Improved Approximate Degree Bounds for k-Distinctness

Nikhil S. Mande
Georgetown University, Washington DC, USA
nikhil.mande@georgetown.edu

Justin Thaler
Georgetown University, Washington DC, USA
justin.thaler@georgetown.edu

Shuchen Zhu
Georgetown University, Washington DC, USA
shuchen.zhu@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 $k \leq \text{polylog}(N)$.

2012 ACM Subject Classification Theory of computation → Quantum query complexity

Keywords and phrases Quantum Query Complexity, Approximate Degree, Dual Polynomials, k-distinctness

Digital Object Identifier 10.4230/LIPIcs.TQC.2020.2

Related Version A full version of the paper is available at `https://arxiv.org/abs/2002.08389`

Funding *Justin Thaler*: Supported by the National Science Foundation CAREER award (grant CCF-1845125).
Shuchen Zhu: Supported by the National Science Foundation CAREER award (grant CCF-1845125).

Acknowledgements JT is grateful to Robin Kothari for extremely useful suggestions and discussions surrounding Theorem 2, and to Mark Bun for essential discussions regarding Theorem 19. SZ would like to thank Yao Ji for several helpful conversations.

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

15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020).
Editor: Steven T. Flammia; Article No. 2; pp. 2:1–2:22
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

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 [18] 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 $\Theta(N^{2/3})$ [4, 1].

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 [8]. 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 [1], with refinements given by Kutin [15] and Ambainis [2]. This lower bound is tight for $k = 2$ (matching Ambainis' upper bound [4]), but it is not known to be tight for any $k > 2$. Recently, Bun, Kothari, and Thaler [11] 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

This paper 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 [11], 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 $f: \{-1,1\}^n \to \{-1,1\}$, 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. [6], 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 [1, 15, 2, 11], our k-distinctness lower bound is in fact an approximate

[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.

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

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. [11].

▶ **Theorem 1** (Informal version of Theorem 17 and Corollary 18). *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 $\tilde{\Omega}(N^{3/4-1/(4k)})$.*

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 is $O\left(N^{3/4-1/(2^{k+2}-4)}\right)$ [8]. For non-constant values of k, the upper bound implied by Belovs' 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' bound 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 2** (Informal). *For any $k \leq \mathrm{polylog}(N)$, 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 [4, 8, 24] 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 2 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 2 does *not* yield a quantum query upper bound, only an approximate degree upper bound. It remains an interesting open question whether the quantum query complexity of k-distinctness is sublinear in N for all $k = \mathrm{polylog}(N)$ (see Section 1.1 for further discussion).

Our proof of Theorem 2 is a simple extension of a result of Sherstov [24, Theorem 1.3] that yielded an $O(N^{3/4})$ approximate degree upper bound for a different function called Surjectivity.[5] A formal statement and proof can be found in the full version of this paper.

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

[4] Belovs' approximate degree upper bound was recently reproved by Sherstov [24], who made the exponential dependence on k explicit (see, e.g., [24, 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.

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 1.2), 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 [18] 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).

It would also be interesting to resolve the quantum query complexity of k-distinctness for $k = \text{polylog}(N)$. Although this question may 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. [6], no transformation in the reverse direction is possible in general [3]. This can be seen, for example, because the quantum query complexity of Surjectivity is known to be $\Omega(N)$ [7, 25], but its approximate degree is $O(N^{3/4})$ [24, 11]. 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 $(\text{polylog}(N))$-distinctness could be derived in this manner. Such an upper bound (even for $(\log N)$-distinctness) would yield improved quantum query upper bounds for min-entropy estimation [17]. On the other hand, due to our Theorem 2, any $N^{3/4+\Omega(1)}$ *lower bound* for $(\text{polylog}(N))$-distinctness would require moving beyond the polynomial method.[7]

1.2 Overview of 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 $x \in (\{-1,1\}^n)^{\leq N}$.

[6] There are works in this general direction, notably [5], 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.

Simplifying very slightly, prior work by Bun and Thaler [13] (building on an important lemma of Ambainis [2]) implied that for $k \geq 2$ the approximate degree of k-distinctness is equivalent to $\widetilde{\deg}(f^{\leq N}_{RN})$ for $f = \mathrm{OR}_R \circ \mathrm{THR}^k_N$. Here, $g_n \circ h_m$ denotes the function on $n \cdot m$ bits obtained by block-composing g and h, i.e., $g \circ h$ evaluates h on n disjoint inputs and feeds the outputs of all n copies of h into g.

Bun et al. [11] proved their $\tilde{\Omega}(N^{3/4-1/(2k)})$ lower bound for $\widetilde{\deg}(f^{\leq N}_{RN})$ 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 $\widetilde{\deg}(f^{\leq N}_{RN}) \geq d$ is a function $\psi\colon \{-1,1\}^{RN} \to \mathbb{R}$ satisfying the following properties (this dual formulation is standard, and can be found, for example, in [21]).

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. [11] constructed their dual witness for $\left(\mathrm{OR}_R \circ \mathrm{THR}^k_N\right)^{\leq N}$ roughly as follows. They took a dual witness Ψ for the fact that $\widetilde{\deg}(\mathrm{OR}_R) = \Omega(R^{1/2})$ [19, 28, 12] and a dual witness ϕ for the fact that THR^k_N also has large approximate degree, and they combined Ψ and ϕ in a certain manner (introduced in prior works [27, 23, 16]) to get a dual witness for the composed function $\left(\mathrm{OR}_R \circ \mathrm{THR}^k_N\right)^{\leq N}$. 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 $\left(\mathrm{OR}_R \circ \mathrm{THR}^k_N\right)^{\leq N}$ 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 $\left(\mathrm{OR}_R \circ \mathrm{THR}^k_N\right)^{\leq N}$. 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 [11]. 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 high level, what we do is replace the dual block composition $\Psi \star \phi$ from the construction of [11] with a *variant* of dual block composition introduced by Sherstov [22]. Sherstov introduced this variant to address the correlation issues that arise when attempting

[9] To clarify, this entire outline is a major simplification of the actual dual witness construction in [11]. 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 [11] 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 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 $i \in \{1, \ldots, R\}$.

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 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 $\langle \phi, h \rangle = \|\phi\|_1$ [23, 16]. 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 $(\text{sgn}(\phi(x_1)), \ldots, \text{sgn}(\phi(x_R)))$ into Ψ. One can think of $\text{sgn}(\phi(x_i))$ 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 $\text{sgn}(\phi(x_i)) \neq h(x_i)$, meaning that ϕ's predictions can sometimes be wrong. In this case, when feeding $\text{sgn}(\phi(x_i))$ 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 $\text{sgn}(\phi(x_i))$ 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, $\text{sgn}(\phi(x_i))$ 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, $g = \text{OR}_R$ has low (-1)-*certificate complexity*, meaning that on inputs x in $\text{OR}_R^{-1}(-1)$, to certify that indeed $x \in \text{OR}_R^{-1}(-1)$, 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

[11] Variants of dual block composition related to the one introduced in [22] have played important roles in other recent works on approximate degree lower bounds, e.g., [14, 26].

[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 $\Psi \star \phi$.

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.

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. [11] 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 [11] 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 [22] 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 $U_\eta := \{x = (x_1, \dots, x_R) \colon \mathrm{sgn}(\phi(x_i)) \neq h(x_i)$ for between 1 and η 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 [11]. 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^k_N satisfying

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

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 3 to help the reader make sense of them.

2 Preliminaries

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. For any function $f : \{-1,1\}^n \to \mathbb{R}$, define $\|f\|_1 := \sum_{x \in \{-1,1\}^n} |f(x)|$. All logarithms in this paper are base 2 unless otherwise specified. Let $\mathbf{1}^n$ (respectively, $-\mathbf{1}^n$) denote the n-bit string $(1,1,\ldots,1)$ (respectively, $(-1,-1,\ldots,-1)$). We use the notation $[n]$ to denote the set $\{1,2,\ldots,n\}$.

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

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 $\mu_\psi(x) = |\psi(x)|$. 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 $x \in \{-1,1\}^n$.

▶ **Definition 3** (k-distinctness). *For integers k, N, R with $k \leq N$, define the function* $\mathrm{DIST}^k_{N,R} : [R]^N \to \{-1,1\}$ *by* $\mathrm{DIST}^k_{N,R}(s_1, \ldots, s_N) = -1$ *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 $\{-1,1\}^{N \log R}$.*

▶ **Definition 4** (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.

Note that this definition places no constraints on an approximating polynomial on inputs outside the promise domain.

We require the following relation between approximate degree of k-distinctness and a related Boolean function; this relationship follows from [10, Proposition 21 and Corollary 26].

▷ Claim 5 ([10]). Let $N, R \in \mathbb{N}$ and $2 \leq k \leq N$ be any integer. Then for any $\epsilon > 0$,

$$\widetilde{\deg}_\epsilon(\mathrm{DIST}^k_{N,R+N}) = \tilde{\Omega}(\widetilde{\deg}_\epsilon(\mathrm{OR}_R \circ \mathrm{THR}^k_N)^{\leq N}). \tag{7}$$

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

▶ **Lemma 6** ([9]). *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any (possibly partial) Boolean function and let $0 < \epsilon < 1$. Then, $\widetilde{\deg}_\epsilon(f) = \widetilde{\deg}(f) \cdot O(\log(1/\epsilon))$.*[13]

▶ **Definition 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 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.*

By linear programming duality, we have the following standard equivalence between lower bounds on approximate degree and existence of "dual polynomials". See, for example, [10].

▶ **Lemma 9.** *Let $f : \{-1,1\}^n \to \{-1,1\}$ be any function. For any integer $0 \leq j \leq n$, we have $\widetilde{\deg}_\epsilon(f^{\leq j}) \geq d$ if and only if there exists a "dual polynomial" $\phi : \{-1,1\}^n \to \mathbb{R}$ satisfying the following properties: $\phi(x) = 0$ for all $|x| > j$, $\langle f, \phi \rangle > \epsilon$, $\sum_{x \in \{-1,1\}^n} |\phi(x)| = 1$, and* $\mathrm{phd}(\phi) \geq d$. *We say that ϕ is a dual polynomial witnessing the fact that $\widetilde{\deg}_\epsilon(f^{\leq j}) \geq d$. For brevity, when ϵ and d are clear from context, we say that ϕ is a dual polynomial for $f^{\leq j}$.*

Špalek [28] exhibited an explicit dual witness for OR (the existence of a dual witness for OR was already implicit from the work of Nisan and Szegedy [19]).

▷ Claim 10 (Implicit in [19]). There exists a constant $c \in (0,1]$ such that for any integer $n \geq 0$, there exists a function $\theta : \{-1,1\}^n \to \mathbb{R}$ satisfying $\|\theta\|_1 = 1$, $\mathrm{phd}(\theta) \geq c\sqrt{n}$, and $\langle \theta, \mathrm{OR}_n \rangle \geq 3/5$.

[13] The statement in [9] only deals with total functions. It can be seen that the proof works for partial functions too.

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 [27, 16, 23] introduced the notion of "dual block composition", which is a powerful method of combining dual witnesses.

▶ **Definition 11** (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*

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

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 [12].

▷ Claim 12 ([12]). Define $\phi : \{-1,1\}^n \to \mathbb{R}$ as $\phi(x) = -1/2$ if $x = -\mathbf{1}^n$, $\phi(x) = 1/2$ if $x = \mathbf{1}^n$ and $\phi(x) = 0$ otherwise. Then, $\mathrm{phd}(\phi) = 1$.

Next we require a lemma, implicit in a result of Razborov and Sherstov [20] (also see [13] for a formulation similar to the one we require).

▶ **Lemma 13** (Implicit in [20]). *Let $N \geq R$ be positive integers, $\Delta \in \mathbb{R}^+$, and $\theta : \{-1,1\}^{RN} \to \mathbb{R}$ be any polynomial such that*

$$\sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |\theta(x)| \leq (2NR)^{-\Delta}. \tag{8}$$

For any positive integer $D < \Delta$, there exists a function $\nu : \{-1,1\}^{RN} \to \mathbb{R}$ such that $\mathrm{phd}(\nu) > D$, $\|\nu\|_1 \leq 1/10$, and $|x| > N \Rightarrow \nu(x) = \theta(x)$.

Lemma 13 helps us convert a dual polynomial θ with little mass on large Hamming weight inputs to a dual polynomial $(\theta - \nu)/\|\theta - \nu\|_1$ with no mass on large Hamming weight inputs without affecting the pure high degree by much.

▶ **Definition 14.** *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 $1 - \eta_i$.*

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

$$\epsilon^+_{f,\psi} := \Pr_{\mu_\psi}[f(x)\psi(x) < 0 | \psi(x) > 0],\ \epsilon^-_{f,\psi} := \Pr_{\mu_\psi}[f(x)\psi(x) < 0 | \psi(x) < 0],\ \epsilon_{f,\psi} = \epsilon^+_{f,\psi} + \epsilon^-_{f,\psi}. \tag{9}$$

▶ **Definition 15.** *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 $\delta^-_{f,\psi} := \sum_{x \in E^-(f,\psi)} |\psi(x)|$.*

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 (9). Define the function $\alpha_{f,\psi} : \{-1,1\}^m \to \mathbb{R}$ as

$$\alpha_{f,\psi}(x) := \begin{cases} 1 =: a^+ & \text{if } \psi(x)f(x) > 0, \psi(x) > 0 \\ \frac{1-2\epsilon+\epsilon^-}{1-\epsilon^-} =: a^- & \text{if } \psi(x)f(x) > 0, \psi(x) < 0 \\ -1 & \text{if } \psi(x)f(x) < 0, \psi(x) > 0 \\ 1 & \text{if } \psi(x)f(x) < 0, \psi(x) < 0. \end{cases} \tag{10}$$

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

▶ **Lemma 16** ([22, 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*

$$p_\eta(z) = (-1)^\eta \prod_{i=1}^{\eta} (|z| - i), \forall z \in \{-1,1\}^n. \tag{11}$$

Then,

$$p_\eta(\mathbf{1}^n) = \eta!, \tag{12}$$

$$\|\hat{p}\|_1 \leq \eta! \binom{n+\eta}{\eta}, \tag{13}$$

$$\mathbb{E}_\nu[|p_\eta(z)|] \leq p_\eta(\mathbf{1}^n)\nu(\mathbf{1}^n)\,(1+A)\,, \quad \textit{where } A := \binom{n}{\eta+1} \frac{\tau^{\eta+1}}{(1-\tau)^n}. \tag{14}$$

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

3 Detailed Outline of Proof of Main Theorem

Our main theorem is as follows.

▶ **Theorem 17.** *For $R \in \mathbb{N}$ sufficiently large, $2 \leq k \leq \frac{\log R}{4}$, and some $N = \Theta(k^{k/2}R)$,*

$$\widetilde{\deg}(\mathrm{DIST}^k_{N,R+N}) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\log^{5/2} R} \cdot R^{\frac{3}{4} - \frac{1}{4k}}\right). \tag{15}$$

Ambainis [2] showed that the approximate degree[14] 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 18.** *For $R \in \mathbb{N}$ sufficiently large, $2 \leq k \leq \frac{\log R}{4}$, and some $N = \Theta(k^{k/2}R)$,*

$$\widetilde{\deg}(\mathrm{DIST}^k_{N,N}) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\log^{5/2} R} \cdot R^{\frac{3}{4} - \frac{1}{4k}}\right). \tag{16}$$

[14] There are several different conventions used in the literature when defining the domain of functions such as k-distinctness. The convention used by Ambainis [2] 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 $\lceil \log_2 R \rceil$.

To prove Theorem 17, Claim 5 implies that it suffices to prove a lower bound on $\widetilde{\deg}(\mathrm{OR}_R \circ \mathrm{THR}_N^k)^{\leq N}$.

▶ **Theorem 19.** *For $R \in \mathbb{N}$ sufficiently large, $2 \leq k \leq \frac{\log R}{4}$, and some $N = \Theta(k^{k/2}R)$,*

$$\widetilde{\deg}((\mathrm{OR}_R \circ \mathrm{THR}_N^k)^{\leq N}) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\log^{5/2} R} \cdot R^{\frac{3}{4}-\frac{1}{4k}}\right). \tag{17}$$

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.

Towards proving Theorem 19, 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:** $\langle \Gamma, (\mathrm{OR}_R \circ \mathrm{THR}_N^k)\rangle > 1/3$,
- **Exponentially little mass on inputs of large Hamming weight:** $\sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |\Gamma(x)| \leq (2NR)^{-\tilde{\Omega}\left(R^{\frac{3}{4}-\frac{1}{4k}}\right)}$.

Next, Lemma 13 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 $\mathcal{W}(x) := \frac{\Gamma(x)-\nu(x)}{\|\Gamma-\nu\|_1}$ places no mass on inputs of Hamming weight larger than N and satisfies $\|\mathcal{W}\|_1 = 1$, $\langle \mathcal{W}, (\mathrm{OR}_R \circ \mathrm{THR}_N^k)\rangle > 7/33$, and $\mathrm{phd}(\mathcal{W}) = \tilde{\Omega}\left(R^{\frac{3}{4}-\frac{1}{4k}}\right)$. Theorem 19 then follows by Lemma 9 and Lemma 6.

In the next subsection we provide a sketch of how we construct such a dual witness Γ and where our approach differs from [11].

3.1 Our Construction of Γ

Our construction of Γ is based on three dual witnesses θ, ϕ and ψ. The function θ is constructed as in Claim 10 with $n = R/4^k$. The function ϕ is defined on 4^k inputs, and is defined as in Claim 12. Our ψ is a fairly straightforward modification of [10, Proposition 55], 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 ψ satisfies the properties described in the following claim, with $T = \sqrt{R}$.

▷ Claim 20 (Modification of [10, Proposition 55]). Let $k, T, N \in \mathbb{N}$ with $2 \leq k \leq T \leq N$, and let ω_T be as constructed in Claim 27, with constants c_1, c_2. Define[15] $\psi : \{-1,1\}^N \to \mathbb{R}$ by $\psi(x) = \omega_T(|x|)/\binom{N}{|x|}$ for $x \in (\{-1,1\}^N)^{\leq T}$ and $\psi(x) = 0$ otherwise. Then

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

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

[15] Note that we suppress the dependence of ψ on T for convenience.

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

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

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

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

The false positive error between THR^k_N and ψ is $\tilde{O}(1/\sqrt{N})$ (as compared to $O(1/N)$ in [11]). The pure high degree of ψ is $\tilde{\Omega}(R^{1/4}N^{-1/(4k)})$ (as compared to $\tilde{\Omega}(R^{1/4}N^{-1/(2k)})$ in [11]). ψ satisfies a "weak decay condition", viz. $\sum_{|x|=t}|\psi(x)| \le \sigma\exp(-\beta t)/t^2$ for some constant σ (for general k, the value of σ only depends on k), and $\beta = \tilde{\Omega}(R^{1/4}N^{1/(4k)})$ (as compared to $\beta = \tilde{\Omega}(R^{1/4}N^{1/(2k)})$ in [11]).

If we were to define $\Gamma = \theta \star \phi \star \psi$, all the analyses from [11] 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 [22]. 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, $\Gamma(x_1, \ldots, x_{R/4^k})$ equals $\theta \bullet (\phi \star \psi)(x)$, which equals

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

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, p_η is defined as in Lemma 16, and α in a function whose definition 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 [22, 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 33).
- **Pure high degree:** Using our definition of p_η, and α, one can show (Claim 34) 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:** Since ψ satisfies $\sum_{|x|=t}|\psi(x)| \le \sigma\exp(-\beta t)/t^2$ for some constant σ and $\beta = \tilde{\Omega}(R^{1/4}N^{1/(4k)})$, Claim 29 implies that $\theta \star (\phi \star \psi) = (\theta \star \phi) \star \psi$ places exponentially small (in $R^{\frac{3}{4}-\frac{1}{4k}}$) mass on inputs in $\{-1,1\}^{RN}$ of Hamming weight larger than N. By the definition of Γ, 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 $R^{\frac{3}{4}-\frac{1}{4k}}$, for which we require Claim 30.
- **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 $x \in E^+(\mathrm{OR}_{4^k} \circ \mathrm{THR}^k_N, \phi \star \psi)$ and outputs either 1 or a value very close to 1 for $x \notin E^+(\mathrm{OR}_{4^k} \circ \mathrm{THR}^k_N, \phi \star \psi)$. This definition of α is our biggest departure from

Sherstov's construction in [22]; 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 Definition 11). 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^k_N 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 21, which in turn is a formalization of Equation (1) that holds with the setting of parameters mentioned above.

The key technical lemma that we use for the correlation analysis is the following, and a sketch of its proof is deferred to Appendix B.

▷ Claim 21. 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 16, let $\alpha=\alpha_{f,\xi}:\{-1,1\}^m\to\mathbb{R}$ be as defined in Equation (10), and consider the distribution μ_ξ over $\{-1,1\}^{nm}$. Let $\epsilon^+=\epsilon^+_{f,\xi}$, $\epsilon^-=\epsilon^-_{f,\xi}$, $\epsilon=\epsilon^++\epsilon^-$, and $A=\binom{n}{\eta+1}\frac{(\epsilon^+)^{\eta+1}}{(1-\epsilon^+)^n}$. If $A<1$, then,
$$\langle\mathrm{OR}\circ f,(\zeta\star\xi)(p_\eta\circ\alpha)\rangle\geq p_\eta(1-2\epsilon^+,\ldots,1-2\epsilon^+)\cdot\left(\delta-\left(2-2\frac{1-\epsilon}{1-\epsilon^+}(1-A)\right)\right). \tag{23}$$

4 Proof of Theorem 19

Due to space constraints, we omit some proofs henceforth. The reader is referred to the full version for complete proofs.

Towards proving Theorem 19, it suffices to exhibit a dual polynomial (see Lemma 9) that has ℓ_1-norm 1, sufficiently large pure high degree, good correlation with $(\mathrm{OR}_R\circ\mathrm{THR}^k_N)^{\leq N}$, and places no mass outside $(\{-1,1\}^{RN})^{\leq N}$. We first define a function Γ (Definition 23) that satisfies the first three properties above, and additionally satisfies a strong decay condition as we described in Section 3.1. In Section 4.1 we use Γ to construct a dual polynomial $\mathcal{W}$, via Lemma 13, 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 10 (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, $\beta = \frac{c_2}{\sqrt{4^k kTN^{1/(2k)}\log N}}$, $\Delta = \frac{\beta\sqrt{\sigma}R}{4\ln^2 R} = \frac{c_2 R}{4\ln^2 R}\sqrt{\frac{(2k)^k}{4^k kTN^{1/(2k)}\log N}}$, $N = \lceil 20\sqrt{\sigma}R\rceil$.
- Let $\omega_T : [T] \cup \{0\} \to \mathbb{R}$ be a function that satisfies the conditions in Claim 27 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 so that ψ satisfies the conditions in Claim 20.
- Let $\theta : \{-1,1\}^{R/4^k} \to \mathbb{R}$ be any function satisfying the conditions in Claim 10 for $n = R/4^k$ (note that $R/4^k > 0$ since $k < (\log R)/2$), and let $\phi : \{-1,1\}^{4^k} \to \mathbb{R}$ be the function defined in Claim 12 with $n = 4^k$.
- Let $p_\eta : \{-1,1\}^{R/4^k} \to \mathbb{R}$ be as defined in Lemma 16 and $\alpha := \alpha_{\phi\star\psi,\mathrm{OR}_{4^k}\circ\mathrm{THR}_N^k} : \{-1,1\}^{4^k N} \to \mathbb{R}$ be as defined in Equation (10).
- Let $\epsilon^+ := \epsilon^+_{,\mathrm{OR}_{4^k}\circ\mathrm{THR}_N^k,\phi\star\psi}$, $\epsilon^- := \epsilon^-_{\phi\star\psi}$, and $\epsilon := \epsilon^+ + \epsilon^-$.

We first show that the function $\phi \star \psi$ has large correlation with $\mathrm{OR}_{4^k} \circ \mathrm{THR}_N^k$, via an analysis that is essentially the same as in [10, Proposition 55].

▷ Claim 22.

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

We next define the function Γ.

▶ **Definition 23.** *Let* $\Gamma : \{-1,1\}^{NR} \to \mathbb{R}$ *be defined by*

$$\Gamma(x_1, \ldots, x_{R/4^k}) := \frac{(\theta \star (\phi \star \psi))(x_1, \ldots, x_{R/4^k}) \cdot p_\eta(\alpha(x_1), \ldots, \alpha(x_{R/4^k}))}{p_\eta(1-2\epsilon^+, \ldots, 1-2\epsilon^+)}, \tag{24}$$

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

▷ Claim 24.

$$\|\Gamma\|_1 = 1, \tag{25}$$
$$\mathrm{phd}(\Gamma) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\sqrt{\log R}} \cdot R^{3/4-1/(4k)}\right), \tag{26}$$
$$\langle \Gamma, (\mathrm{OR}_R \circ \mathrm{THR}_N^k)\rangle > 1/3 \tag{27}$$
$$\sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |\Gamma(x)| \leq (2NR)^{-2(\Delta-\sqrt{R})}. \tag{28}$$

Sketch of Proof of Claim 24

We require certain properties of dual block composition, and of the functions p_η and α, which are listed in Appendix A and Appendix B, respectively.

- The fact that $\|\Gamma\|_1 = 1$ follows from the definition of Γ and Claim 33.
- By the definition of Γ, we have $\mathrm{phd}(\Gamma) = \mathrm{phd}((\theta \star (\phi \star \psi)(p_\eta \circ \alpha))$. By Claim 34, this is at least $(\mathrm{phd}(\theta) - \eta) \cdot (\mathrm{phd}(\phi \star \psi))$. Next, using the facts that $\mathrm{phd}(\psi) = 1$ (Claim 12), multiplicativity of pure high degree under dual block composition (Equation (45)), and our choices of parameters, it can be shown that $\mathrm{phd}(\Gamma) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\sqrt{\log R}} \cdot R^{3/4-1/(4k)}\right)$.

- Recall from our choice of parameters and Claim 22 that $\epsilon^+ \leq \frac{1}{24\sqrt{R}\log R}$ and $\epsilon^- \leq e^{-4}$. Define $A = \binom{R/4^k}{\eta+1}\frac{(\epsilon^+)^{\eta+1}}{(1-\epsilon^+)^{R/4^k}}$. The above upper bounds on ϵ^+ and ϵ^-, and standard computations reveal that $A < 1/16$. Hence the conditions of Claim 21 are satisfied with the parameters fixed in the beginning of this section. Using Claim 21 with $\delta > 3/5$ and the above upper bounds on ϵ^+ and ϵ^-, we are able to show that $\langle \Gamma, (\mathrm{OR}_R \circ \mathrm{THR}_N^k)\rangle > 1/3$.
- We first show, using Lemma 16 and Lemma 26, that $p_\eta(1-2\epsilon^+, \ldots, 1-2\epsilon^+) \geq (1-\epsilon^+)^{R/4^k}\eta!$. Standard computations reveal that, for our choice of parameters, this quantity is at least 1. Hence, it suffices to show that $\sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |(\theta \star (\phi \star \psi)) \cdot (p_\eta \circ \alpha)(x)| \leq (2NR)^{-2(\Delta-\sqrt{R})}$. Next we observe that, using Claim 29 with $\Phi = \theta \star \phi$ and associativity of dual block composition (Equation (46)), that $\sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |((\theta \star \phi) \star \psi)(x)| \leq (2NR)^{-2\Delta}$. Since $\alpha(y) \in [-1,1]$ for all $y \in [-1,1]^{4^k N}$ (Equation (10)), it suffices to show a suitable bound on $\max_{y \in [-1,1]^{R/4^k}} |p_\eta(y)|$, which we are able to do using Claim 30.

4.1 Final Dual Polynomial

We now prove Theorem 19.

Proof of Theorem 19. We exhibit a function $\mathcal{W} : \{-1,1\}^{RN} \to \mathbb{R}$ satisfying

$$\mathcal{W}(x) = 0, \forall x \notin (\{-1,1\}^{RN})^{\leq N}, \tag{29}$$

$$\|\mathcal{W}\|_1 = 1 \tag{30}$$

$$\langle \mathcal{W}, (\mathrm{OR}_R \circ \mathrm{THR}_N^k)\rangle > 7/33, \tag{31}$$

$$\mathrm{phd}(\mathcal{W}) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\log^{5/2} R} \cdot R^{\frac{3}{4}-\frac{1}{4k}}\right). \tag{32}$$

The theorem then follows by Lemma 9 and Lemma 6. Towards the construction of such a $\mathcal{W}$, first note that by Equation (28) and Lemma 13 there exists a function $\nu : \{-1,1\}^{RN} \to \mathbb{R}$ that satisfies the following properties.

$$|x| > N \Rightarrow \nu(x) = \Gamma(x), \tag{33}$$

$$\mathrm{phd}(\nu) \geq 2(\Delta - \sqrt{R}) - 1, \tag{34}$$

$$\|\nu\|_1 \leq 1/10. \tag{35}$$

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

$$\mathcal{W}(x) := \frac{\Gamma(x) - \nu(x)}{\|\Gamma - \nu\|_1}. \tag{36}$$

Clearly Equation (29) and Equation (30) are satisfied. We show in Appendix C that the function $\mathcal{W}$ also satisfies Equation (31), and Equation (32). ◀

References

1 Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *J. ACM*, 51(4):595–605, 2004. doi:10.1145/1008731.1008735.

2 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. doi:10.4086/toc.2005.v001a003.

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

4 Andris Ambainis. Quantum walk algorithm for element distinctness. *SIAM J. Comput.*, 37(1):210–239, 2007. doi:10.1137/S0097539705447311.

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

6 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.

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

8 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. doi:10.1109/FOCS.2012.18.

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

10 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.

11 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.

12 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.

13 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.

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

15 Samuel Kutin. Quantum lower bound for the collision problem with small range. *Theory of Computing*, 1(1):29–36, 2005. doi:10.4086/toc.2005.v001a002.

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

17 Tongyang Li and Xiaodi Wu. Quantum query complexity of entropy estimation. *IEEE Trans. Inf. Theory*, 65(5):2899–2921, 2019.

18 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. doi:10.1007/978-3-030-17659-4_7.

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

20 Alexander A. Razborov and Alexander A. Sherstov. The sign-rank of AC^0. *SIAM J. Comput.*, 39(5):1833–1855, 2010. doi:10.1137/080744037.

21 Alexander A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011.

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

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

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

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

26 Alexander A Sherstov and Justin Thaler. Vanishing-error approximate degree and QMA complexity. *arXiv preprint arXiv:1909.07498*, 2019.

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

28 Robert Špalek. A dual polynomial for OR. *CoRR*, abs/0803.4516, 2008.

A Preliminaries

▶ **Definition 25.** *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

$$\mu_z(x_1,\ldots,x_n) = \prod_{i=1}^{n} \mu_{z_i}(x_i). \tag{37}$$

▶ **Lemma 26.** *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 14, we have*

$$\mathbb{E}_{\Pi(\eta_1,\ldots,\eta_n)}[p(x_1,\ldots,x_n)] = p(1-2\eta_1,\ldots,1-2\eta_n). \tag{38}$$

A.1 Dual Polynomials and Dual Block Composition

Bun et al. [11] 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 is a mild modification of [10, Proposition 54].

▷ Claim 27 (Modification of [10, 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.

$$\sum_{\omega_T(t)>0,\, t\ge k} |\omega_T(t)| \le \frac{1}{48 \cdot 4^k \sqrt{N} \log N}. \tag{39}$$

$$\sum_{\omega_T(t)<0,\, t<k} |\omega_T(t)| \le \left(\frac{1}{2} - \frac{2}{4^k}\right). \tag{40}$$

$$\|\omega_T\|_1 := \sum_{t=0}^{T} |\omega_T(t)| = 1. \tag{41}$$

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

$$\deg(q) < c_1 \sqrt{4^{-k} k^{-1} T N^{-1/(2k)} \log^{-1} N} \implies \sum_{t=0}^{T} \omega_T(t) q(t) = 0. \tag{42}$$

$$\text{For all } t \in [T], |\omega_T(t)| \le \frac{\sigma \exp(-\beta t)}{t^2} \qquad \text{for } \sigma = (2k)^k, \quad \beta = c_2 / \sqrt{4^k k T N^{1/(2k)} \log N}. \tag{43}$$

Sherstov [23] showed that dual block composition (see Definition 11) preserves ℓ_1-norm and that pure high degree is multiplicative (also see [16]). Bun and Thaler [13] observed that dual block composition is associative.

▶ **Lemma 28.** *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*

$$\|\theta \star \phi\|_1 = 1. \tag{44}$$

Multiplicativity of pure high degree:

$$\mathrm{phd}(\theta) > D, \mathrm{phd}(\phi) > d \implies \mathrm{phd}(\theta \star \phi) > Dd. \tag{45}$$

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

$$(\phi \star \theta) \star \psi = \phi \star (\theta \star \psi). \tag{46}$$

It was shown in [10] that for any dual polynomial Φ, and ψ as constructed in Claim 20, the dual block composed function $\Phi \star \psi$ satisfies a "strong dual decay" condition.[16]

▷ Claim 29 ([10, 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 $\psi : \{-1,1\}^N \to \mathbb{R}$ as defined in Claim 20. Then

$$\sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |(\Phi \star \psi)(x)| \leq (2NR)^{-2\Delta} \tag{47}$$

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

B Properties of Auxiliary Functions

It is easy to show that any multilinear polynomial $p : \mathbb{R}^n \to \mathbb{R}$ satisfies $\max_{y \in [-1,1]^n} |p(y)| \leq \|\hat{p}\|_1$. When applied to the function in Lemma 16, we obtain

▷ Claim 30. For p_η defined as in Lemma 16, $\max_{y \in [-1,1]^n} |p_\eta(y)| \leq \eta!\binom{n+\eta}{\eta}$.

We now state the setting for our next few claims.

Assumptions for Claim 31, Claim 32, Claim 33: 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 16, let $\alpha = \alpha_{f,\xi} : \{-1,1\}^m \to \mathbb{R}$ be as defined in Equation (10), and consider the distribution μ_ξ over $\{-1,1\}^{nm}$. 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 31.

$$\begin{aligned}&\zeta(\mathbf{1}^n)\mathbb{E}_{x \sim \mu_{\mathbf{1}^n}}[p_\eta(\alpha(x_1), \ldots, \alpha(x_n))\mathrm{OR}(f(x_1), \ldots, f(x_n))] \\ \geq & p_\eta(1-2\epsilon^+, \ldots, 1-2\epsilon^+)\left(\zeta(\mathbf{1}^n) - |\zeta(\mathbf{1}^n)|2A\right).\end{aligned} \tag{48}$$

[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 ψ as constructed in Claim 20.

▷ Claim 32.

$$\begin{aligned}&\sum_{z\neq\mathbf{1}^n}\zeta(z)\mathbb{E}_{\mu_z}[p_\eta(\alpha(x_1),\ldots,\alpha(x_n))\mathrm{OR}(f(x_1),\ldots,f(x_n))]\\ \geq& p_\eta(1-2\epsilon^+,\ldots,1-2\epsilon^+)\left(\sum_{z\neq\mathbf{1}^n}\zeta(z)\mathrm{OR}(z)-\left(2-2\frac{1-\epsilon}{1-\epsilon^+}(1-A)\right)\sum_{z\neq\mathbf{1}^n}|\zeta(z)|\right).\end{aligned} \tag{49}$$

Due to space constraints we do not prove Claim 31 and Claim 32 here, and refer the reader to the full version for these proofs. We now prove Claim 21 using Claim 31 and Claim 32.

Proof of Claim 21.

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

$$=\sum_{x\in\{-1,1\}^{mn}}\mathrm{OR}(f(x_1),\ldots,f(x_n))$$

$$\cdot\, 2^n\zeta\left(\mathrm{sgn}(\xi(x_1)),\ldots,\mathrm{sgn}(\xi(x_n))\right)p_\eta(\alpha(x_1),\ldots,\alpha(x_n))\prod_{i=1}^n|\xi(x_i)| \qquad \text{by Definition 11}$$

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

$$\left.\mathrm{OR}(f(x_1),\ldots,f(x_n))\prod_{i=1}^n|\xi(x_i)|\right)$$

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

by Definition 25 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 $\mathrm{phd}(\xi)\geq 1$

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

$$\left.+\sum_{z\neq\mathbf{1}^n}\zeta(z)\mathrm{OR}(z)-\left(2-2\frac{1-\epsilon}{1-\epsilon^+}(1-A)\right)\sum_{z\neq\mathbf{1}^n}|\zeta(z)|\right)$$

by Claim 31, 32 and $\mathrm{OR}(\mathbf{1}^n)=1$

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

since $\|\zeta\|_1=1$ and $\langle\zeta,\mathrm{OR}\rangle>\delta$

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

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

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

▷ Claim 33.

$$\left\|(\zeta\star\xi)(p_\eta\circ\alpha)\right\|_1=p_\eta(1-2\epsilon^+,\ldots,1-2\epsilon^+). \tag{50}$$

The proof of the claim follows along the lines as that of [22, Claim 6.2].

▷ Claim 34. 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 (10), and $p_\eta : \{-1,1\}^n \to \mathbb{R}$ defined in Lemma 16. Then

$$\mathrm{phd}((\Psi \star \Lambda) \cdot (p_\eta \circ \alpha)) > (\mathrm{phd}(\Psi) - \eta) \cdot \mathrm{phd}(\Lambda). \tag{51}$$

The proof follows along the same lines as that of [22, Equation (6.7)] and we omit it.

C Main Theorem

Recall from the proof of Theorem 19 in Section 4.1 that it remains to show $\langle \mathcal{W}, (\mathrm{OR}_R \circ \mathrm{THR}_N^k) \rangle > 7/33$ and $\mathrm{phd}(\mathcal{W}) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\log^{5/2} R} \cdot R^{\frac{3}{4}-\frac{1}{4k}}\right)$.

Remaining proof of Theorem 19. To justify Equation (31), we have

$$\begin{aligned}
\langle \mathcal{W}, \mathrm{OR}_R \circ \mathrm{THR}_N^k \rangle &= \frac{1}{\|\Gamma - \nu\|_1} \left(\langle \Gamma, \mathrm{OR}_R \circ \mathrm{THR}_N^k \rangle - \langle \nu, \mathrm{OR}_R \circ \mathrm{THR}_N^k \rangle \right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{by Equation (36)} \\
&\geq \frac{1}{\|\Gamma - \nu\|_1} \left(1/3 - \langle \nu, \mathrm{OR}_R \circ \mathrm{THR}_N^k \rangle \right) \qquad \text{by Claim 24} \\
&\geq \frac{1}{\|\Gamma - \nu\|_1} \{ 1/3 - \|\nu\|_1 \} \\
&\geq \frac{1}{\|\Gamma - \nu\|_1} \frac{7}{30} \qquad \text{by Equation (35)} \\
&\geq \frac{7}{33}. \qquad \text{since } \|\Gamma - \nu\|_1 \leq \tfrac{11}{10} \text{ by triangle inequality}
\end{aligned}$$

We have from Equation (36) that

$$\begin{aligned}
\mathrm{phd}(\mathcal{W}) &= \mathrm{phd}\left(\frac{\Gamma(x) - \nu(x)}{\|\Gamma - \nu\|_1} \right) && (52) \\
&= \mathrm{phd}(\Gamma(x) - \nu(x)) && (53) \\
&\geq \min\{\mathrm{phd}(\Gamma), \mathrm{phd}(\nu)\}. && (54)
\end{aligned}$$

From Equation (34) we have

$$\begin{aligned}
\mathrm{phd}(\nu) &\geq 2(\Delta - \sqrt{R}) - 1 && (55) \\
&= 2\left(\frac{c_2 R}{4 \ln^2 R} \sqrt{\frac{(2k)^k}{4^k k T N^{1/(2k)} \log N}} - \sqrt{R} \right) - 1 && \text{substituting the value of } \Delta \\
&\geq 2\left(\frac{c_2}{4} \cdot \frac{1}{\log^2 R \sqrt{\log N}} \cdot \left(\frac{k}{2}\right)^{k/2} \frac{1}{k^{1/2}} \cdot \frac{R^{3/4}}{N^{1/(4k)}} - \sqrt{R} \right) - 1 \\
&\qquad\qquad\qquad\qquad \text{taking } T = \sqrt{R} \text{ and } \ln R < \log R \\
&= 2\left(\frac{c_2}{4} \cdot \frac{1}{\log^2 R \sqrt{k \log R}} \cdot \left(\frac{k}{2}\right)^{k/2} \frac{1}{k^{1/2}} \cdot \frac{R^{3/4}}{20^{1/(4k)} 2^{1/8} k^{1/8} R^{1/(4k)}} - \sqrt{R} \right) - 1 \\
&\qquad \text{substituting the value of } N \text{ and using } k \log R > \log N \text{ for sufficiently large } R \\
&\geq 2\left(\frac{c_2}{2^{25/24}} \cdot \frac{1}{\log^2 R \cdot \sqrt{\log R}} \cdot \left(\frac{k}{2}\right)^{k/2} \frac{1}{k^{9/8} \cdot 20^{1/(4k)}} \cdot R^{3/4 - 1/(4k)} - \sqrt{R} \right) - 1 \\
&&& (56)
\end{aligned}$$

$$
\begin{aligned}
&\geq 2\left(\frac{c_2}{3} \cdot \frac{1}{\log^{5/2} R} \cdot \frac{1}{2^{9/8} \cdot 20^{1/(4k)}} \cdot R^{3/4-1/(4k)} - \sqrt{R}\right) - 1 \\
&\qquad\qquad\qquad\qquad \text{since } \left(\tfrac{k}{2}\right)^{k/2} \tfrac{1}{k^{9/8}} \geq \tfrac{1}{2^{9/8}} \text{ for all } k \geq 2 \\
&\geq \frac{c_2}{180} \cdot \frac{1}{\log^{5/2} R} \cdot R^{3/4-1/(4k)} - 1 \\
&\qquad \text{since } \tfrac{c_2}{3} \cdot \tfrac{1}{\log^{5/2} R} \cdot R^{3/4-1/(4k)} > 2\sqrt{R} \text{ for } k \geq 2, \text{ for sufficiently large } R \\
&= \Omega\left(\frac{1}{\log^{5/2} R} \cdot R^{3/4-1/(4k)}\right). \qquad (57)
\end{aligned}
$$

Therefore by Claim 24 and Equation (54), 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 (32) and finishing the proof. ◀