optimal $\frac{1}{2} - \frac{1}{n^{\Theta(1)}}$.

THEOREM 1.2 (Main result for large error). *Let $\delta > 0$ and $C \geq 1$ be any constants. Then for each $n \geq 1$, there is an (explicitly given) function $f: \{0, 1\}^n \rightarrow \{0, 1\}$ that has approximate degree*

$$\deg_{\frac{1}{2} - \frac{1}{n^C}}(f) = \Omega(n^{1-\delta})$$

and is computable by a DNF formula of size $n^{O(1)}$ and width $O(1)$.

To rephrase Theorem 1.2, polynomial-size DNF formulas require degree $\Omega(n^{1-\delta})$ for approximation not only to constant error but even to error $\frac{1}{2} - \frac{1}{n^C}$, where $C \geq 1$ is an arbitrarily large constant. The error parameter in Theorem 1.2 cannot be relaxed further to $\frac{1}{2} - \frac{1}{n^{\omega(1)}}$ because any DNF formula with m terms can be approximated to error $\frac{1}{2} - \Omega(\frac{1}{m})$ by a polynomial of degree $O(\sqrt{n \log m})$.

Negating a function has no effect on the approximate degree. Indeed, if f is approximated to error ϵ by a polynomial p , then the negated function $\neg f = 1 - f$ is approximated to the same error ϵ by the polynomial $1 - p$. With this observation, Theorems 1.1 and 1.2 carry over to CNF formulas:

COROLLARY 1.3. *Let $\delta > 0$ and $C \geq 1$ be any constants. Then for each $n \geq 1$, there is an (explicitly given) function $g: \{0, 1\}^n \rightarrow \{0, 1\}$ that has approximate degree*

$$\deg_{\frac{1}{2} - \frac{1}{n^C}}(g) = \Omega(n^{1-\delta})$$

and is computable by a CNF formula of size $n^{O(1)}$ and width $O(1)$.

1.3. One-sided approximation. There is a natural notion of *one-sided* approximation for Boolean functions. Specifically, the *one-sided ϵ -approximate degree* of a function $f: \{0, 1\}^n \rightarrow \{0, 1\}$ is defined as the minimum degree of a real polynomial p such that

$$\begin{aligned} f(x) = 0 &\quad \Rightarrow \quad p(x) \in [-\epsilon, \epsilon], \\ f(x) = 1 &\quad \Rightarrow \quad p(x) \in [1 - \epsilon, +\infty) \end{aligned}$$

for every $x \in \{0, 1\}^n$. This complexity measure is denoted $\deg_\epsilon^+(f)$. It plays a considerable role [23, 15, 40, 16, 45, 44, 43] in the area, both in its own right and due to its applications to other asymmetric notions of computation such as non-determinism and Merlin–Arthur protocols. One-sided approximation is meaningful for any error parameter $\epsilon \in [0, 1/2)$, and as before the standard setting is $\epsilon = 1/3$. By definition, one-sided approximate degree is always at most n . Observe that the definitions of ϵ -approximate degree $\deg_\epsilon(f)$ and its one-sided variant $\deg_\epsilon^+(f)$ impose the same requirement for inputs $x \in f^{-1}(0)$: the approximating polynomial

must approximate f within ϵ at each such x . For inputs $x \in f^{-1}(1)$, on the other hand, the definitions of $\deg_\epsilon(f)$ and $\deg_\epsilon^+(f)$ diverge dramatically, with one-sided ϵ -approximate degree not requiring any upper bound on the approximating polynomial p . As a result, one always has $\deg_\epsilon^+(f) \leq \deg_\epsilon(f)$, and it is reasonable to expect a large gap between the two quantities for some f . Moreover, the one-sided approximate degree of a function is in general not equal to that of its negation: $\deg_\epsilon^+(f) \neq \deg_\epsilon^+(-f)$. This contrasts with the equality $\deg_\epsilon(f) = \deg_\epsilon(-f)$ for two-sided approximation.

In this light, there are three particularly natural questions to ask about one-sided approximate degree:

- (i) What is the one-sided approximate degree of AC^0 circuits?
- (ii) What is the largest possible gap between approximate degree and one-sided approximate degree?
- (iii) What is the largest possible gap between the one-sided approximate degree of a function f and that of its negation $\neg f$?

In this paper, we resolve all three questions in detail. For question (i), we prove that polynomial-size CNF formulas achieve essentially the maximum possible one-sided approximate degree. In fact, our result holds even for approximation to error vanishingly close to random guessing, $\frac{1}{2} - o(1)$:

THEOREM 1.4. *Let $\delta > 0$ and $C \geq 1$ be any constants. Then for each $n \geq 1$, there is an (explicitly given) function $g: \{0, 1\}^n \rightarrow \{0, 1\}$ that has one-sided approximate degree*

$$\deg_{\frac{1}{2} - \frac{1}{nC}}^+(g) = \Omega(n^{1-\delta})$$

and is computable by a CNF formula of size $n^{O(1)}$ and width $O(1)$.

Theorem 1.4 essentially settles the one-sided approximate degree of AC^0 . The theorem is optimal with respect to circuit depth; recall that depth-1 circuits have approximate degree $O(\sqrt{n})$ and hence also one-sided approximate degree $O(\sqrt{n})$. Previous work on the one-sided approximate degree of AC^0 was suboptimal with respect to the degree bound and/or circuit depth. Specifically, the best previous lower bounds were $\Omega(n/\log n)^{2/3}$ due to Bun and Thaler [16] for a polynomial-size CNF formula, and $\Omega(n^{1-\delta})$ due to Sherstov and Wu [49] for AC^0 circuits of depth that grows with $1/\delta$.

As an application of Theorem 1.4, we resolve questions (ii) and (iii) in full, establishing a gap of $O(1)$ versus $\Omega(n^{1-\delta})$ in each case. Moreover, we prove that these gaps remain valid well beyond the standard error regime of $\epsilon = 1/3$. A detailed statement of our separations follows.

COROLLARY 1.5. *Let $\delta > 0$ and $C \geq 1$ be any constants. Then for each $n \geq 1$, there is an (explicitly given) function $f: \{0, 1\}^n \rightarrow \{0, 1\}$ with*

$$\deg_0^+(f) = O(1) \tag{1.1}$$

but

$$\deg_{\frac{1}{2} - \frac{1}{n^c}}(f) = \Omega(n^{1-\delta}), \quad (1.2)$$

$$\deg_{\frac{1}{2} - \frac{1}{n^c}}^+(\neg f) = \Omega(n^{1-\delta}). \quad (1.3)$$

Moreover, f is computable by a DNF formula of size $n^{O(1)}$ and width $O(1)$.

Equations (1.1) and (1.2) in this result give the promised $O(1)$ versus $\Omega(n^{1-\delta})$ separation for question (ii). Analogously, (1.1) and (1.3) give an $O(1)$ versus $\Omega(n^{1-\delta})$ separation for question (iii). Of particular note in both separations is the error regime: the upper bound remains valid even under the stronger requirement of zero error, whereas the lower bounds remain valid even under the weaker requirement of error $\frac{1}{2} - o(1)$. Our separations improve on previous work. For question (ii), the best previous separation was $(\log n)^{O_\delta(1)}$ versus $\Omega(n^{1-\delta})$ for any fixed $\delta > 0$, implicit in [17]. For the harder question (iii), the best previous separation [16] was $O(\log n)$ versus $\Omega(n/\log n)^{2/3}$, which is polynomially weaker than ours.

The derivation of Corollary 1.5 from Theorem 1.4 is short and illustrative, and we include it here.

Proof of Corollary 1.5. Let g be the function from Theorem 1.4, and set $f = \neg g$. Then (1.3) is immediate. Equation (1.2) follows from (1.3) in light of the basic relations $\deg_\epsilon(f) = \deg_\epsilon(\neg f) \geq \deg_\epsilon^+(\neg f)$, valid for all f and ϵ . Finally, (1.1) can be seen as follows. Since g is a CNF formula of width $O(1)$, its negation f is a DNF formula of width $O(1)$. Thus, every term of f can be represented exactly by a polynomial of degree $O(1)$. Summing these polynomials gives a 0-error one-sided approximant for f . \square

We now discuss applications of our results on approximate degree and one-sided approximate degree to fundamental questions in communication complexity.

1.4. Randomized multiparty communication. We adopt the *number-on-the-forehead* model of Chandra, Furst, and Lipton [19], which is the most powerful formalism of multiparty communication. The model features k communicating players and a Boolean function $F: X_1 \times X_2 \times \cdots \times X_k \rightarrow \{0, 1\}$ with k arguments. An input (x_1, x_2, \dots, x_k) is distributed among the k players by giving the i -th player the arguments $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k$ but not x_i . This arrangement can be visualized as having the k players seated in a circle with x_i written on the i -th player's forehead, whence the name of the model. Number-on-the-forehead is the canonical model in the area because any other way of assigning arguments to players results in a less powerful model—provided of course that one does not assign all the arguments to some player, in which case there is never a need to communicate.

The players communicate according to a protocol agreed upon in advance. The communication occurs in the form of broadcasts, with a message sent by any given player instantly reaching everyone else. The players' objective is to compute F on any given input with minimal communication. To this end, the players have access to an unbounded supply of shared random bits which they can use in deciding what message to send at any given point in the protocol. The *cost* of a protocol is the total bit length of all the messages broadcast in a worst-case execution. The ϵ -error randomized communication complexity $R_\epsilon(F)$ of a given function F is the least cost

of a protocol that computes F with probability of error at most ϵ on every input. As with approximate degree, the standard setting of the error parameter is $\epsilon = 1/3$.

The number-on-the-forehead communication complexity of constant-depth circuits is a challenging question that has been the focus of extensive research, e.g., [12, 28, 20, 36, 11, 44, 43, 17]. In contrast to the two-party model, where a lower bound of $\Omega(\sqrt{n})$ for AC^0 circuits is straightforward to prove from first principles [4], the first $n^{\Omega(1)}$ multiparty lower bound [44] for AC^0 was obtained only in 2012. The strongest known multiparty lower bounds for AC^0 are obtained using the *pattern matrix method* of [43], which transforms approximate degree lower bounds in a black-box manner into communication lower bounds. In the most recent application of this method, Bun and Thaler [17] gave a k -party communication problem $F: (\{0, 1\}^n)^k \rightarrow \{0, 1\}$ in AC^0 with communication complexity $\Omega(n/4^k k^2)^{1-\delta}$, where the constant $\delta > 0$ can be taken arbitrarily small at the expense of increasing the depth of the AC^0 circuit. This shows that AC^0 has essentially the maximum possible multiparty communication complexity—as long as one is willing to use circuits of arbitrarily large constant depth. For circuits of small depth, the best lower bound is polynomially weaker: $\Omega(n/4^k k^2)^{3/4-\delta}$ for the k -party communication complexity of polynomial-size DNF formulas, which can be proved by applying the pattern matrix method to the approximate degree lower bounds in [14, 29]. This fragmented state of the art closely parallels that for approximate degree prior to our work.

We resolve the multiparty communication complexity of AC^0 in detail in the following theorem.

THEOREM 1.6. *Fix any constants $\delta \in (0, 1]$ and $C \geq 1$. Then for all integers $n, k \geq 2$, there is an (explicitly given) k -party communication problem $F_{n,k}: (\{0, 1\}^n)^k \rightarrow \{0, 1\}$ with*

$$R_{1/3}(F_{n,k}) \geq \left(\frac{n}{c' 4^k k^2} \right)^{1-\delta},$$

$$R_{\frac{1}{2} - \frac{1}{n^C}}(F_{n,k}) \geq \frac{n^{1-\delta}}{c' 4^k},$$

where $c' \geq 1$ is a constant independent of n and k . Moreover, $F_{n,k}$ is computable by a DNF formula of size $n^{c'}$ and width $c'k$.

Theorem 1.6 essentially represents the state of the art for multiparty communication lower bounds. Indeed, the best communication lower bound to date for any explicit function $F: (\{0, 1\}^n)^k \rightarrow \{0, 1\}$, whether or not F is computable by an AC^0 circuit, is $\Omega(n/2^k)$ [6]. Theorem 1.6 comes close to matching the trivial upper bound of $n+1$ for any communication problem, thereby showing that AC^0 circuits of depth 2 achieve nearly the maximum possible communication complexity. Moreover, our result holds not only for bounded-error communication but also for communication with error $\frac{1}{2} - \frac{1}{n^C}$ for any $C \geq 1$. The error parameter in Theorem 1.6 is optimal and cannot be further increased to $\frac{1}{2} - \frac{1}{n^{\omega(1)}}$; indeed, it is straightforward to see that any DNF formula with m terms has a communication protocol with error $\frac{1}{2} - \Omega(\frac{1}{m})$ and cost 2 bits. Theorem 1.6 is also optimal with respect to circuit depth because the multiparty communication complexity of AC^0 circuits of depth 1 is at most 2 bits.

Since randomized communication complexity is invariant under function negation, Theorem 1.6 remains valid with the word “DNF” replaced with “CNF.”

1.5. Nondeterministic and Merlin–Arthur multiparty communication.

Here again, we adopt the k -party number-on-the-forehead model of Chandra, Furst, and Lipton [19]. Nondeterministic communication is defined in complete analogy with computational complexity. Specifically, a nondeterministic protocol starts with a guess string, whose length counts toward the protocol’s communication cost, and proceeds deterministically thenceforth. A nondeterministic protocol for a given communication problem $F: X_1 \times X_2 \times \cdots \times X_k \rightarrow \{0, 1\}$ is required to output the correct answer for *all* guess strings when presented with a negative instance of F , and for *some* guess string when presented with a positive instance. We further consider *Merlin–Arthur protocols* [3, 5], a communication model that combines the power of randomization and nondeterminism. As before, a Merlin–Arthur protocol for a given problem F starts with a guess string, whose length counts toward the communication cost. From then on, the parties run an ordinary randomized protocol. The randomized phase in a Merlin–Arthur protocol must produce the correct answer with probability at least $2/3$ for *all* guess strings when presented with a negative instance of F , and for *some* guess string when presented with a positive instance. Thus, the cost of a nondeterministic or Merlin–Arthur protocol is the sum of the costs of the guessing phase and communication phase. The minimum cost of a valid protocol for F in these models is called the *nondeterministic communication complexity of F* , denoted $N(F)$, and *Merlin–Arthur communication complexity of F* , denoted $MA_{1/3}(F)$. The quantity $N(-F)$ is called the *co-nondeterministic communication complexity of F* .

Nondeterministic and Merlin–Arthur protocols have been extensively studied for $k = 2$ parties but are much less understood in the multiparty setting [10, 23, 44, 43]. Prior to our paper, the best lower bounds in these models for an AC^0 circuit $F: (\{0, 1\}^n)^k \rightarrow \{0, 1\}$ were $\Omega(\sqrt{n}/2^k k)$ for nondeterministic communication and $\Omega(\sqrt{n}/2^k k)^{1/2}$ for Merlin–Arthur communication, obtained in [43] for the set disjointness problem. We give a quadratic improvement on these lower bounds. In particular, our result for nondeterminism essentially matches the trivial upper bound. Moreover, we obtain our result for a particularly simple function in AC^0 , namely, a polynomial-size CNF formula of constant width. A detailed statement follows.

THEOREM 1.7. *Let $\delta > 0$ be arbitrary. Then for all integers $n, k \geq 2$, there is an (explicitly given) k -party communication problem $G_{n,k}: (\{0, 1\}^n)^k \rightarrow \{0, 1\}$ with*

$$N(-G_{n,k}) \leq c \log n$$

but

$$N(G_{n,k}) \geq \left(\frac{n}{c4^k k^2} \right)^{1-\delta}, \quad (1.4)$$

$$R_{1/3}(G_{n,k}) \geq \left(\frac{n}{c4^k k^2} \right)^{1-\delta}, \quad (1.5)$$

$$MA_{1/3}(G_{n,k}) \geq \left(\frac{n}{c4^k k^2} \right)^{\frac{1-\delta}{2}}, \quad (1.6)$$

where $c \geq 1$ is a constant independent of n and k . Moreover, $G_{n,k}$ is computable by a CNF formula of width ck and size n^c .

This result can be viewed as a far-reaching generalization of Theorem 1.6 to non-deterministic and Merlin–Arthur protocols. To obtain Theorem 1.7, we adapt the pattern matrix method [43] to be able to transform any lower bound on one-sided approximate degree into a multiparty communication lower bound in the nondeterministic and Merlin–Arthur models. With this tool in hand, we obtain Theorem 1.7 from our one-sided approximate degree lower bound (Theorem 1.4).

1.6. Multiparty communication classes. Theorem 1.7 sheds new light on communication complexity classes, defined in the seminal work of Babai, Frankl, and Simon [4]. An infinite family $\{F_n\}_{n=1}^\infty$, where each $F_n: (\{0,1\}^n)^k \rightarrow \{0,1\}$ is a k -party number-on-the-forehead communication problem, is said to be *efficiently solvable* in a given model of communication if F_n has communication complexity at most $\log^c n$ in that model, for a large enough constant $c > 1$ and all $n > c$. One defines BPP_k , NP_k , coNP_k , and MA_k as the classes of families that are efficiently solvable in the randomized, nondeterministic, co-nondeterministic, and Merlin–Arthur models, respectively. In particular, MA_k is a superset of NP_k and BPP_k . In these definitions, $k = k(n)$ can be any function of n , including constant functions such as $k = 3$. The relations among these multiparty classes have been actively studied over the past decade [9, 28, 20, 22, 11, 10, 23, 44, 43]. In particular, for $k \leq \Theta(\log n)$, it is known that coNP_k is not contained in BPP_k , NP_k , or even MA_k . Quantitatively, these results can be summarized as follows.

- (i) Prior to our work, the strongest k -party separation of co-nondeterministic versus randomized communication complexity was $O(\log n)$ versus $\Omega(\sqrt{n}/2^k k)$, proved in [43] for the set disjointness function.
- (ii) The best previous k -party separations of co-nondeterministic versus nondeterministic communication complexity were: $O(\log n)$ versus $\Omega(n)$, proved in [43] nonconstructively by the probabilistic method; and $O(\log n)$ versus $\Omega(\sqrt{n}/2^k k)$, proved in [43] for the set disjointness problem.
- (iii) The best previous k -party separation of co-nondeterministic versus Merlin–Arthur communication complexity was $O(\log n)$ versus $\Omega(\sqrt{n}/2^k k)^{1/2}$, proved in [43] for the set disjointness problem.

Theorem 1.7 gives a quadratic improvement on these previous separations, excluding the nonconstructive separation of coNP_k from NP_k in [10]. Moreover, our quadratically improved separations are achieved for a particularly simple function, namely, the polynomial-size constant-width CNF formula $G_{n,k}$. In the regime $k \leq \Theta(\log n)$, our separations of coNP_k from BPP_k and NP_k are essentially optimal, and our separation of coNP_k from MA_k is within a square of optimal. Recall that no explicit lower bounds at all are currently known in the regime $k \geq \log n$, even for deterministic communication. We state our contributions for communication complexity classes as a corollary below.

COROLLARY 1.8. *Let $k = k(n)$ be a function with $k(n) \leq (\frac{1}{2} - \epsilon) \log n$ for some constant $\epsilon > 0$. Then the communication problem $G_{n,k}$ from Theorem 1.7 satisfies*

$$\{G_{n,k}\}_{n=1}^\infty \in \text{coNP}_k \setminus \text{BPP}_k,$$

$$\begin{aligned} \{G_{n,k}\}_{n=1}^{\infty} &\in \text{coNP}_k \setminus \text{NP}_k, \\ \{G_{n,k}\}_{n=1}^{\infty} &\in \text{coNP}_k \setminus \text{MA}_k. \end{aligned}$$

Analogously, the communication problem $F_{n,k}$ from Theorem 1.6 satisfies

$$\{F_{n,k}\}_{n=1}^{\infty} \in \text{NP}_k \setminus \text{BPP}_k.$$

Proof. The claims for $G_{n,k}$ are immediate from Theorem 1.7 and the definitions of NP_k , coNP_k , BPP_k , MA_k . For the remaining separation, we need only prove the upper bound $N(F_{n,k}) = O(\log n)$. Recall from Theorem 1.6 that $F_{n,k}$ is a DNF formula with $n^{c'}$ terms. This gives the desired nondeterministic protocol: the parties “guess” one of the terms in $F_{n,k}$ (for a cost of $\lceil \log n^{c'} \rceil$ bits), evaluate it (using another 2 bits of communication), and output the result. \square

1.7. Quantum communication complexity. We adopt the standard model of quantum communication, where two parties exchange quantum messages according to an agreed-upon protocol in order to solve a two-party communication problem $F: X \times Y \rightarrow \{0, 1\}$. As usual, an input $(x, y) \in X \times Y$ is split between the parties, with one party knowing only x and the other party knowing only y . We allow arbitrary prior entanglement at the start of the communication. A measurement at the end of the protocol produces a single-bit answer, which is interpreted as the protocol output. An ϵ -error protocol for F is required to output, on every input $(x, y) \in X \times Y$, the correct value $F(x, y)$ with probability at least $1 - \epsilon$. The cost of a quantum protocol is the total number of quantum bits exchanged in the worst case on any input. The ϵ -error quantum communication complexity of F , denoted $Q_{\epsilon}^*(F)$, is the least cost of an ϵ -error quantum protocol for F . The asterisk in $Q_{\epsilon}^*(F)$ indicates that the parties share arbitrary prior entanglement. The standard setting of the error parameter is $\epsilon = 1/3$, which is as usual without loss of generality. For a detailed formal description of the quantum model, we refer the reader to [51, 33, 38].

Proving lower bounds for bounded-error quantum communication is significantly more challenging than for randomized communication. An illustrative example is the set disjointness problem on n bits. Babai, Frankl, and Simon [4] obtained an $\Omega(\sqrt{n})$ randomized communication lower bound for this function in 1986 using a short and elementary proof, which was later improved to a tight $\Omega(n)$ in [25, 32, 7]. This is in stark contrast with the quantum model, where the best lower bound for set disjointness was for a long time a trivial $\Omega(\log n)$ until a tight $\Omega(\sqrt{n})$ was proved by Razborov [33] in 2002.

A completely different proof of the $\Omega(\sqrt{n})$ lower bound for set disjointness was given in [38] by introducing the pattern matrix method. Since then, the method has produced the strongest known quantum lower bounds for AC^0 . Of these, the best lower bound prior to our work was $\Omega(n^{1-\delta})$ due to Bun and Thaler [17], where the constant $\delta > 0$ can be taken arbitrarily small at the expense of circuit depth. In the following theorem, we resolve the quantum communication complexity of AC^0 in full by proving that polynomial-size DNF formulas achieve near-maximum communication complexity.

THEOREM 1.9. *Let $\delta > 0$ and $C \geq 1$ be any constants. Then for each $n \geq 1$, there is an (explicitly given) two-party communication problem $F: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$*

that has quantum communication complexity

$$Q_{\frac{1}{2} - \frac{1}{n^C}}^*(F) = \Omega(n^{1-\delta})$$

and is representable by a DNF formula of size $n^{O(1)}$ and width $O(1)$.

This theorem remains valid for CNF formulas since quantum communication complexity is invariant under function negation. As in all of our results, Theorem 1.9 essentially matches the trivial upper bound, showing that AC^0 circuits of depth 2 achieve nearly the maximum possible complexity. Again analogous to our other results, Theorem 1.9 holds not only for bounded-error communication but also for communication with error $\frac{1}{2} - \frac{1}{n^C}$ for any $C \geq 1$. The error parameter in Theorem 1.9 is optimal and cannot be further increased to $\frac{1}{2} - \frac{1}{n^{\omega(1)}}$: as remarked above, any DNF formula with m terms has a classical communication protocol with error $\frac{1}{2} - \Omega(\frac{1}{m})$ and cost 2 bits. Lastly, Theorem 1.9 is optimal with respect to circuit depth because AC^0 circuits of depth 1 have communication complexity at most 2 bits even in the classical deterministic model.

In our overview so far, we have separately considered the classical multiparty model and the quantum two-party model. By combining the features of these models, one arrives at the k -party number-on-the-forehead model with quantum players. Our results readily generalize to this setting. Specifically, for any constants $\delta > 0$ and $C \geq 1$, we give an explicit DNF formula $F_{n,k}: (\{0, 1\}^n)^k \rightarrow \{0, 1\}$ of size $n^{O(1)}$ and width $O(k)$ such that computing $F_{n,k}$ in the k -party quantum number-on-the-forehead model with error $\frac{1}{2} - \frac{1}{n^C}$ requires $\Omega(n^{1-\delta}/4^k k)$ quantum bits. For more details, see Remark 5.9.

1.8. Previous approaches. In the remainder of the introduction, we sketch our proof of Theorem 1.1. To properly set the stage for our work, we start by reviewing the relevant background and previous approaches. The notation that we adopt below is standard, and we defer its formal review to Section 2.

Dual view of approximation. Let $f: X \rightarrow \{0, 1\}$ be a Boolean function of interest, where X is an arbitrary finite subset of Euclidean space. The approximate degree of f is defined analogously to functions on the Boolean hypercube: $\text{deg}_\epsilon(f)$ is the minimum degree of a real polynomial p such that $|f(x) - p(x)| \leq \epsilon$ for every $x \in X$. A valuable tool in the analysis of approximate degree is linear programming duality, which gives a powerful *dual* view of approximation [38]. This dual characterization states that $\text{deg}_\epsilon(f) \geq d$ if and only if there is a function $\phi: X \rightarrow \mathbb{R}$ with the following two properties: $\langle \phi, f \rangle > \epsilon \|\phi\|_1$; and $\langle \phi, p \rangle = 0$ for every polynomial p of degree less than d . Rephrasing, ϕ must be correlated with f but completely uncorrelated with any polynomial of degree less than d . Such a function ϕ is variously referred to in the literature as a “dual object,” “dual polynomial,” or “witness” for f . The dual characterization makes it possible to prove any approximate degree lower bound by constructing the corresponding witness ϕ . This good news comes with a caveat: for all but the simplest functions, the construction of ϕ is very demanding, and linear programming duality gives no guidance in this regard.

Componentwise composition. The construction of a dual object is more approachable for composed functions since one can hope to break them up into constituent parts, construct a dual object for each, and recombine these results. Formally, define

the *componentwise composition* of functions $f: \{0, 1\}^n \rightarrow \{0, 1\}$ and $g: X \rightarrow \{0, 1\}$ as the Boolean function $f \circ g: X^n \rightarrow \{0, 1\}$ given by $(f \circ g)(x_1, \dots, x_n) = f(g(x_1), \dots, g(x_n))$. To construct a dual object for $f \circ g$, one starts by obtaining dual objects ϕ and ψ for the constituent functions f and g , respectively, either by direct construction or by appeal to linear programming duality. They are then combined to yield a dual object Φ for the composed function, using *dual componentwise composition* [41, 26]:

$$\Phi(x_1, x_2, \dots, x_n) = \phi(\mathbf{I}[\psi(x_1) > 0], \dots, \mathbf{I}[\psi(x_n) > 0]) \prod_{i=1}^n |\psi(x_i)|. \quad (1.7)$$

This composed dual object typically requires additional work to ensure strong enough correlation with the composed function $f \circ g$. Among the generic tools available to assist in this process is a “corrector” object ζ due to Razborov and Sherstov [34], with the following four properties: (i) ζ is orthogonal to low-degree polynomials; (ii) ζ takes on 1 at a prescribed point of the hypercube; (iii) ζ is bounded at inputs of low Hamming weight; and (iv) ζ vanishes at all other points of the hypercube. Using ζ , suitably shifted and scaled, one can surgically correct the behavior of a given dual object Φ at a substantial fraction of the inputs without affecting Φ ’s orthogonality to low-degree polynomials. This technique played an important role in previous work, e.g., [17, 14, 18, 49].

Componentwise composition by itself does not allow one to construct hard-to-approximate functions from easy ones. To see why, consider arbitrary functions $f: \{0, 1\}^{n_1} \rightarrow \{0, 1\}$ and $g: \{0, 1\}^{n_2} \rightarrow \{0, 1\}$ with approximate degrees at most n_1^α and n_2^α , respectively, for some $0 < \alpha < 1$. It is well-known [42] that the composed function $f \circ g$ on $n_1 n_2$ variables has approximate degree $O(n_1^\alpha n_2^\alpha) = O(n_1 n_2)^\alpha$. This means that relative to the new number of variables, the composed function $f \circ g$ is asymptotically no harder to approximate than the constituent functions f and g . In particular, one cannot use componentwise composition to transform functions on n bits with $1/3$ -approximate degree at most n^α into functions on N bits with $1/3$ -approximate degree $\omega(N^\alpha)$.

Previous best bound for AC^0 . In the previous best result on the $1/3$ -approximate degree of AC^0 , Bun and Thaler [17] approached the componentwise composition $f \circ g$ in an ingenious way to amplify the approximate degree for a careful choice of g . Let $f: \{0, 1\}^n \rightarrow \{0, 1\}$ be given, with $1/3$ -approximate degree n^α for some $0 \leq \alpha < 1$. Bun and Thaler consider the componentwise composition $F = f \circ (\text{AND}_{\Theta(\log m)} \circ \text{OR}_m)$, for a small enough parameter $m = \text{poly}(n)$. It was shown in earlier work [41, 16] that dual componentwise composition witnesses the lower bound $\deg_{1/3}(F) = \Omega(\deg_{1/3}(\text{OR}_m) \deg_{1/3}(f)) = \Omega(\sqrt{m} \deg_{1/3}(f))$. Bun and Thaler make the crucial observation that the dual object for OR_m has most of its ℓ_1 mass on inputs of Hamming weight $O(1)$, which in view of (1.7) implies that the dual object for F places most of its ℓ_1 mass on inputs of Hamming weight $\tilde{O}(n)$. The authors of [17] then use the Razborov–Sherstov corrector object to transfer the small amount of ℓ_1 mass that the dual object for F places on inputs of high Hamming weight, to inputs of low Hamming weight. The resulting dual object is supported entirely on inputs of low Hamming weight and therefore witnesses a lower bound on the approximate degree of the *restriction* F' of F to inputs of low Hamming weight.

The restriction F' takes as input $N := \Theta(nm \log m)$ variables but is defined only when its input string has Hamming weight $\tilde{O}(n)$. This makes it possible to represent the input to F' more economically, by specifying the locations of the $\tilde{O}(n)$ nonzero bits inside the array of N variables. Since each such location can be specified using $\lceil \log N \rceil$ bits, the entire input to F' can be specified using $\lceil \log N \rceil \cdot \tilde{O}(n) = \tilde{O}(n)$ bits. This yields a function F'' on $\tilde{O}(n)$ variables. A careful calculation shows that this “input compression” does not hurt the approximate degree. Thus, the approximate degree of F'' is at least the approximate degree of F' , which as discussed above is $\Omega(\sqrt{m} \deg_{1/3}(f))$. With m set appropriately, the approximate degree of F'' is polynomially larger than that of f .

This passage from f to F'' is the desired hardness amplification for approximate degree. To obtain an $\Omega(n^{1-\delta})$ lower bound on the approximate degree of AC^0 , the authors of [17] start with a trivial circuit and apply the hardness amplification step a constant number of times, until approximate degree $\Omega(n^{1-\delta})$ is reached.

Limitations of previous approaches to AC^0 . Bun and Thaler’s hardness amplification for approximate degree rests on two pillars. The first is componentwise composition, whereby the given function $f: \{0, 1\}^n \rightarrow \{0, 1\}$ is composed componentwise with n independent copies of the gadget $\text{AND}_{\Theta(\log m)} \circ \text{OR}_m$. In this gadget, the $\text{AND}_{\Theta(\log m)}$ gate is necessary to control the accumulation of error and to ensure the correlation property of the dual polynomial. The resulting composed function $F = f \circ (\text{AND}_{\Theta(\log m)} \circ \text{OR}_m)$ is defined on $N = \Theta(nm \log m)$ variables. The second pillar of [17] is input compression, where the length- N input to F is represented compactly as an array of $\tilde{O}(n)$ strings of length $\lceil \log N \rceil$ each. The circuitry to implement these two pillars is expensive, requiring in both cases a polynomial-size DNF formula of width $\Theta(\log n + \log m)$. As a result, even a *single* iteration of the Bun–Thaler hardness amplification cannot be implemented as a polynomial-size DNF or CNF formula.

To prove an $\Omega(n^{1-\delta})$ approximate degree lower bound for small $\delta > 0$ in the framework of [17], one needs a number of iterations that grows with $1/\delta$. Thus, the overall circuit produced in [17] has a large constant number of alternating layers of AND and OR gates of logarithmic and polynomial fan-in, respectively, and in particular cannot be flattened into a polynomial-size DNF or CNF formula. Proving Theorem 1.1 within this framework would require reducing the fan-in of the AND gates from $\Theta(\log n + \log m)$ to $O(1)$, which would completely destroy the componentwise composition and input compression pillars of [17]. These pillars are present in all follow-up papers [17, 14, 18, 49] and seem impossible to get around, prompting the authors of [18, p. 14] to entertain the possibility that the approximate degree of AC^0 at any given depth is much smaller than once conjectured. We show that this is not the case.

1.9. Our proof. In this paper, we design hardness amplification from first principles, without using componentwise composition or input compression. Our approach efficiently amplifies the approximate degree even for functions with sparse input, while ensuring that each hardness amplification stage is implementable by a monotone circuit of constant depth with AND gates of constant fan-in and OR gates of polynomial fan-in. As a result, repeating our process any constant number of times produces a polynomial-size DNF formula of constant width.

Our approach at a high level. Let $f: \{0, 1\}^N \rightarrow \{0, 1\}$ be a given function. Let $f|_{\leq \theta}$ denote the restriction of f to inputs of Hamming weight at most θ , and let $d = \deg_{1/3}(f|_{\leq \theta})$ be the approximate degree of this restriction. The total number of variables N can be vastly larger than θ ; in the actual proof, we will set $N = \theta^C$ for a constant $C \geq 1$. Since an input $y \in \{0, 1\}^N$ to $f|_{\leq \theta}$ is guaranteed to have Hamming weight at most θ , we can think of y as the disjunction of θ vectors of Hamming weight at most 1 each:

$$y = y_1 \vee y_2 \vee \cdots \vee y_\theta,$$

where each y_i is either the zero vector 0^N or a basis vector e_1, e_2, \dots, e_N , and the disjunction on the right-hand side is applied coordinate-wise. Our approach centers around encoding each y_i as a string of $n \ll N$ bits so as to make the decoding difficult for polynomials but easy for circuits. Ideally, we would like a decoding function $h: \{0, 1\}^n \rightarrow \{0, 1\}^N$ with the following properties:

- (i) the sets $h^{-1}(v)$ for $v \in \{e_1, e_2, \dots, e_N, 0^N\}$ are indistinguishable by polynomials of degree up to D , for some parameter D ;
- (ii) the sets $h^{-1}(v)$ for $v \in \{e_1, e_2, \dots, e_N, 0^N\}$ contain only strings of Hamming weight $O(1)$;
- (iii) h is computable by a constant-depth monotone circuit with AND gates of constant fan-in and OR gates of polynomial fan-in.

With such h in hand, define $F: (\{0, 1\}^n)^\theta \rightarrow \{0, 1\}$ by

$$F(x_1, x_2, \dots, x_\theta) = f\left(\bigvee_{i=1}^{\theta} h(x_i)\right).$$

Then, one can reasonably expect that approximating F is harder than approximating $f|_{\leq \theta}$. Indeed, an approximating polynomial has access only to the encoded input $(x_1, x_2, \dots, x_\theta)$. Decoding this input presumably involves computing $(x_1, x_2, \dots, x_\theta) \mapsto (h(x_1), h(x_2), \dots, h(x_\theta))$ one way or another, which by property (i) requires a polynomial of degree greater than D . Once the decoded string $h(x_1) \vee h(x_2) \vee \cdots \vee h(x_\theta)$ is available, the polynomial supposedly needs to compute f on that input, which in and of itself requires degree d . Altogether, we expect F to have approximate degree on the order of Dd . Moreover, property (ii) ensures that F is hard to approximate even on inputs of Hamming weight $O(\theta)$, putting us in a strong position for another round of hardness amplification. Finally, property (iii) guarantees that the result of constantly many rounds of hardness amplification is computable by a DNF formula of polynomial size and constant width.

Actual implementation. As one might suspect, the above program is too bold and cannot be implemented literally. Our actual construction of h achieves (i)–(iii) only approximately. In more detail, let k be a sufficiently large constant. For each $v \in \{e_1, e_2, \dots, e_N, 0^N\}$, we construct a probability distribution λ_v on $\{0, 1\}^n$ that has all but a vanishing fraction of its mass on inputs of Hamming weight exactly k , and moreover any two such distributions λ_v and $\lambda_{v'}$ are indistinguishable by polynomials of low degree. We are further able to ensure that an input of Hamming weight k belongs to the support of at most one of the distributions λ_v . Thus, the λ_v are in essence supported on pairwise disjoint sets of strings of Hamming weight k , and are pairwise indistinguishable by polynomials of low degree. The

decoding function h works by taking an input $x \in \{0, 1\}^n$ of Hamming weight k and determining which of the distributions has x in its support—a highly efficient computation realizable as a monotone k -DNF formula. With small probability, h will receive as input a string of Hamming weight larger than k , in which case the decoding may fail.

Construction of the λ_v . Central to our work is the number-theoretic notion of m -discrepancy, which is a measure of pseudorandomness or aperiodicity of a given set of integers modulo m . Formally, the m -discrepancy of a nonempty finite set $S \subseteq \mathbb{Z}$ is defined as

$$\text{disc}_m(S) = \max_{k=1,2,\dots,m-1} \left| \frac{1}{|S|} \sum_{s \in S} \xi^{ks} \right|,$$

where ξ is a primitive m -th root of unity. The construction of sparse sets with low discrepancy is a well-studied problem in combinatorics and theoretical computer science. By building on previous work [2, 48], we construct a sparse set of integers with small discrepancy in our regime of interest. For our application, we set the modulus $m = N + 1$.

Continuing, let $\binom{[n]}{k}$ denote the family of cardinality- k subsets of $[n] = \{1, 2, \dots, n\}$. To design the distributions λ_v , we need an explicit coloring $\gamma: \binom{[n]}{k} \rightarrow [N + 1]$ that is *balanced*, in the sense that for nearly all large enough subsets $A \subseteq \{1, 2, \dots, n\}$ and all $i \in [N + 1]$, the family $\gamma^{-1}(i)$ accounts for almost exactly a $1/(N + 1)$ fraction of all cardinality- k subsets of A . The existence of a highly balanced coloring follows by the probabilistic method, and we construct one explicitly using the sparse set of integers with small $(N + 1)$ -discrepancy constructed earlier in the proof.

Our next ingredient is a dual polynomial ω for the OR function, a staple in approximate degree lower bounds. An important property of ω is that it places a constant fraction of its ℓ_1 mass on the point 0^n . Translating ω from 0^n to a point z of slightly larger Hamming weight results in a new dual polynomial, call it ω_z . Analogous to ω , the new dual polynomial has a constant fraction of its ℓ_1 mass on z and the rest on inputs that are greater than or equal to z componentwise.

For notational convenience, let us now rename γ 's range elements $1, 2, \dots, N + 1$ to $e_1, e_2, \dots, e_N, 0^N$, respectively. For $v \in \{e_1, e_2, \dots, e_N, 0^N\}$, define Φ_v to be the average of the dual polynomials ω_z where z ranges over all characteristic vectors of the sets in $\gamma^{-1}(v)$. Being a convex combination of dual polynomials, each Φ_v is a dual object orthogonal to polynomials of low degree. Observe further that each Φ_v is supported on inputs of Hamming weight at least k , and any input of Hamming weight exactly k belongs to the support of exactly one Φ_v . For inputs x of Hamming weight greater than k , a remarkable thing happens: $\Phi_v(x)$ is almost the same for all v . We prove this by exploiting the fact that γ is highly balanced. As a result, the “common part” of the Φ_v for inputs of Hamming weight greater than k can be subtracted out to obtain a function $\widetilde{\Phi}_v$ for each $v \in \{e_1, e_2, \dots, e_N, 0^N\}$. While these new functions are not dual polynomials, the *difference* of any two of them is since $\widetilde{\Phi}_v - \widetilde{\Phi}_{v'} = \Phi_v - \Phi_{v'}$. Put another way, the $\widetilde{\Phi}_v$ are pairwise indistinguishable by low-degree polynomials. By defining the $\widetilde{\Phi}_v$ in a somewhat more subtle way, we further ensure that each $\widetilde{\Phi}_v$ is nonnegative. The distribution λ_v can then be taken to be the normalized function $\widetilde{\Phi}_v / \|\widetilde{\Phi}_v\|_1$. This construction ensures all the properties

that we need: λ_v has nearly all of its mass on inputs of Hamming weight k ; an input of Hamming weight k belongs to the support of at most one distribution λ_v ; and any pair of distributions $\lambda_v, \lambda_{v'}$ are indistinguishable by a low-degree polynomial. Observe that in our construction, λ_v is close to the uniform probability distribution on the characteristic vectors of the sets in $\gamma^{-1}(v)$.

2. PRELIMINARIES

2.1. General notation. For a string $x \in \{0, 1\}^n$ and a set $S \subseteq \{1, 2, \dots, n\}$, we let $x|_S$ denote the restriction of x to the indices in S . In other words, $x|_S = x_{i_1}x_{i_2}\dots x_{i_{|S|}}$, where $i_1 < i_2 < \dots < i_{|S|}$ are the elements of S . The *characteristic vector* $\mathbf{1}_S$ of a set $S \subseteq \{1, 2, \dots, n\}$ is given by

$$(\mathbf{1}_S)_i = \begin{cases} 1 & \text{if } i \in S, \\ 0 & \text{otherwise.} \end{cases}$$

Given an arbitrary set X and elements $x, y \in X$, the Kronecker delta $\delta_{x,y}$ is defined by

$$\delta_{x,y} = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{otherwise.} \end{cases}$$

For a logical condition C , we use the Iverson bracket

$$\mathbf{I}[C] = \begin{cases} 1 & \text{if } C \text{ holds,} \\ 0 & \text{otherwise.} \end{cases}$$

We let $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ denote the set of natural numbers. We use the comparison operators in a unary capacity to denote one-sided intervals of the real line. Thus, $<a, \leq a, >a, \geq a$ stand for $(-\infty, a), (-\infty, a], (a, \infty), [a, \infty)$, respectively. We let $\ln x$ and $\log x$ stand for the natural logarithm of x and the logarithm of x to base 2, respectively. The term *Euclidean space* refers to \mathbb{R}^n for some positive integer n . We let e_i denote the vector whose i -th component is 1 and the others are 0. Thus, the vectors e_1, e_2, \dots, e_n form the standard basis for \mathbb{R}^n . For a complex number x , we denote the real part, imaginary part, and complex conjugate of x as usual by $\operatorname{Re}(x)$, $\operatorname{Im}(x)$, and \bar{x} , respectively. We typeset the imaginary unit \mathbf{i} in boldface to distinguish it from the index variable i . For an arbitrary integer a and a positive integer m , recall that $a \bmod m$ denotes the unique element of $\{0, 1, 2, \dots, m-1\}$ that is congruent to a modulo m .

For a set X , we let \mathbb{R}^X denote the linear space of real-valued functions on X . The *support* of a function $f \in \mathbb{R}^X$ is denoted $\operatorname{supp} f = \{x \in X : f(x) \neq 0\}$. For real-valued functions with finite support, we adopt the usual norms and inner product:

$$\begin{aligned} \|f\|_\infty &= \max_{x \in \operatorname{supp} f} |f(x)|, \\ \|f\|_1 &= \sum_{x \in \operatorname{supp} f} |f(x)|, \\ \langle f, g \rangle &= \sum_{x \in \operatorname{supp} f \cap \operatorname{supp} g} f(x)g(x). \end{aligned}$$

This covers as a special case functions on finite sets. Analogous to functions, we adopt the familiar norms for vectors $x \in \mathbb{R}^n$ in Euclidean space: $\|x\|_\infty = \max_{i=1, \dots, n} |x_i|$ and $\|x\|_1 = \sum_{i=1}^n |x_i|$. The *tensor product* of $f \in \mathbb{R}^X$ and $g \in \mathbb{R}^Y$ is denoted $f \otimes g \in \mathbb{R}^{X \times Y}$ and given by $(f \otimes g)(x, y) = f(x)g(y)$. The tensor product $f \otimes f \otimes \dots \otimes f$ (n times) is abbreviated $f^{\otimes n}$. We frequently omit the argument in equations and inequalities involving functions, as in $\text{sgn } p = (-1)^f$. Such statements are to be interpreted pointwise. For example, the statement “ $f \geq 2|g|$ on X ” means that $f(x) \geq 2|g(x)|$ for every $x \in X$. For vectors x and y , the notation $x \leq y$ means that $x_i \leq y_i$ for each i .

We adopt the standard notation for function composition, with $f \circ g$ defined by $(f \circ g)(x) = f(g(x))$. In addition, we use the \circ operator to denote the *componentwise composition* of Boolean functions. Formally, the componentwise composition of $f: \{0, 1\}^n \rightarrow \{0, 1\}$ and $g: X \rightarrow \{0, 1\}$ is the function $f \circ g: X^n \rightarrow \{0, 1\}$ given by $(f \circ g)(x_1, x_2, \dots, x_n) = f(g(x_1), g(x_2), \dots, g(x_n))$. Componentwise composition is consistent with standard composition, which in the context of Boolean functions is only defined for $n = 1$. Thus, the meaning of $f \circ g$ is determined by the range of g and is never in doubt.

For a natural number n , we abbreviate $[n] = \{1, 2, \dots, n\}$. For a set S and an integer k , we let $\binom{S}{k}$ stand for the family of cardinality- k subsets of S :

$$\binom{S}{k} = \{A \subseteq S : |A| = k\}.$$

Analogously, for any set I , we define

$$\binom{S}{I} = \{A \subseteq S : |A| \in I\}.$$

To illustrate, $\binom{S}{\leq k}$ denotes the family of subsets of S that have cardinality at most k . Analogously, we have the symbols $\binom{S}{< k}$, $\binom{S}{\geq k}$, $\binom{S}{> k}$. Throughout this manuscript, we use brace notation as in $\{z_1, z_2, \dots, z_n\}$ to specify *multisets* rather than *sets*, the distinction being that the number of times an element occurs is taken into account. The *cardinality* $|Z|$ of a finite multiset Z is defined to be the total number of element occurrences in Z , with each element counted as many times as it occurs. The equality and subset relations on multisets are defined analogously, with the number of element occurrences taken into account. For example, $\{1, 1, 2\} = \{1, 2, 1\}$ but $\{1, 1, 2\} \neq \{1, 2\}$. Similarly, $\{1, 2\} \subseteq \{1, 1, 2\}$ but $\{1, 1, 2\} \not\subseteq \{1, 2\}$.

2.2. Boolean strings and functions. We identify the Boolean values “true” and “false” with 1 and 0, respectively, and view Boolean functions as mappings $X \rightarrow \{0, 1\}$ for a finite set X . The familiar functions $\text{OR}_n: \{0, 1\}^n \rightarrow \{0, 1\}$ and $\text{AND}_n: \{0, 1\}^n \rightarrow \{0, 1\}$ are given by $\text{OR}_n(x) = \bigvee_{i=1}^n x_i$ and $\text{AND}_n(x) = \bigwedge_{i=1}^n x_i$. We abbreviate $\text{NOR}_n = \neg \text{OR}_n$. For Boolean strings $x, y \in \{0, 1\}^n$, we let $x \oplus y$ denote their bitwise XOR. The strings $x \wedge y$ and $x \vee y$ are defined analogously, with the binary operator applied bitwise.

For a vector $v \in \mathbb{N}^n$, we define its *weight* $|v|$ to be $|v| = v_1 + v_2 + \dots + v_n$. If $x \in \{0, 1\}^n$ is a Boolean string, then $|x|$ is precisely the Hamming weight of x . For any sets $X \subseteq \mathbb{N}^n$ and $W \subseteq \mathbb{R}$, we define $X|_W$ to be the subset of vectors in X whose weight belongs to W :

$$X|_W = \{x \in X : |x| \in W\}.$$

In the case of a one-element set $W = \{w\}$, we further shorten $X|_{\{w\}}$ to $X|_w$. For example, $\mathbb{N}^n|_{\leq w}$ denotes the set of vectors whose n components are natural numbers and sum to at most w , whereas $\{0, 1\}^n|_w$ denotes the set of Boolean strings of length n and Hamming weight exactly w . For a function $f: X \rightarrow \mathbb{R}$ on a subset $X \subseteq \{0, 1\}^n$, we let $f|_W$ denote the restriction of f to $X|_W$. Thus, $f|_W$ is a function with domain $X|_W$ given by $f|_W(x) = f(x)$. A typical instance of this notation would be $f|_{\leq w}$ for some real number w , corresponding to the restriction of f to Boolean strings of Hamming weight at most w .

2.3. Concentration of measure. Throughout this manuscript, we view probability distributions as real functions. This convention makes available the shorthand notation introduced above. In particular, for probability distributions μ and λ , the symbol $\text{supp } \mu$ denotes the support of μ , and $\mu \otimes \lambda$ denotes the probability distribution given by $(\mu \otimes \lambda)(x, y) = \mu(x)\lambda(y)$. We use the notation $\mu \times \lambda$ interchangeably with $\mu \otimes \lambda$, the former being more standard for probability distributions. If μ is a probability distribution on X , we consider μ to be defined also on any superset of X with the understanding that $\mu = 0$ outside X .

We recall the following multiplicative form of the Chernoff bound [21].

THEOREM 2.1 (Chernoff bound). *Let $X_1, X_2, \dots, X_n \in \{0, 1\}$ be i.i.d. random variables with $\mathbf{E} X_i = p$. Then for all $0 \leq \delta \leq 1$,*

$$\mathbf{P} \left[\left| \sum_{i=1}^n X_i - pn \right| \geq \delta pn \right] \leq 2 \exp \left(-\frac{\delta^2 pn}{3} \right).$$

Theorem 2.1 assumes i.i.d. Bernoulli random variables. Hoeffding's inequality [24], stated next, is a more general concentration-of-measure result that applies to any independent bounded random variables.

THEOREM 2.2 (Hoeffding's inequality). *Let X_1, X_2, \dots, X_n be independent random variables with $X_i \in [a_i, b_i]$. Define $p = \sum_{i=1}^n \mathbf{E} X_i$. Then for all $\delta \geq 0$,*

$$\mathbf{P} \left[\left| \sum_{i=1}^n X_i - p \right| \geq \delta \right] \leq 2 \exp \left(-\frac{2\delta^2}{\sum_{i=1}^n (b_i - a_i)^2} \right).$$

The standard version of Hoeffding's inequality, stated above, requires X_1, X_2, \dots, X_n to be independent. Less known are Hoeffding's results for dependent random variables, which he obtained along with Theorem 2.2 in his original paper [24]. We will specifically need the following concentration inequality for sampling without replacement [24, Section 6].

THEOREM 2.3 (Hoeffding's sampling without replacement). *Let $\omega_1, \omega_2, \dots, \omega_N$ be given reals, with $\omega_i \in [a, b]$ for all i . Let $J_1, J_2, \dots, J_n \in [N]$ be uniformly random integers that are pairwise distinct. Let $X_i = \omega_{J_i}$ for $i = 1, 2, \dots, n$, and define $p = \sum_{i=1}^n \mathbf{E} X_i$. Then for all $\delta \geq 0$,*

$$\mathbf{P} \left[\left| \sum_{i=1}^n X_i - p \right| \geq \delta \right] \leq 2 \exp \left(-\frac{2\delta^2}{n(b-a)^2} \right).$$

Hoeffding's two theorems are clearly incomparable. On the one hand, Theorem 2.2 requires independence and therefore does not apply to sampling without replacement. On the other hand, each random variable X_i in Theorem 2.3 must be uniformly distributed on a finite multiset of values, which must further be the same multiset for all X_i ; none of this is assumed in Theorem 2.2.

Finally, we will need a concentration-of-measure result due to Bun and Thaler [17, Lemma 4.7] for product distributions on \mathbb{N}^n .

LEMMA 2.4 (cf. Bun and Thaler). *Let $\lambda_1, \lambda_2, \dots, \lambda_n$ be distributions on \mathbb{N} with finite support such that*

$$\lambda_i(t) \leq \frac{C\alpha^t}{(t+1)^2}, \quad t \in \mathbb{N},$$

where $C \geq 0$ and $0 \leq \alpha \leq 1$. Then for all $T \geq 8Cen(1 + \ln n)$,

$$\mathbf{P}_{v \sim \lambda_1 \times \lambda_2 \times \dots \times \lambda_n} [\|v\|_1 \geq T] \leq \alpha^{T/2}.$$

Bun and Thaler's result in [17, Lemma 4.7] differs slightly from the statement above. The proof of Lemma 2.4 as stated can be found in [49, Lemma 3.6]. By leveraging Lemma 2.4, we obtain the following concentration result for probability distributions that are supported on the Boolean hypercube, rather than \mathbb{N} , and are shifted from the origin.

LEMMA 2.5. *Fix integers $B \geq k \geq 0$. Let $\lambda_1, \lambda_2, \dots, \lambda_\ell$ be probability distributions on $\{0, 1\}^B$ with support contained in $\{0, 1\}^B|_{\geq k}$. Suppose further that*

$$\lambda_i(\{0, 1\}^B|_t) \leq \frac{C\alpha^{t-k}}{(t-k+1)^2}, \quad i \in [\ell], \quad t \in \{k, k+1, \dots, B\},$$

where $C \geq 0$ and $0 \leq \alpha \leq 1$. Then for all $T \geq 8C\ell(1 + \ln \ell) + \ell k$,

$$\mathbf{P}_{(x_1, \dots, x_\ell) \sim \lambda_1 \times \dots \times \lambda_\ell} \left[\sum_{i=1}^{\ell} |x_i| \geq T \right] \leq \alpha^{(T-\ell k)/2}.$$

Proof. For $i = 1, 2, \dots, \ell$, consider the distribution μ_i on $\{0, 1, \dots, B-k\}$ given by $\mu_i(t) = \lambda_i(\{0, 1\}^B|_{t+k})$. Then

$$\mu_i(t) \leq \frac{C\alpha^t}{(t+1)^2}, \quad i \in [\ell], \quad t \geq 0. \quad (2.1)$$

Moreover, the random variable $|x_i|$ with $x_i \sim \lambda_i$ has the same distribution as the random variable $u_i + k$ for $u_i \sim \mu_i$. As a result,

$$\begin{aligned} \mathbf{P}_{(x_1, \dots, x_\ell) \sim \lambda_1 \times \dots \times \lambda_\ell} \left[\sum_{i=1}^{\ell} |x_i| \geq T \right] &= \mathbf{P}_{u \sim \mu_1 \times \mu_2 \times \dots \times \mu_\ell} \left[\sum_{i=1}^{\ell} (u_i + k) \geq T \right] \\ &= \mathbf{P}_{u \sim \mu_1 \times \mu_2 \times \dots \times \mu_\ell} [\|u\|_1 \geq T - k\ell] \\ &\leq \alpha^{(T-\ell k)/2}, \end{aligned}$$

where the last step uses Lemma 2.4 along with (2.1) and the hypothesis that $T \geq 8C\ell(1 + \ln \ell) + \ell k$. \square

2.4. Orthogonal content. For a multivariate polynomial $p: \mathbb{R}^n \rightarrow \mathbb{R}$, we let $\deg p$ denote the total degree of p , i.e., the largest degree of any monomial of p . We use the terms *degree* and *total degree* interchangeably in this paper. It will be convenient to define the degree of the zero polynomial by $\deg 0 = -\infty$. For a real-valued function ϕ supported on a finite subset of \mathbb{R}^n , the *orthogonal content* of ϕ , denoted $\text{orth } \phi$, is the minimum degree of a real polynomial p for which $\langle \phi, p \rangle \neq 0$. We adopt the convention that $\text{orth } \phi = \infty$ if no such polynomial exists. It is clear that $\text{orth } \phi \in \mathbb{N} \cup \{\infty\}$, with the extremal cases $\text{orth } \phi = 0 \Leftrightarrow \langle \phi, 1 \rangle \neq 0$ and $\text{orth } \phi = \infty \Leftrightarrow \phi = 0$. Additional facts about orthogonal content are given by the following two propositions.

PROPOSITION 2.6. *Let X and Y be nonempty finite subsets of Euclidean space. Then:*

- (i) $\text{orth}(\phi + \psi) \geq \min\{\text{orth } \phi, \text{orth } \psi\}$ for all $\phi, \psi: X \rightarrow \mathbb{R}$;
- (ii) $\text{orth}(\phi \otimes \psi) = \text{orth}(\phi) + \text{orth}(\psi)$ for all $\phi: X \rightarrow \mathbb{R}$ and $\psi: Y \rightarrow \mathbb{R}$.

A proof of Proposition 2.6 can be found in [49, Proposition 2.1].

PROPOSITION 2.7. *Define $V = \{0^N, e_1, e_2, \dots, e_N\} \subseteq \mathbb{R}^N$. Fix functions $\phi_v: X \rightarrow \mathbb{R}$ ($v \in V$), where X is a finite subset of Euclidean space. Suppose that*

$$\text{orth}(\phi_u - \phi_v) \geq D, \quad u, v \in V, \quad (2.2)$$

where D is a positive integer. Then for every polynomial $p: X^\ell \rightarrow \mathbb{R}$, the mapping $z \mapsto \langle \bigotimes_{i=1}^\ell \phi_{z_i}, p \rangle$ is a polynomial on V^ℓ of degree at most $(\deg p)/D$.

Proof. By linearity, it suffices to consider factored polynomials $p(x_1, \dots, x_\ell) = \prod_{i=1}^\ell p_i(x_i)$, where each p_i is a nonzero polynomial on X . In this setting,

$$\left\langle \bigotimes_{i=1}^\ell \phi_{z_i}, p \right\rangle = \prod_{i=1}^\ell \langle \phi_{z_i}, p_i \rangle. \quad (2.3)$$

By (2.2), we have $\langle \phi_{0^N}, p_i \rangle = \langle \phi_{e_1}, p_i \rangle = \langle \phi_{e_2}, p_i \rangle = \dots = \langle \phi_{e_N}, p_i \rangle$ for any index i with $\deg p_i < D$. As a result, polynomials p_i with $\deg p_i < D$ do not contribute to the degree of the right-hand side of (2.3) as a function of z . For the other polynomials p_i , the inner product $\langle \phi_{z_i}, p_i \rangle$ is a linear polynomial in z_i , namely,

$$\begin{aligned} \langle \phi_{z_i}, p_i \rangle &= z_{i,1} \langle \phi_{e_1}, p_i \rangle + z_{i,2} \langle \phi_{e_2}, p_i \rangle + \dots + z_{i,N} \langle \phi_{e_N}, p_i \rangle \\ &\quad + \left(1 - \sum_{j=1}^N z_{i,j} \right) \langle \phi_{0^N}, p_i \rangle. \end{aligned}$$

Thus, polynomials p_i with $\deg p_i \geq D$ contribute at most 1 each to the degree. Summarizing, the right-hand side of (2.3) is a real polynomial in z_1, z_2, \dots, z_ℓ of degree at most $|\{i : \deg p_i \geq D\}| \leq \frac{\deg p}{D}$. \square

Proposition 2.7 generalizes an analogous result in [49, Proposition 2.2], where the special case $N = 1$ was treated.

2.5. Polynomial approximation. For a real number $\epsilon \geq 0$ and a function $f: X \rightarrow \mathbb{R}$ on a finite subset X of Euclidean space, the ϵ -approximate degree of f is denoted $\deg_\epsilon(f)$ and is defined to be the minimum degree of a polynomial p such that $\|f - p\|_\infty \leq \epsilon$. For $\epsilon < 0$, it will be convenient to define $\deg_\epsilon(f) = +\infty$ since no polynomial satisfies $\|f - p\|_\infty \leq \epsilon$ in this case. We focus on the approximate degree of Boolean functions $f: X \rightarrow \{0, 1\}$. In this setting, the standard choice of the error parameter is $\epsilon = 1/3$. This choice is without loss of generality since $\deg_\epsilon(f) = \Theta(\deg_{1/3}(f))$ for every Boolean function f and every constant $0 < \epsilon < 1/2$. In what follows, we refer to $1/3$ -approximate degree simply as “approximate degree.” The notion of approximate degree has the following dual characterization [38, 39].

FACT 2.8. *Let $f: X \rightarrow \mathbb{R}$ be given, for a finite set $X \subset \mathbb{R}^n$. Let $d \geq 0$ be an integer and $\epsilon \geq 0$ a real number. Then $\deg_\epsilon(f) \geq d$ if and only if there exists a function $\psi: X \rightarrow \mathbb{R}$ such that*

$$\begin{aligned} \langle f, \psi \rangle &> \epsilon \|\psi\|_1, \\ \text{orth } \psi &\geq d. \end{aligned}$$

This characterization of approximate degree can be verified using linear programming duality, cf. [38, 39]. We now recall a variant of approximate degree for one-sided approximation. For a Boolean function $f: X \rightarrow \{0, 1\}$ and $\epsilon \geq 0$, the *one-sided ϵ -approximate degree* of f is denoted $\deg_\epsilon^+(f)$ and defined to be the minimum degree of a real polynomial p such that

$$\begin{aligned} f(x) - \epsilon &\leq p(x) \leq f(x) + \epsilon, & x \in f^{-1}(0), \\ f(x) - \epsilon &\leq p(x), & x \in f^{-1}(1). \end{aligned}$$

We refer to any such polynomial as a *one-sided approximant* for f with error ϵ . As usual, the canonical setting of the error parameter is $\epsilon = 1/3$. In the pathological case $\epsilon < 0$, it will be convenient to define $\deg_\epsilon^+(f) = +\infty$. Observe the asymmetric treatment of $f^{-1}(0)$ and $f^{-1}(1)$ in this formalism. In particular, the one-sided approximate degree of Boolean functions is in general not invariant under negation. One-sided approximate degree enjoys the following dual characterization [16].

FACT 2.9. *Let $f: X \rightarrow \{0, 1\}$ be given, for a finite set $X \subset \mathbb{R}^n$. Let $d \geq 0$ be an integer and $\epsilon \geq 0$ a real number. Then $\deg_\epsilon^+(f) \geq d$ if and only if there exists a function $\psi: X \rightarrow \mathbb{R}$ such that*

$$\begin{aligned} \langle f, \psi \rangle &> \epsilon \|\psi\|_1, \\ \text{orth } \psi &\geq d, \\ \psi(x) &\geq 0 \text{ whenever } f(x) = 1. \end{aligned}$$

2.6. Dual polynomials. Facts 2.8 and 2.9 make it possible to prove lower bounds on approximate degree in a constructive manner, by exhibiting a dual object ψ that serves as a witness. This object is referred to as a *dual polynomial*. Often, a dual polynomial for a composed function f can be constructed by combining dual objects for various components of f . Of particular importance in the study of AC^0 is the dual object for the OR function. The first dual polynomial for OR was constructed

by Špalek [50], with many refinements and generalizations obtained in follow-up work [15, 45, 46, 17, 14, 49]. We will use the following construction from [49, Lemma B.2].

LEMMA 2.10. *Let ϵ be given, $0 < \epsilon < 1$. Then for some constant $c = c(\epsilon) \in (0, 1)$ and every integer $n \geq 1$, there is an (explicitly given) function $\omega: \{0, 1, 2, \dots, n\} \rightarrow \mathbb{R}$ such that*

$$\begin{aligned} \omega(0) &> \frac{1-\epsilon}{2} \cdot \|\omega\|_1, \\ |\omega(t)| &\leq \frac{1}{ct^2 2^{ct/\sqrt{n}}} \cdot \|\omega\|_1 && (t = 1, 2, \dots, n), \\ (-1)^t \omega(t) &\geq 0 && (t = 0, 1, 2, \dots, n), \\ \text{orth } \omega &\geq c\sqrt{n}. \end{aligned}$$

A useful tool in the construction of dual polynomials is the following lemma due to Razborov and Sherstov [34].

LEMMA 2.11 (Razborov and Sherstov). *Fix integers D and n , where $0 \leq D < n$. Then there is an (explicitly given) function $\zeta: \{0, 1\}^n \rightarrow \mathbb{R}$ such that*

$$\text{supp } \zeta \subseteq \{0, 1\}^n|_{\leq D} \cup \{1^n\}, \quad (2.4)$$

$$\zeta(1^n) = 1, \quad (2.5)$$

$$\|\zeta\|_1 \leq 1 + 2^D \binom{n}{D}, \quad (2.6)$$

$$\text{orth } \zeta > D. \quad (2.7)$$

In more detail, this result corresponds to taking $k = D$ and $\zeta = (-1)^n g$ in the proof of Lemma 3.2 of [34]. We will need the following natural generalization of Lemma 2.11.

LEMMA 2.12. *Fix integers D and B , where $0 \leq D < B$. Let $y \in \{0, 1\}^B$ be a string with $|y| > D$. Then there is an (explicitly given) function $\zeta_y: \{0, 1\}^B \rightarrow \mathbb{R}$ such that*

$$\text{supp } \zeta_y \subseteq \{x : x \leq y \text{ and } |x| \leq D\} \cup \{y\}, \quad (2.8)$$

$$\zeta_y(y) = 1, \quad (2.9)$$

$$\|\zeta_y\|_1 \leq 1 + 2^D \binom{B}{D}, \quad (2.10)$$

$$\text{orth } \zeta_y > D. \quad (2.11)$$

Proof. Set $n = |y|$. Lemma 2.11 gives an explicit function $\zeta: \{0, 1\}^n \rightarrow \mathbb{R}$ that satisfies (2.4)–(2.7). Define $\zeta_y: \{0, 1\}^B \rightarrow \mathbb{R}$ by

$$\zeta_y(x) = \zeta(x|_S) \prod_{i \notin S} (1 - x_i),$$

where $S = \{i : y_i = 1\}$. Then (2.9) and (2.10) are immediate from (2.5) and (2.6), respectively. Property (2.11) follows from (2.7) in light of Proposition 2.6 (ii). To verify the remaining property (2.8), fix any input x with $\zeta_y(x) \neq 0$. Then the

definition of ζ_y implies that $x|_{\bar{y}}$ is the zero vector, whereas (2.4) implies that $x|_S$ is either 1^n or a string of Hamming weight at most D . In the former case, we have $x = y$; in the latter case, $x \leq y$ and $|x| \leq D$. \square

Informally, Lemmas 2.11 and 2.12 are useful when one needs to adjust a dual object's metric properties while preserving its orthogonality to low-degree polynomials. These lemmas play a basic role in several recent papers [34, 17, 14, 18, 49] as well as our work. For the reader's benefit, we encapsulate this procedure as Lemma 2.13 below and provide a detailed proof.

LEMMA 2.13. *Let $\Phi: \{0, 1\}^B \rightarrow \mathbb{R}$ be given. Fix integers $T \geq D \geq 0$. Then there is an (explicitly given) function $\tilde{\Phi}: \{0, 1\}^B \rightarrow \mathbb{R}$ such that*

$$\text{supp } \tilde{\Phi} \subseteq \{0, 1\}^B|_{\leq T}, \quad (2.12)$$

$$\text{orth}(\Phi - \tilde{\Phi}) > D, \quad (2.13)$$

$$\|\Phi - \tilde{\Phi}\|_1 \leq \left(1 + 2^D \binom{B}{D}\right) \sum_{x:|x|>T} |\Phi(x)|. \quad (2.14)$$

Proof (adapted from [34, 17, 14, 18, 49]). For $T \geq B$, the lemma holds trivially with $\tilde{\Phi} = \Phi$. In what follows, we treat the complementary case $T < B$.

For each $y \in \{0, 1\}^B|_{>T}$, Lemma 2.12 constructs a function $\zeta_y: \{0, 1\}^B \rightarrow \mathbb{R}$ that obeys (2.8)–(2.11). Define

$$\tilde{\Phi} = \Phi - \sum_{y \in \{0, 1\}^B|_{>T}} \Phi(y) \zeta_y.$$

Then for $x \in \{0, 1\}^B|_{>T}$, properties (2.8) and (2.9) force $\zeta_y(x) = \delta_{x,y}$ and consequently $\tilde{\Phi}(x) = \Phi(x) - \Phi(x) = 0$. This settles (2.12). Property (2.13) is justified by

$$\text{orth}(\Phi - \tilde{\Phi}) = \text{orth} \left(\sum_{y \in \{0, 1\}^B|_{>T}} \Phi(y) \zeta_y \right) \geq \min_{y \in \{0, 1\}^B|_{>T}} \text{orth } \zeta_y > D,$$

where the last two steps use Proposition 2.6(i) and (2.11), respectively. The final property (2.14) can be derived as follows:

$$\begin{aligned} \|\Phi - \tilde{\Phi}\|_1 &= \left\| \sum_{y \in \{0, 1\}^B|_{>T}} \Phi(y) \zeta_y \right\|_1 \\ &\leq \sum_{y \in \{0, 1\}^B|_{>T}} |\Phi(y)| \|\zeta_y\|_1 \\ &\leq \left(1 + 2^D \binom{B}{D}\right) \sum_{y \in \{0, 1\}^B|_{>T}} |\Phi(y)|, \end{aligned}$$

where the last two steps use the triangle inequality and (2.10), respectively. \square

2.7. Symmetrization. Let S_n denote the symmetric group on n elements. For a permutation $\sigma \in S_n$ and an arbitrary sequence $x = (x_1, x_2, \dots, x_n)$, we adopt the shorthand $\sigma x = (x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)})$. A function $f(x_1, x_2, \dots, x_n)$ is called *symmetric* if it is invariant under permutation of the input variables: $f(x_1, x_2, \dots, x_n) = f(x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(n)})$ for all x and σ . Symmetric functions on $\{0, 1\}^n$ are intimately related to univariate polynomials, as was first observed by Minsky and Papert in their *symmetrization argument* [30].

PROPOSITION 2.14 (Minsky and Papert). *Let $p: \mathbb{R}^n \rightarrow \mathbb{R}$ be a given polynomial. Then the mapping*

$$t \mapsto \mathbf{E}_{x \in \{0,1\}^n | t} p(x)$$

is a univariate polynomial on $\{0, 1, 2, \dots, n\}$ of degree at most $\deg p$.

The next result, proved in [49, Corollary 2.13], generalizes Minsky and Papert's symmetrization to the setting when x_1, x_2, \dots, x_n are vectors rather than bits.

FACT 2.15 (Sherstov and Wu). *Let $p: (\mathbb{R}^N)^\theta \rightarrow \mathbb{R}$ be a given polynomial. Then the mapping*

$$v \mapsto \mathbf{E}_{\substack{x \in \{0^N, e_1, e_2, \dots, e_N\}^\theta: \\ x_1 + x_2 + \dots + x_\theta = v}} p(x) \tag{2.15}$$

is a polynomial on $\mathbb{N}^N |_{\leq \theta}$ of degree at most $\deg p$.

Minsky and Papert's symmetrization corresponds to $N = 1$ in Fact 2.15.

2.8. Number theory. For positive integers a and b that are relatively prime, we let $(a^{-1})_b \in \{1, 2, \dots, b-1\}$ denote the multiplicative inverse of a modulo b . The following fact is well-known and straightforward to verify; see, e.g., [48, Fact 2.8].

FACT 2.16. *For any positive integers a and b that are relatively prime,*

$$\frac{(a^{-1})_b}{b} + \frac{(b^{-1})_a}{a} - \frac{1}{ab} \in \mathbb{Z}.$$

The *prime counting function* $\pi(x)$ for a real argument $x \geq 0$ evaluates to the number of prime numbers less than or equal to x . In this manuscript, it will be clear from the context whether π refers to 3.14159... or the prime counting function. The asymptotic growth of the latter is given by the *prime number theorem*, which states that $\pi(n) \sim n/\ln n$. The following explicit bound on $\pi(n)$ is due to Rosser [35].

FACT 2.17 (Rosser). *For $n \geq 55$,*

$$\frac{n}{\ln n + 2} < \pi(n) < \frac{n}{\ln n - 4}.$$

The number of distinct prime divisors of a natural number n is denoted $\nu(n)$. The following first-principles bound on $\nu(n)$ is asymptotically tight for infinitely many n ; see [48, Fact 2.11] for details.

FACT 2.18. *The number of distinct prime divisors of n obeys*

$$(\nu(n) + 1)! \leq n.$$

In particular,

$$\nu(n) \leq (1 + o(1)) \frac{\ln n}{\ln \ln n}.$$

3. BALANCED COLORINGS

For integers $n \geq k \geq 1$ and $r \geq 1$, consider a mapping $\gamma: \binom{[n]}{k} \rightarrow [r]$. We refer to any such γ as a *coloring of $\binom{[n]}{k}$ with r colors*. An important ingredient in our work is the construction of a *balanced* coloring, in the following technical sense.

DEFINITION 3.1. Let $\gamma: \binom{[n]}{k} \rightarrow [r]$ be a given coloring. For a subset $A \subseteq [n]$, we say that γ is ϵ -balanced on A iff for each $i \in [r]$,

$$\frac{1 - \epsilon}{r} \binom{|A|}{k} \leq \left| \gamma^{-1}(i) \cap \binom{A}{k} \right| \leq \frac{1 + \epsilon}{r} \binom{|A|}{k}.$$

We define γ to be (ϵ, δ, m) -balanced iff

$$\mathbf{P}_{A \in \binom{[n]}{\ell}} [\gamma \text{ is } \epsilon\text{-balanced on } A] \geq 1 - \delta$$

for all $\ell \in \{m, m + 1, \dots, n\}$.

As one might expect, a uniformly random coloring is balanced with high probability; we establish this fact in Section 3.1. In Sections 3.2–3.5 that follow, we construct a highly balanced coloring based on an integer set with low discrepancy. The reader who is interested only in the quantitative aspect of our theorems and is not concerned about explicitness, may read Section 3.1 and skip without loss of continuity to Section 4.

3.1. Existence of balanced colorings. The next lemma uses the probabilistic method to establish the existence of balanced colorings with excellent parameters.

LEMMA 3.2. *Let $\epsilon, \delta \in (0, 1]$ be given. Let n, m, k, r be positive integers with $n \geq m \geq k$ and*

$$\binom{m}{k} \geq \frac{3r}{\epsilon^2} \cdot \ln \frac{2rn}{\delta}. \quad (3.1)$$

Then there exists an (ϵ, δ, m) -balanced coloring $\gamma: \binom{[n]}{k} \rightarrow [r]$.

Proof. Let $\gamma: \binom{[n]}{k} \rightarrow [r]$ be a uniformly random coloring. For fixed i and $A \in \binom{[n]}{\geq m}$, the cardinality $|\gamma^{-1}(i) \cap \binom{A}{k}|$ is the sum of $\binom{|A|}{k}$ independent Bernoulli random

variables, each with expected value $1/r$. As a result,

$$\begin{aligned}
& \mathbf{P}_\gamma[\gamma \text{ is not } \epsilon\text{-balanced on } A] \\
&= \mathbf{P}_\gamma \left[\max_{i \in [r]} \left| \left| \gamma^{-1}(i) \cap \binom{A}{k} \right| - \frac{1}{r} \binom{|A|}{k} \right| > \frac{\epsilon}{r} \binom{|A|}{k} \right] \\
&\leq r \max_{i \in [r]} \mathbf{P}_\gamma \left[\left| \left| \gamma^{-1}(i) \cap \binom{A}{k} \right| - \frac{1}{r} \binom{|A|}{k} \right| > \frac{\epsilon}{r} \binom{|A|}{k} \right] \\
&\leq r \cdot 2 \exp \left(-\frac{\epsilon^2}{3r} \binom{|A|}{k} \right) \\
&\leq r \cdot 2 \exp \left(-\frac{\epsilon^2}{3r} \binom{m}{k} \right) \\
&\leq \frac{\delta}{n}, \tag{3.2}
\end{aligned}$$

where the second step applies the union bound over $i \in [r]$, the third step uses the Chernoff bound (Theorem 2.1), and the fifth step uses (3.1). Now

$$\begin{aligned}
& \mathbf{E}_\gamma \max_{\ell \in \{m, m+1, \dots, n\}} \mathbf{P}_{A \in \binom{[n]}{\ell}} [\gamma \text{ is not } \epsilon\text{-balanced on } A] \\
&\leq \mathbf{E}_\gamma \sum_{\ell=m}^n \mathbf{P}_{A \in \binom{[n]}{\ell}} [\gamma \text{ is not } \epsilon\text{-balanced on } A] \\
&= \sum_{\ell=m}^n \mathbf{E}_{A \in \binom{[n]}{\ell}} \mathbf{P}_\gamma [\gamma \text{ is not } \epsilon\text{-balanced on } A] \\
&\leq \sum_{\ell=m}^n \frac{\delta}{n} \\
&\leq \delta,
\end{aligned}$$

where the next-to-last step uses (3.2). We conclude that there exists a coloring γ with

$$\max_{\ell \in \{m, m+1, \dots, n\}} \mathbf{P}_{A \in \binom{[n]}{\ell}} [\gamma \text{ is not } \epsilon\text{-balanced on } A] \leq \delta,$$

which is the definition of an (ϵ, δ, m) -balanced coloring. \square

For our purposes, the following consequence of Lemma 3.2 will be sufficient.

COROLLARY 3.3. *Let n, m, k, r be positive integers with $n \geq m \geq k^2$. Let $\epsilon \in (0, 1]$ be given with*

$$\epsilon \geq \frac{3r\sqrt{k \ln(n+1)}}{m^{k/4}}.$$

Then there exists an (ϵ, ϵ, m) -balanced coloring $\gamma: \binom{[n]}{k} \rightarrow [r]$.

Proof. We have

$$\begin{aligned}
\frac{3r}{\epsilon^2} \cdot \ln \frac{2rn}{\epsilon} &\leq \frac{3rm^{k/2}}{9r^2k \ln(n+1)} \cdot \ln \left(\frac{2rn}{3r\sqrt{k \ln(n+1)}} \cdot m^{k/4} \right) \\
&\leq \frac{m^{k/2}}{3rk \ln(n+1)} \cdot \ln(n \cdot m^{k/4}) \\
&\leq \frac{m^{k/2}}{3rk \ln(n+1)} \cdot 2k \ln n \\
&\leq m^{k/2} \\
&\leq \left(\frac{m}{k}\right)^k \\
&\leq \binom{m}{k},
\end{aligned}$$

where the next-to-last step uses the hypothesis $m \geq k^2$. By Lemma 3.2, we conclude that there is an (ϵ, ϵ, m) -balanced coloring $\gamma: \binom{[n]}{k} \rightarrow [r]$. \square

In Sections 3.2–3.5 below, we will give an explicit coloring with parameters essentially matching Corollary 3.3.

3.2. Discrepancy defined. Discrepancy is a measure of pseudorandomness or aperiodicity of a multiset of integers with respect to a given modulus M . Formally, let $M \geq 2$ be a given integer. The M -discrepancy of a nonempty multiset $Z = \{z_1, z_2, \dots, z_n\}$ of arbitrary integers is defined as

$$\text{disc}_M(Z) = \max_{k=1,2,\dots,M-1} \left| \frac{1}{n} \sum_{j=1}^n \omega^{kz_j} \right|,$$

where ω is a primitive M -th root of unity; the right-hand side is obviously the same for any such ω . Equivalently, we may write

$$\text{disc}_M(Z) = \max_{\omega \neq 1: \omega^M=1} \left| \frac{1}{n} \sum_{j=1}^n \omega^{z_j} \right|,$$

where the maximum is over M -th roots of unity ω other than 1. Yet another way to think of M -discrepancy is in terms of the discrete Fourier transform on \mathbb{Z}_M . Specifically, consider the *frequency vector* $(f_0, f_1, \dots, f_{M-1})$ of Z , where f_j is the total number of element occurrences in Z that are congruent to j modulo M . Applying the discrete Fourier transform to $(f_j)_{j=0}^{M-1}$ produces the sequence $(\sum_{j=0}^{M-1} f_j \exp(-2\pi i k j / M))_{k=0}^{M-1} = (\sum_{j=1}^n \exp(-2\pi i k z_j / M))_{k=0}^{M-1}$, which is a permutation of $(n, \sum_{j=1}^n \omega^{z_j}, \dots, \sum_{j=1}^n \omega^{(M-1)z_j})$ for a primitive M -th root of unity ω . Thus, the M -discrepancy of Z coincides up to a normalizing factor with the largest absolute value of a nonconstant Fourier coefficient of the frequency vector of Z . The notion of m -discrepancy has a long history in combinatorics and theoretical computer science; see [48] for a bibliographic overview.

LEMMA 3.4 (Discrepancy under sampling without replacement). *Fix integers $n \geq t \geq 1$ and $M \geq 2$. Let $Z = \{z_1, z_2, \dots, z_n\}$ be a multiset of integers. Then for all $\alpha \in [0, 1]$,*

$$\mathbf{P}_{S \in \binom{[n]}{t}} [\text{disc}_M(\{z_i : i \in S\}) - \text{disc}_M(Z) \geq \alpha] \leq 4M \exp\left(-\frac{t\alpha^2}{8}\right),$$

where $\{z_i : i \in S\}$ is understood to be a multiset of cardinality t .

Proof. Fix an M -th root of unity ω . Then $\text{Re}(\omega^{z_1}), \text{Re}(\omega^{z_2}), \dots, \text{Re}(\omega^{z_n})$ range in $[-1, 1]$. Now, let $S \in \binom{[n]}{t}$ be a uniformly random subset. Then the Hoeffding inequality for sampling without replacement (Theorem 2.3) implies that

$$\mathbf{P}_{S \in \binom{[n]}{t}} \left[\left| \frac{1}{t} \sum_{j \in S} \text{Re}(\omega^{z_j}) - \frac{1}{n} \sum_{j=1}^n \text{Re}(\omega^{z_j}) \right| \geq \frac{\alpha}{2} \right] \leq 2 \exp\left(-\frac{t\alpha^2}{8}\right).$$

Analogously,

$$\mathbf{P}_{S \in \binom{[n]}{t}} \left[\left| \frac{1}{t} \sum_{j \in S} \text{Im}(\omega^{z_j}) - \frac{1}{n} \sum_{j=1}^n \text{Im}(\omega^{z_j}) \right| \geq \frac{\alpha}{2} \right] \leq 2 \exp\left(-\frac{t\alpha^2}{8}\right).$$

Combining these two equations shows that for every M -th root of unity ω ,

$$\mathbf{P}_{S \in \binom{[n]}{t}} \left[\left| \frac{1}{t} \sum_{j \in S} \omega^{z_j} - \frac{1}{n} \sum_{j=1}^n \omega^{z_j} \right| \geq \alpha \right] \leq 4 \exp\left(-\frac{t\alpha^2}{8}\right). \quad (3.3)$$

Now

$$\begin{aligned} & \text{disc}_M(\{z_i : i \in S\}) - \text{disc}_M(Z) \\ &= \max_{\omega} \left| \frac{1}{t} \sum_{j \in S} \omega^{z_j} \right| - \max_{\omega} \left| \frac{1}{n} \sum_{j=1}^n \omega^{z_j} \right| \\ &\leq \max_{\omega} \left\{ \left| \frac{1}{t} \sum_{j \in S} \omega^{z_j} \right| - \left| \frac{1}{n} \sum_{j=1}^n \omega^{z_j} \right| \right\} \\ &\leq \max_{\omega} \left| \frac{1}{t} \sum_{j \in S} \omega^{z_j} - \frac{1}{n} \sum_{j=1}^n \omega^{z_j} \right|, \end{aligned} \quad (3.4)$$

where the maximum in all equations is taken over M -th roots of unity $\omega \neq 1$. Using (3.3) and the union bound over ω , we see that the right-hand side of (3.4) is bounded by α with probability at least $1 - 4M \exp(-t\alpha^2/8)$. \square

3.3. A low-discrepancy set. The construction of sparse integer sets with small discrepancy relative to a given modulus M is a well-studied problem. There is an inherent trade-off between the size of the set and the discrepancy it achieves, and different works have focused on different regimes depending on the application at hand. We work in a regime not considered previously: for any constant $\epsilon > 0$, we

construct a set of cardinality at most M^ϵ that has M -discrepancy at most $M^{-\delta}$ for some constant $\delta = \delta(\epsilon) > 0$. We construct such a set based on the following result.

THEOREM 3.5 (cf. [2, 48]). *Fix an integer $R \geq 1$ and reals $P \geq 2$ and $\Delta \geq 1$. Let M be an integer with*

$$M \geq P^2(R+1).$$

Fix a set $S_p \subseteq \{1, 2, \dots, p-1\}$ for each prime $p \in (P/2, P]$ with $p \nmid M$. Suppose further that the cardinalities of any two sets from among the S_p differ by a factor of at most Δ . Consider the multiset

$$S = \{(r + s \cdot (p^{-1})_M) \bmod M : \\ r = 1, \dots, R; \quad p \in (P/2, P] \text{ prime with } p \nmid M; \quad s \in S_p\}. \quad (3.5)$$

Then the elements of S are pairwise distinct and nonzero. Moreover, if $S \neq \emptyset$ then

$$\text{disc}_M(S) \leq \frac{c}{\sqrt{R}} + \frac{c \log M}{\log \log M} \cdot \frac{\log P}{P} \cdot \Delta + \max_p \{\text{disc}_p(S_p)\}$$

for some (explicitly given) constant $c \geq 1$ independent of P, R, M, Δ .

Ajtai et al. [2] proved a special case of Theorem 3.5 for M prime and $\Delta = 1$. Their argument was generalized in [48, Theorem 3.6] to arbitrary moduli M , again in the setting of $\Delta = 1$. The treatment in [48] in turn readily generalizes to any $\Delta \geq 1$, and for the reader's convenience we provide a complete proof of Theorem 3.5 in Appendix A. With this result in hand, we obtain the low-discrepancy set with the needed parameters:

THEOREM 3.6 (Explicit low-discrepancy set). *For all integers $M \geq 2$ and $t \geq 2$, there is an (explicitly given) nonempty set $S \subseteq \{1, 2, \dots, M\}$ with*

$$|S| \leq t, \quad (3.6)$$

$$\text{disc}_M(S) \leq \frac{C^* \log t}{t^{1/4}} \cdot \frac{\log M}{1 + \log \log M}, \quad (3.7)$$

where $C^ \geq 1$ is an (explicitly given) absolute constant independent of M and t .*

Proof. Facts 2.17 and 2.18 imply that

$$\pi(P) - \pi\left(\frac{P}{2}\right) \geq \frac{P}{C \log P} \quad \text{for all } P \geq C, \quad (3.8)$$

$$\nu(M) \leq \frac{C \log M}{1 + \log \log M} \quad \text{for all } M \geq 2, \quad (3.9)$$

for some integer $C \geq 1$ that is an absolute constant. Moreover, C can be easily calculated from the explicit bounds in Facts 2.17 and 2.18. We will show that the theorem holds for some constant $C^* \geq 4C^2$.

For $t \geq M$, the theorem is trivial since the set $S = \{1, 2, \dots, M\}$ achieves $\text{disc}_M(S) = 0$. Also, if the right-hand side of (3.7) exceeds 1, then (3.7) holds

trivially for the set $S = \{1\}$. In what follows, we treat the remaining case when

$$t < M, \tag{3.10}$$

$$\frac{4C^2 \log t}{t^{1/4}} \cdot \frac{\log M}{1 + \log \log M} \leq 1. \tag{3.11}$$

The latter condition forces

$$t \geq \max\{81, C^8\}. \tag{3.12}$$

Set $P = \lfloor t^{1/4} \rfloor$ and $R = \lfloor \sqrt{t} - 1 \rfloor$. Then (3.10) and (3.12) imply that $P \geq \max\{3, C\}$, $R \geq 1$, and $M \geq P^2(R + 1)$. As a result, Theorem 3.5 is applicable with the sets $S_p = \{1, 2, \dots, p - 1\}$ for prime $p \in (P/2, P]$. The discrepancy of these sets is given by $\text{disc}_p(S_p) = 1/(p - 1)$. Define S by (3.5). The interval $(P/2, P]$ contains $\pi(P) - \pi(P/2)$ prime numbers, of which at most $\nu(M)$ are divisors of M . We have

$$\begin{aligned} \pi(P) - \pi\left(\frac{P}{2}\right) - \nu(M) &\geq \frac{P}{C \log P} - \frac{C \log M}{1 + \log \log M} \\ &\geq \frac{t^{1/4}}{C \log t} - \frac{C \log M}{1 + \log \log M} \\ &> 0, \end{aligned}$$

where the first step uses (3.8), (3.9), and $P \geq C$, and the last step uses (3.11). We conclude that $(P/2, P]$ contains a prime that does not divide M , which in turn implies that S is nonempty. Continuing, $P \geq 3$ forces $\Delta \leq (P - 1)/(\lceil P/2 \rceil - 1) \leq 3$ in the notation of Theorem 3.5. As a result, Theorem 3.5 guarantees (3.7) for a large enough constant C^* . We note that C^* can be easily calculated from the constant c in Theorem 3.5. Since $|S| \leq RP^2 \leq t$ by definition, the proof is complete. \square

3.4. Discrepancy and balanced colorings. We will leverage the low-discrepancy integer set in Theorem 3.6 to construct a balanced coloring of $\binom{[n]}{k}$. For this, we now develop a connection between these two notions of pseudorandomness. We will henceforth denote the modulus by r since in our construction, the modulus is set equal to the number of colors in the coloring of $\binom{[n]}{k}$. We start with a technical lemma.

LEMMA 3.7. *Fix integers ℓ, k, r with $\ell \geq k \geq 1$ and $r \geq 2$. Let $Z = \{z_1, z_2, \dots, z_\ell\}$ be a multiset of integers. Then for all $\alpha \in [0, 1]$,*

$$\begin{aligned} \max_{a \in \mathbb{Z}} \left| \mathbf{P}_{S \in \binom{[\ell]}{k}} \left[\sum_{i \in S} z_i \equiv a \pmod{r} \right] - \frac{1}{r} \right| \\ \leq 4rk \exp\left(-\frac{\lfloor \ell/k \rfloor \alpha^2}{8}\right) + (\text{disc}_r(Z) + \alpha)^k. \end{aligned}$$

Proof. Let ω be a primitive r -th root of unity. Then

$$\begin{aligned}
\mathbf{P}_{S \in \binom{[\ell]}{k}} \left[\sum_{i \in S} z_i \equiv a \pmod{r} \right] &= \mathbf{E}_{S \in \binom{[\ell]}{k}} \mathbf{I} \left[\sum_{i \in S} z_i \equiv a \pmod{r} \right] \\
&= \mathbf{E}_{S \in \binom{[\ell]}{k}} \frac{1}{r} \sum_{t=0}^{r-1} \omega^{t(\sum_{i \in S} z_i - a)} \\
&= \frac{1}{r} \sum_{t=0}^{r-1} \mathbf{E}_{S \in \binom{[\ell]}{k}} \omega^{t(\sum_{i \in S} z_i - a)} \\
&= \frac{1}{r} + \frac{1}{r} \sum_{t=1}^{r-1} \mathbf{E}_{S \in \binom{[\ell]}{k}} \omega^{t(\sum_{i \in S} z_i - a)} \\
&= \frac{1}{r} + \frac{1}{r} \sum_{t=1}^{r-1} \mathbf{E}_{i_1, i_2, \dots, i_k} \omega^{t(z_{i_1} + z_{i_2} + \dots + z_{i_k} - a)},
\end{aligned}$$

where the final expectation is taken over a uniformly random tuple of indices $i_1, i_2, \dots, i_k \in [\ell]$ that are pairwise distinct. Therefore,

$$\begin{aligned}
\left| \mathbf{P}_{S \in \binom{[\ell]}{k}} \left[\sum_{i \in S} z_i \equiv a \pmod{r} \right] - \frac{1}{r} \right| &= \left| \frac{1}{r} \sum_{t=1}^{r-1} \mathbf{E}_{i_1, i_2, \dots, i_k} \omega^{t(z_{i_1} + z_{i_2} + \dots + z_{i_k} - a)} \right| \\
&\leq \frac{1}{r} \sum_{t=1}^{r-1} \left| \mathbf{E}_{i_1, i_2, \dots, i_k} \omega^{t(z_{i_1} + z_{i_2} + \dots + z_{i_k} - a)} \right| \\
&= \frac{1}{r} \sum_{t=1}^{r-1} \left| \mathbf{E}_{i_1, i_2, \dots, i_k} \prod_{j=1}^k \omega^{tz_{i_j}} \right|. \quad (3.13)
\end{aligned}$$

We now introduce conditioning to make i_1, i_2, \dots, i_k independent random variables. Specifically, i_1, i_2, \dots, i_k can be generated by the following two-step procedure:

- (i) pick uniformly random sets $S_1, S_2, \dots, S_k \in \binom{[\ell]}{[\ell/k]}$ that are pairwise disjoint;
- (ii) for $j = 1, 2, \dots, k$, pick i_j uniformly at random from among the elements of S_j .

By symmetry, this procedure generates every tuple (i_1, i_2, \dots, i_k) of pairwise distinct integers with equal probability. Importantly, conditioning on S_1, S_2, \dots, S_k makes i_1, i_2, \dots, i_k independent. Now (3.13) gives

$$\begin{aligned}
\max_{a \in \mathbb{Z}} \left| \mathbf{P}_{S \in \binom{[\ell]}{k}} \left[\sum_{i \in S} z_i \equiv a \pmod{r} \right] - \frac{1}{r} \right| &\leq \frac{1}{r} \sum_{t=1}^{r-1} \left| \mathbf{E}_{i_1, i_2, \dots, i_k} \prod_{j=1}^k \omega^{tz_{i_j}} \right| \\
&= \frac{1}{r} \sum_{t=1}^{r-1} \left| \mathbf{E}_{S_1, S_2, \dots, S_k} \prod_{j=1}^k \mathbf{E}_{i_j \in S_j} \omega^{tz_{i_j}} \right|
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{r} \sum_{t=1}^{r-1} \mathbf{E}_{S_1, S_2, \dots, S_k} \prod_{j=1}^k \left| \mathbf{E}_{i_j \in S_j} \omega^{tz_{i_j}} \right| \\
&\leq \frac{1}{r} \sum_{t=1}^{r-1} \mathbf{E}_{S_1, S_2, \dots, S_k} \prod_{j=1}^k \text{disc}_r(\{z_i : i \in S_j\}) \\
&\leq \mathbf{E}_{S_1, S_2, \dots, S_k} \prod_{j=1}^k \text{disc}_r(\{z_i : i \in S_j\}), \tag{3.14}
\end{aligned}$$

where $\{z_i : i \in S_j\}$ for each j is a multiset of cardinality $\lfloor \ell/k \rfloor$.

Let B_j be the event that $\{z_i : i \in S_j\}$ has r -discrepancy greater than $\text{disc}_r(Z) + \alpha$, and let $B = B_1 \vee B_2 \vee \dots \vee B_k$. Conditioned on B , we get $\prod_j \text{disc}_r(\{z_i : i \in S_j\}) \leq 1$ since r -discrepancy is at most 1. Conditioned on \overline{B} , we have by definition that $\prod_j \text{disc}_r(\{z_i : i \in S_j\}) \leq (\text{disc}_r(Z) + \alpha)^k$. Thus,

$$\mathbf{E}_{S_1, S_2, \dots, S_k} \prod_{j=1}^k \text{disc}_r(\{z_i : i \in S_j\}) \leq \mathbf{P}_{S_1, S_2, \dots, S_k} [B] + (\text{disc}_r(Z) + \alpha)^k. \tag{3.15}$$

Recall that S_1, S_2, \dots, S_k are identically distributed, namely, each S_j has the distribution of a uniformly random subset of $[\ell]$ of cardinality $\lfloor \ell/k \rfloor$. As a result, Lemma 3.4 guarantees that B_j occurs with probability at most $4r \exp(-\lfloor \ell/k \rfloor \alpha^2/8)$. Applying the union bound over all j ,

$$\mathbf{P}_{S_1, S_2, \dots, S_k} [B] \leq 4rk \exp\left(-\frac{\lfloor \ell/k \rfloor \alpha^2}{8}\right). \tag{3.16}$$

Combining (3.14)–(3.16) concludes the proof. \square

We are now in a position to give our general transformation of a low-discrepancy integer set into a balanced coloring of $\binom{[n]}{k}$.

THEOREM 3.8 (From a low-discrepancy set to a balanced coloring). *Let n, m, k, r be integers with $n \geq m \geq k \geq 1$ and $r \geq 2$. Let $Z = \{z_1, z_2, \dots, z_n\}$ be a multiset of integers. Define $\gamma: \binom{[n]}{k} \rightarrow [r]$ by*

$$\gamma(S) = 1 + \left(\left(\sum_{i \in S} z_i \right) \bmod r \right). \tag{3.17}$$

Let $\beta, \zeta \in [0, 1]$ be arbitrary. Then γ is (ϵ, δ, m) -balanced, where

$$\begin{aligned}
\epsilon &= 4r^2k \exp\left(-\frac{\lfloor m/k \rfloor \zeta^2}{8}\right) + r(\text{disc}_r(Z) + \beta + \zeta)^k, \\
\delta &= 4r \exp\left(-\frac{m\beta^2}{8}\right).
\end{aligned}$$

Proof. Let $\ell \in \{m, m+1, \dots, n\}$ be arbitrary. Then Lemma 3.4 implies that for all but a δ fraction of the sets $A \in \binom{[n]}{\ell}$,

$$\text{disc}_r(\{z_i : i \in A\}) \leq \text{disc}_r(Z) + \beta. \tag{3.18}$$

It remains to prove that γ is ϵ -balanced on every set $A \in \binom{[n]}{\ell}$ that satisfies (3.18). We have

$$\begin{aligned}
& \max_{a \in [r]} \left| \frac{|\gamma^{-1}(a) \cap \binom{A}{k}|}{\binom{|A|}{k}} - \frac{1}{r} \right| \\
&= \max_{a \in [r]} \left| \mathbf{P}_{S \in \binom{A}{k}} [\gamma(S) = a] - \frac{1}{r} \right| \\
&= \max_{a \in \mathbb{Z}} \left| \mathbf{P}_{S \in \binom{A}{k}} \left[\sum_{i \in S} z_i \equiv a \pmod{r} \right] - \frac{1}{r} \right| \\
&\leq 4rk \exp\left(-\frac{\lfloor \ell/k \rfloor \zeta^2}{8}\right) + (\text{disc}_r(\{z_i : i \in A\}) + \zeta)^k \\
&\leq 4rk \exp\left(-\frac{\lfloor m/k \rfloor \zeta^2}{8}\right) + (\text{disc}_r(Z) + \beta + \zeta)^k \\
&= \frac{\epsilon}{r},
\end{aligned}$$

where the second step uses the definition of γ , the third step applies Lemma 3.7, the fourth step uses (3.18) and $\ell \geq m$, and the fifth step uses the definition of ϵ . We have shown that γ is ϵ -balanced on A , thereby completing the proof. \square

3.5. An explicit balanced coloring. Theorem 3.8 transforms any integer set with small r -discrepancy into a balanced coloring with r colors. We now apply this transformation to the low-discrepancy integer set constructed earlier, resulting in an explicit balanced coloring.

THEOREM 3.9 (Explicit balanced coloring). *Let n, m, k, r be integers with $n/2 \geq m \geq k \geq 1$ and $r \geq 2$. Let $\beta, \zeta \in [0, 1]$ be arbitrary. Then there is an (explicitly given) integer $n' \in (n/2, n]$ and an (explicitly given) (ϵ, δ, m) -balanced coloring $\gamma: \binom{[n']}{k} \rightarrow [r]$, where*

$$\epsilon = 4r^2k \exp\left(-\frac{\lfloor m/k \rfloor \zeta^2}{8}\right) + r \left(\frac{C^* \log n}{n^{1/4}} \cdot \frac{\log r}{1 + \log \log r} + \beta + \zeta \right)^k, \quad (3.19)$$

$$\delta = 4r \exp\left(-\frac{m\beta^2}{8}\right), \quad (3.20)$$

and $C^* \geq 1$ is the absolute constant from Theorem 3.6.

Proof. By hypothesis, $n \geq 2$. Invoke Theorem 3.6 with $M = r$ and $t = n$ to obtain an explicit nonempty set $S \subseteq \{1, 2, \dots, r\}$ with

$$\begin{aligned}
& |S| \leq n, \\
& \text{disc}_r(S) \leq \frac{C^* \log n}{n^{1/4}} \cdot \frac{\log r}{1 + \log \log r}.
\end{aligned}$$

Let Z be the union of $\lfloor n/|S| \rfloor$ copies of S . Then $\text{disc}_r(Z) = \text{disc}_r(S)$ by the definition of r -discrepancy. Letting $n' = |Z|$, we claim that $n' \in (n/2, n]$. Indeed, the upper

bound is justified by $n' = |S| \cdot \lfloor n/|S| \rfloor \leq n$, whereas the lower bound is the arithmetic mean of the bounds $n' \geq |S|$ and $n' > |S|(n/|S| - 1)$.

Now, let $z_1, z_2, \dots, z_{n'}$ be the elements of Z and define $\gamma: \binom{[n']}{k} \rightarrow [r]$ by (3.17). Then Theorem 3.8 implies that γ is (ϵ, δ, m) -balanced with ϵ, δ given by (3.19) and (3.20), respectively. \square

Taking $\beta = \zeta = m^{-1/4}$ in Theorem 3.9, we obtain:

COROLLARY 3.10 (Explicit balanced coloring). *Let n, m, k, r be integers with $n/2 \geq m \geq k \geq 1$ and $r \geq 2$. Then there is an (explicitly given) integer $n' \in (n/2, n]$ and an (explicitly given) (ϵ, δ, m) -balanced coloring $\gamma: \binom{[n']}{k} \rightarrow [r]$, where*

$$\epsilon = 4r^2 k \exp\left(-\frac{\sqrt{m}}{16k}\right) + r \left(\frac{3C^* \log^2(n+r)}{m^{1/4}}\right)^k, \quad (3.21)$$

$$\delta = 4r \exp\left(-\frac{\sqrt{m}}{8}\right), \quad (3.22)$$

and $C^* \geq 1$ is the absolute constant from Theorem 3.6.

The parameters in Corollary 3.10 generously meet our requirements. In our setting of interest, the integers n, m, r are polynomially related. Thus, we obtain an (m^{-K}, m^{-K}, m) -balanced coloring for any desired constant $K \geq 1$ by invoking Corollary 3.10 with a large enough constant $k = k(K)$.

4. HARDNESS AMPLIFICATION

In Section 3, we laid the foundation for our main result by constructing an explicit integer set with small discrepancy and transforming it into a highly balanced coloring of $\binom{[n]}{k}$. In this section, we use this coloring to design a hardness amplification method for approximate degree and its one-sided variant.

4.1. Pseudodistributions from balanced colorings. Recall from the introduction that our approach centers around encoding the vectors $e_1, e_2, \dots, e_N, 0^N$ as n -bit strings with $n \ll N$ so as to make the decoding easy for circuits but hard for low-degree polynomials. The construction of this code requires several steps. As a first step, we show how to convert any balanced coloring of $\binom{[n]}{k}$ with r colors into an explicit sequence of functions $\phi_1, \phi_2, \dots, \phi_r: \{0, 1\}^n \rightarrow \mathbb{R}$ that are almost everywhere nonnegative, are supported almost entirely on pairwise disjoint sets of strings of Hamming weight k , and are pairwise indistinguishable by low-degree polynomials. We call them *pseudodistributions* to highlight the fact that each ϕ_i has ℓ_1 norm approximately 1, nearly all of it coming from the points where ϕ_i is nonnegative.

THEOREM 4.1. *Let $\epsilon, \delta \in [0, 1)$ be given. Let n, m, k, r be positive integers with $n \geq m > k$. Let $\gamma: \binom{[n]}{k} \rightarrow [r]$ be a given (ϵ, δ, m) -balanced coloring. Then there are (explicitly given) functions $\phi_1, \phi_2, \dots, \phi_r: \{0, 1\}^n \rightarrow \mathbb{R}$ with the following properties.*

- (i) **Support:** $\text{supp } \phi_i \subseteq \{x \in \{0, 1\}^n : |x| = k \text{ or } |x| \geq m\}$;
- (ii) **Essential support:** $\{0, 1\}^n|_k \cap \text{supp } \phi_i = \{\mathbf{1}_S : S \in \gamma^{-1}(i)\}$;

- (iii) **Nonnegativity:** $\phi_i \geq 0$ on $\{0, 1\}^n|_k$;
- (iv) **Normalization:** $\sum_{x:|x|=k} \phi_i(x) = 1$;
- (v) **Tail bound:** $\sum_{x:|x| \neq k} |\phi_i(x)| \leq (8\epsilon + 4r\delta)/(1 - \epsilon)$;
- (vi) **Graded bound:** for some absolute constant $c' \in (0, 1)$,

$$\sum_{x:|x|=\ell} |\phi_i(x)| \leq \frac{\epsilon + r\delta}{1 - \epsilon} \cdot \frac{m^2}{c'\ell^2} \cdot \exp\left(-\frac{c'(\ell - k)}{\sqrt{nm}}\right), \quad \ell > k;$$

- (vii) **Orthogonality:** for some absolute constant $c'' \in (0, 1)$,

$$\text{orth}(\phi_i - \phi_j) \geq c'' \sqrt{\frac{n}{m}}, \quad i, j \in [r].$$

Proof. Define

$$\Delta = m - k, \tag{4.1}$$

$$D = \left\lfloor \frac{n - k}{\Delta} \right\rfloor. \tag{4.2}$$

Setting $\epsilon = 1/2$ in Lemma 2.10 gives an explicit function $\omega: \{0, 1, 2, \dots, D\} \rightarrow \mathbb{R}$ with

$$\omega(0) > \frac{1}{4} \|\omega\|_1, \tag{4.3}$$

$$|\omega(t)| \leq \frac{1}{ct^2 2^{ct/\sqrt{D}}} \cdot \|\omega\|_1 \quad (t = 1, 2, \dots, D), \tag{4.4}$$

$$\text{orth } \omega \geq c\sqrt{D}, \tag{4.5}$$

where $0 < c < 1$ is an absolute constant. For convenience of notation, we will extend ω to all of \mathbb{R} by setting $\omega(t) = 0$ for $t \notin \{0, 1, 2, \dots, D\}$. With this extension, (4.4) gives

$$|\omega(t)| \leq \frac{1}{ct^2 2^{ct/\sqrt{D}}} \cdot \|\omega\|_1, \quad t \in [1, \infty). \tag{4.6}$$

For $S \in \binom{[n]}{k}$, define an auxiliary dual object $\phi_S: \{0, 1\}^n \rightarrow \mathbb{R}$ by

$$\phi_S(x) = \binom{n - k}{|x| - k}^{-1} \omega(0)^{-1} \omega\left(\frac{|x| - k}{\Delta}\right) \prod_{i \in S} x_i. \tag{4.7}$$

Then

$$\phi_S(\mathbf{1}_S) = 1, \quad S \in \binom{[n]}{k}. \tag{4.8}$$

Since $\phi_S(x) = 0$ unless $x|_S = \mathbf{1}^k$, we see that $\mathbf{1}_S$ is in fact the only input of Hamming weight k at which ϕ_S is nonzero:

$$\phi_S(\mathbf{1}_T) = \delta_{S,T}, \quad S, T \in \binom{[n]}{k}. \tag{4.9}$$

Since $\text{supp } \omega \subseteq \{0, 1, 2, \dots, D\}$, the only inputs x other than $\mathbf{1}_S$ in the support of ϕ_S have Hamming weight $|x| \in \{k + \Delta, k + 2\Delta, \dots, k + D\Delta\}$, so that in particular $|x| \geq m$. In summary,

$$\text{supp } \phi_S \subseteq \{x : x = \mathbf{1}_S \text{ or } |x| \geq m\}, \quad S \in \binom{[n]}{k}, \quad (4.10)$$

$$\text{supp } \phi_S \subseteq \bigcup_{i=0}^D \{x : |x| = k + i\Delta\}, \quad S \in \binom{[n]}{k}. \quad (4.11)$$

We now turn to the construction of the ϕ_i . By definition of an (ϵ, δ, m) -balanced coloring, the given coloring $\gamma: \binom{[n]}{k} \rightarrow [r]$ satisfies

$$\mathbf{P}_{A \in \binom{[n]}{\ell}} \left[\frac{1 - \epsilon}{r} \binom{|A|}{k} \leq \left| \gamma^{-1}(i) \cap \binom{A}{k} \right| \leq \frac{1 + \epsilon}{r} \binom{|A|}{k} \right] \geq 1 - \delta, \quad \ell = m, m + 1, \dots, n. \quad (4.12)$$

Since $\delta < 1$, taking $\ell = n$ in this equation leads to

$$\left| |\gamma^{-1}(i)| - \frac{1}{r} \binom{n}{k} \right| \leq \frac{\epsilon}{r} \binom{n}{k}, \quad i \in [r], \quad (4.13)$$

and in particular

$$|\gamma^{-1}(i)| \geq \frac{1 - \epsilon}{r} \binom{n}{k}, \quad i \in [r]. \quad (4.14)$$

For $i = 1, 2, \dots, r$, we define $\phi_i: \{0, 1\}^n \rightarrow \mathbb{R}$ by

$$\phi_i(x) = \mathbf{E}_{S \in \gamma^{-1}(i)} \phi_S(x) - \mathbf{I}[|x| \geq m] \mathbf{E}_{S \in \binom{[n]}{k}} \phi_S(x).$$

This definition is legitimate since $\gamma^{-1}(i) \neq \emptyset$ for every i due to (4.14) and $\epsilon < 1$.

CLAIM 4.2. *For all $i \in [r]$ and $\ell \in \{m, m + 1, \dots, n\}$,*

$$\mathbf{E}_{A \in \binom{[n]}{\ell}} \left| \mathbf{P}_{S \in \gamma^{-1}(i)} [S \subseteq A] - \mathbf{P}_{S \in \binom{[n]}{k}} [S \subseteq A] \right| \leq \frac{2\epsilon + r\delta}{1 - \epsilon} \binom{n}{k}^{-1} \binom{\ell}{k}.$$

Proof. Fix $i \in [r]$ and $\ell \in \{m, m + 1, \dots, n\}$ arbitrarily for the remainder of the proof. Let $A \in \binom{[n]}{\ell}$ be uniformly random. If γ is ϵ -balanced on A , then by definition

$$\left| \left| \gamma^{-1}(i) \cap \binom{A}{k} \right| - \frac{1}{r} \binom{|A|}{k} \right| \leq \frac{\epsilon}{r} \binom{|A|}{k}.$$

If γ is not ϵ -balanced on A , we have the trivial bound

$$\left| \left| \gamma^{-1}(i) \cap \binom{A}{k} \right| - \frac{1}{r} \binom{|A|}{k} \right| \leq \binom{|A|}{k}.$$

Combining these two equations, we arrive at

$$\left| \left| \gamma^{-1}(i) \cap \binom{A}{k} \right| - \frac{1}{r} \binom{|A|}{k} \right| \leq \left(\frac{\epsilon}{r} + Y_A \right) \binom{|A|}{k} \quad (4.15)$$

for all A , where Y_A is the indicator random variable for the event that γ is not ϵ -balanced on A . Since γ is (ϵ, δ, m) -balanced, we further have

$$\mathbf{E}_{A \in \binom{[n]}{\ell}} Y_A \leq \delta. \quad (4.16)$$

Now

$$\begin{aligned} & \left| \mathbf{P}_{S \in \gamma^{-1}(i)} [S \subseteq A] - \mathbf{P}_{S \in \binom{[n]}{k}} [S \subseteq A] \right| \\ &= \left| \frac{|\gamma^{-1}(i) \cap \binom{A}{k}|}{|\gamma^{-1}(i)|} - \frac{\binom{|A|}{k}}{\binom{n}{k}} \right| \\ &= \frac{1}{|\gamma^{-1}(i)|} \binom{n}{k}^{-1} \left| \left| \gamma^{-1}(i) \cap \binom{A}{k} \right| \binom{n}{k} - |\gamma^{-1}(i)| \binom{|A|}{k} \right| \\ &\leq \frac{r}{1-\epsilon} \binom{n}{k}^{-2} \left| \left| \gamma^{-1}(i) \cap \binom{A}{k} \right| \binom{n}{k} - |\gamma^{-1}(i)| \binom{|A|}{k} \right| \\ &\leq \frac{r}{1-\epsilon} \binom{n}{k}^{-2} \left| \left| \gamma^{-1}(i) \cap \binom{A}{k} \right| \binom{n}{k} - \frac{1}{r} \binom{|A|}{k} \binom{n}{k} \right| \\ &\quad + \frac{r}{1-\epsilon} \binom{n}{k}^{-2} \left| \frac{1}{r} \binom{|A|}{k} \binom{n}{k} - |\gamma^{-1}(i)| \binom{|A|}{k} \right| \\ &\leq \frac{r}{1-\epsilon} \binom{n}{k}^{-2} \left(\frac{\epsilon}{r} + \frac{\epsilon}{r} + Y_A \right) \binom{|A|}{k} \binom{n}{k} \\ &= \frac{r}{1-\epsilon} \binom{n}{k}^{-1} \left(\frac{2\epsilon}{r} + Y_A \right) \binom{\ell}{k}, \end{aligned} \quad (4.17)$$

where the third step is valid by (4.14), the fourth step applies the triangle inequality, the fifth step uses (4.13) and (4.15), and the last step uses $|A| = \ell$. It remains to pass to expectations with respect to A :

$$\begin{aligned} & \mathbf{E}_{A \in \binom{[n]}{\ell}} \left| \mathbf{P}_{S \in \gamma^{-1}(i)} [S \subseteq A] - \mathbf{P}_{S \in \binom{[n]}{k}} [S \subseteq A] \right| \\ &\leq \mathbf{E}_{A \in \binom{[n]}{\ell}} \frac{r}{1-\epsilon} \binom{n}{k}^{-1} \left(\frac{2\epsilon}{r} + Y_A \right) \binom{\ell}{k} \\ &= \frac{r}{1-\epsilon} \binom{n}{k}^{-1} \left(\frac{2\epsilon}{r} + \mathbf{E}_{A \in \binom{[n]}{\ell}} Y_A \right) \binom{\ell}{k} \\ &\leq \frac{2\epsilon + r\delta}{1-\epsilon} \binom{n}{k}^{-1} \binom{\ell}{k}, \end{aligned}$$

where the last step uses (4.16). \square

CLAIM 4.3. For each $i \in [r]$ and $\ell \in \{m, m+1, \dots, n\}$,

$$\sum_{x:|x|=\ell} |\phi_i(x)| \leq \frac{2\epsilon + r\delta}{1-\epsilon} \cdot \left| \frac{\omega(\frac{\ell-k}{\Delta})}{\omega(0)} \right|.$$

Proof. Fix $i \in [r]$ and $\ell \in \{m, m+1, \dots, n\}$ arbitrarily for the remainder of the proof. Consider any input $x = \mathbf{1}_A$ with $|A| = \ell$. In this case, the definition of ϕ_i simplifies to

$$\phi_i(\mathbf{1}_A) = \mathbf{E}_{S \in \gamma^{-1}(i)} \phi_S(\mathbf{1}_A) - \mathbf{E}_{S \in \binom{[n]}{k}} \phi_S(\mathbf{1}_A).$$

Recall from (4.7) that

$$\phi_S(\mathbf{1}_A) = \frac{\omega\left(\frac{\ell-k}{\Delta}\right)}{\omega(0)\binom{n-k}{\ell-k}} \cdot \mathbf{I}[S \subseteq A].$$

As a result,

$$\phi_i(\mathbf{1}_A) = \frac{\omega\left(\frac{\ell-k}{\Delta}\right)}{\omega(0)\binom{n-k}{\ell-k}} \left(\mathbf{P}_{S \in \gamma^{-1}(i)}[S \subseteq A] - \mathbf{P}_{S \in \binom{[n]}{k}}[S \subseteq A] \right).$$

Passing to absolute values and summing over $A \in \binom{[n]}{\ell}$, we obtain

$$\begin{aligned} \sum_{A \in \binom{[n]}{\ell}} |\phi_i(\mathbf{1}_A)| &= \left| \frac{\omega\left(\frac{\ell-k}{\Delta}\right)}{\omega(0)\binom{n-k}{\ell-k}} \right| \sum_{A \in \binom{[n]}{\ell}} \left| \mathbf{P}_{S \in \gamma^{-1}(i)}[S \subseteq A] - \mathbf{P}_{S \in \binom{[n]}{k}}[S \subseteq A] \right| \\ &\leq \left| \frac{\omega\left(\frac{\ell-k}{\Delta}\right)}{\omega(0)\binom{n-k}{\ell-k}} \right| \cdot \binom{n}{\ell} \cdot \frac{2\epsilon + r\delta}{1-\epsilon} \cdot \binom{n}{k}^{-1} \binom{\ell}{k} \\ &= \left| \frac{\omega\left(\frac{\ell-k}{\Delta}\right)}{\omega(0)} \right| \cdot \frac{2\epsilon + r\delta}{1-\epsilon}, \end{aligned}$$

where the second step applies Claim 4.2, and the final step is justified by

$$\begin{aligned} &\binom{n-k}{\ell-k}^{-1} \binom{n}{\ell} \binom{n}{k}^{-1} \binom{\ell}{k} \\ &= \frac{(\ell-k)! (n-\ell)!}{(n-k)!} \cdot \frac{n!}{\ell! (n-\ell)!} \cdot \frac{k! (n-k)!}{n!} \cdot \frac{\ell!}{k! (\ell-k)!} = 1. \quad \square \end{aligned}$$

We now turn to the verification of properties (i)–(vii) in the theorem statement.

Properties (i)–(iv). Equation (4.10) shows that ϕ_i is a linear combination of functions whose support is contained in $\{x : |x| = k \text{ or } |x| \geq m\}$. This settles the support requirement (i). For $T \in \binom{[n]}{k}$,

$$\begin{aligned} \phi_i(\mathbf{1}_T) &= \mathbf{E}_{S \in \gamma^{-1}(i)} \phi_S(\mathbf{1}_T) \\ &= \mathbf{E}_{S \in \gamma^{-1}(i)} \delta_{S,T} \\ &= \frac{\mathbf{I}[T \in \gamma^{-1}(i)]}{|\gamma^{-1}(i)|}, \end{aligned} \quad T \in \binom{[n]}{k}, \quad (4.18)$$

where the first step is immediate from the defining equation for ϕ_i , and the second step applies (4.9). The essential support property (ii) and nonnegativity property (iii) are now immediate from (4.18). The normalization requirement (iv) follows by summing (4.18) over $T \in \binom{[n]}{k}$.

Properties (v) and (vi). The tail bound (v) for $i \in [r]$ can be seen as follows:

$$\begin{aligned}
\sum_{x:|x|\neq k} |\phi_i(x)| &= \sum_{x:|x|\geq m} |\phi_i(x)| \\
&= \sum_{\ell=m}^n \sum_{x:|x|=\ell} |\phi_i(x)| \\
&\leq \frac{2\epsilon + r\delta}{1-\epsilon} \cdot \sum_{\ell=m}^n \left| \omega(0)^{-1} \omega\left(\frac{\ell-k}{\Delta}\right) \right| \\
&\leq \frac{2\epsilon + r\delta}{1-\epsilon} \cdot \frac{\|\omega\|_1}{|\omega(0)|} \\
&\leq \frac{8\epsilon + 4r\delta}{1-\epsilon},
\end{aligned}$$

where the first step uses the support property (i), the third step is valid by Claim 4.3, and the last step applies (4.3).

The graded bound (vi) for $\ell \in (k, m)$ holds trivially since ϕ_i vanishes on inputs of Hamming weight in (k, m) , by the support property (i). The validity of (vi) for $\ell \geq m$ is borne out by

$$\begin{aligned}
\sum_{x:|x|=\ell} |\phi_i(x)| &\leq \frac{2\epsilon + r\delta}{1-\epsilon} \cdot \left| \omega(0)^{-1} \omega\left(\frac{\ell-k}{\Delta}\right) \right| \\
&\leq \frac{8\epsilon + 4r\delta}{1-\epsilon} \cdot \frac{1}{\|\omega\|_1} \cdot \left| \omega\left(\frac{\ell-k}{\Delta}\right) \right| \\
&\leq \frac{8\epsilon + 4r\delta}{1-\epsilon} \cdot \frac{1}{c \left(\frac{\ell-k}{\Delta}\right)^2 2^{c(\ell-k)/(\Delta\sqrt{D})}} \\
&= \frac{8\epsilon + 4r\delta}{1-\epsilon} \cdot \frac{1}{c \left(\frac{\ell-k}{m-k}\right)^2 2^{c(\ell-k)/((m-k)\sqrt{\lfloor (n-k)/(m-k) \rfloor})}} \\
&\leq \frac{8\epsilon + 4r\delta}{1-\epsilon} \cdot \frac{m^2}{c\ell^2 2^{c(\ell-k)/\sqrt{nm}}},
\end{aligned}$$

where the first step restates Claim 4.3, the second step is justified by (4.3), the third step appeals to (4.6), and the fourth step substitutes the values from (4.1) and (4.2).

Property (vii). To begin with, we claim that

$$\text{orth } \phi_S \geq c\sqrt{D}, \quad S \in \binom{[n]}{k}. \quad (4.19)$$

Indeed, let p be a real polynomial on $\{0, 1\}^n$ with $\deg p < c\sqrt{D}$. By linearity, it suffices to consider polynomials p that factor as $p(x) = p_1(x|_S)p_2(x|_{\bar{S}})$ for some nonzero polynomials p_1, p_2 . Now, Minsky and Papert's symmetrization argument (Proposition 2.14) guarantees that

$$\mathbf{E}_{\substack{y \in \{0,1\}^{n-k} \\ |y|=i}} p_2(y) = p_2^*(i), \quad i = 0, 1, 2, \dots, n-k, \quad (4.20)$$

for some univariate polynomial p_2^* of degree at most $\deg p_2$. As a result,

$$\begin{aligned}
\langle \phi_S, p \rangle &= \sum_{\substack{x \in \{0,1\}^n: \\ x|_S = 1^k}} \phi_S(x) p(x) \\
&= \sum_{i=0}^D \sum_{\substack{x \in \{0,1\}^n: \\ |x|=k+i\Delta, x|_S=1^k}} \phi_S(x) p(x) \\
&= \sum_{i=0}^D \sum_{\substack{x \in \{0,1\}^n: \\ |x|=k+i\Delta, x|_S=1^k}} \binom{n-k}{i\Delta}^{-1} \frac{\omega(i)}{\omega(0)} \cdot p(x) \\
&= \sum_{i=0}^D \mathbf{E}_{\substack{y \in \{0,1\}^{n-k}: \\ |y|=i\Delta}} \left[\frac{\omega(i)}{\omega(0)} \cdot p_1(1^k) p_2(y) \right] \\
&= \frac{p_1(1^k)}{\omega(0)} \sum_{i=0}^D \omega(i) p_2^*(i\Delta) \\
&= 0,
\end{aligned}$$

where the first and third steps use the definition of ϕ_S , the second step is justified by (4.11), the next-to-last step uses (4.20), and the last step is valid by (4.5) since $\deg p_2^* \leq \deg p_2 \leq \deg p < c\sqrt{D}$. This settles (4.19).

Now the orthogonality requirement (vii) can be seen as follows:

$$\begin{aligned}
\text{orth}(\phi_i - \phi_j) &= \text{orth} \left(\mathbf{E}_{S \in \gamma^{-1}(i)} \phi_S - \mathbf{E}_{S \in \gamma^{-1}(j)} \phi_S \right) \\
&\geq \min_{S \in \binom{[n]}{k}} \text{orth} \phi_S \\
&\geq c\sqrt{D} \\
&= c\sqrt{\left\lfloor \frac{n-k}{m-k} \right\rfloor} \\
&\geq c\sqrt{\left\lfloor \frac{n}{m} \right\rfloor},
\end{aligned}$$

where the second step uses Proposition 2.6(i), the third step is valid by (4.19), the fourth step applies the definition of D , and the last step uses $n \geq m$. \square

4.2. Encoding via indistinguishable distributions. As our next step, we will show that the pseudodistributions $\phi_1, \phi_2, \dots, \phi_r$ in Theorem 4.1 can be turned into actual probability distributions $\lambda_1, \lambda_2, \dots, \lambda_r$ provided that the underlying coloring of $\binom{[n]}{k}$ is sufficiently balanced. The resulting distributions λ_i inherit all the desirable analytic properties established for the ϕ_i in Theorem 4.1. Specifically, the λ_i are supported almost entirely on pairwise disjoint sets of inputs of Hamming weight k and are pairwise indistinguishable by low-degree polynomials.

THEOREM 4.4. *Let $0 < \beta < 1$ be given. Let n, n', m, k, r be positive integers with $n \geq n' \geq m > k$. Let $\gamma: \binom{[n']}{k} \rightarrow [r]$ be a given $(\frac{\beta}{16rm^2}, \frac{\beta}{16r^2m^2}, m)$ -balanced coloring.*

Then there are (explicitly given) probability distributions $\lambda_1, \lambda_2, \dots, \lambda_r$ on $\{0, 1\}^n$ such that

$$\text{supp } \lambda_i \subseteq \{x \in \{0, 1\}^n : |x| = k \text{ or } |x| \geq m\}, \quad i \in [r], \quad (4.21)$$

$$\{0, 1\}^n|_k \cap \text{supp } \lambda_i = \{\mathbf{1}_S : S \in \gamma^{-1}(i)\}, \quad i \in [r], \quad (4.22)$$

$$\lambda_i(\{0, 1\}^n|_k) \geq 1 - \beta, \quad i \in [r], \quad (4.23)$$

$$\lambda_i(\{0, 1\}^n|_\ell) \leq \frac{\exp(-c(\ell - k)/\sqrt{n'm})}{c(\ell - k + 1)^2}, \quad i \in [r], \ell \geq k, \quad (4.24)$$

$$\text{orth}(\lambda_i - \lambda_j) \geq c\sqrt{\frac{n'}{m}}, \quad i, j \in [r], \quad (4.25)$$

where $c \in (0, 1)$ is an absolute constant, independent of n, n', m, k, r, β .

Proof. By hypothesis, γ is (ϵ, δ, m) -balanced with

$$\epsilon = \frac{\beta}{16rm^2},$$

$$\delta = \frac{\beta}{16r^2m^2}.$$

Applying Theorem 4.1 with these parameters gives functions $\phi_1, \phi_2, \dots, \phi_r : \{0, 1\}^{n'} \rightarrow \mathbb{R}$ that obey

$$\text{supp } \phi_i \subseteq \{x \in \{0, 1\}^{n'} : |x| = k \text{ or } |x| \geq m\}, \quad (4.26)$$

$$\{0, 1\}^{n'}|_k \cap \text{supp } \phi_i = \{\mathbf{1}_S : S \in \gamma^{-1}(i)\}, \quad (4.27)$$

$$\phi_i \geq 0 \text{ on } \{0, 1\}^{n'}|_k, \quad (4.28)$$

$$\sum_{x:|x|=k} \phi_i(x) = 1, \quad (4.29)$$

$$\sum_{x:|x|\neq k} |\phi_i(x)| \leq \frac{\beta}{r}, \quad (4.30)$$

$$\sum_{x:|x|=\ell} |\phi_i(x)| \leq \frac{\beta}{rc'\ell^2} \cdot \exp\left(-\frac{c'(\ell - k)}{\sqrt{n'm}}\right), \quad \ell > k, \quad (4.31)$$

$$\text{orth}(\phi_i - \phi_j) \geq c''\sqrt{\frac{n'}{m}} \quad \text{for all } i, j \in [r], \quad (4.32)$$

where $c', c'' \in (0, 1)$ are the absolute constants defined in Theorem 4.1. For $i \in [r]$, define $\tilde{\phi}_i : \{0, 1\}^{n'} \rightarrow \mathbb{R}$ by

$$\tilde{\phi}_i(x) = \phi_i(x) - \mathbf{I}[|x| > k] \min_{j \in [r]} \phi_j(x). \quad (4.33)$$

Equation (4.26) shows that $\tilde{\phi}_i$ is a linear combination of functions whose support is contained in $\{x \in \{0, 1\}^{n'} : |x| = k \text{ or } |x| \geq m\}$. As a result,

$$\text{supp } \tilde{\phi}_i \subseteq \{x \in \{0, 1\}^{n'} : |x| = k \text{ or } |x| \geq m\}. \quad (4.34)$$

Since $\tilde{\phi}_i = \phi_i$ on $\{0, 1\}^{n'}|_k$, we obtain from (4.27) and (4.29) that

$$\{0, 1\}^{n'}|_k \cap \text{supp } \tilde{\phi}_i = \{\mathbf{1}_S : S \in \gamma^{-1}(i)\}, \quad (4.35)$$

$$\sum_{x:|x|=k} \tilde{\phi}_i(x) = 1. \quad (4.36)$$

In particular,

$$\|\tilde{\phi}_i\|_1 \geq 1. \quad (4.37)$$

We further claim that

$$\tilde{\phi}_i(x) \geq 0, \quad x \in \{0, 1\}^{n'}. \quad (4.38)$$

Indeed, the nonnegativity of $\tilde{\phi}_i(x)$ for $x \in \{0, 1\}^{n'}|_k$ follows from $\tilde{\phi}_i(x) = \phi_i(x)$ and (4.28), whereas the nonnegativity of $\tilde{\phi}_i(x)$ for $x \in \{0, 1\}^{n'}|_{>k}$ follows from (4.33) via $\tilde{\phi}_i(x) = \phi_i(x) - \min_{j \in [r]} \phi_j(x) \geq \phi_i(x) - \phi_i(x) \geq 0$.

On $\{0, 1\}^{n'}|_{>k}$, we have

$$\tilde{\phi}_i = \phi_i - \min_{j \in [r]} \phi_j = \max_{j \in [r]} \{\phi_i - \phi_j\} \leq \max_{j \in [r]} |\phi_i - \phi_j| \leq \sum_{j=1}^r |\phi_j|.$$

This conclusion is also valid on $\{0, 1\}^{n'}|_{<k}$ due to (4.34). Thus,

$$\tilde{\phi}_i(x) \leq \sum_{j=1}^r |\phi_j(x)|, \quad |x| \neq k. \quad (4.39)$$

Summing over x gives

$$\begin{aligned} \sum_{x:|x| \neq k} \tilde{\phi}_i(x) &\leq \sum_{x:|x| \neq k} \sum_{j=1}^r |\phi_j(x)| \\ &= \sum_{j=1}^r \sum_{x:|x| \neq k} |\phi_j(x)| \\ &\leq \beta, \end{aligned} \quad (4.40)$$

where the third step applies (4.30).

For all $i \in [r]$ and $\ell \in \{m, m+1, \dots, n'\}$, we have the graded bound

$$\begin{aligned} \sum_{x:|x|=\ell} \tilde{\phi}_i(x) &\leq \sum_{j=1}^r \sum_{x:|x|=\ell} |\phi_j(x)| \\ &\leq \frac{1}{c'\ell^2} \cdot \exp\left(-\frac{c'(\ell-k)}{\sqrt{n'm}}\right), \end{aligned} \quad (4.41)$$

where the first step uses (4.39), and the second step uses (4.31). Finally, for $i, j \in [r]$, we have

$$\begin{aligned} \text{orth}(\tilde{\phi}_i - \tilde{\phi}_j) &= \text{orth}(\phi_i - \phi_j) \\ &\geq c'' \sqrt{\frac{n'}{m}}, \end{aligned} \quad (4.42)$$

where the first step uses the definition (4.33), and the second step uses (4.32).

Define $c = \min\{c', c''\}$. Equations (4.36) and (4.38) show that each $\tilde{\phi}_i$ is a nonnegative function and is not identically zero, making it possible to define a probability distribution λ_i on $\{0, 1\}^n$ by

$$\lambda_i(x) = \frac{1}{\|\tilde{\phi}_i\|_1} \tilde{\phi}_i(x_1 x_2 \dots x_{n'}) \prod_{j=n'+1}^n (1 - x_j).$$

In other words, λ_i is nonzero only on inputs x with $x_{n'+1} = x_{n'+2} = \dots = x_n = 0$, and on such inputs $\lambda_i(x)$ is the properly normalized version of the nonnegative function $\tilde{\phi}_i(x_1 x_2 \dots x_{n'})$. Then properties (4.21) and (4.22) are immediate from (4.34) and (4.35), respectively. Property (4.23) follows from

$$\begin{aligned} \lambda_i(\{0, 1\}^n |_k) &= \frac{1}{\|\tilde{\phi}_i\|_1} \sum_{x \in \{0, 1\}^n |_k} \tilde{\phi}_i(x_1 x_2 \dots x_{n'}) \prod_{j=n'+1}^n (1 - x_j) \\ &= \frac{1}{\|\tilde{\phi}_i\|_1} \sum_{x \in \{0, 1\}^{n'} |_k} \tilde{\phi}_i(x_1 x_2 \dots x_{n'}) \\ &= \frac{1}{\|\tilde{\phi}_i\|_1} \\ &\geq \frac{1}{1 + \beta} \\ &\geq 1 - \beta, \end{aligned}$$

where the third step uses (4.36), and the fourth step uses (4.36) and (4.40). Property (4.24) is trivial for $\ell = k$ and follows for $\ell > k$ from

$$\begin{aligned} \lambda_i(\{0, 1\}^n |_\ell) &= \frac{1}{\|\tilde{\phi}_i\|_1} \sum_{x \in \{0, 1\}^n |_\ell} \tilde{\phi}_i(x_1 x_2 \dots x_{n'}) \prod_{j=n'+1}^n (1 - x_j) \\ &= \frac{1}{\|\tilde{\phi}_i\|_1} \sum_{x \in \{0, 1\}^{n'} |_\ell} \tilde{\phi}_i(x_1 x_2 \dots x_{n'}) \\ &\leq \frac{1}{\|\tilde{\phi}_i\|_1} \cdot \frac{\exp(-c'(\ell - k)/\sqrt{n'm})}{c'\ell^2} \\ &\leq \frac{\exp(-c(\ell - k)/\sqrt{n'm})}{c\ell^2}, \end{aligned}$$

where the third step uses (4.41), and the fourth step uses (4.37) and $c = \min\{c', c''\}$.

It remains to verify (4.25). For this, fix $i, j \in [r]$ arbitrarily. Then $\|\tilde{\phi}_i\|_1 - \|\tilde{\phi}_j\|_1 = \langle \tilde{\phi}_i, 1 \rangle - \langle \tilde{\phi}_j, 1 \rangle = \langle \tilde{\phi}_i - \tilde{\phi}_j, 1 \rangle = 0$, where the first step uses (4.38), and the third step uses (4.42). We thus see that

$$\|\tilde{\phi}_i\|_1 = \|\tilde{\phi}_j\|_1. \tag{4.43}$$

Next, observe that λ_i can be written as the product of two functions on disjoint sets of variables, and likewise for λ_j . Namely,

$$\begin{aligned}\lambda_i &= \frac{1}{\|\tilde{\phi}_i\|_1} \tilde{\phi}_i \otimes \text{NOR}_{n-n'}, \\ \lambda_j &= \frac{1}{\|\tilde{\phi}_j\|_1} \tilde{\phi}_j \otimes \text{NOR}_{n-n'}.\end{aligned}$$

Now

$$\begin{aligned}\text{orth}(\lambda_i - \lambda_j) &= \text{orth}\left(\left(\frac{\tilde{\phi}_i}{\|\tilde{\phi}_i\|_1} - \frac{\tilde{\phi}_j}{\|\tilde{\phi}_j\|_1}\right) \otimes \text{NOR}_{n-n'}\right) \\ &\geq \text{orth}\left(\frac{\tilde{\phi}_i}{\|\tilde{\phi}_i\|_1} - \frac{\tilde{\phi}_j}{\|\tilde{\phi}_j\|_1}\right) \\ &= \text{orth}\left(\frac{\tilde{\phi}_i - \tilde{\phi}_j}{\|\tilde{\phi}_i\|_1}\right) \\ &= \text{orth}(\tilde{\phi}_i - \tilde{\phi}_j) \\ &\geq c'' \sqrt{\frac{n'}{m}},\end{aligned}$$

where the second step uses Proposition 2.6(ii), the third step applies (4.43), and the last step is justified by (4.42). In view of $c = \min\{c', c''\}$, this settles (4.25) and completes the proof. \square

4.3. Hardness amplification for approximate degree. We have reached the crux of our proof, a hardness amplification theorem for approximate degree. Unlike previous work, our hardness amplification is directly applicable to Boolean functions with sparse input and does not use componentwise composition or input compression. The theorem statement below has a large number of parameters, for maximum generality and black-box integration with the auxiliary results of previous sections. We will later derive a succinct and easy-to-apply corollary that will suffice for our hardness amplification purposes.

THEOREM 4.5. *Let $C^* \geq 1$ and $c \in (0, 1)$ be the absolute constants from Theorems 3.6 and 4.4, respectively. Fix a real number $0 < \beta < 1$ and positive integers n, m, k, N, θ, D, T such that*

$$n/2 \geq m > k, \tag{4.44}$$

$$\begin{aligned}4(N+1)^2 k \exp\left(-\frac{\sqrt{m}}{16k}\right) + (N+1) \left(\frac{3C^* \log^2(n+N+1)}{m^{1/4}}\right)^k \\ \leq \frac{\beta}{16(N+1)^2 m^2},\end{aligned} \tag{4.45}$$

$$T \geq \frac{8e}{c} \cdot \theta(1 + \ln \theta) + \theta k, \tag{4.46}$$

$$T \geq D. \tag{4.47}$$

Define

$$\Delta = \left(1 + 2^D \binom{n\theta}{D}\right) \exp\left(-\frac{c(T - \theta k)}{2\sqrt{nm}}\right). \quad (4.48)$$

Then there is an (explicitly given) mapping $H: (\{0, 1\}^n)^\theta \rightarrow \{0, 1\}^N$ such that:

- (i) each output bit of H is computable by a monotone $(k+1)$ -DNF formula;
- (ii) for every $\epsilon \in [0, 1]$ and every $f: \{0, 1\}^N \rightarrow \{0, 1\}$, one has

$$\deg_{\epsilon - \beta\theta - 2\Delta}((f \circ H)|_{\leq T}) \geq \min\left\{c \deg_\epsilon(f|_{\leq \theta}) \sqrt{\frac{n}{2m}}, D\right\}.$$

Proof. We may assume that

$$\epsilon - \beta\theta - 2\Delta \geq 0 \quad (4.49)$$

since otherwise the left-hand side in the approximate degree lower bound of (ii) is by definition $+\infty$. Define $V \subseteq \mathbb{R}^N$ by $V = \{0^N, e_1, e_2, \dots, e_N\}$ and set $r = N+1$. In view of (4.44) and (4.45), Corollary 3.10 gives an explicit integer $n' \in (n/2, n]$ and an explicit $(\frac{\beta}{16r^2m^2}, \frac{\beta}{16r^2m^2}, m)$ -balanced coloring $\gamma: \binom{[n']}{k} \rightarrow [r]$. Alternatively, if one is not concerned about explicitness, the existence of γ can be deduced from the much simpler Corollary 3.3. Specifically, (4.45) forces $\sqrt{m} \geq k$ and in particular $n \geq m \geq k^2 \geq 1$. Moreover, (4.45) implies that $3r\sqrt{k \ln(n+1)}/m^{k/4} \leq \frac{\beta}{16r^2m^2}$. Now Corollary 3.3 guarantees the existence of a $(\frac{\beta}{16r^2m^2}, \frac{\beta}{16r^2m^2}, m)$ -balanced coloring $\gamma: \binom{[n']}{k} \rightarrow [r]$.

Since $n' \geq m > k$, Theorem 4.4 gives explicit distributions $\lambda_{0^N}, \lambda_{e_1}, \lambda_{e_2}, \dots, \lambda_{e_N}$ on $\{0, 1\}^n$ such that

$$\text{supp } \lambda_v \subseteq \{x \in \{0, 1\}^n : |x| = k \text{ or } |x| \geq m\}, \quad v \in V, \quad (4.50)$$

$$\{0, 1\}^n|_k \cap \text{supp } \lambda_{e_i} = \{\mathbf{1}_S : S \in \gamma^{-1}(i)\}, \quad i \in [N], \quad (4.51)$$

$$\{0, 1\}^n|_k \cap \text{supp } \lambda_{0^N} = \{\mathbf{1}_S : S \in \gamma^{-1}(N+1)\}, \quad (4.52)$$

$$\lambda_v(\{0, 1\}^n|_k) \geq 1 - \beta, \quad v \in V, \quad (4.53)$$

$$\lambda_v(\{0, 1\}^n|_t) \leq \frac{\exp(-c(t-k)/\sqrt{nm})}{c(t-k+1)^2}, \quad v \in V, t \geq k, \quad (4.54)$$

$$\text{orth}(\lambda_v - \lambda_u) \geq c\sqrt{\frac{n}{2m}}, \quad v, u \in V. \quad (4.55)$$

Properties (4.51) and (4.52) imply that

$$\{0, 1\}^n|_k \cap \text{supp } \lambda_u \cap \text{supp } \lambda_v = \emptyset, \quad u, v \in V, u \neq v. \quad (4.56)$$

For $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_\theta) \in V^\theta$, define

$$\Lambda_{\mathbf{v}} = \bigotimes_{i=1}^{\theta} \lambda_{\mathbf{v}_i}. \quad (4.57)$$

CLAIM 4.6. For each $\mathbf{v} \in V^\theta$, there is a function $\widetilde{\Lambda}_{\mathbf{v}}: (\{0, 1\}^n)^\theta \rightarrow \mathbb{R}$ such that

$$\text{supp } \widetilde{\Lambda}_{\mathbf{v}} \subseteq (\{0, 1\}^n)^\theta|_{\leq T}, \quad (4.58)$$

$$\text{orth}(\Lambda_{\mathbf{v}} - \widetilde{\Lambda}_{\mathbf{v}}) > D, \quad (4.59)$$

$$\|\Lambda_{\mathbf{v}} - \widetilde{\Lambda}_{\mathbf{v}}\|_1 \leq \Delta. \quad (4.60)$$

We will settle Claim 4.6, and all other claims, after the proof of the theorem.

We now turn to the construction of the monotone mapping H in the theorem statement. Define $h: \{0, 1\}^n \rightarrow \{0, 1\}^N$ by

$$(h(z))_j = \bigvee_{S \in \binom{[n]}{k+1} \cup \gamma^{-1}(j)} \bigwedge_{s \in S} z_s, \quad j = 1, 2, \dots, N. \quad (4.61)$$

Clearly, this is a monotone DNF formula of width $k + 1$. Define $H: (\{0, 1\}^n)^\theta \rightarrow \{0, 1\}^N$ by

$$H(x_1, x_2, \dots, x_\theta) = \bigvee_{i=1}^{\theta} h(x_i), \quad x_1, x_2, \dots, x_\theta \in \{0, 1\}^n, \quad (4.62)$$

where the right-hand side is the componentwise disjunction of the Boolean vectors $h(x_1), h(x_2), \dots, h(x_\theta)$. Observe that both h and H are monotone and are given explicitly in closed form in terms of the coloring γ constructed at the beginning of the proof. This settles (i).

For (ii), fix an arbitrary function $f: \{0, 1\}^N \rightarrow \{0, 1\}$ and abbreviate

$$d = \deg_\epsilon(f|_{\leq \theta}).$$

By the dual characterization of approximate degree (Fact 2.8), there is a function $\psi: \{0, 1\}^N|_{\leq \theta} \rightarrow \mathbb{R}$ such that

$$\|\psi\|_1 = 1, \quad (4.63)$$

$$\langle f, \psi \rangle > \epsilon, \quad (4.64)$$

$$\text{orth } \psi \geq d. \quad (4.65)$$

Define $\Psi: (\{0, 1\}^n)^\theta \rightarrow \mathbb{R}$ by

$$\Psi = \sum_{u \in \{0, 1\}^N|_{\leq \theta}} \psi(u) \mathbf{E}_{\substack{\mathbf{v} \in V^\theta: \\ \mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_\theta = u}} \widetilde{\Lambda}_{\mathbf{v}}. \quad (4.66)$$

We will now use (4.44)–(4.66) to prove a sequence of claims.

CLAIM 4.7. One has

$$\text{supp } \Psi \subseteq (\{0, 1\}^n)^\theta|_{\leq T}, \quad (4.67)$$

$$\|\Psi\|_1 \leq 1 + \Delta, \quad (4.68)$$

$$\text{orth } \Psi \geq \min \left\{ cd \sqrt{\frac{n}{2m}}, D \right\}. \quad (4.69)$$

CLAIM 4.8. Let $v \in V$ be given. Then for all $z \in \{0, 1\}^n|_k \cap \text{supp } \lambda_v$, one has $h(z) = v$.

CLAIM 4.9. Let $u \in \{0, 1\}^N|_{\leq \theta}$ and $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_\theta) \in V^\theta$ be given such that $\mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_\theta = u$. Then

$$|f(u) - \langle \widetilde{\Lambda}_{\mathbf{v}}, f \circ H \rangle| \leq \beta\theta + \Delta. \quad (4.70)$$

CLAIM 4.10. One has

$$\langle f \circ H, \Psi \rangle > (\epsilon - \beta\theta - 2\Delta) \|\Psi\|_1. \quad (4.71)$$

Note from (4.67) that Ψ is supported on inputs of Hamming weight at most T and can therefore be regarded as a function on $(\{0, 1\}^n)^\theta|_{\leq T}$. Now the claimed bound in (ii) follows by Fact 2.8 in view of (4.69) and (4.71). The proof of the theorem is complete. \square

Proof of Claim 4.6. Equations (4.46), (4.50), and (4.54) ensure that Lemma 2.5 is applicable to the distributions $\lambda_{\mathbf{v}_1}, \lambda_{\mathbf{v}_2}, \dots, \lambda_{\mathbf{v}_\theta}$ with parameters $\ell = \theta$, $B = n$, $C = 1/c$, and $\alpha = \exp(-c/\sqrt{nm})$, whence

$$\Lambda_{\mathbf{v}}((\{0, 1\}^n)^\theta|_{> T}) \leq \exp\left(-\frac{c(T - \theta k)}{2\sqrt{nm}}\right), \quad \mathbf{v} \in V^\theta.$$

In view of (4.47), we can now invoke Lemma 2.13 with parameter $B = n\theta$ to obtain a function $\widetilde{\Lambda}_{\mathbf{v}}: (\{0, 1\}^n)^\theta \rightarrow \mathbb{R}$ that satisfies (4.58)–(4.60). \square

Proof of Claim 4.7. Observe from (4.58) that Ψ is a linear combination of functions supported on inputs of Hamming weight at most T . This settles the support property (4.67). Property (4.68) can be verified as follows:

$$\begin{aligned} \|\Psi\|_1 &\leq \sum_{u \in \{0, 1\}^N|_{\leq \theta}} |\psi(u)| \mathbf{E}_{\substack{\mathbf{v} \in V^\theta: \\ \mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_\theta = u}} \|\widetilde{\Lambda}_{\mathbf{v}}\|_1 \\ &\leq \left(\sum_{u \in \{0, 1\}^N|_{\leq \theta}} |\psi(u)| \right) \max_{\mathbf{v} \in V^\theta} \|\widetilde{\Lambda}_{\mathbf{v}}\|_1 \\ &= \|\psi\|_1 \max_{\mathbf{v} \in V^\theta} \|\widetilde{\Lambda}_{\mathbf{v}}\|_1 \\ &\leq \|\psi\|_1 \max_{\mathbf{v} \in V^\theta} \{\|\Lambda_{\mathbf{v}}\|_1 + \|\widetilde{\Lambda}_{\mathbf{v}} - \Lambda_{\mathbf{v}}\|_1\} \\ &\leq 1 + \Delta, \end{aligned}$$

where the first and fourth steps apply the triangle inequality, and the last step uses (4.60) and (4.63).

To settle (4.69), consider an arbitrary polynomial $P: (\{0, 1\}^n)^\theta \rightarrow \mathbb{R}$ of degree less than $\min\{cd\sqrt{n/(2m)}, D\}$. Then

$$\begin{aligned} \langle \Psi, P \rangle &= \sum_{u \in \{0, 1\}^N|_{\leq \theta}} \psi(u) \mathbf{E}_{\substack{\mathbf{v} \in V^\theta: \\ \mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_\theta = u}} \langle \widetilde{\Lambda}_{\mathbf{v}}, P \rangle \\ &= \sum_{u \in \{0, 1\}^N|_{\leq \theta}} \psi(u) \mathbf{E}_{\substack{\mathbf{v} \in V^\theta: \\ \mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_\theta = u}} [\langle \Lambda_{\mathbf{v}}, P \rangle + \langle \widetilde{\Lambda}_{\mathbf{v}} - \Lambda_{\mathbf{v}}, P \rangle] \\ &= \sum_{u \in \{0, 1\}^N|_{\leq \theta}} \psi(u) \mathbf{E}_{\substack{\mathbf{v} \in V^\theta: \\ \mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_\theta = u}} \langle \Lambda_{\mathbf{v}}, P \rangle, \end{aligned} \quad (4.72)$$

where the first and second steps use the linearity of inner product, and the third step is valid by (4.59). Equation (4.55) allows us to invoke Proposition 2.7 with $\ell = \theta$ and $\phi_v = \lambda_v$ to infer that the inner product $\langle \Lambda_{\mathbf{v}}, P \rangle$ is a polynomial in \mathbf{v} of degree less than d . As a result, Fact 2.15 implies that the expected value in (4.72) is a polynomial in u of degree less than d . In summary, (4.72) is the inner product of ψ with a polynomial of degree less than d and is therefore zero by (4.65). The proof of (4.69) is complete. \square

Proof of Claim 4.8. Consider an arbitrary string $z \in \{0, 1\}^n|_k$. Then

$$(h(z))_j = \bigvee_{S \in \gamma^{-1}(j)} \bigwedge_{s \in S} z_s = \mathbf{I}[z \in \text{supp } \lambda_{e_j}],$$

where the first step uses the defining equation (4.61) together with $|z| = k$, and the second step applies (4.51) along with $|z| = k$. Thus, $h(z)$ can be written out explicitly as

$$h(z) = (\mathbf{I}[z \in \text{supp } \lambda_{e_1}], \mathbf{I}[z \in \text{supp } \lambda_{e_2}], \dots, \mathbf{I}[z \in \text{supp } \lambda_{e_N}]). \quad (4.73)$$

Now recall from (4.56) that a string z of Hamming weight k can belong to at most one of the sets $\text{supp } \lambda_{0^N}, \text{supp } \lambda_{e_1}, \text{supp } \lambda_{e_2}, \dots, \text{supp } \lambda_{e_N}$. As a result, if $z \in \text{supp } \lambda_{e_i}$ then $z \notin \text{supp } \lambda_{e_j}$ for all $j \neq i$ and consequently $h(z) = e_i$ by (4.73). Analogously, if $z \in \text{supp } \lambda_{0^N}$ then $z \notin \text{supp } \lambda_{e_j}$ for all j and consequently $h(z) = 0^N$ by (4.73). This settles the claim for all $v \in V$. \square

Proof of Claim 4.9. Since u is a Boolean vector, the equality $\mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_\theta = u$ forces

$$\mathbf{v}_1 \vee \mathbf{v}_2 \vee \dots \vee \mathbf{v}_\theta = u, \quad (4.74)$$

where the disjunction is applied componentwise. For any input $(x_1, x_2, \dots, x_\theta)$ where $x_i \in \{0, 1\}^n|_k \cap \text{supp } \lambda_{\mathbf{v}_i}$, we have

$$(f \circ H)(x_1, x_2, \dots, x_\theta) = f \left(\bigvee_{i=1}^{\theta} h(x_i) \right) = f \left(\bigvee_{i=1}^{\theta} \mathbf{v}_i \right) = f(u),$$

where the second and third steps use Claim 4.8 and (4.74), respectively. Since $\text{supp } \Lambda_{\mathbf{v}} = \prod_{i=1}^{\theta} \text{supp } \lambda_{\mathbf{v}_i}$, we have shown that

$$f \circ H \equiv f(u) \quad \text{on } (\{0, 1\}^n|_k)^\theta \cap \text{supp } \Lambda_{\mathbf{v}}. \quad (4.75)$$

Furthermore,

$$\Lambda_{\mathbf{v}}((\{0, 1\}^n|_k)^\theta) = \prod_{i=1}^{\theta} \lambda_{\mathbf{v}_i}(\{0, 1\}^n|_k) \geq (1 - \beta)^\theta \geq 1 - \beta\theta, \quad (4.76)$$

where the second step uses (4.53). Now

$$\begin{aligned}
& |f(u) - \langle \widetilde{\Lambda}_{\mathbf{v}}, f \circ H \rangle| \\
& \leq |f(u) - \langle \Lambda_{\mathbf{v}}, f \circ H \rangle| + |\langle \Lambda_{\mathbf{v}} - \widetilde{\Lambda}_{\mathbf{v}}, f \circ H \rangle| \\
& \leq |f(u) - \langle \Lambda_{\mathbf{v}}, f \circ H \rangle| + \|\Lambda_{\mathbf{v}} - \widetilde{\Lambda}_{\mathbf{v}}\|_1 \\
& = \left| f(u) - \mathbf{E}_{\Lambda_{\mathbf{v}}} f \circ H \right| + \|\Lambda_{\mathbf{v}} - \widetilde{\Lambda}_{\mathbf{v}}\|_1 \\
& \leq \mathbf{E}_{\Lambda_{\mathbf{v}}} |f(u) - f \circ H| + \|\Lambda_{\mathbf{v}} - \widetilde{\Lambda}_{\mathbf{v}}\|_1 \\
& \leq 0 \cdot \Lambda_{\mathbf{v}}(\{0, 1\}^n |_k)^\theta + 1 \cdot \Lambda_{\mathbf{v}}(\overline{\{0, 1\}^n |_k})^\theta + \|\Lambda_{\mathbf{v}} - \widetilde{\Lambda}_{\mathbf{v}}\|_1 \\
& \leq \beta\theta + \|\Lambda_{\mathbf{v}} - \widetilde{\Lambda}_{\mathbf{v}}\|_1 \\
& \leq \beta\theta + \Delta,
\end{aligned}$$

where the last three steps use (4.75), (4.76), and (4.60), respectively. \square

Proof of Claim 4.10. To begin with,

$$\begin{aligned}
\langle f, \psi \rangle - \langle f \circ H, \Psi \rangle &= \sum_{u \in \{0, 1\}^N |_{\leq \theta}} \psi(u) f(u) \\
&\quad - \sum_{u \in \{0, 1\}^N |_{\leq \theta}} \psi(u) \mathbf{E}_{\substack{\mathbf{v} \in V^\theta: \\ \mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_\theta = u}} \langle \widetilde{\Lambda}_{\mathbf{v}}, f \circ H \rangle \\
&= \sum_{u \in \{0, 1\}^N |_{\leq \theta}} \psi(u) \mathbf{E}_{\substack{\mathbf{v} \in V^\theta: \\ \mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_\theta = u}} [f(u) - \langle \widetilde{\Lambda}_{\mathbf{v}}, f \circ H \rangle] \\
&\leq \sum_{u \in \{0, 1\}^N |_{\leq \theta}} |\psi(u)| \mathbf{E}_{\substack{\mathbf{v} \in V^\theta: \\ \mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_\theta = u}} |f(u) - \langle \widetilde{\Lambda}_{\mathbf{v}}, f \circ H \rangle| \\
&\leq \|\psi\|_1 \max_{u \in \{0, 1\}^N |_{\leq \theta}} \max_{\substack{\mathbf{v} \in V^\theta: \\ \mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_\theta = u}} |f(u) - \langle \widetilde{\Lambda}_{\mathbf{v}}, f \circ H \rangle| \\
&\leq \|\psi\|_1 (\beta\theta + \Delta) \\
&= \beta\theta + \Delta, \tag{4.77}
\end{aligned}$$

where the last two steps use Claim 4.9 and (4.63), respectively. Then

$$\begin{aligned}
\langle f \circ H, \Psi \rangle &> \epsilon - \beta\theta - \Delta \\
&\geq \frac{\epsilon - \beta\theta - \Delta}{1 + \Delta} \cdot \|\Psi\|_1 \\
&\geq (\epsilon - \beta\theta - 2\Delta) \|\Psi\|_1,
\end{aligned}$$

where the first step uses (4.64) and (4.77), the second step is justified by (4.49) and (4.68), and the third step is legitimate since $a/(1+b) \geq a-b$ for all $a \in [0, 1]$ and $b \geq 0$. This completes the proof of (4.71). \square

4.4. Hardness amplification for one-sided approximate degree. In this section, we will prove that the construction of Theorem 4.5 amplifies not only approximate degree but also its one-sided variant. We start with a technical lemma.

LEMMA 4.11. *Let n, m, k, θ, D, T be positive integers with*

$$T \geq n + D, \quad (4.78)$$

$$T \geq \theta k. \quad (4.79)$$

Let $y \in (\{0, 1\}^n)_{>T}^\theta$ be given. Then there exists $\zeta_y: (\{0, 1\}^n)^\theta \rightarrow \mathbb{R}$ such that

$$\text{supp } \zeta_y \subseteq (\{0, 1\}^n)_{\leq T}^\theta \cup \{y\}, \quad (4.80)$$

$$\zeta_y(y) = 1, \quad (4.81)$$

$$\text{orth } \zeta_y > D, \quad (4.82)$$

$$\|\zeta_y\|_1 \leq 1 + 2^D \binom{n(\theta-1)}{D}, \quad (4.83)$$

$$\zeta_y = 0 \quad \text{on} \quad (\{0, 1\}^n)_{\leq k}^\theta. \quad (4.84)$$

Proof. It follows from (4.79) that $y = (y_1, y_2, \dots, y_\theta)$ has a coordinate with Hamming weight greater than k . By symmetry, we may assume that

$$|y_1| > k. \quad (4.85)$$

We have $|y_2 y_3 \dots y_\theta| = |y| - |y_1| > T - n \geq D$, where the second step uses the hypothesis $|y| > T$ along with the trivial bound $|y_1| \leq n$, whereas the third step is legitimate by (4.78). Thanks to the newly obtained inequality $|y_2 y_3 \dots y_\theta| > D$, Lemma 2.12 is applicable with $B = n(\theta - 1)$ and gives a function $\zeta: (\{0, 1\}^n)^{\theta-1} \rightarrow \mathbb{R}$ such that

$$\text{supp } \zeta \subseteq (\{0, 1\}^n)^{\theta-1}_{\leq D} \cup \{y_2 y_3 \dots y_\theta\}, \quad (4.86)$$

$$\zeta(y_2 y_3 \dots y_\theta) = 1, \quad (4.87)$$

$$\|\zeta\|_1 \leq 1 + 2^D \binom{n(\theta-1)}{D}, \quad (4.88)$$

$$\text{orth } \zeta > D. \quad (4.89)$$

We will prove that the claimed properties (4.80)–(4.84) are enjoyed by the function

$$\zeta_y(x) = \delta_{x_1, y_1} \zeta(x_2 x_3 \dots x_\theta).$$

To verify the support property (4.80), fix any x with $\zeta_y(x) \neq 0$. Then necessarily $\delta_{x_1, y_1} = 1$, forcing $x_1 = y_1$. Now (4.86) implies that x either equals y or has Hamming weight at most $|y_1| + D$. Since $|y_1| + D \leq n + D \leq T$ by (4.78), this completes the proof of (4.80).

The remaining properties are straightforward. Property (4.81) follows from the corresponding property (4.87) of ζ . Likewise, property (4.82) follows from (4.89) in light of Proposition 2.6 (ii). Property (4.83) is immediate from (4.88). Finally, (4.84) is a consequence of (4.85). \square

We are now ready to state and prove our hardness amplification result, which is a far-reaching generalization of Theorem 4.5.

THEOREM 4.12. *Let $C^* \geq 1$ and $c \in (0, 1)$ be the absolute constants from Theorems 3.6 and 4.4, respectively. Fix a real number $0 < \beta < 1$ and positive integers*

n, m, k, N, θ, D, T such that

$$n/2 \geq m > k, \quad (4.90)$$

$$4(N+1)^2 k \exp\left(-\frac{\sqrt{m}}{16k}\right) + (N+1) \left(\frac{3C^* \log^2(n+N+1)}{m^{1/4}}\right)^k \leq \frac{\beta}{16(N+1)^2 m^2}, \quad (4.91)$$

$$T \geq \frac{8e}{c} \cdot \theta(1 + \ln \theta) + \theta k, \quad (4.92)$$

$$T \geq D + n. \quad (4.93)$$

Define

$$\Delta = \left(1 + 2^D \binom{n\theta}{D}\right) \exp\left(-\frac{c(T - \theta k)}{2\sqrt{nm}}\right). \quad (4.94)$$

Then there is an (explicitly given) mapping $H: \{0, 1\}^\theta \rightarrow \{0, 1\}^N$ such that:

- (i) each output bit of H is computable by a monotone $(k+1)$ -DNF formula;
- (ii) for every $\epsilon \in [0, 1]$ and every $f: \{0, 1\}^N \rightarrow \{0, 1\}$, one has

$$\deg_{\epsilon - \beta\theta - 2\Delta}((f \circ H)|_{\leq T}) \geq \min\left\{c \deg_\epsilon(f|_{\leq \theta}) \sqrt{\frac{n}{2m}}, D\right\};$$

- (iii) for every $\epsilon \in [0, 1]$ and every $f: \{0, 1\}^N \rightarrow \{0, 1\}$ with $f(1^N) = 0$, one has

$$\deg_{\epsilon - \beta\theta - 2\Delta}^+((f \circ H)|_{\leq T}) \geq \min\left\{c \deg_\epsilon^+(f|_{\leq \theta}) \sqrt{\frac{n}{2m}}, D\right\}.$$

Proof. As in the proof of Theorem 4.5, we may assume that

$$\epsilon - \beta\theta - 2\Delta \geq 0 \quad (4.95)$$

since otherwise the left-hand side in the lower bounds of (ii) and (iii) is by definition $+\infty$. Define $V \subseteq \mathbb{R}^N$ by $V = \{0^N, e_1, e_2, \dots, e_N\}$ and set $r = N + 1$. Arguing as in the proof of Theorem 4.5, we obtain an explicit integer $n' \in (n/2, n]$ and an explicit $(\frac{\beta}{16r^2 m^2}, \frac{\beta}{16r^2 m^2}, m)$ -balanced coloring $\gamma: \binom{[n']}{k} \rightarrow [r]$, which in turn results in explicit distributions $\lambda_{0^N}, \lambda_{e_1}, \lambda_{e_2}, \dots, \lambda_{e_N}$ on $\{0, 1\}^n$ such that

$$\text{supp } \lambda_v \subseteq \{x \in \{0, 1\}^n : |x| = k \text{ or } |x| \geq m\}, \quad v \in V, \quad (4.96)$$

$$\{0, 1\}^n|_k \cap \text{supp } \lambda_{e_i} = \{\mathbf{1}_S : S \in \gamma^{-1}(i)\}, \quad i \in [N], \quad (4.97)$$

$$\{0, 1\}^n|_k \cap \text{supp } \lambda_{0^N} = \{\mathbf{1}_S : S \in \gamma^{-1}(N+1)\}, \quad (4.98)$$

$$\lambda_v(\{0, 1\}^n|_k) \geq 1 - \beta, \quad v \in V, \quad (4.99)$$

$$\lambda_v(\{0, 1\}^n|_t) \leq \frac{\exp(-c(t-k)/\sqrt{nm})}{c(t-k+1)^2}, \quad v \in V, t \geq k, \quad (4.100)$$

$$\text{orth}(\lambda_v - \lambda_u) \geq c\sqrt{\frac{n}{2m}}, \quad v, u \in V. \quad (4.101)$$

Properties (4.97) and (4.98) imply that

$$\{0, 1\}^n|_k \cap \text{supp } \lambda_u \cap \text{supp } \lambda_v = \emptyset, \quad u, v \in V, u \neq v. \quad (4.102)$$

For $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_\theta) \in V^\theta$, define

$$\Lambda_{\mathbf{v}} = \bigotimes_{i=1}^{\theta} \lambda_{\mathbf{v}_i}. \quad (4.103)$$

CLAIM 4.13. *For each $\mathbf{v} \in V^\theta$, there is a function $\widetilde{\Lambda}_{\mathbf{v}}: (\{0, 1\}^n)^\theta \rightarrow \mathbb{R}$ such that*

$$\text{supp } \widetilde{\Lambda}_{\mathbf{v}} \subseteq (\{0, 1\}^n)^\theta|_{\leq T}, \quad (4.104)$$

$$\text{orth}(\Lambda_{\mathbf{v}} - \widetilde{\Lambda}_{\mathbf{v}}) > D, \quad (4.105)$$

$$\|\Lambda_{\mathbf{v}} - \widetilde{\Lambda}_{\mathbf{v}}\|_1 \leq \Delta, \quad (4.106)$$

$$\widetilde{\Lambda}_{\mathbf{v}} = \Lambda_{\mathbf{v}} \quad \text{on } (\{0, 1\}^n|_{\leq k})^\theta. \quad (4.107)$$

We will settle Claim 4.13 after the proof of the theorem. We now define the monotone mapping H exactly the same way as in the proof of Theorem 4.5. Specifically, define $h: \{0, 1\}^n \rightarrow \{0, 1\}^N$ by

$$(h(z))_j = \bigvee_{S \in \binom{[n]}{k+1} \cup \gamma^{-1}(j)} \bigwedge_{s \in S} z_s, \quad j = 1, 2, \dots, N. \quad (4.108)$$

Define $H: (\{0, 1\}^n)^\theta \rightarrow \{0, 1\}^N$ by

$$H(x_1, x_2, \dots, x_\theta) = \bigvee_{i=1}^{\theta} h(x_i), \quad x_1, x_2, \dots, x_\theta \in \{0, 1\}^n, \quad (4.109)$$

where the right-hand side is the componentwise disjunction of the Boolean vectors $h(x_1), h(x_2), \dots, h(x_\theta)$. With these definitions, items (i) and (ii) are immediate because they are restatements of Theorem 4.5 (i), (ii). To prove the remaining item (iii), fix an arbitrary function $f: \{0, 1\}^N \rightarrow \{0, 1\}$ with

$$f(1^N) = 0, \quad (4.110)$$

and abbreviate

$$d = \text{deg}_\epsilon^+(f|_{\leq \theta}).$$

By the dual characterization of one-sided approximate degree (Fact 2.9), there is a function $\psi: \{0, 1\}^N|_{\leq \theta} \rightarrow \mathbb{R}$ such that

$$\|\psi\|_1 = 1, \quad (4.111)$$

$$\langle f, \psi \rangle > \epsilon, \quad (4.112)$$

$$\text{orth } \psi \geq d, \quad (4.113)$$

$$\psi(x) \geq 0 \quad \text{whenever } f(x) = 1. \quad (4.114)$$

Define $\Psi: (\{0, 1\}^n)^\theta \rightarrow \mathbb{R}$ by

$$\Psi = \sum_{u \in \{0, 1\}^N|_{\leq \theta}} \psi(u) \mathbf{E}_{\substack{\mathbf{v} \in V^\theta: \\ \mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_\theta = u}} \widetilde{\Lambda}_{\mathbf{v}}. \quad (4.115)$$

Equations (4.90)–(4.115) subsume the corresponding equations (4.44)–(4.66) in the proof of Theorem 4.5. Recall that from (4.44)–(4.66), we deduced Claims 4.7–4.10. As a result, Claims 4.7–4.10 remain valid here as well. In particular, we have

$$\text{supp } \Psi \subseteq (\{0, 1\}^n)^\theta|_{\leq T}, \quad (4.116)$$

$$\text{orth } \Psi \geq \min \left\{ cd\sqrt{\frac{n}{2m}}, D \right\}, \quad (4.117)$$

$$h \equiv v \quad \text{on } \{0, 1\}^n|_k \cap \text{supp } \lambda_v \quad (v \in V), \quad (4.118)$$

$$\langle f \circ H, \Psi \rangle > (\epsilon - \beta\theta - 2\Delta)\|\Psi\|_1. \quad (4.119)$$

Moreover, we will shortly prove the following new claim.

CLAIM 4.14. $\Psi(x) \geq 0$ whenever $(f \circ H)(x) = 1$.

The lower bound on the one-sided approximate degree in (iii) now follows from the dual characterization of one-sided approximate degree (Fact 2.9) in view of (4.116)–(4.119) and Claim 4.14. This completes the proof of Theorem 4.12. \square

Proof of Claim 4.13. Fix $\mathbf{v} \in V^\theta$ arbitrarily for the remainder of the proof. Equations (4.92), (4.96), and (4.100) ensure that Lemma 2.5 is applicable to the distributions $\lambda_{\mathbf{v}_1}, \lambda_{\mathbf{v}_2}, \dots, \lambda_{\mathbf{v}_\theta}$ with parameters $\ell = \theta$, $B = n$, $C = 1/c$, and $\alpha = \exp(-c/\sqrt{nm})$, whence

$$\Lambda_{\mathbf{v}}((\{0, 1\}^n)^\theta|_{>T}) \leq \exp\left(-\frac{c(T - \theta k)}{2\sqrt{nm}}\right). \quad (4.120)$$

Recall from (4.92) and (4.93) that $T \geq D + n$ and $T \geq \theta k$, which makes Lemma 4.11 applicable. Define $\widetilde{\Lambda}_{\mathbf{v}}: (\{0, 1\}^n)^\theta \rightarrow \mathbb{R}$ by

$$\widetilde{\Lambda}_{\mathbf{v}} = \Lambda_{\mathbf{v}} - \sum_{y \in (\{0, 1\}^n)^\theta|_{>T}} \Lambda_{\mathbf{v}}(y)\zeta_y, \quad (4.121)$$

where ζ_y is as given by Lemma 4.11. To verify the support property (4.104), fix any input x of Hamming weight $|x| > T$. For all y in the summation with $y \neq x$, we have $\zeta_y(x) = 0$ in view of (4.80). As a result, (4.121) simplifies to $\widetilde{\Lambda}_{\mathbf{v}}(x) = \Lambda_{\mathbf{v}}(x) - \Lambda_{\mathbf{v}}(x)\zeta_x(x)$. In view of (4.81), we conclude that $\widetilde{\Lambda}_{\mathbf{v}}(x) = 0$.

The orthogonality property (4.105) follows from

$$\text{orth}(\Lambda_{\mathbf{v}} - \widetilde{\Lambda}_{\mathbf{v}}) \geq \min_{y \in (\{0, 1\}^n)^\theta|_{>T}} \text{orth } \zeta_y > D,$$

where the first step uses the defining equation (4.121) and Proposition 2.6 (i), and the second step is legitimate by (4.82).

Property (4.106) can be verified as follows:

$$\begin{aligned}
\|\Lambda_{\mathbf{v}} - \widetilde{\Lambda}_{\mathbf{v}}\|_1 &\leq \sum_{y \in (\{0,1\}^n)^\theta |_{>T}} \Lambda_{\mathbf{v}}(y) \|\zeta_y\|_1 \\
&\leq \left(1 + 2^D \binom{n(\theta-1)}{D}\right) \sum_{y \in (\{0,1\}^n)^\theta |_{>T}} \Lambda_{\mathbf{v}}(y) \\
&\leq \left(1 + 2^D \binom{n(\theta-1)}{D}\right) \exp\left(-\frac{c(T-\theta k)}{2\sqrt{nm}}\right) \\
&\leq \Delta,
\end{aligned}$$

where the first step uses the triangle inequality along with the defining equation (4.121), the second step applies (4.83), the third step is valid by (4.120), and the fourth step uses the definition (4.94).

Finally, (4.107) follows from the definition (4.121) in view of (4.84). \square

Proof of Claim 4.14. We will prove the claim in contrapositive form. Specifically, fix an arbitrary string $x = (x_1, x_2, \dots, x_\theta) \in (\{0,1\}^n)^\theta$ with $\Psi(x) < 0$. Our objective is to deduce that $(f \circ H)(x) = 0$.

There are two cases to consider. If $|x_i| > k$ for some i , then the defining equation (4.108) implies that $h(x_i) = 1^N$. As a result,

$$(f \circ H)(x) = f(H(x)) = f\left(\bigvee_{i=1}^{\theta} h(x_i)\right) = f(1^N) = 0,$$

where the last step uses (4.110).

We now treat the complementary case $x \in (\{0,1\}^n |_{\leq k})^\theta$. By (4.107) and (4.115),

$$\begin{aligned}
\Psi(x) &= \sum_{u \in \{0,1\}^N |_{\leq \theta}} \psi(u) \mathbf{E}_{\substack{\mathbf{v} \in V^\theta: \\ \mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_\theta = u}} \Lambda_{\mathbf{v}}(x) \\
&= \sum_{u \in \{0,1\}^N |_{\leq \theta}} \psi(u) \mathbf{E}_{\substack{\mathbf{v} \in V^\theta: \\ \mathbf{v}_1 + \mathbf{v}_2 + \dots + \mathbf{v}_\theta = u}} \prod_{i=1}^{\theta} \lambda_{\mathbf{v}_i}(x_i). \tag{4.122}
\end{aligned}$$

It follows from $\Psi(x) < 0$ that the summation in (4.122) contains at least one negative term, corresponding to a string $u \in \{0,1\}^N |_{\leq \theta}$. This forces

$$\psi(u) < 0 \tag{4.123}$$

and additionally implies the existence of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_\theta \in V$ with

$$x_i \in \text{supp } \lambda_{\mathbf{v}_i}, \quad i = 1, 2, \dots, \theta, \tag{4.124}$$

$$\sum_{i=1}^{\theta} \mathbf{v}_i = u. \tag{4.125}$$

Since $x \in (\{0,1\}^n |_{\leq k})^\theta$ in the case under consideration, it follows from (4.96) and (4.124) that $|x_i| = k$ for all i . Now (4.118) ensures that $h(x_i) = \mathbf{v}_i$ for all i , which in turn makes it possible to rewrite (4.125) as $\sum_{i=1}^{\theta} h(x_i) = u$. Since

$u, h(x_1), h(x_2), \dots, h(x_\theta) \in \{0, 1\}^N$, we conclude that $\bigvee_{i=1}^\theta h(x_i) = u$. As a result,

$$(f \circ H)(x) = f\left(\bigvee_{i=1}^\theta h(x_i)\right) = f(u) = 0,$$

where the last step is immediate from (4.114) and (4.123). \square

4.5. Specializing the parameters. Theorems 4.5 and 4.12 have a large number of parameters that one can adjust to produce various hardness amplification theorems. We do so in this section. For any constants $\alpha \in (0, 1]$ and $C \geq 1$, we show how to transform a function f on θ^C bits with approximate degree

$$\deg_\epsilon(f|_{\leq \theta}) \geq \theta^{1-\alpha} \tag{4.126}$$

into a function F on $T^{1+\alpha}$ bits with approximate degree

$$\deg_{\epsilon - \frac{1}{T}}(F|_{\leq T}) \geq T^{1 - \frac{2}{3}\alpha}. \tag{4.127}$$

Comparing the exponents in (4.126) and (4.127), we see that F is harder to approximate than f relative to the Hamming weight of the inputs for F and f , respectively. Moreover, we show that F is expressible as $F = f \circ H$ for some mapping H whose output bits are computable by monotone DNF formulas of constant width. In particular, if f is a monotone DNF formula of constant width, then so is F . The formal statement follows.

COROLLARY 4.15. *Fix reals $\alpha \in (0, 1]$, $A \geq 1$, and $C \geq 1$ arbitrarily. Then for all large enough integers θ , there is an (explicitly given) mapping $H: \{0, 1\}^{\lfloor T^{1+\alpha} \rfloor} \rightarrow \{0, 1\}^{\lfloor \theta^C \rfloor}$ with $T = \lfloor \theta \log^2 \theta \rfloor$ such that the output bits of H are computable by monotone $\lceil 50(A+C)/\alpha \rceil$ -DNF formulas and*

$$\deg_{\epsilon - \frac{1}{T^A}}((f \circ H)|_{\leq T}) \geq T^{1 - \frac{2}{3}\alpha} \tag{4.128}$$

for every $\epsilon \in [0, 1]$ and every function $f: \{0, 1\}^{\lfloor \theta^C \rfloor} \rightarrow \{0, 1\}$ with $\deg_\epsilon(f|_{\leq \theta}) \geq \theta^{1-\alpha}$.

Proof. Invoke Theorem 4.5 with parameters

$$\beta = \frac{1}{2\theta \lfloor \theta \log^2 \theta \rfloor^A}, \tag{4.129}$$

$$N = \lfloor \theta^C \rfloor, \tag{4.130}$$

$$n = \lfloor \theta^\alpha \rfloor, \tag{4.131}$$

$$m = \lfloor \theta^{\alpha/4} \rfloor, \tag{4.132}$$

$$k = \left\lceil \frac{50(A+C)}{\alpha} \right\rceil - 1, \tag{4.133}$$

$$D = \lceil \theta^{1 - \frac{5}{8}\alpha} \rceil, \tag{4.134}$$

$$T = \lfloor \theta \log^2 \theta \rfloor. \tag{4.135}$$

Provided that θ is large enough, these parameter settings satisfy the theorem hypotheses (4.44)–(4.47), whereas (4.48) gives

$$\Delta \leq \frac{1}{4T^A}. \quad (4.136)$$

As a result, Theorem 4.5 guarantees that

$$\deg_{\epsilon - \frac{1}{T^A}}((f \circ H)|_{\leq T}) \geq \frac{c}{2} \cdot \theta^{1 - \frac{5}{8}\alpha}, \quad (4.137)$$

where $c \in (0, 1)$ is the absolute constant from Theorem 4.4 and $H: \{0, 1\}^{\lceil T^{1+\alpha} \rceil} \rightarrow \{0, 1\}^{\lceil \theta^C \rceil}$ is an explicit mapping whose output bits are computable by monotone $\lceil 50(A + C)/\alpha \rceil$ -DNF formulas. (In fact, H uses only $n\theta \approx (T/\log^2 T)^{1+\alpha}$ input bits, but this improvement is not relevant for our purposes.) Provided that θ is large enough relative to the absolute constant c , we infer (4.128) immediately from (4.137). \square

Analogously, we have the following hardness amplification result for *one-sided* approximate degree.

COROLLARY 4.16. *Fix reals $\alpha \in (0, 1]$, $A \geq 1$, and $C \geq 1$ arbitrarily. Then for all large enough integers θ , there is an (explicitly given) mapping $H: \{0, 1\}^{\lceil T^{1+\alpha} \rceil} \rightarrow \{0, 1\}^{\lceil \theta^C \rceil}$ with $T = \lfloor \theta \log^2 \theta \rfloor$ such that the output bits of H are computable by monotone $\lceil 50(A + C)/\alpha \rceil$ -DNF formulas and*

$$\deg_{\epsilon - \frac{1}{T^A}}^+((f \circ H)|_{\leq T}) \geq T^{1 - \frac{2}{3}\alpha} \quad (4.138)$$

for every $\epsilon \in [0, 1]$ and every function $f: \{0, 1\}^{\lceil \theta^C \rceil} \rightarrow \{0, 1\}$ such that $\deg_{\epsilon}^+(f|_{\leq \theta}) \geq \theta^{1-\alpha}$ and $f(1^{\lceil \theta^C \rceil}) = 0$.

Proof. The proof is the same, mutatis mutandis, as that of Corollary 4.15. Specifically, invoke Theorem 4.12 with parameters (4.129)–(4.135). Provided that θ is large enough, these parameter settings satisfy the theorem hypotheses (4.90)–(4.93), whereas (4.94) gives (4.136). As a result, Theorem 4.12 guarantees that

$$\deg_{\epsilon - \frac{1}{T^A}}^+((f \circ H)|_{\leq T}) \geq \frac{c}{2} \cdot \theta^{1 - \frac{5}{8}\alpha}, \quad (4.139)$$

where $c \in (0, 1)$ is the absolute constant from Theorem 4.4 and $H: \{0, 1\}^{\lceil T^{1+\alpha} \rceil} \rightarrow \{0, 1\}^{\lceil \theta^C \rceil}$ is an explicit mapping whose output bits are computable by monotone $\lceil 50(A + C)/\alpha \rceil$ -DNF formulas. Provided that θ is large enough relative to the absolute constant c , this settles (4.138). \square

5. MAIN RESULTS

In this section, we will settle our main results on approximate degree and present their applications to communication complexity.

5.1. Approximate degree of DNF and CNF formulas. We will start with the two-sided case. Our proof here amounts to taking the trivial one-variable formula x_1 and iteratively applying the hardness amplification of Corollary 4.15.

THEOREM 5.1. *For every $\delta \in (0, 1]$ and $\Delta \geq 1$, there is a constant $c \geq 1$ and an (explicitly given) family $\{f_n\}_{n=1}^\infty$ of functions $f_n: \{0, 1\}^n \rightarrow \{0, 1\}$ such that each f_n is computable by a monotone c -DNF formula and satisfies*

$$\deg_{\frac{1}{2} - \frac{1}{n^\Delta}}(f_n) \geq \frac{1}{c} \cdot n^{1-\delta}, \quad n = 1, 2, 3, \dots \quad (5.1)$$

Proof. Let $K \geq 1$ be the smallest integer such that

$$\frac{1 - (2/3)^K}{1 + (2/3)^{K-1}} > 1 - \delta. \quad (5.2)$$

Define

$$A = 2\Delta + 3. \quad (5.3)$$

Now, let $n \geq 1$ be any large enough integer. Define $T_0, T_1, T_2, \dots, T_K$ recursively by $T_0 = \lfloor n / \log^{2K} n \rfloor$ and $T_i = \lfloor T_{i-1} \log^2 T_{i-1} \rfloor$ for $i \geq 1$. Thus,

$$T_i \leq \frac{n}{\log^{2(K-i)} n}, \quad i = 0, 1, 2, \dots, K, \quad (5.4)$$

$$T_i \sim \frac{n}{\log^{2(K-i)} n}, \quad i = 0, 1, 2, \dots, K, \quad (5.5)$$

where \sim denotes equality up to lower-order terms. Provided that n is larger than a certain constant, inductive application of Corollary 4.15 gives functions

$$g_{n,i}: \{0, 1\}^{\lfloor T_i^{1+(2/3)^{i-1}} \rfloor} \rightarrow \{0, 1\}, \quad i = 0, 1, 2, \dots, K, \quad (5.6)$$

such that

$$\deg_{\frac{1}{2} - \frac{1}{T_0^A} - \frac{1}{T_1^A} - \dots - \frac{1}{T_i^A}}(g_{n,i}|_{\leq T_i}) \geq T_i^{1-(2/3)^i}, \quad i = 0, 1, 2, \dots, K, \quad (5.7)$$

and each $g_{n,i}$ is an explicitly constructed monotone c_i -DNF formula for some constant c_i independent of n . In more detail, the requirement (5.7) for $i = 0$ is equivalent to $\deg_{\frac{1}{2} - \frac{1}{T_0^A}}(g_{n,0}|_{\leq T_0}) > 0$ and is trivially satisfied by the ‘‘dictator’’ function $g_{n,0}(x) = x_1$, whereas for $i \geq 1$ the function $g_{n,i}$ is obtained constructively from $g_{n,i-1}$ by invoking Corollary 4.15 with

$$\begin{aligned} \alpha &= \left(\frac{2}{3}\right)^{i-1}, \\ C &= 1 + \left(\frac{2}{3}\right)^{i-2}, \\ \theta &= T_{i-1}, \\ f &= g_{n,i-1}, \\ \epsilon &= \frac{1}{2} - \frac{1}{T_0^A} - \frac{1}{T_1^A} - \dots - \frac{1}{T_{i-1}^A}. \end{aligned}$$

Specializing (5.4)–(5.7) to $i = K$, the function $g_{n,K}$ is a monotone c_K -DNF formula for some constant c_K independent of n , takes at most $N := n^{1+(2/3)^{K-1}}$ input variables, and has approximate degree

$$\begin{aligned} \deg_{\frac{1}{2} - \frac{1}{N^{\Delta+1}}}(g_{n,K}) &\geq \deg_{\frac{1}{2} - \frac{1}{T_0^A} - \frac{1}{T_1^A} - \dots - \frac{1}{T_K^A}}(g_{n,K}) \\ &\geq \deg_{\frac{1}{2} - \frac{1}{T_0^A} - \frac{1}{T_1^A} - \dots - \frac{1}{T_K^A}}(g_{n,K} |_{\leq T_K}) \\ &= \Omega(n^{1-(2/3)^K}) \\ &= \omega(N^{1-\delta}), \end{aligned}$$

where the first and last steps hold for all large enough n due to (5.3) and (5.2), respectively. The desired function family $\{f_n\}_{n=1}^\infty$ can then be defined by setting

$$f_n = g_{\lfloor n^{1/(1+(2/3)^{K-1})} \rfloor, K}$$

for all n larger than a certain constant n_0 , and taking the remaining functions f_1, f_2, \dots, f_{n_0} to be the dictator function $x \mapsto x_1$. \square

Theorem 5.1 immediately implies Theorems 1.1 and 1.2 from the introduction. We now move on to the one-sided case.

THEOREM 5.2. *For every $\delta \in (0, 1]$ and $\Delta \geq 1$, there is a constant $c \geq 1$ and an (explicitly given) family $\{f_n\}_{n=1}^\infty$ of functions $f_n: \{0, 1\}^n \rightarrow \{0, 1\}$ such that each f_n is computable by a monotone c -DNF formula and satisfies*

$$\deg_{\frac{1}{2} - \frac{1}{n^\Delta}}(\neg f_n) \geq \frac{1}{c} \cdot n^{1-\delta}, \quad n = 1, 2, 3, \dots \quad (5.8)$$

This result subsumes Theorem 5.1 and settles Theorem 1.4 in the introduction. The proof below makes repeated use of the following observation: if one applies Corollary 4.16 to a function f that is the *negation* of a constant-width monotone DNF formula, then the resulting composition $f \circ H$ is again the *negation* of a constant-width monotone DNF formula. This is easy to see by writing $\neg(f \circ H) = (\neg f) \circ H$ and noting that both $\neg f$ and H are computable by constant-width monotone DNF formulas.

Proof of Theorem 5.2. Much of the proof is identical to that of Theorem 5.1. As before, let $K \geq 1$ be the smallest integer such that

$$\frac{1 - (2/3)^K}{1 + (2/3)^{K-1}} > 1 - \delta. \quad (5.9)$$

Define

$$A = 2\Delta + 3. \quad (5.10)$$

Now, let $n \geq 1$ be any large enough integer. Define $T_0, T_1, T_2, \dots, T_K$ recursively by $T_0 = \lfloor n / \log^{2K} n \rfloor$ and $T_i = \lfloor T_{i-1} \log^2 T_{i-1} \rfloor$ for $i \geq 1$. Thus,

$$T_i \leq \frac{n}{\log^{2(K-i)} n}, \quad i = 0, 1, 2, \dots, K, \quad (5.11)$$

$$T_i \sim \frac{n}{\log^{2(K-i)} n}, \quad i = 0, 1, 2, \dots, K, \quad (5.12)$$

where \sim denotes equality up to lower-order terms. Provided that n is larger than a certain constant, inductive application of Corollary 4.16 gives functions

$$g_{n,i} : \{0, 1\}^{\lfloor T_i^{1+(2/3)^{i-1}} \rfloor} \rightarrow \{0, 1\}, \quad i = 0, 1, 2, \dots, K, \quad (5.13)$$

such that

$$\deg_{\frac{1}{2} - \frac{1}{T_0^A} - \frac{1}{T_1^A} - \dots - \frac{1}{T_i^A}}^+(\neg g_{n,i} |_{\leq T_i}) \geq T_i^{1-(2/3)^i}, \quad i = 0, 1, 2, \dots, K, \quad (5.14)$$

and each $g_{n,i}$ is an explicitly constructed monotone c_i -DNF formula for some constant c_i independent of n . In more detail, the requirement (5.14) for $i = 0$ is equivalent to $\deg_{\frac{1}{2} - \frac{1}{T_0^A}}^+(\neg g_{n,0} |_{\leq T_0}) > 0$ and is trivially satisfied by the ‘‘dictator’’ function $g_{n,0}(x) = x_1$. For $i \geq 1$, we obtain $g_{n,i}$ from $g_{n,i-1}$ by applying Corollary 4.16 with

$$\begin{aligned} \alpha &= \left(\frac{2}{3}\right)^{i-1}, \\ C &= 1 + \left(\frac{2}{3}\right)^{i-2}, \\ \theta &= T_{i-1}, \\ f &= \neg g_{n,i-1}, \\ \epsilon &= \frac{1}{2} - \frac{1}{T_0^A} - \frac{1}{T_1^A} - \dots - \frac{1}{T_{i-1}^A}. \end{aligned}$$

This appeal to Corollary 4.16 is legitimate because $g_{n,i-1}$ is a monotone DNF formula and therefore its negation $f = \neg g_{n,i-1}$ evaluates to 0 on the all-ones input.

Specializing (5.11)–(5.14) to $i = K$, the function $g_{n,K}$ is a monotone c_K -DNF formula for some constant c_K independent of n , takes at most $N := n^{1+(2/3)^{K-1}}$ input variables, and has one-sided approximate degree

$$\begin{aligned} \deg_{\frac{1}{2} - \frac{1}{N^{\Delta+1}}}^+(\neg g_{n,K}) &\geq \deg_{\frac{1}{2} - \frac{1}{T_0^A} - \frac{1}{T_1^A} - \dots - \frac{1}{T_K^A}}^+(\neg g_{n,K}) \\ &\geq \deg_{\frac{1}{2} - \frac{1}{T_0^A} - \frac{1}{T_1^A} - \dots - \frac{1}{T_K^A}}^+(\neg g_{n,K} |_{\leq T_K}) \\ &= \Omega(n^{1-(2/3)^K}) \\ &= \omega(N^{1-\delta}), \end{aligned}$$

where the first and last steps hold for all large enough n due to (5.10) and (5.9), respectively. The desired function family $\{f_n\}_{n=1}^\infty$ can then be defined by setting

$$f_n = g_{\lfloor n^{1/(1+(2/3)^{K-1})} \rfloor, K}$$

for all n larger than a certain constant n_0 , and taking the remaining functions f_1, f_2, \dots, f_{n_0} to be the dictator function $x \mapsto x_1$. \square

5.2. Quantum communication complexity. Using the pattern matrix method, we will ‘‘lift’’ our approximate degree results to a near-optimal lower bound on the communication complexity of DNF formulas in the two-party quantum model. Before we can apply the pattern matrix method, there is a technicality to address with regard to the representation of Boolean values as real numbers. In this paper,

we have followed the standard convention of representing “true” and “false” as 1 and 0, respectively. There is another common encoding, inspired by Fourier analysis and used in the pattern matrix method [38, 43], whereby “true” and “false” are represented as -1 and 1 , respectively. To switch back and forth between these representations, we will use the following proposition.

PROPOSITION 5.3. *For any function $f: X \rightarrow \mathbb{R}$ on a finite subset X of Euclidean space, and any reals $\epsilon \geq 0$ and $c \neq 0$,*

$$\begin{aligned} \deg_\epsilon(f + c) &= \deg_\epsilon(f), \\ \deg_{|c|\epsilon}(cf) &= \deg_\epsilon(f). \end{aligned}$$

Proof. For any polynomial p , we have the following equivalences:

$$\begin{aligned} \|f - p\|_\infty \leq \epsilon &\Leftrightarrow \|(f + c) - (p + c)\|_\infty \leq \epsilon, \\ \|f - p\|_\infty \leq \epsilon &\Leftrightarrow \|cf - cp\|_\infty \leq |c|\epsilon, \end{aligned}$$

where the second line uses $c \neq 0$. □

As a corollary, we can relate in a precise way the approximate degree of a Boolean function $f: X \rightarrow \{0, 1\}$ and the approximate degree of the associated ± 1 -valued function $f': X \rightarrow \{-1, +1\}$ given by $f' = (-1)^f$.

COROLLARY 5.4. *For any Boolean function $f: X \rightarrow \{0, 1\}$ and any $\epsilon \geq 0$,*

$$\deg_\epsilon((-1)^f) = \deg_{\epsilon/2}(f).$$

Proof. Since f is Boolean-valued, we have the equality of functions $(-1)^f = 1 - 2f$. Now $\deg_\epsilon((-1)^f) = \deg_\epsilon(1 - 2f) = \deg_\epsilon(-2f) = \deg_{2 \cdot \epsilon/2}(-2f) = \deg_{\epsilon/2}(f)$, where the second and fourth steps apply Proposition 5.3. □

Corollary 5.4 makes it easy to convert approximate degree results between the $0, 1$ representation and ± 1 representation. For communication complexity, no conversion is necessary in the first place:

$$Q_\epsilon^*(F) = Q_\epsilon^*((-1)^F), \quad F: X \times Y \rightarrow \{0, 1\}, \quad (5.15)$$

where Q_ϵ^* denotes ϵ -error quantum communication complexity with arbitrary prior entanglement. This equality holds because the representation of “true” and “false” in a communication protocol is a purely syntactic matter, and one can relabel the output values $0, 1$ as $1, -1$, respectively, without affecting the protocol’s correctness or communication cost. We note that (5.15) and Corollary 5.4 pertain to the encoding of the *output* of a Boolean function f . How “true” and “false” bits are represented in the *input* to f is immaterial both for communication complexity and approximate degree because the bijection $(0, 1) \leftrightarrow (1, -1)$ is a linear map.

We are now in a position to prove the promised communication lower bounds. The pattern matrix method for two-party quantum communication is given by the following theorem [38, Theorem 1.1].

THEOREM 5.5 (Sherstov). *Let $f: \{0, 1\}^t \rightarrow \{0, 1\}$ be given. Define $F: \{0, 1\}^{4t} \times \{0, 1\}^{4t} \rightarrow \{0, 1\}$ by*

$$F(x, y) = f \left(\bigvee_{i=1}^4 (x_{1,i} \wedge y_{1,i}), \dots, \bigvee_{i=1}^4 (x_{t,i} \wedge y_{t,i}) \right).$$

Then for all $\alpha \in [0, 1]$ and $\beta < \alpha/2$,

$$Q_{\beta}^*(F) \geq \frac{1}{4} \deg_{\alpha/2}(f) - \frac{1}{2} \log \left(\frac{3}{\alpha - 2\beta} \right).$$

The original statement in [38, Theorem 1.1] uses the ± 1 representation for the range of f and F . We translated it to the $0, 1$ representation, as stated in Theorem 5.5, by applying (5.15) to F and Corollary 5.4 to f . By combining Theorems 5.1 and 5.5, we obtain our main result on the quantum communication complexity of DNF formulas:

THEOREM 5.6. *For all $\delta \in (0, 1]$ and $A \geq 1$, there is a constant $c \geq 1$ and an (explicitly given) family $\{F_n\}_{n=1}^{\infty}$ of two-party communication problems $F_n: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$ such that each F_n is computable by a monotone c -DNF formula and satisfies*

$$Q_{\frac{1}{2} - \frac{1}{n^A}}^*(F_n) = \Omega(n^{1-\delta}). \quad (5.16)$$

Proof. Theorem 5.1 gives a constant $c' \geq 1$ and an explicit family $\{f_n\}_{n=1}^{\infty}$ of functions $f_n: \{0, 1\}^n \rightarrow \{0, 1\}$ such that each f_n is computable by a monotone c' -DNF formula and satisfies

$$\deg_{\frac{1}{2} - \frac{1}{n^{2A}}}(f_n) \geq \frac{1}{c'} \cdot n^{1-\delta}, \quad n = 1, 2, 3, \dots \quad (5.17)$$

For $n \geq 4$, define $F_n: \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}$ by

$$F_n(x, y) = f_{\lfloor n/4 \rfloor} \left(\bigvee_{i=1}^4 (x_{1,i} \wedge y_{1,i}), \dots, \bigvee_{i=1}^4 (x_{\lfloor n/4 \rfloor, i} \wedge y_{\lfloor n/4 \rfloor, i}) \right),$$

where we index the strings x and y as arrays of $\lfloor n/4 \rfloor \times 4$ bits. Clearly, F_n is computable by a monotone $2c'$ -DNF formula. We now invoke the pattern matrix method for quantum communication (Theorem 5.5) with parameters

$$\begin{aligned} \alpha &= 1 - \frac{2}{\lfloor n/4 \rfloor^{2A}}, \\ \beta &= \frac{1}{2} - \frac{1}{n^A}, \\ f &= f_{\lfloor n/4 \rfloor}, \end{aligned}$$

which satisfy $\beta < \alpha/2$ for all $n \geq 24$. As a result,

$$\begin{aligned} Q_{\frac{1}{2} - \frac{1}{n^A}}^*(F_n) &\geq \frac{1}{4} \cdot \deg_{\frac{1}{2} - \frac{1}{\lfloor n/4 \rfloor^{2A}}}(f_{\lfloor n/4 \rfloor}) - \frac{1}{2} \log \left(\frac{3}{\frac{2}{n^A} - \frac{2}{\lfloor n/4 \rfloor^{2A}}} \right) \\ &\geq \frac{1}{4} \cdot \frac{1}{c'} \cdot \left\lfloor \frac{n}{4} \right\rfloor^{1-\delta} - \frac{1}{2} \log \left(\frac{3}{\frac{2}{n^A} - \frac{2}{\lfloor n/4 \rfloor^{2A}}} \right) \end{aligned}$$

for all $n \geq 24$, where the first inequality applies the pattern matrix method, and the second inequality uses (5.17). Now (5.16) follows since A, c', δ are constants. \square

Theorem 5.6 settles Theorem 1.9 from the introduction.

5.3. Randomized multiparty communication. We now turn to communication lower bounds for DNF formulas in the k -party number-on-the-forehead model. Analogous to (5.15), we have

$$R_\epsilon(F) = R_\epsilon((-1)^F), \quad F: X_1 \times X_2 \times \cdots \times X_k \rightarrow \{0, 1\}, \quad (5.18)$$

where R_ϵ denotes ϵ -error number-on-the-forehead randomized communication complexity. The k -party set disjointness problem $\text{DISJ}_{n,k}: (\{0, 1\}^n)^k \rightarrow \{0, 1\}$ is given by

$$\text{DISJ}_{n,k}(x_1, x_2, \dots, x_k) = \bigwedge_{j=1}^n \bigvee_{i=1}^k x_{i,j}.$$

In other words, the problem asks whether there is a coordinate j in which each of the Boolean vectors x_1, x_2, \dots, x_k has a 1. If one views x_1, x_2, \dots, x_k as the characteristic vectors of corresponding sets S_1, S_2, \dots, S_k , then the set disjointness function evaluates to true if and only if $S_1 \cap S_2 \cap \cdots \cap S_k = \emptyset$. For a communication problem $g: X_1 \times X_2 \times \cdots \times X_k \rightarrow \{0, 1\}$ and a function $f: \{0, 1\}^n \rightarrow \{0, 1\}$, we view the componentwise composition $f \circ g$ as a k -party communication problem on $X_1^n \times X_2^n \times \cdots \times X_k^n$. The multiparty pattern matrix method [43, Theorem 5.1] gives a lower bound on the communication complexity of $f \circ \text{DISJ}_{m,k}$ in terms of the approximate degree of f :

THEOREM 5.7 (Sherstov). *Let $f: \{0, 1\}^n \rightarrow \{0, 1\}$ be given. Consider the k -party communication problem F defined by $F = f \circ \text{DISJ}_{m,k}$. Then for all $\alpha, \beta \geq 0$ with $\beta < \alpha/2$, one has*

$$R_\beta(F) \geq \frac{\deg_{\alpha/2}(f)}{2} \cdot \log \left(\frac{\sqrt{m}}{C2^k k} \right) - \log \frac{1}{\alpha - 2\beta},$$

where $C > 0$ is an absolute constant.

The actual statement of the pattern matrix method in [43, Theorem 5.1] is for functions f and F with range $\{-1, +1\}$. Theorem 5.7 above, stated for functions with range $\{0, 1\}$, is immediate from [43, Theorem 5.1] by applying (5.18) to F and Corollary 5.4 to f . We are now ready for our main result on the randomized multiparty communication complexity of DNF formulas.

THEOREM 5.8. *Fix arbitrary constants $\delta \in (0, 1]$ and $A \geq 1$. Then for all integers $n, k \geq 2$, there is an (explicitly given) k -party communication problem $F_{n,k}: (\{0, 1\}^n)^k \rightarrow \{0, 1\}$ with*

$$R_{1/3}(F_{n,k}) \geq \left(\frac{n}{cA^k k^2} \right)^{1-\delta}, \quad (5.19)$$

$$R_{\frac{1}{2} - \frac{1}{n^A}}(F_{n,k}) \geq \frac{n^{1-\delta}}{cA^k}, \quad (5.20)$$

where $c \geq 1$ is a constant independent of n and k . Moreover, each $F_{n,k}$ is computable by a monotone DNF formula of width ck and size n^c .

It will be helpful to keep in mind that the conclusion of Theorem 5.8 is “monotone” in c , in the sense that proving Theorem 5.8 for a given constant c proves it for all larger constants as well.

Proof. Theorem 5.1 gives a constant $c' \geq 1$ and an explicit family $\{f_n\}_{n=1}^\infty$ of functions $f_n: \{0,1\}^n \rightarrow \{0,1\}$ such that each f_n is computable by a monotone DNF formula of width c' and satisfies

$$\deg_{\frac{1}{2} - \frac{1}{n^{2A/\delta}}}(f_n) \geq \frac{1}{c'} \cdot n^{1-\frac{\delta}{2}}, \quad n = 1, 2, 3, \dots \quad (5.21)$$

Let $C > 0$ be the absolute constant from Theorem 5.7. For arbitrary integers $n, k \geq 2$, define

$$F_{n,k} = \begin{cases} \text{AND}_k & \text{if } n < \lceil C2^{k+1}k \rceil^2, \\ f_{\lfloor n/\lceil C2^{k+1}k \rceil^2} \circ \neg\text{DISJ}_{\lceil C2^{k+1}k \rceil^2, k} & \text{otherwise.} \end{cases}$$

We first analyze the cost of representing $F_{n,k}$ as a DNF formula. If $n < \lceil C2^{k+1}k \rceil^2$, then by definition $F_{n,k}$ is a monotone DNF formula of width k and size 1. In the complementary case, $f_{\lfloor n/\lceil C2^{k+1}k \rceil^2}$ is by construction a monotone DNF formula of width c' and hence of size at most $n^{c'}$, whereas $\neg\text{DISJ}_{\lceil C2^{k+1}k \rceil^2, k}$ is by definition a monotone DNF formula of width k and size at most $\lceil C2^{k+1}k \rceil^2 \leq n$. As a result, the composed function $F_{n,k}$ is a monotone DNF formula of width $c'k$ and size at most $n^{c'} \cdot n^{c'} = n^{2c'}$. In particular, the claim in the theorem statement regarding the width and size of $F_{n,k}$ as a monotone DNF formula is valid for any constant $c \geq 2c'$.

We now turn to the communication complexity of $F_{n,k}$. Since $F_{n,k}$ is nonconstant, we have the trivial bound

$$R_\epsilon(F_{n,k}) \geq 1, \quad 0 \leq \epsilon < \frac{1}{2}. \quad (5.22)$$

We further claim that

$$R_\epsilon(F_{n,k}) \geq \frac{1}{2c'} \cdot \left[\frac{n}{\lceil C2^{k+1}k \rceil^2} \right]^{1-\frac{\delta}{2}} + \log \left(1 - \frac{2}{\lfloor n/\lceil C2^{k+1}k \rceil^2 \rfloor^{2A/\delta}} - 2\epsilon \right) \quad (5.23)$$

whenever the logarithmic term is well-defined. For $n < \lceil C2^{k+1}k \rceil^2$, this claim is vacuous. In the complementary case $n \geq \lceil C2^{k+1}k \rceil^2$, consider the family $\{g_n\}_{n=1}^\infty$ of functions $g_n: \{0,1\}^n \rightarrow \{0,1\}$ given by $g_n(x_1, x_2, \dots, x_n) = f_n(\neg x_1, \neg x_2, \dots, \neg x_n)$. For each n , it is clear that g_n and f_n have the same approximate degree. Since $F_{n,k} = g_{\lfloor n/\lceil C2^{k+1}k \rceil^2} \circ \text{DISJ}_{\lceil C2^{k+1}k \rceil^2, k}$, one now obtains (5.23) directly from (5.21) and the multiparty pattern matrix method (Theorem 5.7).

For a sufficiently large constant $c \geq 1$, the communication lower bound (5.19) follows from (5.23) for $n \geq c4^k k^2$ and follows from (5.22) for $n < c4^k k^2$.

The proof of (5.20) is more tedious. Take the constant $c \geq 1$ large enough that the following relations hold:

$$\left\lfloor \frac{n^\delta}{\lceil 2C \rceil^2 \log^2 n} \right\rfloor \geq 2n^{\delta/2} \quad \text{for all } n \geq c, \quad (5.24)$$

$$n^{\delta^2/4} \geq 1 + A \log n \quad \text{for all } n \geq c, \quad (5.25)$$

$$c \geq 2c'. \quad (5.26)$$

If $n^{1-\delta} < c4^k$, then (5.20) holds due to (5.22). In what follows, we treat the complementary case when

$$\frac{n^{1-\delta}}{c4^k} \geq 1 \quad (5.27)$$

and in particular

$$n \geq c, \quad (5.28)$$

$$k \leq \log n. \quad (5.29)$$

Then

$$\begin{aligned} \left\lfloor \frac{n}{\lceil C2^{k+1}k \rceil^2} \right\rfloor^{1-\frac{\delta}{2}} &\geq \left\lfloor \frac{n}{\lceil 2C \rceil^2 4^k k^2} \right\rfloor^{1-\frac{\delta}{2}} \\ &\geq \left[\frac{n^{1-\delta}}{4^k} \cdot \frac{n^\delta}{\lceil 2C \rceil^2 \log^2 n} \right]^{1-\frac{\delta}{2}} \\ &\geq \left[\frac{n^{1-\delta}}{4^k} \cdot 2n^{\delta/2} \right]^{1-\frac{\delta}{2}} \\ &\geq \left(\frac{n^{1-\delta}}{4^k} \cdot n^{\delta/2} \right)^{1-\frac{\delta}{2}} \\ &\geq \frac{n^{1-\delta}}{4^k} \cdot n^{\delta^2/4} \\ &\geq \frac{n^{1-\delta}}{4^k} \cdot (1 + A \log n), \end{aligned} \quad (5.30)$$

where the second step uses (5.29), the third step uses (5.24) and (5.28), the fourth step is valid by (5.27), and the last step uses (5.25) and (5.28). Continuing,

$$\begin{aligned} &\log \left(1 - \frac{2}{\lfloor n / \lceil C2^{k+1}k \rceil^2 \rfloor^{2A/\delta}} - 2 \left(\frac{1}{2} - \frac{1}{n^A} \right) \right) \\ &= \log \left(\frac{2}{n^A} - \frac{2}{\lfloor n / \lceil C2^{k+1}k \rceil^2 \rfloor^{2A/\delta}} \right) \\ &\geq \log \left(\frac{2}{n^A} - \frac{2}{\lfloor n / (\lceil 2C \rceil^2 4^k k^2) \rfloor^{2A/\delta}} \right) \\ &\geq \log \left(\frac{2}{n^A} - \frac{2}{\lfloor n / (\lceil 2C \rceil^2 n^{1-\delta} \log^2 n) \rfloor^{2A/\delta}} \right) \\ &\geq \log \left(\frac{2}{n^A} - \frac{2}{(2n^{\delta/2})^{2A/\delta}} \right) \end{aligned}$$

$$\begin{aligned}
&\geq \log \left(\frac{2}{n^A} - \frac{1}{n^A} \right) \\
&= -A \log n,
\end{aligned} \tag{5.31}$$

where the third step uses (5.27) and (5.29), and the fourth step uses (5.24) and (5.28). Now

$$\begin{aligned}
R_{\frac{1}{2} - \frac{1}{n^A}}(F_{n,k}) &\geq \frac{1}{2c'} \cdot \frac{n^{1-\delta}}{4^k} \cdot (1 + A \log n) - A \log n \\
&\geq \frac{n^{1-\delta}}{c4^k} \cdot (1 + A \log n) - A \log n \\
&\geq \frac{n^{1-\delta}}{c4^k},
\end{aligned}$$

where the first step substitutes the bounds (5.30) and (5.31) into (5.23), the second step uses (5.26), and the third step is valid by (5.27). This completes the proof of (5.20). \square

Theorem 5.8 settles Theorem 1.6 from the introduction.

REMARK 5.9. In this section, we considered k -party number-on-the-forehead bounded-error communication complexity with *classical* players. The model naturally extends to *quantum* players, and our lower bound in Theorem 5.8 implies an $\Omega(n^{1-\delta}/4^k k)$ communication lower bound in this quantum k -party number-on-the-forehead model for computing an explicit DNF formula $F: (\{0, 1\}^n)^k \rightarrow \{0, 1\}$ of size $n^{O(1)}$ and width $O(k)$ with error probability $\frac{1}{2} - \frac{1}{n^A}$, where the constants $\delta > 0$ and $A \geq 1$ can be set arbitrarily. In more detail, the multiparty pattern matrix method actually gives a bound on the *generalized discrepancy* of the composed communication problem F . By the results of [27], generalized discrepancy leads in turn to a lower bound on the communication complexity of F in the quantum k -party number-on-the-forehead model. Quantitatively, the authors of [27] show that any classical communication lower bound obtained via generalized discrepancy carries over to the quantum model with only a factor of $\Theta(k)$ loss.

5.4. Nondeterministic and Merlin–Arthur multiparty communication.

To obtain our results on nondeterminism and Merlin–Arthur communication, we will now develop a general technique for transforming lower bounds on one-sided approximate degree into lower bounds in these communication models. The technique in question is implicit in the papers [23, 43] but has not been previously formalized in our sought generality.

Consider a k -party communication problem $F: X_1 \times X_2 \times \cdots \times X_k \rightarrow \{0, 1\}$, for some finite sets X_1, X_2, \dots, X_k . A fundamental notion in the study of multiparty communication is that of a *cylinder intersection* [6], defined as any function $\chi: X_1 \times X_2 \times \cdots \times X_k \rightarrow \{0, 1\}$ of the form

$$\chi(x_1, \dots, x_k) = \prod_{i=1}^k \phi_i(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k)$$

for some $\phi_i: X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots \times X_k \rightarrow \{0, 1\}$, $i = 1, 2, \dots, k$. In other words, a cylinder intersection is the product of k Boolean functions, where the i -th

function does not depend on the i -th coordinate. For a probability distribution μ on the domain of F , the *discrepancy of F with respect to μ* is denoted $\text{disc}_\mu(F)$ and defined as

$$\text{disc}_\mu(F) = \max_{\chi} \left| \mathbf{E}_{(x_1, \dots, x_k) \sim \mu} (-1)^{F(x_1, \dots, x_k)} \chi(x_1, \dots, x_k) \right|,$$

where the maximum is taken over all cylinder intersections χ . This notion of discrepancy was defined by Babai, Nisan, and Szegedy [6] and is unrelated to the one that we encountered in Section 3.2. It is of interest to us because of the following theorem [23, Theorem 4.1], which gives a lower bound on nondeterministic and Merlin–Arthur communication complexity in terms of discrepancy.

THEOREM 5.10 (Gavinsky and Sherstov). *Let $F : X \rightarrow \{0, 1\}$ be a given k -party communication problem, where $X = X_1 \times \dots \times X_k$. Fix a function $H : X \rightarrow \{0, 1\}$ and a probability distribution Π on X . Put*

$$\begin{aligned} \alpha &= \Pi(F^{-1}(1) \cap H^{-1}(1)), \\ \beta &= \Pi(F^{-1}(1) \cap H^{-1}(0)), \\ Q &= \log \frac{\alpha}{\beta + \text{disc}_\Pi(H)}. \end{aligned}$$

Then

$$N(F) \geq Q, \tag{5.32}$$

$$MA_{1/3}(F)^2 \geq \min \left\{ \Omega(Q), \Omega \left(\frac{Q}{\log\{2/\alpha\}} \right)^2 \right\} \tag{5.33}$$

$$\geq \Omega(Q) - \left(\log \frac{2}{\alpha} \right)^2. \tag{5.34}$$

We note that the original statement in [23] is for functions with range $\{-1, +1\}$. The above version for $\{0, 1\}$ follows immediately because the output values of a communication protocol serve as textual labels that can be changed at will. Equation (5.34), which is also not part of the statement in [23], follows from (5.33) in view of the inequality $(q/a)^2 \geq 2q - a^2$ for all reals q, a with $a \neq 0$. (Start with $(q/a - a)^2 \geq 0$ and multiply out the left-hand side.)

We will need yet another notion of discrepancy, introduced in [43] and called “repeated discrepancy.” Let G be a k -party communication problem on $X = X_1 \times X_2 \times \dots \times X_k$. A probability distribution π on the domain of G is called *balanced* if $\pi(G^{-1}(0)) = \pi(G^{-1}(1)) = 1/2$. For such π , the *repeated discrepancy of G with respect to π* is given by

$$\text{rdisc}_\pi(G) = \sup_{d, r \in \mathbb{Z}^+} \max_{\chi} \left| \mathbf{E}_{\dots, x_{i,j}, \dots} \left[\chi(\dots, x_{i,j}, \dots) \prod_{i=1}^d (-1)^{G(x_{i,1})} \right] \right|^{1/d},$$

where the maximum is over k -dimensional cylinder intersections χ on $X^{dr} = X_1^{dr} \times X_2^{dr} \times \dots \times X_k^{dr}$, and the arguments $x_{i,j}$ ($i = 1, 2, \dots, d$; $j = 1, 2, \dots, r$) are chosen independently according to π conditioned on $G(x_{i,1}) = G(x_{i,2}) = \dots = G(x_{i,r})$ for each i . The repeated discrepancy of a communication problem is much harder

to bound from above than standard discrepancy. The following result from [43, Theorem 4.27] bounds the repeated discrepancy of set disjointness.

THEOREM 5.11 (Sherstov). *Let m and k be positive integers. Then there is a balanced probability distribution π on the domain of $\text{DISJ}_{m,k}$ such that*

$$\text{rdisc}_\pi(\text{DISJ}_{m,k}) \leq \left(\frac{ck2^k}{\sqrt{m}} \right)^{1/2},$$

where $c > 0$ is an absolute constant independent of m, k, π .

It was shown in [43] that repeated discrepancy gives a highly efficient way to transform multiparty communication protocols into polynomials. For a nonnegative integer d and a function f on a finite subset of Euclidean space, define

$$E(f, d) = \min_p \{ \|f - p\|_\infty : \deg p \leq d \},$$

where the minimum is taken over polynomials of degree at most d . In other words, $E(f, d)$ stands for the minimum error in an ℓ_∞ -norm approximation of f by a polynomial of degree at most d . The following result was proved in [43, Theorem 4.2].

THEOREM 5.12 (Sherstov). *Let $G: X \rightarrow \{0, 1\}$ be a k -party communication problem, where $X = X_1 \times X_2 \times \cdots \times X_k$. For an integer $n \geq 1$ and a balanced probability distribution π on the domain of G , consider the linear operator $L_{\pi,n}: \mathbb{R}^{X^n} \rightarrow \mathbb{R}^{\{0,1\}^n}$ given by*

$$(L_{\pi,n}\chi)(z) = \mathbf{E}_{x_1 \sim \pi_{z_1}} \cdots \mathbf{E}_{x_n \sim \pi_{z_n}} \chi(x_1, \dots, x_n), \quad z \in \{0, 1\}^n, \quad (5.35)$$

where π_0 and π_1 are the probability distributions induced by π on $G^{-1}(0)$ and $G^{-1}(1)$, respectively. Then for some absolute constant $c > 0$ and every k -dimensional cylinder intersection χ on $X^n = X_1^n \times X_2^n \times \cdots \times X_k^n$,

$$E(L_{\pi,n}\chi, d-1) \leq (c \text{rdisc}_\pi(G))^d, \quad d = 1, 2, \dots, n.$$

We are now in a position to derive the promised lower bound on nondeterministic and Merlin–Arthur communication complexity in terms of one-sided approximate degree. Our proof combines Theorems 5.10–5.12 in a way closely analogous to the proof of [43, Theorem 6.9].

THEOREM 5.13. *Let $f: \{0, 1\}^n \rightarrow \{0, 1\}$ be given. Let m and k be positive integers, and put $F = f \circ \text{DISJ}_{m,k}$. Then for all $\epsilon \in (0, 1/2]$,*

$$N(F) \geq \frac{\deg_\epsilon^+(f)}{2} \log \left(\frac{\sqrt{m}}{Ck2^k} \right) - \log \frac{1}{\epsilon}, \quad (5.36)$$

$$MA_{1/3}(F)^2 \geq \frac{\deg_\epsilon^+(f)}{C} \log \left(\frac{\sqrt{m}}{Ck2^k} \right) - \left(\log \frac{2}{\epsilon} \right)^2, \quad (5.37)$$

where $C \geq 1$ is an absolute constant, independent of f, n, m, k, ϵ .

Proof. Abbreviate $d = \deg_\epsilon^+(f)$. Let $X = (\{0,1\}^m)^k$ denote the domain of $\text{DISJ}_{m,k}$. By Theorem 5.11, there is a probability distribution π on X such that

$$\pi(\text{DISJ}_{m,k}^{-1}(0)) = \pi(\text{DISJ}_{m,k}^{-1}(1)) = \frac{1}{2}, \quad (5.38)$$

$$\text{rdisc}_\pi(\text{DISJ}_{m,k}) \leq \left(\frac{c' k 2^k}{\sqrt{m}} \right)^{1/2}, \quad (5.39)$$

where $c' > 0$ is an absolute constant independent of m and k . By the dual characterization of one-sided approximate degree (Fact 2.9), there exists a function $\psi: \{0,1\}^n \rightarrow \mathbb{R}$ such that

$$\langle f, \psi \rangle > \epsilon, \quad (5.40)$$

$$\|\psi\|_1 = 1, \quad (5.41)$$

$$\text{orth } \psi \geq d, \quad (5.42)$$

$$\psi \geq 0 \text{ on } f^{-1}(1). \quad (5.43)$$

Define $\Psi: X^n \rightarrow \mathbb{R}$ by

$$\Psi(x) = 2^n \psi(\text{DISJ}_{m,k}(x_1), \dots, \text{DISJ}_{m,k}(x_n)) \prod_{i=1}^n \pi(x_i). \quad (5.44)$$

CLAIM 5.14. Ψ satisfies

$$\langle F, \Psi \rangle > \epsilon, \quad (5.45)$$

$$\|\Psi\|_1 = 1, \quad (5.46)$$

$$\Psi \geq 0 \text{ on } F^{-1}(1). \quad (5.47)$$

We will carry on with the theorem proof and settle the claims later. Equation (5.46) allows us to write

$$\Psi(x) = \Pi(x) \cdot (-1)^{1-H(x)} \quad (5.48)$$

for some Boolean function $H: X^n \rightarrow \{0,1\}$ and a probability distribution Π on X^n . Indeed, one can explicitly define $\Pi(x) = |\Psi(x)|$ and $H(x) = \mathbf{I}[\Psi(x) \geq 0]$.

CLAIM 5.15. One has

$$\Pi(F^{-1}(1) \cap H^{-1}(0)) = 0, \quad (5.49)$$

$$\Pi(F^{-1}(1) \cap H^{-1}(1)) > \epsilon. \quad (5.50)$$

CLAIM 5.16. There is an absolute constant $c > 0$ such that

$$\text{disc}_\Pi(H) \leq \left(\frac{ck2^k}{\sqrt{m}} \right)^{d/2}.$$

The sought communication bounds (5.36) and (5.37) follow from Theorem 5.10 in view of Claims 5.15 and 5.16. \square

We now settle the claims used in the proof of Theorem 5.13.

Proof of Claim 5.14. We have

$$\begin{aligned}
\langle F, \Psi \rangle &= 2^n \mathbf{E}_{x_1, \dots, x_n \sim \pi} f(\dots, \text{DISJ}_{m,k}(x_i), \dots) \psi(\dots, \text{DISJ}_{m,k}(x_i), \dots) \\
&= 2^n \mathbf{E}_{z \in \{0,1\}^n} f(z) \psi(z) \\
&= \langle f, \psi \rangle \\
&> \epsilon,
\end{aligned}$$

where the second step uses (5.38), and the third step is legitimate by (5.40). Analogously,

$$\begin{aligned}
\|\Psi\|_1 &= 2^n \mathbf{E}_{x_1, \dots, x_n \sim \pi} |\psi(\text{DISJ}_{m,k}(x_1), \dots, \text{DISJ}_{m,k}(x_n))| \\
&= 2^n \mathbf{E}_{z \in \{0,1\}^n} |\psi(z)| \\
&= 1,
\end{aligned}$$

where the last two steps are valid by (5.38) and (5.41), respectively. The final property (5.47) can be seen from the following chain of implications:

$$\begin{aligned}
x \in F^{-1}(1) &\Rightarrow (\text{DISJ}_{m,k}(x_1), \dots, \text{DISJ}_{m,k}(x_n)) \in f^{-1}(1) \\
&\Rightarrow \psi(\text{DISJ}_{m,k}(x_1), \dots, \text{DISJ}_{m,k}(x_n)) \geq 0 \\
&\Rightarrow \Psi(x) \geq 0,
\end{aligned}$$

where the first and third steps use the definitions of F and Ψ , respectively, and the second step is valid by (5.43). \square

Proof of Claim 5.15. Fix any point $x \in F^{-1}(1) \cap H^{-1}(0)$. Then (5.47) implies that $\Psi(x) \geq 0$, or equivalently $\Pi(x) \cdot (-1)^{1-H(x)} \geq 0$. This forces $\Pi(x) \leq 0$ due to $H(x) = 0$. Since Π is a probability distribution, we conclude that $\Pi(x) = 0$. The proof of (5.49) is complete. The remaining relation (5.50) can be seen as follows:

$$\begin{aligned}
\Pi(F^{-1}(1) \cap H^{-1}(1)) &= \Pi(F^{-1}(1) \cap H^{-1}(1)) - \Pi(F^{-1}(1) \cap H^{-1}(0)) \\
&= \sum_{X^n} \Pi(x) F(x) (-1)^{1-H(x)} \\
&= \langle \Psi, F \rangle \\
&> \epsilon,
\end{aligned}$$

where the first step exploits (5.49), and the last step applies (5.45). \square

Proof of Claim 5.16. Let π_0 and π_1 be the probability distributions induced by π on $\text{DISJ}_{m,k}^{-1}(0)$ and $\text{DISJ}_{m,k}^{-1}(1)$, respectively, and let $L_{\pi,n}: \mathbb{R}^{X^n} \rightarrow \mathbb{R}^{\{0,1\}^n}$ be the linear operator given by (5.35). Then for any cylinder intersection $\chi: X^n \rightarrow \{0,1\}$,

we have

$$\begin{aligned}
\left| \mathbf{E}_{x \sim \Pi} (-1)^{H(x)} \chi(x) \right| &= \left| \sum_{X^n} \Pi(x) (-1)^{H(x)} \chi(x) \right| \\
&= \left| \sum_{X^n} \Psi(x) \chi(x) \right| \\
&= \left| 2^n \mathbf{E}_{x_1, \dots, x_n \sim \pi} \psi(\dots, \text{DISJ}_{m,k}(x_i), \dots) \chi(x) \right| \\
&= \left| \sum_{z \in \{0,1\}^n} \psi(z) \mathbf{E}_{x_1 \sim \pi_{z_1}} \dots \mathbf{E}_{x_n \sim \pi_{z_n}} \chi(x) \right| \\
&= |\langle \psi, L_{\pi,n} \chi \rangle|, \tag{5.51}
\end{aligned}$$

where the second step uses (5.48), the third step invokes the definition (5.44), the fourth step is justified by (5.38), and the last step is valid by the definition of $L_{\pi,n}$.

For every polynomial $p: \{0,1\}^n \rightarrow \mathbb{R}$ of degree less than d , we have

$$\begin{aligned}
|\langle \psi, L_{\pi,n} \chi \rangle| &= |\langle \psi, L_{\pi,n} \chi - p \rangle + \langle \psi, p \rangle| \\
&= |\langle \psi, L_{\pi,n} \chi - p \rangle| \\
&\leq \|\psi\|_1 \|L_{\pi,n} \chi - p\|_\infty \\
&= \|L_{\pi,n} \chi - p\|_\infty, \tag{5.52}
\end{aligned}$$

where the second step uses (5.42), the third step applies Hölder's inequality, and the fourth step substitutes (5.41). Taking the infimum in (5.52) over all polynomials p of degree less than d , we arrive at

$$|\langle \psi, L_{\pi,n} \chi \rangle| \leq E(L_{\pi,n} \chi, d-1). \tag{5.53}$$

Now

$$\begin{aligned}
\text{disc}_\Pi(H) &= \max_\chi \left| \mathbf{E}_{x \sim \Pi} (-1)^{H(x)} \chi(x) \right| \\
&\leq \max_\chi E(L_{\pi,n} \chi, d-1) \\
&\leq (c'' \text{rdisc}_\pi(\text{DISJ}_{m,k}))^d \\
&\leq \left(c'' \left(\frac{c' k 2^k}{\sqrt{m}} \right)^{1/2} \right)^d,
\end{aligned}$$

where the first step maximizes over all cylinder intersections χ , the second step combines (5.51) and (5.53), the third step is valid for some absolute constant $c'' > 0$ by Theorem 5.12, and the fourth step holds by (5.39). \square

This completes the proof of Theorem 5.13. By combining it with our main result on one-sided approximate degree, we now obtain our sought lower bounds for nondeterministic and Merlin–Arthur multiparty communication.

THEOREM 5.17. *Let $\delta > 0$ be arbitrary. Then for all integers $n, k \geq 2$, there is an (explicitly given) k -party communication problem $F_{n,k}: (\{0, 1\}^n)^k \rightarrow \{0, 1\}$ with*

$$N(F_{n,k}) \leq c \log n, \quad (5.54)$$

$$N(\neg F_{n,k}) \geq \left(\frac{n}{c4^k k^2} \right)^{1-\delta}, \quad (5.55)$$

$$R_{1/3}(\neg F_{n,k}) \geq \left(\frac{n}{c4^k k^2} \right)^{1-\delta}, \quad (5.56)$$

$$MA_{1/3}(\neg F_{n,k}) \geq \left(\frac{n}{c4^k k^2} \right)^{\frac{1-\delta}{2}}, \quad (5.57)$$

where $c \geq 1$ is a constant independent of n and k . Moreover, each $F_{n,k}$ is computable by a monotone DNF formula of width ck and size n^c .

Proof. Theorem 5.2 gives a constant $c' \geq 1$ and an explicit family $\{f_n\}_{n=1}^\infty$ of functions $f_n: \{0, 1\}^n \rightarrow \{0, 1\}$ such that each f_n is computable by a monotone DNF formula of width c' and satisfies

$$\deg_{3/8}^+(\neg f_n) \geq \frac{1}{c'} \cdot n^{1-\delta}, \quad n = 1, 2, 3, \dots \quad (5.58)$$

In particular,

$$\deg_{3/8}(\neg f_n) \geq \frac{1}{c'} \cdot n^{1-\delta}, \quad n = 1, 2, 3, \dots \quad (5.59)$$

Let $C \geq 1$ be the maximum of the absolute constants from Theorems 5.7 and 5.13. For arbitrary integers $n, k \geq 2$, define

$$F_{n,k} = \begin{cases} \text{AND}_k & \text{if } n < \lceil C2^{k+1}k \rceil^2, \\ f_{\lfloor n/\lceil C2^{k+1}k \rceil^2} \circ \neg \text{DISJ}_{\lceil C2^{k+1}k \rceil^2, k} & \text{otherwise.} \end{cases}$$

We first analyze the cost of representing $F_{n,k}$ as a DNF formula. If $n < \lceil C2^{k+1}k \rceil^2$, then by definition $F_{n,k}$ is a monotone DNF formula of width k and size 1. In the complementary case, $f_{\lfloor n/\lceil C2^{k+1}k \rceil^2}$ is by construction a monotone DNF formula of width c' and hence of size at most $n^{c'}$, whereas $\neg \text{DISJ}_{\lceil C2^{k+1}k \rceil^2, k}$ is by definition a monotone DNF formula of width k and size at most $\lceil C2^{k+1}k \rceil^2 \leq n$. As a result, the composed function $F_{n,k}$ is a monotone DNF formula of width $c'k$ and size at most $n^{c'} \cdot n^{c'} = n^{2c'}$. In particular, the claim in the theorem statement regarding the width and size of $F_{n,k}$ as a monotone DNF formula is valid for any large enough c . This in turn implies the upper bound in (5.54): consider the nondeterministic protocol in which the parties “guess” one of the terms of the DNF formula for $F_{n,k}$ (for a cost of $\lceil \log n^{2c'} \rceil$ bits), evaluate it (using another 2 bits of communication), and output the result.

We now turn to the communication lower bounds. Since $F_{n,k}$ is nonconstant, we have the trivial bounds

$$N(\neg F_{n,k}) \geq 1, \quad (5.60)$$

$$R_{1/3}(\neg F_{n,k}) \geq 1, \quad (5.61)$$

$$MA_{1/3}(\neg F_{n,k}) \geq 1. \quad (5.62)$$

We further claim that

$$N(\neg F_{n,k}) \geq \frac{1}{2c'} \cdot \left\lfloor \frac{n}{\lceil C2^{k+1}k \rceil^2} \right\rfloor^{1-\delta} - \log \frac{8}{3}, \quad (5.63)$$

$$R_{1/3}(\neg F_{n,k}) \geq \frac{1}{2c'} \cdot \left\lfloor \frac{n}{\lceil C2^{k+1}k \rceil^2} \right\rfloor^{1-\delta} - \log 12, \quad (5.64)$$

$$MA_{1/3}(\neg F_{n,k})^2 \geq \frac{1}{Cc'} \cdot \left\lfloor \frac{n}{\lceil C2^{k+1}k \rceil^2} \right\rfloor^{1-\delta} - \left(\log \frac{16}{3} \right)^2. \quad (5.65)$$

For $n < \lceil C2^{k+1}k \rceil^2$, these claims are trivial since communication complexity is nonnegative. In the complementary case $n \geq \lceil C2^{k+1}k \rceil^2$, consider the family $\{g_n\}_{n=1}^\infty$ of functions $g_n: \{0,1\}^n \rightarrow \{0,1\}$ given by $g_n(x_1, x_2, \dots, x_n) = \neg f_n(\neg x_1, \neg x_2, \dots, \neg x_n)$. For each n , it is clear that g_n and $\neg f_n$ have the same one-sided approximate degree. Since $\neg F_{n,k} = g_{\lfloor n/\lceil C2^{k+1}k \rceil^2} \circ \text{DISJ}_{\lceil C2^{k+1}k \rceil^2, k}$, one now obtains (5.63) and (5.65) directly from (5.58) and Theorem 5.13. Analogously, g_n and $\neg f_n$ have the same two-sided approximate degree for each n , and one obtains (5.64) from (5.59) and Theorem 5.7.

For a large enough constant $c \geq 1$, the communication lower bounds (5.55)–(5.57) follow from (5.63)–(5.65) for $n \geq c4^k k^2$, and from (5.60)–(5.62) for $n < c4^k k^2$. \square

Theorem 5.17 settles Theorem 1.7 from the introduction.

ACKNOWLEDGMENTS

The author is thankful to Justin Thaler and Mark Bun for useful comments on an earlier version of this paper.

REFERENCES

- [1] S. AARONSON AND Y. SHI, *Quantum lower bounds for the collision and the element distinctness problems*, J. ACM, 51 (2004), pp. 595–605, doi:10.1145/1008731.1008735.
- [2] M. AJTAI, H. IWANIEC, J. KOMLÓS, J. PINTZ, AND E. SZEMERÉDI, *Construction of a thin set with small Fourier coefficients*, Bulletin of the London Mathematical Society, 22 (1990), pp. 583–590, doi:10.1112/blms/22.6.583.
- [3] L. BABAI, *Trading group theory for randomness*, in *Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing (STOC)*, 1985, pp. 421–429, doi:10.1145/22145.22192.
- [4] L. BABAI, P. FRANKL, AND J. SIMON, *Complexity classes in communication complexity theory*, in *Proceedings of the Twenty-Seventh Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 1986, pp. 337–347, doi:10.1109/SFCS.1986.15.
- [5] L. BABAI AND S. MORAN, *Arthur-Merlin games: A randomized proof system, and a hierarchy of complexity classes*, J. Comput. Syst. Sci., 36 (1988), pp. 254–276, doi:10.1016/0022-0000(88)90028-1.
- [6] L. BABAI, N. NISAN, AND M. SZEGEDY, *Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs*, J. Comput. Syst. Sci., 45 (1992), pp. 204–232, doi:10.1016/0022-0000(92)90047-M.
- [7] Z. BAR-YOSEF, T. S. JAYRAM, R. KUMAR, AND D. SIVAKUMAR, *An information statistics approach to data stream and communication complexity*, J. Comput. Syst. Sci., 68 (2004), pp. 702–732, doi:10.1016/j.jcss.2003.11.006.
- [8] R. BEALS, H. BUHRMAN, R. CLEVE, M. MOSCA, AND R. DE WOLF, *Quantum lower bounds by polynomials*, J. ACM, 48 (2001), pp. 778–797, doi:10.1145/502090.502097.
- [9] P. BEAME, M. DAVID, T. PITASSI, AND P. WOELFEL, *Separating deterministic from nondeterministic NOF multiparty communication complexity*, in *Proceedings of the Thirty-Fourth International Colloquium on Automata, Languages and Programming (ICALP)*, 2007, pp. 134–145, doi:10.1007/978-3-540-73420-8_14.

- [10] P. BEAME, M. DAVID, T. PITASSI, AND P. WOELFEL, *Separating deterministic from randomized multiparty communication complexity*, Theory of Computing, 6 (2010), pp. 201–225, doi:10.4086/toc.2010.v006a009.
- [11] P. BEAME AND T. HUYNH, *Multiparty communication complexity and threshold circuit size of AC^0* , SIAM J. Comput., 41 (2012), pp. 484–518, doi:10.1137/100792779.
- [12] P. BEAME, T. PITASSI, N. SEGERLIND, AND A. WIGDERSON, *A strong direct product theorem for corruption and the multiparty communication complexity of disjointness*, Computational Complexity, 15 (2006), pp. 391–432, doi:10.1007/s00037-007-0220-2.
- [13] H. BUHRMAN AND R. DE WOLF, *Communication complexity lower bounds by polynomials*, in *Proceedings of the Sixteenth Annual IEEE Conference on Computational Complexity (CCC)*, 2001, pp. 120–130, doi:10.1109/CCC.2001.933879.
- [14] M. BUN, R. KOTHARI, AND J. THALER, *The polynomial method strikes back: Tight quantum query bounds via dual polynomials*, Theory Comput., 16 (2020), pp. 1–71, doi:10.4086/toc.2020.v016a010.
- [15] M. BUN AND J. THALER, *Dual lower bounds for approximate degree and Markov–Bernstein inequalities*, Inf. Comput., 243 (2015), pp. 2–25, doi:10.1016/j.ic.2014.12.003.
- [16] M. BUN AND J. THALER, *Hardness amplification and the approximate degree of constant-depth circuits*, in *Proceedings of the Forty-Second International Colloquium on Automata, Languages and Programming (ICALP)*, 2015, pp. 268–280, doi:10.1007/978-3-662-47672-7_22.
- [17] M. BUN AND J. THALER, *A nearly optimal lower bound on the approximate degree of AC^0* , SIAM J. Comput., 49 (2020), doi:10.1137/17M1161737.
- [18] M. BUN AND J. THALER, *The large-error approximate degree of AC^0* , Theory of Computing, 17 (2021), pp. 1–46, doi:10.4086/toc.2021.v017a007.
- [19] A. K. CHANDRA, M. L. FURST, AND R. J. LIPTON, *Multi-party protocols*, in *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing (STOC)*, 1983, pp. 94–99, doi:10.1145/800061.808737.
- [20] A. CHATTOPADHYAY AND A. ADA, *Multiparty communication complexity of disjointness*, in *Electronic Colloquium on Computational Complexity (ECCC)*, January 2008. Report TR08-002.
- [21] H. CHERNOFF, *A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations*, Ann. Math. Statist., 23 (1952), pp. 493–507.
- [22] M. DAVID, T. PITASSI, AND E. VIOLA, *Improved separations between nondeterministic and randomized multiparty communication*, ACM Transactions on Computation Theory (TOCT), 1 (2009), doi:10.1145/1595391.1595392.
- [23] D. GAVINSKY AND A. A. SHERSTOV, *A separation of NP and coNP in multiparty communication complexity*, Theory of Computing, 6 (2010), pp. 227–245, doi:10.4086/toc.2010.v006a010.
- [24] W. HOEFFDING, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association, 58 (1963), pp. 13–30, doi:10.1080/01621459.1963.10500830.
- [25] B. KALYANASUNDARAM AND G. SCHNITGER, *The probabilistic communication complexity of set intersection*, SIAM J. Discrete Math., 5 (1992), pp. 545–557, doi:10.1137/0405044.
- [26] T. LEE, *A note on the sign degree of formulas*, 2009. Available at <http://arxiv.org/abs/0909.4607>.
- [27] T. LEE, G. SCHECHTMAN, AND A. SHRAIBMAN, *Lower bounds on quantum multiparty communication complexity*, in *Proceedings of the Twenty-Fourth Annual IEEE Conference on Computational Complexity (CCC)*, 2009, pp. 254–262, doi:10.1109/CCC.2009.24.
- [28] T. LEE AND A. SHRAIBMAN, *Disjointness is hard in the multiparty number-on-the-forehead model*, Computational Complexity, 18 (2009), pp. 309–336, doi:10.1007/s00037-009-0276-2.
- [29] N. S. MANDE, J. THALER, AND S. ZHU, *Improved approximate degree bounds for k -distinctness*, in *Proceedings of the 15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC)*, vol. 158, 2020, pp. 2:1–2:22, doi:10.4230/LIPIcs.TQC.2020.2.
- [30] M. L. MINSKY AND S. A. PAPERT, *Perceptrons: An Introduction to Computational Geometry*, MIT Press, Cambridge, Mass., 1969.
- [31] N. NISAN AND M. SZEGEDY, *On the degree of Boolean functions as real polynomials*, Computational Complexity, 4 (1994), pp. 301–313, doi:10.1007/BF01263419.
- [32] A. A. RAZBOROV, *On the distributional complexity of disjointness*, Theor. Comput. Sci., 106 (1992), pp. 385–390, doi:10.1016/0304-3975(92)90260-M.

- [33] A. A. RAZBOROV, *Quantum communication complexity of symmetric predicates*, Izvestiya of the Russian Academy of Sciences, Mathematics, 67 (2002), pp. 145–159.
- [34] A. A. RAZBOROV AND A. A. SHERSTOV, *The sign-rank of AC^0* , SIAM J. Comput., 39 (2010), pp. 1833–1855, doi:10.1137/080744037.
- [35] B. ROSSER, *Explicit bounds for some functions of prime numbers*, American Journal of Mathematics, 63 (1941), pp. 211–232.
- [36] A. A. SHERSTOV, *Communication lower bounds using dual polynomials*, Bulletin of the EATCS, 95 (2008), pp. 59–93.
- [37] A. A. SHERSTOV, *Separating AC^0 from depth-2 majority circuits*, SIAM J. Comput., 38 (2009), pp. 2113–2129, doi:10.1137/08071421X.
- [38] A. A. SHERSTOV, *The pattern matrix method*, SIAM J. Comput., 40 (2011), pp. 1969–2000, doi:10.1137/080733644.
- [39] A. A. SHERSTOV, *Strong direct product theorems for quantum communication and query complexity*, SIAM J. Comput., 41 (2012), pp. 1122–1165, doi:10.1137/110842661.
- [40] A. A. SHERSTOV, *Approximating the AND-OR tree*, Theory of Computing, 9 (2013), pp. 653–663, doi:10.4086/toc.2013.v009a020.
- [41] A. A. SHERSTOV, *The intersection of two halfspaces has high threshold degree*, SIAM J. Comput., 42 (2013), pp. 2329–2374, doi:10.1137/100785260.
- [42] A. A. SHERSTOV, *Making polynomials robust to noise*, Theory of Computing, 9 (2013), pp. 593–615, doi:10.4086/toc.2013.v009a018.
- [43] A. A. SHERSTOV, *Communication lower bounds using directional derivatives*, J. ACM, 61 (2014), pp. 1–71, doi:10.1145/2629334.
- [44] A. A. SHERSTOV, *The multiparty communication complexity of set disjointness*, SIAM J. Comput., 45 (2016), pp. 1450–1489, doi:10.1137/120891587.
- [45] A. A. SHERSTOV, *Breaking the Minsky–Papert barrier for constant-depth circuits*, SIAM J. Comput., 47 (2018), pp. 1809–1857, doi:10.1137/15M1015704.
- [46] A. A. SHERSTOV, *The power of asymmetry in constant-depth circuits*, SIAM J. Comput., 47 (2018), pp. 2362–2434, doi:10.1137/16M1064477.
- [47] A. A. SHERSTOV, *Algorithmic polynomials*, SIAM J. Comput., 49 (2020), pp. 1173–1231, doi:10.1137/19M1278831.
- [48] A. A. SHERSTOV, *The hardest halfspace*, Comput. Complex., 30 (2021), pp. 1–85, doi:10.1007/s00037-021-00211-4.
- [49] A. A. SHERSTOV AND P. WU, *Near-optimal lower bounds on the threshold degree and sign-rank of AC^0* , in *Proceedings of the Fifty-First Annual ACM Symposium on Theory of Computing (STOC)*, 2019, pp. 401–412, doi:10.1145/3313276.3316408.
- [50] R. ŠPALEK, *A dual polynomial for OR*. Available at <http://arxiv.org/abs/0803.4516>, 2008.
- [51] R. DE WOLF, *Quantum Computing and Communication Complexity*, PhD thesis, University of Amsterdam, 2001.

APPENDIX A. CONSTRUCTING LOW-DISCREPANCY INTEGER SETS

The purpose of this appendix is to provide a detailed and self-contained proof of Theorem 3.5, restated below.

THEOREM. *Fix an integer $R \geq 1$ and reals $P \geq 2$ and $\Delta \geq 1$. Let m be an integer with*

$$m \geq P^2(R + 1).$$

Fix a set $S_p \subseteq \{1, 2, \dots, p - 1\}$ for each prime $p \in (P/2, P]$ with $p \nmid m$. Suppose further that the cardinalities of any two sets from among the S_p differ by a factor of at most Δ . Consider the multiset

$$S = \{(r + s \cdot (p^{-1})_m) \bmod m : \\ r = 1, \dots, R; \quad p \in (P/2, P] \text{ prime with } p \nmid m; \quad s \in S_p\}.$$

Then the elements of S are pairwise distinct and nonzero. Moreover, if $S \neq \emptyset$ then

$$\text{disc}_m(S) \leq \frac{c}{\sqrt{R}} + \frac{c \log m}{\log \log m} \cdot \frac{\log P}{P} \cdot \Delta + \max_p \{\text{disc}_p(S_p)\} \quad (\text{A.1})$$

for some (explicitly given) constant $c \geq 1$ independent of P, R, m, Δ .

The special case $\Delta = 1$ in this result was proved in [48, Theorem 3.6], and that proof applies with cosmetic changes to any $\Delta \geq 1$. As a service to the reader, we provide the complete derivation below; the treatment here is the same word for word as in [48] except for one minor point of departure to handle arbitrary $\Delta \geq 1$. We use the same notation as in [48] and in particular denote the modulus by lowercase m , as opposed to the uppercase M in the main body of our paper (Theorem 3.5). Analogous to [48], the presentation is broken down into five key milestones, corresponding to Sections A.1–A.5 below.

A.1. Exponential notation. In the remainder of this manuscript, we adopt the shorthand

$$e(x) = \exp(2\pi x \mathbf{i}),$$

where \mathbf{i} is the imaginary unit. We will need the following bounds [48, Section 6.1]:

$$|1 - e(x)| \leq 2\pi x, \quad 0 \leq x \leq 1, \quad (\text{A.2})$$

$$|1 - e(x)| \geq 4 \min(x, 1 - x), \quad 0 \leq x \leq 1. \quad (\text{A.3})$$

Let \mathcal{P} denote the set of prime numbers $p \in (P/2, P]$ with $p \nmid m$. In this notation, the multiset S is given by

$$S = \{(r + s \cdot (p^{-1})_m) \bmod m : p \in \mathcal{P}, s \in S_p, r = 1, 2, \dots, R\}.$$

There are precisely $\pi(P) - \pi(P/2)$ primes in $(P/2, P]$, of which at most $\nu(m)$ are prime divisors of m . Therefore,

$$|\mathcal{P}| \geq \pi(P) - \pi\left(\frac{P}{2}\right) - \nu(m). \quad (\text{A.4})$$

A.2. Elements of S are nonzero and distinct. As our first step, we verify that the elements of S are nonzero and distinct modulo m . This part of the argument is reproduced word for word from [48, Section 6.2].

Specifically, consider any $r \in \{1, 2, \dots, R\}$, any prime $p \in (P/2, P]$ with $p \nmid m$, and any $s \in S_p$. Then $pr + s \in [1, PR + P - 1] \subseteq [1, m)$. This means that $pr + s \not\equiv 0 \pmod{m}$, which in turn implies that $r + s \cdot (p^{-1})_m \not\equiv 0 \pmod{m}$.

We now show that the multiset S contains no repeated elements. For this, consider any $r, r' \in \{1, 2, \dots, R\}$, any primes $p, p' \in \mathcal{P}$, and any $s \in S_p$ and $s' \in S_{p'}$ such that

$$r + s \cdot (p^{-1})_m \equiv r' + s' \cdot (p'^{-1})_m \pmod{m}. \quad (\text{A.5})$$

Our goal is to show that $p = p', r = r', s = s'$. To this end, multiply (A.5) through by pp' to obtain

$$r \cdot pp' + s \cdot p' \equiv r' \cdot pp' + s' \cdot p \pmod{m}. \quad (\text{A.6})$$

The left-hand side and right-hand side of (A.6) are integers in $[1, RP^2 + (P-1)P] \subseteq [1, m)$, whence

$$r \cdot pp' + s \cdot p' = r' \cdot pp' + s' \cdot p. \quad (\text{A.7})$$

This implies that $p \mid s \cdot p'$, which in view of $s < p$ and the primality of p and p' forces $p = p'$. Now (A.7) simplifies to

$$r \cdot p + s = r' \cdot p + s', \quad (\text{A.8})$$

which in turn yields $s \equiv s' \pmod{p}$. Recalling that $s, s' \in \{1, 2, \dots, p-1\}$, we arrive at $s = s'$. Finally, substituting $s = s'$ in (A.8) gives $r = r'$.

A.3. Correlation for k small. So far, we have shown that the elements of S are distinct and nonzero. To bound the m -discrepancy of this set, we must bound the exponential sum

$$\left| \sum_{s \in S} e\left(\frac{k}{m} \cdot s\right) \right| \quad (\text{A.9})$$

for all $k = 1, 2, \dots, m-1$. This subsection and the next provide two complementary bounds on (A.9). The first bound, presented below, is preferable when k is close to zero modulo m .

CLAIM A.1. *Let $k \in \{1, 2, \dots, m-1\}$ be given. Then*

$$\left| \sum_{s \in S} e\left(\frac{k}{m} \cdot s\right) \right| \leq \left(\frac{2\pi \min(k, m-k)}{m} + \max_{p \in \mathcal{P}} \{\text{disc}_p(S_p)\} + \frac{\nu(k) + \nu(m-k)}{|\mathcal{P}|} \cdot \Delta \right) |S|.$$

This claim generalizes the analogous statement in [48, Claim 6.10], where the special case $\Delta = 1$ was considered.

Proof. Let \mathcal{P}' be the set of those primes in \mathcal{P} that divide neither k nor $m-k$. Then clearly

$$|\mathcal{P} \setminus \mathcal{P}'| \leq \nu(k) + \nu(m-k). \quad (\text{A.10})$$

Exactly as in [48], we have

$$\begin{aligned}
& \left| \sum_{s \in S} e\left(\frac{k}{m} \cdot s\right) \right| \\
&= \left| \sum_{r=1}^R \sum_{p \in \mathcal{P}} \sum_{s \in S_p} e\left(\frac{k}{m} \cdot (r + s \cdot (p^{-1})_m)\right) \right| \\
&\leq \sum_{r=1}^R \sum_{p \in \mathcal{P}} \left| \sum_{s \in S_p} e\left(\frac{k}{m} \cdot (r + s \cdot (p^{-1})_m)\right) \right| \\
&= R \sum_{p \in \mathcal{P}} \left| \sum_{s \in S_p} e\left(\frac{ks \cdot (p^{-1})_m}{m}\right) \right| \\
&\leq R \sum_{p \in \mathcal{P}'} \left| \sum_{s \in S_p} e\left(\frac{ks \cdot (p^{-1})_m}{m}\right) \right| + R \sum_{p \in \mathcal{P} \setminus \mathcal{P}'} \left| \sum_{s \in S_p} e\left(\frac{ks \cdot (p^{-1})_m}{m}\right) \right| \\
&\leq R \sum_{p \in \mathcal{P}'} \left| \sum_{s \in S_p} e\left(\frac{ks \cdot (p^{-1})_m}{m}\right) \right| + R \sum_{p \in \mathcal{P} \setminus \mathcal{P}'} |S_p|. \tag{A.11}
\end{aligned}$$

We proceed to bound the two summations in (A.11). Bounding the second summation is straightforward:

$$\begin{aligned}
R \sum_{p \in \mathcal{P} \setminus \mathcal{P}'} |S_p| &\leq R \cdot \frac{|\mathcal{P} \setminus \mathcal{P}'|}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \Delta |S_p| \\
&= \frac{|\mathcal{P} \setminus \mathcal{P}'|}{|\mathcal{P}|} \cdot \Delta |S| \\
&\leq \frac{\nu(k) + \nu(m-k)}{|\mathcal{P}|} \cdot \Delta |S|, \tag{A.12}
\end{aligned}$$

where the first step is valid because the cardinalities of any two sets S_p differ by a factor of at most Δ , and the last step uses (A.10). This three-line derivation is our only point of departure from the treatment in [48].

The other summation in (A.11) is analyzed exactly as in [48]. For $p \in \mathcal{P}'$ and $K \in \{k, k-m\}$, we have

$$\begin{aligned}
& \left| \sum_{s \in S_p} e\left(\frac{ks \cdot (p^{-1})_m}{m}\right) \right| \\
&= \left| \sum_{s \in S_p} e\left(\frac{Ks \cdot (p^{-1})_m}{m}\right) \right| \\
&= \left| \sum_{s \in S_p} e\left(-\frac{Ks \cdot (m^{-1})_p}{p}\right) e\left(\frac{Ks}{pm}\right) \right|
\end{aligned}$$

$$\begin{aligned}
&\leq \left| \sum_{s \in S_p} e\left(-\frac{Ks \cdot (m^{-1})_p}{p}\right) \left(e\left(\frac{Ks}{pm}\right) - 1\right) \right| + \left| \sum_{s \in S_p} e\left(-\frac{Ks \cdot (m^{-1})_p}{p}\right) \right| \\
&\leq \left| \sum_{s \in S_p} e\left(-\frac{Ks \cdot (m^{-1})_p}{p}\right) \left(e\left(\frac{Ks}{pm}\right) - 1\right) \right| + \text{disc}_p(S_p) \cdot |S_p| \\
&\leq \sum_{s \in S_p} \left| e\left(\frac{Ks}{pm}\right) - 1 \right| + \text{disc}_p(S_p) \cdot |S_p| \\
&= \sum_{s \in S_p} \left| e\left(\frac{|K|s}{pm}\right) - 1 \right| + \text{disc}_p(S_p) \cdot |S_p| \\
&\leq |S_p| \cdot \frac{2\pi|K|}{m} + \text{disc}_p(S_p) \cdot |S_p|,
\end{aligned}$$

where the second step uses Fact 2.16 and the relative primality of p and m ; the third step applies the triangle inequality; the fourth step follows from $p \nmid |K|$, and the last step is valid by (A.2) and $s < p$. We have shown that

$$\left| \sum_{s \in S_p} e\left(\frac{ks \cdot (p^{-1})_m}{m}\right) \right| \leq \frac{2\pi \min(k, m-k)}{m} \cdot |S_p| + \text{disc}_p(S_p) \cdot |S_p|$$

for $p \in \mathcal{P}'$. Summing over \mathcal{P}' ,

$$\begin{aligned}
R \sum_{p \in \mathcal{P}'} \left| \sum_{s \in S_p} e\left(\frac{ks \cdot (p^{-1})_m}{m}\right) \right| &\leq R \sum_{p \in \mathcal{P}'} \left(\frac{2\pi \min(k, m-k)}{m} \cdot |S_p| + \text{disc}_p(S_p) \cdot |S_p| \right) \\
&\leq R \sum_{p \in \mathcal{P}} \left(\frac{2\pi \min(k, m-k)}{m} \cdot |S_p| + \text{disc}_p(S_p) \cdot |S_p| \right) \\
&\leq \left(\frac{2\pi \min(k, m-k)}{m} + \max_{p \in \mathcal{P}} \{\text{disc}_p(S_p)\} \right) R \sum_{p \in \mathcal{P}} |S_p| \\
&= \left(\frac{2\pi \min(k, m-k)}{m} + \max_{p \in \mathcal{P}} \{\text{disc}_p(S_p)\} \right) |S|. \tag{A.13}
\end{aligned}$$

By (A.11)–(A.13), the proof of the claim is complete. \square

A.4. Correlation for k large. We now present an alternative bound on the exponential sum (A.9), which is preferable to the bound of Claim A.1 when k is far from zero modulo m . This part of the proof is reproduced verbatim from [48, Section 6.4].

CLAIM A.2. *Let $k \in \{1, 2, \dots, m-1\}$ be given. Then*

$$\left| \sum_{s \in S} e\left(\frac{k}{m} \cdot s\right) \right| \leq \frac{m}{2R \min(k, m-k)} \cdot |S|.$$

Proof:

$$\begin{aligned}
\left| \sum_{s \in S} e\left(\frac{k}{m} \cdot s\right) \right| &= \left| \sum_{p \in \mathcal{P}} \sum_{s \in S_p} \sum_{r=1}^R e\left(\frac{k}{m} \cdot (r + s \cdot (p^{-1})_m)\right) \right| \\
&\leq \sum_{p \in \mathcal{P}} \sum_{s \in S_p} \left| \sum_{r=1}^R e\left(\frac{k}{m} \cdot (r + s \cdot (p^{-1})_m)\right) \right| \\
&= \sum_{p \in \mathcal{P}} \sum_{s \in S_p} \left| \sum_{r=1}^R e\left(\frac{kr}{m}\right) \right| \\
&= \sum_{p \in \mathcal{P}} \sum_{s \in S_p} \frac{|1 - e(kR/m)|}{|1 - e(k/m)|} \\
&\leq \sum_{p \in \mathcal{P}} \sum_{s \in S_p} \frac{2}{|1 - e(k/m)|} \\
&\leq \sum_{p \in \mathcal{P}} \sum_{s \in S_p} \frac{m}{2 \min(k, m-k)} \\
&= \frac{m}{2R \min(k, m-k)} \cdot |S|,
\end{aligned}$$

where the last two steps use (A.3) and $|S| = R \sum_{p \in \mathcal{P}} |S_p|$, respectively. \square

A.5. Finishing the proof. The remainder of the proof is reproduced without changes from [48, Section 6.5], except for the use of the updated bound in Claim A.1 for arbitrary $\Delta \geq 1$.

Specifically, Facts 2.17 and 2.18 imply that

$$\pi(P) - \pi\left(\frac{P}{2}\right) \geq \frac{P}{C \log P} \quad (P \geq C), \quad (\text{A.14})$$

$$\max_{k=1,2,\dots,m} \nu(k) \leq \frac{C \log m}{\log \log m}, \quad (\text{A.15})$$

where $C \geq 1$ is a constant independent of R, P, m, Δ . Moreover, C can be easily calculated from the explicit bounds in Facts 2.17 and 2.18. We will show that the theorem conclusion (A.1) holds with $c = 4C^2$. We may assume that

$$P \geq C, \quad (\text{A.16})$$

$$\frac{C \log m}{\log \log m} \leq \frac{P}{2C \log P}, \quad (\text{A.17})$$

since otherwise the right-hand side of (A.1) exceeds 1 and the theorem is trivially true. By (A.4) and (A.14)–(A.17), we obtain

$$|\mathcal{P}| \geq \frac{P}{2C \log P},$$

which along with (A.15) gives

$$\begin{aligned} \max_{k=1,2,\dots,m-1} \frac{\nu(k) + \nu(m-k)}{|\mathcal{P}|} &\leq \frac{2C \log m}{\log \log m} \cdot \frac{2C \log P}{P} \\ &= \frac{c \log m}{\log \log m} \cdot \frac{\log P}{P}. \end{aligned} \quad (\text{A.18})$$

Claims A.1 and A.2 ensure that for every $k = 1, 2, \dots, m-1$,

$$\begin{aligned} \left| \sum_{s \in S} e\left(\frac{k}{m} \cdot s\right) \right| &\leq \left(\min\left(\frac{2\pi \min(k, m-k)}{m}, \frac{m}{2R \min(k, m-k)}\right) \right. \\ &\quad \left. + \max_{p \in \mathcal{P}} \{\text{disc}_p(S_p)\} + \frac{\nu(k) + \nu(m-k)}{|\mathcal{P}|} \cdot \Delta \right) |S| \\ &\leq \left(\sqrt{\frac{\pi}{R}} + \max_{p \in \mathcal{P}} \{\text{disc}_p(S_p)\} + \frac{\nu(k) + \nu(m-k)}{|\mathcal{P}|} \cdot \Delta \right) |S| \\ &\leq \left(\frac{c}{\sqrt{R}} + \max_{p \in \mathcal{P}} \{\text{disc}_p(S_p)\} + \frac{\nu(k) + \nu(m-k)}{|\mathcal{P}|} \cdot \Delta \right) |S|; \end{aligned}$$

here we are using the updated bound from Claim A.1 in this paper for general Δ . Substituting the estimate from (A.18), we conclude that

$$\begin{aligned} \max_{k=1,2,\dots,m-1} \left| \sum_{s \in S} e\left(\frac{k}{m} \cdot s\right) \right| \\ \leq \left(\frac{c}{\sqrt{R}} + \max_{p \in \mathcal{P}} \{\text{disc}_p(S_p)\} + \frac{c \log m}{\log \log m} \cdot \frac{\log P}{P} \cdot \Delta \right) |S|. \end{aligned}$$

This conclusion is equivalent to (A.1). The proof of Theorem 3.5 is complete.