Electronic Colloquium on Computational Complexity, Report No. 20 (2020)

Improved Approximate Degree Bounds For k-distinctness

Nikhil S. Mande
Georgetown University
nikhil.mande@georgetown.edu

Justin Thaler
Georgetown University
justin.thaler@georgetown.edu

Shuchen Zhu
Georgetown University
sz424@georgetown.edu

Abstract

An open problem that is widely regarded as one of the most important in quantum query complexity is to resolve the quantum query complexity of the k-distinctness function on inputs of size N. While the case of $k = 2$ (also called Element Distinctness) is well-understood, there is a polynomial gap between the known upper and lower bounds for all constants $k > 2$. Specifically, the best known upper bound is $O\left(N^{(3/4)-1/(2^{k+2}-4)}\right)$ (Belovs, FOCS 2012), while the best known lower bound for $k \geq 2$ is $\tilde{\Omega}\left(N^{2/3} + N^{(3/4)-1/(2k)}\right)$ (Aaronson and Shi, J. ACM 2004; Bun, Kothari, and Thaler, STOC 2018).

For any constant $k \geq 4$, we improve the lower bound to $\tilde{\Omega}\left(N^{(3/4)-1/(4k)}\right)$. This yields, for example, the first proof that 4-distinctness is strictly harder than Element Distinctness. Our lower bound applies more generally to approximate degree.

As a secondary result, we give a simple construction of an approximating polynomial of degree $\tilde{O}(N^{3/4})$ that applies whenever $k \leq \text{polylog}(N)$.

1 Introduction

In quantum query complexity, a quantum algorithm is given query access to the bits of an unknown input x, and the goal is to compute some (known) function f of x while minimizing the number of bits of x that are queried. In contrast to classical query complexity, quantum query algorithms are allowed to make queries in superposition, and the algorithm is not charged for performing unitary operations that are independent of x. Quantum query complexity is a rich model that allows for the design of highly sophisticated algorithms and captures much of the power of quantum computing. Indeed, most quantum algorithms were discovered in or can easily be described in the query setting.

An open problem that is widely regarded as one of the most important in quantum query complexity [LZ19] is to resolve the complexity of the *k-distinctness* function. For this function, the input x specifies a list of N numbers from a given range of size R,[1] and the function evaluates to TRUE[2] if there is any range item that appears k or more times in the list. The case $k = 2$ corresponds to the complement of the widely-studied *Element Distinctness* function, whose complexity is known to be $\Theta(N^{2/3})$ [Amb07, AS04].

For general values of k, the best known upper bound on the quantum query complexity of k-distinctness is $O\left(N^{3/4-1/(2^{k+2}-4)}\right)$, due to a highly sophisticated algorithm of Belovs [Bel12]. Belovs' algorithm is based on the so-called learning graph framework in quantum algorithm design, and improves over an earlier upper bound of $O(N^{k/(k+1)})$ due to Ambainis [Amb07] that is based on quantum walks over the Johnson graph.

[1] For purposes of this introduction, N and R are assumed to be of the same order of magnitude (up to a factor depending on k alone). For simplicity throughout this section, we state our bounds purely in terms of N, leaving unstated the assumption that R and N are of the same order of magnitude.

[2] Throughout this manuscript, we associate -1 with logical TRUE and $+1$ with logical FALSE.

ISSN 1433-8092

For a long time, the best known lower bound on the quantum query complexity of k-distinctness was $\Omega(N^{2/3})$ for any $k \geq 2$, due to Aaronson and Shi [AS04], with refinements given by Kutin [Kut05] and Ambainis [Amb05]. This lower bound is tight for $k = 2$ (matching Ambainis' upper bound [Amb07]), but it is not known to be tight for any $k > 2$. Recently, Bun, Kothari, and Thaler [BKT18] proved a lower bound of $\tilde{\Omega}(N^{3/4-1/(2k)})$ for constant k.[3] This improved over the prior lower bound of $\Omega(N^{2/3})$ for any constant $k \geq 7$. Furthermore, combined with Belovs' upper bound, this established that for sufficiently large constants k, the exponent in the quantum query complexity of k-distinctness approaches $3/4$ from below. However, the precise rate at which the quantum query complexity approaches $N^{3/4}$ remains open: there is a polynomial gap between the upper and lower bounds for any constant k, and indeed there is a qualitative difference between the inverse-exponential dependence on k in the exponent of $N^{3/4-1/(2^{k+2}-4)}$ (the known upper bound), and the inverse-linear dependence in the known lower bound of $N^{3/4-1/(2k)}$.

Main Result. In this paper, our main result improves the lower bound from $\tilde{\Omega}(N^{3/4-1/(2k)})$ to $\tilde{\Omega}(N^{3/4-1/(4k)})$. While this bound is qualitatively similar to the lower bound of [BKT18], it offers a polynomial improvement for every constant $k \geq 4$. Perhaps more significantly, for $k \in \{4, 5, 6\}$, it is the first improvement over Aaronson and Shi's $\Omega(N^{2/3})$ lower bound that has stood for nearly 20 years.

Approximate Degree. The *ϵ-error approximate degree* of a Boolean function $f\colon \{-1,1\}^n \to \{-1,1\}$, denoted $\widetilde{\deg}_\epsilon(f)$, is the least degree of a real polynomial p such that $|p(x) - f(x)| \leq \epsilon$ for all $x \in \{-1,1\}^n$. The standard setting of the error parameter is $\epsilon = 1/3$, and the $(1/3)$-approximate degree of f is denoted $\widetilde{\deg}(f)$ for brevity.

As famously observed by Beals et al. [BBC+01], the quantum query complexity of a function f is lower bounded by (one half times) the approximate degree of f. Hence, any lower bound on the approximate degree of f implies that (up to a factor of 2) the same lower bound holds for the quantum query complexity of f.

As with prior lower bounds for k-distinctness [AS04, Kut05, Amb05, BKT18], our k-distinctness lower bound is in fact an approximate degree lower bound (on the natural Boolean function induced by k-distinctness on $N\lceil \log_2 R \rceil$ bits, where R denotes the size of the range). Our analysis is a substantial refinement of the lower bound analysis of Bun et al. [BKT18].

Theorem 1.1 (Informal version of Theorem 4.1 and Corollary 4.2)**.** *For any constant $k \geq 2$, the approximate degree and quantum query complexity of the k-distinctness function with domain size N and range size $R \geq N$ is $\tilde{\Omega}(N^{3/4-1/(4k)})$.*

Remark 1.2. *Theorem 1.1 provides an approximate degree lower bound for constant error $\epsilon = 1/3$. A recent result of Sherstov and Thaler [ST19, Theorem 3.4] transforms any constant-error approximate degree lower bound for k-distinctness, into a lower bound for vanishing error $\epsilon = o(1)$. Specifically, combining Theorem 1.1 and [ST19, Theorem 3.4] yields that for constant k, the ϵ-error approximate degree of k-distinctness is at least $\tilde{\Omega}\left(N^{3/4-1/(4k)} \log^{1/4+1/(4k)}(1/\epsilon)\right)$, for all $\epsilon \in [(1/3)^N, 1/3]$.*

A Secondary Result: The Approximate Degree for Super-Constant Values of k. Recall that for constant k, the best known approximate degree upper bound for k-distinctness, due to Belovs, is $O\left(N^{3/4-1/(2^{k+2}-4)}\right)$. For non-constant values of k, the upper bound implied by Belovs'

[3] Throughout this manuscript, $\tilde{O}$, $\tilde{\Omega}$ and $\tilde{\Theta}$ notations are used to hide factors that are polylogarithmic in N.

algorithm grows exponentially with k. That is, the Big-Oh notation in the upper bound hides a leading factor of at least 2^{ck} for some positive constant c.[4] Consequently Belovs' result is $N^{3/4+\Omega(1)}$ for any $k \geq \Omega(\log N)$. Furthermore, the bound becomes vacuous (i.e., linear in N) for $k \geq c \log N$ for a large enough constant $c > 0$.

Our secondary result improves this state of affairs by giving a $\tilde{O}(N^{3/4})$ approximate degree upper bound that holds for any value of k that grows at most polylogarithmically with N.

Theorem 1.3 (Informal)**.** *For any $k \leq \text{polylog}(N)$, the approximate degree of k-distinctness is $\tilde{O}(N^{3/4})$.*

We mention that for *any* $k \geq 2$, the approximating polynomials for k-distinctness that follow from prior works [Amb07, Bel12, She18a] are quite complicated, and in our opinion there has not been a genuinely simple construction of any $O(N^{3/4})$-degree approximating polynomials recorded in the literature, even for the case of $k = 2$ (i.e., Element Distinctness). Accordingly, we feel that Theorem 1.3 has didactic value even for constant values of k (though the $\tilde{O}(N^{3/4})$ approximate degree upper bound that it achieves is not tight for any constant $k \geq 2$).

To clarify, Theorem 1.3 does *not* yield a quantum query upper bound, but only an approximate degree upper bound. Indeed, it remains an interesting open question whether the quantum query complexity of k-distinctness is sublinear in N for all $k = \text{polylog}(N)$ (see Section 1.1 for further discussion).

Our proof of Theorem 1.3 is a simple extension of a result of Sherstov [She18a, Theorem 1.3] that yielded an $O(N^{3/4})$ approximate degree upper bound for a different function called Surjectivity.[5] In Section 2.2 below, we explain the main observations necessary to obtain Theorem 1.3 via the technique used to prove the upper bound for Surjectivity.

1.1 Discussion and Open Problems

The most obvious and important open question is to finish resolving the approximate degree and quantum query complexity of k-distinctness for any $k > 2$. Currently, the upper and lower bounds qualitatively differ in their dependence on k, with the upper bound having an exponent of the form $3/4 - \exp(-O(k))$ and the lower bound having an exponent of the from $3/4 - \Omega(1/k)$. It seems very likely that major new techniques will be needed to qualitatively change the form of *either* the upper or lower bound. In particular, on the lower bounds side, our analysis is based on a variant of a technique called *dual block composition* (see Section 2.1), and we suspect that we have reached the limit of what is provable for k-distinctness using this technique and its variants.

We remark here that Liu and Zhandry [LZ19] recently showed that the quantum query complexity of a certain *search* version of k-distinctness (defined over randomly generated inputs) is $\Theta(n^{1/2-1/(2^k-1)})$. This inverse-exponential dependence on k is tantalizingly reminsicent of Belovs' upper bound for k-distinctness. This may be construed as mild evidence that $3/4 - \exp(-O(k))$ is the right qualitative bound for k-distinctness itself.

A very interesting intermediate goal is to establish any polynomial improvement over the long-standing $\Omega(n^{2/3})$ lower bound for 3-distinctness. This would finally establish that 3-distinctness is strictly harder than Element Distinctness (such a result is now known for all $k \geq 4$ due to Theorem 1.1).

[4] Belovs' approximate degree upper bound was recently reproved by Sherstov [She18a], who made the exponential dependence on k explicit (see, e.g., [She18a, Theorem 6.6]). To clarify, Belovs' result is in fact a quantum query upper bound, which in turn implies an approximate degree upper bound. Sherstov's proof avoids quantum algorithms, and hence does not yield a quantum query upper bound.

[5] Surjectivity is the function that interprets its input as a list of N numbers from a given range of size R, and evaluates to TRUE if and only if every range element appears at least once in the list.

It would also be interesting to resolve the quantum query complexity of k-distinctness for $k = \text{polylog}(N)$. Although this question my appear to be of specialized interest, we believe that resolving it could shed light on the relationship between approximate degree and quantum query complexity. Indeed, while any quantum algorithm for a function f can be turned into an approximating polynomial for f via the transformation of Beals et al. [BBC+01], no transformation in the reverse direction is possible in general [Amb06]. This can be seen, for example, because the quantum query complexity of Surjectivity is known to be $\Omega(N)$ [BM12, She18b], but its approximate degree is $O(N^{3/4})$ [She18a, BKT18]. Nonetheless, approximate degree and quantum query complexity turn out to coincide for most functions that arise naturally (Surjectivity remains the only function that exhibits a separation, without having been specifically constructed for that purpose). In our opinion, this phenomenon remains mysterious, and it would be interesting to demystify it. For example, could one identify special properties of approximating polynomials that would permit a reverse-Beals-et-al. transformation to turn that polynomial into a quantum query algorithm?[6] Perhaps an $\tilde{O}(N^{3/4})$ upper bound for (polylog(N))-distinctness could be derived in this manner. On the other hand, due to our Theorem 1.3, any $N^{3/4+\Omega(1)}$ *lower bound* for (polylog(N))-distinctness would require moving beyond the polynomial method.[7]

1.2 Paper Roadmap

We give a high-level overview of the proofs of our lower bound and upper bound in Sections 2.1 and 2.2, respectively. Section 3 covers preliminaries. The proof of our main theorem (Theorem 1.1) is spread over Sections 4-6. Section 4 gives a detailed, technical outline of the proof, Section 5 establishes some auxiliary lemmas, and Section 6 contains the heart of the proof. Finally, Section 7 proves Theorem 1.3.

2 Overview of the Proofs

In this section we give an overview of the proofs of our lower bound and upper bound.

2.1 The Lower Bound

Throughout this subsection we assume that $k \geq 2$ is an arbitrary but fixed constant.

Let THR_N^k denote the function on N-bit inputs that evaluates to -1 on inputs of Hamming weight at least k, and evaluates to 1 otherwise. For $N \leq n$, let $(\{-1,1\}^n)^{\leq N}$ denote the subset of $\{-1,1\}^n$ consisting of all inputs of Hamming weight at most N. For any function $f_n \colon \{-1,1\}^n \to \{-1,1\}$,[8] let $f_n^{\leq N}$ denote the partial function obtained by restricting the domain of f to $(\{-1,1\}^n)^{\leq N}$, and let $\widetilde{\deg}(f_n^{\leq N})$ denote the least degree of a real polynomial p such that $|p(x) - f_n(x)| \leq 1/3$ for all $x \in (\{-1,1\}^n)^{\leq N}$.

Simplifying very slightly, prior work by Bun and Thaler [BT17] (building on an important lemma of Ambainis [Amb05]) implied that for $k \geq 2$ the approximate degree of k-distinctness is equivalent to $\widetilde{\deg}(f_{RN}^{\leq N})$ for $f = \text{OR}_R \circ \text{THR}_N^k$. Here, $g_n \circ h_m$ denotes the function on $n \cdot m$ bits

[6] There are works in this general direction, notably [ABP19], which shows that a certain technical refinement of approximate degree, called approximation by completely bounded forms, characterizes quantum query complexity. But to our knowledge these works have not yielded any novel quantum query upper bounds for any specific function.

[7] We remark that the positive-weights adversary method is also incapable of proving such a result due to the certificate complexity barrier.

[8] Throughout, we use subscripts where appropriate to clarify the number of bits over which a function is defined.

obtained by block-composing g and h, i.e., $g \circ h$ evaluates h on n disjoint inputs and feeding the outputs of all n copies of h into g.

Bun et al. [BKT18] proved their $\tilde{\Omega}(N^{3/4-1/(2k)})$ lower bound for $\widetilde{\deg}(f_{RN}^{\leq N})$ via the *method of dual polynomials*. This is a technique for proving approximate degree lower bounds that works by constructing an explicit solution to a certain linear program capturing the approximate degree of any function. Specifically, a dual witness to the fact that $\widetilde{\deg}(f_{RN}^{\leq N}) \geq d$ is a function $\psi\colon \{-1,1\}^{RN} \to \mathbb{R}$ satisfying the following properties.

First, ψ must be uncorrelated with all polynomials p of degree at most d, i.e., $\langle \psi, p \rangle = 0$ for all such polynomials p, where $\langle \psi, p \rangle = \sum_{x \in \{-1,1\}^{RN}} \psi(x)p(x)$. Such a ψ is said to have *pure high degree* at least d.

Second, ψ must be well-correlated with f, i.e., $\langle \psi, f \rangle \geq (1/3) \cdot \|\psi\|_1$, where $\|\psi\|_1 := \sum_{x \in \{-1,1\}^{RN}} |\psi(x)|$.

Finally, ψ must equal 0 on inputs in $\{-1,1\}^{RN} \setminus \left(\{-1,1\}^{RN}\right)^{\leq N}$.

To simplify greatly, Bun et al. [BKT18] constructed their dual witness for $\left(\mathrm{OR}_R \circ \mathrm{THR}_N^k\right)^{\leq N}$ roughly as follows. They took a dual witness Ψ for the fact that $\widetilde{\deg}(\mathrm{OR}_R) \geq \Omega(R^{1/2})$ [NS94, Š08, BT15] and a dual witness ϕ for the fact that THR_N^k also has large approximate degree, and they combined Ψ and ϕ in a certain manner (introduced in prior works [SZ09, She13, Lee09]) to get a dual witness for the composed function $\left(\mathrm{OR}_R \circ \mathrm{THR}_N^k\right)^{\leq N}$. The technique used to combine Ψ and ϕ is often called *dual block composition*, and is denoted $\Psi \star \phi$.[9] Dual block composition is defined as follows (below, each $x_i \in \{-1,1\}^N$):

$$(\Psi \star \phi)(x_1, \ldots, x_R) = 2^R \cdot \Psi(\mathrm{sgn}(\phi(x_1)), \ldots, \mathrm{sgn}(\phi(x_R))) \cdot \prod_{i=1}^{R} |\phi(x_i)|/\|\phi\|_1.$$

Here, $\mathrm{sgn}(r)$ equals -1 if $r < 0$ and equals $+1$ if $r > 0$.[10] To show that $\Psi \star \phi$ is a dual witness for the fact that the approximate degree of $\left(\mathrm{OR}_R \circ \mathrm{THR}_N^k\right)^{\leq N}$ is at least d, it is necessary to show that $\Psi \star \phi$ has pure high degree at least d, and that $\Psi \star \phi$ is well-correlated with $\left(\mathrm{OR}_R \circ \mathrm{THR}_N^k\right)^{\leq N}$. It is known that pure high degree increases multiplicatively under the $\star$ operation, and hence the pure high degree calculation for $\Psi \star \phi$ is straightforward. In contrast, the correlation calculation is the key technical challenge and bottleneck in the analysis of [BKT18]. Our key improvement over their work is to modify the construction of the dual witness in a manner that allows for an improved correlation bound.

At a very high level, what we do is replace the dual block composition $\Psi \star \phi$ from the construction of [BKT18] with a *variant* of dual block composition introduced by Sherstov [She12]. Sherstov specifically introduced this variant to address the correlation issues that arise when attempting to use dual block composition to prove approximate degree lower bounds for composed functions, and he used it to prove direct sum and direct product theorems for approximate degree.[11] However, we have to modify even Sherstov's variant of dual block composition in significant ways to render it

[9] To clarify, this entire outline is a major simplification of the actual dual witness construction in [BKT18]. The details provided in the outline of this introduction are chosen to highlight the key technical issues that we must address in this work. Amongst other simplifications in this outline, the actual dual witness from [BKT18] is not $\Psi \star \phi$, but rather a "post-processed" version of $\Psi \star \phi$, where the post-processing step is used to ensure that the dual witness evaluates to 0 on all inputs of Hamming weight more than N.

[10] It is irrelevant how one defines $\mathrm{sgn}(0)$ because if $\phi(x_i) = 0$ for any i, the product $\prod_{i=1}^{R} |\phi(x_i)|/\|\phi\|_1$ forces $\Psi \star \phi$ to 0. For this reason, the remainder of the discussion in this section implicitly assumes that $\phi(x_i) \neq 0$ for all $i \in \{1, \ldots, R\}$.

[11] Variants of dual block composition related to the one introduced in [She12] have played important roles in other recent works on approximate degree lower bounds, e.g., [BT19, ST19].

useful in our context. We now attempt to give an informal sense of our modification and why it is necessary.

For block-composed functions $g \circ h$, the rough idea of any proof attempting to show that $\langle \Psi \star \phi, g \circ h \rangle$ is large is to hope that the following approximate equality holds:

$$\langle \Psi \star \phi, g \circ h \rangle \approx \langle \Psi, g \rangle. \tag{1}$$

If Equation (1) holds even approximately, then the correlation analysis of $\Psi \star \phi$ is complete, since the assumption that Ψ is a dual witness for the high approximate degree of g implies that the right hand side is large.

Equation (1) in fact holds with *exact* equality if ϕ agrees in sign with h at all inputs, i.e., if $\langle \phi, h \rangle = \|\phi\|_1$ [She13, Lee09]. Unfortunately, the fact that ϕ is a dual witness for the large approximate degree of h implies only a much weaker lower bound on $\langle \phi, h \rangle$, namely that

$$\langle \phi, h \rangle \geq (1/3) \cdot \|\phi\|_1. \tag{2}$$

In general, Equation (2) is not enough to ensure that Equation (1) holds even approximately.

A rough intuition for why Equation (1) may fail to hold is the following. The definition of $\Psi \star \phi$ feeds $(\mathrm{sgn}(\phi(x_1)), \ldots, \mathrm{sgn}(\phi(x_R)))$ into Ψ. One can think of $\mathrm{sgn}(\phi(x_i))$ as ϕ's "prediction" about $h(x_i)$, and the fact that $\langle \phi, h \rangle \geq (1/3) \cdot \|\phi\|_1$ means that for an x_i chosen at random from the probability distribution $|\phi|/\|\phi\|_1$, this prediction is correct with probability at least $2/3$. Unfortunately, there are values of x_i for which $\mathrm{sgn}(\phi(x_i)) \neq h(x_i)$, meaning that ϕ's predictions can sometimes be wrong. In this case, in feeding $\mathrm{sgn}(\phi(x_i))$ into Ψ, dual block composition is "feeding an error" into Ψ, and this can cause $\Psi \star \phi$ to "make more errors" (i.e, output a value on an input that disagrees in sign with $g \circ h$ on that same input) than Ψ itself.

That is, there are two reasons $\Psi \star \phi$ may make an error: either Ψ itself may make an error (let us call this Source 1 for errors), and/or one or more copies of ϕ may make an error (let us call this Source 2 for errors).[12]

The first source of error is already fully accounted for in the right hand side of Equation (1). The second source of error is not, and this is the reason that Equation (1) may fail to hold even approximately.

Roughly speaking, while Equation (2) guarantees that $\mathrm{sgn}(\phi(x_i))$ is not "an error" for each i with good probability (i.e., probability at least $2/3$), that still means that with very high probability, $\mathrm{sgn}(\phi(x_i))$ will be in error (i.e., not equal to $h(x_i)$) for *a constant fraction* of blocks $i \in \{1, \ldots, R\}$. Any one of these errors could be enough to cause a Source 2 error.

Fortunately for us, $g = \mathrm{OR}_R$ has low (-1)-*certificate complexity*, meaning that on inputs x in $\mathrm{OR}_R^{-1}(-1)$, to certify that indeed $x \in \mathrm{OR}_R^{-1}(-1)$, it is sufficient to identify just one coordinate of x that equals -1. This renders certain kinds of sign-errors made by ϕ benign. Specifically, letting $S = \{x \colon \phi(x) < 0\}$ and $E^- = S \cap f^{-1}(1)$ denote the false-negative errors made by ϕ, the low (-1)-certificate complexity of OR_R means that it is okay if "a constant fraction of the negative values output by ϕ are in error". That is, so long as

$$\left(\sum_{E^-} |\phi(x)| \right) / \left(\sum_{x \in S} |\phi(x)| \right) = 1 - \Omega(1), \tag{3}$$

the contribution of "false negative errors made by ϕ" to actual Source 2 errors made by $\Psi \star \phi$ is low.

[12] There may be inputs $x = (x_1, \ldots, x_n)$ to $\Psi \star \phi$ that could be classified as *both* Source 1 and Source 2 errors. For purposes of this high-level introduction, it is not important whether such inputs get classified as Source 1 or Source 2 errors for $\Psi \star \phi$.

However, the situation is starkly different for "false positive errors" made by ϕ; while OR_R has certificates of size 1 for inputs in $\mathrm{OR}_R^{-1}(-1)$, the certificate complexity of the (unique) input in $\mathrm{OR}_R^{-1}(+1)$ is n. That is, letting $T = \{x\colon \phi(x) > 0\}$ and $E^+ = T \cap f^{-1}(-1)$, for Equation (1) to hold even approximately for $g = \mathrm{OR}_R$, it is essential that

$$\left(\sum_{E^+} |\phi(x)|\right) / \left(\sum_{x \in T} |\phi(x)|\right) \ll 1/R. \tag{4}$$

Accordingly, Bun et al. [BKT18] obtain their lower bound for k-distinctness by using a dual witness ϕ for $h = \mathrm{THR}_N^k$ that satisfies Equation (4). Using a dual with such few false positive errors causes [BKT18] to lose an additive $1/(2k)$ term in the exponent of N in their final degree bound, relative to what they would obtain if Equation (2) were sufficient to ensure that Equation (1) approximately held.

As previously mentioned, Sherstov [She12] introduced a variant of dual block composition intended to handle Source 2 errors that might have otherwise rendered Equation (1) false. Specifically, Sherstov proposed multiplying $(\Psi \star \phi)(x)$ by a low-degree polynomial $p_\eta(x)$ intended to "kill" any inputs x that may contribute Source 2 errors (here, η is a parameter, and we will explain shortly how the value of η is ultimately chosen). Specifically, p_η "counts" the number of blocks x_i of x such that $\mathrm{sgn}(\phi(x_i)) \neq h(x_i)$, and p_η is defined (through polynomial interpolation) to evaluate to 0 if this number is any integer between 1 and η. This has the effect of eliminating all Source 2 errors made by $\Psi \star \phi$ on inputs x for which at most η copies of ϕ make an error. That is, p_η kills all inputs x in the set

$$U_\eta := \{x = (x_1, \ldots, x_R)\colon \mathrm{sgn}(\phi(x_i)) \neq h(x_i) \text{ for between 1 and } \eta \text{ values of } i\}.$$

Note that multiplying $\Psi \star \phi$ by p_η has the additional, unfortunate effect of distorting the values that $\Psi \star \phi$ takes on other inputs; bounding the effect of this distortion is one challenge that Sherstov's analysis (as well as our own analysis in this work) has to address.

The intuition is that, so long as most Source 2 errors made by $\Psi \star \phi$ are caused by inputs in the set U_η, then multiplying $\Psi \star \phi$ by p_η should eliminate the otherwise devastating effects of most Source 2 errors. So the remaining challenge is to choose a dual witness ϕ for h guaranteeing that indeed most Source 2 errors are caused by inputs in U_η. More precisely, ϕ must be chosen to ensure that, with respect to the product distribution $\prod_{i=1}^R |\phi(x_i)|/\|\phi\|_1$, it is very unlikely that more than η copies of ϕ make an error on their input x_i.

To this end, it is implicit in Sherstov's analysis that Equation (1) approximately holds with $(\Psi \star \phi) \cdot p_\eta$ in place of $\Psi \star \phi$ so long as

$$\left(\sum_{x \in E^- \cup E^+} |\phi(x)|\right) / \|\phi\|_1 \ll \eta/R. \tag{5}$$

Notice that this is exactly Equation (4), except that the right hand side has crucially increased by a factor of η (also, Equation (5) counts both false-positive and false-negative errors, as opposed to just false-positive errors, which is a key discrepancy that we address below). The bigger that η is set, the less stringent is the requirement of Equation (5). However, it turns out that, in order to ensure that $(\Psi \star \phi) \cdot p_\eta$ has pure high degree close to that of $\Psi \star \phi$ itself, η must be set to a value that is noticeably smaller than the pure high degree of Ψ. Ultimately, to obtain the strongest possible results, η gets set to some constant $C < 1$ times the pure high degree of Ψ.

In order to bring Sherstov's ideas to bear on k-distinctness, we have to modify his construction as follows. The key issue (alluded to above) is that Sherstov's construction is not targeted at functions $g \circ h$ where g has low (-1)-certificate complexity, and it is essential that we exploit this low certificate complexity in the correlation analysis to improve on the k-distinctness lower bound from [BKT18]. Essentially, we modify Sherstov's definition of p_η to "ignore" all false negative errors (which as explained above are benign in our setting because $g = \mathrm{OR}_R$ has low (-1)-certificate complexity). Rather we have p_η only "count" the false positive errors and kill any inputs where this number is between 1 and η.

We are able to show that with this modification, it is sufficient to choose a dual witness ϕ for THR_N^k satisfying

$$\left(\sum_{E^+} |\phi(x)|\right) / \left(\sum_{x \in T} |\phi(x)|\right) \ll \eta / R. \tag{6}$$

We end up setting $\eta \approx O(\sqrt{R})$ for our lower bound, hence the denominator on the right hand side of this inequality represents a quadratic improvement compared to that on the right hand side of Equation (4). This improvement ultimately enables us to improve the lower bound from $\tilde{\Omega}(N^{3/4-1/(2k)})$ to $\tilde{\Omega}(N^{3/4-1/(4k)})$.

The actual calculations required to establish the sufficiency of Equation (6) are quite involved, and we provide a more detailed proof overview in Section 4 to help the reader make sense of them.

2.2 The Upper Bound

Recall from Section 2.1 that the approximate degree of k-distinctness is (essentially) equivalent to $\widetilde{\deg}(f_{RN}^{\leq N})$ for $f = \mathrm{OR}_R \circ \mathrm{THR}_N^k$. Similarly, the approximate degree of the Surjectivity function is (essentially) equivalent to $\widetilde{\deg}(f_{RN}^{\leq N})$ for $f = \mathrm{AND}_R \circ \mathrm{OR}_N$. Sherstov proved an upper bound of $O(R^{1/4} \cdot N^{1/2})$ for this latter quantity.

Up to polylogarithmic factors, in Theorem 1.3 we achieve an identical upper bound for k-distinctness, for any $k \leq \mathrm{polylog}(N)$. To do so, we make the following easy observations. First, in order to apply Sherstov's construction to a function $f = g \circ h$, it is enough that g have approximate degree $O(\sqrt{R})$,[13] and that h be exactly computed as a linear combination of conjunctions, where the coefficients in the linear combination have ℓ_1-norm at most quasipolynomially large in N. Second, we observe that for $k \leq \mathrm{polylog}(N)$, THR_N^k is exactly computed by such a linear combination of conjunctions. Together, these observations are enough to apply Sherstov's construction for Surjectivity to obtain the approximate degree upper bound of Theorem 1.3 for k-distinctness.

3 Preliminaries

Notation. Let N, n and m be positive integers, $N \leq n$. For $z \in \{-1,1\}^n$, let $|z|$ represent the *Hamming weight* of z, i.e., the number of -1's in z. Define $(\{-1,1\}^n)^{\leq N} := \{x \in \{-1,1\}^n : |x| \leq N\}$. For any function $f : \{-1,1\}^n \to \mathbb{R}$, denote by $f^{\leq N}$ the partial function that is defined on $(\{-1,1\}^n)^{\leq N}$ and agrees with f on all such inputs. Define $\mathrm{sgn} : \mathbb{R} \to \{-1,1\}$ by $\mathrm{sgn}(x) = 1$ for all non-negative x, and -1 otherwise. All logarithms in this paper are base 2 unless otherwise specified. Let 1^n (respectively, $(-1)^n$) denote the n-bit string $(1,1,\ldots,1)$ (respectively,

[13] More precisely, it should be possible to approximate g by a linear combination of monotone conjunctions, where the ℓ_1-norm of the coefficients of the linear combination is $2^{\tilde{O}(\sqrt{R})}$. It is not hard to show, by Parseval's identity, that this is guaranteed if g has approximate degree $\tilde{O}(\sqrt{R})$.

$(-1,-1,\ldots,-1)$). For strings $a \in \{-1,1\}^m$ and $b \in \{-1,1\}^n$, we denote by a,b the $(m+n)$-bit string formed by the concatenation of a and b. We use the notation $[n]$ to denote the set $\{1,2,\ldots,n\}$.

For any function $f : \{-1,1\}^n \to \mathbb{R}$, define $\|f\|_1 = \sum_{x\in\{-1,1\}^n} |f(x)|$. For an event E, the corresponding indicator function is

$$I[E] = \begin{cases} 1 & \text{if } E \text{ holds,} \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

For any function $\psi : \{-1,1\}^m \to \mathbb{R}$ such that $\|\psi\|_1 = 1$, let μ_ψ be the distribution on $\{-1,1\}^m$, defined by

$$\mu_\psi(x) = |\psi(x)|. \tag{8}$$

Definition 3.1. *For any integer $n > 0$, any function $\psi : \{-1,1\}^m \to \mathbb{R}$ such that $\|\psi\|_1 = 1$, and any $w \in \{-1,1\}$, let μ_w be the probability distribution μ_ψ conditioned on the event that $\operatorname{sgn}(\psi(x)) = w$. For any $z \in \{-1,1\}^n$, let μ_z denote the probability distribution $(\mu_\psi)^{\otimes n}$ conditioned on the event that $\operatorname{sgn}(\psi(x_i)) = z_i$ for all $i \in [n]$.*

We omit the dependence of μ_z on ψ since ψ will typically be clear from context. Note that μ_z as defined above is a product distribution given by

$$\mu_z(x_1,\ldots,x_n) = \prod_{i=1}^{n} \mu_{z_i}(x_i). \tag{9}$$

Definition 3.2. *For $\eta_i \in [0,1]$, let $\Pi(\eta_1,\ldots,\eta_n)$ be the product distribution on $\{-1,1\}^n$ where the ith bit of the string equals -1 with probability η_i, and 1 with probability $1-\eta_i$.*

Lemma 3.3. *Let n be any positive integer, $p : \{-1,1\}^n \to \mathbb{R}$ be a multilinear polynomial, and $\eta_1,\ldots,\eta_n \in [0,1]$. For $x = (x_1,\ldots,x_n)$ drawn from the product distribution $\Pi(\eta_1,\ldots,\eta_n)$ defined in Definition 3.2, we have*

$$\mathbb{E}_{\Pi(\eta_1,\ldots,\eta_n)}[p(x_1,\ldots,x_n)] = p(1-2\eta_1,\ldots,1-2\eta_n). \tag{10}$$

Any function $f : \{-1,1\}^n \to \mathbb{R}$ has a unique multilinear representation $f = \sum_{S\subseteq[n]} \hat{f}(S)\chi_S$, where for any $S \subseteq [n]$, the function $\chi_S : \{-1,1\}^n \to \{-1,1\}$ is defined by $\chi_S(x) = \prod_{i\in S} x_i$. Hence, $\|\hat{f}\|_1 = \sum_{S\subseteq[n]} |\hat{f}(S)|$. It follows that for any function $\phi : \{-1,1\}^n \to \mathbb{R}$, there exists a unique multilinear polynomial $\tilde{\phi} : \mathbb{R}^n \to \mathbb{R}$ such that $\tilde{\phi}(x) = \phi(x)$ for all $x \in \{-1,1\}^n$.

3.1 Functions of Interest

Define the function $\mathrm{OR}_N : \{-1,1\}^N \to \{-1,1\}$ to equal 1 if $x = 1^N$, and -1 otherwise. Define the *Threshold* function $\mathrm{THR}_N^k : \{-1,1\}^N \to \{-1,1\}$ to equal 1 for inputs of Hamming weight less than k, and -1 otherwise.

Definition 3.4 (k-distinctness)**.** *For integers k, N, R with $k \le N$, define the function $\mathrm{DIST}_{N,R}^k : [R]^N \to \{-1,1\}$ by $\mathrm{DIST}_{N,R}^k(s_1,\ldots,s_N) = -1$ iff there exists an $r \in [R]$ and distinct indices $i_1,\ldots,i_k$ such that $s_{i_1} = \cdots = s_{i_k} = r$. When necessary, the domain of the function can be viewed as $\{-1,1\}^{N\log R}$.*

Given any functions $f_n : \{-1,1\}^n \to \{-1,1\}$ and $g_m : \{-1,1\}^m \to \{-1,1\}$, we define the function $f_n \circ g_m : \{-1,1\}^{mn} \to \{-1,1\}$ as $f_n \circ g_m(x_{11},\ldots,x_{1m},x_{21},\ldots,x_{2m},\ldots,x_{n1},\ldots,x_{nm}) = f_n(g_m(x_1),g_m(x_2),\ldots,g_m(x_n)), x_i \in \{-1,1\}^m$ for all $i \in [n]$. We drop subscripts when the arities of the constituent functions are clear.

3.2 Notions of Approximation

Definition 3.5 (Approximate degree)**.** *For any function $f : \{-1,1\}^n \to \mathbb{R}$, any integer $N \leq n$, and any $\epsilon \in [0,1]$, define the ϵ-approximate degree of $f^{\leq N}$ to be*

$$\widetilde{\deg}_\epsilon(f^{\leq N}) = \min_{\substack{p:|p(x)-f(x)|\leq\epsilon \\ \forall x\in\{-1,1\}^n,|x|\leq N}} \deg(p).$$

When the subscript is dropped, ϵ is assumed to equal $1/3$. When the superscript is dropped in $f^{\leq N}$, then N is assumed to equal n.[14]

Definition 3.6. *For any finite subset $X \subseteq \mathbb{R}^n$, any function $f : X \to \mathbb{R}$, and any integer $d \geq 0$, define*

$$E(f,d) := \min_{p:\deg(p)\leq d} \left\{ \max_{x\in X} |f(x) - p(x)| \right\}.$$

Definition 3.7 (Correlation)**.** *Consider any function $f : \{-1,1\}^n \to \mathbb{R}$ and $\psi : \{-1,1\}^n \to \mathbb{R}$. Define the* correlation *between f and ψ to be*

$$\langle f, \psi \rangle = \sum_{x\in\{-1,1\}^n} f(x)\psi(x).$$

Definition 3.8 (Pure high degree)**.** *For $\phi : \{-1,1\}^n \to \mathbb{R}$, we say that the* pure high degree *of ϕ, which we denote by* $\mathrm{phd}(\phi)$*, is d if $d \geq 0$ is the largest integer for which $\langle \phi, p \rangle = 0$ for any polynomial $p : \{-1,1\}^n \to \mathbb{R}$ of degree strictly less than d.*

For any Boolean function $f : \{-1,1\}^m \to \{-1,1\}$ and function $\psi : \{-1,1\}^m \to \mathbb{R}, \|\psi\|_1 = 1$, let

$$\epsilon^+_{f,\psi} := \Pr_{\mu_\psi}[f(x)\psi(x) < 0 | \psi(x) > 0], \qquad \epsilon^-_{f,\psi} := \Pr_{\mu_\psi}[f(x)\psi(x) < 0 | \psi(x) < 0]. \tag{11}$$

Define $\epsilon_{f,\psi} = \epsilon^+_{f,\psi} + \epsilon^-_{f,\psi}$.

Definition 3.9. *For any functions $f : \{-1,1\}^n \to \{-1,1\}$ and $\psi : \{-1,1\}^n \to \mathbb{R}$, let*

$$E^+(f,\psi) := \{x \in \{-1,1\}^n : f(x)\psi(x) < 0, \psi(x) > 0\},$$
$$E^-(f,\psi) := \{x \in \{-1,1\}^n : f(x)\psi(x) < 0, \psi(x) < 0\}.$$

We define the false positive error *between f and ψ to be*

$$\delta^+_{f,\psi} := \sum_{x\in E^+(f,\psi)} |\psi(x)|$$

and false negative error *to be*

$$\delta^-_{f,\psi} := \sum_{x\in E^-(f,\psi)} |\psi(x)|.$$

We observe the following simple connection between $\delta^+_{f,\psi}$ ($\delta^-_{f,\psi}$) and $\epsilon^+_{f,\psi}$ ($\epsilon^-_{f,\psi}$).

Claim 3.10. *For any Boolean function $f : \{-1,1\}^m \to \{-1,1\}$ and any function $\psi : \{-1,1\}^m \to \mathbb{R}$ with $\|\psi\|_1 = 1, \mathrm{phd}(\psi) \geq 1$,*

$$\epsilon^+_{f,\psi} = 2\delta^+_{f,\psi}, \quad \epsilon^-_{f,\psi} = 2\delta^-_{f,\psi}. \tag{12}$$

[14] Note that this definition places no constraints on an approximating polynomial on inputs outside the promise domain. In other contexts, an approximating polynomial may be required to be bounded outside the promise domain.

Proof.

$$\begin{aligned}
\delta^+_{f,\psi} &= \sum_{x \in E^+(f,\psi)} |\psi(x)| && \text{by Definition 3.9} \\
&= \Pr_{x \sim \mu_\psi}[x \in E^+(f,\psi)] && \text{by Equation (8)} \\
&= \Pr_{x \sim \mu_\psi}[f(x)\psi(x) < 0 \wedge \psi(x) > 0] && \text{by Definition 3.9} \\
&= \Pr_{x \sim \mu_\psi}[\psi(x) > 0] \cdot \Pr_{x \sim \mu_\psi}[f(x)\psi(x) < 0 | \psi(x) > 0] \\
&= \frac{\epsilon^+_{f,\psi}}{2}. && \text{since } \langle \psi, 1 \rangle = 0 \text{ and } \textstyle\sum_x |\psi(x)| = 1 \text{ implies } \Pr_{\mu_\psi}[\psi(x) > 0] = 1/2
\end{aligned}$$

The equality $\epsilon^-_{f,\psi} = 2\delta^-_{f,\psi}$ can be proved similarly. □

By linear programming duality, we have the following standard equivalence between lower bounds on approximate degree and existence of "dual polynomials". See, for example, [BKT17].

Lemma 3.11. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any function. For any integer $0 \le j \le n$, we have $\widetilde{\deg}_\epsilon(f^{\le j}) \ge d$ if and only if there exists a "dual polynomial" $\phi : \{-1,1\}^n \to \mathbb{R}$ satisfying the following properties.*

- $\sum_{x \in \{-1,1\}^n} |\phi(x)| = 1$.
- $\mathrm{phd}(\phi) > d$.
- $\langle f, \phi \rangle > \epsilon$.
- $\phi(x) = 0$ *for all* $|x| > j$.

We say that ϕ is a dual polynomial witnessing the fact that $\widetilde{\deg}_\epsilon(f^{\le j}) > d$. For brevity, when ϵ and d are clear from context, we say that ϕ is a dual polynomial for $f^{\le j}$.

Špalek [Š08] exhibited an explicit dual witness for OR (existence of a dual witness for OR was already implicit from the work of Nisan and Szegedy [NS94]).

Claim 3.12 (Implicit in [NS94])**.** *There exists a constant $c \in (0,1]$ such that for any integer $n \ge 0$, there exists a function $\theta : \{-1,1\}^n \to \mathbb{R}$ satisfying*

- $\|\theta\|_1 = 1$,
- $\mathrm{phd}(\theta) \ge c\sqrt{n}$,
- $\langle \theta, \mathrm{OR}_n \rangle \ge 3/5$.

We also require the following error reduction theorem for approximate degree.

Lemma 3.13 ([BNRdW07])**.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any (possibly partial) Boolean function and let $0 < \epsilon < 1$. Then,*

$$\widetilde{\deg}_\epsilon(f) = \widetilde{\deg}(f) \cdot O(\log(1/\epsilon)).$$[15]

[15] The statement in [BNRdW07] only deals with total functions. It can be seen that the proof works for partial functions too.

3.3 Dual Polynomials and Dual Block Composition

Bun et al. [BKT18] exhibited a dual witness for the approximate degree of the k-threshold function. Their dual witness additionally satisfies a decay condition, meaning that it places very little mass on inputs of large Hamming weight. The following claim, which gives a preliminary construction towards their dual witness for THR_N^k, is a mild modification of [BKT17, Proposition 54].

Claim 3.14 (Modification of [BKT17, Proposition 54])**.** *Let $k, T, N \in \mathbb{N}$ with $2 \le k \le T$. There exist constants $c_1, c_2 \in (0,1]$ and a function $\omega_T : [T] \cup \{0\} \to \mathbb{R}$ such that all of the following hold.*

$$\sum_{\omega_T(t)>0, t\ge k} |\omega_T(t)| \le \frac{1}{48 \cdot 4^k \sqrt{N} \log N}. \tag{13}$$

$$\sum_{\omega_T(t)<0, t<k} |\omega_T(t)| \le \left(\frac{1}{2} - \frac{2}{4^k}\right). \tag{14}$$

$$\|\omega_T\|_1 := \sum_{t=0}^{T} |\omega_T(t)| = 1. \tag{15}$$

For all polynomials $q : \mathbb{R} \to \mathbb{R}$,

$$\deg(q) < c_1 \sqrt{4^{-k} k^{-1} T N^{-1/(2k)} \log^{-1} N} \implies \sum_{t=0}^{T} \omega_T(t) q(t) = 0. \tag{16}$$

$$\text{For all } t \in [T], |\omega_T(t)| \le \frac{\sigma \exp(-\beta t)}{t^2} \qquad \text{for } \sigma = (2k)^k, \quad \beta = c_2 / \sqrt{4^k k T N^{1/(2k)} \log N}. \tag{17}$$

Although the proof follows along the same lines as that of [BKT17], we provide a proof in the appendix for completeness.

The next claim yields a dual polynomial for THR_N^k, and we omit its proof.

Claim 3.15 (Modification of [BKT17, Proposition 55])**.** *Let $k, T, N \in \mathbb{N}$ with $2 \le k \le T \le N$, and let ω_T be as constructed in Claim 3.14, with constants c_1, c_2. Define $\psi_T : \{-1,1\}^N \to \mathbb{R}$ by $\psi_T(x) = \omega_T(|x|)/\binom{N}{|x|}$ for $x \in (\{-1,1\}^N)^{\le T}$ and $\psi_T(x) = 0$ otherwise. Then*

$$\delta^+_{\mathrm{THR}_N^k, \psi_T} \le \frac{1}{48 \cdot 4^k \sqrt{N} \log N} \tag{18}$$

$$\delta^-_{\mathrm{THR}_N^k, \psi_T} \le \frac{1}{2} - \frac{2}{4^k} \tag{19}$$

$$\|\psi_T\|_1 = 1 \tag{20}$$

For any polynomial p: $\{-1,1\}^N \to \mathbb{R}$,

$$\deg(p) < c_1 \sqrt{4^{-k} k^{-1} T N^{-1/(2k)} \log^{-1} N} \implies \langle \psi_T, p \rangle = 0 \tag{21}$$

$$\text{For all } t \in [n], \qquad \sum_{|x|=t} |\psi_T(x)| \le \frac{(2k)^k \exp\left(-c_2 t / \sqrt{4^k k T N^{1/(2k)} \log N}\right)}{t^2}. \tag{22}$$

Towards proving approximate degree lower bounds for composed functions, one might hope to combine dual polynomials of the constituent functions in some way to obtain a dual polynomial for the composed function. A series of works [SZ09, Lee09, She13] introduced the notion of "dual block composition", which is a powerful method of combining dual witnesses.

Definition 3.16 (Dual block composition)**.** *Let $\theta : \{-1,1\}^n \to \mathbb{R}, \phi : \{-1,1\}^m \to \mathbb{R}$ be any functions satisfying $\|\theta\|_1 = \|\phi\|_1 = 1$ and $\mathrm{phd}(\phi) \geq 1$. Let $x = (x_1, \ldots, x_n)$ where each $x_i \in \{-1,1\}^m$. Define the* dual block composition *of θ and ϕ, denoted $\theta \star \phi$, to be*

$$\theta \star \phi(x) = 2^n \theta(\mathrm{sgn}(\phi(x_1)), \ldots, \mathrm{sgn}(\phi(x_n))) \prod_{i=1}^{n} |\phi(x_i)|.$$

Sherstov [She13] showed that dual block composition preserves ℓ_1-norm and that pure high degree is multiplicative (also see [Lee09]). Bun and Thaler [BT17] observed that dual block composition is associative.

Lemma 3.17. *Let $\phi : \{-1,1\}^{m_\phi} \to \mathbb{R}, \theta : \{-1,1\}^{m_\theta} \to \mathbb{R}$ be any functions. Then,*
***Preservation of ℓ_1-norm**: If $\|\theta\|_1 = 1, \|\phi\|_1 = 1$ and $\langle \phi, 1 \rangle = 0$, then*

$$\|\theta \star \phi\|_1 = 1. \tag{23}$$

***Multiplicativity of pure high degree**:*

$$\mathrm{phd}(\theta) > D, \mathrm{phd}(\phi) > d \implies \mathrm{phd}(\theta \star \phi) > Dd. \tag{24}$$

***Associativity**: For every $\psi : \{-1,1\}^{m_\psi} \to \mathbb{R}$, we have*

$$(\phi \star \theta) \star \psi = \phi \star (\theta \star \psi). \tag{25}$$

It was shown in [BKT17] that for any dual polynomial Φ, and ψ_T as constructed in Claim 3.15, the dual block composed function $\Phi \star \psi_T$ satisfies a "strong dual decay" condition.[16]

Claim 3.18 ([BKT17, Proposition 31])**.** *Let R be sufficiently large and $k \leq T \leq R$ be any positive integer. Fix $\sigma = (2k)^k$ and let $N = \lceil 20\sqrt{\sigma} R \rceil$. Let $\Phi : \{-1,1\}^R \to \mathbb{R}$ be any function with $\|\Phi\|_1 = 1$ and $\psi_T : \{-1,1\}^N \to \mathbb{R}$ as defined in Claim 3.15. Then*

$$\sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |(\Phi \star \psi_T)(x)| \leq (2NR)^{-2\Delta} \tag{26}$$

for some $\Delta \geq \frac{\beta\sqrt{\sigma}R}{4 \ln^2 R}$ for $\beta = c_2 / \sqrt{4^k k T N^{1/(2k)} \log N}$.

We now define a simple but important function ϕ that we use in our construction of a dual witness for $\mathrm{DIST}^k_{N,R}$. This function was first used in the context of dual block composition by Bun and Thaler [BT15].

Claim 3.19 ([BT15])**.** *Define $\phi : \{-1,1\}^n \to \mathbb{R}$ as*

$$\phi(x) = \begin{cases} -1/2 & \text{if } x = -1^n \\ 1/2 & \text{if } x = 1^n \\ 0 & \text{otherwise.} \end{cases} \tag{27}$$

Then, $\mathrm{phd}(\phi) = 1$.

[16] They in fact showed that $\Psi \star \psi$ satisfies this strong decay condition for *any* ψ satisfying a corresponding "weak decay" condition. However for this paper, we only require this statement for $\psi = \psi_T$ as constructed in Claim 3.15.

Bun et al. [BKT17], slightly extending a result in [BT15], showed that on dual block composing ϕ and ψ, where ϕ is defined as in Claim 3.19, the correlation of the dual block composed witness $\phi \star \psi$ with $\mathrm{OR}_M \circ f$ amplifies the correlation of f with ψ as follows.

Lemma 3.20 ([BKT17, Proposition 56]). *Let $f : \{-1,1\}^n \to \{-1,1\}$ and $\psi : \{-1,1\}^n \to \mathbb{R}$ be any functions with $\|\psi\|_1 = 1$. For every $M \in \mathbb{N}$ and $\phi : \{-1,1\}^M \to \mathbb{R}$ as defined in Claim 3.19, we have*

$$\delta^+_{\mathrm{OR}_M \circ f, \phi\star\psi} \leq M\delta^+_{f,\psi}, \tag{28}$$

$$\delta^-_{\mathrm{OR}_M \circ f, \phi\star\psi} \leq \frac{1}{2}(2\delta^-_{f,\psi})^M. \tag{29}$$

3.4 Some Polynomials

In this section we list out a few polynomials that we require, along with their properties.

Lemma 3.21 ([She12, Lemma 3.1]). *For any $\tau_1, \ldots, \tau_n \in [0,1)$, define $\nu = \Pi(\tau_1, \ldots, \tau_n)$ and $\tau = \max\{\tau_1, \ldots, \tau_n\}$. For any $\eta = 0, 1, \ldots, n-1$, let $p_\eta : [-1,1]^n \to \mathbb{R}$ be the unique degree-η multilinear polynomial that satisfies*

$$p_\eta(z) = (-1)^\eta \prod_{i=1}^{\eta}(|z| - i), \forall z \in \{-1,1\}^n\,. \tag{30}$$

Then,

$$p_\eta(1^n) = \eta!, \tag{31}$$

$$\|\hat{p}_\eta\|_1 \leq \eta!\binom{n+\eta}{\eta}, \tag{32}$$

$$\mathbb{E}_\nu[|p_\eta(z)|] \leq p_\eta(1^n)\nu(1^n)\,(1+A)\,, \quad \text{where } A := \binom{n}{\eta+1}\frac{\tau^{\eta+1}}{(1-\tau)^n}. \tag{33}$$

Furthermore, $p_\eta(z) \geq 0$ for all $z \in \{-1,1\}^n$ provided that η is even.

It is easy to show that for any multilinear polynomial $p : \mathbb{R}^n \to \mathbb{R}$, we have $\max_{y\in[-1,1]^n} |p(y)| \leq \|\hat{p}\|_1$. When applied to the function in the previous lemma, we obtain

Claim 3.22. *For p_η defined as in Lemma 3.21, $\max_{y\in[-1,1]^n} |p_\eta(y)| \leq \eta!\binom{n+\eta}{\eta}$.*

Finally, we require a lemma, implicit in a result of Razborov and Sherstov [RS10] (also see [BT17, Proposition 21] for a formulation similar to the one we require), that helps us convert a dual polynomial with little mass on large Hamming weight inputs to a dual polynomial with no mass on large Hamming weight inputs without affecting the pure high degree by much.

Lemma 3.23 (Implicit in [RS10]). *Let $N \geq R$ be positive integers, $\Delta \in \mathbb{R}^+$, and $\theta : \{-1,1\}^{RN} \to \mathbb{R}$ be any polynomial such that*

$$\sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |\theta(x)| \leq (2NR)^{-\Delta}.$$

For any positive integer $D < \Delta$, there exists a function $\nu : \{-1,1\}^{RN} \to \mathbb{R}$ such that

- $\mathrm{phd}(\nu) > D$

- $\|\nu\|_1 \leq 1/10$
- $|x| > N \Rightarrow \nu(x) = \theta(x)$.

Definition 3.24. *For any integer $d \geq 0$, let $T_d : \mathbb{R} \to \mathbb{R}$ denote the degree-d Chebyshev polynomial, defined recursively as follows.*

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \\ T_d(x) &= 2xT_{d-1}(x) - T_{d-2}(x). \end{aligned}$$

We now observe a simple well-known fact about Chebyshev polynomials whose proof we include for completeness.

Claim 3.25. *For any $d \geq 0$, consider the d'th Chebyshev polynomial $T_d : \mathbb{R} \to \mathbb{R}$ as defined in Definition 3.24, and write its expansion $T_d(x) = \sum_{i=0}^{d} a_i x^i$. Then,*

$$\sum_{i=0}^{d} |a_i| \leq 3^d. \tag{34}$$

Proof. We prove this by induction.

Let S_d denote $\sum_{i=0}^{d} |a_i|$ where a_i's are the coefficients in the expansion $T_d(x) = \sum_{i=0}^{d} a_i x^i$. By Definition 3.24, the hypothesis is satisfied for $d = 0, 1$. Next suppose the hypothesis is true for all $d \leq k$ for some $k \geq 1$. By the recursive definition in Definition 3.24, we have $S_{k+1} \leq 2S_k + S_{k-1} \leq 2 \cdot 3^k + 3^{k-1} = 3^{k-1}(6+1) < 3^{k+1}$. □

We also require the following well-known properties of Chebyshev polynomials.

Fact 3.26. *For any integer $d \geq 0$,*

$$|T_d(x)| \leq 1 \qquad |x| \leq 1 \tag{35}$$

$$T_d(1+\epsilon) \geq 1 + d^2\epsilon \qquad \epsilon \geq 0. \tag{36}$$

Definition 3.27. *For any positive integer n, any polynomial $p : \{-1,1\}^n \to \{0,1\}$ that is of the form*

$$\left(\prod_{i \in A} \frac{1+x_i}{2}\right)\left(\prod_{j \in B} \frac{1-x_j}{2}\right) \tag{37}$$

for some sets $A, B \subseteq [n]$, is called a conjunction.

It can be observed that the product of conjunctions is a conjunction.

Claim 3.28 ([She18a, Corollary 4.7])**.** *Let $n \leq N$ be any positive integers, and A, B be any subsets of $[N]$. Define $f : \left(\{-1,1\}^N\right)^{\leq n} \to \{0,1\}$[17] by*

$$f(x) = \left(\prod_{i \in A} \frac{1+x_i}{2}\right)\left(\prod_{j \in B} \frac{1-x_j}{2}\right).$$

[17]The version in [She18a] deals with functions whose domain is $\left(\{0,1\}^N\right)^{\leq n}$. The statement there can easily be seen to imply the statement in this paper.

Then, for any integer $d \geq 0$, we have

$$E(f,d) \leq \exp\left(-\frac{cd^2}{n}\right)$$

for some absolute constant c.

Definition 3.29. *Consider any positive integer n and any function $f : \{-1,1\}^n \to \mathbb{R}$. Define the* conjunction norm *of f, which we denote by $\rho(f)$, to be*

$$\min\left\{\sum_{A\subseteq[n]}\sum_{B\subseteq[n]}|C_{A,B}| : f(x) = \sum_{A\subseteq[n]}\sum_{B\subseteq[n]} C_{A,B}\left(\prod_{i\in A}\frac{1+x_i}{2}\right)\left(\prod_{j\in B}\frac{1-x_j}{2}\right),\quad C_{A,B}\in\mathbb{R}\right\}.$$

We now state some simple observations about the conjunction norm which we do not prove here. See, for example, [She18a, Proposition 2.4].

Fact 3.30. *Let m, n be positive integers, $f, g : \{-1,1\}^n \to \mathbb{R}$ be any functions, and $p : \mathbb{R} \to \mathbb{R}$ be any degree-m polynomial of the form $p(x) = \sum_{i=0}^m a_i x^i, a_i \in \mathbb{R}$. Then ρ is well defined and satisfies*

$$\rho(a \cdot f) = |a|\rho(f), \quad \text{for any } a \in \mathbb{R}, \tag{38}$$

$$\rho(f+g) \leq \rho(f) + \rho(g), \tag{39}$$

$$\rho(f \cdot g) \leq \rho(f) \cdot \rho(g), \tag{40}$$

$$\rho(p \circ g) \leq (\max\{1, \rho(g)\})^m \cdot \sum_{i=0}^m |a_i|. \tag{41}$$

4 Outline of Proof of Main Theorem

Our main theorem is as follows.

Theorem 4.1. *For $R \in \mathbb{N}$ sufficiently large, $2 \leq k \leq \frac{\log R}{4}$, and some $N = \Theta(k^{k/2}R)$,*

$$\widetilde{\deg}(\mathrm{DIST}^k_{N,R+N}) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\log^{7/2} R} \cdot R^{\frac{3}{4}-\frac{1}{4k}}\right). \tag{42}$$

Ambainis [Amb05] showed that the approximate degree[18] of functions that are symmetric (both with respect to range elements and with respect to domain elements) is the same for all range sizes greater than or equal to N. As a corollary, we obtain the following.

Corollary 4.2. *For $R \in \mathbb{N}$ sufficiently large, $2 \leq k \leq \frac{\log R}{4}$, and some $N = \Theta(k^{k/2}R)$,*

$$\widetilde{\deg}(\mathrm{DIST}^k_{N,N}) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\log^{7/2} R} \cdot R^{\frac{3}{4}-\frac{1}{4k}}\right). \tag{43}$$

We require the following relation between approximate degree of k-distinctness and a related Boolean function; this relationship follows from [BKT17, Proposition 21 and Corollary 26].

[18] There are several different conventions used in the literature when defining the domain of functions such as k-distinctness. The convention used by Ambainis [Amb05] considers the input to be specified by $N \cdot R$ variables $y_{1,1}, \ldots, y_{N,R}$, where $y_{i,j} = -1$ if and only if the ith list item in the input equals range element j (i.e., it is promised that for each i, $y_{i,j} = -1$ for exactly one j). We use the convention that the input is specified by $N\lceil \log_2 R \rceil$ bits. It is well known (and not hard to show) that conversion between the two conventions affects approximate degree by at most a factor of $\lceil \log_2 R \rceil$.

Claim 4.3 ([BKT17]). *Let $N, R \in \mathbb{N}$ and $2 \leq k \leq N$ be any integer. Then for any $\epsilon > 0$,*

$$\widetilde{\deg}_{\epsilon}(\mathrm{DIST}^{k}_{N,R+N}) = \Omega\left(\frac{1}{\log R} \cdot \widetilde{\deg}_{\epsilon}(\mathrm{OR}_R \circ \mathrm{THR}^{k}_{N})^{\leq N}\right). \tag{44}$$

To prove Theorem 4.1, Claim 4.3 implies that it suffices to prove a lower bound on $\widetilde{\deg}(\mathrm{OR}_R \circ \mathrm{THR}^{k}_{N})^{\leq N}$.

Theorem 4.4. *For $R \in \mathbb{N}$ sufficiently large, $2 \leq k \leq \frac{\log R}{4}$, and some $N = \Theta(k^{k/2}R)$,*

$$\widetilde{\deg}((\mathrm{OR}_R \circ \mathrm{THR}^{k}_{N})^{\leq N}) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\log^{5/2} R} \cdot R^{\frac{3}{4}-\frac{1}{4k}}\right). \tag{45}$$

Note that the theorems above continue to yield non-trivial lower bounds for some values of $k = \omega(1)$. However for ease of exposition, we assume throughout this section that $k \geq 2$ is an arbitrary but fixed constant.

Outline of the Proof of Theorem 4.4. Towards proving Theorem 4.4, we construct a dual witness Γ satisfying the following four conditions.

- **Normalization:** $\|\Gamma\|_1 = 1$,
- **Pure high degree:** There exists a $D = \tilde{\Omega}\left(R^{\frac{3}{4}-\frac{1}{4k}}\right)$ such that for every polynomial $p : \{-1,1\}^{RN} \to \mathbb{R}$ of degree less than D, we have $\langle p, \Gamma \rangle = 0$,
- **Correlation:** $\langle \Gamma, (\mathrm{OR}_R \circ \mathrm{THR}^{k}_{N}) \rangle > 1/3$,
- **Exponentially little mass on inputs of large Hamming weight:** $\sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |\Gamma(x)| \leq (2NR)^{-\tilde{\Omega}\left(R^{\frac{3}{4}-\frac{1}{4k}}\right)}$ for all $x \notin (\{-1,1\}^{RN})^{\leq N}$.

Next, Lemma 3.23 implies existence of a function ν that equals Γ on $x \notin (\{-1,1\}^{RN})^{\leq N}$, has pure high degree $\tilde{\Omega}\left(R^{\frac{3}{4}-\frac{1}{4k}}\right)$, and $\|\nu\|_1 \leq 1/10$. The function $\mathcal{W} : \{-1,1\}^{RN} \to \mathbb{R}$ defined by $\mathcal{W}(x) := \frac{\Gamma(x)-\nu(x)}{\|\Gamma-\nu\|_1}$ then satisfies the conditions in Equations (75), (76), (77) and (78) (see Section 6.2 for proofs). Theorem 4.4 then follows by Lemma 3.11 and Lemma 3.13.

Organization of the rest of this section and the proof of Theorem 4.4. The rest of this section is devoted towards providing a sketch of how we construct such a dual witness Γ. In the next subsection we first sketch an outline of the approximate degree lower bound in [BKT18], and in the subsequent subsection we elaborate on where our approach differs from theirs. Section 5 presents auxiliary lemmas that will be used in the formal proof of Theorem 4.4, while Section 6 contains the proof itself.

4.1 Prior Work

At a high level, we follow the same outline as followed in [BKT18], who exhibited a dual witness Λ witnessing $\widetilde{\deg}(\mathrm{DIST}^{k}_{N,R}) = \tilde{\Omega}\left(R^{\frac{3}{4}-\frac{1}{2k}}\right)$ for the same ranges of k, N, R that we consider. In this section we sketch their construction. Their dual witness takes the form $\Lambda = \theta \star \phi \star \psi$, where θ, ϕ, ψ each have ℓ_1-norm 1 and additionally satisfy the properties below.

- The function ψ satisfies:

  - The false positive error between THR_N^k and ψ is $O(1/N)$.
  - The false negative error between THR_N^k and ψ is at most $\frac{1}{2} - \frac{2}{4^k}$.
  - The pure high degree of ψ is $\tilde{\Omega}(\sqrt{R}N^{-1/(2k)})$.
  - ψ satisfies a "weak decay condition", viz. $\sum_{|x|=t} |\psi(x)| \leq \sigma \exp(-\beta t)/t^2$ for some constant σ (for general k, the value of σ only depends on k), and $\beta = \tilde{\Omega}(\sqrt{R}N^{1/(2k)})$.

- The function ϕ is defined on 4^k inputs, and is defined as in Claim 3.19.

- θ is constructed as in Claim 3.12 with $n = R/4^k$.

The facts that $\|\Lambda\|_1 = 1$ and $\mathrm{phd}(\Lambda) = \tilde{\Omega}(R^{3/4}N^{-1/(2k)})$ follow immediately from the definitions of θ, ϕ, ψ, and the fact that dual block composition preserves ℓ_1-norm and causes pure high degree to increase multiplicatively (Lemma 3.17).

Next they use the fact that dual block composition is associative (Equation (25)) to express Λ as $(\theta \star \phi) \star \psi$ and conclude using Claim 3.18 that Λ places exponentially small (in $R^{\frac{3}{4}-\frac{1}{2k}}$) mass on inputs in $\{-1,1\}^{RN}$ of Hamming weight larger than N.

It remains to show the correlation bound, i.e., $\langle \Lambda, \mathrm{OR}_R \circ \mathrm{THR}_N^k \rangle > 1/3$. For the correlation analysis it is convenient to view Λ as $\theta \star (\phi \star \psi)$. The following is the outline of their correlation analysis.

1. By construction, $\delta^+_{\mathrm{THR}_N^k, \psi} = O(1/N)$ and $\delta^-_{\mathrm{THR}_N^k, \psi} \leq \frac{1}{2} - \frac{2}{4^k}$.

2. By Lemma 3.20, the false positive error between $\mathrm{OR}_{4^k} \circ \mathrm{THR}_N^k$ and $\phi \star \psi$ remains $O(1/N)$, whereas the the false negative error between $\mathrm{OR}_{4^k} \circ \mathrm{THR}_N^k$ and $\phi \star \psi$ becomes a small enough constant.

3. As mentioned in Section 2.1, the very low (-1)-certificate complexity of OR_R renders false-negative errors benign. Thus the false-negative and false-positive error rates achieved in the last bullet point are sufficient to ensure $\langle \theta \star (\phi \star \psi), \mathrm{OR}_{R/4^k} \circ (\mathrm{OR}_{4^k} \circ \mathrm{THR}_N^k) \rangle \geq 1/3$ by showing $\langle \theta \star (\phi \star \psi), \mathrm{OR}_{R/4^k} \circ (\mathrm{OR}_{4^k} \circ \mathrm{THR}_N^k) \rangle \approx \langle \theta, \mathrm{OR}_{R/4^k} \rangle$.

Roughly, where we improve over this prior work is in item 3 above. Whereas [BKT18] needed a false-positive error rate for $\phi \star \psi$ of $O(1/N)$ to ensure that their final dual witness Λ is well-correlated with $\mathrm{OR}_R \circ \mathrm{THR}_N^k$, we modify the construction of Λ so that a false-positive error rate of roughly $1/\sqrt{N}$ suffices to ensure good correlation of the final dual witness with $\mathrm{OR}_R \circ \mathrm{THR}_N^k$.

4.2 Our Construction

As in the previous section, our construction of Γ is also based on three dual witnesses. The functions θ, ϕ are exactly the same as in the previous section. Our ψ is a fairly straightforward modification of the one described in the previous section, that has a larger pure high degree, at the cost of a worse false positive error. A little more formally, our functions θ, ϕ, ψ have ℓ_1-norm equal to 1, and additionally satisfy the following.

- The function ψ satisfies:
  - The false positive error between THR_N^k and ψ is $\tilde{O}(1/\sqrt{N})$.
  - The false negative error between THR_N^k and ψ is at most $\frac{1}{2} - \frac{2}{4^k}$.
  - The pure high degree of ψ is $\tilde{\Omega}(\sqrt{R}N^{-1/(4k)})$.

– ψ satisfies a "weak decay condition", viz. $\sum_{|x|=t} |\psi(x)| \leq \sigma \exp(-\beta t)/t^2$ for some constant σ (for general k, the value of σ only depends on k), and $\beta = \tilde{\Omega}(\sqrt{R}N^{1/(4k)})$.

- The function ϕ is defined on 4^k inputs, and is defined as in Claim 3.19.
- θ is constructed as in Claim 3.12 with $n = R/4^k$.

If we were to define $\Gamma = \theta \star \phi \star \psi$, all the analyses from the previous section would work, except for the correlation analysis, which fails. To fix this, our main technical contribution is to not use dual block composition, but rather a variant of it inspired by a result of Sherstov [She12]. Our function Γ takes the form $\Gamma = \theta \bullet (\phi \star \psi)$, where $\bullet$ denotes our variant of dual block composition. In a little more detail,

$$\Gamma(x_1, \ldots, x_{R/4^k}) :=$$
$$\theta \bullet (\phi \star \psi)(x) = \frac{1}{p_\eta(1-2\epsilon^+, \ldots, 1-2\epsilon^+)} \cdot (\theta \star (\phi \star \psi))(x_1, \ldots, x_{R/4^k}) \cdot p_\eta(\alpha(x_1), \ldots, \alpha(x_{R/4^k})),$$

for

$$\epsilon^+ = \epsilon^+_{\phi\star\psi, \mathrm{OR}_{4^k} \circ \mathrm{THR}^k_N},$$
$$\epsilon^- = \epsilon^-_{\phi\star\psi, \mathrm{OR}_{4^k} \circ \mathrm{THR}^k_N},$$

η is a parameter that we set later, and p_η and α are functions whose definitions we elaborate on later in this section.

We first give a very high-level idea of how we prove the required properties of Γ, and then elaborate on the definitions of η, p_η and α.

- **Normalization:** Following along similar lines as [She12, Claim 6.2], we prove that $\|\Gamma\|_1 = 1$ by modifying the proof that dual block composition preserves ℓ_1-norm, crucially exploiting properties of p_η and α (see Claim 5.5).
- **Pure high degree:** Using our definition of p_η, and α, one can show (Claim 5.6) that the pure high degree of $\theta \bullet (\phi \star \psi)$ is at least $(\mathrm{phd}(\theta) - \eta)\mathrm{phd}(\phi \star \psi)$. The value of η is chosen to be $\mathrm{phd}(\theta)/2$ so that this quantity is the same order of magnitude as $\mathrm{phd}(\theta)\mathrm{phd}(\phi \star \psi) = \mathrm{phd}(\theta)\mathrm{phd}(\psi)$, which is $\tilde{\Omega}(R^{3/4}N^{-1/(4k)})$.
- **Exponentially little mass on inputs of large Hamming weight:** By a similar argument as sketched in the last section, it can be shown that the mass placed by $(\theta \star \phi) \star \psi$ on inputs of Hamming weight larger than N is exponentially small in $\tilde{\Omega}(R^{\frac{3}{4}-\frac{1}{4k}})$. Since $\theta \bullet (\phi \star \psi) := \frac{1}{p_\eta(1-2\epsilon^+, \ldots, 1-2\epsilon^+)} \cdot (\theta \star (\phi \star \psi))(x_1, \ldots, x_{R/4^k}) \cdot p_\eta(\alpha(x_1), \ldots, \alpha(x_{R/4^k}))$, it suffices to show that the maximum absolute value of $\frac{p_\eta(\alpha(x_1), \ldots, \alpha(x_{R/4^k}))}{p_\eta(1-2\epsilon^+, \ldots, 1-2\epsilon^+)}$ is at most exponentially large in $R^{\frac{3}{4}-\frac{1}{4k}}$, which we do in Claim 6.6.
- **Correlation:** Conceptually, the function $p_\eta : \{-1, 1\}^{R/4^k} \to \mathbb{R}$ can be viewed as one that "corrects" $\theta \star (\phi \star \psi)$: it "counts" the number of false positives fed to it by $\phi \star \psi$, and changes the output of $\theta \star (\phi \star \psi)$ to 0 on inputs where this number is any integer between 1 and η. The function $\alpha : \{-1, 1\}^N \to \mathbb{R}$ acts as the function that, in a sense, indicates whether or not $\phi \star \psi$ is making a *false positive* error.

– **Detecting errors:** The function α takes three possible output values: it outputs -1 for $x \in E^+(\mathrm{OR}_{4^k} \circ \mathrm{THR}_N^k, \phi \star \psi)$ and outputs either 1 or a value very close to 1 for $x \notin E^+(\mathrm{OR}_{4^k} \circ \mathrm{THR}_N^k, \phi \star \psi)$. This definition of α is our biggest departure from Sherstov's construction in [She12]; Sherstov defined α to output -1 for *both* false-positive and false-negative errors, whereas our α only outputs -1 for false-positive errors.

– **Zeroing out errors:** Define the function p_η to be (the unique multilinear extension of) the function that outputs 0 if its input has Hamming weight between 1 and η. Recall that our construction considers the dual witness

$$\frac{1}{p_\eta(1-2\epsilon^+, \ldots, 1-2\epsilon^+)} \cdot (\theta \star (\phi \star \psi))(x_1, \ldots, x_{R/4^k}) \cdot p_\eta(\alpha(x_1), \ldots, \alpha(x_{R/4^k})),$$

and the purpose of multiplying $\theta \star (\phi \star \psi)$ by p_η is for p_η to zero out most inputs in which one or more false-positive errors are being fed by $\phi \star \psi$ into θ (see Equation (3.16)).

Unfortunately, p_η is nonzero on inputs of Hamming weight more than η. Hence, in terms of the correlation analysis, a key question that must be addressed is: what fraction of the ℓ_1-mass of $\theta \star (\phi \star \psi)$ is placed on inputs where more than η copies of $\phi \star \psi$ make a false-positive error? We need this fraction to be very small, because multiplying by p_η fails to zero out such inputs.

Note that under the distribution defined by $|\phi \star \psi|$, the *expected* number of false positive errors fed into θ is $(R/4^k) \cdot \epsilon^+$. Since we have set $\eta = O(\sqrt{R/(4 \cdot 4^k)})$, it suffices to have $\epsilon^+ \ll 1/(c\eta)$ for some large enough constant c to conclude that with high probability (over the distribution $|\phi \star \psi|$), the number of false positive errors fed into θ is at most a small constant times η. It turns out that this value of ϵ^+ is indeed attained by $\phi \star \psi$, since the false positive error between THR_N^k and ψ was set to be $\tilde{O}(1/\sqrt{N}) = \tilde{O}(1/\sqrt{R})$ to begin with. Thus, with high probability, multiplying $\theta \star (\phi \star \psi)$ by p_η successfully zeros out all but an exponentially small fraction of the errors made by $\theta \star (\phi \star \psi)$ that can be attributed to false-positive errors made by $\phi \star \psi$. This intuitive proof outline is formalized in Claim 6.4, which in turn is a formalization of Equation (1) that holds with the setting of parameters mentioned above.

5 Properties of Auxiliary Functions

Given any function $f : \{-1,1\}^m \to \{-1,1\}$ and $\psi : \{-1,1\}^m \to \mathbb{R}$, $\|\psi\|_1 = 1$, let $\epsilon^+ = \epsilon^+_{f,\psi}$ and $\epsilon^- = \epsilon^-_{f,\psi}$ as defined in Equation (11). Define the function $\alpha_{f,\psi} : \{-1,1\}^m \to \mathbb{R}$ as

$$\alpha_{f,\psi}(x) := \begin{cases} 1 =: a^+ & \text{if } \psi(x)f(x) > 0, \psi(x) > 0 \\ \frac{1-2\epsilon^+-\epsilon^-}{1-\epsilon^-} =: a^- & \text{if } \psi(x)f(x) > 0, \psi(x) < 0 \\ -1 & \text{if } \psi(x)f(x) < 0, \psi(x) > 0 \\ 1 & \text{if } \psi(x)f(x) < 0, \psi(x) < 0. \end{cases} \tag{46}$$

For the remaining sections, for $z_i \in \{-1,1\}$, $a^{z_i} = a^+$ if $z_i = 1$, and $a^{z_i} = a^-$ if $z_i = -1$.

Claim 5.1. *For any integer $m > 0$, any functions $f : \{-1,1\}^m \to \{-1,1\}$ and $\psi : \{-1,1\}^m \to \mathbb{R}$ such that $\|\psi\|_1 = 1$, let $\alpha = \alpha_{f,\psi} : \{-1,1\}^m \to \mathbb{R}$ be as defined in Equation (46). Then for any integer $n > 0$, any z in $\{-1,1\}^n$, and all $i \in [n]$,*

$$\mathbb{E}_{(x_1,\ldots,x_n)\sim\mu_z}[\alpha(x_i)] = 1 - 2\epsilon^+_{f,\psi}. \tag{47}$$

Proof. Let $\epsilon^+ = \epsilon^+_{f,\psi}$ and $\epsilon^- = \epsilon^-_{f,\psi}$.

$$\begin{aligned}
\mathbb{E}_{\mu_z}[\alpha(x_i)] &= \mathbb{E}_{\mu_{z_i}}[\alpha(x_i)] && \text{by Equation (9)}\\
&= \begin{cases} \epsilon^+ \cdot -1 + (1-\epsilon^+) & \text{if } z_i = 1 \\ \epsilon^- \cdot 1 + (1-\epsilon^-)\frac{1-2\epsilon^+-\epsilon^-}{1-\epsilon^-} & \text{if } z_i = -1 \end{cases} && \text{by Definition 3.1 and Equation (46)}\\
&= 1 - 2\epsilon^+.
\end{aligned}$$

□

Consider any positive integer m, functions $f : \{-1,1\}^m \to \{-1,1\}$ and $\psi : \{-1,1\}^m \to \mathbb{R}$, and any integers $\eta < n$. By Claim 5.1, Equation (9) and the fact that $p_\eta : [-1,1]^n \to \mathbb{R}$ as defined in Lemma 3.21 is multilinear, it holds for all $z \in \{-1,1\}^n$ that

$$\mathbb{E}_{\mu_z}[p_\eta(\alpha(x_1), \ldots, \alpha(x_n))] = p_\eta(1 - 2\epsilon^+_{f,\psi}, \ldots, 1 - 2\epsilon^+_{f,\psi}). \tag{48}$$

Let

$$b^+ := 0, \quad b^- := \frac{\epsilon^+_{f,\psi}}{1 - \epsilon^-_{f,\psi}}, \tag{49}$$

and a^+, a^- be as defined in Equation (46). For the remaining sections, for $z_i \in \{-1,1\}$, $b^{z_i} := b^+$ if $z_i = 1$ and $b^{z_i} := b^-$ if $z_i = -1$. Then, by multilinearity of p_η and Definition 3.2, for any $i \in [n]$ and any $c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_n \in [-1,1]$ we have

$$\mathbb{E}_{w \sim \Pi(b^{z_i})}[p_\eta(c_1, \ldots, c_{i-1}, w, c_{i+1}, \ldots, c_n)] = p_\eta(c_1, \ldots, c_{i-1}, a^{z_i}, c_{i+1}, \ldots, c_n), \tag{50}$$

since $1 - 2b^+ = 1 = a^+$ and $1 - 2b^- = \frac{1-\epsilon^-_{f,\psi}-2\epsilon^+_{f,\psi}}{1-\epsilon^-_{f,\psi}} = a^-$. We also obtain that

$$\mathbb{E}_{(w_1,\ldots,w_n) \sim \Pi(b^{z_1},\ldots,b^{z_n})}[p_\eta(w_1, \ldots, w_n)] = p_\eta(a^{z_1}, \ldots, a^{z_n}), \tag{51}$$

by Lemma 3.3. We now state the setting for our next few claims.

Assumptions for Claim 5.2, Claim 5.3, Claim 5.4, Claim 5.5: Let m, n be any positive integers, $\eta < n$ be any even positive integer, and $f : \{-1,1\}^m \to \{-1,1\}$ be any function. Let $\zeta : \{-1,1\}^n \to \mathbb{R}$ be such that $\langle \zeta, \mathrm{OR}_n \rangle > \delta$ and $\|\zeta\|_1 = 1$, and $\xi : \{-1,1\}^m \to \mathbb{R}$ be any function such that $\|\xi\|_1 = 1$ and $\mathrm{phd}(\xi) \geq 1$. Let $p_\eta : \{-1,1\}^n \to \mathbb{R}$ be as defined in Lemma 3.21, let $\alpha = \alpha_{f,\xi} : \{-1,1\}^m \to \mathbb{R}$ be as defined in Equation (46), and define the distribution μ_ξ over $\{-1,1\}^{nm}$ as in Equation (8). Let $\epsilon^+ = \epsilon^+_{f,\xi}$, $\epsilon^- = \epsilon^-_{f,\xi}$, $\epsilon = \epsilon^+ + \epsilon^-$, and $A = \binom{n}{\eta+1}\frac{(\epsilon^+)^{\eta+1}}{(1-\epsilon^+)^n}$.

Claim 5.2.

$$\begin{aligned}
&\zeta(1^n)\mathbb{E}_{x \sim \mu_{1^n}}[p_\eta(\alpha(x_1), \ldots, \alpha(x_n))\mathrm{OR}(f(x_1), \ldots, f(x_n))]\\
\geq &p_\eta(1 - 2\epsilon^+, \ldots, 1 - 2\epsilon^+)\left(\zeta(1^n) - |\zeta(1^n)|2A\right).
\end{aligned} \tag{52}$$

Claim 5.3.

$$\begin{aligned}
&\sum_{z \neq 1^n} \zeta(z)\mathbb{E}_{\mu_z}[p_\eta(\alpha(x_1), \ldots, \alpha(x_n))\mathrm{OR}(f(x_1), \ldots, f(x_n))]\\
\geq &p_\eta(1 - 2\epsilon^+, \ldots, 1 - 2\epsilon^+)\left(\sum_{z \neq 1^n} \zeta(z)\mathrm{OR}(z) - \left(2 - 2\left(\frac{1-\epsilon^+-\epsilon^-}{1-\epsilon^+}\right)(1-A)\right)\sum_{z \neq 1^n}|\zeta(z)|\right).
\end{aligned} \tag{53}$$

Claim 5.4. *If $A < 1$, then,*

$$\langle \mathrm{OR} \circ f, (\zeta \star \xi)(p_\eta \circ \alpha) \rangle \geq p_\eta(1-2\epsilon^+, \ldots, 1-2\epsilon^+) \cdot \left(\delta - \left(2 - 2\left(\frac{1-\epsilon^+-\epsilon^-}{1-\epsilon^+} \right)(1-A) \right) \right). \tag{54}$$

We first prove Claim 5.4 using Claim 5.2 and Claim 5.3, and prove those claims later.

Proof.

$$\langle \mathrm{OR} \circ f, (\zeta \star \xi)(p_\eta \circ \alpha) \rangle = \sum_{x \in \{-1,1\}^{mn}} (\mathrm{OR} \circ f)(x)(\zeta \star \xi)(p_\eta \circ \alpha)(x)$$

$$= \sum_{x \in \{-1,1\}^{mn}} \mathrm{OR}(f(x_1), \ldots, f(x_n)) 2^n \zeta\left(\mathrm{sgn}(\xi(x_1)), \ldots, \mathrm{sgn}(\xi(x_n))\right) p_\eta(\alpha(x_1), \ldots, \alpha(x_n)) \prod_{i=1}^n |\xi(x_i)|$$

by Definition 3.16

$$= \sum_{z \in \{-1,1\}^n} \zeta(z) \left(\sum_{x : \mathrm{sgn}(\xi(x_i)) = z_i \forall i \in [n]} 2^n p_\eta(\alpha(x_1), \ldots, \alpha(x_n)) \mathrm{OR}(f(x_1), \ldots, f(x_n)) \prod_{i=1}^n |\xi(x_i)| \right)$$

$$= \sum_{z \in \{-1,1\}^n} \zeta(z) \mathbb{E}_{\mu_z}[p_\eta(\alpha(x_1), \ldots, \alpha(x_n)) \mathrm{OR}(f(x_1), \ldots, f(x_n))]$$

by Definition 3.1 and $\Pr_{x_i \sim \mu_\xi}[\mathrm{sgn}(x_i) = 1] = \Pr_{x_i \sim \mu_\xi}[\mathrm{sgn}(x_i) = -1] = 1/2$ since $\mathrm{phd}(\xi) \geq 1$

$$\geq p_\eta(1-2\epsilon^+, \ldots, 1-2\epsilon^+) \left(\zeta(1^n)\mathrm{OR}(1^n) - 2|\zeta(1^n)|A + \sum_{z \neq 1^n} \zeta(z)\mathrm{OR}(z) \right.$$

$$\left. - \left(2 - 2\left(\frac{1-\epsilon^+-\epsilon^-}{1-\epsilon^+} \right)(1-A) \right) \sum_{z \neq 1^n} |\zeta(z)| \right) \qquad \text{by Claims 5.2, 5.3 and } \mathrm{OR}(1^n) = 1$$

$$= p_\eta(1-2\epsilon^+, \ldots, 1-2\epsilon^+) \left(\sum_{z \in \{-1,1\}^n} \zeta(z)\mathrm{OR}(z) - 2|\zeta(1^n)|A - \left(2 - 2\left(\frac{1-\epsilon^+-\epsilon^-}{1-\epsilon^+} \right)(1-A) \right) \sum_{z \neq 1^n} |\zeta(z)| \right)$$

$$\geq p_\eta(1-2\epsilon^+, \ldots, 1-2\epsilon^+) \left(\delta - \max\left\{ 2A, 2 - 2\left(\frac{1-\epsilon^+-\epsilon^-}{1-\epsilon^+} \right)(1-A) \right\} \right)$$

since $|\zeta(1^n)| + \sum_{z \neq 1^n} |\zeta(z)| = 1$ and $\langle \zeta, \mathrm{OR} \rangle > \delta$

$$\geq p_\eta(1-2\epsilon^+, \ldots, 1-2\epsilon^+) \left(\delta - \left(2 - 2\left(\frac{1-\epsilon^+-\epsilon^-}{1-\epsilon^+} \right)(1-A) \right) \right),$$

where the last inequality holds because $\left(2 - 2\left(\frac{1-\epsilon^+-\epsilon^-}{1-\epsilon^+} \right)(1-A) \right) - 2A = (1-A)\left(2 - 2\left(\frac{1-\epsilon^+-\epsilon^-}{1-\epsilon^+} \right) \right) > 0$ since $\left(\frac{1-\epsilon^+-\epsilon^-}{1-\epsilon^+} \right) < 1$, and $A < 1$. □

Next we prove Claim 5.2.

Proof of Claim 5.2. Recall that μ_{1^n} is the distribution μ_ξ conditioned on the event that $\mathrm{sgn}(\xi(x_i)) = 1$ for all $i \in [n]$. Note that for all $x_1, \ldots, x_n$ in the support of μ_{1^n} such that $I[(f(x_1), \ldots, f(x_n)) = 1^n]$ (which means $f(x_i) = 1$ for all $i \in [n]$), we have by the definition of a^{z_i} in Equation (46), that

$\alpha(x_i) = a^+$ for all $i \in [n]$. Hence,

$$\begin{aligned}
&\mathbb{E}_{\mu_1{}^n}[p_\eta(\alpha(x_1),\ldots,\alpha(x_n))I[(f(x_1),\ldots,f(x_n)) = 1^n]] \\
=&\Pr_{x\sim\mu_1{}^n}[(f(x_1),\ldots,f(x_n)) = 1^n]p_\eta(a^+,\ldots,a^+) \\
=&\left(\prod_{i=1}^n \Pr_{\mu_1}[f(x_i)=1]\right)p_\eta(a^+,\ldots,a^+) && \text{by Equation (9)} \\
=&\left(\prod_{i=1}^n (1-\epsilon^+)\right)\mathbb{E}_{w\sim\Pi(b^+,\ldots,b^+)}[p_\eta(w)] && \text{by Equation (11) and Equation (51)} \\
\geq&(1-\epsilon^+)^n \Pr_{w\sim\Pi(b^+,\ldots,b^+)}[w = 1^n]p_\eta(1^n) && \text{since } p_\eta(w)\geq 0 \text{ for all } w\in\{-1,1\}^n \text{ by Lemma 3.21} \\
=&(1-\epsilon^+)^n p_\eta(1^n), && (55)
\end{aligned}$$

where the last line follows by Definition 3.2 and Equation (49). Next,

$$\begin{aligned}
&|\mathbb{E}_{\mu_1{}^n}[p_\eta(\alpha(x_1),\ldots,\alpha(x_n))\mathrm{OR}(f(x_1),\ldots,f(x_n))] - p_\eta(1-2\epsilon^+,\ldots,1-2\epsilon^+)| \\
=&|\mathbb{E}_{\mu_1{}^n}[p_\eta(\alpha(x_1),\ldots,\alpha(x_n))\left(\mathrm{OR}(f(x_1),\ldots,f(x_n)) - 1\right)]| && \text{by Equation (48)} \\
=&2\mathbb{E}_{\mu_1{}^n}[p_\eta(\alpha(x_1),\ldots,\alpha(x_n))(1 - I[(f(x_1),\ldots,f(x_n)) = 1^n])] \\
\leq&2\mathbb{E}_{\mu_1{}^n}[p_\eta(\alpha(x_1),\ldots,\alpha(x_n))] - 2(1-\epsilon^+)^n p_\eta(1^n) && (56)
\end{aligned}$$

by Equation (55). Hence, by Equation (56),

$$\begin{aligned}
&\zeta(1^n)\mathbb{E}_{\mu_1{}^n}[p_\eta(\alpha(x_1),\ldots,\alpha(x_n))\mathrm{OR}(f(x_1),\ldots,f(x_n))] \\
\geq&\zeta(1^n)p_\eta(1-2\epsilon^+,\ldots,1-2\epsilon^+) - |\zeta(1^n)|\left(2\mathbb{E}_{\mu_1{}^n}[p_\eta(\alpha(x_1),\ldots,\alpha(x_n))] - 2(1-\epsilon^+)^n p_\eta(1^n)\right) \\
=&p_\eta(1-2\epsilon^+,\ldots,1-2\epsilon^+)\left(\zeta(1^n) - |\zeta(1^n)|\left(2 - \frac{2(1-\epsilon^+)^n p_\eta(1^n)}{p_\eta(1-2\epsilon^+,\ldots,1-2\epsilon^+)}\right)\right) && \text{by Equation (48)} \\
\geq&p_\eta(1-2\epsilon^+,\ldots,1-2\epsilon^+)\left(\zeta(1^n) - |\zeta(1^n)|2A\right), && (57)
\end{aligned}$$

where the last inequality follows as we have by Equation (33) and the fact that p_η is non-negative on all Boolean inputs (Lemma 3.21) that

$$\mathbb{E}_{\Pi(\epsilon^+,\ldots,\epsilon^+)}[p_\eta(z)] \leq p_\eta(1^n)(1-\epsilon^+)^n(1+A), \tag{58}$$

which by Lemma 3.3 implies that

$$\frac{2p_\eta(1^n)(1-\epsilon^+)^n}{p_\eta(1-2\epsilon^+,\ldots,1-2\epsilon^+)} \geq 2(1+A)^{-1} \geq 2(1-A), \tag{59}$$

for all $A \geq 0$. □

We now prove Claim 5.3.

Proof of Claim 5.3. We first introduce some notation that we use in this proof. For any $i \in [n]$ and $r \in [-1,1]$, let $y^i(r)$ denote the n-bit string $(1-2\epsilon^+,\ldots,1-2\epsilon^+,r,1-2\epsilon^+,\ldots,1-2\epsilon^+)$, where r is in the i'th position. It is easy to verify from its definition that p_η is symmetric on $\{-1,1\}^n$. Hence, for any $i \in [n]$,

$$\begin{aligned}
&p_\eta(y^i(1))(1-\epsilon^+) = \mathbb{E}_{w\sim\Pi(\epsilon^+,\ldots,\epsilon^+)}[p_\eta(w,1)](1-\epsilon^+) && \text{by Lemma 3.3} \\
\geq&\Pr_{\Pi(\epsilon^+,\ldots,\epsilon^+)}\left[w = 1^{n-1}\right]p_\eta(1^n)(1-\epsilon^+) && \text{since } p_\eta(w,1)\geq 0 \text{ for all } w\in\{-1,1\}^{n-1} \text{ by Lemma 3.21} \\
=&p_\eta(1^n)(1-\epsilon^+)^n, && (60)
\end{aligned}$$

where the last equality follows from Definition 3.2. For any $z \neq 1^n$, let $i \in [n]$ be an index such that $z_i = -1$.[19] Fix any $z \neq 1^n$. Note that by Equation (46) and Equation (9), $\alpha(x_i) = a^-$ for all x_i in the support of μ_{z_i} satisfying $f(x_i) = -1$. Hence,

$$\begin{aligned}
&\mathbb{E}_{x\sim\mu_z}[p_\eta(\alpha(x_1),\dots,\alpha(x_n))I[f(x_i)=-1]] = \Pr_{\mu_z}[f(x_i)=-1]p_\eta(y^i(a^-)) && \text{by Claim 5.1}\\
=&\Pr_{\mu_z}[f(x_i)=-1]\mathbb{E}_{w\sim\Pi(b^-)}[p_\eta(y^i(w))] && \text{by Equation (50)}\\
=&(1-\epsilon^-)\mathbb{E}_{w\sim\Pi(b^-)}[p_\eta(y^i(w))] && \text{by Definition 3.1 the definition of } \epsilon^- \text{ in Equation (11)}\\
\geq&(1-\epsilon^-)\left(1-\frac{\epsilon^+}{1-\epsilon^-}\right)p_\eta(y^i(1)) &&\\
& && \text{by Equation (49) and Definition 3.2 and } p_\eta \text{ is non-negative on } \{-1,1\}^n \text{ (Lemma 3.21)}\\
=&(1-\epsilon^- -\epsilon^+)p_\eta(y^i(1)). && (61)
\end{aligned}$$

Next,

$$\begin{aligned}
&|\mathbb{E}_{\mu_z}\left[p_\eta(\alpha(x_1),\dots,\alpha(x_n))\left(\mathrm{OR}(f(x_1),\dots,f(x_n))-\mathrm{OR}(z)\right)\right]|\\
=&|\mathbb{E}_{\mu_z}\left[p_\eta(\alpha(x_1),\dots,\alpha(x_n))\left(\mathrm{OR}(f(x_1),\dots,f(x_n))+1\right)\right]| && \text{since } \mathrm{OR}(z)=-1, \forall z\neq 1^n\\
\leq&2\mathbb{E}_{\mu_z}\left[p_\eta(\alpha(x_1),\dots,\alpha(x_n))(1-I[f(x_i)=-1])\right] &&\\
& && \text{since } p_\eta \text{ is non-negative on } \{-1,1\}^n \text{ by Lemma 3.21}\\
\leq&2\mathbb{E}_{\mu_z}\left[p_\eta(\alpha(x_1),\dots,\alpha(x_n))\right]-2(1-\epsilon^- -\epsilon^+)p_\eta(y^i(1)), && (62)
\end{aligned}$$

where the last inequality follows by next applying Equation (61). Finally,

$$\begin{aligned}
&\sum_{z\neq 1^n}\mathbb{E}_{\mu_z}[p_\eta(\alpha(x_1),\dots,\alpha(x_n))\mathrm{OR}(f(x_1),\dots,f(x_n))]\zeta(z)\\
\geq&\sum_{z\neq 1^n}\zeta(z)\mathrm{OR}(z)p_\eta(y^i(1-2\epsilon^+))-\sum_{z\neq 1^n}|\zeta(z)|\left(2p_\eta(y^i(1-2\epsilon^+))-2(1-\epsilon^- -\epsilon^+)p_\eta(y^i(1))\right)\\
& && \text{by Equation (62) and Equation (48)}\\
=&p_\eta(1-2\epsilon^+,\dots,1-2\epsilon^+)\left(\sum_{z\neq 1^n}\zeta(z)\mathrm{OR}(z)-\left(2-2\left(\left(\frac{1-\epsilon^+-\epsilon^-}{1-\epsilon^+}\right)\right)\frac{(1-\epsilon^+)p_\eta(y^i(1))}{p_\eta(y^i(1-2\epsilon^+))}\right)\sum_{z\neq 1^n}|\zeta(z)|\right) && (63)\\
\geq&p_\eta(1-2\epsilon^+,\dots,1-2\epsilon^+)\left(\sum_{z\neq 1^n}\zeta(z)\mathrm{OR}(z)-\left(2-2\left(\left(\frac{1-\epsilon^+-\epsilon^-}{1-\epsilon^+}\right)\right)\frac{(1-\epsilon^+)^n p_\eta(1^n)}{p_\eta(y^i(1-2\epsilon^+))}\right)\sum_{z\neq 1^n}|\zeta(z)|\right)\\
& && \text{by Equation (60)}\\
\geq&p_\eta(1-2\epsilon^+,\dots,1-2\epsilon^+)\left(\sum_{z\neq 1^n}\zeta(z)\mathrm{OR}(z)-\left(2-2\left(\left(\frac{1-\epsilon^+-\epsilon^-}{1-\epsilon^+}\right)\right)(1-A)\right)\sum_{z\neq 1^n}|\zeta(z)|\right).\\
& && \text{by Equation (59)}
\end{aligned}$$

□

Finally, we require a closed form expression for $\|(\zeta\star\xi)(p_\eta\circ\alpha)\|_1$.

[19] The notation i_z is more accurate, but we drop the dependence on z to avoid clutter. The underlying z will be clear from context.

Claim 5.5.

$$\|(\zeta \star \xi)(p_\eta \circ \alpha)\|_1 = p_\eta(1-2\epsilon^+,\dots,1-2\epsilon^+). \tag{64}$$

The proof of the claim follows along the lines as that of [She12, Claim 6.2], but we provide the proof for completeness.

Proof. Consider the distribution μ on $\{-1,1\}^{mn}$ defined by $\mu(x_1,\dots,x_n) = \prod_{i=1}^n \mu_\xi(x_i)$. Since $\mathrm{phd}(\xi) \geq 1$, we conclude that the string $(\mathrm{sgn}(\xi(x_1)),\dots,\mathrm{sgn}(\xi(x_n)))$ is uniformly distributed in $\{-1,1\}^n$ when $(x_1,\dots,x_n)$ is sampled from μ. Hence, we have

$$\begin{aligned}
\|(\zeta \star \xi)(p_\eta \circ \alpha)\|_1 &= \sum_{x\in\{-1,1\}^{mn}} 2^n \zeta(\mathrm{sgn}(\xi(x_1)),\dots,\mathrm{sgn}(\xi(x_n)))p_\eta(\alpha(x_1),\dots,\alpha(x_n))\prod_{i=1}^n |\xi(x_i)| \\
&= \sum_{z\in\{-1,1\}^n} |\zeta(z)\mathbb{E}_{\mu_z}[p_\eta(\alpha(x_1),\dots,\alpha(x_n))]| \\
&= \sum_{z\in\{-1,1\}^n} |\zeta(z)|\mathbb{E}_{\mu_z}[p_\eta(\alpha(x_1),\dots,\alpha(x_n))] \\
&\qquad\qquad \text{since } p_\eta \text{ is non-negative on } \{-1,1\}^n \text{ by Lemma 3.21} \\
&= p_\eta(1-2\epsilon^+,\dots,1-2\epsilon^+) \sum_{z\in\{-1,1\}^n} |\zeta(z)| \qquad \text{by Equation (48)} \\
&= p_\eta(1-2\epsilon^+,\dots,1-2\epsilon^+). \qquad \text{since } \|\zeta\|_1 = 1
\end{aligned}$$

□

Claim 5.6. *Let $\Psi : \{-1,1\}^n \to \mathbb{R}$, $\Lambda : \{-1,1\}^m \to \mathbb{R}$, and $f : \{-1,1\}^m \to \mathbb{R}$ be any functions. For any positive integer η, let $\alpha = \alpha_{f,\Lambda} : \{-1,1\}^m \to \mathbb{R}$ be as defined in Equation (46), and $p_\eta : \{-1,1\}^n \to \mathbb{R}$ defined in Lemma 3.21. Then*

$$\mathrm{phd}((\Psi \star \Lambda)\cdot(p_\eta \circ \alpha)) > (\mathrm{phd}(\Psi) - \eta)\cdot \mathrm{phd}(\Lambda). \tag{65}$$

The proof follows along the same lines as that of [She12, Equation (6.7)] and we omit it.

6 Proof of Theorem 4.4

Towards proving Theorem 4.4, it suffices to exhibit a dual polynomial (see Lemma 3.11) that has ℓ_1-norm 1, sufficiently large pure high degree, good correlation with $(\mathrm{OR}_R \circ \mathrm{THR}_N^k)^{\leq N}$, and places no mass outside $(\{-1,1\}^{RN})^{\leq N}$. We first define a function Γ (Definition 6.1) that satisfies the first three properties above, and additionally satisfies a strong decay condition. In Section 6.2 we use Γ to construct a dual polynomial $\mathcal{W}$, via Lemma 3.23, satisfying all the requisite properties. We now set several key variables.

- Let R be sufficiently large and fix $k \leq (\log R)/4$. Set $T = \sqrt{R}$, $\eta = \left(\frac{c}{2}\sqrt{\frac{R}{4^k}}\right) - 1$ where $c \in (0,1]$ is the constant from Claim 3.12 (assume without loss of generality that η is even), $\sigma = (2k)^k$, $c_1, c_2 \in (0,1]$ are constants fixed in the next bullet point, $\beta = \frac{c_2}{\sqrt{4^k k T N^{1/(2k)} \log N}}$, $\Delta = \frac{\beta\sqrt{\sigma}R}{4\ln^2 R} = \frac{c_2 R}{4\ln^2 R}\sqrt{\frac{(2k)^k}{4^k k T N^{1/(2k)}\log N}}$, $N = \lceil 20\sqrt{\sigma}R \rceil$.
- Let $\omega_T : [T] \cup \{0\} \to \mathbb{R}$ be a function that satisfies the conditions in Claim 3.14 and let c_1, c_2 be the constants for which the claim holds. Let $\psi : \{-1,1\}^N \to \mathbb{R}$ be defined by $\psi(x) = \omega_T(|x|)/\binom{N}{|x|}$ if $|x| \leq T$, and 0 otherwise.

- Let $\theta : \{-1,1\}^{R/4^k} \to \mathbb{R}$ be any function satisfying the conditions in Claim 3.12 for $n = R/4^k$ (note that $R/4^k > 0$ since $k < (\log R)/2$).
- Let $\phi : \{-1,1\}^{4^k} \to \mathbb{R}$ be the function defined in Claim 3.19 with $n = 4^k$.
- Let $p_\eta : [-1,1]^{R/4^k} \to \mathbb{R}$ be as defined in Lemma 3.21.
- Let $\alpha := \alpha_{\phi\star\psi,\mathrm{OR}_{4^k}\circ\mathrm{THR}_N^k} : \{-1,1\}^{4^k N} \to \mathbb{R}$ be as defined in Equation (46).
- Let $\epsilon^+ := \epsilon^+_{\phi\star\psi,\mathrm{OR}_{4^k}\circ\mathrm{THR}_N^k}, \epsilon^- := \epsilon^-_{\phi\star\psi,\mathrm{OR}_{4^k}\circ\mathrm{THR}_N^k}$, and $\epsilon := \epsilon^+ + \epsilon^-$.

We next define the function Γ.

Definition 6.1. *Let* $\Gamma : \{-1,1\}^{NR} \to \mathbb{R}$ *be defined by*

$$\Gamma(x_1,\ldots,x_{R/4^k}) := \frac{(\theta \star (\phi \star \psi))(x_1,\ldots,x_{R/4^k}) \cdot p_\eta(\alpha(x_1),\ldots,\alpha(x_{R/4^k}))}{p_\eta(1-2\epsilon^+,\ldots,1-2\epsilon^+)}, \tag{66}$$

where each $x_i \in \{-1,1\}^{4^k N}$.

6.1 Properties of Γ

We now show in Section 6.1.1, Section 6.1.2 and Section 6.1.3 that Γ satisfies the following four properties.

- $\langle \Gamma, \mathrm{OR}_R \circ \mathrm{THR}_N^k \rangle > 1/3$.
- $\|\Gamma\|_1 = 1$.
- $\mathrm{phd}(\Gamma) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\sqrt{\log R}} \cdot R^{\frac{3}{4}-\frac{1}{4k}}\right)$.
- $\sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |\Gamma(x)| \leq (2NR)^{-2(\Delta-\sqrt{R})}$.

6.1.1 Pure High Degree

In this section we show the required lower bound on $\mathrm{phd}(\Gamma)$.

Claim 6.2.

$$\mathrm{phd}(\Gamma) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\sqrt{\log R}} \cdot R^{3/4-1/(4k)}\right). \tag{67}$$

Proof.

$$
\begin{aligned}
\mathrm{phd}(\Gamma) &= \mathrm{phd}((\theta \star (\phi \star \psi))(p_\eta \circ \alpha)) && \text{by Definition 6.1} \\
&\geq (\mathrm{phd}(\theta) - \eta) \cdot \mathrm{phd}(\phi \star \psi) \\
& && \text{by Claim 5.6, using } \Psi = \theta, \Lambda = \phi \star \psi, \text{ and } f = \mathrm{OR}_{4^k} \circ \mathrm{THR}_N^k \\
&\geq \frac{c}{2}\sqrt{\frac{R}{4^k}} \cdot \mathrm{phd}(\phi) \cdot \mathrm{phd}(\psi) && \text{by Claim 3.17, Claim 3.12, and since } \eta = \left(\tfrac{c}{2}\sqrt{\tfrac{R}{4^k}}\right) - 1 \\
&\geq \frac{c}{2}\sqrt{\frac{R}{4^k}} \cdot c_1\sqrt{4^{-k}k^{-1}TN^{-1/(2k)}\log^{-1} N} && \text{by Claim 3.19 and Equation (21)} \\
&= \frac{cc_1}{2}\sqrt{\frac{RT}{4^{2k}kN^{1/(2k)}\log N}} \\
&= \frac{cc_1}{2}\sqrt{\frac{1}{\log N}}\sqrt{\frac{1}{4^{2k}k}}\sqrt{\frac{RT}{N^{1/2k}}} \\
&= \frac{cc_1}{2}\sqrt{\frac{1}{\log 20 + (1/2)\cdot k\log(2k) + \log R}}\sqrt{\frac{1}{4^{2k}k}}\sqrt{\frac{R\sqrt{R}}{(20\sqrt{(2k)^k}R)^{1/2k}}} \\
& && \text{using } T = \sqrt{R} \text{ and } N = 20\sqrt{(2k)^k}R \\
&\geq \frac{cc_1}{2^{9/8}} \cdot \frac{1}{\sqrt{k\log R}} \cdot \frac{1}{4^k \cdot 20^{1/(4k)} \cdot k^{5/8}} \cdot R^{3/4-1/(4k)} \\
& && \text{since } k\log R > \log 20 + 1/2\cdot k\log(2k) + \log R \text{ for sufficiently large } R \\
&= \frac{cc_1}{2^{9/8}} \cdot \frac{1}{4^k \cdot 20^{1/(4k)} \cdot k^{9/8}} \cdot \frac{1}{\sqrt{\log R}} \cdot R^{3/4-1/(4k)} \\
&\geq \frac{cc_1}{60} \cdot \frac{1}{\sqrt{\log R}} \cdot \frac{1}{4^k k^2} \cdot R^{3/4-1/(4k)}. && \text{since } 1/20^{1/(4k)} > 1/20, \text{ for all } k \geq 2
\end{aligned}
$$

□

6.1.2 Correlation

We first show that the function $\phi\star\psi$ has large correlation with $\mathrm{OR}_{4^k}\circ\mathrm{THR}_N^k$, the following analysis is essentially the same as in [BKT17, Proposition 55].

Claim 6.3.

$$\epsilon^+_{\mathrm{OR}_{4^k}\circ\mathrm{THR}_N^k,\phi\star\psi} \leq \frac{1}{24\sqrt{R}\log R}, \tag{68}$$

$$\epsilon^-_{\mathrm{OR}_{4^k}\circ\mathrm{THR}_N^k,\phi\star\psi} \leq e^{-4}. \tag{69}$$

Proof.

$$
\begin{aligned}
\epsilon^+_{\mathrm{OR}_{4^k}\circ\mathrm{THR}_N^k,\phi\star\psi} &= 2\delta^+_{\mathrm{OR}_{4^k}\circ\mathrm{THR}_N^k,\phi\star\psi} && \text{by Claim 3.10} \\
&\leq 2\cdot 4^k \cdot \delta^+_{\mathrm{THR}_N^k,\psi} && \text{by Equation (28), using } M = 4^k \\
&\leq \frac{1}{24\sqrt{N}\log N} && \text{by Equation (18)} \\
&\leq \frac{1}{24\sqrt{R}\log R}. && \text{since } N = \lceil 20\sqrt{\sigma}R\rceil > R
\end{aligned}
$$

Next,

$$\begin{aligned}
\epsilon^{-}_{\mathrm{OR}_{4^k}\circ\mathrm{THR}^k_N,\phi\star\psi} &= 2\delta^{-}_{\mathrm{OR}_{4^k}\circ\mathrm{THR}^k_N,\phi\star\psi} && \text{by Claim 3.10}\\
&\le (2\delta^{-}_{\mathrm{THR}^k_N,\psi})^{4^k} && \text{by Equation (29) using } M = 4^k\\
&\le \left(1-\frac{4}{4^k}\right)^{4^k} && \text{by Equation (19)}\\
&\le e^{-4}. && \text{since } (1-1/n)^n \le 1/e \text{ for all } n \ge 1
\end{aligned}$$

□

Claim 6.4. *The function* Γ *satisfies*

$$\begin{aligned}
\|\Gamma\|_1 &= 1,\\
\langle \Gamma, (\mathrm{OR}_R \circ \mathrm{THR}^k_N)\rangle &> 1/3.
\end{aligned}$$

Proof. The conditions of Claim 5.5 are satisfied with $n = R/4^k, m = 4^k N, f = \mathrm{OR}_{4^k}\circ\mathrm{THR}^k_N, \zeta = \theta, \xi = \phi\star\psi, \eta = \left(\frac{c}{2}\sqrt{\frac{R}{4^k}}\right) - 1$. Hence by Claim 5.5,

$$\|\Gamma\|_1 = \frac{\|(\theta\star(\phi\star\psi))(p_\eta\circ\alpha)\|_1}{p_\eta(1-2\epsilon^+,\ldots,1-2\epsilon^+)} = 1.$$

Define $A = \binom{R/4^k}{\eta+1}\frac{(\epsilon^+)^{\eta+1}}{(1-\epsilon^+)^{R/4^k}}$. If $A < 1$, then the conditions of Claim 5.4 are satisfied with the same parameters mentioned in the beginning on this proof.

We first show that $A < 1$, and then invoke Claim 5.4. To avoid clutter, define $\gamma = \eta+1 = \frac{c}{2}\sqrt{\frac{R}{4^k}}$.

$$\begin{aligned}
A &= \binom{R/4^k}{\gamma}\frac{(\epsilon^+)^\gamma}{(1-\epsilon^+)^{R/4^k}}\\
&\le \left(\frac{Re}{4^k\gamma}\right)^\gamma\left(\frac{1}{24\sqrt{R}\log R}\right)^\gamma\left(1-\frac{1}{24\sqrt{R}\log R}\right)^{-R/4^k} && \text{using Claim 6.3 and } \tbinom{m}{n}\le\left(\tfrac{me}{n}\right)^n\\
&\le \left(\frac{e}{24}\right)^\gamma\left(\frac{\sqrt{R}}{4^k\gamma\log R}\right)^\gamma\cdot 3^{\sqrt{R}/(4^k 24\log R)}\\
& && \text{rearranging terms and using } (1-1/n)^n\le 1/e \text{ for } n>0\\
&= \left(\frac{e}{12}\right)^\gamma\left(\frac{1}{c\sqrt{4^k}\log R}\right)^\gamma\cdot 3^{\sqrt{R}/(4^k 24\log R)} && \text{since } \gamma = \tfrac{c}{2}\sqrt{R/4^k}\\
&= \left(\frac{e\cdot 3^{1/(12c\sqrt{4^k}\log R)}}{12c\sqrt{4^k}\log R}\right)^\gamma && \text{again using } \gamma = \tfrac{c}{2}\sqrt{R/4^k}\\
&\le (e/48)^\gamma && \text{since } 3^{\frac{1}{12c\sqrt{4^k}\log R}}\le 2 \text{ and } 12c\sqrt{4^k}\log R\ge 8 \text{ for sufficiently large } R\\
&\le 1/16. && (70)
\end{aligned}$$

Thus, the conditions in Claim 5.4 are satisfied. By the definition of Γ, we have

$$\begin{aligned}
\langle \Gamma, (\mathrm{OR}_R \circ \mathrm{THR}_N^k) \rangle &= \frac{\langle ((\theta \star (\phi \star \psi)) \cdot (p_\eta \circ \alpha), \mathrm{OR}_{R/4^k} \circ (\mathrm{OR}_{4^k} \circ \mathrm{THR}_N^k) \rangle}{p_\eta(1-2\epsilon^+, \ldots, 1-2\epsilon^+)} \\
&\geq \delta - \left(2 - 2\left(\frac{1-\epsilon^+-\epsilon^-}{1-\epsilon^+}\right)(1-A)\right) && \text{by Claim 5.4} \\
&\geq 3/5 - \left(2\frac{\epsilon^-}{1-\epsilon^+} + 2A\left(\frac{1-\epsilon^+-\epsilon^-}{1-\epsilon^+}\right)\right) && \text{since } \delta \geq 3/5 \text{ by Claim 3.12} \\
&\geq 3/5 - \left(2e^{-4}\frac{1}{1-\frac{1}{24\sqrt{R}\log R}} + 2A\right) && \text{by Claim 6.3 and } \tfrac{1-\epsilon^+-\epsilon^-}{1-\epsilon^+} < 1 \\
&> 3/5 - 1/8 - 1/8 && \text{by Equation (70) and since } R \text{ is sufficiently large} \\
&> 1/3.
\end{aligned}$$

□

6.1.3 Strong Decay

We first state and prove a property of p_η that we require.

Claim 6.5.

$$p_\eta(1-2\epsilon^+, \ldots, 1-2\epsilon^+) > 1. \tag{71}$$

Proof.

$$\begin{aligned}
&p_\eta(1-2\epsilon^+, \ldots, 1-2\epsilon^+) = \mathbb{E}_{w \sim \Pi(\epsilon^+, \ldots, \epsilon^+)}[p_\eta(w)] && \text{by Lemma 3.3} \\
&\geq \Pr_{\Pi(\epsilon^+, \ldots, \epsilon^+)}\left[w = 1^{R/4^k}\right] p_\eta(1^{R/4^k}) && \text{since } p_\eta \text{ is non-negative on } \{-1,1\}^{R/4^k} \text{ by Lemma 3.21} \\
&= (1-\epsilon^+)^{R/4^k}\eta! && \text{by Equation (31)} \\
&\geq \left(1 - \frac{1}{24\sqrt{R}\log R}\right)^{R/4^k} \cdot 2^{\frac{c}{2}\sqrt{\frac{R}{4^k}}-1} && \text{by Equation (68) and using } \eta = \tfrac{c}{2}\sqrt{\tfrac{R}{4^k}} - 1 \\
&> \left(\frac{1}{3}\right)^{\left(\sqrt{R}/(4^k 24 \log R)\right)} \cdot 2^{c\sqrt{R/(4^k \cdot 4)}-1} \\
& && \text{since } R \text{ is sufficiently large and } (1-1/n)^n < 1/e \text{ for } n > 0 \\
&= \frac{2^{\sqrt{R}\left(\frac{c}{2^{k+1}} - \frac{\log 3}{4^k 24 \log R}\right)}}{2} > 2^{\sqrt{R}\left(\frac{c}{2^{k+2}}\right)-1} \geq 1,
\end{aligned} \tag{72}$$

since R is sufficiently large and $k < (\log R)/4$.

□

We next show that Γ satisfies a particular decay property.

Claim 6.6. *The function Γ defined in Definition 6.1 satisfies*

$$\sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |\Gamma(x)| \leq (2NR)^{-2(\Delta - \sqrt{R})}. \tag{73}$$

Proof. First note that by Definition 6.1 and Claim 6.5, it suffices to show the same decay property for $(\theta \star (\phi \star \psi)) \cdot (p_\eta \circ \alpha)(x)$, that is, $\sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |(\theta \star (\phi \star \psi)) \cdot (p_\eta \circ \alpha)(x)| \leq (2NR)^{-2(\Delta - \sqrt{R})}$.

By associativity of dual block composition (Equation (25)), $\theta \star \phi \star \psi = (\theta \star \phi) \star \psi$. Recall that $\psi : \{-1,1\}^N \to \mathbb{R}$ is defined as $\psi(x) = \omega_T(|x|)/\binom{N}{|x|}$ if $|x| \leq T$, and 0 otherwise, for ω_T satisfying the conditions in Claim 3.14. Hence, ψ satisfies the conditions of Claim 3.15 and also those in Claim 3.18. Hence using Claim 3.18 with $\Phi = \theta \star \phi$, we have

$$\sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |((\theta \star \phi) \star \psi)(x)| \leq (2NR)^{-2\Delta}. \tag{74}$$

For any $x \in \{-1,1\}^{RN}$, we write $x = (x_1, \ldots, x_{R/4^k})$, where $x_i \in \{-1,1\}^{4^k N}$, for all i.

$$\begin{aligned}
\sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |(\theta \star (\phi \star \psi)) \cdot (p_\eta \circ \alpha)(x)| &= \sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |((\theta \star \phi) \star \psi)(x)||p_\eta(\alpha(x_1), \ldots, \alpha(x_{R/4^k}))| \\
&\leq \max_{y \in [-1,1]^{R/4^k}} |p_\eta(y)| \sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |((\theta \star \phi) \star \psi)(x)| \\
&\qquad \text{since } \alpha(w) \in [-1,1] \text{ for all } w \in \{-1,1\}^{4^k N} \text{ by Equation (46)} \\
&\leq (2NR)^{-2\Delta} \eta! \binom{R/4^k + \eta}{\eta} \\
&\qquad \text{by Claim 3.22 and Equation (74)} \\
&\leq (2NR)^{-2\Delta} \left(c\sqrt{\frac{R}{4^{k+1}}}\right)! \left(\frac{2eR/4^k}{c\sqrt{R/4^{k+1}}}\right)^{c\sqrt{R/4^{k+1}}} \\
&\qquad \text{since } \eta = c\sqrt{R/4^{k+1}} - 1 < R/4^k, \text{ and } \tbinom{a}{b} \leq (ae/b)^b \\
&\leq (2NR)^{-2\Delta} \sqrt{R}^{\sqrt{R}} \left(\frac{8e}{c}\sqrt{\frac{R}{4^{k+1}}}\right)^{\sqrt{R/4^{k+1}}} \\
&\leq (2NR)^{-2\Delta} (8eR/c)^{\sqrt{R}} \\
&\leq (2NR)^{-2(\Delta - \sqrt{R})}. \\
&\qquad \text{since } R \text{ (and hence } N) \text{ is sufficiently large}
\end{aligned}$$

□

6.2 Final Dual Polynomial

We now prove Theorem 4.4.

Proof of Theorem 4.4. We exhibit a function $\mathcal{W} : \{-1,1\}^{RN} \to \mathbb{R}$ satisfying

$$\mathcal{W}(x) = 0, \forall x \notin (\{-1,1\}^{RN})^{\leq N}, \tag{75}$$

$$\|\mathcal{W}\|_1 = 1 \tag{76}$$

$$\langle \mathcal{W}, (\mathrm{OR}_R \circ \mathrm{THR}_N^k) \rangle > 7/33, \tag{77}$$

$$\mathrm{phd}(\mathcal{W}) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\log^{5/2} R} \cdot R^{\frac{3}{4} - \frac{1}{4k}}\right). \tag{78}$$

The theorem then follows by Lemma 3.11 and Lemma 3.13. Towards the construction of such a $\mathcal{W}$, first note that by Claim 6.6 and Lemma 3.23 there exists a function $\nu : \{-1,1\}^{RN} \to \mathbb{R}$ that satisfies the following properties.

$$|x| > N \Rightarrow \nu(x) = \Gamma(x), \tag{79}$$

$$\mathrm{phd}(\nu) \geq 2(\Delta - \sqrt{R}) - 1, \tag{80}$$

$$\|\nu\|_1 \leq \frac{1}{10}. \tag{81}$$

Define $\mathcal{W} : \{-1,1\}^{RN} \to \mathbb{R}$ by

$$\mathcal{W}(x) := \frac{\Gamma(x) - \nu(x)}{\|\Gamma - \nu\|_1}. \tag{82}$$

For any $x \notin (\{-1,1\}^{RN})^{\leq N}$, we have $\mathcal{W}(x) = \frac{\Gamma(x)-\nu(x)}{\|\Gamma-\nu\|_1} = 0$ by Equation (79). This justifies Equation (75).

Equation (76) immediately follows from Equation (82).

To justify Equation (77), we have

$$\begin{aligned}
\langle \mathcal{W}, \mathrm{OR}_R \circ \mathrm{THR}_N^k \rangle &= \frac{1}{\|\Gamma - \nu\|_1} \left(\langle \Gamma, \mathrm{OR}_R \circ \mathrm{THR}_N^k \rangle - \langle \nu, \mathrm{OR}_R \circ \mathrm{THR}_N^k \rangle \right) && \text{by Equation (82)} \\
&\geq \frac{1}{\|\Gamma - \nu\|_1} \left(1/3 - \langle \nu, \mathrm{OR}_R \circ \mathrm{THR}_N^k \rangle \right) && \text{by Claim 6.4} \\
&\geq \frac{1}{\|\Gamma - \nu\|_1} \{1/3 - \|\nu\|_1\} \geq \frac{1}{\|\Gamma - \nu\|_1} \frac{7}{30} && \text{by Equation (81)} \\
&\geq \frac{7}{33}. && \text{since } \|\Gamma - \nu\|_1 \leq \|\Gamma\|_1 + \|\nu\|_1 \leq \tfrac{11}{10} \text{ by the triangle inequality}
\end{aligned}$$

We have from Equation (82) that

$$\begin{aligned}
\mathrm{phd}(\mathcal{W}) &= \mathrm{phd}\left(\frac{\Gamma(x) - \nu(x)}{\|\Gamma - \nu\|_1} \right) && (83) \\
&= \mathrm{phd}(\Gamma(x) - \nu(x)) && (84) \\
&\geq \min\{\mathrm{phd}(\Gamma), \mathrm{phd}(\nu)\}. && (85)
\end{aligned}$$

From Equation (80) we have

$$\begin{aligned}
\mathrm{phd}(\nu) &\geq 2(\Delta - \sqrt{R}) - 1 && (86)\\
&= 2\left(\frac{c_2 R}{4\ln^2 R}\sqrt{\frac{(2k)^k}{4^k k T N^{1/(2k)}\log N}} - \sqrt{R}\right) - 1 && \text{substituting the value of } \Delta\\
&\geq 2\left(\frac{c_2}{4}\cdot\frac{1}{\log^2 R\sqrt{\log N}}\cdot\left(\frac{k}{2}\right)^{k/2}\frac{1}{k^{1/2}}\cdot\frac{R^{3/4}}{N^{1/(4k)}} - \sqrt{R}\right) - 1\\
& && \text{using } T = \sqrt{R} \text{ and } \ln R < \log R\\
&\geq 2\left(\frac{c_2}{4}\cdot\frac{1}{\log^2 R\sqrt{k\log R}}\cdot\left(\frac{k}{2}\right)^{k/2}\frac{1}{k^{1/2}}\cdot\frac{R^{3/4}}{20^{1/(4k)}2^{1/8}k^{1/8}R^{1/(4k)}} - \sqrt{R}\right) - 1\\
& && \text{substituting the value of } N \text{ and using } k\log R > \log N \text{ for sufficiently large } R\\
&= 2\left(\frac{c_2}{2^{17/8}}\cdot\frac{1}{\log^2 R\cdot\sqrt{\log R}}\cdot\left(\frac{k}{2}\right)^{k/2}\frac{1}{k^{9/8}\cdot 20^{1/(4k)}}\cdot R^{3/4-1/(4k)} - \sqrt{R}\right) - 1 && (87)\\
&\geq 2\left(\frac{c_2}{2^{26/8}}\cdot\frac{1}{\log^{5/2} R}\cdot\frac{1}{20^{1/(4k)}}\cdot R^{3/4-1/(4k)} - \sqrt{R}\right) - 1\\
& && \text{since } \left(\tfrac{k}{2}\right)^{k/2}\tfrac{1}{k^{9/8}} \geq \tfrac{1}{2^{9/8}} \text{ for all } k \geq 2\\
&\geq 2\left(\frac{c_2}{320}\cdot\frac{1}{\log^{5/2} R}\cdot R^{3/4-1/(4k)} - \sqrt{R}\right) - 1 && (88)\\
&\geq \frac{c_2}{320}\cdot\frac{1}{\log^{5/2} R}\cdot R^{3/4-1/(4k)} - 1\\
& && \text{since } \sqrt{R} \leq \tfrac{c_2}{640}\cdot\tfrac{1}{\log^{5/2} R}\cdot R^{3/4-1/(4k)} \text{ for } k \geq 2, \text{ for sufficiently large } R\\
&= \Omega\left(\frac{1}{\log^{5/2} R}\cdot R^{3/4-1/(4k)}\right). && (89)
\end{aligned}$$

Therefore by Claim 6.2 and Equation (85), we have $\mathrm{phd}(\mathcal{W}) = \Omega\left(\frac{1}{4^k k^2}\cdot\frac{1}{\log^{5/2} R}\cdot R^{\frac{3}{4}-\frac{1}{4k}}\right)$, justifying Equation (78) and finishing the proof.

□

7 An Upper Bound

We extend ideas from Sherstov's upper bound on the approximate degree of surjectivity [She18a] to prove an approximate degree upper bound for k-distinctness, where k is not necessarily a constant. We first note that it suffices to show an approximate degree upper bound on $\left(\mathrm{OR}_R \circ \mathrm{THR}_N^k\right)^{\leq N}$.

Claim 7.1. *For any positive integers k, R, N,*

$$\widetilde{\deg}\left(\mathrm{DIST}_{N,R}^k\right) \leq \widetilde{\deg}\left(\left(\mathrm{OR}_R \circ \mathrm{THR}_N^k\right)^{\leq N}\right)\cdot O(\log R). \tag{90}$$

Claim 7.1 has essentially appeared in multiple prior works, e.g., [BT17, Equation 4], [BKT17, Section 3.4.1], [She18a, Section 6]. Claim 7.1 is a converse to Claim 4.3, but is far more straightforward to prove than Claim 4.3. Claim 7.1 follows from the fact that $\mathrm{DIST}_{N,R}^k$ can be written as an OR over all R range items j of the function that tests whether k or more copies of i appear in

the input list. In more detail, for $i \in [N]$ and $j \in [R]$, let $y_{i,j}(x) = -1$ if the ith item of the input list equals range item j. Note that $y_{ij}(x)$ is a function of degree at most $\lceil \log_2 R \rceil$ in x. Moreover,

$$\mathrm{DIST}^k_{N,R} = (\mathrm{OR}_R \circ \mathrm{THR}^k_N)(y_{1,1}(x), y_{2,1}(x), \ldots, y_{R,N}(x)).$$

Claim 7.1 follows.

The following is our main theorem in this section.

Theorem 7.2. *For any positive integers k, R, N, with $k \leq N/2$,*

$$\widetilde{\deg}\left(\left(\mathrm{OR}_R \circ \mathrm{THR}^k_N\right)^{\leq N}\right) = O(N^{1/2} R^{1/4} \sqrt{k \log N}).$$

For any integers $N \geq i \geq 0$, define the function $\mathrm{EXACT}^i_N : \{-1,1\}^N \to \{0,1\}$ by

$$\mathrm{EXACT}^i_N(x) = \begin{cases} 1 & |x| = i \\ 0 & \text{otherwise.} \end{cases}$$

Note that

$$\mathrm{EXACT}^i_N(x) = \sum_{S \subseteq [N] : |S| = i} \prod_{u \in S} \left(\frac{1 - x_u}{2}\right) \prod_{v \notin S} \left(\frac{1 + x_v}{2}\right). \tag{91}$$

Recall that for integers $N \geq k \geq 0$, the function $\mathrm{THR}^k_N : \{-1,1\}^N \to \{-1,1\}$ is defined by

$$\mathrm{THR}^k_N(x) = \begin{cases} -1 & |x| \geq k \\ 1 & \text{otherwise.} \end{cases}$$

We have

$$\mathrm{THR}^k_N(x) = 2\left(\sum_{i=0}^{k-1} \mathrm{EXACT}^i_N(x)\right) - 1 \tag{92}$$

since exactly one summand outputs 1 if the Hamming weight of x is less than k, and all summands output 0 otherwise.

For integers $m, R \geq 0$, define a degree-m polynomial $p : \{-1,1\}^R \to \mathbb{R}$ by

$$p(x) = \frac{2}{T_m\left(1 + \frac{1}{R}\right)} \cdot T_m\left(\frac{\sum_{i=1}^R x_i}{R} + \frac{1}{R}\right) - 1. \tag{93}$$

Note that when $|x| = 0$, we have $\sum_{i=1}^R x_i = R$, and hence $p(x) = 1$. When $|x| > 0$, we have $\frac{\sum_{i=1}^R x_i}{R} + \frac{1}{R} \in [-1, 1]$, and by Equation (35) this implies $T_m\left(\frac{\sum_{i=1}^R x_i}{R} + \frac{1}{R}\right) \in [-1, 1]$, and thus $p(x) \in \left[-1 - \frac{2}{1+(m^2/R)}, -1 + \frac{2}{1+(m^2/R)}\right]$ by Equation (36). The next claim immediately follows.

Claim 7.3. *The degree-m polynomial p defined in Equation (93) uniformly approximates OR_R to error $\frac{2}{1+(m^2/R)}$.*

We are now ready to prove our final upper bound.

Proof of Theorem 7.2. Let $m \geq 1$ be an integer parameter to be fixed later and let T_m be the degree-m Chebyshev polynomial. Thus by Claim 7.3, the function $\mathrm{OR}_R \circ \mathrm{THR}_N^k$ is approximated pointwise to error $\frac{2}{1+(m^2/R)}$ by the degree-m polynomial $p : \{-1,1\}^{RN} \to \mathbb{R}$ defined by

$$\begin{aligned}
p(x) &= \frac{2}{T_m\left(1+\frac{1}{R}\right)} \cdot T_m\left(\frac{1}{R} + \frac{1}{R}\sum_{j=1}^{R} \mathrm{THR}_N^k(x_{j,1},\ldots,x_{j,N})\right) - 1 \\
&= \frac{2}{T_m\left(1+\frac{1}{R}\right)} \cdot T_m\left(\frac{1}{R} - 1 + \frac{2}{R}\sum_{j=1}^{R}\sum_{i=0}^{k-1} \mathrm{EXACT}_N^i(x_{j,1},\ldots,x_{j,N})\right) - 1 \quad \text{by Equation (92)} \\
&= \frac{2}{T_m\left(1+\frac{1}{R}\right)} \cdot T_m\left(\frac{1}{R} - 1 + \frac{2}{R}\sum_{j=1}^{R}\sum_{i=0}^{k-1}\left(\sum_{S\subseteq[N]:|S|=i}\prod_{u\in S}\left(\frac{1-x_{j,u}}{2}\right)\prod_{v\notin S}\left(\frac{1+x_{j,v}}{2}\right)\right)\right). \\
&\qquad\qquad \text{by Equation (91)}
\end{aligned}$$

For simplicity of notation, define

$$C_{j,S} := \prod_{u\in S}\left(\frac{1-x_{j,u}}{2}\right)\prod_{v\notin S}\left(\frac{1+x_{j,v}}{2}\right). \tag{94}$$

We next show an upper bound on $\rho(p)$ (recall that $\rho(p)$ is the conjunction norm of p defined in Definition 3.29).

$$\begin{aligned}
\rho(p) &= \left|\frac{2}{T_m(1+\frac{1}{R})}\right| \cdot \rho\left(T_m\left(\frac{1}{R} - 1 + \frac{2}{R}\sum_{j=1}^{R}\sum_{i=0}^{k-1}\left(\sum_{S\subseteq[N]:|S|=i} C_{j,S}\right)\right)\right) \qquad \text{by Equation (38)} \\
&\leq 2\cdot 3^m \cdot \rho\left(\frac{1}{R} - 1 + \frac{2}{R}\sum_{j=1}^{R}\sum_{i=0}^{k-1}\left(\sum_{S\subseteq[N]:|S|=i} C_{j,S}\right)\right)^m \\
&\qquad\qquad \text{by Equation (41), Equation (34), and } T_m(1+\tfrac{1}{R}) > 1 \\
&\leq 2\cdot 3^m \cdot \left(\left|\frac{1}{R} - 1\right| + \rho\left(\frac{2}{R}\sum_{j=1}^{R}\sum_{i=0}^{k-1}\left(\sum_{S\subseteq[N]:|S|=i} C_{j,S}\right)\right)\right)^m \qquad \text{by Equation (39)} \\
&\leq 2\cdot 3^m \cdot \left(1 + \frac{2}{R}\sum_{j=1}^{R}\rho\left(\sum_{i=0}^{k-1}\left(\sum_{S\subseteq[N]:|S|=i} C_{j,S}\right)\right)\right)^m \qquad \text{by Equation (38) and Equation (39)} \\
&\leq 2\cdot 3^m \cdot \left(1 + 2\cdot k\binom{N}{k}\right)^m \qquad \text{by Equation (39) and } \rho(C_{j,S}) \text{ is at most 1} \\
&\leq \left(c_1 \cdot k\binom{N}{k}\right)^m, \qquad (95)
\end{aligned}$$

for some positive constant c_1. By Claim 3.28, we have the following. For each conjunction f there is a degree-d polynomial p_f such that $|p_f(x) - f(x)| \leq 2^{-c\cdot d^2/N}$ for all $x \in \left(\{-1,1\}^{RN}\right)^{\leq N}$ for some positive constant c. By construction, $\deg(p) = m$ and $\left|p(x) - \left(\mathrm{OR}_R \circ \mathrm{THR}_N^k\right)(x)\right| \leq 2/\left(1+\frac{m^2}{R}\right)$ for all $x \in \{-1,1\}^{RN}$. By the triangle inequality, we obtain that for any integers

$m, d \geq 0$,

$$\begin{aligned}
E\left(\left(\mathrm{OR}_R \circ \mathrm{THR}_N^k\right)^{\leq N}, d\right) &\leq \frac{2}{1+\frac{m^2}{R}} + \rho(p) \cdot 2^{-c \cdot d^2/N} \\
&\leq \frac{2}{1+\frac{m^2}{R}} + 2^{m \log\left(c_1 k \binom{N}{k}\right)} \cdot 2^{-c \cdot d^2/N} && \text{by Equation (95)} \\
&\leq \frac{2}{7} + 2^{\sqrt{6R} \log\left(c_1 k \binom{N}{k}\right) - c \cdot d^2/N} && \text{setting } m = \sqrt{6R} \\
&\leq \frac{2}{7} + 2^{\sqrt{6R}(\log(c_1) + \log(k) + \log(N^k)) - c \cdot d^2/N} && \text{since } \tbinom{N}{k} \leq N^k \\
&\leq \frac{2}{7} + 2^{3\sqrt{6R} k \log(N) - cd^2/N} && \text{for sufficiently large } N \\
&\leq \frac{1}{3}. && \text{for } d = \tfrac{4}{c} \cdot R^{1/4} \sqrt{Nk \log N}
\end{aligned}$$

Hence there is a polynomial of degree $\frac{4}{c} \cdot R^{1/4} \sqrt{Nk \log N}$ that approximates $\left(\mathrm{OR}_R \circ \mathrm{THR}_N^k\right)^{\leq N}$ within error $1/3$, and the theorem follows.

□

Combining Claim 7.1 and Theorem 7.2 immediately yields an upper bound on the approximate degree of k-distinctness.

Corollary 7.4. *For any positive integers R, N and $k \leq N/2$,*

$$\widetilde{\deg}\left(\mathrm{DIST}_{N,R}^k\right) = O(\sqrt{k} N^{1/2} R^{1/4} \log R \sqrt{\log N}).$$

Recall (cf. Corollary 4.2) that Ambainis [Amb05] showed that, for all functions that are symmetric both with respect to range elements and with respect to domain elements, the approximate degree is the same for all range sizes greater than or equal to N. This implies that the upper bound in Corollary 7.4 can be refined to

$$\widetilde{\deg}\left(\mathrm{DIST}_{N,R}^k\right) = O(\sqrt{k} N^{1/2} \min\{N, R\}^{1/4} \log R \sqrt{\log N}).$$

Acknowledgements

JT and SZ are supported by the National Science Foundation CAREER award (grant CCF-1845125). JT is grateful to Robin Kothari for extremely useful suggestions and discussions surrounding Theorem 1.3, and to Mark Bun for essential discussions regarding Theorem 4.4. SZ would like to thank Yao Ji for several helpful conversations.

References

[ABP19] Srinivasan Arunachalam, Jop Briët, and Carlos Palazuelos. Quantum query algorithms are completely bounded forms. *SIAM Journal on Computing*, 48(3):903–925, 2019.

[Amb05] Andris Ambainis. Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. *Theory of Computing*, 1(1):37–46, 2005.

[Amb06] Andris Ambainis. Polynomial degree vs. quantum query complexity. *Journal of Computer and System Sciences*, 72(2):220–238, 2006.

[Amb07] Andris Ambainis. Quantum walk algorithm for element distinctness. *SIAM J. Comput.*, 37(1):210–239, 2007.

[AS04] Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *J. ACM*, 51(4):595–605, 2004.

[BBC+01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald De Wolf. Quantum lower bounds by polynomials. *Journal of the ACM (JACM)*, 48(4):778–797, 2001.

[Bel12] Aleksandrs Belovs. Learning-graph-based quantum algorithm for k-distinctness. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 207–216, 2012.

[BKT17] Mark Bun, Robin Kothari, and Justin Thaler. The polynomial method strikes back: Tight quantum query bounds via dual polynomials. *CoRR*, abs/1710.09079, version 3, 2017.

[BKT18] Mark Bun, Robin Kothari, and Justin Thaler. The polynomial method strikes back: tight quantum query bounds via dual polynomials. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 297–310, 2018.

[BM12] Paul Beame and Widad Machmouchi. The quantum query complexity of AC^0. *Quantum Information & Computation*, 12(7-8):670–676, 2012.

[BNRdW07] Harry Buhrman, Ilan Newman, Hein Röhrig, and Ronald de Wolf. Robust polynomials and quantum algorithms. *Theory Comput. Syst.*, 40(4):379–395, 2007.

[BT15] Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, pages 268–280, 2015.

[BT17] Mark Bun and Justin Thaler. A nearly optimal lower bound on the approximate degree of AC^0. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 1–12, 2017.

[BT19] Mark Bun and Justin Thaler. The large-error approximate degree of ac$^\wedge$ 0. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[Kut05] Samuel Kutin. Quantum lower bound for the collision problem with small range. *Theory of Computing*, 1(1):29–36, 2005.

[Lee09] Troy Lee. A note on the sign degree of formulas. *CoRR*, abs/0909.4607, 2009.

[LZ19] Qipeng Liu and Mark Zhandry. On finding quantum multi-collisions. In *Advances in Cryptology - EUROCRYPT 2019 - 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19-23, 2019, Proceedings, Part III*, pages 189–218, 2019.

[NS94] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[OS10] Ryan O'Donnell and Rocco A. Servedio. New degree bounds for polynomial threshold functions. *Combinatorica*, 30(3):327–358, 2010.

[RS10] Alexander A. Razborov and Alexander A. Sherstov. The sign-rank of AC^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010.

[She12] Alexander A. Sherstov. Strong direct product theorems for quantum communication and query complexity. *SIAM J. Comput.*, 41(5):1122–1165, 2012.

[She13] Alexander A Sherstov. The intersection of two halfspaces has high threshold degree. *SIAM Journal on Computing*, 42(6):2329–2374, 2013.

[She18a] Alexander A Sherstov. Algorithmic polynomials. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 311–324, 2018.

[She18b] Alexander A. Sherstov. The power of asymmetry in constant-depth circuits. *SIAM J. Comput.*, 47(6):2362–2434, 2018.

[ST19] Alexander A Sherstov and Justin Thaler. Vanishing-error approximate degree and qma complexity. *arXiv preprint arXiv:1909.07498*, 2019.

[SZ09] Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation*, 9(5):444–460, 2009.

[Š08] Robert Špalek. A dual polynomial for OR. *CoRR*, abs/0803.4516, 2008.

A A Dual Polynomial for Threshold Function

In this section, we prove Claim 3.14. We require the following well-known combinatorial identity. For a proof, see, for example, [OS10].

Fact A.1. *Let $N \in \mathbb{N}$ and let $p : \mathbb{R} \to \mathbb{R}$ be any polynomial of degree less than N. Then,*

$$\sum_{i=0}^{N} (-1)^i \binom{N}{i} p(i) = 0.$$

Proof of Claim 3.14. Let $E_+ := \{t \colon \omega(t) > 0, t \geq k\}$, and $E_- := \{t \colon \omega(t) < 0, t < k\}$. By normalizing, it suffices to construct a function $\omega : [T] \cup \{0\} \to \mathbb{R}$ such that

$$\sum_{t\in E_+} |\omega(t)| \le \frac{1}{48\cdot 4^k\sqrt{N}\log N}\cdot\|\omega\|_1 \tag{96}$$

$$\sum_{t\in E_-} |\omega(t)| \le \left(\frac{1}{2}-\frac{2}{4^k}\right)\cdot\|\omega\|_1 \tag{97}$$

For all univariate polynomials $q\colon \mathbb{R}\to\mathbb{R}$,

$$\deg(q) < c_1\sqrt{4^{-k}k^{-1}TN^{-1/(2k)}\log^{-1}N} \implies \sum_{t=0}^{T}\omega(t)q(t) = 0 \tag{98}$$

$$|\omega(t)| \le \frac{(2k)^k\exp\left(-c_2 t/\sqrt{4^k kTN^{1/(2k)}\log N}\right)\|\omega\|_1}{t^2} \qquad \forall t = 1,2,\ldots,T. \tag{99}$$

Let $\ell = 100k\lceil N^{1/(2k)}4^k\log N\rceil$, and let $m = \lfloor\sqrt{T/\ell}\rfloor$. Define the set

$$S = \{1,2,\ldots,k\}\cup\{\ell i^2 : 0\le i\le m\}.$$

Note that $|S| = k+m+1 \ge m = (1/10)\sqrt{4^{-k}k^{-1}TN^{-1/(2k)}\log^{-1}N}$. Define the polynomial $\omega : [T]\cup\{0\}\to\mathbb{R}$ by

$$\omega(t) = \frac{(-1)^{T-t-m+1}}{T!}\binom{T}{t}\prod_{r\in([T]\cup\{0\})\setminus S}(t-r).$$

The signs are chosen so that $\omega(k) < 0$, because in the expression

$$\omega(k) = \frac{(-1)^{T-k-m+1}}{T!}\binom{T}{k}\prod_{r\in([T]\cup\{0\})\setminus S}(k-r),$$

the number of terms in the product is $|([T]\cup\{0\})\setminus S| = T-k-m$, and each term in the product is negative for $k=0$.

Let q be any univariate polynomial of degree less than $|S|-1$. Then,

$$\begin{aligned}
\sum_{t=1}^{T}\omega(t)q(t) &= \frac{(-1)^{T-m+1}}{T!}\sum_{t=1}^{T}(-1)^t\binom{T}{t}\prod_{r\in([T]\cup\{0\})\setminus S}(t-r)\cdot q(t) \\
&\qquad\qquad \text{since } (-1)^{-t} = (-1)^t \text{ for all integer } t \\
&= \frac{(-1)^{T-m+1}}{T!}\sum_{t=1}^{T}(-1)^t\binom{T}{t}p(t) \qquad \text{where } p(t) := \textstyle\prod_{r\in([T]\cup\{0\})\setminus S}(t-r)\cdot q(t) \\
&= 0 \qquad\qquad \text{by Fact A.1}
\end{aligned}$$

where we could use Fact A.1 since $\deg(p) \le \deg(q) + \deg(\prod_{r\in([T]\cup\{0\})\setminus S}(t-r)) \le |[T]\cup\{0\}| - |S| + \deg(q) < T+1-|S|+|S|-1 = T$.

Since $|S|-1 = k+m > m = (1/10)\sqrt{4^{-k}k^{-1}TN^{-1/(2k)}\log^{-1}N}$, we conclude that ω satisfies Equation (98) for $c_1 = 1/10$. We now show that Equation (99) holds. For $t = 1,\ldots,k$, we have

$$\frac{(2k)^k\exp\left(-c_2 t/\sqrt{4^k kTN^{1/(2k)}\log N}\right)}{t^2} \ge \frac{(2k)^k\exp\left(-c_2\sqrt{k}\right)}{k^2} \ge 1$$

as long as $c_2 \le 1/2$ and $k \ge 2$. Since $|\omega(t)| \le \|\omega\|_1$, the bound holds for $t = 1, \dots, k$.

Next, note that $\omega(t) = 0$ for $t \notin S$. For $t \in S$, we have

$$\begin{aligned}
|\omega(t)| &= \frac{1}{T!} \cdot \frac{T!}{t!(T-t)!} \cdot \frac{\prod_{r \in ([T] \cup \{0\}) \setminus S} |t-r| \cdot \prod_{r \in S \setminus \{t\}} |t-r|}{\prod_{r \in S \setminus \{t\}} |t-r|} \\
&= \frac{1}{t!(T-t)!} \cdot \frac{\prod_{r \in ([T] \cup \{0\}) \setminus \{t\}} |t-r|}{\prod_{r \in S \setminus \{t\}} |t-r|} \\
&= \frac{1}{t!(T-t)!} \cdot \frac{t!(T-t)!}{\prod_{r \in S \setminus \{t\}} |t-r|} = \prod_{r \in S \setminus \{t\}} \frac{1}{|t-r|}.
\end{aligned}$$

Thus,

$$|\omega(t)| = \begin{cases} \prod_{r \in S \setminus \{t\}} \frac{1}{|t-r|} & \text{for } t \in S, \\ 0 & \text{otherwise.} \end{cases}$$

For $t \in \{0, 1, \dots, k\}$, we observe that

$$\frac{|\omega(t)|}{|\omega(k)|} = \frac{k! \cdot \prod_{i=1}^{m} (\ell i^2 - k)}{t! \cdot (k-t)! \cdot \prod_{i=1}^{m} (\ell i^2 - t)} \le \binom{k}{t}. \tag{100}$$

Meanwhile, for $t = \ell j^2$ with $j \ge 1$, we get

$$\begin{aligned}
\frac{|\omega(t)|}{|\omega(k)|} &= \frac{k! \cdot \prod_{i=1}^{m} (\ell i^2 - k)}{\prod_{i=1}^{k} (\ell j^2 - i) \cdot \prod_{i \in \{[m] \cup \{0\}\} \setminus \{j\}} |\ell i^2 - \ell j^2|} \\
&\le \frac{k! \cdot \prod_{i=1}^{m} \ell i^2}{(\ell j^2 - k)^k \cdot \prod_{i \in \{[m] \cup \{0\}\} \setminus \{j\}} \ell (i+j)|i-j|} \\
&= \frac{2 \cdot k!}{(\ell j^2 - k)^k} \cdot \frac{(m!)^2}{(m+j)!(m-j)!}.
\end{aligned}$$

The first factor is bounded above by

$$\frac{2 \cdot k!}{(\ell - k)^k j^{2k}}.$$

Since $\ell \ge 2k$ by our choice of ℓ, and $k \ge 2$, this expression is at most

$$\frac{k^k}{(\ell/2)^k j^4} = \frac{(2k)^k}{\ell^k \cdot j^4}.$$

We control the second factor by

$$\begin{aligned}
\frac{(m!)^2}{(m+j)!(m-j)!} &= \frac{m}{m+j} \cdot \frac{m-1}{m+j-1} \cdot \dots \cdot \frac{m-j+1}{m+1} \\
&\le \left(\frac{m}{m+j} \right)^j \\
&\le \left(1 - \frac{j}{2m} \right)^j \\
&\le e^{-j^2/2m},
\end{aligned}$$

where the last inequality uses the fact that $1 - x \le e^{-x}$ for all x. Hence,

$$\frac{|\omega(\ell j^2)|}{|\omega(k)|} \le \frac{(2k)^k}{\ell^k \cdot j^4} \cdot e^{-j^2/2m}. \tag{101}$$

This immediately yields

$$\frac{|\omega(\ell j^2)|}{\|\omega\|_1} \le \frac{|\omega(\ell j^2)|}{|\omega(k)|} \le \frac{(2k)^k}{(\ell j^2)^2} \cdot e^{-\ell j^2/(2\ell m)},$$

If we choose $c_2 < 1/20$, we have $\frac{1}{2\ell m} \ge \frac{1}{2\sqrt{T\ell}} > \frac{c_2}{\sqrt{4^k k T N^{1/(2k)} \log N}}$ since $\ell = 100k\lceil N^{1/(2k)} 4^k \log N\rceil$. This establishes Equation (99) for all $t = \ell j^2 > k$. Moreover, by Equation (101),

$$\begin{aligned}
\sum_{t>k} |\omega(t)| &\le |\omega(k)| \cdot \sum_{j=1}^{m} \frac{(2k)^k}{\ell^k \cdot j^4} \cdot e^{-j^2/2m} \\
&\le \left(\frac{2k}{\ell}\right)^k \cdot |\omega(k)| \cdot \sum_{j=1}^{m} \frac{1}{j^4} \\
&< \frac{(2k)^k}{(100kN^{1/(2k)}4^k \log N)^k} \cdot \frac{\pi^4}{50} \cdot |\omega(k)| && \text{since } \textstyle\sum_{j=1}^{\infty} 1/j^4 < \pi^4/50 \\
&= \frac{1}{50^k \sqrt{N} (4^k)^k \log^k N} \cdot \frac{1}{50/\pi^4} \cdot |\omega(k)| \\
&\le \frac{1}{50^2 \sqrt{N} 4^k \log N} \cdot \frac{1}{50/\pi^4} \cdot |\omega(k)| && \text{for all } k \ge 2 \\
&\le \frac{|\omega(k)|}{48 \cdot 4^k \cdot \sqrt{N} \log N}. && (102)
\end{aligned}$$

Hence, since $\omega(k) < 0$,

$$\sum_{t \in E_+} |\omega(t)| \le \sum_{t>k} |\omega(t)| \le \frac{|\omega(k)|}{48 \cdot 4^k \cdot \sqrt{N} \log N} \le \frac{\|\omega\|_1}{48 \cdot 4^k \cdot \sqrt{N} \log N},$$

which gives Equation (96).

Finally, to establish Equation (97), we combine Equation (100) and Equation (102) to obtain

$$\frac{\|\omega\|_1}{|\omega(k)|} \le \sum_{t=0}^{k} \binom{k}{t} + \frac{1}{48 \cdot 4^k \cdot \sqrt{N} \log N} < 2^k + 1 < \frac{1}{2} \cdot 4^k. \tag{103}$$

We calculate

$$\begin{aligned}
\frac{\|\omega\|_1}{2} - \sum_{t \in E_-} |\omega(t)| &= \frac{1}{2} \left(\sum_{t:\omega(t)<0} |\omega(t)| + \sum_{t:\omega(t)>0} |\omega(t)| \right) - \sum_{t \in E_-} (-\omega(t)) \\
&= \sum_{t:\omega(t)<0} (-\omega(t)) - \sum_{t \in E_-} (-\omega(t)) \\
&\qquad \text{since } \langle \omega, \mathbf{1} \rangle = 0, \text{ so } \textstyle\sum_{t:\omega(t)<0} |\omega(t)| = \sum_{t:\omega(t)>0} |\omega(t)| \\
&= \sum_{t\colon \omega(t)<0, t \ge k} (-\omega(t)) && \text{since } E_- = \{t\colon \omega(t) < 0, t < k\} \\
&\ge -\omega(k).
\end{aligned}$$

Rearranging and applying the bound in Equation (103),

$$\sum_{t \in E_-} |\omega(t)| \leq \left(\frac{1}{2} + \frac{\omega(k)}{\|\omega\|_1}\right) \cdot \|\omega\|_1 \leq \left(\frac{1}{2} - 2 \cdot 4^{-k}\right) \cdot \|\omega\|_1.$$

□

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il