Electronic Colloquium on Computational Complexity, Report No. 172 (2025)

Restriction Trees for Sparsity and Applications

Arkadev Chattopadhyay * Yogesh Dahiya † Shachar Lovett ‡

Abstract

Exact and point-wise approximating representations of Boolean functions by real polynomials have been of great interest in the theory of computing. We focus on the study of sparsity of such representations. Our results include the following:

- We show that for every total Boolean function, its exact and approximate sparsity in the De Morgan basis, are polynomially related to each other in the log scale, ignoring poly-log(n) factors. This answers an open question posed by Knop, Lovett, McGuire and Yuan (STOC 2021). It builds on and is analogous to the seminal result of Nisan and Szegedy (Computational Complexity 1994) who proved the same for degree and approximate degree.
- We consider more powerful representations using *generalized monomials*, where each monomial is an indicator of a sub-cube. There are 3^n such monomials, where n is the number of variables. We prove that even for these representations, the sparsity and approximate sparsity of total Boolean functions remain polynomially related to each other in the log scale, ignoring poly-log(n) factors.
- We show that for every total Boolean function f, the log of its De Morgan sparsity characterizes upto polynomial loss and ingnoring poly-log(n) factors, the quantum and classical 2-party bounded-error communication complexity of $f \circ \mathrm{EQ}_4$, where EQ_4 is Equality of two 2-bit strings, one held by Alice and the other by Bob. As a consequence, we show that bounded-error quantum protocols cannot exhibit super-polynomial cost advantage over their classical counterparts, for computing such functions.

At the core of all our results, lies a novel characterization of non-sparse functions. This characterization is in terms of a combinatorial object that we call *max-degree restriction trees*. These objects locally certify high sparsity, in the same sense that block-sensitivity locally certifies degree.

*TIFR, Mumbai. Email: arkadev.c@tifr.res.in. Supported by the Department of Atomic Energy, Govmt. of India, under project #RTI4001 and by a Google India Faculty Award.

†TIFR, Mumbai. Email: yogesh.dahiya@tifr.res.in. Supported by the Department of Atomic Energy, Govmt. of India, under project #RTI4001 and the R. Narasimhan postdoctoral fellowship.

‡UC San Diego. Email: slovett@ucsd.edu. Supported by Simons Investigator Award #929894 and NSF award CCF-2425349.

ISSN 1433-8092

Contents

1 Introduction

Polynomial representations of Boolean functions have been invaluable in theoretical computer science and discrete mathematics. While the representation could use any field, in this work we consider only polynomials over the reals. Two bases are particularly prominent.

The first arises naturally by viewing the domain of Boolean functions as $\{0,1\}^n$, and hence, every multilinear monomial just represents the Boolean AND of a subset of variables. This is known as the *De Morgan basis*. The other basis comes about by viewing the domain as $\{1,-1\}^n$ which is a simple linear transformation of $\{0,1\}^n$ that maps $0 \mapsto 1$ and $1 \mapsto -1$. In this basis, called the *Fourier basis*, each monomial represents the Boolean parity of a subset of variables.

Two natural complexity measures show up in either basis: the degree and sparsity of the representation. As every Boolean function has a unique representation in either basis, it is usually quite straightforward to determine the degree and sparsity of the unique representation for a function f, which we denote by $\deg(f)$ and $\mathrm{spar}(f)$ in the De Morgan basis, and by $\deg^{\oplus}(f)$, and $\mathrm{spar}^{\oplus}(f)$ in the Fourier basis. The linear invertible mapping from one basis to the other ensures that for every f, $\deg(f) = \deg^{\oplus}(f)$. But sparsity can be very sensitive to the basis chosen. For example, the n-bit AND function has sparsity 1 in the De Morgan basis and 2^n in the Fourier basis; and the n-bit PARITY function has sparsity 1 in the Fourier basis and 2^n in the De Morgan basis. We understand reasonably satisfactorily exact polynomial representations of Boolean functions. However, when we turn to approximations, the picture becomes subtler.

Classical approximation theory deals with polynomials that point-wise approximate functions. In an influential work, Nisan and Szegedy [NS94] introduced this notion to the study of Boolean functions. In particular, they defined the complexity measure of approximate degree of a Boolean function f, denoted by $\widetilde{\deg}(f)$, to be the smallest degree needed by a polynomial to point-wise approximate f to within a constant distance that is, by default, taken to be $1/3$. Observe that the same reasoning as applied above for exact degree implies that approximate degree of a function is also a measure that is independent of the basis. The notion of approximate degree has had tremendous impact in computer science as it is related to many other complexity measures including the randomized and quantum query complexity of f [BdW02, BBC+01, AS04, BKT20, BT22] and the quantum and classical communication complexity of appropriately lifted functions [BdW01, Raz03, SZ09, She11, CA08, LS09b, BH09, She14]. It has also found applications in learning theory [KS04, KKMS08], differential privacy [TUV12, CTUW14], secret sharing [BIVW16, BMTW19], and many other areas. Unlike exact degree, getting tight bounds on approximate degree often turns out to be challenging. However, Nisan and Szegedy proved a remarkable structural result that for every total Boolean function f, the approximate and exact degree of f are polynomially related to each other. One of the striking applications of this result is the polynomial equivalence of quantum and classical query models for total Boolean functions, first derived by Beals et al. [BBC+01]. In a much more recent work, building upon Huang's breakthrough proof [Hua19] of the sensitivity conjecture, Aaronson, Ben-David, Kothari, Rao and Tal [ABDK+21] finally gave a tight relationship between the two measures by showing that $\deg(f) = O(\widetilde{\deg}(f))^2$. The tightness is witnessed by the Boolean AND and OR functions.

Given Nisan and Szegedy's result, one naturally wonders if approximation could reduce *sparsity* needed for total functions. In the Fourier basis, it is known that approximation can reduce sparsity exponentially. For instance, the Fourier sparsity of the n-bit AND function is 2^n. However, it can be shown that its Fourier approximate-sparsity is $O(n^2)$ (implicit in Bruck and Smolesnky [BS90, appendix] and explicit in [CMS20, Lemma 2.8]). Surprisingly, the question if approximation helps significantly in the De Morgan basis remained unaddressed. Recently, Knop et al. [KLMY21a] conjectured that in the De Morgan basis, approximation should not significantly reduce sparsity for any total Boolean function. Our main result, stated below, confirms this conjecture. We denote by $\widetilde{\mathrm{spar}}(f)$, the approximate-sparsity of f in the De Morgan basis.

Theorem 1.1. *For every total Boolean function $f : \{0,1\}^n \to \{0,1\}$, we have*

$$\log(spar(f)) = O\big(\log(\widetilde{spar}(f))^2 \cdot \log n\big).$$

One way to interpret the theorem is that the logarithm of the sparsity plays the role of degree when comparing exact and approximate representations. Before we continue, let us remark on the tightness of this result.

Remark 1.2. *The n-bit* OR *function has sparsity* $2^n - 1$ *and approximate sparsity* $2^{O(\sqrt{n}\log n)}$, *showing that exponential gaps may exist between the two measures in the absolute scale. It is thus necessary to consider the log scale as done in Theorem 1.1 for seeking polynomial relationship between the two measures. The example of* OR *also demonstrates the tightness of our bound up to poly-logarithmic factors. Finally, the appearance of the ambient dimension n in our result is unavoidable. Consider the function* $\text{THR}^n_{n-1} : \{0,1\}^n \to \{0,1\}$ *defined by*

$$\text{THR}^n_{n-1}(x) = 1 \quad \textit{if and only if} \quad |x| \geqslant n-1,$$

namely, the function evaluates to 1 if the input has at most one zero. It's simple to verify that its exact sparsity is $n+1$, *and we show in section 4.3 that its approximate sparsity is* $O(\log n)$, *implying that an additive* $O(\log n)$ *or multiplicative* $O\left(\frac{\log n}{\log\log n}\right)$ *factor is necessary in Theorem 1.1.*

One of the motivations of Nisan and Szegedy to study approximate degree was to relate this measure with decision tree complexity. Let $\mathrm{D}^{dt}(f)$ and $\mathrm{R}^{dt}(f)$ denote the deterministic and randomized bounded-error decision tree complexities of f respectively. Their result, along with the recent improvement of [ABDK+21] yields the following relationship.

Theorem 1.3 (Nisan-Szegedy [NS94, Theorem 1.5] + Aaronson et al. [ABDK+21])**.** *For every total Boolean function f, the following holds:*

$$\widetilde{deg}(f) \leqslant R^{dt}(f) \leqslant D^{dt}(f) \leqslant O(\widetilde{deg}(f)^4).$$

Just as $\widetilde{\deg}(f)$ lower bounds $\mathrm{R}^{dt}(f)$, it is straightforward to verify that $\log(\widetilde{\mathrm{spar}}(f))$, up to an additive $\log n$ term, lower bounds the randomized AND-decision tree (ADT) complexity of f. In an ADT, each internal node queries the AND of a subset of variables. ADT's have connections to combinatorial group testing algorithms and have also been the subject of several recent works [Nis21, BN21, KLMY21b, CDM+23]. We denote the deterministic and randomized ADT complexities of f by $\mathrm{D}^{\wedge dt}(f)$ and $\mathrm{R}^{\wedge dt}(f)$ respectively.

Combining our main result with the recent result of Knop et al. [KLMY21b] yields the following ADT analog of Theorem 1.3.

Theorem 1.4. *For every total Boolean function* $f : \{0,1\}^n \to \{0,1\}$, *the following holds:*

$$\Omega(\log(\widetilde{spar}(f)) - \log n) \overset{(1)}{=} R^{\wedge dt}(f) \leqslant D^{\wedge dt}(f) \overset{(2)}{=} O\left((\log \widetilde{spar}(f))^6 \cdot \log n\right).$$

It is worth noting that one could go one step further in the chain of inequalities to show that $\mathrm{D}^{\wedge dt}(f)$ is upper bounded by $O(\mathrm{R}^{\wedge dt}(f)^6)$ upto polylog factors, thus concluding that randomization doesn't yield more than polynomial savings over the cost of deterministic ADT algorithms. Such a conclusion, in fact with a better polynomial bound, was first derived recently by Chattopadhyay, Dahiya, Mande, Radhakrishnan and Sanyal [CDM+23]. But our current technique is quite different and the previous result could not give an upper bound on ADT complexity in terms of approximate sparsity as we do here.

Apart from degree and sparsity, there is a third complexity measure that has been well investigated in the Fourier basis. This is the Fourier ℓ_1 norm, also called the spectral norm of a Boolean function f. Denoted by $\|\hat{f}\|_1$, it is defined as the sum of the magnitude of the Fourier coefficients of f. It appeared in the context of additive combinatorics [GS08], communication complexity of XOR functions [CM17a, CMS20, CHH+25] and analysis of Boolean functions [BS90, AFH12a, CM17b]. One naturally defines its ε-approximate version, denoted by $\|\hat{f}\|_{1,\varepsilon}$, to be the amount of Fourier ℓ_1 mass needed by any real-valued function g to point-wise approximate f within ε. Can approximation reduce significantly the needed ℓ_1 mass? Very recently, Cheung, Hatami, Hosseini, Nikolov, Pitassi and Shirley [CHH+25], constructed a Boolean function f such that $\log(\|\hat{f}\|_{1,1/3})$ is exponentially smaller than $\log(\|\hat{f}\|_1)$, which implies that the Fourier basis yields exponential advantage to approximation even with respect to the spectral norm.

In contrast, the proof method that we develop to establish Theorem 1.1 shows that approximation does not significantly reduce even the ℓ_1 mass of a total Boolean function in the De Morgan basis. More precisely, let $\mathrm{wt}(f)$ and $\widetilde{\mathrm{wt}}_\varepsilon(f)$ represent the exact and ε-approximate ℓ_1 norm of f in the De Morgan basis (we write $\widetilde{\mathrm{wt}}(f) := \widetilde{\mathrm{wt}}_{1/3}(f)$ when $\varepsilon = 1/3$).

Theorem 1.5. *For every total Boolean function* $f : \{0,1\}^n \to \{0,1\}$, *we have*

$$\log wt(f) = O\left((\log \widetilde{wt}(f))^2 \cdot \log n\right).$$

General representations. A natural question emerges from our results. Let $\mathcal{F}$ be a family of elementary real-valued functions defined over the n-ary Boolean cube B_n, such that $\mathcal{F}$ spans the vector space $\mathbb{R}^{B_n}$ of all real-valued functions. The *sparsity* (resp., *weight*) of a (Boolean) function f wrt $\mathcal{F}$ is defined as the minimum integer (resp., non-negative real number) $k \geqslant 0$ such that f can be written as a linear combination of at most k functions from $\mathcal{F}$ (resp., with total absolute coefficient sum at most k)[1]. Denote these complexity measures by $\mathrm{spar}_{\mathcal{F}}(f)$ and $\mathrm{wt}_{\mathcal{F}}(f)$ respectively. For example, when $\mathcal{F}$ is the family of all AND functions, these measures correspond to the De Morgan sparsity and ℓ_1 norm of f, and when $\mathcal{F}$ corresponds to all parities, these correspond to the Fourier sparsity and the spectral norm of f. Likewise, one defines the approximate sparsity and weight of f with respect to $\mathcal{F}$, denoting them by $\widetilde{\mathrm{spar}}_{\mathcal{F}}(f)$ and $\widetilde{\mathrm{wt}}_{\mathcal{F}}(f)$ respectively.

Question 1.6. *What properties of $\mathcal{F}$ ensure that approximation doesn't help reduce sparsity or the weight of a Boolean function, i.e. do there exist constants α and β such that for all Boolean functions f, $\log(spar_{\mathcal{F}}(f)) = O((\log(\widetilde{spar}_{\mathcal{F}}(f))^{\alpha})$ and/or $\log(wt_{\mathcal{F}}(f)) = O((\log(\widetilde{wt}_{\mathcal{F}}(f))^{\beta})$?*

This question is quite broad. For instance, if one views the input domain of the functions as the set of $m \times m$ Boolean matrices, and $\mathcal{F}$ be the set of all rank one matrices, then Question 1.6 specializes to asking if log of the rank of a Boolean matrix is always at most a fixed polynomial of the log of its approximate-rank. In general, this is false. For example, the identity matrix has rank m but its approximate rank is $(\log m)^{O(1)}$. However, it is unknown for special classes of Boolean matrices like those that are the truth table of AND-functions (i.e., functions composed with 2-bit AND gadgets). Understanding the power of approximation for such special classes of matrices is of significant interest, given its connection to quantum communication complexity. We return to this aspect later in the section.

Summarizing what we have seen, if $\mathcal{F}$ is the set of all parities, i.e. the Fourier monomials, then approximation can significantly help, and reduce sparsity exponentially. We showed that if $\mathcal{F}$ is the set of all monotone Boolean AND functions, i.e. the De Morgan basis, then approximations do not help, and reduce sparsity by at most a polynomial factor on the log scale. In fact our main result gives us slightly more: let $[n]$ be partitioned into two sets, the set of positive literals denoted by $\mathcal{P}$ and the set of negated literals $\mathcal{N}$. Each such partition defines a *shifted* De Morgan basis, where a shifted monomial is given by a pair of sets $P \subseteq \mathcal{P}$ and $N \subseteq \mathcal{N}$, and corresponds to the Boolean function $M_{P,N} := \prod_{i \in P} x_i \prod_{j \in N}(1 - x_j)$. Observe that while OR has full De Morgan sparsity, it has sparsity just 2 in the fully shifted De Morgan basis, i.e., the basis that corresponds to $\mathcal{P} = \varnothing$ ans $\mathcal{N} = [n]$. There are 2^n such shifted bases, and it is straightforward to verify that our results—Theorem 1.1 and Theorem 1.5—imply that, in each shifted basis, the approximate sparsity and approximate ℓ_1-norm are polynomially related to the exact sparsity and exact ℓ_1-norm.

Generalized Monomials. A natural generalization of the case when $\mathcal{F}$ is just a shifted De Morgan basis, is the case when we populate the set $\mathcal{F}$ with all shifted monomials. More precisely, consider $\mathcal{F} := \{M_{P,N} : P, N \subseteq [n], P \cap N = \varnothing\}$, where each $M_{P,N} := \prod_{i \in P} x_i \prod_{j \in N}(1 - x_j)$ is called a *generalized monomial*. Observe that any shifted De Morgan basis is a strict subset of $\mathcal{F}$, the set of generalized monomials, whose size is 3^n. As expected, generalized monomials can significantly reduce sparsity compared to any shifted De Morgan basis. For instance, consider the following function that mixes two shifted OR's by a monotone addressing scheme: let $f_{\mathrm{mixed}} : \{0,1\}^2 \times \{0,1\}^n \to \{0,1\}$, where $f_{\mathrm{mixed}}(x, y)$ outputs 0 if $x = 00$, outputs 1 if $x = 11$, computes the Boolean OR of y if $x = 10$, and computes the Boolean AND of y if $x = 01$. It is easy to verify that f_{mixed} can be represented as a generalized polynomial of sparsity $O(1)$, while the sparsity of f_{mixed} in any shifted De Morgan basis is $2^{\Omega(n)}$.

It is thus natural to consider the power of approximations when $\mathcal{F}$ is the set of all generalized monomials. This set can also be viewed to be the set of indicator functions of subcubes of the Boolean cube. For this $\mathcal{F}$, we provide a complete answer to Question 1.6 below.

For the family $\mathcal{F}$ of all generalized monomials, let $\mathrm{gspar}(f)$, $\widetilde{\mathrm{gspar}}(f)$, $\mathrm{gwt}(f)$, and $\widetilde{\mathrm{gwt}}(f)$ denote the sparsity, approximate sparsity, ℓ_1-norm, and approximate ℓ_1-norm of f, respectively.

Theorem 1.7. *For every total Boolean function $f : \{0,1\}^n \to \{0,1\}$, we have*

$$\max\big\{\log gspar(f), \log gwt(f)\big\} = O\Big(\big(\min\{\log \widetilde{gspar}(f), \log \widetilde{gwt}(f)\}\big)^2 \cdot \log n\Big).$$

[1]Note that there may be more than one way of doing that.

The measure $\mathrm{gspar}(f)$ is also connected to the standard query complexity model. Recall that the *size* of a decision tree is the number of its leaves. The deterministic (resp. randomized) decision tree size complexity of a function f is the minimum size of a deterministic (resp. randomized) decision tree computing f. The quantities $\mathrm{gspar}(f)$ and $\widetilde{\mathrm{gspar}}(f)$ (up to a multiplicative $O(n)$ factor) provide natural lower bounds on the deterministic and randomized decision tree sizes, respectively, since any decision tree of size s computing f can be transformed into a generalized polynomial for f with sparsity at most s. Furthermore, for *monotone* functions, as observed in [CDL25], the deterministic and randomized decision tree size complexities are characterized—up to polynomial loss and poly-log(n) factors—by $\mathrm{gspar}(f)$ on the logarithmic scale. This correspondence, however, fails for general functions. Notably, the so-called *Sink function* on n bits satisfies $\mathrm{gspar}(f) = O(\sqrt{n})$ but $\mathrm{RSize}^{dt}(f) = 2^{\Omega(\sqrt{n})}$ [CMS20].

Applications to Quantum-Classical Equivalence. A major open question at the interface of quantum computing and communication complexity is whether there exists a *total* Boolean function for which the 2-party quantum bounded-error communication complexity is exponentially smaller than its classical randomized counterpart. This question remains open despite intensive research efforts. For example, Shi and Zhu [SZ09] formulated what they call the *Log-Equivalence Conjecture* (LEC) which asserts that such a total function cannot exist. One of their stated intuitions for believing in this conjecture was the approximate analogue of the classical Log-Rank Conjecture, now known as the *Log-Approximate-Rank Conjecture* (LARC), a term first coined by Lee and Shraibman [LS$^+$09c]. The LARC asserts that there exists a fixed polynomial p such that for every total Boolean function F, the bounded-error randomized communication complexity of F is asymptotically upper-bounded by $p(\log \widetilde{\mathrm{rank}}(M_F))$, where $\widetilde{\mathrm{rank}}(M_F)$ denotes the approximate rank of the communication matrix of F. As it is well known [LS$^+$09c] that the bounded-error quantum communication complexity of F (even with prior entanglement) is lower bounded by $\log(\widetilde{\mathrm{rank}}(M_F))$, upto an additive $O(\log n)$ term, the LARC essentially implies the LEC (in the communication complexity regime of $\Omega(\log n)$). However, a surprising but natural counter-example to the LARC was obtained in the more recent work of Chattopadhyay, Mande and Sherif [CMS20]. On the other hand, two later independent works [ABT19, SdW] showed that this counter-example has also large quantum bounded-error communication complexity and therefore doesn't falsify the LEC. On the whole, the correctness of the LEC remains far from clear. In fact, it's not clear if the known LARC counter-example is an oddity: are there other quite different counter-examples to the LARC?

In a breakthrough more than two decades ago, Razborov [Raz03] showed that the quantum bounded-error communication complexity of the promised Set-Disjointness problem is $\Omega(\sqrt{n})$, when the universe size is n. In the promised Set-Disjointness problem, Alice and Bob are promised that either their sets are disjoint or they intersect exactly at one element of the universe. Razborov's proof of this result implied two things. First, as observed by Razborov himself, for every symmetric function f, the quantum bounded-error communication complexity of the composed function $f \circ \textsc{And}_2$ is polynomially related to even its classical deterministic complexity. Second, consider any gadget $g : \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$, whose truth table contains both the And and Or functions as submatrices. Razborov's result implied that for *every total* Boolean function f, the bounded-error quantum communication complexity of $f \circ g$ is $\Omega(\sqrt{\mathrm{bs}(f)})$, where $\mathrm{bs}(f)$ is the block-sensitivity of f. This follows because one can embed a promised Set-Disjointness instance of size $\mathrm{bs}(f)$ within $f \circ g$ when g satisfies the above property. Combining this with Nisan's classical result [Nis91] that $\mathrm{D}^{dt}(f) = O(\mathrm{bs}^3(f))$ for every total Boolean function f, one obtains a quantum–classical equivalence for such composed functions when the gadget size b is constant. This is summed up below, and was also explicitly noted by Sherstov [She10]. We denote by $\mathrm{D}^{cc}(F)$, $\mathrm{R}^{cc}(F)$, and $\mathrm{Q}^{cc}(F)$ the deterministic, randomized, and quantum (with prior entanglement) two-party communication complexities of a function F, respectively.

Theorem 1.8 (Razborov [Raz03] + Nisan [Nis91]). *Let $g : \{0,1\}^b \times \{0,1\}^b \to \{0,1\}$ be any gadget whose truth table embeds both the 2-bit* And *and* Or *communication gadgets as submatrices. Then, for every total Boolean function f, the following chain of inequalities holds:*

$$\Omega\big(\sqrt{bs(f)}\big) = Q^{cc}(f \circ g) = O(D^{cc}(f \circ g)) = O\big(bs^3(f) \cdot b\big) = O_b\big(\big(Q^{cc}(f \circ g)\big)^6\big).$$

Several well known gadgets used in communication complexity, like Inner-Product on 4 bits, Indexing with one address-bit (and therefore pattern matrices) all embed both the 2-bit And and Or functions, simultaneously. In particular, Theorem 1.8 has so far provided the largest class of functions lifted by *constant-size* (or even small, i.e. poly-log(n) size) gadgets for which LEC is known to be true (upto poly-log(n) losses for small gadgets). Indeed, confirming the LEC for all lifted functions employing constant-size gadgets would constitute major progress. The two simplest and perhaps the most natural constant-size gadgets are the 2-bit AND and XOR functions. It is not known if the LEC holds for either of them. In particular, we know that the LARC is false for XOR lifted functions as the counter-example from [CMS20] is of the form $f \circ \text{XOR}$, where f is the so called Sink function.

Given this situation with 2-bit gadgets, we explore a natural question: what happens for 4-bit gadgets that themselves arise as compositions of the 2-bit AND and XOR functions. There are two such gadgets: first, $\text{XOR}_2 \circ \text{AND}_2$ which is also Inner-Product on 4 bits. This, as we saw before, by virtue of embedding both 2-bit AND and OR, is completely handled by Theorem 1.8 yielding polynomial equivalence of bounded-error quantum and even classical deterministic protocols. The second gadget is $\text{AND}_2 \circ \text{XOR}_2$, which is essentially Equality on four bits, denoted henceforth by EQ_4. It is simple to verify that EQ_4 embeds the 2-bit AND gadget but not the 2-bit OR, placing functions of the form $f \circ \text{EQ}_4$ outside the scope of Theorem 1.8. Indeed, the bounded-error randomized communication complexity of $\text{AND} \circ \text{EQ}_4$—which is simply the Equality function on a larger number of bits—is only $O(1)$, even though AND itself has full block sensitivity. As a result, the gadget EQ_4 may be regarded as *weak*, since EQ_4-lifted functions exhibit an *exponential* separation between randomized and deterministic communication complexities. This stands in sharp contrast to the class of gadgets covered by Theorem 1.8, where even deterministic and quantum communication complexities remain polynomially related. To analyze EQ_4-lifted functions, one must therefore rely on a complexity measure for f that differs fundamentally from block sensitivity and approximate degree, both of which are polynomially related for all total functions.

In this work, we show that (log of) the De Morgan sparsity of f is such a complexity measure.

Theorem 1.9. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. Then,*

1. $\log spar(f) = O\Big((\log \widetilde{rank}(f \circ \text{EQ}_4))^2 \cdot \log n\Big)$.
2. $\log spar(f) = O\big(Q^{cc}(f \circ \text{EQ}_4)^2 \cdot \log n\big)$.

The above shows that lower bound on the log of the De Morgan sparsity of f lifts with a quadratic loss to both the log-approximate-rank and the quantum, bounded-error communication complexity of $f \circ \text{EQ}_4$. Combining this result with the recent result of Knop et al [KLMY21b] who showed that the deterministic AND-decision tree complexity of f is upper bounded by a polynomial of the log of the De Morgan sparsity of f, we establish the LARC, and therefore, the polynomial equivalence of bounded-error quantum and classical randomized communication complexity of $f \circ \text{EQ}_4$ (both up to poly-log n terms).

Theorem 1.10. *Let $f : \{0,1\}^n \to \{0,1\}$ be any total Boolean function. Then,*

$$R^{cc}(f \circ \text{EQ}_4) \leqslant D^{\wedge dt}(f) \log D^{\wedge dt}(f) = O\big((\log spar(f))^5 (\log n)^2\big) = O\Big((\log \widetilde{rank}(f \circ \text{EQ}_4))^{10} (\log n)^7\Big),$$

$$R^{cc}(f \circ \text{EQ}_4) = O\big((\log spar(f))^5 (\log n)^2\big) = O\big(Q^{cc}(f \circ \text{EQ}_4)^{10} (\log n)^7\big).$$

We make serveral remarks. First, the LEC for XOR-lifted functions directly implies the LEC for EQ_4-lifted functions. Hence, our proof of the LEC for the latter constitutes a necessary step toward resolving the XOR-gadget case. Second, to the best of our knowledge, Theorem 1.10 yields the first polynomial equivalence of the powers of quantum and classical bounded-error protocols for a natural class of functions for which quantum bounded-error protocols' power provably does not collapse all the way to those of deterministic protocols. Observe that one of the key ingredients in obtaining the equivalence for this class is the lifting of De Morgan sparsity established in Theorem 1.9 to log-approximate-rank and quantum communication complexity. It is plausible, as suggested by Buhrman and de Wolf [BdW01], that such a lifting takes place for even the simpler gadget of 2-bit AND which would then resolve the LEC for AND functions, a longstanding open problem. However, our result seems to be the first that provably lifts sparsity (either De Morgan or Fourier) to any form of bounded-error (classical or quantum) communication complexity for a class of functions where (log of) sparsity also yields an upper bound. We therefore view our result as a natural step toward resolving the bounded-error quantum–classical equivalence for AND functions as well.

1.1 Main Technical Tool - Restriction Trees

Nisan and Szegedy [NS94] showed that both degree and approximate degree, upto a polynomial loss, can be certified by the combinatorial measure of *block-sensitivity* that had been introduced earlier by Nisan [Nis91]. This combinatorial characterization was key to showing the structural result that approximate degree of any total Boolean function cannot be super-polynomially lower than its exact degree. In this work, our main technical contribution is to find an analogously effective combinatorial characterization of not just exact De Morgan sparsity of a Boolean function but even its point-wise approximate version.

To get an intuition about what this combinatorial structure could be, we start with some simple examples of functions of large exact sparsity for which we can prove lower bounds on their approximate sparsity. Lower bounds on approximate sparsity were known for specific functions such as OR_n and PARITY_n (these are folklore results), typically established using random restrictions and approximate degree lower bounds. However, these results apply only to specific functions or restricted classes of functions. The general idea is to apply a random restriction ρ, which selects a random subset of variables and fixes each to 0, with the goal of eliminating high-degree monomials from a candidate sparse polynomial P that approximates f, such that $f|_\rho$ still has large approximate degree while $P|_\rho$ has degree that is too small, yielding a contradiction.

This can be illustrated by considering the n-bit OR function. Let P be any sparse polynomial approximating OR_n. Consider a random restriction ρ that, independently for each of the n variables, fixes it to 0 with probability 1/2 and leaves it free with probability 1/2. With high probability, at least $n/3$ variables are left free. On the other hand, any monomial of degree larger than $\sqrt{n}$ survives (i.e., none of its variables are set to 0) with probability at most $2^{-\sqrt{n}}$. If the number of monomials in P is s, then the probability that $P|_\rho$ contains a monomial of degree larger than $\sqrt{n}$ is at most $s \cdot 2^{-\sqrt{n}}$, which is less than 1/2 if $s < 2^{\sqrt{n}}/2$. With high probability, $(\text{OR}_n)|_\rho$ is an r-bit OR function with $r \geqslant n/3$. Since $\text{OR}_n|_\rho$ is still approximated by $P|_\rho$ (for every ρ), and recalling that the approximate degree of OR_r is $\Omega(\sqrt{r})$, we conclude that $s = 2^{\Omega(\sqrt{n})}$.

While this works for the OR function, there are functions which are very different from OR and yet have large sparsity in De Morgan basis. An example of that is $\text{AND}_n \circ \text{OR}_2$, where the bottom ORs are 2-bit functions. It is simple to verify that this function has sparsity $2^{\Omega(n)}$. But there is no way to induce a large OR in this function. If one tried to apply a random restriction like the one that worked for OR, one concludes easily that it won't work as with high probability one of the bottom OR_2 will have both its input variables fixed to 0, thereby killing the entire function. One way to fix this is to consider a slightly more careful restriction. For each of the bottom ORs, one selects one of its two input variables at random and fixes it to 0, leaving the other variable free. It is not hard to show that in this case the restricted function is always the AND over the remaining n free variables, and if the approximating polynomial for the $\text{AND}_n \circ \text{OR}_2$ had sparsity $2^{o(\sqrt{n})}$, then with nonzero probability, the restricted polynomial would give an $o(\sqrt{n})$-degree approximation to AND_n, contradicting known lower bounds. The important thing to note here is that our random restriction is no longer done independently for each variable, as the restrictions on the two variables in each OR_2 block are correlated.

While the random restrictions are somewhat different for OR_n and $\text{AND}_n \circ \text{OR}_2$, one common feature seems to be the following: an iterative/adaptive process in which, at each step, a variable is selected and either left free or fixed to 0. In some cases, the choice of fixing a variable to 0 (or, more generally, to 1) may be forced by earlier assignments. This naturally suggests to depict the variety of restriction choices as a binary tree that we call a *restriction tree*. In such a tree, each node is labeled with a variable that is either going to be fixed to 0 or is going to be left free. These two possible decisions are represented by two outgoing edges from the node, one labeled with 0, the other labeled with $*$, much like in a decision tree. Note that the label of the two child nodes could be different, just as in a decision tree, depicting the adaptivity of the restriction process. Every root to leaf path of this tree describes a specific restriction, where edges labeled $*$ identify the free variables, while edges labeled 0 denote variables fixed to 0. Additionally, the leaf contains assignment to all remaining variables, some of which were forced to prevent our function from simplifying. Thus, each leaf represents a restriction of the function, which has not 'simplified' the function by too much. A sufficient way of ensuring that and something that we were doing in the two examples above, is to say that the exact degree of the restricted function is maximum possible, i.e. equal to the number of free variables. We call such a restriction tree to be a *max-degree restriction tree*. Obviously, the deeper the tree, the richer and harder the original function is!

Let us express the structure of the two example cases in terms of restriction trees. Consider OR_n

on $x_1, \ldots, x_n$. A max-degree restriction tree of depth n for $\textsc{Or}_n$ is constructed as follows. The tree is a full binary tree of depth n, where each node at level i (root at level 1) is labeled by x_i, and its two outgoing edges are labeled 0 and $*$. For a leaf v, the restriction ρ_v labeled at that leaf assigns each variable according to the edge labels on the path from the root to v: variables following an edge labeled 0 are set to 0, while those following edge labeled $*$ remain free. For every leaf v, the restricted function $\textsc{Or}_n|_{\rho_v}$ computes $\textsc{Or}$ on the $*$-variables of ρ_v, making this a max-degree restriction tree for $\textsc{Or}_n$.

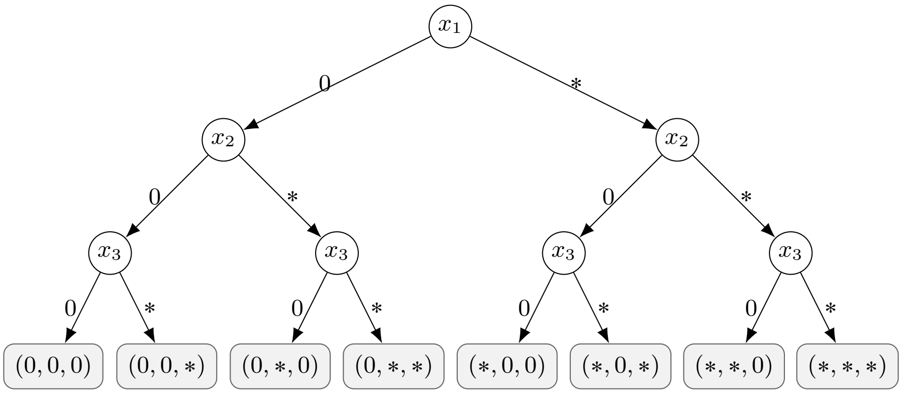

Figure 1: A one-sided max-degree restriction tree for $\textsc{Or}_3$. Leaves show the restriction ρ_v on (x_1, x_2, x_3).

It is not hard to phrase the set of restrictions that we considered for the $\textsc{And}_n \circ \textsc{Or}_2$ function in this way as well. And our *random* restrictions could just be described as taking a uniformly random walk on the restriction tree starting from the root and outputting the restriction associated with the reached leaf. Note that, so far, our restriction tree has been binary and at each node we only fixed variables to 0. Such trees will be called *one-sided* max-degree restriction trees. A special case of our central structural result is the following:

Lemma 1.11 (Special Case of First Structural Result)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function of De Morgan sparsity s. Then f admits a one-sided max-degree restriction tree of depth $\Omega(\log s/\log n)$.*

Conversely, we will show that the existence of a deep one-sided max-degree restriction tree implies not only large exact sparsity but also large approximate sparsity; that is, such trees serve as certificates for both measures.

When dealing with generalized sparsity, however, we'd need the flexibility of deciding to fix variables to either 0 or 1. This gives rise naturally to ternary restriction trees, that we call *two-sided* max-degree restriction trees. It turns out that the function

$$(\textsc{And}_n \circ \textsc{Or}_2)(x_1, \ldots, x_n, y_1, \ldots, y_n) = \textsc{And}_n(\textsc{Or}_2(x_1, y_1), \ldots, \textsc{Or}_2(x_n, y_n)).$$

admits a depth n two-sided max-degree restriction tree that can be constructed as follows: The tree is a full ternary tree of depth n, where each node at level i (root at level 1) is labeled by x_i and has edges labeled 0, 1, and $*$. For a leaf v, the restriction ρ_v labeled at that leaf is defined as follows: ρ_v assigns the x-variables according to the edge labels along the path from the root to v; variables following an edge labeled 0 are set to 0, those labeled 1 to 1, and those labeled $*$ remain free. The assignment of y-variables in ρ_v depends on the corresponding x-assignments:

$$\rho_v(y_i) = \begin{cases} 1, & \text{if } \rho_v(x_i) = 0, \\ 0, & \text{otherwise.} \end{cases}$$

It is immediate that this is a two-sided restriction tree. Moreover, for every leaf v, the restricted function $(\textsc{And}_n \circ \textsc{Or}_2)|_{\rho_v}$ computes $\textsc{And}$ on the $*$-variables of ρ_v, making it a max-degree restriction tree.

It is not hard to verify, given this two-sided max-degree restriction tree for $\textsc{And}_n \circ \textsc{Or}_2$, that even its approximate generalized sparsity is $2^{\Omega(\sqrt{n})}$. This shows that such trees can certify large (approximate) generalized sparsity. Remarkably, our second structural result shows that such certificates always exist for generalized sparsity.

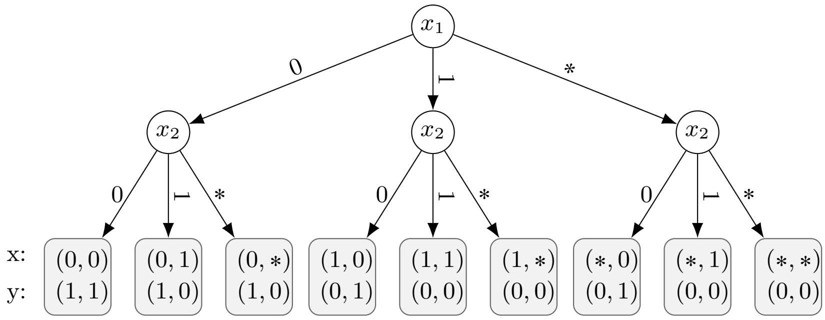

Figure 2: Two-sided max-degree restriction tree for $\text{AND}_2 \circ \text{OR}_2$. Leaves show restrictions in block form: top line lists (x_1, x_2), bottom line lists (y_1, y_2).

Lemma 1.12 (Special Case of Second Structural Result)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function of generalized sparsity s. Then f admits a two-sided max-degree restriction tree of depth $\Omega(\log s/\log n)$.*

Conversely, we will show that the existence of a deep two-sided max-degree restriction tree implies large approximate generalized sparsity as well, thereby serving as a certificate for both generalized and approximate sparsity.

The set of De Morgan monomials forms a basis of the vector space $\mathbb{R}^{\{0,1\}^n}$. Indeed this fact could be used to design local certificates of high sparsity, exploiting linear algebra. For instance, the rank of the communication matrix M of $f \circ \text{AND}_2$ is exactly s where $s = \text{spar}(f)$. Hence, the locality of rank can be exploited to certify De Morgan sparsity of f by exposing only the entries of an $s \times s$ submatrix of M. This certificate doesn't, however, seem useful to certify the approximate-sparsity as well. This is what our max-degree restriction trees, guaranteed to exist by Lemma 1.11, achieve. But, a generalized representation is not unique and the set of generalized monomials doesn't seem to have nice linear algebraic structure as it contains a union of several different bases. Hence, it seems far less clear that local certificates for generalized sparsity ought to exist. Indeed, we do not know of any other local certificates, algebraic or combinatorial, than what our Lemma 1.12 yields.

It turns out that a convenient point of view on both De Morgan sparsity and generalized sparsity is by viewing these as complexity measures with some nice closure properties. These are abstracted in Definition 3.5 as *one-sided nice measures* and in Definition 3.16 as *two-sided nice measures*. These properties guide us into extracting one-sided max-degree restriction tree in Lemma 3.12 and two-sided max-degree restriction tree in Lemma 3.19 respectively. Simple observations show that De Morgan sparsity is a one-side nice measure and generalized sparsity is even a two-sided nice measure, and from these observations our results stated here follow as special cases immediately.

Application to Sparsity vs Approximate Sparsity. To show that if a function f has large (De Morgan or generalized) sparsity, then it must also have large approximate (De Morgan or generalized) sparsity, we use the max-degree restriction trees guaranteed by Lemma 1.11 and Lemma 1.12. Given such a (one-sided or two-sided) max-degree restriction tree for f, we sample a random restriction by performing a random walk from the root to a leaf and returning the restriction associated with that leaf. Assuming toward contradiction that f admits a sparse approximating polynomial, we analyze the restricted function and its approximator under a random restriction. In contrast to our earlier examples, we cannot claim that all high-degree monomials in the approximator are "killed." Instead, we show that if the approximating polynomial were sparse to begin with, then with high probability its restriction contains no monomial of large degree in the remaining free variables. Each surviving monomial either collapses to a constant or remains nonzero but of low degree on the free variables. Thus, there exists a restriction under which the restricted approximator has small degree, whereas, by the definition of a max-degree restriction tree, the restricted function has max degree on the free variables. Invoking the quadratic relationship between degree and approximate degree proved by Aaronson et al. [ABDK+21] (see Theorem 2.7), we obtain a contradiction. It turns out that this argument naturally extends to give lower bounds on approximate (De Morgan or generalized) weight as well. Finally, the implications for deterministic and randomized AND-decision trees follow from known connections.

Application to Quantum–Classical Protocol Equivalence. Lee and Shraibman [LS+09c] observed that to prove lower bounds on the logarithm of the approximate rank of the communication matrix, as well as the quantum communication complexity, of an XOR-lifted function $f \circ \text{XOR}_2$, it suffices to show lower bounds on the (logarithm of the) approximate spectral norm, i.e., the Fourier ℓ_1-mass, of f. This connection implies that strong lower bounds on the approximate spectral norm of $f \circ \text{AND}_2$ yield strong quantum communication lower bounds for $f \circ \text{EQ}_4$. Using one-sided max-degree restriction tree of f, we lift (log of) the De Morgan sparsity of f to (log of) the approximate spectral norm of $f \circ \text{AND}_2$. This seems surprising given the general wisdom that Fourier monomials do not get killed/simplified by 0/1 restrictions. The key is that we are not proving a lower bound for f but for its AND-lift. Indeed, Krause and Pudlák [KP95] devised a modified random restriction argument to prove lower bounds on Fourier sparsity of lifted functions. Their technique was later adapted by Chattopadhyay and Mande [CM17b] to bound approximate spectral norm for a certain class of functions. In Section 5.1, we observe those ideas can be made to work with max-degree restriction trees as follows: taking a random walk on the max-degree restriction tree for f, we sample a *lifted-restriction* for $f \circ \text{AND}_2$. Such a random lifted restriction does not *kill* a high-degree monomial but with high probability turns it 'irrelevant' for the restricted function. In a second step, all irrelevant monomials can be effectively killed by applying an expectation operator leaving a low degree polynomial approximating the high degree restricted function at a leaf of the restriction tree, thus yielding again a contradiction.

Relation to Earlier Work. This report subsumes the results of our earlier ECCC report *Exact versus Approximate Representations of Boolean Functions in the De Morgan Basis* [CDL25], apart from a single result which showed that, for monotone functions, the complexity measures $\text{gspar}(f)$, $\widetilde{\text{gspar}}(f)$, $\text{gwt}(f)$, $\widetilde{\text{gwt}}(f)$, $\text{DSize}^{dt}(f)$, and $\text{RSize}^{dt}(f)$ are all polynomially related on the logarithmic scale, up to polylogarithmic factors in n. All other results in that report are either improved or subsumed here, and the present work contains several additional results.

Organization. In Section 2, we present the necessary preliminaries and define the complexity measures used throughout the paper. In Section 3, we formally define restriction trees and show how to construct them for functions with large one-sided and two-sided "nice" measures, with sparsity and generalized sparsity serving as canonical examples of such measures. In Section 4, we use restriction trees to relate various measures—sparsity, approximate sparsity, ℓ_1-norm, and approximate ℓ_1-norm—for both De Morgan and generalized polynomial representations. Finally, in Section 5, we show how to lift the sparsity measure of a Boolean function to the approximate rank of its lifted version $f \circ \text{EQ}_4$, and discuss the resulting consequences for EQ_4-lifted functions.

2 Preliminaries

In this section, we collect notation, definitions, and known results that will be used throughout the paper. All functions considered are defined on the Boolean hypercube $\{0,1\}^n$, and all polynomials are assumed to be multilinear real polynomials.

De Morgan Basis.

Definition 2.1 (Multilinear Polynomial Representation)**.** *A polynomial $Q \in \mathbb{R}[x_1, x_2, \ldots, x_n]$ is* multilinear *if each variable appears with degree at most one in every monomial. Every function $f : \{0,1\}^n \to \mathbb{R}$ has a unique multilinear polynomial Q satisfying $Q(x) = f(x)$ for all $x \in \{0,1\}^n$, and conversely every multilinear polynomial defines a function on the Boolean hypercube. Hence, functions on $\{0,1\}^n$ and multilinear polynomials are in one-to-one correspondence, and we use the two notions interchangeably.*

Definition 2.2 (Support, Range, Degree, Sparsity, and Norm of a Polynomial)**.** *Let $Q \in \mathbb{R}[x_1, \ldots, x_n]$ be a multilinear polynomial written as*

$$Q(x) = \sum_{S \subseteq [n]} a_S \prod_{i \in S} x_i.$$

- *The* support *of Q, denoted $\text{vars}(Q)$, is the set of variables that appear in some monomial with a nonzero coefficient.*

- *The* range *of* Q*, denoted* *range*(Q)*, is* $\max_{x\in\{0,1\}^n}|Q(x)|$.
- *The* degree *of* Q*, denoted* *deg*(Q)*, is* $\max\{|S| \mid a_S \neq 0\}$.
- *The* sparsity *of* Q*, denoted* *spar*(Q)*, is the number of nonzero coefficients* a_S.
- *The* ℓ_1*-norm of* Q*, denoted* *wt*(Q)*, is given by* $\sum_{S\subseteq[n]}|a_S|$.

Definition 2.3 (Complexity Measures for Functions via Polynomials)**.** *Let* $f:\{0,1\}^n\to\mathbb{R}$ *be a function, and let* $\mathcal{P}(f)$ *denote its unique multilinear polynomial representation. We define the following complexity measures:*

$$\mathit{deg}(f) := \mathit{deg}(\mathcal{P}(f)), \quad \mathit{spar}(f) := \mathit{spar}(\mathcal{P}(f)), \quad \mathit{wt}(f) := \mathit{wt}(\mathcal{P}(f)).$$

Remark 2.4. *For Boolean functions* $f:\{0,1\}^n\to\{0,1\}$*, the coefficients in* $\mathcal{P}(f)$ *are integers (see for example [Juk12, Chapter 2]), and hence* $\mathit{spar}(f)\leqslant\mathit{wt}(f)$*.*

Definition 2.5 (Complexity Measures for Functions via Approximating Polynomials)**.** *Let* $f:\{0,1\}^n\to\mathbb{R}$ *and let* $\varepsilon>0$*. We define:*

$$\begin{aligned}\widetilde{\mathit{deg}}_\varepsilon(f) &:= \min\{\mathit{deg}(Q) \mid Q \text{ satisfies } |Q(x)-f(x)|\leqslant\varepsilon \text{ for all } x\in\{0,1\}^n\},\\ \widetilde{\mathit{spar}}_\varepsilon(f) &:= \min\{\mathit{spar}(Q) \mid Q \text{ satisfies } |Q(x)-f(x)|\leqslant\varepsilon \text{ for all } x\in\{0,1\}^n\},\\ \widetilde{\mathit{wt}}_\varepsilon(f) &:= \min\{\mathit{wt}(Q) \mid Q \text{ satisfies } |Q(x)-f(x)|\leqslant\varepsilon \text{ for all } x\in\{0,1\}^n\}.\end{aligned}$$

When $\varepsilon=1/3$*, we write* $\widetilde{\mathit{deg}}(f):=\widetilde{\mathit{deg}}_{1/3}(f)$*,* $\widetilde{\mathit{spar}}(f):=\widetilde{\mathit{spar}}_{1/3}(f)$*, and* $\widetilde{\mathit{wt}}(f):=\widetilde{\mathit{wt}}_{1/3}(f)$*.*

Theorem 2.6 ([BT22], Theorem 10, Section 3.4)**.** *Let* $f:\{0,1\}^n\to\{0,1\}$ *be a Boolean function. Then for any* $0<\varepsilon<\frac{1}{2}$*,*

$$\widetilde{\mathit{deg}}_\varepsilon(f) = O\left(\widetilde{\mathit{deg}}_{1/3}(f)\cdot\log(1/\varepsilon)\right).$$

Theorem 2.7 ([ABDK+21, Theorem 4])**.** *For every Boolean function* $f:\{0,1\}^n\to\{0,1\}$*,*

$$\mathit{deg}(f) = O(\widetilde{\mathit{deg}}(f)^2).$$

Remark 2.8. *The bound in Theorem 2.7 is tight; for example, the* OR_n *function satisfies* $\mathit{deg}(\mathrm{OR}_n)=n$ *and* $\widetilde{\mathit{deg}}(\mathrm{OR}_n)=\Theta(\sqrt{n})$*.*

Generalized Polynomials. In standard polynomial representations, even simple functions like $\mathrm{OR}_n = 1-\prod_{i=1}^n(1-x_i)$ can have high sparsity: the standard expansion of OR_n contains 2^n-1 monomials. To address this and allow for more compact representations, we consider *generalized polynomials*, which extend standard polynomials by introducing formal complements $\bar{x}_i$ for each variable x_i. For example, OR_n can be written more succinctly as

$$\mathrm{OR}_n(x_1,\ldots,x_n) = 1-\prod_{i=1}^n\bar{x}_i,$$

where each $\bar{x}_i$ acts as a stand-in for $1-x_i$. This representation uses only two monomials, offering exponential savings in sparsity.

We now define generalized polynomials formally.

Definition 2.9 (Generalized Polynomial)**.** *A* generalized polynomial *is a polynomial over the ring*

$$\mathbb{R}[x_1,\ldots,x_n,\bar{x}_1,\ldots,\bar{x}_n]/I,$$

where $\bar{x}_i$ *denotes the formal complement of* x_i*, and* I *is the ideal generated by the relations:*

$$x_i^2-x_i=0 \quad \text{and} \quad x_i+\bar{x}_i-1=0 \quad \text{for all } i\in[n].$$

Definition 2.10 (Generalized Representation of Boolean Functions)**.** *A generalized polynomial* $Q\in\mathbb{R}[x_1,\ldots,x_n,\bar{x}_1,\ldots,\bar{x}_n]/I$ represents *a function* $f:\{0,1\}^n\to\mathbb{R}$ *if* $Q(x,\bar{x})=f(x)$ *for all* $x\in\{0,1\}^n$*, where* $\bar{x}_i=1-x_i$*.*

Definition 2.11 (Generalized Complexity Measures)**.** *As in the standard case, one can define the degree, sparsity, and ℓ_1-norm of a generalized polynomial. For a function $f : \{0,1\}^n \to \mathbb{R}$, we define gdeg$(f)$, gspar$(f)$, and gwt$(f)$ as the minimum degree, sparsity, and ℓ_1-norm, respectively, over all generalized polynomials that represent f exactly.*

Analogously, the approximate measures $\widetilde{gdeg}(f)$, $\widetilde{gspar}(f)$, and $\widetilde{gwt}(f)$ denote the minimum degree, sparsity, and ℓ_1-norm among all generalized polynomials that approximate f pointwise within error $1/3$.

Remark 2.12. *Using generalized polynomials offers no advantage in terms of degree. Indeed, each dual variable $\bar{x}_i$ can be replaced by $1 - x_i$, yielding a standard polynomial of the same degree. Therefore, $\deg(f) = gdeg(f)$ and $\widetilde{\deg}(f) = \widetilde{gdeg}(f)$. Since the degree measures coincide, we will simply write $\deg(f)$ and $\widetilde{\deg}(f)$ and avoid using the generalized notation gdeg(f) and $\widetilde{gdeg}(f)$.*

On the other hand, generalized representations are not unique and can be exponentially more succinct. As discussed above, OR_n has a generalized representation with just two monomials, while its standard representation requires $2^n - 1$.

Fourier Basis. Another fundamental basis for representing Boolean functions is the *Fourier basis*, where each monomial $\chi_S(x) = (-1)^{\sum_{i \in S} x_i}$ represents the ±1-valued parity function on the subset $S \subseteq [n]$ of variables.

Definition 2.13 (Fourier Complexity Measures)**.** *Let $f : \{0,1\}^n \to \mathbb{R}$ have the Fourier expansion*

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\, \chi_S(x).$$

The Fourier degree *of f is the size of the largest subset S with a nonzero Fourier coefficient $\widehat{f}(S)$. Since $\chi_S(x) = \prod_{i \in S}(1 - 2x_i)$, the Fourier degree of f equals its ordinary degree. The* Fourier sparsity *of f, denoted $\left\|\widehat{f}\right\|_0$, is the number of nonzero coefficients $\widehat{f}(S)$. The* Fourier ℓ_1-norm *(also called the* Fourier spectral norm*) is*

$$\left\|\widehat{f}\right\|_1 = \sum_{S \subseteq [n]} |\widehat{f}(S)|.$$

For any $\varepsilon > 0$, the ε-approximate Fourier sparsity *and* ε-approximate Fourier ℓ_1-norm *of f are*

$$\left\|\widehat{f}\right\|_{0,\varepsilon} := \min_p \left\{ \|\widehat{p}\|_0 : p : \{0,1\}^n \to \mathbb{R},\ |p(x) - f(x)| \leqslant \varepsilon \text{ for all } x \in \{0,1\}^n \right\},$$
$$\left\|\widehat{f}\right\|_{1,\varepsilon} := \min_p \left\{ \|\widehat{p}\|_1 : p : \{0,1\}^n \to \mathbb{R},\ |p(x) - f(x)| \leqslant \varepsilon \text{ for all } x \in \{0,1\}^n \right\}.$$

Communication Complexity. We assume familiarity with the standard model of communication complexity and refer the reader to [KN97] for detailed definitions. In this model, two parties, Alice and Bob, aim to compute a Boolean function $F : X \times Y \to \{0,1\}$, where Alice receives $x \in X$ and Bob receives $y \in Y$. They exchange messages according to a pre-agreed protocol in order to compute $F(x,y)$, while minimizing the total number of bits communicated in the worst case. The *deterministic communication complexity* of F, denoted $\mathrm{D}^{cc}(F)$, is the minimum number of bits exchanged by any deterministic protocol that always outputs $F(x,y)$. In the *randomized communication model*, Alice and Bob have access to shared public randomness and must compute $F(x,y)$ correctly with probability at least $2/3$ for all inputs. The corresponding measure is the *randomized communication complexity*, denoted $\mathrm{R}^{cc}(F)$.

We also assume familiarity with quantum communication complexity [DW02]. We use $\mathrm{Q}^{cc}(F)$ to denote the bounded-error (error at most $1/3$) quantum communication complexity of a two-party function F in the model with unlimited shared entanglement. In this model, Alice and Bob may share, at the start of the protocol, an entangled state of their choice (independent of their inputs x and y) at no cost—for example, a collection of EPR pairs. This shared entanglement can be viewed as the quantum analogue of shared randomness.

Definition 2.14 (Rank and Approximate Rank)**.** *For a Boolean function $F : X \times Y \to \{0,1\}$, let its* communication matrix *be $M_F(x,y) = F(x,y)$. The* rank *of F (over the reals) is $rank(F) = rank(M_F)$. For $0 < \varepsilon < 1/2$, the ε-*approximate rank *of F is*

$$\widetilde{rank}_\varepsilon(F) \;=\; \min\Big\{ rank(A) : A \in \mathbb{R}^{|X|\times|Y|},\ \|A - M_F\|_\infty \leqslant \varepsilon \Big\}.$$

We write $\widetilde{rank}(F)$ for $\widetilde{rank}_{1/3}(F)$.

3 Restriction Trees

Definition 3.1 (Restrictions)**.** *A* restriction *ρ on a set of variables $V \subseteq \{x_1, \ldots, x_n\}$ is a partial assignment $\rho : V \to \{0,1,*\}$, where for $x_i \in V$, $\rho(x_i) \in \{0,1\}$ indicates that x_i is fixed, and $\rho(x_i) = *$ means x_i is left free. For a polynomial $Q \in \mathbb{R}[x_1, \ldots, x_n]$, we write $Q|_\rho$ for the polynomial obtained by substituting $x_i = \rho(x_i)$ for all fixed variables x_i. The* size *of ρ, denoted $|\rho|_*$, is the number of variables in V left free.*

An *adaptive random restriction* on a polynomial proceeds by successively fixing variables, where each decision may depend on how previous assignments affect a chosen complexity measure (for example, the sparsity of the polynomial). This process is naturally captured by a tree structure, which we call a *restriction tree.* In a restriction tree, each node specifies a variable to be restricted, and each outgoing edge corresponds to a possible value assigned to that variable. We illustrated this concept informally in the introduction; here, we define it formally.

Definition 3.2 (Restriction tree)**.** *Let $\Sigma \subseteq \{0,1,*\}$ and let $V \subseteq \{x_1, \ldots, x_n\}$. A Σ-*restriction tree *on V of depth d is a full $|\Sigma|$-ary tree of depth d satisfying:*

1. *Each internal node is labeled by a variable $x_i \in V$, and the labeling is* read-once*—that is, no variable appears more than once along any root-to-leaf path.*
2. *Each internal node has one outgoing edge for every symbol in Σ, each labeled by a distinct element of Σ.*
3. *Each leaf v is associated with a restriction $\rho_v : V \to \{0,1,*\}$ of the following form. Let $A_\sigma(v)$ be the set of variables whose edge labeled $\sigma \in \Sigma$ was taken on the path from the root to v. Then ρ_v has the form,*

$$\rho_v(x_i) = \begin{cases} \sigma & \text{if } x_i \in A_\sigma(v) \text{ for some } \sigma \in \Sigma, \\ \in \{0,1\} & \text{if } x_i \in V \backslash \bigcup_{a \in \Sigma} A_a(v). \end{cases}$$

In our work, we consider two kinds of restriction trees, each suited to a different purpose: the $\{0,*\}$-restriction tree and the $\{0,1,*\}$-restriction tree.

Definition 3.3 (Max-degree restriction)**.** *Let $Q \in \mathbb{R}[x_1, \ldots, x_n]$ be a nonzero multilinear polynomial, and let $\rho : V \to \{0,1,*\}$ be a restriction satisfying $vars(Q) \subseteq V$. We say that ρ is a* max-degree restriction *of Q if the restricted polynomial $Q|_\rho$ is nonzero and has maximal degree, i.e., $deg(Q|_\rho) = |\rho|_*$.*

Definition 3.4 (Max-degree restriction tree)**.** *Let $Q \in \mathbb{R}[x_1, \ldots, x_n]$ be a multilinear polynomial, and let T be a Σ-restriction tree on variables $V \supseteq vars(Q)$. We say that T is a* max-degree Σ-restriction tree *of Q if every leaf v of T is labeled by a max-degree restriction ρ_v of Q.*

For a Boolean function $f : \{0,1\}^n \to \{0,1\}$, we call T a max-degree Σ-restriction tree of f if it is a max-degree Σ-restriction tree of the unique multilinear polynomial Q representing f.

We refer to $\{0,*\}$-restriction trees as *one-sided restriction trees* and $\{0,1,*\}$-restriction trees as *two-sided restriction trees.* Similarly, for a polynomial Q (or a Boolean function f), we call a max-degree $\{0,*\}$-restriction tree a *one-sided max-degree restriction tree*, and a max-degree $\{0,1,*\}$-restriction tree a *two-sided max-degree restriction tree.*

In the next two subsections, we describe how to construct deep one-sided max-degree restriction trees for functions with high sparsity or ℓ_1-norm, and deep two-sided max-degree restriction trees for functions with high generalized sparsity or generalized ℓ_1-norm. In fact, we abstract the essential properties of a measure that ensure that a large value of the measure yields a deep restriction tree.

3.1 Restriction Trees from One-Sided Measures

Algorithm 1 OneSidedMaxDegreeRestrictionTree

Input: Non-zero multilinear polynomial $Q \in \mathbb{R}[x_1, \ldots, x_n]$; set $V \subseteq \{x_1, \ldots, x_n\}$ with $\text{vars}(Q) \subseteq V$.
Output: A max-degree $\{0, *\}$-restriction tree T for Q.
if $|V| = 0$ **then**
    return a tree consisting of a single leaf labeled by the empty restriction.
else
    if there exists $x_i \in V$, $b \in \{0, 1\}$ such that $\mu(Q|_{x_i=b}) \geqslant \left(1 - \frac{1}{n}\right) \cdot \mu(Q)$ **then**
        Let $T' \leftarrow$ OneSidedMaxDegreeRestrictionTree($Q|_{x_i=b}, V \backslash \{x_i\}$)
        Extend each leaf label ρ_v in T' by $x_i = b$
        Set $T \leftarrow T'$
    else
        Choose $x_i \in V$ arbitrarily
        Express Q as $Q = R_1 \cdot x_i + R_0$
        Let $T_0 =$ OneSidedMaxDegreeRestrictionTree($R_0, V \backslash \{x_i\}$)
        Let $T_* =$ OneSidedMaxDegreeRestrictionTree($R_1, V \backslash \{x_i\}$)
        Extend each leaf label ρ_v in T_0 by $x_i = 0$
        Extend each leaf label ρ_v in T_* by $x_i = *$
        Create T: root labeled by x_i, 0-edge going to T_0, $*$-edge to T_*
    end if
    return T
end if

For a complexity measure $\mu : \mathbb{R}[x_1, \ldots, x_n] \to \mathbb{R}_{\geqslant 0}$ mapping real polynomials to non-negative values, Algorithm 1 describes a procedure for constructing a one-sided max-degree restriction tree for a given polynomial Q with large μ-measure. As long as μ behaves "nicely", starting from a polynomial with large measure, the algorithm produces a one-sided max-degree restriction tree of large depth. We present the algorithm for a general measure μ to highlight the properties required for the construction. Later, we will instantiate μ with specific measures such as sparsity and the ℓ_1-norm. Next, we formalize the notion of a "nicely behaved" measure.

Definition 3.5 (One-sided nice measures)**.** *Let $\mu : \mathbb{R}[x_1, \ldots, x_n] \to \mathbb{R}_{\geqslant 0}$ be a complexity measure on polynomials. We say that μ is* one-sided nice *if it satisfies the following properties:*

1. *(Subadditivity) For all polynomials P, Q, $\mu(P + Q) \leqslant \mu(P) + \mu(Q)$.*
2. *(Variable monotonicity) For any polynomial P and variable x, $\mu(xP) \leqslant \mu(P)$.*
3. *(Sign invariance) For any polynomial P, $\mu(P) = \mu(-P)$.*
4. *(Normalization) For any constant polynomial c, $\mu(c) \leqslant |c|$. In particular, $\mu(0) = 0$.*

Remark 3.6. *Examples of one-sided nice measures include spar$(\cdot)$ and wt$(\cdot)$. Throughout, we will run Algorithm 1 with a fixed one-sided nice measure in mind. We also note that the normalization condition in the above definition can be relaxed—it is included primarily to capture measures such as wt$(\cdot)$.*

For a variable x_i and a multilinear polynomial Q expressed as $Q = R_1 x_i + R_0$, our algorithm is concerned with how the measure μ behaves under the restrictions

$$Q|_{x_i=0} = R_0 \quad \text{and} \quad Q|_{x_i=1} = R_0 + R_1.$$

When μ is one-sided nice, it satisfies the following structural property, which is crucial for constructing deep restriction trees starting from polynomials having large value under the measure μ. It shows that either μ doesn't decrease much under some restriction of x_i, or otherwise that μ is lower bounded for both R_0 and R_1. Apart from normalization, this is the only aspect of one-sided niceness that our analysis relies on.

Claim 3.7. *Let $\varepsilon > 0$, and let $Q \in \mathbb{R}[x_1, \ldots, x_n]$ be a multilinear polynomial. If μ is one-sided nice, then for every $x_i \in \text{vars}(Q)$, writing $Q = R_1 x_i + R_0$, one of the following holds:*

1. $\max\{\mu(R_0), \mu(R_0 + R_1)\} \geqslant (1-\varepsilon)\,\mu(Q)$, *or*
2. $\min\{\mu(R_0), \mu(R_1)\} \geqslant \frac{\varepsilon}{2}\,\mu(Q)$.

Proof. Fix a variable $x_i \in \text{vars}(Q)$, and write $Q = R_1 x_i + R_0$. Suppose

$$\max\{\mu(R_0), \mu(R_0 + R_1)\} < (1-\varepsilon)\,\mu(Q).$$

Then,

$$\mu(Q) = \mu(R_1 x_i + R_0) \overset{(1)}{\leqslant} \mu(x_i R_1) + \mu(R_0) \overset{(2)}{\leqslant} \mu(R_1) + \mu(R_0) \overset{(3)}{<} \mu(R_1) + (1-\varepsilon)\mu(Q),$$

which implies $\mu(R_1) \geqslant \varepsilon\,\mu(Q)$. Here, (1) uses the subadditivity of μ, (2) uses its variable monotonicity, and (3) uses the assumption.

Next, we have

$$\begin{aligned}\mu(Q) &\overset{(1)}{\leqslant} \mu(R_1) + \mu(R_0) = \mu(R_1 + R_0 - R_0) + \mu(R_0)\\ &\overset{(2)}{\leqslant} \mu(R_1 + R_0) + \mu(-R_0) + \mu(R_0) \overset{(3)}{=} \mu(R_0 + R_1) + 2\mu(R_0) \overset{(4)}{<} (1-\varepsilon)\mu(Q) + 2\mu(R_0),\end{aligned}$$

which yields $\mu(R_0) \geqslant \frac{\varepsilon}{2}\,\mu(Q)$. Here, (1) reuses the inequality above, (2) and (3) use subadditivity and sign invariance of μ, and (4) uses the assumption.

Combining the two bounds proves the claim. ∎

Observation 3.8. *By the above claim, if the input polynomial Q to Algorithm 1 has $\mu(Q) > 0$, where μ is a one-sided nice measure, then all recursive calls in the algorithm operate on polynomials with non-zero μ-measure. Since $\mu(0) = 0$, this also implies that all polynomials encountered during the execution are non-zero.*

Our construction algorithm terminates when $|V| = 0$, at which point the input polynomial becomes a nonzero constant. In our analysis, we will need an upper bound on the magnitude of this constant at termination. The following simple fact about discrete derivatives will be useful.

Claim 3.9. *Let $Q \in \mathbb{R}[x_1, \ldots, x_n]$ be a multilinear polynomial with range$(Q) = r$. For any variable x_i, write $Q = R_1 x_i + R_0$. Then the range of R_1 is at most $2r$.*

Proof. Fix any $w \in \{0,1\}^n$, and let w_0, w_1 be the assignments obtained from w by setting $x_i = 0$ and $x_i = 1$, respectively. Then $R_1(w) = Q(w_1) - R_0(w) = Q(w_1) - Q(w_0)$. Since both $Q(w_0)$ and $Q(w_1)$ lie in $[-r, r]$, their difference lies in $[-2r, 2r]$, as claimed. ∎

On an input polynomial with large one-sided nice measure μ, we now show that the restriction tree produced by Algorithm 1 is both deep and a one-sided max-degree restriction tree, up to the minor technicality that each root-to-leaf path has depth at least d rather than exactly d. This detail can be fixed without affecting any properties of the tree (see Claim 3.12). We begin by analyzing the depth.

The algorithm constructs the restriction tree recursively, following one of three possible cases:

- If $|V| = 0$, the recursion terminates and the algorithm backtracks.
- If the condition on line 6 holds, the algorithm makes a single recursive call (line 7).
- Otherwise, the condition on line 10 holds, and the algorithm makes two recursive calls (lines 13 and 14).

The recursion proceeds while $|V| \geqslant 1$ and halts when $|V| = 0$, after which the final tree is assembled during backtracking. We classify recursive calls as *passive* if line 6 is satisfied, and *active* if line 10 is satisfied. We will argue that any execution of the algorithm must include a substantial number of active calls. Each active call increases the depth of the resulting tree by one. Hence, if every execution path includes at least ℓ active calls, the resulting tree has depth at least ℓ along every root-to-leaf path.

Claim 3.10. *Let Q be a multilinear polynomial with $\mu(Q) \geqslant s$ for a one-sided nice measure μ and range$(Q) = r$, and let $T = $ OneSidedMaxDegreeRestrictionTree$(Q, \{x_1, \ldots, x_n\})$ be the tree output by Algorithm 1. Then every root-to-leaf path in T has length at least $\Omega\left(\frac{\log(s/r)}{\log n}\right)$.*

Proof. It suffices to show that every execution path of the algorithm contains at least $\Omega\left(\frac{\log(s/r)}{\log n}\right)$ active calls. Fix an execution path of the algorithm. Since the algorithm removes one variable per recursive call, the path comprises exactly n calls before reaching the base case ($|V| = 0$). Let ℓ of these calls be *active*. Track how the measure μ and the range of the input polynomial evolve during the recursion:

- In a *passive* call (line 6), the next polynomial is $Q|_{x_i=u}$ with $\mu(Q|_{x_i=u}) \geqslant (1 - 1/n)\,\mu(Q)$, while the range does not increase as range$(Q|_{x_i=u}) \leqslant$ range(Q).
- In an *active* call (line 10), Claim 3.7 guarantees that μ decreases by at most a factor of $1/(2n)$. If the recursion proceeds via line 13, the range does not increase; if it proceeds via line 14, the next polynomial is the discrete derivative $\partial_{x_i} Q$, whose range increases by at most a factor of 2 (Claim 3.9).

At the base case ($|V| = 0$), the polynomial is a nonzero constant. Since each active call can at most double the range, the resulting constant has magnitude at most $2^\ell r$, where $r = $ range(Q). Using $\mu(c) \leqslant |c|$ for constants c, the cumulative change in μ along the path satisfies

$$2^\ell r \;\geqslant\; (1 - 1/n)^{n-\ell} (1/2n)^\ell \,\mu(Q) \;\geqslant\; (1 - 1/n)^n (1/2n)^\ell s \;\geqslant\; \tfrac{1}{10}\,(1/2n)^\ell s,$$

where the last inequality uses $(1 - 1/n)^n \geqslant 1/10$ for $n \geqslant 2$.

Rearranging and taking logarithms yields

$$\ell \;\geqslant\; \frac{\log(s/10r)}{\log(4n)} \;=\; \Omega\Big(\frac{\log(s/r)}{\log n}\Big).$$

Hence every execution path of the algorithm contains at least $\Omega(\log(s/r)/\log n)$ active calls, as claimed. ◻

Next, we show that the restrictions labeling the leaves of the tree output by Algorithm 1 satisfy the max-degree condition for the input polynomial Q.

Claim 3.11. *Let $Q \in \mathbb{R}[x_1, \ldots, x_n]$ be a nonzero multilinear polynomial with vars$(Q) \subseteq V$, and let $T = $ OneSidedMaxDegreeRestrictionTree(Q, V) be the tree produced by Algorithm 1. Then for every leaf restriction ρ of T, the polynomial $Q|_\rho$ is nonzero and has full degree.*

Proof. We proceed by induction on $|V|$.

Base Case ($|V| = 0$): Here, Q must be a non-zero constant polynomial. The claim holds trivially.

Inductive Step ($|V| \geqslant 1$). We consider the two possible branches of the algorithm, depending on which condition is satisfied (line 6 or line 10):

1. **Case 1:** The condition at line 6 is satisfied. Suppose this occurs for some variable $x_i \in V$ and value $b \in \{0, 1\}$. The algorithm returns a tree T obtained from T'

$$T' \leftarrow \text{OneSidedMaxDegreeRestrictionTree}(Q|_{x_i=b},\, V \backslash \{x_i\})$$

by extending every leaf restriction of T' with $x_i = b$. Let $\rho_v = \rho'_v \cup \{x_i = b\}$ denote a leaf restriction of the resulting tree T, where ρ'_v is a leaf restriction of T'. By the induction hypothesis, T' is a max-degree restriction tree for $Q|_{x_i=b}$; hence $(Q|_{x_i=b})|_{\rho'_v}$ is nonzero and has full degree. Therefore,

$$Q|_{\rho_v} = (Q|_{x_i=b})|_{\rho'_v} \quad \text{and} \quad \deg(Q|_{\rho_v}) = \deg((Q|_{x_i=b})|_{\rho'_v}) \overset{(1)}{=} |\rho'_v|_* \overset{(2)}{=} |\rho_v|_*,$$

where (1) follows from the induction hypothesis, and (2) because ρ_v and ρ'_v leave the same number of variables free. Thus $Q|_{\rho_v}$ is nonzero and has full degree.

2. **Case 2:** The "else" branch (line 10) is executed. Let $x_i \in V$ be the variable chosen in line 11, and write $Q = x_i R_1 + R_0$. The algorithm constructs the subtrees

$$T_0 = \textsc{OneSidedMaxDegreeRestrictionTree}(R_0,\, V \backslash \{x_i\}),$$
$$T_* = \textsc{OneSidedMaxDegreeRestrictionTree}(R_1,\, V \backslash \{x_i\}),$$

and combines them into the output tree T by extending each leaf of T_0 with $x_i = 0$ and each leaf of T_* with $x_i = *$. For a leaf restriction $\rho_v = \rho_v^0 \cup \{x_i = 0\}$, where ρ_v^0 is a leaf restriction of T_0, the argument is identical to Case 1. Now consider a leaf restriction $\rho_v = \rho_v^* \cup \{x_i = *\}$, where ρ_v^* is a leaf restriction of T_*. Then $Q|_{\rho_v} = x_i \cdot (R_1|_{\rho_v^*}) + (R_0|_{\rho_v^*})$, and hence

$$\deg(Q|_{\rho_v}) \overset{(1)}{=} \max\left(1 + \deg(R_1|_{\rho_v^*}),\, \deg(R_0|_{\rho_v^*})\right) \geqslant 1 + \deg(R_1|_{\rho_v^*}) \overset{(2)}{=} 1 + |\rho_v^*|_* = |\rho_v|_*.$$

where (1) and (2) follow from the induction hypothesis, which ensures that $R_1|_{\rho_v^*}$ is non-zero and of full degree. Since $\deg(Q|_{\rho_v})$ cannot exceed $|\rho_v|_*$, equality must hold. Moreover, as $R_1|_{\rho_v^*}$ is nonzero, so is $Q|_{\rho_v}$.

In both cases, the restriction of Q under any leaf restriction is a nonzero polynomial of full degree. ☐

Combining Claim 3.10 and Claim 3.11, we obtain the following:

Claim 3.12. *Let Q be a multilinear polynomial with $\mu(Q) \geqslant s$ for a one-sided -nice measure μ and range$(Q) = r$. Then there exists a one-sided max-degree restriction tree for Q of depth $d = \Omega\left(\frac{\log(s/r)}{\log n}\right)$.*

Proof. Let $T = \textsc{OneSidedMaxDegreeRestrictionTree}(Q, \{x_1, \ldots, x_n\})$. By construction and by Claim 3.10, Claim 3.11, T is a one-sided max-degree restriction tree for Q, up to the minor technicality that each root-to-leaf path has depth at least d rather than exactly d. To obtain a tree T' of exact depth d, modify T as follows: for each node v at depth d, convert v into a leaf and label it with the restriction ρ corresponding to the leaf of T reached by continuing from v along only 0-edges. ☐

Since both sparsity and the ℓ_1-norm are one-sided nice measures, applying Algorithm 1 to the multilinear polynomial representing a Boolean function yields the following corollaries.

Corollary 3.13. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function of sparsity s. Then f admits a one-sided max-degree restriction tree of depth $\Omega(\log s / \log n)$.*

Corollary 3.14. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function of ℓ_1-norm w. Then f admits a one-sided max-degree restriction tree of depth $\Omega(\log w / \log n)$.*

Remark 3.15. *Since for every Boolean function $f : \{0,1\}^n \to \{0,1\}$, we have spar$(f) \leqslant$ wt(f) (see Remark 2.4), Corollary 3.13 also follows as a special case of Corollary 3.14.*

We note that Corollary 3.13 corresponds to Lemma 1.11, as stated in the introduction.

3.2 Restriction Trees from Two-sided Measures

Algorithm 2 TwoSidedMaxDegreeRestrictionTree

Input: Non-zero multilinear polynomial $Q \in \mathbb{R}[x_1, \ldots, x_n]$; set $V \subseteq \{x_1, \ldots, x_n\}$ with $\mathrm{vars}(Q) \subseteq V$.
Output: A max-degree $\{0, 1, *\}$-restriction tree T for Q.
if $|V| = 0$ **then**
  return a tree consisting of a single leaf labeled by the empty restriction.
else
  if there exists $x_i \in V$, $b \in \{0, 1\}$ such that $\mu(Q|_{x_i=b}) \geqslant \left(1 - \frac{1}{n}\right) \cdot \mu(Q)$ **then**
    Let $T' \leftarrow$ TwoSidedMaxDegreeRestrictionTree$(Q|_{x_i=b}, V\backslash\{x_i\})$
    Extend each leaf label ρ_v in T' by $x_i = b$
    Set $T \leftarrow T'$
  else
    Choose $x_i \in V$ arbitrarily
    Express Q as $Q = R_1 \cdot x_i + R_0$
    Let $T_0 =$ TwoSidedMaxDegreeRestrictionTree$(R_0, V\backslash\{x_i\})$
    Let $T_1 =$ TwoSidedMaxDegreeRestrictionTree$(R_1 + R_0, V\backslash\{x_i\})$
    Let $T_* =$ TwoSidedMaxDegreeRestrictionTree$(R_1, V\backslash\{x_i\})$
    Extend each leaf label ρ_v in T_0 by $x_i = 0$
    Extend each leaf label ρ_v in T_1 by $x_i = 1$
    Extend each leaf label ρ_v in T_* by $x_i = *$
    Create T: root labeled by x_i, 0-edge going to T_0, 1-edge going to T_1 and $*$-edge going to T_*
  end if
  return T
end if

For a complexity measure $\mu : \mathbb{R}[x_1, \ldots, x_n] \to \mathbb{R}_{\geqslant 0}$, Algorithm 2 constructs a two-sided max-degree restriction tree for a polynomial Q with large μ-value. Analogous to the one-sided construction in Section 3.1, if μ satisfies suitable niceness conditions, then starting from a polynomial of large measure, the algorithm produces a two-sided max-degree restriction tree of large depth. The notion of niceness here encompasses measures such as sparsity and the ℓ_1-norm of generalized polynomial representations of Q.

Definition 3.16 (Two-sided nice measures)**.** *Let $\mu : \mathbb{R}[x_1, \ldots, x_n] \to \mathbb{R}_{\geqslant 0}$ be a complexity measure on polynomials. We call μ* two-sided nice *if it is one-sided nice and additionally satisfies*

$$\textit{(Literal monotonicity)} \qquad \mu((1 - x)P) \leqslant \mu(P) \quad \textit{for all polynomials } P \textit{ and variables } x.$$

Remark 3.17. *Examples of two-sided nice measures include gspar$(\cdot)$ and gwt$(\cdot)$. Throughout, Algorithm 2 is analyzed under a fixed two-sided nice measure.*

The following claim, analogous to Claim 3.7, describes how a two-sided nice measure μ behaves under variable restrictions. Since two-sided niceness strengthens one-sided niceness, we obtain a stronger conclusion—crucial for constructing deep two-sided restriction trees from polynomials with large two-sided nice measure. Apart from normalization, this is the only property of two-sided niceness used in our analysis.

Claim 3.18. *Let $\varepsilon > 0$, and let $Q \in \mathbb{R}[x_1, \ldots, x_n]$ be a multilinear polynomial. If μ is two-sided nice, then for every $x_i \in \mathrm{vars}(Q)$, writing $Q = R_1 x_i + R_0$, one of the following holds:*

1. $\max\{\mu(R_0),\, \mu(R_0 + R_1)\} \geqslant (1 - \varepsilon)\, \mu(Q)$, *or*
2. $\min\{\mu(R_0),\, \mu(R_1),\, \mu(R_1 + R_0)\} \geqslant \frac{\varepsilon}{2}\, \mu(Q)$.

Proof. Fix a variable $x_i \in \mathrm{vars}(Q)$, and write $Q = R_1 x_i + R_0$. Suppose

$$\max\{\mu(R_0),\, \mu(R_0 + R_1)\} < (1 - \varepsilon)\, \mu(Q).$$

Since μ is one-sided nice, we have $\min\{\mu(R_0), \mu(R_1)\} \geqslant \frac{\varepsilon}{2}\mu(Q)$. Now,

$$\begin{aligned}
\mu(Q) &= \mu(R_1 + R_0 - (1-x_i)R_1) \\
&\leqslant \mu(R_1 + R_0) + \mu((1-x_i)R_1) && \text{(subadditivity, sign invariance)} \\
&\leqslant \mu(R_1 + R_0) + \mu(R_1) && \text{(literal monotonicity)} \\
&= \mu(R_1 + R_0) + \mu(R_1 + R_0 - R_0) \\
&\leqslant 2\mu(R_1 + R_0) + \mu(R_0) && \text{(subadditivity, sign invariance)} \\
&< 2\mu(R_1 + R_0) + (1-\varepsilon)\mu(Q),
\end{aligned}$$

which implies $\mu(R_1 + R_0) \geqslant \frac{\varepsilon}{2}\mu(Q)$. □

We claim that when Algorithm 2 is applied to a polynomial Q with large two-sided -nice measure μ, it produces a tree that is both deep and a two-sided max-degree restriction tree for Q. The proof closely follows the one-sided case (Algorithm 1): the depth argument proceeds as in Claim 3.10, with Claim 3.18 taking the place of Claim 3.7, while the max-degree condition is established inductively, as in Claim 3.11. Since the argument uses no new ideas, we omit the details and state the result.

Claim 3.19. *Let Q be a multilinear polynomial with $\mu(Q) \geqslant s$ for a two-sided nice measure μ and range$(Q) = r$. Then there exists a two-sided max-degree restriction tree for Q of depth $d = \Omega\left(\frac{\log(s/r)}{\log n}\right)$.*

Both generalized sparsity and generalized ℓ_1-norm are two-sided nice measures. Applying Algorithm 2 to the multilinear polynomial representing a Boolean function yields the following corollaries.

Corollary 3.20. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function of generalized sparsity s. Then f admits a two-sided max-degree restriction tree of depth $\Omega(\log s/\log n)$.*

Corollary 3.21. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function of generalized ℓ_1-norm w. Then f admits a two-sided max-degree restriction tree of depth $\Omega(\log w/\log n)$.*

We note that Corollary 3.13 corresponds to Lemma 1.12, as stated in the introduction.

4 Sparsity versus Approximate Sparsity

In this section, we show that for every Boolean function $f : \{0,1\}^n \to \{0,1\}$, the logarithms of its exact and approximate sparsity and ℓ_1-norm—both for ordinary and generalized polynomial representations—are polynomially related.

Proof Overview. We outline the proof relating the exact and approximate sparsity of ordinary polynomial representations; the same ideas extend naturally to the other cases.

Let f be a Boolean function with large exact sparsity. We aim to show that any polynomial approximating f within error $1/3$ must also have large sparsity. Using the large sparsity of f, we construct a deep one-sided max-degree restriction tree T of depth $d = \Omega(\log(\mathrm{spar}(f))/\log n)$, as described in Section 3. For a random restriction ρ sampled uniformly from the leaves of T, we show that ρ satisfies the following properties:

1. With probability at least 0.9, $\deg(f|_\rho) \geqslant d/4$.
2. For any monomial M and any $t \geqslant 1$, $\Pr_\rho[\deg(M|_\rho) \geqslant t] \leqslant 2^{-t}$.

With the above properties of ρ, it becomes evident why the approximate sparsity of f must be large. Suppose, for contradiction, that there exists a polynomial $\tilde{Q}$ that 1/3-approximates f and has sparsity less than $\frac{1}{10}\cdot 2^{\sqrt{d/4c}}$ for some constant $c > 0$. By Property (2) and a union bound, with probability at least 0.9, the restriction ρ eliminates all monomials of degree at least $\sqrt{d/4c}$ in $\tilde{Q}$; hence, $\deg(\tilde{Q}|_\rho) < \sqrt{d/4c}$ with probability at least 0.9. At the same time, by Property (1), $\deg(f|_\rho) \geqslant d/4$ with probability at least 0.9. Therefore, there exists a restriction ρ such that

$$\deg(f|_\rho) \geqslant d/4 \quad \text{and} \quad \deg(\tilde{Q}|_\rho) < \sqrt{d/4c}.$$

But since $\tilde{Q}|_\rho$ approximates $f|_\rho$, this contradicts the known relationship between degree and approximate degree for Boolean functions—namely, that $\deg(f) \leqslant c \cdot \widetilde{\deg}(f)^2$ for some universal constant c [ABDK+21].

We conclude that any polynomial approximating f must have sparsity at least $\frac{1}{10} \cdot 2^{\sqrt{d/4c}}$. Since $d = \Omega(\log(\mathrm{spar}(f))/\log n)$, it follows that the logarithms of exact and approximate sparsity are quadratically related (up to a $\log n$ factor).

A similar argument shows that the exact and approximate ℓ_1-norms of ordinary polynomial representations are also quadratically related (up to a $\log n$ factor).

For generalized measures arising from generalized polynomial representations, the same reasoning applies, with one-sided restriction trees replaced by two-sided max-degree restriction trees. As shown in Section 3.2, a large generalized measure yields a deep two-sided max-degree restriction tree. Sampling a random restriction ρ from its leaves satisfies analogous properties:

1. With probability at least 0.9, $\deg(f|_\rho) \geqslant d/6$.
2. For any generalized monomial M and any $t \geqslant 1$, $\Pr_\rho[\deg(M|_\rho) \geqslant t] \leqslant 2^{-t}$.

The strengthened Property (2) then gives the same type of quadratic relationship between exact and approximate generalized measures, showing that the exact and approximate sparsity and ℓ_1-norms of generalized polynomial representations of Boolean functions are also quadratically related (up to a $\log n$ factor).

4.1 De Morgan Basis

We show Properties (1) and (2) from the proof overview for restrictions sampled uniformly from the leaves of one-sided restriction trees, and then use them to relate the various measures associated with ordinary polynomial representations.

For a restriction tree T, let $\rho \sim \mathsf{Leaf}(T)$ denote the process of sampling a leaf v of T uniformly at random and returning the restriction ρ labeled at that leaf.

Claim 4.1. *Let T be a one-sided max-degree restriction tree of depth $d \geqslant 40$ for a multilinear polynomial Q. Then*

$$\Pr_{\rho\sim\mathsf{Leaf}(T)}[\mathit{deg}(Q|_\rho) \geqslant d/4] \geqslant 0.9.$$

Proof. Since T is a max-degree restriction tree for Q, for every leaf restriction ρ we have $\deg(Q|_\rho) = |\rho|_*$. Hence, it suffices to show that $|\rho|_* \geqslant d/4$ with high probability.

Sampling ρ uniformly from the leaves of T is equivalent to performing a random root-to-leaf walk, where each step independently follows a $*$-edge with probability 1/2, and outputting the restriction labeling the leaf reached. The number of free variables in the resulting restriction equals the number of $*$-edges encountered, implying that $\mathbb{E}[|\rho|_*] = d/2$. Applying a Chernoff bound,

$$\Pr[|\rho|_* \leqslant d/4] \leqslant e^{-d/16} \leqslant 0.1,$$

where the last inequality uses $d \geqslant 40$. □

Claim 4.2. *Let T be a one-sided restriction tree over variables V. Then, for any monomial M with $\mathit{vars}(M) \subseteq V$ and any $t \in \mathbb{N}$,*

$$\Pr_{\rho\sim\mathsf{Leaf}(T)}[\, \mathit{deg}(M|_\rho) \geqslant t \,] \leqslant 2^{-t}.$$

Proof. We prove the claim by induction on the depth d of the restriction tree T.

Base Case ($d = 0$): When T has depth zero, it consists of a single leaf labeled with a restriction that assigns all variables in V. Hence, M becomes a constant under this restriction, and the claim holds trivially.

Inductive Step ($d > 0$): Let the root of T query variable x_i, and let the 0- and $*$-edges lead to subtrees T_0 and T_*, which are one-sided restriction trees on $V\backslash\{x_i\}$ of depth $d-1$. Let $\rho \sim \mathsf{Leaf}(T)$, and define $\rho_0 \sim \mathsf{Leaf}(T_0)$ and $\rho_* \sim \mathsf{Leaf}(T_*)$. We consider two cases:

- **Case 1:** $x_i \notin \mathrm{vars}(M)$. Then

$$\Pr_{\rho}[\deg(M|_{\rho}) \geqslant t] = \tfrac{1}{2}\Pr_{\rho_0}[\deg(M|_{\rho_0}) \geqslant t] + \tfrac{1}{2}\Pr_{\rho_*}[\deg(M|_{\rho_*}) \geqslant t] \leqslant \tfrac{1}{2}\cdot 2^{-t} + \tfrac{1}{2}\cdot 2^{-t} = 2^{-t},$$

  by the inductive hypothesis applied to T_0 and T_*, since $\mathrm{vars}(M) \subseteq V\backslash\{x_i\}$.

- **Case 2:** $x_i \in \mathrm{vars}(M)$. With probability 1/2, the restriction ρ sets $x_i = 0$, making $M|_{\rho} = 0$. With the remaining probability 1/2, x_i remains free in ρ, and $M|_{\rho} = x_i \cdot (M'|_{\rho_*})$, where $M' = M/x_i$. Therefore,

$$\Pr_{\rho}[\deg(M|_{\rho}) \geqslant t] = \tfrac{1}{2}\Pr_{\rho_*}[\deg(M'|_{\rho_*}) \geqslant t-1] \leqslant \tfrac{1}{2}\cdot 2^{-(t-1)} = 2^{-t},$$

  again by the inductive hypothesis applied to T_*, since $\mathrm{vars}(M') \subseteq V\backslash\{x_i\}$.

This completes the proof. □

Claim 4.3. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function, and suppose that f admits a one-sided max-degree restriction tree of depth $d \geqslant 40$. Then*

$$\log \widetilde{spar}(f) = \Omega(\sqrt{d}).$$

Proof. Suppose, for the sake of contradiction, that the claim does not hold. Let $k = \sqrt{d/4c}$, where $c > 0$ is a constant to be chosen later. Assume there exists a real polynomial $\tilde{Q}$ that 1/3-approximates f and has sparsity

$$\mathrm{spar}(\tilde{Q}) \leqslant \tfrac{1}{10}\cdot 2^{k}.$$

We will argue that such a polynomial cannot exist, thereby proving the claim.

Let Q be the exact multilinear polynomial representing f, and let T be a one-sided max-degree restriction tree of depth d for Q. Sample a restriction $\rho \sim \mathsf{Leaf}(T)$, and consider the restricted polynomials $Q|_{\rho}$ and $\tilde{Q}|_{\rho}$.

By Claim 4.2, for any fixed monomial in $\tilde{Q}$, the probability that it survives with degree at least k under ρ is at most 2^{-k}. Applying a union bound over all monomials of $\tilde{Q}$, we obtain

$$\Pr_{\rho}\Big[\deg(\tilde{Q}|_{\rho}) \geqslant k\Big] \leqslant \mathrm{spar}(\tilde{Q})\cdot 2^{-k} \leqslant \tfrac{1}{10}.$$

By Claim 4.1, with probability at least 0.9, we have $\deg(Q|_{\rho}) \geqslant d/4$. Hence, with probability at least 0.8, both events hold simultaneously:

$$\deg(Q|_{\rho}) \geqslant d/4 \quad \text{and} \quad \deg(\tilde{Q}|_{\rho}) < k.$$

Fix such a restriction ρ. Then $\tilde{Q}|_{\rho}$ is a polynomial of degree less than k that 1/3-approximates $f|_{\rho}$. We therefore have

$$\deg(f|_{\rho}) = \deg(Q|_{\rho}) \geqslant d/4 = c\cdot k^2 > c\cdot(\deg(\tilde{Q}|_{\rho}))^2 \geqslant c\cdot(\widetilde{\deg}(f|_{\rho}))^2,$$

contradicting the known relationship between degree and approximate degree for Boolean functions—namely, that for all Boolean functions g, $\deg(g) \leqslant c\cdot\widetilde{\deg}(g)^2$ for some universal constant c (see Theorem 2.7).

Hence, our assumption was false, and the claim follows. □

The same reasoning as in the proof above also gives a better lower bound on the exact sparsity of f. Indeed, if f admits a one-sided max-degree restriction tree of depth $d \geqslant 40$, then starting with an exact polynomial $\tilde{Q}$ of small sparsity leads to a similar contradiction. Suppose $\mathrm{spar}(\tilde{Q}) \leqslant \frac{1}{10}\cdot 2^{d/4}$. By the properties of the restriction tree (as established in the proof above), there exists a restriction ρ such that $\deg(f|_{\rho}) \geqslant d/4$ while $\deg(\tilde{Q}|_{\rho}) < d/4$, contradicting the fact that $\tilde{Q}|_{\rho}$ must compute $f|_{\rho}$ exactly. Hence, we obtain the following as well.

Claim 4.4. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function that admits a one-sided max-degree restriction tree of depth $d \geqslant 40$. Then*

$$\log spar(f) = \Omega(d).$$

Claim 4.5. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function, and suppose that f admits a one-sided max-degree restriction tree of depth $d \geqslant 40$. Then*

$$\log \widetilde{wt}(f) = \Omega(\sqrt{d}).$$

Proof. Suppose, for the sake of contradiction, that the claim does not hold. Let $k = (1/c_1) \cdot \sqrt{d/4c}$ for appropriate positive constants c and c_1 to be determined later. Assume there exists a real polynomial $\tilde{Q} = \sum_{S \subseteq [n]} q_S \prod_{i \in S} x_i$ that 1/3-approximates f and has ℓ_1-norm

$$\mathrm{wt}(\tilde{Q}) \leqslant \frac{1}{100} \cdot 2^k.$$

We will argue that such a polynomial cannot exist, thereby proving the claim.

Let Q be the exact multilinear polynomial representing f, and let T be a one-sided max-degree restriction tree of depth d for Q. Sample a restriction $\rho \sim \mathsf{Leaf}(T)$, and consider the restricted polynomials $Q|_\rho$ and $\tilde{Q}|_\rho$.

We analyze the expected ℓ_1-mass of high-degree monomials in $\tilde{Q}|_\rho$. For any polynomial $P = \sum_{S \subseteq [n]} a_S \prod_{i \in S} x_i$, define the degree-$k$ tail of its ℓ_1-norm as

$$\mathrm{wt}(P)^{\geqslant k} := \sum_{\substack{S \subseteq [n] \\ |S| \geqslant k}} |a_S|.$$

Using Claim 4.2, we obtain

$$\mathbb{E}_\rho \left[\mathrm{wt}(\tilde{Q}|_\rho)^{\geqslant k}\right] \leqslant \sum_{\substack{S \subseteq [n] \\ |S| \geqslant k}} |q_S| \cdot \Pr_\rho \left[\deg\left(\prod_{i \in S} x_i|_\rho\right) \geqslant k\right] \leqslant \mathrm{wt}(\tilde{Q}) \cdot 2^{-k} \leqslant \frac{1}{100}.$$

By Markov's inequality, with probability at least 0.9, we have $\mathrm{wt}(\tilde{Q}|_\rho)^{\geqslant k} < 0.1$. Moreover, by Claim 4.1, with probability at least 0.9, $\deg(Q|_\rho) \geqslant d/4$. Therefore, with probability at least 0.8, a random $\rho \sim \mathsf{Leaf}(T)$ satisfies both:

$$\deg(Q|_\rho) \geqslant d/4 \quad \text{and} \quad \mathrm{wt}(\tilde{Q}|_\rho)^{\geqslant k} < 0.1.$$

Fix such a restriction ρ. Let $\bar{Q}$ be the polynomial obtained from $\tilde{Q}|_\rho$ by discarding all monomials of degree at least k. Since $\tilde{Q}|_\rho$ 1/3-approximates $f|_\rho$ and the total weight of the discarded tail is at most 0.1, it follows that $\bar{Q}$ 0.44-approximates $f|_\rho$, with $\deg(\bar{Q}) < k$.

By standard error reduction (see Theorem 2.6), we can boost the success probability of $\bar{Q}$ to obtain a polynomial that 1/3-approximates $f|_\rho$ with degree at most $c_1 k$. Thus, $\widetilde{\deg}(f|_\rho) < c_1 k = \sqrt{d/4c}$. On the other hand, we have $\deg(f|_\rho) = \deg(Q|_\rho) \geqslant d/4$. This contradicts the known relationship between degree and approximate degree for Boolean functions, which asserts that for any Boolean function g, $\deg(g) \leqslant c \cdot \widetilde{\deg}(g)^2$ for some universal constant c (see Theorem 2.7).

Hence, our assumption was false, and the claim follows. □

Theorem 4.6. *For every total Boolean function $f : \{0,1\}^n \to \{0,1\}$,*

$$\log spar(f) = O\big((\log \widetilde{spar}(f))^2 \cdot \log n\big) \quad \textit{and} \quad \log spar(f) = O\big((\log \widetilde{wt}(f))^2 \cdot \log n\big).$$

Proof. By Corollary 3.13, there exists a one-sided max-degree restriction tree for f of depth $d = \Omega\big(\frac{\log(\mathrm{spar}(f))}{\log n}\big)$. Assume $d \geqslant 40$, as the claim is trivial otherwise. Applying Claim 4.3 and Claim 4.5, we obtain

$$\log \widetilde{\mathrm{spar}}(f) = \Omega(\sqrt{d}) = \Omega\left(\sqrt{\frac{\log \mathrm{spar}(f)}{\log n}}\right),$$

$$\log \widetilde{\mathrm{wt}}(f) = \Omega(\sqrt{d}) = \Omega\left(\sqrt{\frac{\log \mathrm{spar}(f)}{\log n}}\right). \quad \square$$

Theorem 4.7. *For every total Boolean function $f : \{0,1\}^n \to \{0,1\}$, we have*

$$\log wt(f) = O\big((\log \widetilde{spar}(f))^2 \cdot \log n\big) . \quad \text{and} \quad \log wt(f) = O\Big((\log \widetilde{wt}(f))^2 \cdot \log n\Big)$$

Proof. By Corollary 3.14, there exists a one-sided max-degree restriction tree for f of depth $d = \Omega\big(\frac{\log \mathrm{wt}(f)}{\log n}\big)$. Assume $d \geqslant 40$, as the claim is trivial otherwise. Applying Claim 4.3 and Claim 4.5, we obtain

$$\log \widetilde{\mathrm{spar}}(f) = \Omega(\sqrt{d}) = \Omega\Big(\sqrt{\frac{\log \mathrm{wt}(f)}{\log n}}\Big),$$

$$\log \widetilde{\mathrm{wt}}(f) = \Omega(\sqrt{d}) = \Omega\Big(\sqrt{\frac{\log \mathrm{wt}(f)}{\log n}}\Big).$$

□

Thus, for every Boolean function f, the measures $\mathrm{spar}(f)$, $\widetilde{\mathrm{spar}}(f)$, $\mathrm{wt}(f)$, and $\widetilde{\mathrm{wt}}(f)$ are all polynomially related on the logarithmic scale, up to poly-logarithmic factors in n, thereby proving Theorem 1.1 and Theorem 1.5.

4.2 Generalized Polynomials

Analogous to Section 4.1, where we used one-sided restriction trees, here we study the properties of restrictions sampled uniformly from the leaves of two-sided restriction trees. Using these properties, we then relate the complexity measures associated with generalized polynomial representations.

Claim 4.8. *Let T be a two-sided max-degree restriction tree of depth $d \geqslant 60$ for a multilinear polynomial Q. Then*

$$\Pr_{\rho \sim \mathsf{Leaf}(T)}[deg(Q|_\rho) \geqslant d/6] \geqslant 0.9.$$

Proof. Since T is a max-degree restriction tree for Q, for every leaf restriction ρ we have $\deg(Q|_\rho) = |\rho|_*$. Hence, it suffices to show that $|\rho|_* \geqslant d/6$ with high probability.

Sampling ρ uniformly from the leaves of T is equivalent to performing a random root-to-leaf walk, where each step independently follows a $*$-edge with probability $1/3$, and outputting the restriction labeling the leaf reached. The number of free variables in the resulting restriction equals the number of $*$-edges encountered, implying that $\mathbb{E}[|\rho|_*] = d/3$. Applying a Chernoff bound,

$$\Pr[|\rho|_* \leqslant d/4] \leqslant e^{-d/24} \leqslant 0.1,$$

where the last inequality uses $d \geqslant 60$. □

Claim 4.9. *Let T be a two-sided restriction tree over variables V. Then, for any generalized monomial M with $vars(M) \subseteq V$ and any $t \in \mathbb{N}$,*

$$\Pr_{\rho \sim \mathsf{Leaf}(T)}[\, deg(M|_\rho) \geqslant t\,] \leqslant 2^{-t}.$$

Proof. We prove the claim by induction on the depth d of the restriction tree T.

Base Case ($d = 0$): When T has depth zero, it consists of a single leaf labeled with a restriction that assigns all variables in V. Hence, M becomes a constant under this restriction, and the claim holds trivially.

Inductive Step ($d > 0$): Let the root of T query variable x_i, and let the 0, 1 and $*$-edges lead to subtrees T_0, T_1 and T_*, which are two-sided restriction trees on $V\backslash\{x_i\}$ of depth $d-1$. Let $\rho \sim \mathsf{Leaf}(T)$, and define $\rho_0 \sim \mathsf{Leaf}(T_0)$, $\rho_1 \sim \mathsf{Leaf}(T_0)$ and $\rho_* \sim \mathsf{Leaf}(T_*)$. We consider two cases:

- **Case 1:** M does not contain x_i or $\bar{x}_i$: Then

$$\Pr_{\rho}[\deg(M|_\rho) \geqslant t] = \tfrac{1}{3}\Pr_{\rho_0}[\deg(M|_{\rho_0}) \geqslant t] + \tfrac{1}{3}\Pr_{\rho_1}[\deg(M|_{\rho_1}) \geqslant t] + \tfrac{1}{3}\Pr_{\rho_*}[\deg(M|_{\rho_*}) \geqslant t] \leqslant 2^{-t}.$$

  by the inductive hypothesis applied to T_0, T_1 and T_*, since $\mathrm{vars}(M) \subseteq V\backslash\{x_i\}$.

- **Case 2:** M contains x_i or $\bar{x}_i$: Without loss of generality, suppose $x_i \in M$ (the case $\bar{x}_i \in M$ is analogous). Let $M' = M/x_i$. Then:

$$\Pr_{\rho}[\deg(M|_\rho) \geqslant t] = \frac{1}{3} \cdot 0 + \frac{1}{3} \Pr_{\rho_1}[\deg(M'|_{\rho_1}) \geqslant t] + \frac{1}{3} \Pr_{\rho*}[\deg(M'|_{\rho*}) \geqslant t-1],$$

where the first term is zero because setting $x_i \leftarrow 0$ kills the monomial. Using the inductive hypothesis:

$$\Pr_{\rho}[\deg(M|_\rho) \geqslant t] \leqslant \frac{1}{3} \cdot 2^{-t} + \frac{1}{3} \cdot 2^{-(t-1)} = 2^{-t}.$$

This completes the proof. □

With the properties of restrictions shown above, it must be evident that the existence of a deep two-sided restriction tree implies large values of $\text{gspar}(f)$ and $\widetilde{\text{gwt}}(f)$. The argument proceeds analogously to that of Claim 4.3 and Claim 4.5, with the only difference being that we invoke Claim 4.8 and Claim 4.9 in place of Claim 4.1 and Claim 4.2.

Combining this with Corollary 3.20 and Corollary 3.21, which establish the existence of deep two-sided restriction trees from large exact generalized measures, we obtain relations between the exact and approximate versions of these measures. Since the proofs follow almost identical lines to those in Section 4.1, we omit the details and state the results below.

Claim 4.10. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function, and suppose that f admits a two-sided max-degree restriction tree of depth $d \geqslant 60$. Then*

$$\log \widetilde{gspar}(f) = \Omega(\sqrt{d}). \quad \text{and} \quad \log gspar(f) = \Omega(d).$$

Claim 4.11. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function, and suppose that f admits a two-sided max-degree restriction tree of depth $d \geqslant 60$. Then*

$$\log \widetilde{gwt}(f) = \Omega(\sqrt{d}).$$

Theorem 4.12. *For every total Boolean function $f : \{0,1\}^n \to \{0,1\}$, we have*

$$\log gspar(f) = O\Big((\log(\widetilde{gspar}(f)))^2 \cdot \log n\Big) \quad \text{and} \quad \log gspar(f) = O\Big((\log \widetilde{gwt}(f))^2 \cdot \log n\Big).$$

Theorem 4.13. *For every total Boolean function $f : \{0,1\}^n \to \{0,1\}$, we have*

$$\log gwt(f) = O\Big((\log(\widetilde{gspar}(f)))^2 \cdot \log n\Big) \quad \text{and} \quad \log gwt(f) = O\Big((\log \widetilde{gwt}(f))^2 \cdot \log n\Big).$$

Thus, for every Boolean function f, the measures $gspar(f)$, $\widetilde{gspar}(f)$, $gwt(f)$, and $\widetilde{gwt}(f)$ are all polynomially related on the logarithmic scale, up to poly-logarithmic factors in n, thereby proving Theorem 1.7.

4.3 Discussion on the Optimality of Our Results

Optimality of the bounds in Theorem 4.6 and Theorem 4.7. Our results in Theorem 4.6 and Theorem 4.7 are optimal up to polynomial factors in $\log n$, as witnessed by the OR_n function. The exact sparsity and ℓ_1-norm of OR_n are both $2^n - 1$. Since its approximate degree is $\Theta(\sqrt{n})$ [NS94], the approximate sparsity is at most $n^{O(\sqrt{n})}$. Moreover, the approximate ℓ_1-norm of OR_n is also bounded by $n^{O(\sqrt{n})}$, as shown using a standard Chebyshev polynomial approximator.

Observation 4.14 ($\widetilde{\text{wt}}(\text{OR}_n) \leqslant n^{O(\sqrt{n})}$). *Let T_d denote the degree-d Chebyshev polynomial defined recursively by $T_0(z) = 1$, $T_1(z) = z$, and $T_d(z) = 2zT_{d-1}(z) - T_{d-2}(z)$ for $d \geqslant 2$. For $d = 2\sqrt{n}$, define*

$$p(z) = 1 - \frac{T_d\left(\frac{n-z}{n-1}\right)}{T_d\left(\frac{n}{n-1}\right)}, \quad \text{and} \quad q(x_1, \ldots, x_n) = p\left(\sum_{i=1}^{n} x_i\right).$$

Then q 1/3-approximates OR_n (see [NS94, Example 2]). Furthermore, using the recursive definition, it is easy to verify that the coefficients of T_d are bounded in absolute value by 3^d. Therefore, for $d = 2\sqrt{n}$, the ℓ_1-norm of q is at most $n^{O(\sqrt{n})}$.

Thus, $\log \mathrm{spar}(\textsc{Or}_n) = \Theta(n), \log \mathrm{wt}(\textsc{Or}_n) = \Theta(n), \log \widetilde{\mathrm{spar}}(\textsc{Or}_n) = O(\sqrt{n}\log n)$, and $\log \widetilde{\mathrm{wt}}(\textsc{Or}_n) = O(\sqrt{n}\log n)$, showing that the bounds in Theorem 4.6 and Theorem 4.7 are essentially tight up to polynomial factors in $\log n$.

The dependence on n is also unavoidable. Consider the function $\textsc{Thr}^n_{n-1} : \{0,1\}^n \to \{0,1\}$, defined as $\textsc{Thr}^n_{n-1}(x) = 1$ iff $|x| \geqslant n-1$, namely, the function evaluates to 1 if the input has at most one zero. Its exact sparsity is $n+1$, and its exact ℓ_1-norm is $2n-1$, via

$$\textsc{Thr}^n_{n-1}(x) = \sum_{\substack{S\subseteq[n]\\|S|=n-1}} \prod_{i\in S} x_i - (n-1)\prod_{i\in[n]} x_i.$$

In contrast, we show that the approximate sparsity of $\textsc{Thr}^n_{n-1}$ is only $O(\log n)$, while its approximate ℓ_1-norm is $O(1)$. This implies that an additive $O(\log n)$ or multiplicative $O\left(\frac{\log n}{\log\log n}\right)$ factor in Theorems 4.6 and 4.7 is unavoidable for bounds involving approximate sparsity, and a $\log n$ factor—either additive or multiplicative—is similarly necessary for those involving the approximate ℓ_1-norm.

Claim 4.15. *$\widetilde{spar}(\textsc{Thr}^n_{n-1}) = O(\log n)$ and $\widetilde{wt}(\textsc{Thr}^n_{n-1}) = O(1)$.*

Proof. We construct an approximator for $\textsc{Thr}^n_{n-1}$ with sparsity $O(\log n)$ and ℓ_1-norm $O(1)$. The construction relies on a combinatorial notion we call a *separating collection.*

Separating collections. Let $\{i,j\} \in \binom{[n]}{2}$ be an unordered pair of distinct indices. A set $S \subseteq [n]$ is said to *separate* $\{i,j\}$ if exactly one of i or j belongs to S. A pair $(S_1,S_2) \in 2^{[n]} \times 2^{[n]}$ is said to separate $\{i,j\}$ if at least one of S_1 or S_2 separates it.

We say that a collection $F \subseteq 2^{[n]} \times 2^{[n]}$ is *δ-separating* if, for every pair $\{i,j\} \in \binom{[n]}{2}$, at least a δ-fraction of the elements in F separate it. Formally,

$$\forall \{i,j\} \in \binom{[n]}{2}, \quad |\{(S_1,S_2)\in F : (S_1,S_2) \text{ separates } \{i,j\}\}| \geqslant \delta|F|.$$

We will show that there exists a $2/3$-separating collection F of size $O(\log n)$. Assuming such a collection exists, we describe a sparse approximating polynomial for $\textsc{Thr}^n_{n-1}$.

Approximator construction. Let $F \subseteq 2^{[n]} \times 2^{[n]}$ be a $2/3$-separating collection of size $O(\log n)$. For each pair $(S_1,S_2) \in F$, define

$$f_{(S_1,S_2)}(x) = \left(1-(1-\prod_{i\in S_1} x_i)(1-\prod_{i\notin S_1} x_i)\right)\cdot\left(1-(1-\prod_{i\in S_2} x_i)(1-\prod_{i\notin S_2} x_i)\right).$$

Each function $f_{(S_1,S_2)}$ evaluates to 1 if the input $x \in \{0,1\}^n$ contains at most one zero, and evaluates to 0 if, for some $S \in \{S_1,S_2\}$, the input x contains zeros in both S and its complement. Define

$$g(x) := \frac{1}{|F|}\sum_{(S_1,S_2)\in F} f_{(S_1,S_2)}(x).$$

We claim that g is a $1/3$-approximator for $\textsc{Thr}^n_{n-1}$.

- If x is a 1-input, i.e., x has at most one zero, then for every $S \subseteq [n]$, either $\prod_{i\in S} x_i = 1$ or $\prod_{i\notin S} x_i = 1$. Thus, each term $f_{(S_1,S_2)}(x) = 1$, so $g(x) = 1$.
- If x is a 0-input, i.e., it contains at least two zeros, let $i,j \in [n]$ be distinct positions where $x_i = x_j = 0$. For any $(S_1,S_2) \in F$ that separates $\{i,j\}$, one of S_1 or S_2 contains exactly one of i, j, so one of the products in the corresponding $f_{(S_1,S_2)}(x)$ vanishes, and hence $f_{(S_1,S_2)}(x) = 0$. Since F is $2/3$-separating, at least $2/3$ of the terms in the sum are 0, so $g(x) \leqslant 1/3$.

Hence, g is a $1/3$-approximator for $\textsc{Thr}^n_{n-1}$. Each $f_{(S_1,S_2)}$ has constant sparsity, and there are $O(\log n)$ such terms, so g has total sparsity $O(\log n)$. Moreover, as g is an average of constant-ℓ_1-norm functions, its ℓ_1-norm is $O(1)$.

Existence of separating collections. It remains to show that a 2/3-separating collection of size $O(\log n)$ exists. We do this via the probabilistic method.

Let $t = 216\ln(n^2) = O(\log n)$, and sample $F = \{(S_1^{(k)}, S_2^{(k)})\}_{k=1}^t$, where each set $S_u^{(k)} \subseteq [n]$ (for $u \in \{1,2\}$) is formed by including each element independently with probability 1/2. We show that with positive probability, F is 2/3-separating.

Fix a pair $\{i,j\} \in \binom{[n]}{2}$, and let X_i be the indicator that $(S_1^{(i)}, S_2^{(i)})$ separates $\{i,j\}$. Each X_i has expectation $\mathbb{E}[X_i] = 3/4$, so the sum $\sum_{i=1}^t X_i$ has expectation $\frac{3t}{4}$. By a Chernoff bound,

$$\Pr\left[\sum_{i=1}^t X_i \leqslant \frac{2t}{3}\right] \leqslant e^{-t/216} \leqslant \frac{1}{n^2}.$$

Taking a union bound over all $\binom{n}{2} < n^2$ pairs, the probability that F fails to be 2/3-separating for some pair is less than 1/2. Hence, with positive probability, a 2/3-separating set F of size $t = O(\log n)$ exists. □

Optimality of the bounds in Theorem 4.12 and Theorem 4.13. Our results in Theorem 4.12 and Theorem 4.13 are also optimal up to polynomial factors in $\log n$. This optimality is witnessed by the function $\textsc{And}_n \circ \textsc{Or}_2$.

The approximate sparsity and approximate ℓ_1-norm of $\textsc{And}_n \circ \textsc{Or}_2$ are both $n^{O(\sqrt{n})}$. To see this, take an approximating polynomial for $\textsc{And}_n$ of degree $\Theta(\sqrt{n})$ [NS94] and replace each variable with the exact polynomial for $\textsc{Or}_2$. The resulting polynomial approximates $\textsc{And}_n \circ \textsc{Or}_2$ and has approximate sparsity $n^{O(\sqrt{n})}$. Moreover, since the degree-$\Theta(\sqrt{n})$ approximator for $\textsc{And}_n$ has ℓ_1-norm $n^{O(\sqrt{n})}$ (similar to the argument in Observation 4.14), the constructed polynomial inherits the same asymptotic ℓ_1-norm. Hence, both the approximate generalized sparsity and approximate generalized ℓ_1-norm of $\textsc{And}_n \circ \textsc{Or}_2$ are $n^{O(\sqrt{n})}$.

To show the optimality of our results, we show that the exact generalized sparsity and generalized ℓ_1-norm of $\textsc{And}_n \circ \textsc{Or}_2$ are $2^{\Omega(n)}$.

Claim 4.16. *gspar*$(\textsc{And}_n \circ \textsc{Or}_2) = 2^{\Omega(n)}$.

Proof. As observed in Section 1.1, the function $\textsc{And}_n \circ \textsc{Or}_2$ admits a two-sided max-degree restriction tree of depth n. Applying Claim 4.4 then yields the claimed lower bound. □

Claim 4.17. *gwt*$(\textsc{And}_n \circ \textsc{Or}_2) = 2^{\Omega(n)}$.

Proof. We first observe that the Fourier ℓ_1-norm is always a lower bound on the generalized ℓ_1-norm, i.e., $\left\|\widehat{f}\right\|_1 \leqslant \mathrm{gwt}(f)$. This follows since every generalized monomial (an $\textsc{And}$ of literals) can be expressed in the Fourier basis with ℓ_1-norm 1. Hence, it suffices to show that $\left\|\widehat{(\textsc{And}_n \circ \textsc{Or}_2)}\right\|_1$ is large. We have

$$\left\|\widehat{(\textsc{And}_n \circ \textsc{Or}_2)}\right\|_1 \overset{(1)}{=} \left(\left\|\widehat{\textsc{Or}_2}\right\|_1\right)^n \overset{(2)}{=} (3/2)^n,$$

where (1) follows from the multiplicativity of the Fourier ℓ_1-norm for functions over disjoint variable sets, and (2) uses the fact that $\left\|\widehat{\textsc{Or}_2}\right\|_1 = 3/2$. Therefore,

$$\mathrm{gwt}(\textsc{And}_n \circ \textsc{Or}_2) \geqslant \left\|\widehat{(\textsc{And}_n \circ \textsc{Or}_2)}\right\|_1 = (3/2)^n = 2^{\Omega(n)}.$$ □

The dependence on n is also unavoidable in Theorem 4.6 and Theorem 4.7. Again, witnessed by the $\textsc{Thr}^n_{n-1}$. By Claim 4.15, we have $\widetilde{\mathrm{gspar}}(\textsc{Thr}^n_{n-1}) = O(\log n)$ and $\widetilde{\mathrm{gwt}}(\textsc{Thr}^n_{n-1}) = O(1)$. We now show that both $\log \mathrm{gspar}(\textsc{Thr}^n_{n-1})$ and $\log \mathrm{gwt}(\textsc{Thr}^n_{n-1})$ are $\Omega(\log n)$.

Claim 4.18. *gspar*$(\textsc{Thr}^n_{n-1}) = \Omega(n)$.

Proof. Let Q be the generalized polynomial computing $\textsc{Thr}^n_{n-1}$. We claim that its sparsity must be at least $n/2$. Suppose not, and let there be k generalized monomials in Q that contain at least one negated literal, with $k < n/2$. We can eliminate all such monomials by fixing at most k variables to 1: for each such monomial, choose one variable that appears negated and set it to 1. Let ρ be the restriction that sets these $\ell \leqslant k$ variables to 1. Then $Q|_\rho$ is an ordinary polynomial computing $\textsc{Thr}^{n-\ell}_{n-\ell-1}$. However, any ordinary polynomial computing $\textsc{Thr}^{n-\ell}_{n-\ell-1}$ must have sparsity $n - \ell + 1 \geqslant n - k + 1 > n/2$, a contradiction. ☐

To show that $\log \mathrm{gwt}(\textsc{Thr}^n_{n-1}) = \Omega(\log n)$, we observe, by a simple calculation, that $\log \left\| \widehat{\textsc{Thr}^n_{n-1}} \right\|_1 = \Theta(\log n)$ (this also follows from [AFH12b, Theorem 1.1]). Since the Fourier ℓ_1-norm lower bounds the generalized ℓ_1-norm, it follows that $\log \mathrm{gwt}(\textsc{Thr}^n_{n-1}) = \Omega(\log n)$.

4.4 Implications for the And Query Model

The measure $\log \mathrm{spar}(f)$ naturally connects to the And-query model—a variant of the standard decision tree model where each query computes the And of an arbitrary subset of input bits. Just as polynomial degree characterizes ordinary deterministic query complexity up to polynomial loss, Knop et al. [KLMY21b] showed that $\log \mathrm{spar}(f)$ characterizes deterministic query complexity in the And-query model, up to polynomial loss and poly-logarithmic factors in n.

In the randomized setting, it is easy to see that $\log \widetilde{\mathrm{spar}}(f)$ lower bounds randomized And-query complexity. Let $\mathrm{R}^{\wedge dt}(f)$ denote the randomized And-query complexity of f. The following is easy to verify:

Claim 4.19 ([KLMY21a, Claim 3.20]). *For every total Boolean function $f : \{0,1\}^n \to \{0,1\}$, we have*

$$\log \widetilde{spar}(f) = O(R^{\wedge dt}(f) + \log n).$$

However, it was unknown whether it also characterizes the randomized query complexity up to polynomial loss. Our results, combined with those of [KLMY21b], establish that this is indeed the case.

Knop et al. [KLMY21b] showed that for any Boolean function f,

$$\mathrm{D}^{\wedge dt}(f) = O\left((\log \mathrm{spar}(f))^5 \cdot \log n\right),$$

which, when combined with Theorem 1.1, implies

$$\mathrm{R}^{\wedge dt}(f) \leqslant \mathrm{D}^{\wedge dt}(f) = O\left((\log \widetilde{\mathrm{spar}}(f))^{10} \cdot (\log n)^6\right).$$

A tighter bound can be obtained using a structural result of Knop et al., which relates deterministic And-query complexity to sparsity and a combinatorial measure called *monotone block sensitivity*:

Definition 4.20 (Monotone Block Sensitivity). *The* monotone block sensitivity *of a Boolean function $f : \{0,1\}^n \to \{0,1\}$, denoted $MBS(f)$, is a variant of block sensitivity that only considers flipping 0's to 1's. A subset $B \subseteq [n]$ is called a* sensitive 0-block of f at input x *if $x_i = 0$ for all $i \in B$, and $f(x) \neq f(x \oplus 1_B)$, where $x \oplus 1_B$ denotes the input obtained by flipping all bits in B from 0 to 1. For an input $x \in \{0,1\}^n$, let $MBS(f,x)$ denote the maximum number of pairwise disjoint sensitive 0-blocks of f at x. Then, $MBS(f) = \max_{x \in \{0,1\}^n} MBS(f,x)$.*

Claim 4.21 ([KLMY21b, Lemma 3.2, Claim 4.4, Lemma 4.6]). *For any Boolean function f,*

$$D^{\wedge dt}(f) = O\left((\log MBS(f))^2 \cdot \log spar(f) \cdot \log n\right).$$

Intuitively, a large value of $\mathrm{MBS}(f)$ indicates that a large-arity promise-Or function can be embedded into f via suitable restrictions and identifications of variables.

To tighten our upper bound on $\mathrm{R}^{\wedge dt}(f)$, we now upper bound $\mathrm{MBS}(f)$ in terms of $\log \widetilde{\mathrm{spar}}(f)$. While Knop et al. showed $\mathrm{MBS}(f) = O((\log \mathrm{spar}(f))^2)$, the same proof idea gives a similar bound in terms of approximate sparsity. We use the following classical result of Nisan and Szegedy [NS94], which lower bounds the approximate degree via monotone block sensitivity:

Theorem 4.22 (Nisan and Szegedy [NS94]). *For any Boolean function $f : \{0,1\}^n \to \{0,1\}$,*

$$\widetilde{deg}(f) \geqslant \sqrt{\frac{MBS(f)}{6}}.$$

Remark 4.23. *The original result of Nisan and Szegedy lower bounds the approximate degree in terms of the* block sensitivity *of f, which considers all sensitive blocks, not just sensitive 0-blocks. Here, we only apply it to the monotone variant MBS(f) defined above.*

Claim 4.24. *For any Boolean function f,*

$$MBS(f) = O\left((\log \widetilde{spar}(f))^2\right).$$

Proof. Assume $\mathrm{MBS}(f) = k \geqslant 40$; otherwise, the claim is trivial. Let this be witnessed by an input $z \in \{0,1\}^n$ and disjoint 0-blocks $B_1, \ldots, B_k \subseteq [n]$, such that $f(z) \neq f(z \oplus 1_{B_i})$ for all $i \in [k]$.

Define $g : \{0,1\}^k \to \{0,1\}$ by identifying variables within each B_i, fixing all others according to z, and letting g be the resulting function. Then $g(0^k) = f(z)$ and $g(x) \neq f(z)$ for all x with Hamming weight 1. Thus, g has sensitivity k at 0^k. Since restrictions and identifications do not increase approximate sparsity, we have $\widetilde{\mathrm{spar}}(g) \leqslant \widetilde{\mathrm{spar}}(f)$, so it suffices to show $\widetilde{\mathrm{spar}}(g)$ is large.

Suppose, for contradiction, that g is 1/3-approximated by a polynomial Q of sparsity

$$\mathrm{spar}(Q) \leqslant \frac{1}{10} \cdot 2^{\ell},$$

for $\ell = c \cdot \sqrt{k/4}$, where $c > 0$ is a constant to be fixed later. We will argue that such a polynomial cannot exist, thereby proving the claim.

Define a distribution $\mathcal{D}$ over restrictions $\rho : \{x_1, \ldots, x_k\} \to \{0, *\}$, where each variable is independently set to 0 with probability 1/2 and left free with probability 1/2. This distribution satisfies the following:

1. By a standard Chernoff bound,

$$\Pr_{\rho}\left[|\rho|_* \leqslant k/4\right] \leqslant e^{-k/16} \leqslant 0.1,$$

   where the last inequality uses $k \geqslant 40$. Thus, with probability at least 0.9, $|\rho|_* \geqslant k/4$.

2. For every ρ in the support of $\mathcal{D}$, the restricted function $g|_\rho$ has sensitivity $|\rho|_*$ at the all-zero input. Hence, by Theorem 4.22, $\widetilde{\deg}(g|_\rho) \geqslant c \cdot \sqrt{|\rho|_*}$.

3. For any monomial M over $\{x_1, \ldots, x_k\}$, we have $\Pr_\rho[\deg(M|_\rho) > 0] = 2^{-\deg(M)}$.

Now consider the restricted polynomial $Q|_\rho$. By property (3) of $\mathcal{D}$, the probability that any fixed monomial in Q of degree at least ℓ survives is at most $2^{-\ell}$, so by a union bound:

$$\Pr_{\rho}[\deg(Q|_\rho) \geqslant \ell] \leqslant \mathrm{spar}(Q) \cdot 2^{-\ell} \leqslant \frac{1}{10}.$$

By property (1) of $\mathcal{D}$, with probability at least 0.9, ρ leaves at least $k/4$ variables free. Thus, with probability at least 0.8, both of the following hold:

$$|\rho|_* \geqslant k/4 \quad \text{and} \quad \deg(Q|_\rho) < \ell.$$

Fix such a restriction ρ. Then $Q|_\rho$ is a polynomial of degree less than ℓ that 1/3-approximates $f|_\rho$. Hence, $\widetilde{\deg}(f|_\rho) < \ell = c \cdot \sqrt{k/4} \leqslant c \cdot \sqrt{|\rho|_*}$, contradicting property (2) of $\mathcal{D}$. Hence, our assumption was false, and the claim follows. □

Theorem 1.4 (Restated)**.** *For every total Boolean function $f : \{0,1\}^n \to \{0,1\}$, the following holds:*

$$\Omega(\log(\widetilde{spar}(f)) - \log n) \overset{(1)}{=} R^{\wedge dt}(f) \leqslant D^{\wedge dt}(f) \overset{(2)}{=} O\left((\log \widetilde{spar}(f))^6 \cdot \log n\right).$$

Proof. The bound in (1) follows from Claim 4.19. For (2), combining Claim 4.21, Claim 4.24, and Theorem 1.1, we get:

$$\mathrm{D}^{\wedge dt}(f) = O\left((\log \mathrm{MBS}(f))^2 \cdot \log \mathrm{spar}(f) \cdot \log n\right) = O\left((\log \widetilde{\mathrm{spar}}(f))^6 \cdot \log n\right).$$ □

This parallels the classical setting, where deterministic and randomized query complexity, degree, and approximate degree are all polynomially related. In the AND-query model, $\log \mathrm{spar}(f)$ plays the role of degree, while $\log \widetilde{\mathrm{spar}}(f)$ plays the role of approximate degree. Combined with the results of [KLMY21b], our work shows that deterministic and randomized AND-query complexities, log sparsity, and log approximate sparsity are all polynomially related—up to poly-logarithmic loss factors.

5 Lifting with Equality Gadget

In this section, we lift the sparsity measure of a Boolean function $f : \{0,1\}^n \to \{0,1\}$ to the approximate rank of the lifted function $f \circ \mathrm{EQ}_4$, where EQ_4 denotes the two-party equality function on 2-bit inputs.

This lifting has several consequences. For instance, the *log-approximate-rank conjecture* holds up to poly-logarithmic factors in n for EQ_4-lifted functions. The randomized and quantum communication complexities of such functions are polynomially related (again, up to poly-logarithmic factors in n). Moreover, the gadget EQ_4 is, in a sense, weak—since for some f, the randomized and deterministic communication complexities of $f \circ \mathrm{EQ}_4$ are exponentially separated, as the equality function itself belongs to this class.

Our proof proceeds in two stages. We first lift the sparsity of f to the approximate Fourier ℓ_1-norm of $f \circ \textsc{And}_2$. We then apply known results connecting the approximate Fourier ℓ_1-norm of $f \circ \textsc{And}_2$ to the approximate rank of $(f \circ \textsc{And}_2) \circ \textsc{Xor}_2$. Since the gadgets $\textsc{And}_2 \circ \textsc{Xor}_2$ and EQ_4 are equivalent, this yields the desired lifting from sparsity to approximate rank. The main technical contribution of this section is therefore a lifting theorem relating the sparsity of f to the approximate Fourier ℓ_1-norm of $f \circ \textsc{And}_2$.

Proof Overview. We outline the proof that lifts the sparsity of f to the approximate Fourier sparsity of $f \circ \textsc{And}_2$; the same ideas extend to the approximate Fourier ℓ_1-norm.

Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function with large sparsity. We aim to show that $f \circ \textsc{And}_2$ also has large approximate Fourier sparsity. The proof proceeds via a random restriction argument designed to simplify Fourier monomials while preserving the "hardness" of $f \circ \textsc{And}_2$ under the restriction (in terms of degree).

Unlike ordinary (or generalized) monomials, Fourier monomials are not "killed" by simply fixing variables. However, since we are working with the lifted function $f \circ \textsc{And}_2$, we can mimic this effect by constructing restrictions that *mask* certain variables. Specifically, we construct random restrictions for $f \circ \textsc{And}_2$ by lifting restrictions from a one-sided max-degree restriction tree T for f, which exists due to the large sparsity of f.

Formally, for $f : \{0,1\}^n \to \{0,1\}$ on inputs $(z_1, \ldots, z_n)$, we define the lifted function

$$(f \circ \textsc{And}_2)(x_1, \ldots, x_n, y_1, \ldots, y_n) = f(\textsc{And}_2(x_1, y_1), \ldots, \textsc{And}_2(x_n, y_n)).$$

To sample a restriction ρ for $f \circ \textsc{And}_2$, we first sample $\rho_f \sim \mathsf{Leaf}(T)$ and lift it as follows:

$$\rho(z_i) = \begin{cases} 1 \;\Rightarrow\; (x_i, y_i) = (1, 1), \\ * \;\Rightarrow\; (x_i, y_i) = (*, 1), \\ 0 \;\Rightarrow\; (x_i, y_i) = (\Delta, 0), \end{cases}$$

where both $*$ and Δ denote free variables—Δ being a *masked variable.* Note that $(f \circ \textsc{And}_2)|_\rho = f|_{\rho_f}$, and the restricted function $(f \circ \textsc{And}_2)|_\rho$ is independent of masked variables. Algorithm 3 formalizes this sampling process.

By construction and properties of max-degree restrictions, the sampled ρ satisfies:

1. With probability at least 0.9, $\deg((f \circ \textsc{And}_2)|_\rho) \geqslant d = \Omega\big(\frac{\log \mathrm{spar}(f)}{\log n}\big)$;

2. For any Fourier monomial $\chi_S(x)$ on the x-variables, the probability that k of its variables are assigned $*$ and none are masked is at most 2^{-k}.

These properties together imply that the approximate Fourier sparsity of $f \circ \textsc{And}_2$ must be large. Suppose, toward a contradiction, that there exists a polynomial $\tilde{Q}$ that 1/3-approximates $f \circ \textsc{And}_2$ with Fourier sparsity smaller than $\frac{1}{10} \cdot 2^{\sqrt{d/c}}$ for some constant $c > 0$. Under ρ, the y-part of each Fourier term becomes constant. By Property (2) and a union bound, with probability at least 0.9, every Fourier monomial in $\tilde{Q}$ of degree at least $\sqrt{d/c}$ must contain at least one masked variable. Meanwhile, by Property (1), with probability at least 0.9, $\deg((f \circ \textsc{And}_2)|_\rho) \geqslant d$. Hence, there exists a restriction ρ satisfying both conditions.

Fix such a ρ. Then $\tilde{Q}|_\rho$ approximates $(f \circ \textsc{And}_2)|_\rho$, and all its high-degree terms involve at least one masked variable. Since $(f \circ \textsc{And}_2)|_\rho$ is independent of masked variables, taking the expectation of $\tilde{Q}|_\rho$

over masked variables set uniformly to 0/1 removes all high-degree Fourier terms, yielding a low-degree polynomial Q' of degree $< \sqrt{d/c}$ that still 1/3-approximates $(f \circ \textsc{And}_2)|_\rho$. This contradicts the known relationship $\deg(g) \leqslant c \cdot \widetilde{\deg}(g)^2$ for all Boolean functions g [ABDK+21].

Therefore, any polynomial approximating $f \circ \textsc{And}_2$ must have Fourier sparsity at least $\frac{1}{10} \cdot 2^{\sqrt{d/4c}}$. A similar argument also yields a lower bound on the approximate Fourier ℓ_1-norm.

5.1 Lifting De Morgan Sparsity to the Approximate Fourier ℓ_1-Norm via the $\textsc{And}_2$ Gadget

We now formalize the random restriction procedure introduced in the proof overview. Given a one-sided max-degree restriction tree T for $f : \{0,1\}^n \to \{0,1\}$, Algorithm 3 describes how to sample lifted random restrictions for $f \circ \textsc{And}_2$.

Algorithm 3 LiftedRestriction

1: **Input:** One-sided max-degree restriction tree T for $f : \{0,1\}^n \to \{0,1\}$
2: **Output:** Restriction $\rho : \{x_i, y_i\}_{i=1}^n \to \{0, 1, *, \Delta\}$
3: Sample $\rho_f \sim \mathsf{Leaf}(T)$ over variables $(z_1, \ldots, z_n)$
4: **for** each $i \in [n]$ **do**
5: Set $(\rho(x_i), \rho(y_i)) = \begin{cases} (\Delta, 0), & \text{if } \rho_f(z_i) = 0, \\ (1,1), & \text{if } \rho_f(z_i) = 1, \\ (*, 1), & \text{if } \rho_f(z_i) = *. \end{cases}$
6: **end for**
7: **return** ρ

A restriction sampled by LiftedRestriction assigns each variable a value in $\{0, 1, *, \Delta\}$. When applied to $f \circ \textsc{And}_2$, the symbols $*$ and Δ are treated as free variables. We refer to variables marked by Δ as *masked variables*, distinguishing them from ordinary free variables, since the restricted function does not depend on them. It will be convenient to keep track of these masked variables separately in the subsequent analysis.

From the construction of LiftedRestriction and the properties of max-degree restriction trees, we now establish Properties (1) and (2) stated in the proof overview.

Claim 5.1. *Let T be a one-sided max-degree restriction tree of depth $d \geqslant 40$ for $f : \{0,1\}^n \to \{0,1\}$. Then:*

1. *For any ρ in the support of* LiftedRestriction(T)*, let $M = \{x_i \mid \rho(x_i) = \Delta\}$ be the set of masked variables. Then the restricted function $(f \circ \textsc{And}_2)|_\rho$ does not depend on the variables in M.*
2. *For any ρ in the support of* LiftedRestriction(T)*, $\deg\big((f \circ \textsc{And}_2)|_\rho\big) = |\rho^{-1}(*)|$.*
3. *With probability at least 0.9 over $\rho \sim$* LiftedRestriction(T)*, $\deg\big((f \circ \textsc{And}_2)|_\rho\big) \geqslant d/4$.*

Proof. We verify each:

1. If x_i is masked in ρ (i.e., $\rho(x_i) = \Delta$), then by construction $\rho(y_i) = 0$. Hence, $\textsc{And}_2(x_i, y_i)\big|_\rho = 0$ regardless of the value of x_i, so $(f \circ \textsc{And}_2)|_\rho$ is independent of all masked variables.
2. For any sampled ρ, let ρ_f be the restriction from $\mathsf{Leaf}(T)$ (see line 3) used to generate ρ. Then
$$(\textsc{And}_2(x_1, y_1)|_\rho, \ldots, \textsc{And}_2(x_n, y_n)|_\rho) = (\rho_f(z_1), \ldots, \rho_f(z_n)),$$
and hence$(f \circ \textsc{And}_2)|_\rho = f|_{\rho_f}$. Since ρ_f is a max-degree restriction of f, it follows that
$$\deg\big((f \circ \textsc{And}_2)|_\rho\big) = \deg(f|_{\rho_f}) = |\rho_f^{-1}(*)| = |\rho^{-1}(*)|.$$
3. By (2) and Claim 4.1,
$$\Pr_\rho\big[\deg\big((f \circ \textsc{And}_2)|_\rho\big) \geqslant \tfrac{d}{4}\big] = \Pr_{\rho_f \sim \mathsf{Leaf}(T)}\big[\deg(f|_{\rho_f}) \geqslant \tfrac{d}{4}\big] \geqslant 0.9. \qquad \square$$

Claim 5.2. *Let T be a one-sided max-degree restriction tree for $f : \{0,1\}^n \to \{0,1\}$, and let ρ be sampled from* LiftedRestriction(T)*. For any $S \subseteq \{x_1, \ldots, x_n\}$ and $t \in \mathbb{N}$,*

$$\Pr_{\rho}[|\{x_i \in S : \rho(x_i) = *\}| \geqslant t \text{ and } \forall x_i \in S,\ \rho(x_i) \neq \Delta] \leqslant 2^{-t}.$$

Proof. For any sampled ρ, let ρ_f denote the restriction on variables $(z_1, \ldots, z_n)$ obtained from $\mathsf{Leaf}(T)$ (see line 3) used to generate ρ. Then,

$$\Pr_{\rho}[|\{x_i \in S : \rho(x_i) = *\}| \geqslant t \text{ and } \forall x_i \in S,\ \rho(x_i) \neq \Delta] \overset{(1)}{=} \Pr_{\rho_f \sim \mathsf{Leaf}(T)}\left[\deg\Big(\big(\prod_{x_i \in S} z_i\big)|_{\rho_f}\Big) \geqslant t\right] \overset{(2)}{\leqslant} 2^{-t},$$

where (1) follows from the construction of ρ and (2) from Claim 4.2. $\square$

Combining these properties of random restrictions, we obtain the following results.

Claim 5.3. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function admitting a one-sided max-degree restriction tree of depth $d \geqslant 40$. Then*

$$\log \left\| \widehat{f \circ \textsc{And}_2} \right\|_{1,1/3} = \Omega(\sqrt{d}).$$

Proof. Assume for contradiction that the claim fails. Let $k = \frac{1}{c_1}\sqrt{d/(4c)}$ for suitable positive constants $c, c_1 > 0$ to be fixed later. Suppose there exists a real polynomial in the Fourier basis

$$\tilde{Q}(x, y) = \sum_{S_1, S_2 \subseteq [n]} q_{S_1,S_2}\, \chi_{S_1}(x)\chi_{S_2}(y)$$

that 1/3-approximates $f \circ \textsc{And}_2$ and satisfies $\sum_{S_1,S_2} |q_{S_1,S_2}| \leqslant \frac{1}{100}\, 2^k$. We will derive a contradiction.

Let T be a one-sided max-degree restriction tree of depth d for f, and sample $\rho \sim$ LiftedRestriction(T). Consider the restricted functions $(f \circ \textsc{And}_2)|_\rho$ and $\tilde{Q}|_\rho$.

Simplification of $\tilde{Q}|_\rho$. Under ρ, all y-variables are fixed, so each $\chi_{S_2}(y)$ becomes a constant. For the x-part, we call a term $\chi_{S_1}(x)$ *relevant high degree under* ρ if none of its variables are masked ($\rho(x_i) \neq \Delta$ for all $x_i \in S_1$) and at least k of them are assigned $*$ by ρ. A term $\chi_{S_1}(x)\chi_{S_2}(y)$ is relevant high degree if its x-part is. We now bound the expected ℓ_1-mass of such terms in $\tilde{Q}|_\rho$:

$$\begin{aligned}
&\mathbb{E}_\rho\Big[\ell_1\text{-mass of relevant high-degree terms in } \tilde{Q}|_\rho\Big] \\
&\leqslant \sum_{S_1,S_2} |q_{S_1,S_2}| \cdot \Pr_{\rho}\Big[|\{x_i \in S_1 : \rho(x_i) = *\}| \geqslant k,\ \rho(x_i) \neq \Delta\ \forall i \in S_1\Big] \\
&\leqslant \sum_{S_1,S_2} |q_{S_1,S_2}| \cdot 2^{-k} \qquad \text{(by Claim 5.2)} \\
&\leqslant 2^{-k} \sum_{S_1,S_2} |q_{S_1,S_2}| \leqslant \tfrac{1}{100}.
\end{aligned}$$

By Markov's inequality, with probability at least 0.9 over ρ, the total ℓ_1-mass of such terms is at most 0.1. Moreover, by Claim 5.1(3), $\deg((f \circ \textsc{And}_2)|_\rho) \geqslant d/4$ with probability at least 0.9. Hence, with probability at least 0.8, both events hold simultaneously. Fix such a restriction ρ.

Constructing a low-degree approximator for $(f \circ \textsc{And}_2)|_\rho$. Let $\bar{Q}$ be $\tilde{Q}|_\rho$ with all relevant high-degree terms removed. Since $\tilde{Q}|_\rho$ 1/3-approximates $(f \circ \textsc{And}_2)|_\rho$ and the discarded ℓ_1-mass is at most 0.1, $\bar{Q}$ 0.44-approximates $(f \circ \textsc{And}_2)|_\rho$. Note that while $\bar{Q}$ may still contain high-degree monomials, all such monomials include at least one masked variable. Since $(f \circ \textsc{And}_2)|_\rho$ is independent of masked variables (see Claim 5.1(1)), taking the expectation of $\bar{Q}$ over masked variables set uniformly at random to 0 or 1 removes all remaining high-degree terms, yielding a polynomial $\bar{Q}'$ of degree $< k$ that still 0.44-approximates $(f \circ \textsc{And}_2)|_\rho$.

Final contradiction. By error reduction (Theorem 2.6), $\bar{Q}'$ can be amplified to a 1/3-approximator of degree at most $c_1 k$. Thus,

$$\widetilde{\deg}((f \circ \textsc{And}_2)|_\rho) < c_1 k = \sqrt{d/(4c)}.$$

However, $\deg((f \circ \textsc{And}_2)|_\rho) \geqslant d/4$, contradicting the general bound $\deg(g) \leqslant c \cdot (\widetilde{\deg}(g))^2$ for Boolean functions g (Theorem 2.7). $\square$

Theorem 5.4. *For every total Boolean function $f : \{0,1\}^n \to \{0,1\}$, we have*

$$\log \mathit{spar}(f) = O\Big((\log \big\| \widehat{f \circ \mathrm{AND}_2} \big\|_{1,1/3})^2 \cdot \log n \Big).$$

Proof. By Corollary 3.13, there exists a one-sided max-degree restriction tree for f of depth $d = \Omega\big(\frac{\log(\mathrm{spar}(f))}{\log n}\big)$. Assume $d \geqslant 40$, as the claim is trivial otherwise. Applying Claim 5.3, we obtain

$$\log \big\| \widehat{f \circ \mathrm{AND}_2} \big\|_{1,1/3} = \Omega(\sqrt{d}) = \Omega\left(\sqrt{\frac{\log \mathrm{spar}(f)}{\log n}} \right). \qquad \square$$

5.2 Consequences for Communication Complexity

We use the following result of Lee and Shraibman [LS+09c], which shows that the logarithm of the approximate Fourier ℓ_1-norm of f lower bounds both the logarithm of the approximate rank and the quantum communication complexity of $f \circ \mathrm{XOR}_2$.

Theorem 5.5 ([LS+09c], Theorem 7.10; see also [She21], Theorem 3.5.4)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. Then*

$$\log \widetilde{\mathit{rank}}(f \circ \mathrm{XOR}_2) = \Omega\big(\log \big\|\widehat{f}\big\|_{1,1/3}\big) \qquad \textit{and} \qquad Q^{cc}(f \circ \mathrm{XOR}_2) = \Omega\big(\log \big\|\widehat{f}\big\|_{1,1/3}\big).$$

Combining the above theorem with Theorem 5.4, we obtain the following consequences.

Theorem 1.9 (Restated)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. Then,*

1. $\log \mathit{spar}(f) = O\Big((\log \widetilde{\mathit{rank}}(f \circ \mathrm{EQ}_4))^2 \cdot \log n \Big)$.
2. $\log \mathit{spar}(f) = O\big(Q^{cc}(f \circ \mathrm{EQ}_4)^2 \cdot \log n \big)$.

Proof. By combining Theorem 5.4 and Theorem 5.5, we have

$$\begin{aligned}
\log \mathrm{spar}(f) &= O\Big((\log \big\| \widehat{f \circ \mathrm{AND}_2} \big\|_{1,1/3})^2 \cdot \log n \Big) \\
&= O\Big((\log \widetilde{\mathrm{rank}}((f \circ \mathrm{AND}_2) \circ \mathrm{XOR}_2))^2 \cdot \log n \Big) \\
&\overset{(1)}{=} O\Big((\log \widetilde{\mathrm{rank}}(f \circ \mathrm{EQ}_4))^2 \cdot \log n \Big),
\end{aligned}$$

where (1) follows from the fact that the communication matrices of $(f \circ \mathrm{AND}_2) \circ \mathrm{XOR}_2$ and $f \circ \mathrm{EQ}_4$ are identical up to a permutation of rows. Furthermore,

$$\log \mathrm{spar}(f) = O\Big((\log \big\| \widehat{f \circ \mathrm{AND}_2} \big\|_{1,1/3})^2 \cdot \log n \Big) = O\big(Q^{cc}((f \circ \mathrm{AND}_2) \circ \mathrm{XOR}_2)^2 \cdot \log n \big). \qquad \square$$

Remark 5.6. *One could alternatively derive the second bound in Theorem 1.9 from the first using $\log \widetilde{\mathit{rank}}(F) = O(Q^{cc}(F) + \log n)$ [LS09a] for any two-party function $F : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. However, this would introduce an additive $(\log n)^{O(1)}$ loss when $Q^{cc}(F) = o(\log n)$. To avoid this, we directly apply $Q^{cc}(f \circ \mathrm{XOR}_2) = \Omega(\log \big\|\widehat{f}\big\|_{1,1/3})$ from Theorem 5.5 to establish the second bound.*

Theorem 5.7. *For any function of the form $f \circ \mathrm{EQ}_4$, where $f : \{0,1\}^n \to \{0,1\}$, the following hold:*

1. *The* log-approximate-rank conjecture *holds up to poly-logarithmic factors in n.*
2. *The randomized and quantum communication complexities differ by at most a polynomial factor (again, up to poly-logarithmic factors in n).*
3. *There exists a function f for which the randomized and deterministic communication complexities are exponentially separated.*

Proof. **(1)** We construct a randomized communication protocol for $f \circ \mathrm{EQ}_4$ by simulating an AND-decision tree for f. Each AND-query in the decision tree corresponds to evaluating an equality function on at most $2n$ bits. Since the equality function admits an $O(1)$-bit randomized protocol with constant error, we can solve each such instance to error $O(1/\mathrm{D}^{\wedge dt}(f))$ to obtain an overall constant-error protocol for $f \circ \mathrm{EQ}_4$. Therefore,

$$\mathrm{R}^{cc}(f \circ \mathrm{EQ}_4) \overset{(1)}{\leqslant} \mathrm{D}^{\wedge dt}(f) \log \mathrm{D}^{\wedge dt}(f) \overset{(2)}{=} O\big((\log \mathrm{spar}(f))^5 (\log n)^2\big) \overset{(3)}{=} O\Big((\log \widetilde{\mathrm{rank}}(f \circ \mathrm{EQ}_4))^{10} (\log n)^7\Big)\,,$$

where (1) follows from the simulation argument, (2) uses the result of Knop et al. [KLMY21b], and (3) follows from Theorem 1.9. This proves that the log-approximate-rank conjecture holds for $f \circ \mathrm{EQ}_4$ up to poly-logarithmic factors in n.

(2) From the bound above,

$$\begin{aligned}
\mathrm{R}^{cc}(f \circ \mathrm{EQ}_4) &= O\big((\log \mathrm{spar}(f))^5 (\log n)^2\big) \\
&\overset{(1)}{=} O\Big(\big(\log \big\| \widehat{f \circ \mathrm{AND}_2} \big\|_{1,1/3}\big)^{10} (\log n)^7\Big) \\
&\overset{(2)}{=} O\big(\mathrm{Q}^{cc}((f \circ \mathrm{AND}_2) \circ \mathrm{XOR}_2)^{10} (\log n)^7\big) \\
&= O\big(\mathrm{Q}^{cc}(f \circ \mathrm{EQ}_4)^{10} (\log n)^7\big)\,,
\end{aligned}$$

where (1) follows from Theorem 5.4, and (2) from Theorem 5.5 Hence, the randomized and quantum communication complexities of $f \circ \mathrm{EQ}_4$ differ by at most a polynomial factor, up to poly-logarithmic factors in n.

(3) For $f = \mathrm{AND}_n$, the lifted function $f \circ \mathrm{EQ}_4$ is the equality function on $2n$ bits, which exhibits an exponential separation between deterministic and randomized communication complexities. ☐

The statement of Theorem 1.10 is encompassed by Theorem 5.7.

References

[ABDK+21] Scott Aaronson, Shalev Ben-David, Robin Kothari, Shravas Rao, and Avishay Tal. Degree vs. approximate degree and quantum implications of huang's sensitivity theorem. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1330–1342, 2021.

[ABT19] Anurag Anshu, Naresh Goud Boddu, and Dave Touchette. Quantum log-approximate-rank conjecture is also false. In David Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 982–994. IEEE Computer Society, 2019.

[AFH12a] Anil Ada, Omar Fawzi, and Hamed Hatami. Spectral norm of symmetric functions. In Anupam Gupta, Klaus Jansen, José D. P. Rolim, and Rocco A. Servedio, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 15th International Workshop, APPROX 2012, and 16th International Workshop, RANDOM 2012, Cambridge, MA, USA, August 15-17, 2012. Proceedings*, volume 7408 of *Lecture Notes in Computer Science*, pages 338–349. Springer, 2012.

[AFH12b] Anil Ada, Omar Fawzi, and Hamed Hatami. Spectral norm of symmetric functions. In *International Workshop on Approximation Algorithms for Combinatorial Optimization*, pages 338–349. Springer, 2012.

[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. *J. ACM*, 48(4):778–797, 2001.

[BdW01] Harry Buhrman and Ronald de Wolf. Communication complexity lower bounds by polynomials. In *Proceedings of the 16th Annual IEEE Conference on Computational Complexity, Chicago, Illinois, USA, June 18-21, 2001*, pages 120–130. IEEE Computer Society, 2001.

[BdW02] Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: a survey. *Theor. Comput. Sci.*, 288(1):21–43, 2002.

[BH09] Paul Beame and Dang-Trinh Huynh-Ngoc. Multiparty communication complexity and threshold circuit size of acˆ0. In *50th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2009, October 25-27, 2009, Atlanta, Georgia, USA*, pages 53–62. IEEE Computer Society, 2009.

[BIVW16] Andrej Bogdanov, Yuval Ishai, Emanuele Viola, and Christopher Williamson. Bounded indistinguishability and the complexity of recovering secrets. In *International Cryptology Conference*, volume LNCS 9816, pages 593–618. Springer, 2016.

[BKT20] Mark Bun, Robin Kothari, and Justin Thaler. The polynomial method strikes back: Tight quantum query bounds via dual polynomials. *Theory Comput.*, 16:1–71, 2020.

[BMTW19] Andrej Bogdanov, Nikhil S. Mande, Justin Thaler, and Christopher Williamson. Approximate degree, secret sharing and concentration phenomena. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM*, volume LIPIcs, 145, pages 71:1–71:21. Schloss-Dagstuhl, 2019.

[BN21] Gal Beniamini and Noam Nisan. Bipartite perfect matching as a real polynomial. In *53rd ACM Symposium on Theory of Computing (STOC)*, pages 1118–1131. ACM, 2021.

[BS90] Jehoshua Bruck and Roman Smolensky. Polyomial threshold functions, AC^0 functions, and spectral norms (extended abstract). In *31st Annual Symposium on Foundations of Computer Science (FOCS)*, volume II, pages 632–641. IEEE, 1990.

[BT22] Mark Bun and Justin Thaler. Approximate degree in classical and quantum computing. *Found. Trends Theor. Comput. Sci.*, 15(3-4):229–423, 2022.

[CA08] Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. *Electron. Colloquium Comput. Complex.*, TR08-002, 2008.

[CDL25] Arkadev Chattopadhyay, Yogesh Dahiya, and Shachar Lovett. Exact versus approximate representations of boolean functions in the de morgan basis. *arXiv preprint arXiv:2507.13963*, 2025.

[CDM+23] Arkadev Chattopadhyay, Yogesh Dahiya, Nikhil S Mande, Jaikumar Radhakrishnan, and Swagato Sanyal. Randomized versus deterministic decision tree size. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, pages 867–880, 2023.

[CHH+25] Tsun-Ming Cheung, Hamed Hatami, Kaave Hosseini, Aleksandar Nikolov, Toniann Pitassi, and Morgan Shirley. A lower bound on the trace norm of boolean matrices and its applications. In *16th Innovations in Theoretical Computer Science Conference (ITCS 2025)*, pages 37–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2025.

[CM17a] Arkadev Chattopadhyay and Nikhil S. Mande. Dual polynomials and communication complexity of XOR functions. *Electron. Colloquium Comput. Complex.*, TR17-062, 2017.

[CM17b] Arkadev Chattopadhyay and Nikhil S. Mande. A lifting theorem with applications to symmetric functions. In Satya V. Lokam and R. Ramanujam, editors, *37th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science, FSTTCS 2017, December 11-15, 2017, Kanpur, India*, volume 93 of *LIPIcs*, pages 23:1–23:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017.

[CMS20] Arkadev Chattopadhyay, Nikhil S. Mande, and Suhail Sherif. The log-approximate-rank conjecture is false. *J.ACM*, 67(4):1–28, 2020.

[CTUW14] Karthekeyan Chandrasekaran, Justin Thaler, Jonathan Ullman, and Andrew Wan. Faster private release of marginals on small databases. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 387–402, 2014.

[DW02] Ronald De Wolf. Quantum communication and complexity. *Theoretical computer science*, 287(1):337–353, 2002.

[GS08] Ben Green and Tom Sanders. Boolean functions with small spectral norm. *Geometric and Functional Analysis*, 18:144 – 162, 2008.

[Hua19] Hao Huang. Induced subgraphs of hypercubes and a proof of the sensitivity conjecture. *CoRR*, abs/1907.00847, 2019.

[Juk12] Stasys Jukna. *Boolean Function Complexity - Advances and Frontiers*, volume 27 of *Algorithms and Combinatorics*. Springer, 2012.

[KKMS08] Adam Tauman Kalai, Adam R Klivans, Yishay Mansour, and Rocco A Servedio. Agnostically learning halfspaces. *SIAM Journal on Computing*, 37(6):1777–1805, 2008.

[KLMY21a] Alexander Knop, Shachar Lovett, Sam McGuire, and Weiqiang Yuan. Guest column: Models of computation between decision trees and communication. *ACM SIGACT News*, 52(2):46–70, 2021.

[KLMY21b] Alexander Knop, Shachar Lovett, Sam McGuire, and Weiqiang Yuan. Log-rank and lifting for and-functions. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 197–208, 2021.

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[KP95] Matthias Krause and Pavel Pudlák. On computing boolean functions by sparse real polynomials. In *36th Annual Symposium on Foundations of Computer Science, FOCS 1995, Milwaukee, Wisconsin, USA, 23-25 October 1995*, pages 682–691. IEEE Computer Society, 1995.

[KS04] Adam R Klivans and Rocco A Servedio. Learning dnf in time 2õ (n1/3). *Journal of Computer and System Sciences*, 68(2):303–318, 2004.

[LS09a] Troy Lee and Adi Shraibman. An approximation algorithm for approximation rank. In *2009 24th Annual IEEE Conference on Computational Complexity*, pages 351–357. IEEE, 2009.

[LS09b] Troy Lee and Adi Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Comput. Complex.*, 18(2):309–336, 2009.

[LS+09c] Troy Lee, Adi Shraibman, et al. Lower bounds in communication complexity. *Foundations and Trends® in Theoretical Computer Science*, 3(4):263–399, 2009.

[Nis91] Noam Nisan. CREW prams and decision trees. *SIAM J. Comput.*, 20(6):999–1007, 1991.

[Nis21] Noam Nisan. The demand query model for bipartite matching. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 592–599. ACM, 2021.

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

[Raz03] Alexander Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya:Mathematics*, 67(1):145–159, 2003.

[SdW] Makrand Sinha and Ronald de Wolf. Exponential separation between quantum communication and logarithm of approximate rank. In David Zuckerman, editor, *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 966–981.

[She10] Alexander A. Sherstov. On quantum-classical equivalence for composed communication problems. *Quantum Inf. Comput.*, 10(5&6):435–455, 2010.

[She11] Alexander A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011.

[She14] Alexander A. Sherstov. Communication lower bounds using directional derivatives. *J. ACM*, 61(6):34:1–34:71, 2014.

[She21] Suhail Sherif. *Communication Complexity and Quantum Optimization Lower Bounds via Query Complexity.* PhD thesis, PhD thesis, Tata Institute of Fundamental Research, Mumbai, 2021.

[SZ09] Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Inf. Comput.*, 9(5&6):444–460, 2009.

[TUV12] Justin Thaler, Jonathan Ullman, and Salil Vadhan. Faster algorithms for privately releasing marginals. In *International Colloquium on Automata, Languages, and Programming*, pages 810–821. Springer, 2012.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il