A Strong Direct Sum Theorem for Distributional Query Complexity

Guy Blanc
Department of Computer Science, Stanford University, CA, USA

Caleb Koch
Department of Computer Science, Stanford University, CA, USA

Carmen Strassle
Department of Computer Science, Stanford University, CA, USA

Li-Yang Tan
Department of Computer Science, Stanford University, CA, USA

Abstract

Consider the expected query complexity of computing the k-fold direct product $f^{\otimes k}$ of a function f to error ε with respect to a distribution μ^k. One strategy is to sequentially compute each of the k copies to error ε/k with respect to μ and apply the union bound. We prove a *strong direct sum theorem* showing that this naive strategy is essentially optimal. In particular, computing a direct product necessitates a blowup in both query complexity and error.

Strong direct sum theorems contrast with results that only show a blowup in query complexity or error but not both. There has been a long line of such results for distributional query complexity, dating back to (Impagliazzo, Raz, Wigderson 1994) and (Nisan, Rudich, Saks 1994), but a strong direct sum theorem that holds for all functions in the standard query model had been elusive.

A key idea in our work is the first use of the Hardcore Theorem (Impagliazzo 1995) in the context of query complexity. We prove a new *resilience lemma* that accompanies it, showing that the hardcore of $f^{\otimes k}$ is likely to remain dense under arbitrary partitions of the input space.

2012 ACM Subject Classification Theory of computation → Computational complexity and cryptography

Keywords and phrases Query complexity, direct product theorem, hardcore theorem

Digital Object Identifier 10.4230/LIPIcs.CCC.2024.16

Funding The authors are supported by NSF awards 1942123, 2211237, 2224246 and a Google Research Scholar award. Guy is also supported by a Jane Street Graduate Research Fellowship, Caleb by an NDSEG fellowship, and Carmen by a Stanford Computer Science Distinguished Fellowship.

1 Introduction

The *direct sum problem* seeks to understand the ways in which the complexity of solving k independent instances of a computational task scales with k. This problem and its variants such as the *XOR problem*, where one only seeks to compute the XOR of the k output values, have a long history in complexity theory. Research on them dates back to Strassen [33] and they have since been studied in all major computational models including boolean circuits [36, 26, 12, 14, 17, 29, 15, 11], communication protocols [16, 30, 22, 25, 35, 21, 31, 18, 20, 4, 8, 37], as well as classical [16, 27, 30, 22, 10, 5, 6, 9, 13] and quantum query complexity [2, 22, 32, 31, 1, 24].

1.1 This work

We focus on classical query complexity, and specifically distributional query complexity. Distributional complexity, also known as average-case complexity, is a basic notion applicable to all models of computation. Direct sum theorems and XOR lemmas for distributional complexity, in addition to being statements of independent interest, have found applications in areas ranging from derandomization [36, 28, 17] to streaming [3] and property testing [7].

Let $f : \{\pm1\}^n \to \{\pm1\}$ be a boolean function, μ be a distribution over $\{\pm1\}^n$, and consider the task of computing the k-fold direct product $f^{\otimes k}(X^{(1)}, \ldots, X^{(k)}) \coloneqq (f(X^{(1)}), \ldots, f(X^{(k)}))$ of f to error ε with respect to μ^k. One strategy is to sequentially compute each $f(X^{(i)})$ to error ε/k with respect to μ and apply the union bound. Writing $\overline{\mathrm{Depth}}^{\mu}(f, \varepsilon)$ to denote the minimum expected depth of any decision tree that computes f to error ε w.r.t. μ, this shows that:

$$\overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes k}, \varepsilon) \leq k \cdot \overline{\mathrm{Depth}}^{\mu}(f, \tfrac{\varepsilon}{k}).$$

Our main result is that this naive strategy is essentially optimal for all functions and distributions:

▶ **Theorem 1** (Strong direct sum theorem for distributional query complexity; special case of Theorem 2)**.** *For every function $f : \{\pm1\}^n \to \{\pm1\}$, distribution μ over $\{\pm1\}^n$, integer $k \in \mathbb{N}$, and $\varepsilon < 1$,*

$$\overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes k}, \varepsilon) \geq \tilde{\Omega}(\varepsilon^2 k) \cdot \overline{\mathrm{Depth}}^{\mu}(f, \Theta(\tfrac{\varepsilon}{k})).$$

Such direct sum theorems are termed *strong*, referring to the fact that they show that computing a direct product necessitates a blowup in *both* the computational resources of interest – in our case, query complexity – *and* error. Strong direct sum theorems contrast with standard ones, which only show a blowup in computational resource, and also with direct *product* theorems, which focus on the blowup in error. We give a detailed overview of prior work in Section 3, mentioning for now that while standard direct sum and direct product theorems for distributional query complexity have long been known, a strong direct sum theorem had been elusive. Prior to our work, it was even open whether $\overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes k}, 1.01\varepsilon) \geq 1.01 \cdot \overline{\mathrm{Depth}}^{\mu}(f, \varepsilon)$ holds. Indeed, the problem is known to be quite subtle, as a striking counterexample of Shaltiel [30] shows that a strong direct sum theorem is badly false if one considers *worst-case* instead of expected query complexity.

A strong XOR lemma

We also obtain a strong XOR lemma (Theorem 3) as a corollary of a simple equivalence between direct sum theorems and XOR lemmas for query complexity. One direction is immediate, since the k-fold XOR $f^{\oplus k}(X^{(1)}, \ldots, X^{(k)}) \coloneqq f(X^{(1)}) \oplus \cdots \oplus f(X^{(k)})$ can only be easier to compute than the k-fold direct product. For the query model the converse also holds: a direct sum theorem implies an XOR lemma with analogous parameters.

2 Broader context: Comparison with the randomized setting

Direct sum theorems are also well-studied in the setting of *randomized* query complexity. Recall that the ε-error randomized query complexity of f, denoted $\overline{\mathrm{R}}(f, \varepsilon)$, is the minimum expected depth of any randomized decision tree that computes f with error at most ε for all inputs. By Yao's minimax principle, direct sum theorems for distributional query complexity imply analogous ones for randomized query complexity. However, as we now elaborate, such theorems are substantially more difficult to prove in the distributional setting.

2.1 A simple and near-optimal strong direct sum theorem for $\overline{\mathrm{R}}$

For randomized query complexity, proving a strong direct sum theorem with near-optimal parameters requires only two observations.

Observation #1

The first is that a standard direct sum theorem, one without error amplification, easily holds in the distributional setting, and hence the randomized setting as well by Yao's principle:

$$\overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes k},\varepsilon) \geq \Omega(k)\cdot\overline{\mathrm{Depth}}^{\mu}(f,\varepsilon) \quad \text{and therefore} \quad \overline{\mathrm{R}}(f^{\otimes k},\varepsilon) \geq \Omega(k)\cdot\overline{\mathrm{R}}(f,\varepsilon). \tag{1}$$

The idea is that given a decision tree T of average depth q that computes $f^{\otimes k}$ with error ε, one can extract a decision tree of average depth q/k that computes f to error ε: place the input in a random block $\boldsymbol{i}\sim[k]$, fill the remaining blocks with independent random draws from μ, and return the $\boldsymbol{i}$th bit of T's output. It is straightforward to show that this reduces the average depth of T by a factor of k while preserving its error.

Observation #2

The second observation is standard error reduction of randomized algorithms by repetition, which in particular implies:

$$\overline{\mathrm{R}}(f,\tfrac{\varepsilon}{k}) \leq O(\log k)\cdot\overline{\mathrm{R}}(f,\varepsilon). \tag{2}$$

Combining Equations (1) and (2) yields a strong direct sum theorem

$$\overline{\mathrm{R}}(f^{\otimes k},\varepsilon) \geq \Omega\left(\frac{k}{\log k}\right)\cdot\overline{\mathrm{R}}(f,\tfrac{\varepsilon}{k})$$

that is within a $O(\log k)$ factor of optimal.

Blais–Brody

Using more sophisticated techniques, Blais and Brody [6] were recently able remove to this $O(\log k)$ factor and obtain an optimal strong direct sum theorem for $\overline{\mathrm{R}}$. Building on their work, Brody, Kim, Lerdputtipongporn, and Srinivasulu [9] then obtained an optimal strong XOR lemma for $\overline{\mathrm{R}}$.

2.2 Error reduction fails in the distributional setting

While the crux of [6] and [9]'s works is the removal of a $O(\log k)$ factor, the situation is very different in the distributional setting. As mentioned, prior to our work even a direct sum theorem where both factor-of-$\tilde{\Omega}(k)$ blowups in Theorem 1 are replaced by 1.01 was not known to hold.

With regards to the argument above, it is Observation #2 that breaks in the distributional setting – not only does error reduction by repetition break, the distributional analogue of Equation (2) is simply false. This points to a fundamental difference between distributional and randomized complexity: while generic error reduction of randomized algorithms is possible in all reasonable models of computation, the analogous statement for distributional complexity is badly false in all reasonable models of computation. For the query model specifically, in Appendix C we give an easy proof of the following:

▶ **Fact 1.** *For any $n \in \mathbb{N}$ and μ being the uniform distribution over $\{\pm 1\}^n$, there is a function $f : \{\pm 1\}^n \to \{\pm 1\}$ such that $\overline{\mathrm{Depth}}^{\mu}(f, \frac{1}{4}) = 0$ and yet $\overline{\mathrm{Depth}}^{\mu}(f, \frac{1}{8}) \geq \Omega(n)$.*

2.3 A brief summary of our approach

We revisit Observation #1 and show how the very same extraction strategy can in fact yield a tree with error $\Theta(\varepsilon/k)$, instead of ε, at the expense of only a slight increase in depth. A key technical ingredient in our analysis is *Impagliazzo's Hardcore Theorem* [14]. For intuition as to why this theorem may be relevant for us, we note that it is tightly connected to the notion of *boosting* from learning theory – they are, in some sense, dual to each other [23]. And boosting is, of course, a form of error reduction, albeit one that is more intricate than error reduction by repetition.

See Section 5 for a detailed overview of our approach, including a discussion of why Impagliazzo's Hardcore Theorem, as is, does not suffice, thereby necessitating our new "resilience lemma" that accompanies it.

3 Prior Work

We now place Theorem 1 within the context of prior work on direct sum and product theorems for distributional query complexity. This is a fairly large body of work that dates back to the 1990s.

3.1 Standard direct sum and product theorems

Standard direct sum theorems

As we sketched in Section 2.1, a simple argument shows that

$$\overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes k}, \varepsilon) \geq \Omega(k) \cdot \overline{\mathrm{Depth}}^{\mu}(f, \varepsilon). \tag{3}$$

This along with an application of Markov's inequality yields:

$$\mathrm{Depth}^{\mu^k}(f^{\otimes k}, \varepsilon - \varepsilon') \geq \Omega(\varepsilon' k) \cdot \mathrm{Depth}^{\mu}(f, \varepsilon), \tag{4}$$

where $\mathrm{Depth}^{\mu}(\cdot, \cdot)$ is the analogue of $\overline{\mathrm{Depth}}^{\mu}(\cdot, \cdot)$ for worst-case instead of expected query complexity. (The details of these arguments are spelt out in [19, 5].)

Note that the error budget is the same on both sides of Equation (3) and the error budget on the RHS of Equation (4) is *larger* than that of the LHS. In a strong direct sum theorem one seeks a lower bound even when the error budget on the RHS is much smaller than that of the LHS, ideally by a multiplicative factor of k to match the naive upper bound.

A direct product theorem

Impagliazzo, Raz, and Wigderson [16] proved a direct product theorem which focuses on the blowup in error. They showed that:

$$\mathrm{Depth}^{\mu^k}(f^{\otimes k}, \varepsilon) \geq \mathrm{Depth}^{\mu}(f, \tfrac{\varepsilon}{k}). \tag{5}$$

While this result has the sought-for factor of k difference between the error budgets on the LHS and RHS, it comes at the price of there no longer being *any* blowup in depth.

3.2 Progress and barriers towards a strong direct sum theorem

These results naturally point to the problem of proving a unifying strong direct sum theorem. We now survey efforts at such a best-of-both-worlds result over the years.

Decision forests

Nisan, Rudich, and Saks [27] proved the following strengthening of [16]'s result. While [16] gives an upper bound on the success probability of a *single* depth-d decision tree for $f^{\otimes k}$, [27] showed that the same bound holds even for *decision forests* where one gets to construct a different depth-d tree for each of the k copies of f.

Since one can always stack the k many depth-d trees in a decision forest to obtain a single tree of depth kd, [27]'s result establishes a special case of a strong direct sum theorem under a structural assumption on the tree for $f^{\otimes k}$. See Figure 3 in Appendix A for an illustration of the stacked decision tree that one gets from a decision forest.

Fair decision trees

Building on the techniques of [27], Shaltiel [30] proved a strong direct sum theorem under a different structural assumption on the tree for $f^{\otimes k}$. He considered decision trees of depth kd that are "fair" in the sense that every path queries each of the k blocks of variables at most d times. ([30] actually proved a strong XOR lemma for fair decision trees, which implies a strong direct sum theorem for such trees.) See Figure 4 in Appendix A for an illustration of a fair decision tree.

Shaltiel's counterexample for worst-case query complexity

These results of [27] and [30] could be viewed as evidence in favor of a general strong direct sum theorem, one that does not impose any structural assumptions on the tree for $f^{\otimes k}$. However, in the same paper Shaltiel also presented an illuminating example: he constructed a function, which we call Shal, and a distribution μ such that for all $k \in \mathbb{N}$,

$$\mathrm{Depth}^{\mu^k}(\mathsf{Shal}^{\otimes k}, \varepsilon) \leq O\big(\mathrm{Depth}^{\mu}(\mathsf{Shal}, \tfrac{\varepsilon}{k})\big). \tag{6}$$

This shows, surprisingly, that for *worst-case* query complexity, the factor-of-$\Omega(k)$ blowup in query complexity that one seeks in a strong direct sum theorem is not always necessary, and in fact sometimes even a constant factor suffices.

Shaltiel's counterexample vs. Theorem 1

This counterexample for worst-case query complexity should be contrasted with our main result, Theorem 1, which shows that a strong direct sum theorem holds for *expected* query complexity. Indeed, the starting point of our work was the encouraging observation that Shaltiel's function does in fact satisfy a strong direct sum theorem if one instead considers expected query complexity. That is, for any $\varepsilon < 1$ and sufficiently large k,

$$\overline{\mathrm{Depth}}^{\mu^k}(\mathsf{Shal}^{\otimes k}, \varepsilon) \geq \Omega(k) \cdot \overline{\mathrm{Depth}}^{\mu}(\mathsf{Shal}, \tfrac{\varepsilon}{k}).$$

This is a simple observation but appears to have been overlooked. As we now overview, subsequent work considered other ways of sidestepping Shaltiel's counterexample.

3.3 Results in light of Shaltiel's counterexample

A strong direct sum theorem for the OR function

Klauck, Špalek, and de Wolf [22] sidestepped Shaltiel's counterexample by considering a *specific* function (and distribution): motivated by applications to time-space tradeoffs, they proved a strong direct sum theorem for the OR function and with μ being its canonical hard distribution. Using this, they also showed, for all functions f a lower bound on $f^{\otimes k}$'s query complexity in terms of f's block sensitivity. This stands in contrast to a strong direct sum theorem where one seeks a lower bound on $f^{\otimes k}$'s query complexity in terms of f's query complexity.

A phase transition in Shaltiel's counterexample

The precise parameters of Shaltiel's counterexample are:

$$\mathrm{Depth}^{\mu^k}(\mathsf{Shal}^{\otimes k}, e^{-\Theta(\delta k)}) \le C\delta k \cdot \mathrm{Depth}^{\mu}(\mathsf{Shal}, \delta)$$

for all sufficiently large constants C. Importantly, the multiplicative factor on the RHS is only δk instead of k, and therefore becomes a constant if the initial hardness parameter is $\delta = \varepsilon/k$ (thereby yielding Equation (6)).

Drucker [10] showed that there is a "phase transition" in Shaltiel's counterexample in the following sense: for all functions f and a sufficiently small constant $c > 0$,

$$\mathrm{Depth}^{\mu^k}(f^{\otimes k}, 1 - e^{-\Theta(\delta k)}) \ge c\delta k \cdot \mathrm{Depth}^{\mu}(f, \delta). \tag{7}$$

Therefore, while [30] showed the existence of a function Shal such that its k-fold direct product can be computed to surprisingly low error if the depth budget is $C\delta k \cdot \mathrm{Depth}^{\mu}(f, \delta)$ for a sufficiently *large* constant C, [10] showed that for all functions f, this stops being the case if the depth budget is instead $c\delta k \cdot \mathrm{Depth}^{\mu}(f, \delta)$ for a sufficiently *small* constant c.

Query complexity with aborts

Blais and Brody [6] showed that Shaltiel's counterexample can be sidestepped in a different way. En route to proving their strong direct sum theorem for randomized query complexity (discussed in Section 2), they considered decision trees $T : \{\pm 1\}^n \to \{\pm 1, \perp\}$ that are allowed to output $\perp$ ("abort") on certain inputs, and where the error of T in computing a function $f : \{\pm 1\}^n \to \{\pm 1\}$ is measured with respect to $T^{-1}(\{\pm 1\})$. In other words, T's output on x is considered correct if $T(x) = \perp$.

Writing $\mathrm{Depth}^{\mu}_{\Pr[\perp] \le \frac{1}{3}}(f, \varepsilon)$ to denote the minimum depth of any decision tree for f that aborts with probability at most $1/3$ and otherwise errs with probability at most ε (both w.r.t. μ), [6] proved that

$$\mathrm{Depth}^{\mu^k}_{\Pr[\perp] \le \frac{1}{3}}(f^{\otimes k}, \varepsilon) \ge \Omega(k) \cdot \mathrm{Depth}^{\mu}_{\Pr[\perp] \le \frac{1}{3}}(f, \tfrac{\varepsilon}{k}). \tag{8}$$

Even though the error budget on non-aborts is only ε/k on the RHS, the fact that the tree is allowed to abort with probability $1/3$ means that it is deemed correct on a $1/3$ fraction of inputs "for free". A decision tree that aborts with probability $1/3$ and otherwise errs with probability ε/k can therefore be much smaller than one that never aborts and errs with probability ε/k, and indeed, it is easy to construct examples witnessing the maximally large separation:

$$n = \mathrm{Depth}^{\mu}(f, \tfrac{\varepsilon}{k}) \gg \mathrm{Depth}^{\mu}_{\Pr[\perp] \le \frac{1}{3}}(f, \tfrac{\varepsilon}{k}) = 1.$$

Building on [6], Brody, Kim, Lerdputtipongporn, and Srinivasulu [9] proved a strong XOR lemma for this model of query complexity with aborts, achieving analogous parameters.

A strong XOR lemma assuming hardness against all depths

A standard strong XOR lemma states that if f is hard against decision trees of certain *fixed* depth d, then $f^{\otimes k}$ is much harder against decision trees of depth $\Omega(dk)$. Recent work of Hoza [13] shows that Shaltiel's counterexample can be sidestepped if one allows for the stronger assumption that f's hardness "scales nicely" with d. (See the paper for the precise statement of the resulting strong XOR lemma.)

3.4 Summary

Summarizing, prior work on direct sum and product theorems for distributional query complexity either: focused on the blowup in error [16] or query complexity [19, 5] but not both; considered restrictions (fair decision trees [30]) or variants (decision forests [27]; allowing for aborts [6, 9]) of the query model; focused on specific functions (the OR function [22]); or imposed additional hardness assumptions about the function [13]. Theorem 1, on the other hand, gives a strong direct sum theorem that holds for all functions in the standard query model. See Table 1.

Table 1 Direct sum and product theorems for distributional query complexity.

Reference	Error Amplification	Query Amplification	Query model/ Assumption
[19, 5]	$\times$	✓	Standard query model
[16]	✓	$\times$	Standard query model
[27]	✓	✓	Decision forests
[30]	✓	✓	Fair decision trees
[22]	✓	✓	$f = \text{OR}$
[6, 9]	✓	✓	Decision trees with aborts
[13]	✓	✓	Hardness against all depths
Theorem 1	✓	✓	Standard query model

4 Formal statements of our results and their tightness

Theorem 1 is a special case of the following result:

▶ **Theorem 2** (Strong direct sum theorem). *For every function $f : \{\pm 1\}^n \to \{\pm 1\}$, distribution μ over $\{\pm 1\}^n$, $k \in \mathbb{N}$, and $\gamma, \delta \in (0, 1)$, we have that*

$$\overline{\text{Depth}}^{\mu^k}(f^{\otimes k}, 1 - e^{-\Theta(\delta k)} - \gamma) \geq \Omega\left(\frac{\gamma^2 k}{\log(1/\delta)}\right) \cdot \overline{\text{Depth}}^{\mu}(f, \delta).$$

We in fact prove a strong *threshold* direct sum theorem which further generalizes Theorem 2: while a direct sum theorem shows that $f^{\otimes k}$ is hard to compute, i.e. it is hard to get *all* k copies of f correct, a threshold direct sum theorem shows that it is hard even to get *most* of the k copies of f correct. See Theorem 22.

By the equivalence between strong direct sum theorems and strong XOR lemmas (Claim 35), we also get:

▶ **Theorem 3** (Strong XOR lemma). *For every function $f : \{\pm 1\}^n \to \{\pm 1\}$ and distribution μ over $\{\pm 1\}^n$, $k \in \mathbb{N}$, and $\gamma, \delta \in (0,1)$, we have that*

$$\overline{\mathrm{Depth}}^{\mu^k}\left(f^{\oplus k}, \tfrac{1}{2}(1 - e^{-\Theta(\delta k)} - \gamma)\right) \geq \Omega\left(\frac{\gamma^2 k}{\log(1/\delta)}\right) \cdot \overline{\mathrm{Depth}}^{\mu}(f, \delta).$$

4.1 Tightness

Theorem 2 amplifies an initial hardness parameter of $\delta = \Theta(1/k)$ to $1 - \gamma$ for any small constant γ with a near-optimal overhead of

$$\Omega\left(\frac{\gamma^2 k}{\log(1/\delta)}\right) = \Omega\left(\frac{k}{\log k}\right).$$

However, due to the polynomial dependence on γ, we cannot achieve a final hardness parameter that is exponentially close to 1 as a function of k. We show that this is unavoidable since at least a linear dependence on γ is necessary:

▷ Claim 4 (Linear dependence on γ is necessary). Let $\mathsf{Par} : \{\pm 1\}^n \to \{\pm 1\}$ be the parity function and μ be the uniform distribution over $\{\pm 1\}^n$. Then for all γ,

$$\overline{\mathrm{Depth}}^{\mu^k}(\mathsf{Par}^{\otimes k}, 1 - \gamma) \leq O(\gamma k) \cdot \overline{\mathrm{Depth}}^{\mu}(\mathsf{Par}, \tfrac{1}{4}).$$

The same example shows that a linear dependence on γ is likewise necessary in the setting of XOR lemmas. Determining the optimal polynomial dependence on γ in both settings, as well as the necessity of the $\log(1/\delta)$ factor, are concrete avenues for future work.

5 Technical Overview for Theorem 2

5.1 Hardcore measures and the Hardcore Theorem

At the heart of our proof is the notion of a *hardcore measure* and *Impagliazzo's Hardcore Theorem* [14], both adapted to the setting of query complexity.

▶ **Definition 5** (Hardcore measure for query complexity). *We say that $H : \{\pm 1\}^n \to [0,1]$ is a (γ, d)-hardcore measure for $f : \{\pm 1\}^n \to \{\pm 1\}$ w.r.t. μ of density δ if:*

1. *H's density is δ: $\mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mu}[H(\boldsymbol{x})] = \delta$.*
2. *d-query algorithms achieve correlation at most γ with f on H:*

$$\mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mu}[f(\boldsymbol{x})T(\boldsymbol{x})H(\boldsymbol{x})] \leq \gamma \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mu}[H(\boldsymbol{x})] = \gamma\delta.$$

for all decision trees T whose expected depth w.r.t. μ is at most d.

▶ **Theorem 6** (Hardcore Theorem for query complexity). *For every function $f : \{\pm 1\}^n \to \{\pm 1\}$, distribution μ over $\{\pm 1\}^n$, and $\gamma, \delta > 0$, there exists a (γ, d)-hardcore measure H for f of density $\delta/2$ w.r.t. μ where*

$$d = \Theta\left(\frac{\gamma^2}{\log(1/\delta)}\right) \overline{\mathrm{Depth}}^{\mu}(f, \delta).$$

The Hardcore Theorem was originally proved, and remains most commonly used, in the setting of circuit complexity where it has long been recognized as a powerful result. (See e.g. [34], where it is described as "one of the bits of magic of complexity theory".) We show in Appendix B that its proof extends readily to the setting of query complexity to establish Theorem 6. Despite its importance in circuit complexity and its straightforward extension to query complexity, our work appears to be the first to consider its applicability in the latter setting.

▶ Remark 7. For intuition regarding Definition 5, note that if $H : \{\pm 1\}^n \to \{0,1\}$ is the indicator of a *set*, the two properties simplify to: $\Pr_{\boldsymbol{x}\sim\mu}[\boldsymbol{x} \in H] = \delta$ and $\mathbb{E}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})T(\boldsymbol{x}) \mid \boldsymbol{x} \in H] \leq \gamma$.

5.2 Two key quantities: hardcore density and hardcore advantage at a leaf

Setup

For the remainder of this section, we fix a function $f : \{\pm 1\}^n \to \{\pm 1\}$, distribution μ over $\{\pm 1\}^n$, and initial hardness parameter δ (which we think of as small, close to 0). Let $T : (\{\pm 1\}^n)^k \to \{\pm 1\}^k$ be a decision tree that seeks to compute $f^{\otimes k}$ w.r.t. μ^k. Our goal is to show that T's error must be large, close to 1, unless its depth is sufficient large.

Definitions of the hardcore density and hardcore advantage at a leaf

Let $H : \{\pm 1\}^n \to [0,1]$ be a (γ, d)-hardcore measure for f of density δ w.r.t. μ given by Theorem 6. Each leaf ℓ of T corresponds to a tuple of restrictions $(\pi_1, \dots, \pi_k)$ to each of the k blocks of inputs. We will be interested in understanding, for a random block $\boldsymbol{i} \in [k]$, the extent to which the restricted function $H_{\pi_{\boldsymbol{i}}}$ retains the two defining properties of a hardcore measure: high density and strong hardness. We therefore define:

▶ **Definition 8** (Hardcore density at ℓ). *For $i \in [k]$, the hardcore density at ℓ in the ith block is the quantity:*

$$\mathrm{Dens}_H(\ell, i) \coloneqq \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}\left[H(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \ell\right].$$

The total hardcore density at ℓ is the quantity $\mathrm{Dens}_H(\ell) \coloneqq \sum_{i=1}^{k} \mathrm{Dens}_H(\ell, i)$.

See Figure 1 for an illustration of Definition 8.

▶ **Definition 9** (Hardcore advantage at ℓ). *For $i \in [k]$, the hardcore advantage at ℓ in the ith block is the quantity:*

$$\mathrm{Adv}_H(\ell, i) \coloneqq \left|\mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}\left[f(\boldsymbol{X}^{(i)})T(\boldsymbol{X})_i H(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \ell\right]\right|.$$

The total hardcore advantage at ℓ is the quantity $\mathrm{Adv}_H(\ell) \coloneqq \sum_{i=1}^{k} \mathrm{Adv}_H(\ell, i)$.

Intuitively, leaves for which $\mathrm{Dens}_H(\ell)$ is large and $\mathrm{Adv}_H(\ell)$ is small contribute significantly to error of T. Lemma 10 below formalizes this:

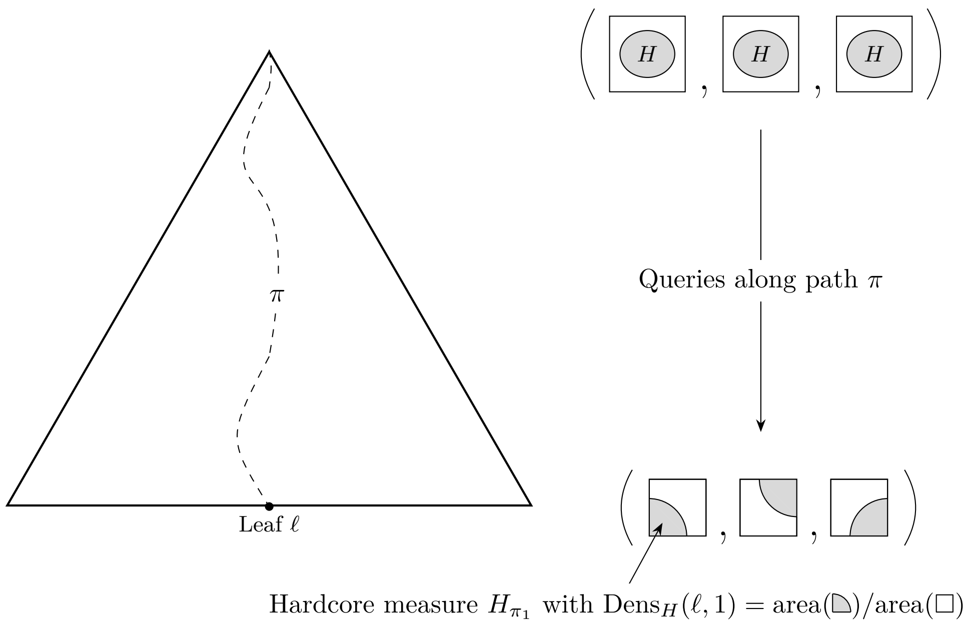


Figure 1 Illustration of a hardcore density. The tree $T : (\{\pm 1\}^n)^3 \to \{\pm 1\}^3$ seeks to compute a function $f^{\otimes 3}$. The tuple of squares at the top of the figure illustrates the set of all inputs to the function while the strings in the support of the hardcore measure are shaded gray. The tuple at the bottom of the figure illustrates the set of inputs reaching the leaf ℓ. Each block is the subcube consistent with the path π and the shaded region denotes the fragment of H which is contained in the corresponding subcube.

Notation

Canonical distribution over leaves. We write $\mu^k(T)$ to denote the distribution over leaves of T where:

$$\Pr_{\boldsymbol{\ell} \sim \mu^k(T)}[\boldsymbol{\ell} = \ell] = \Pr_{\boldsymbol{X} \sim \mu^k}[\boldsymbol{X} \text{ reaches } \ell].$$

▶ **Lemma 10** (Accuracy in terms of hardcore density and advantage at leaves)**.**

$$\Pr_{\boldsymbol{X} \sim \mu^k}[T(\boldsymbol{X}) = f^{\otimes k}(\boldsymbol{X})] \le \mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)}\left[\exp\left(-\frac{\mathrm{Dens}_H(\boldsymbol{\ell}) - \mathrm{Adv}_H(\boldsymbol{\ell})}{4}\right)\right].$$

5.3 Expected total hardcore density and advantage

Lemma 10 motivates understanding the random variables $\mathrm{Dens}_H(\boldsymbol{\ell})$ and $\mathrm{Adv}_H(\boldsymbol{\ell})$ for $\boldsymbol{\ell} \sim \mu^k(T)$. We begin by bounding their expectations:

▷ Claim 11 (Expected total hardcore density). If H is a hardcore measure of density δ then

$$\mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)}[\mathrm{Dens}_H(\boldsymbol{\ell})] = \delta k.$$

Claim 11 is a statement about density preservation. It says that H's expected density at a random leaf $\boldsymbol{\ell} \sim \mu^k(T)$ and in a random block $\boldsymbol{i} \sim [k]$ is equal to H's initial density:

$$\mathop{\mathbb{E}}_{\substack{\boldsymbol{\ell}\sim\mu^k(T)\\ \boldsymbol{i}\sim[k]}}\left[\underbrace{\mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}\left[H(\boldsymbol{X}^{(\boldsymbol{i})}) \mid \boldsymbol{X} \text{ reaches } \boldsymbol{\ell}\right]}_{\mathrm{Dens}_H(\boldsymbol{\ell},\boldsymbol{i})}\right] = \delta = \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}[H(\boldsymbol{X})].$$

▷ Claim 12 (Expected total hardcore advantage). If H is a (γ, d)-hardcore measure for f of density δ w.r.t. μ and the expected depth of T is at most dk, then

$$\mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim\mu^k(T)}[\mathrm{Adv}_H(\boldsymbol{\ell})] \leq \gamma \mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim\mu^k(T)}[\mathrm{Dens}_H(\boldsymbol{\ell})].$$

Claim 12 is a statement about depth amplification. By definition, H being a (γ, d)-hardcore measure for f means that

$$\underbrace{\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})T_{\mathrm{small}}(\boldsymbol{x})H(\boldsymbol{x})]}_{\text{Hardcore advantage}} \leq \gamma \underbrace{\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}[H(\boldsymbol{x})]}_{\text{Hardcore density}}$$

for every tree $T_{\mathrm{small}} : \{\pm1\}^n \to \{\pm1\}$ of expected depth d. Claim 12 says that

$$\underbrace{\sum_{i=1}^{k} \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}\left[f(\boldsymbol{X}^{(i)})T_{\mathrm{large}}(\boldsymbol{X})_i H(\boldsymbol{X}^{(i)})\right]}_{\text{Total hardcore advantage}} \leq \gamma \underbrace{\sum_{i=1}^{k} \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}[H(\boldsymbol{X}^{(i)})]}_{\text{Total hardcore density}}.$$

for every tree $T_{\mathrm{large}} : (\{\pm1\}^n)^k \to \{\pm1\}^k$ of expected depth dk. Crucially, the depth of T_{large} is allowed to be a factor of k larger than that of T_{small}, and yet the ratio of hardcore advantage to hardcore density remains the same (γ in both cases).

5.3.1 Done if Jensen went the other way

For intuition as to why Claim 11 and Claim 12 are relevant yet insufficient for us, note that if it were the case that $\mathbb{E}[\exp(-\boldsymbol{Z})] \leq \exp(-\mathbb{E}[\boldsymbol{Z}])$, which unfortunately is the opposite of what Jensen's inequality gives, we would have the strong bound on the accuracy of T that we seek:

$$\begin{aligned}
\Pr_{\boldsymbol{X}\sim\mu^k}[T(\boldsymbol{X}) = f^{\otimes k}(\boldsymbol{X})] &\leq \mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim\mu^k(T)}\left[\exp\left(-\frac{\mathrm{Dens}_H(\boldsymbol{\ell}) - \mathrm{Adv}_h(\boldsymbol{\ell})}{4}\right)\right] && \text{(Lemma 10)}\\
&\text{“}\leq\text{”} \exp\left(-\mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim\mu^k(T)}\left[\frac{\mathrm{Dens}_H(\boldsymbol{\ell}) - \mathrm{Adv}_H(\boldsymbol{\ell})}{4}\right]\right) && \text{(Wrong direction of Jensen)}\\
&\leq \exp\left(-\frac{\delta k - \gamma\delta k}{4}\right) && \text{(Claim 11 and Claim 12)}\\
&\leq \exp(-\Theta(\delta k)).
\end{aligned}$$

For an actual proof, we need to develop a more refined understanding of the distribution of $\mathrm{Dens}_H(\boldsymbol{\ell})$ beyond just its expectation. (As it turns out, this along with the bound on $\mathbb{E}[\mathrm{Adv}_H(\boldsymbol{\ell})]$ given by Claim 12 suffices.)

5.4 A resilience lemma for hardcore measures

An illustrative bad case to rule out

Suppose T were such that it achieved $\mathbb{E}[\mathrm{Dens}_H(\boldsymbol{\ell})] = \delta k$ by having a δ-fraction of leaves with $\mathrm{Dens}_H(\ell) = k$ and the remaining $1-\delta$ fraction with $\mathrm{Dens}_H(\ell) = 0$. If this were the case then "all the hardness" would be concentrated on a small δ fraction of leaves, and the best lower bound that we would be able to guarantee on error of T with respect to $f^{\otimes k}$ would only be δ. This is our starting assumption on the hardness of f, and so no error amplification has occurred.

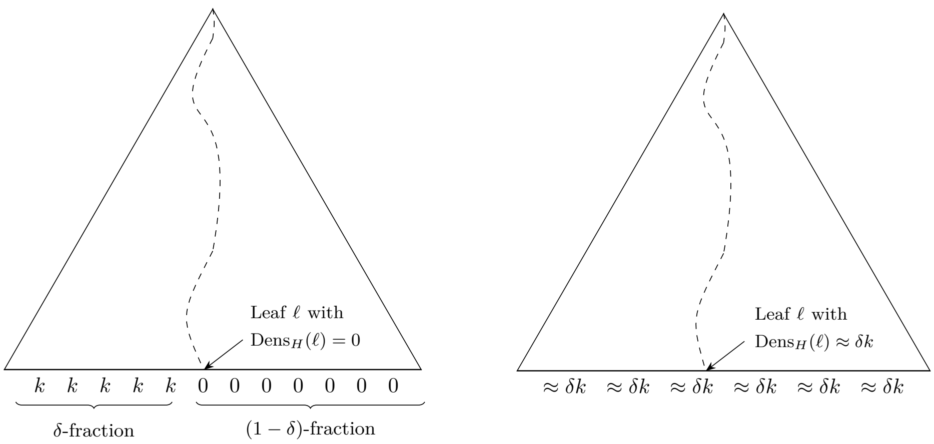
The resilience lemma

We rule out cases like this by showing that T must achieve $\mathbb{E}[\mathrm{Dens}_H(\boldsymbol{\ell})] = \delta k$ by having the vast majority of its leaves with $\mathrm{Dens}_H(\ell) = \Omega(\delta k)$, i.e. that $\mathrm{Dens}_H(\boldsymbol{\ell})$ is tightly concentrated around its expectation:

▶ **Lemma 13** (Resilience lemma). *For any hardcore measure H of density δ w.r.t. μ and tree $T : (\{\pm 1\}^n)^k \to \{\pm 1\}^k$,*

$$\Pr_{\boldsymbol{\ell} \sim \mu^k(T)}[\mathrm{Dens}_H(\boldsymbol{\ell}) \le \delta k/2] \le e^{-\delta k/8}.$$

Similarly, $\Pr_{\boldsymbol{\ell} \sim \mu^k(T)}[\mathrm{Dens}_H(\boldsymbol{\ell}) \ge 2\delta k] \le e^{-\delta k/3}$.



(a) An illustration of the bad case where Dens_H is anti-concentrated away from its mean of δk.

(b) An illustration of the good case where Dens_H is concentrated around its mean of δk.

Figure 2 An illustration of our resilience lemma (Lemma 13). This lemma shows that all trees resemble the one on the right, with $\mathrm{Dens}_H(\boldsymbol{\ell})$ tightly concentrated around its mean of δk. This allows us to rule out bad trees such as those on the left where all of the hardness is concentrated on a small fraction of the leaves.

Comparing Lemma 13 to Claim 12, we see that Claim 12 is a statement about density preservation *in expectation* whereas Lemma 13 is a statement about density preservation *with high probability*. It says that H's density at a random leaf $\boldsymbol{\ell} \sim \mu^k(T)$ and in a random block $\boldsymbol{i} \sim [k]$ remains, with high probability, roughly the same as that of H's initial density – this is why we call Lemma 13 a resilience lemma. (Our proof of Lemma 13 in fact shows that the H's density remains resilient under arbitrary partitions of $(\{\pm 1\}^n)^k$, not just those induced by a decision tree.)

With Lemma 13 in hand, the intuition sketched in Section 5.3.1 can be made formal.

6 Discussion and Future Work

Our main results are a strong direct sum theorem and a strong XOR lemma for distributional query complexity, showing that if f is somewhat hard to approximate with depth-d decision trees, then $f^{\otimes k}$ and $f^{\oplus k}$ are both *much harder* to approximate, even with decision trees of *much larger* depth. These results hold for expected query complexity, and they circumvent a counterexample of Shaltiel showing that such statements are badly false for worst-case query complexity. We view our work as confirming a remark Shaltiel made in his paper, that his counterexample "seems to exploit defects in the formulation of the problem rather than show that our general intuition for direct product assertions is false."

Shaltiel's counterexample applies to many other models including boolean circuits and communication protocols. A broad avenue for future work is to understand how this counterexample can be similarly circumvented in these models by working with more fine-grained notions of computation cost. Consider for example boolean circuit complexity and Yao's XOR lemma [36], which states that if f is mildly hard to approximate with size-s circuits w.r.t. μ, then $f^{\oplus k}$ is extremely hard to approximate with size-s' circuits w.r.t. μ^k. A well-known downside of this important result is that it only holds for $s' \ll s$. Indeed, Shaltiel's counterexample shows that it cannot hold for $s' \gg s$, at least not for the standard notion of circuit size. Extrapolating from our work, can we prove a strong XOR lemma for boolean circuits by considering a notion of the "expected size" of a circuit $C : \{\pm 1\}^n \to \{\pm 1\}$ with respect to a distribution μ over $\{\pm 1\}^n$? A natural approach is to consider the standard notion of the expected runtime of a Turing machine with respect to a distribution over inputs and have the Cook–Levin theorem guide us towards an appropriate analogue for circuit size.

On a more technical level, a crucial ingredient in our work is the first use of Impagliazzo's Hardcore Theorem within the context of query complexity (and indeed, to our knowledge, the first use of it outside of circuit complexity). Could this powerful theorem be useful for other problems in query complexity, possibly when used in conjunction with our new resilience lemma?

7 Preliminaries

We use $[n]$ to denote the set $\{1, 2, \dots, n\}$ and **bold font** (e.g $\boldsymbol{x} \sim \mathcal{D}$) to denote random variables. For any distribution μ, we use μ^k to denote k-fold the product distribution $\mu \times \cdots \times \mu$.

For any function $f : \{\pm 1\}^n \to \{\pm 1\}$, we use $f^{\otimes k} : (\{\pm 1\}^n)^k \to \{\pm 1\}^k$ to denote its k-fold direct product,

$$f^{\otimes k}(X^{(1)}, \dots, X^{(k)}) \coloneqq \big(f(X^{(1)}), \dots, f(X^{(k)})\big).$$

Similarly, we use $f^{\oplus k} : (\{\pm 1\}^n)^k \to \{\pm 1\}$ to denote its k-fold direct sum,

$$f^{\oplus k}(X^{(1)}, \dots, X^{(k)}) \coloneqq \prod_{i \in [k]} f(X^{(i)}).$$

▶ **Definition 14** (Bernoulli distribution). *For any $\delta \in [0,1]$, we write* $\mathrm{Ber}(\delta)$ *to denote the distribution of $\boldsymbol{z}$ where $\boldsymbol{z} = 1$ with probability δ and 0 otherwise.*

▶ **Definition 15** (Binomial distribution). *For any $k \in \mathbb{N}$, $\delta \in [0,1]$, we write* $\mathrm{Bin}(k,\delta)$ *to denote the sum of k independent random variables drawn from* $\mathrm{Ber}(\delta)$.

▶ **Fact 2** (Chernoff bound). *Let $\boldsymbol{z}_1, \dots, \boldsymbol{z}_k$ be independent and each bounded within $[0,1]$ and $\boldsymbol{Z} := \sum_{i\in[k]} \boldsymbol{z}_i$. For any threshold $t \le \mu := \mathbb{E}[\boldsymbol{Z}]$,*

$$\Pr[\boldsymbol{Z} \le t] \le \exp\Big(-\tfrac{(\mu-t)^2}{2\mu}\Big).$$

Similar bounds for the probability $\boldsymbol{Z}$ exceeds its mean hold. For example,

$$\Pr[\boldsymbol{Z} \ge 2\mu] \le \exp\big(-\tfrac{\mu}{3}\big).$$

Furthermore, the above bounds also hold for any random variable $\boldsymbol{Y}$ satisfying $\mathbb{E}[e^{\lambda \boldsymbol{Y}}] \le \mathbb{E}[e^{\lambda \boldsymbol{Z}}]$ for all $\lambda \in \mathbb{R}$.

7.1 Randomized vs. deterministic decision trees

We will prove all of our results with respect to the expected depth of a *randomized* decision tree. In this subsection, we formally define *deterministic* and *randomized* decision trees and prove that our results easily extend to the deterministic setting.

▶ **Definition 16** (Deterministic decision tree). *A deterministic decision tree, $T : \{\pm1\}^n \to \{\pm1\}$, is a binary tree with two types of nodes: Internal nodes each query some x_i for $i \in [n]$ and have two children whereas leaf nodes are labeled by a bit $b \in \{\pm1\}$ and have no children. On input $x \in \{\pm1\}^n$, $T(x)$ is computed as follows: We proceed through T starting at the root. Whenever at an internal node that queries the i^{th} coordinate, we proceed to the left child if $x_i = -1$ and right child if $x_i = +1$. Once we reach a leaf, we output the label of that leaf.*

▶ **Definition 17** (Randomized decision tree). *A randomized decision tree, $\mathcal{T} : \{\pm1\}^n \to \{\pm1\}$, is distribution over deterministic decision trees. On input $x \in \{\pm1\}^n$, it first draws $\boldsymbol{T} \sim \mathcal{T}$ and then outputs $\boldsymbol{T}(x)$.*

▶ **Definition 18** (Expected depth). *For any deterministic decision tree $T : \{\pm1\}^n \to \{\pm1\}$ and distribution μ on $\{\pm1\}^n$, we use $\overline{\mathrm{Depth}}^{\mu}(T)$ to denote the expected depth of T, which is the expected number of coordinates T queries on a random input $\boldsymbol{x} \sim \mu$. Similarly, for a randomized decision tree $\mathcal{T}$, $\overline{\mathrm{Depth}}^{\mu}(\mathcal{T}) := \mathbb{E}_{\boldsymbol{T}\sim\mathcal{T}}[\overline{\mathrm{Depth}}^{\mu}(\boldsymbol{T})]$.*

We write $\overline{\mathrm{Depth}}^{\mu}(\cdot,\cdot)$ to denote the minimum expected depth of *any* decision tree, including randomized decision trees. Thus, all of our main results, as written, hold for *randomized* decision trees; however, equivalent statements are true if we restrict ourselves to only *deterministic* decision trees, with only a small change in constants, as easily seen from the following claim.

▷ Claim 19. For any $f : \{\pm1\}^n \to \{\pm1\}$, distribution μ, and constant ε,

$$\overline{\mathrm{Depth}}^{\mu}(f, 2\varepsilon) \le \overline{\mathrm{Depth}}^{\mu}_{\mathrm{det}}(f, 2\varepsilon) \le 2 \cdot \overline{\mathrm{Depth}}^{\mu}(f, \varepsilon).$$

Proof. The left-most inequality follows immediately from that fact that any deterministic decision tree is also a randomized decision tree. For the second inequality, given any randomized decision tree $\mathcal{T}$ with error ε and expected depth d, we'll construct a deterministic decision tree T with error at most 2ε and expected depth at most $2d$. First, we decompose the expected error of $\mathcal{T}$:

$$\varepsilon = \Pr_{\boldsymbol{x}\sim\mu}[\mathcal{T}(\boldsymbol{x}) \neq f(\boldsymbol{x})] = \mathop{\mathbb{E}}_{\boldsymbol{T}\sim\mathcal{T}}\Big[\Pr_{\boldsymbol{x}\sim\mu}[\boldsymbol{T}(\boldsymbol{x}) \neq f(\boldsymbol{x})]\Big].$$

Applying Markov's inequality, if we sample $\boldsymbol{T} \sim \mathcal{T}$, with probability at least $1/2$, it has error at most ε. Similarly, since the expected depth of $\mathcal{T}$ is d, with probability at least $1/2$, $\boldsymbol{T}$ will have expected depth at most $2d$. By union bound, there is a nonzero probability that we choose a single (deterministic) tree with error at most 2ε and expected depth at most $2d$. ◁

Claim 19 immediately allows direct sum theorems for *randomized* decision trees to also apply to *deterministic* decision trees.

▶ **Corollary 20** (Randomized direct sum theorems imply deterministic ones). *Suppose that a randomized direct sum theorem of the following form holds. For a function $f : \{\pm1\}^n \to \{\pm1\}$, distribution μ over $\{\pm1\}^n$, $k \in \mathbb{N}$, and constants ε, δ, M,*

$$\overline{\mathrm{Depth}}^{\mu^{\otimes k}}(f^{\otimes k}, \varepsilon) \geq M \cdot \overline{\mathrm{Depth}}^{\mu}(f, \delta).$$

Then,

$$\overline{\mathrm{Depth}}_{\mathrm{det}}^{\mu^{\otimes k}}(f^{\otimes k}, \varepsilon) \geq \overline{\mathrm{Depth}}^{\mu^{\otimes k}}(f^{\otimes k}, \varepsilon) \geq M \cdot \overline{\mathrm{Depth}}^{\mu}(f, \delta) \geq \tfrac{M}{2} \cdot \overline{\mathrm{Depth}}_{\mathrm{det}}^{\mu}(f, 2\delta).$$

With Corollary 20 in mind, the remainder of this paper will only consider *randomized* decision trees.

8 Proof of Theorem 2

The purpose of this section is to prove our direct sum theorem (Theorem 2) showing that if f is hard to compute, than $f^{\otimes k}$ is even harder to compute. We will in fact prove a *threshold* direct sum theorem, showing that it is hard even to get *most* of the k copies correct. To formalize this, we generalize the notation $\overline{\mathrm{Depth}}(\cdot,\cdot)$ to take in an additional threshold parameter t specifying how many blocks we allow to be wrong.

▶ **Definition 21.** *For any function $f : \{\pm1\}^n \to \{\pm1\}$, error ε, and threshold $t \in \mathbb{N}$, we use $\overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes}, \varepsilon, t)$ to denote the minimum expected depth of a tree $T : (\{\pm1\}^n)^k \to \{\pm1\}^k$ satisfying*

$$\Pr_{\boldsymbol{X}\sim\mu^k}\big[\|T(\boldsymbol{X}) - f^{\otimes k}(\boldsymbol{X})\|_0] > t\big] \leq \varepsilon.$$

▶ **Theorem 22** (A strong threshold direct sum theorem for query complexity). *For every function $f : \{\pm1\}^n \to \{\pm1\}$, distribution μ over $\{\pm1\}^n$, $k \in \mathbb{N}$, and $\gamma, \delta \in (0,1)$,*

$$\overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes k}, 1 - e^{-\Omega(\delta k)} - \gamma, \Omega(\delta k)) \geq \Omega\left(\frac{\gamma^2 k}{\log(1/\delta)}\right) \cdot \overline{\mathrm{Depth}}^{\mu}(f, \delta).$$

Note that the threshold of $\Omega(\delta k)$ is within a constant factor of optimal, as repeating an algorithm that errs δ fraction of the time k times will lead to an average of δk mistakes. Theorem 22 implies our standard strong direct sum theorem (Theorem 2) because

$$\overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes k}, \varepsilon, t) \geq \overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes k}, \varepsilon, 0) = \overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes k}, \varepsilon)$$

for any $t \geq 0$ and $\varepsilon > 0$.

8.1 The structure of this section

By the Hardcore Theorem (Theorem 6), proving Theorem 22 reduces to proving:

▶ **Theorem 23** (Hardness of $f^{\otimes k}$ in terms of a hardcore measure for f)**.** *Suppose that $f : \{\pm 1\}^n \to \{\pm 1\}$ has an (γ, d)-hardcore measure w.r.t μ of density δ. Then, for any $T : (\{\pm 1\}^n)^k \to \{\pm 1\}^k$ with $\overline{\mathrm{Depth}}^{\mu^k}(T) \le kd$,*

$$\Pr_{\boldsymbol{X} \sim \mu^k}\left[\|T(\boldsymbol{X}) - f^{\otimes k}(\boldsymbol{X})\|_0 \le \tfrac{\delta k}{10}\right] \le e^{-\frac{\delta k}{10}} + 10\gamma.$$

This section is therefore devoted to proving Theorem 23. As discussed in Section 5, our proof tracks two key quantities: we will analyze how *hardcore density* (Definition 8) and *hardcore advantage* (Definition 9) are distributed over the leaves of T. This proof will be broken into three steps:

1. In Section 8.2, we prove Claim 12 and Lemma 13, which aim to understand the distributions of the hardcore density and hardcore advantage of a random leaf of T.
2. In Section 8.3, we derive an expression for the probability T makes surprisingly few mistakes as a function of the hardcore density and hardcore advantage at each leaf. This generalizes Lemma 10.
3. In Section 8.4, we combine the above to prove Theorem 23.

8.2 How hardcore density and advantage distribute over the leaves

We begin with proving our resilience lemma for hardcore density. Roughly speaking, this will say that for any tree $T : (\{\pm 1\}^n)^k \to \{\pm 1\}^k$, the hardcore density of a random leaf concentrates around δk. We recall the definition of hardcore density.

▶ **Definition 24** (Hardcore density at ℓ, Definition 8 restated)**.** *For any tree $T : (\{\pm 1\}^n)^k \to \{\pm 1\}^k$, hardcore measure $H : \{\pm 1\}^n \to [0, 1]$, distribution μ on $\{\pm 1\}^n$, $i \in [k]$, and leaf ℓ of T, the hardcore density at ℓ in the ith block is the quantity:*

$$\mathrm{Dens}_H(\ell, i) \coloneqq \mathop{\mathbb{E}}_{\boldsymbol{X} \sim \mu^k}\left[H(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \ell\right].$$

The total hardcore density at ℓ is the quantity $\mathrm{Dens}_H(\ell) \coloneqq \displaystyle\sum_{i=1}^{k} \mathrm{Dens}_H(\ell, i)$.

The distribution over leaves in the resilience lemma is the canonical distribution.

▶ **Definition 25** (Canonical distribution)**.** *For any tree T and distribution μ over T's domain, we write $\mu(T)$ to denote the distribution over leaves of T where:*

$$\Pr_{\boldsymbol{\ell} \sim \mu^k(T)}[\boldsymbol{\ell} = \ell] = \Pr_{\boldsymbol{X} \sim \mu}[\boldsymbol{X} \text{ reaches } \ell].$$

▶ **Lemma 26** (Resilience lemma, generalization of Lemma 13)**.** *For any $T : (\{\pm 1\}^n)^k \to \{\pm 1\}^k$, hardcore measure H, distribution μ, and convex $\Phi : \mathbb{R} \to \mathbb{R}$,*

$$\mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)}[\Phi(\mathrm{Dens}_H(\boldsymbol{\ell}))] \le \mathop{\mathbb{E}}_{\boldsymbol{z} \sim \mathrm{Bin}(k, \delta)}[\Phi(\boldsymbol{z})]$$

where $\delta \coloneqq \mathbb{E}_{\boldsymbol{x} \sim \mu}[H(\boldsymbol{x})]$ is the density of H w.r.t. μ.

Proof. Draw $\boldsymbol{X} \sim \mu^k$. Then, for each $i \in [k]$, independently draw $\boldsymbol{z}_i \sim \mathrm{Ber}(H(\boldsymbol{X}^{(i)}))$. Note that, for any leaf ℓ and $i \in [k]$,

$$\mathrm{Dens}_H(\ell, i) \coloneqq \mathop{\mathbb{E}}_{\boldsymbol{X} \sim \mu^k}[H(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \boldsymbol{\ell}] = \mathop{\mathbb{E}}_{\boldsymbol{X} \sim \mu^k}[\boldsymbol{z}_i \mid \boldsymbol{X} \text{ reaches } \boldsymbol{\ell}].$$

By the above equality and definition $\mathrm{Dens}_H(\ell) = \sum_{i \in [k]} \mathrm{Dens}_H(\ell, i)$,

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)}[\Phi(\mathrm{Dens}_H(\boldsymbol{\ell}))] &= \mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)}\left[\Phi\left(\mathop{\mathbb{E}}_{\boldsymbol{X} \sim \mu^k}\left[\sum_{i \in [k]} \boldsymbol{z}_i \mid \boldsymbol{X} \text{ reaches } \boldsymbol{\ell}\right]\right)\right] \\
&\leq \mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)}\left[\mathop{\mathbb{E}}_{\boldsymbol{X} \sim \mu^k}\left[\Phi\left(\sum_{i \in [k]} \boldsymbol{z}_i \mid \boldsymbol{X} \text{ reaches } \boldsymbol{\ell}\right)\right]\right] \\
&\qquad\qquad \text{(Jensen's inequality)} \\
&= \mathop{\mathbb{E}}_{\boldsymbol{X} \sim \mu^k}\left[\Phi\left(\sum_{i \in [k]} \boldsymbol{z}_i\right)\right]. \qquad \text{(Law of total expectation)}
\end{aligned}$$

Note that the last line holds precisely for the distribution $\mu^k(T)$ defined in Definition 25, which is why we use that distribution.

Since $\boldsymbol{X}$ is drawn from a product distribution, and $\boldsymbol{z}_i$ depends on only the i^{th} coordinate of $\boldsymbol{X}$, $\boldsymbol{z}_1, \dots, \boldsymbol{z}_k$ are independent. Furthermore, each has mean $\mathbb{E}_{\boldsymbol{x} \sim \mu}[H(\boldsymbol{x})] = \delta$. Therefore, $\sum_{i \in [k]} \boldsymbol{z}_i$ is distributed according to $\mathrm{Bin}(k, \delta)$. ◀

A couple of remarks about the above Lemma: First, it implies that $\mathrm{Dens}_H(\boldsymbol{\ell})$ concentrates around δk. Since $z \mapsto e^{\lambda z}$ is convex for any $\lambda \in \mathbb{R}$, Lemma 26 implies that the moment generating function of $\mathrm{Dens}_H(\boldsymbol{\ell})$ is dominated by that of $\mathrm{Bin}(k, \delta)$. This means that Chernoff bounds that hold for $\mathrm{Bin}(k, \delta)$ also hold for $\mathrm{Dens}_H(\boldsymbol{\ell})$. In particular, the statement of Lemma 13 is a consequence of the Chernoff bound given in Fact 2.

Second, the proof of Lemma 26 does not make heavy use of the decision tree structure of T. It only uses that the leaves of T partition $(\{\pm 1\}^n)^k$, and so may find uses for other models that partition the domain.

Depth amplification for hardcore advantage

While the resilience lemma gives a fairly fine-grained understanding of how hardcore density distributes among the leaves, our guarantee for hardcore advantage are more coarse – that its expectation over the leaves is bounded.

▶ **Definition 27** (Hardcore advantage at ℓ, Definition 9 restated). *For any tree $T : (\{\pm 1\}^n)^k \to \{\pm 1\}^k$, hardcore measure $H : \{\pm 1\}^n \to [0, 1]$, distribution μ on $\{\pm 1\}^n$, $i \in [k]$, and leaf ℓ of T, the hardcore advantage at ℓ in the ith block is the quantity:*

$$\mathrm{Adv}_H(\ell, i) \coloneqq \left| \mathop{\mathbb{E}}_{\boldsymbol{X} \sim \mu^k} \left[f(\boldsymbol{X}^{(i)}) T(\boldsymbol{X})_i H(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \textit{ reaches } \ell \right] \right|. \tag{9}$$

The total hardcore advantage at ℓ is the quantity $\mathrm{Adv}_H(\ell) \coloneqq \sum_{i=1}^{k} \mathrm{Adv}_H(\ell, i)$.

▶ **Lemma 28** (Expected total hardcore advantage, Claim 12 restated). *If H is a (γ, d)-hardcore measure for f of density δ w.r.t. μ and the expected depth of T is at most dk, then*

$$\mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)}[\mathrm{Adv}_H(\boldsymbol{\ell})] \leq \gamma \mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)}[\mathrm{Dens}_H(\boldsymbol{\ell})] = \gamma \delta k.$$

Proof of Lemma 28. By contrapositive. Suppose there exists $T_{\text{large}} : (\{\pm 1\}^n)^k \to \{\pm 1\}^k$ making dk queries on average w.r.t. μ^k for which,

$$\mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)}[\mathrm{Adv}_H(\boldsymbol{\ell})] > \gamma \mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)}[\mathrm{Dens}_H(\boldsymbol{\ell})] = \gamma \delta k.$$

Then, we'll show there exists $T_{\text{small}} : \{\pm 1\}^n \to \{\pm 1\}$ making d queries on average w.r.t. μ for which

$$\mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mu}[f(\boldsymbol{x}) T_{\text{small}}(\boldsymbol{x}) H(\boldsymbol{x})] > \gamma \cdot \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mu}[H(\boldsymbol{x})] = \gamma \delta.$$

Before constructing T_{small}, we observe that we can assume, without loss of generality, that for every leaf ℓ of T, that we can remove the absolute value from Equation (9); i.e. that

$$\mathrm{Adv}_H(\ell, i) = \mathop{\mathbb{E}}_{\boldsymbol{X} \sim \mu^k}\left[f(\boldsymbol{X}^{(i)}) T(\boldsymbol{X})_i H(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \ell\right]$$

Otherwise, we could modify this leaf by flipping the label of $T(X)_i$ whenever X reaches a leaf where the above quantity is negative. This does not change the hardcore advantage, so this new T still satisfies our assumption.

T_{small} will be a randomized algorithm. Upon receiving the input $x \in \{\pm 1\}^n$, it samples $\boldsymbol{X} \sim \mu^k$ and $\boldsymbol{i} \sim \mathrm{Unif}([k])$, and then constructs $\boldsymbol{X}(x, \boldsymbol{i})$ by inserting x into the $\boldsymbol{i}^{\text{th}}$ block of $\boldsymbol{X}$,

$$(\boldsymbol{X}(x, i))^{(j)} = \begin{cases} \boldsymbol{X}^{(j)} & \text{if } j \neq i \\ x & \text{if } j = \boldsymbol{i}. \end{cases}$$

Then, $T_{\text{small}}(x)$ outputs $T_{\text{large}}(\boldsymbol{X}(x, \boldsymbol{i}))_{\boldsymbol{i}}$.

Our analysis of T_{small} relies on the following simple observation: If we sample $\boldsymbol{x} \sim \mu$, then even conditioning on any choice of $\boldsymbol{i} = i$, the distribution of $\boldsymbol{X}(\boldsymbol{x}, \boldsymbol{i})$ is μ^k. This also means that $\boldsymbol{X}(\boldsymbol{x}, \boldsymbol{i})$ and $\boldsymbol{i}$ are independent.

We claim that T_{small} has the two desired properties; low expected number of queries, and high accuracy on H. To bound the expected number of queries T_{small} makes on an input $\boldsymbol{x} \sim \mu$, we use that $\boldsymbol{X}(\boldsymbol{x}, \boldsymbol{i})$ is distributed according to μ^k. Therefore, $T_{\text{large}}(\boldsymbol{X}(\boldsymbol{x}, \boldsymbol{i}))$ makes, on average, dk queries. Expanded, we have that,

$$\sum_{i \in [k], j \in [n]} \Pr[T_{\text{large}}(\boldsymbol{X}(\boldsymbol{x}, \boldsymbol{i})) \text{ queries } \boldsymbol{X}(\boldsymbol{x})_j^{(i)}] = dk.$$

Whereas, the number of queries $T_{\text{small}}(\boldsymbol{x})$ makes only counts queries to the $\boldsymbol{i}^{\text{th}}$ block, and is therefore,

$$\begin{aligned}
&\sum_{i \in [k], j \in [n]} \Pr\left[T_{\text{large}}(\boldsymbol{X}(\boldsymbol{x}, \boldsymbol{i})) \text{ queries } \boldsymbol{X}(\boldsymbol{x}, \boldsymbol{i})_j^{(i)} \cdot \mathbb{1}[\boldsymbol{i} = i]\right] \\
&\qquad = \sum_{i \in [k], j \in [n]} \Pr\left[T_{\text{large}}(\boldsymbol{X}(\boldsymbol{x}, \boldsymbol{i}))) \text{ queries } \boldsymbol{X}(\boldsymbol{x}, \boldsymbol{i})_j^{(i)}\right] \cdot \frac{1}{k} \\
&\qquad = d.
\end{aligned}$$

In the above, the first equality uses that $\boldsymbol{i}$ is independent of $\boldsymbol{X}(\boldsymbol{x}, \boldsymbol{i})$, and so is still uniform on $[k]$ even conditioned on which queries T_{large} makes.

Lastly, we verify that T_{small} has high accuracy on the hardcore measure.

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})T_{\text{small}}(\boldsymbol{x},\boldsymbol{i}))H(\boldsymbol{x})] &= \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})T_{\text{large}}(\boldsymbol{X}(\boldsymbol{x},\boldsymbol{i})))_{\boldsymbol{i}}H(\boldsymbol{x})] && \text{(Definition of } T_{\text{small}})\\
&= \mathop{\mathbb{E}}_{\boldsymbol{i}\sim[k]}\left[\mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}\left[f(\boldsymbol{X}^{(\boldsymbol{i})})T_{\text{large}}(\boldsymbol{X})_{\boldsymbol{i}}H(\boldsymbol{X}^{(\boldsymbol{i})})\right]\right]\\
& && (\boldsymbol{i},\boldsymbol{X}(\boldsymbol{x}) \text{ are independent})\\
&= \mathop{\mathbb{E}}_{\boldsymbol{i}\sim[k]}\left[\mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim\mu^k(T_{\text{large}})}[\mathrm{Adv}_H(\boldsymbol{\ell},\boldsymbol{i})]\right] && \text{(Definition 27)}\\
&= \frac{1}{k}\mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim\mu^k(T_{\text{large}})}[\mathrm{Adv}_H(\boldsymbol{\ell})] > \gamma\delta. && \blacktriangleleft
\end{aligned}$$

8.3 Understanding the error in terms of hardcore density and advantage

To state the main result of this subsection, we'll define the following distribution for the sum of independent Bernoulli random variables.

▶ **Definition 29.** *For any $p \in [0,1]^k$, we use $\mathrm{BerSum}(p)$ to denote the distribution of $\boldsymbol{z} := \boldsymbol{z}_1 + \cdots + \boldsymbol{z}_k$ where each $\boldsymbol{z}_i$ is independently drawn from $\mathrm{Ber}(p_i)$.*

The following generalizes Lemma 10.

▶ **Lemma 30** (Accuracy in terms of hardcore density and advantage of the leaves)**.** *Let H be a (γ,d)-hardcore measure w.r.t. μ for $f : \{\pm1\}^n \to \{\pm1\}$, and $T : (\{\pm1\}^n)^k \to \{\pm1\}$ be any tree. Then, for any $t \geq 0$,*

$$\Pr_{\boldsymbol{X}\sim\mu^k}\left[\left\|T(\boldsymbol{X}) - f^{\otimes k}(\boldsymbol{X})\right\|_0 \leq t\right] \leq \mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim\mu^k(T)}\left[\Pr_{\boldsymbol{z}\sim\mathrm{BerSum}(p(\boldsymbol{\ell}))}[\boldsymbol{z} \leq t]\right]$$

where $p(\ell) \in [0,1]^k$ is the vector where

$$p(\ell)_i := \frac{\mathrm{Dens}_H(\ell,i) - \mathrm{Adv}_H(\ell,i)}{2} \qquad \text{for each } i \in [k].$$

Lemma 30 implies a generalization of Lemma 10.

▶ **Corollary 31.** *Let H be a (γ,d)-hardcore measure w.r.t. μ for $f : \{\pm1\}^n \to \{\pm1\}$, and $T : (\{\pm1\}^n)^k \to \{\pm1\}$ be any tree. Then, for any $t \geq 0$,*

$$\Pr_{\boldsymbol{X}\sim\mu^k}\left[\left\|T(\boldsymbol{X}) - f^{\otimes k}(\boldsymbol{X})\right\|_0 \leq t\right] \leq \mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim\mu^k(T)}\left[\min\left(1, \exp\left(t - \frac{\mathrm{Dens}_H(\boldsymbol{\ell}) - \mathrm{Adv}_H(\boldsymbol{\ell})}{4}\right)\right)\right].$$

Proof. The Chernoff bound of Fact 2 says that, for any $p \in [0,1]^k$ and $\mu := \sum_{i\in[k]} p_i$,

$$\Pr_{\boldsymbol{z}\sim\mathrm{BerSum}(p)}[\boldsymbol{z} \leq t] \leq \begin{cases} \exp\left(-\frac{(\mu-t)^2}{2\mu}\right) & \text{if } \mu \geq t\\ 1 & \text{otherwise.}\end{cases}$$

We want to show that the above is bounded by $\min(1, e^{t-\mu/2})$. Clearly this holds for $\mu < t$, so we need only consider the case where $\mu \geq t$

$$\exp\left(-\frac{(\mu-t)^2}{2\mu}\right) = \exp\left(-\frac{\mu^2-2t\mu+t^2}{2\mu}\right) \leq \exp\left(-\frac{\mu^2-2t\mu}{2\mu}\right) = e^{t-\mu/2}.$$

Since $\exp\left(-\frac{(\mu-t)^2}{2\mu}\right) \leq 1$ as well, it is upper bounded by $\min(1, e^{t-\mu/2})$ as desired. The desired result follows from Lemma 30 as well as $\sum_{i\in[k]} p(\ell)_i = \frac{\mathrm{Dens}_H(\ell)-\mathrm{Adv}_H(\ell)}{2}$ for every leaf ℓ of T. ◀

The main observation underlying Lemma 30 is that, if we choose an input $\boldsymbol{X} \sim \mu^k$ conditioned on reaching a leaf $\ell \in T$, that $\boldsymbol{X}$ is distributed according to a k-wise product distribution (i.e. from $\mu_1(\ell) \times \cdots \times \mu_k(\ell)$ for appropriately defined distributions). The below is essentially the same as Lemma 3.2 of [10], but we include a proof for completeness.

▷ Claim 32. For any (potentially randomized) tree $T : (\{\pm 1\}^n)^k \to \mathcal{Y}$ and leaf ℓ of T, if $\boldsymbol{X} \sim \mu^k$, then the distribution of $\boldsymbol{X}$ conditioned on reaching the leaf ℓ is a product distribution over the k blocks of $\boldsymbol{X}$.

Proof. First, if T is a randomized tree, it is as a distribution over deterministic trees. If the desired result holds for each of those deterministic trees, it also holds for T. Therefore, it suffices to consider the case where T is deterministic.

We'll prove that the distribution of $\boldsymbol{X}$ reaching any internal node *or* leaf of T is product by induction on the depth of that node. If that depth is 0, then all inputs reach it and so the desired result follows from μ^k being product.

For depth $d \geq 1$, let α be the parent of ℓ. Then α has depth $d-1$, so by the inductive hypothesis, the distribution of inputs reaching α is product. Let $i \in [k], j \in [n], b \in \{\pm 1\}$ be chosen so that an input X reaches ℓ iff it reaches α and $X^{(i)}_j = b$. Then,

$$\begin{aligned}
&\Pr[\boldsymbol{X} = X \mid \boldsymbol{X} \text{ reaches } \ell] \\
&\quad = \Pr[\boldsymbol{X} = X \mid \boldsymbol{X} \text{ reaches } \alpha] \cdot \frac{\mathbb{1}[X^{(i)}_j = b]}{\Pr[X^{(i)}_j = b \mid \boldsymbol{X} \text{ reaches } \alpha]} \\
&\quad = \frac{\mathbb{1}[X^{(i)}_j = b]}{\Pr[X^{(i)}_j = b \mid \boldsymbol{X} \text{ reaches } \alpha]} \cdot \prod_{\ell \in [k]} \Pr[\boldsymbol{X}^{(\ell)} = X^{(\ell)} \mid \boldsymbol{X} \text{ reaches } \alpha] \qquad \text{(Inductive hypothesis)} \\
&\quad = \left(\prod_{\ell \neq i} \Pr[\boldsymbol{X}^{(\ell)} = X^{(\ell)} \mid \boldsymbol{X} \text{ reaches } \alpha]\right) \cdot \frac{\Pr[\boldsymbol{X}^{(i)} = X^{(i)} \mid \boldsymbol{X} \text{ reaches } \alpha] \cdot \mathbb{1}[X^{(i)}_j = b]}{\Pr[X^{(i)}_j = b \mid \boldsymbol{X} \text{ reaches } \alpha]} \\
&\quad = \left(\prod_{\ell \neq i} \Pr[\boldsymbol{X}^{(\ell)} = X^{(\ell)} \mid \boldsymbol{X} \text{ reaches } \alpha]\right) \cdot \Pr[\boldsymbol{X}^{(i)} = X^{(i)} \mid \boldsymbol{X} \text{ reaches } \alpha, \boldsymbol{X}^{(i)}_j = b].
\end{aligned}$$

The above is decomposed as a product over the k components of X, so is a product distribution. ◁

We conclude this subsection with a proof of Lemma 30.

Proof of Lemma 30. Consider any leaf ℓ of T. We wish to compute the probability that $T(\boldsymbol{X})$ makes less than t mistakes on $f^{\otimes k}(\boldsymbol{X})$ given that $\boldsymbol{X}$ reaches the leaf ℓ. On this leaf, T outputs a single vector $y \in \{\pm 1\}^k$. Meanwhile, by Claim 32, the distribution of $\boldsymbol{X}$ is product over the blocks, and so $f(\boldsymbol{X}^{(1)}), \ldots, f(\boldsymbol{X}^{(k)})$ are independent. Define $q(\ell) \in [0,1]^k$ as,

$$q(\ell)_i := \Pr_{\boldsymbol{X} \sim \mu^k}[y_i \neq f(\boldsymbol{X}^{(i)})].$$

Then,

$$\Pr_{\boldsymbol{X} \sim \mu^k}\left[\left\|T(\boldsymbol{X}), f^{\otimes k}(\boldsymbol{X})\right\|_0 \leq t \mid \boldsymbol{X} \text{ reaches } \ell\right] = \Pr_{\boldsymbol{z} \sim \mathrm{BerSum}(q(\ell))}[\boldsymbol{z} \leq t].$$

For $\boldsymbol{z} \sim \mathrm{BerSum}(q)$, the probability $\boldsymbol{z} \leq t$ is monotonically decreasing in each q_i. Therefore, it suffices to show that $q(\ell)_i \geq p(\ell)_i$ for each $i \in [k]$. We compute,

$$\begin{aligned}
q(\ell)_i &= \Pr_{\boldsymbol{X} \sim \mu^k}[T(\boldsymbol{X})_i \neq f(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \ell] \\
&= \frac{1 - \mathbb{E}_{\boldsymbol{X} \sim \mu^k}[T(\boldsymbol{X})_i f(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \ell]}{2} \qquad (T(X)_i, f(X^{(i)} \in \{\pm 1\})
\end{aligned}$$

Separating the above expectation into two pieces, for $\boldsymbol{X}$ drawn from μ^k conditioned in $\boldsymbol{X}$ reaching ℓ,

$$\begin{aligned}\mathbb{E}[T(\boldsymbol{X})_i f(\boldsymbol{X}^{(i)})] &= \mathbb{E}[T(\boldsymbol{X})_i f(\boldsymbol{X}^{(i)}) H(\boldsymbol{X}^{(i)})] + \mathbb{E}[T(\boldsymbol{X})_i f(\boldsymbol{X}^{(i)})(1 - H(\boldsymbol{X}^{(i)}))] \\ &\le \mathrm{Adv}_H(\ell, i) + \mathbb{E}[T(\boldsymbol{X})_i f(\boldsymbol{X}^{(i)})(1 - H(\boldsymbol{X}^{(i)}))] \qquad \text{(Definition 9)} \\ &\le \mathrm{Adv}_H(\ell, i) + \mathbb{E}[(1 - H(\boldsymbol{X}^{(i)}))] = 1 + \mathrm{Adv}_H(\ell, i) - \mathrm{Dens}_H(\ell, i).\end{aligned}$$

Therefore,

$$q(\ell)_i \ge \frac{\mathrm{Dens}_H(\ell, i) - \mathrm{Adv}_H(\ell, i)}{2} = p(\ell)_i. \qquad \blacktriangleleft$$

8.4 Completing the proof of the threshold direct sum theorem

In this subsection, we complete the proof of Theorem 23. Throughout this section, we'll use the following function:

$$g_t(z) \coloneqq \min(1, e^{t - z/4}).$$

By using Corollary 31, it suffices to show that

$$\mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)} \left[g_{\delta k/10}(\mathrm{Dens}_H(\boldsymbol{\ell}) - \mathrm{Adv}_H(\boldsymbol{\ell})) \right] \le e^{-\delta k/10} + 1 - \gamma. \tag{10}$$

Recall that we have much information about how $\mathrm{Dens}_H(\boldsymbol{\ell})$ distributes over the leaves via Lemma 26, but a coarser understanding of how $\mathrm{Adv}_H(\boldsymbol{\ell})$ distributes via Lemma 28. Because of this, we will first bound the above equation where the $\mathrm{Adv}_H(\boldsymbol{\ell})$ is set to 0 and analyze how much including that term affects the result.

▶ **Lemma 33.** *For any tree $T : (\{\pm 1\}^n)^k \to \{\pm 1\}^k$ and hardcore measure of density δ w.r.t. distribution μ,*

$$\mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)} \left[g_{\delta k/10}(\mathrm{Dens}_H(\boldsymbol{\ell})) \right] \le e^{-.121\delta k}.$$

Proof. We bound,

$$\mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)} \left[\min\left(1, \exp\left(\delta k/10 - \frac{\mathrm{Dens}_H(\boldsymbol{\ell})}{4}\right)\right) \right] \le e^{\delta k/10} \cdot \mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)} \left[e^{-\mathrm{Dens}_H(\ell)/4} \right].$$

Since $z \mapsto e^{-z/4}$ is convex, we can use Lemma 26 to bound the above using the moment generating function of the binomial distribution,

$$\begin{aligned}\mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)} \left[e^{-\mathrm{Dens}_H(\boldsymbol{\ell})/4} \right] &\le \mathop{\mathbb{E}}_{\boldsymbol{z} \sim \mathrm{Bin}(k, \delta)} [e^{-\boldsymbol{z}/4}] \\ &= (1 - \delta(1 - e^{-1/4}))^k \\ &\le e^{-(1 - e^{-1/4})\delta k}.\end{aligned}$$

Combining the above,

$$\mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)} \left[g_{\delta k/10}(\mathrm{Dens}_H(\boldsymbol{\ell})) \right] \le e^{-\delta k(1 - e^{-1/4} - 1/10)} \le e^{-0.121\delta k}. \qquad \blacktriangleleft$$

Next, we prove a Lipschitz-style bound for g. This will be useful in incorporating $\mathrm{Adv}_H(\boldsymbol{\ell})$ to Equation (10).

▶ **Proposition 34.** *For any* $z, \Delta, t \geq 0$,

$$g_t(z - \Delta) \leq g_t(z) + \Delta/4. \tag{11}$$

Furthermore, if $z \geq 5t$, *then*

$$g_t(z - \Delta) \leq g_t(z) + \Delta/t. \tag{12}$$

Proof. Equation (11) follows from the (1/4)-Lipschitzness of $g_t(z)$.

For Equation (12), fix any choice of $z \geq 5t$. We want to show that for any choice of Δ,

$$\frac{g_t(z-\Delta) - g_t(z)}{\Delta} \leq 1/t.$$

We claim that the left hand side of the above inequality is maximized when $\Delta = z - 4t$. When Δ is increased beyond $z - 4t$, the numerator remains constant (because $g_t(z)$ is constant for any $z \leq 4t$, but the denominator increases, so the maximum cannot occur at at any $\Delta > z - 4t$. On the other hand, $g_t(z)$ is convex when restricted to the domain $[4t, \infty)$, so the maximum cannot occur at any $\Delta < z - 4t$. Therefore, it suffices to consider $\Delta = z - 4t$, in which case,

$$\frac{g_t(z-\Delta) - g_t(z)}{\Delta} = \frac{1 - g_t(z)}{\Delta} \leq \frac{1}{\Delta} \leq 1/t. \qquad ◀$$

We are now ready to prove the main result of this section.

Proof of Theorem 23. By applying Corollary 31,

$$\Pr_{\boldsymbol{X} \sim \mu^k}\left[\left\|T(\boldsymbol{X}) - f^{\otimes k}(\boldsymbol{X})\right\|_0 \leq \delta k/10\right] \leq \mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)}\left[g_{\delta k/10}(\mathrm{Dens}_H(\boldsymbol{\ell}) - \mathrm{Adv}_H(\boldsymbol{\ell}))\right].$$

First, we consider the case where $\delta k \leq 40$. Here, by applying Equation (11),

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)}\left[g_{\delta k/10}(\mathrm{Dens}_H(\boldsymbol{\ell}) - \mathrm{Adv}_H(\boldsymbol{\ell}))\right] \\
&\qquad \leq \mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)}\left[g_{\delta k/10}(\mathrm{Dens}_H(\boldsymbol{\ell}))\right] + \tfrac{1}{4} \cdot \mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim \mu^k(T)}[\mathrm{Adv}_H(\boldsymbol{\ell})] \\
&\qquad \leq e^{-0.121\delta k} + \gamma \delta k/4 && \text{(Lemmas 28 and 33)} \\
&\qquad \leq e^{-\delta k/10} + 10\gamma && (\delta k \leq 40)
\end{aligned}$$

When $\delta k > 40$, we break down the desired result into two pieces, depending on whether $\mathrm{Dens}_H(\boldsymbol{\ell})$ is small or large. For the piece where $\mathrm{Dens}_H(\boldsymbol{\ell})$ is small, we just use that $g(\cdot)$ is bounded between 0 and 1 which means $g(z) - g(z - \Delta) \leq 1$,

$$\begin{aligned}
&\mathbb{E}\left[g_{\delta k/10}(\mathrm{Dens}_H(\boldsymbol{\ell}) - \mathrm{Adv}_H(\boldsymbol{\ell})) \cdot \mathbb{1}[\mathrm{Dens}_H(\boldsymbol{\ell}) \leq \delta k/2]\right] \\
&\qquad \leq \mathbb{E}\left[g_{\delta k/10}(\mathrm{Dens}_H(\boldsymbol{\ell})) \cdot \mathbb{1}[\mathrm{Dens}_H(\boldsymbol{\ell}) \leq \delta k/2]\right] + \Pr[\mathrm{Dens}_H(\boldsymbol{\ell}) \leq \delta k/2] \\
&\qquad \leq \mathbb{E}\left[g_{\delta k/10}(\mathrm{Dens}_H(\boldsymbol{\ell})) \cdot \mathbb{1}[\mathrm{Dens}_H(\boldsymbol{\ell}) \leq \delta k/2]\right] + e^{-\delta k/8}. && \text{(Lemma 13)}
\end{aligned}$$

For the piece where $\mathrm{Dens}_H(\boldsymbol{\ell})$ is large, we use Equation (12)

$$\begin{aligned}
&\mathbb{E}\left[g_{\delta k/10}(\mathrm{Dens}_H(\boldsymbol{\ell}) - \mathrm{Adv}_H(\boldsymbol{\ell})) \cdot \mathbb{1}[\mathrm{Dens}_H(\boldsymbol{\ell}) > \delta k/2]\right] \\
&\quad \leq \mathbb{E}\left[g_{\delta k/10}(\mathrm{Dens}_H(\boldsymbol{\ell})) \cdot \mathbb{1}[\mathrm{Dens}_H(\boldsymbol{\ell}) > \delta k/2]\right] + \tfrac{10}{\delta k} \cdot \mathbb{E}[\mathrm{Adv}_H(\boldsymbol{\ell}) \cdot \mathbb{1}[\mathrm{Dens}_H(\boldsymbol{\ell}) > \delta k/2]] \\
& && \text{(Equation (12))} \\
&\quad \leq \mathbb{E}\left[g_{\delta k/10}(\mathrm{Dens}_H(\boldsymbol{\ell})) \cdot \mathbb{1}[\mathrm{Dens}_H(\boldsymbol{\ell}) > \delta k/2]\right] + \tfrac{10}{\delta k} \cdot \mathbb{E}[\mathrm{Adv}_H(\boldsymbol{\ell})] && (\mathrm{Dens}_H(\boldsymbol{\ell}) \geq 0) \\
&\quad \leq \mathbb{E}\left[g_{\delta k/10}(\mathrm{Dens}_H(\boldsymbol{\ell})) \cdot \mathbb{1}[\mathrm{Dens}_H(\boldsymbol{\ell}) > \delta k/2]\right] + \tfrac{10}{\delta k} \cdot \gamma \delta k && \text{(Lemma 28)}
\end{aligned}$$

Combining the above two pieces,

$$\begin{aligned}\mathbb{E}\left[g_{\delta k/10}(\mathrm{Dens}_H(\boldsymbol{\ell}) - \mathrm{Adv}_H(\boldsymbol{\ell}))\right] &\le \mathbb{E}\left[g_{\delta k/10}(\mathrm{Dens}_H(\boldsymbol{\ell}))\right] + e^{-\delta k/8} + 10\gamma \\ &\le e^{-\delta k/8} + e^{-0.121\delta k} + 10\gamma \qquad \text{(Lemma 33)}\end{aligned}$$

When $\delta k > 40$, $e^{-\delta k/8} + e^{-0.121\delta k} < e^{-\delta k/10}$, so we also recover the desired result in this case. ◀

9 Equivalence between direct sum theorems and XOR lemmas and the proof of Theorem 3

In this section, we prove the following claim which shows that a strong direct sum theorem implies a strong XOR lemma. We then derive Theorem 3 as a consequence of this equivalence and our strong direct sum theorem for query complexity (Theorem 1).

▷ Claim 35 (Equivalence between direct sum theorems and XOR lemmas). For every $f : \{\pm1\}^n \to \{\pm1\}$, distribution μ over $\{\pm1\}^n$, integer $k \in \mathbb{N}$, multiplicative factor $M \in \mathbb{R}$, and $\varepsilon \in (0,1)$, if the following direct sum theorem holds:

$$\overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes k}, \varepsilon) \ge M \cdot \overline{\mathrm{Depth}}^{\mu}(f, \delta).$$

then, the following XOR lemma holds:

$$\overline{\mathrm{Depth}}^{\mu^k}(f^{\oplus k}, \tfrac{\varepsilon}{2}) \ge M \cdot \overline{\mathrm{Depth}}^{\mu}(f, \delta).$$

In order to prove Claim 35 and Theorem 3, we establish a lemma which allows us to convert any decision accurately computing $f^{\oplus k}$ into a decision tree accurately computing $f^{\otimes k}$. The following definition captures this conversion.

▶ **Definition 36** (The product tree). *Given a decision tree $T : (\{\pm1\}^n)^k \to \{\pm1\}$, the k-wise product tree $\tilde{T} : (\{\pm1\}^n)^k \to \{\pm1\}^k$ is defined as follows. For the internal nodes, $\tilde{T}$ has exactly the same structure as T. For a leaf ℓ in T, the leaf vector $(\ell_1, \dots, \ell_k) \in \{\pm1\}^k$ in $\tilde{T}$ is defined by*

$$\ell_i \coloneqq \mathrm{sign}\left(\mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}[f(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \textit{ reaches } \ell]\right)$$

for all $i \in [k]$.

Intuitively, $\tilde{T}$ computes T's best guess for $f(X^{(i)})$ for each $i \in [k]$ on a given input $(X^{(1)}, \dots, X^{(k)})$. If T is really good at computing $f^{\oplus k}$ then at every leaf it should have queried enough variables to pin down f's value on each of the input blocks. The main lemma formalizes this intuition.

▶ **Lemma 37.** *For any $f : \{\pm1\}^n \to \{\pm1\}$, distribution μ over $\{\pm1\}^n$, and tree $T : (\{\pm1\}^n)^k \to \{\pm1\}$, the k-wise product tree $\tilde{T} : (\{\pm1\}^n)^k \to \{\pm1\}^k$ satisfies*

$$\Pr_{\boldsymbol{X}\sim\mu^k}[\tilde{T}(\boldsymbol{X}) = f^{\otimes k}(\boldsymbol{X})] \ge \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}[T(\boldsymbol{X}) f^{\oplus k}(\boldsymbol{X})].$$

9.1 Proofs of Claim 35 and Theorem 3 assuming Lemma 37

The following corollary of Lemma 37 implies Claim 35 and Theorem 3.

▶ **Corollary 38** (Main corollary of Lemma 37). *For all $f : \{\pm 1\}^n \to \{\pm 1\}$, $k \geq 1$, distributions μ over $\{\pm 1\}^n$, and $\varepsilon > 0$, $\overline{\mathrm{Depth}}^{\mu^k}(f^{\oplus k}, \frac{\varepsilon}{2}) \geq \overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes k}, \varepsilon)$.*

Proof. Let $\boldsymbol{A}$ be a randomized query algorithm for $f^{\oplus k}$ with error $\varepsilon/2$ and expected cost $q = \overline{\mathrm{Depth}}^{\mu^k}(f^{\oplus k}, \varepsilon/2)$. Let $\mathcal{T}$ denote the distribution over decision trees determined by $\boldsymbol{A}$. Consider the algorithm $\tilde{\boldsymbol{A}}$ which computes $f^{\otimes k}(X)$ by sampling $\boldsymbol{T} \sim \mathcal{T}$ and returning $\tilde{\boldsymbol{T}}(X)$ where $\tilde{\boldsymbol{T}}$ is the decision tree from Lemma 37. Then, the success of $\tilde{\boldsymbol{A}}$ is

$$\begin{aligned} \mathop{\mathbb{E}}_{\boldsymbol{T}\sim\mathcal{T}}\left[\Pr_{\boldsymbol{X}\sim\mu^k}[\tilde{\boldsymbol{T}}(\boldsymbol{X}) = f^{\otimes k}(\boldsymbol{X})]\right] &\geq \mathop{\mathbb{E}}_{\boldsymbol{T}\sim\mathcal{T}}\left[\mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}[\boldsymbol{T}(\boldsymbol{X}) f^{\oplus k}(\boldsymbol{X})]\right] && \text{(Lemma 37)} \\ &\geq 1 - \varepsilon \end{aligned}$$

where the last step uses the fact that advantage is $1 - 2 \cdot \mathrm{error}$. Since the structure of each $\tilde{T}$ is the same as T, the expected cost of $\tilde{\boldsymbol{A}}$ is q which completes the proof. ◀

Proofs of Claim 35 and Theorem 3. By Corollary 38,

$$\overline{\mathrm{Depth}}^{\mu^k}(f^{\oplus k}, \tfrac{\varepsilon}{2}) \geq \overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes k}, \varepsilon) \geq M \cdot \overline{\mathrm{Depth}}^{\mu}(f, \delta).$$

Theorem 3 follows immediately by applying Claim 35 to Theorem 2. ◀

▶ Remark 39 (On the necessity of the $1/2$ loss in ε in Corollary 38). One may wonder whether the $1/2$ loss in ε parameter in Corollary 38 is necessary. For example, can one show $\overline{\mathrm{Depth}}^{\mu^k}(f^{\oplus k}, 0.51\varepsilon) \geq \overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes k}, \varepsilon)$? The issue is that $\overline{\mathrm{Depth}}^{\mu^k}(f^{\oplus k}, 0.5) = 0$ for all functions $f : \{\pm 1\}^n \to \{\pm 1\}$ because the bias of $f^{\oplus k}$ is at least 0.5. So such a statement cannot hold for all f in all parameter regimes. Concretely, one can show that if f is the parity of n bits and μ is uniform over $\{\pm 1\}^n$, then $\overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes k}, \varepsilon) \geq \Omega(kn)$ for all constant $\varepsilon < 1$. Any path in a decision tree for $f^{\otimes k}$ which queries at most λkn bits for some constant $\lambda < 1$ has success probability $2^{-\Omega(k)}$. So to achieve any constant accuracy requires $\Omega(kn)$ expected depth. On the other hand, $\overline{\mathrm{Depth}}^{\mu^k}(f^{\oplus k}, 0.5) = 1 \ll \overline{\mathrm{Depth}}^{\mu^k}(f^{\otimes k}, \varepsilon)$ which is achieved by the decision tree that outputs a single constant value. Therefore, the $\varepsilon/2$ in Corollary 38 is necessary for such a statement to hold in full generality.

9.2 Proof of Lemma 37

Each ℓ_i for $i \in [k]$ satisfies

$$\begin{aligned} \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}[\ell_i \cdot f(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \ell] &= \left|\mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}[f(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \ell]\right| \\ &\geq \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}[\ell \cdot f(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \ell]. \end{aligned}$$

In particular, for all leaves ℓ of T,

$$\mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}[\tilde{T}(\boldsymbol{X})_i \cdot f(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \ell] \geq \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}[T(\boldsymbol{X}) \cdot f(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \ell]. \tag{13}$$

Therefore:

$$\begin{aligned}
&\Pr_{\boldsymbol{X}\sim\mu^k}[\tilde{T}(\boldsymbol{X}) = f^{\otimes k}(\boldsymbol{X})] \\
&\quad= \mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim\mu^k(T)}\left[\Pr_{\boldsymbol{X}\sim\mu^k}[\tilde{T}(\boldsymbol{X}) = f^{\otimes k}(\boldsymbol{X}) \mid \boldsymbol{X} \text{ reaches } \boldsymbol{\ell}]\right] \\
&\quad= \mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim\mu^k(T)}\left[\prod_{i\in[k]}\Pr_{\boldsymbol{X}\sim\mu^k}[\tilde{T}(\boldsymbol{X})_i = f(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \boldsymbol{\ell}]\right] && \text{(Claim 32)} \\
&\quad\geq \mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim\mu^k(T)}\left[\prod_{i\in[k]}\mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}[\tilde{T}(\boldsymbol{X})_i f(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \boldsymbol{\ell}]\right] \\
&\quad\geq \mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim\mu^k(T)}\left[\prod_{i\in[k]}\mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}[T(\boldsymbol{X}) f(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \boldsymbol{\ell}]\right] && \text{(Equation (13))} \\
&\quad= \mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim\mu^k(T)}\left[\mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}\left[T(\boldsymbol{X})\prod_{i\in[k]} f(\boldsymbol{X}^{(i)}) \mid \boldsymbol{X} \text{ reaches } \boldsymbol{\ell}\right]\right] && \text{(Claim 32)} \\
&\quad= \mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim\mu^k(T)}\left[\mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}[T(\boldsymbol{X}) f^{\otimes k}(\boldsymbol{X}) \mid \boldsymbol{X} \text{ reaches } \boldsymbol{\ell}]\right] && \textstyle(\prod_{i\in[k]} f(\boldsymbol{X}^{(i)}) = f^{\oplus k}(\boldsymbol{X})) \\
&\quad= \mathop{\mathbb{E}}_{\boldsymbol{X}\sim\mu^k}[T(\boldsymbol{X}) f^{\oplus k}(\boldsymbol{X})]
\end{aligned}$$

which completes the proof. ◀

10 Proof of Claim 4

▷ Claim 40 (The γ factor in Theorem 2 is necessary; Claim 4 restated). Let $\mathsf{Par} : \{\pm 1\}^n \to \{\pm 1\}$ be the parity function and μ be the uniform distribution over $\{\pm 1\}^n$. Then for all γ,

$$\overline{\mathrm{Depth}}^{\mu^k}(\mathsf{Par}^{\otimes k}, 1-\gamma) \leq 2\gamma k \cdot \overline{\mathrm{Depth}}^{\mu}(\mathsf{Par}, \tfrac{1}{4}).$$

We will need the following simple proposition, which states that in any tree that seeks to compute the n-variable parity function, leaves of depth strictly less than n contribute $\frac{1}{2}$ to the error:

▶ **Proposition 41.** *For any (potentially randomized) tree $T : \{\pm 1\}^n \to \{\pm 1\}$ and leaf ℓ of T with depth strictly less than n,*

$$\Pr_{\boldsymbol{x}\sim\mathrm{Unif}(\{\pm 1\}^n)}[T(\boldsymbol{x}) \neq \mathsf{Par}(\boldsymbol{x}) \mid \boldsymbol{x} \text{ reaches } \ell] = \tfrac{1}{2}.$$

Proof. Since ℓ is at depth strictly less than n, there must be some index $i \in [n]$ not queried on the path to ℓ. Taking any input x that reaches ℓ, the input x' with the i^{th} bit flipped must also reach ℓ and have the opposite parity. Both of these inputs are equally likely under the uniform distribution and so the value of $\mathsf{Par}(\boldsymbol{x})$ conditioned on $\boldsymbol{x}$ reaching $\boldsymbol{\ell}$ is equally likely to be $+1$ and -1. Therefore, T errs half the time it reaches this leaf regardless of how it labels it. ◀

Proof of Claim 40. The proof proceeds in two parts. First, we show that $\overline{\mathrm{Depth}}^{\mu}(\mathsf{Par}, \frac{1}{4}) = \frac{n}{2}$. Second, we prove that $\overline{\mathrm{Depth}}^{\mu^{\otimes k}}(\mathsf{Par}^{\otimes k}, 1-\gamma) \leq \gamma kn$.

(1) $\overline{\mathrm{Depth}}^{\mu}(\mathsf{Par}, \frac{1}{4}) = \frac{n}{2}$. Let T be an arbitrary randomized decision tree let $p_n(T)$ be the probability that T queries all n variables. Then, the expected depth of T is at least $n \cdot p_n(T)$. Meanwhile, by Proposition 41, the error of T in computing parity is at least $\frac{1}{2} \cdot p_n(T)$ w.r.t. the uniform distribution. Therefore, $\overline{\mathrm{Depth}}^{\mu}(f, \frac{1}{4}) \geq \frac{n}{2}$.

While this direction is not needed for Claim 40 we show for completeness that $\overline{\mathrm{Depth}}^{\mu}(f, \frac{1}{4}) \leq \frac{n}{2}$ by constructing a randomized[1] decision tree T for f. With probability $\frac{1}{2}$, T queries all n variables to compute f exactly. Otherwise, it simply outputs 0. T has expected depth $\frac{n}{2}$, and it errs only when it queries no variables and guesses incorrectly, which happens with probability $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Thus, $\overline{\mathrm{Depth}}^{\mu}(f, \frac{1}{4}) \leq \frac{n}{2}$.

(2) $\overline{\mathrm{Depth}}^{\mu^{\otimes k}}(f^{\otimes k}, 1-\gamma) \leq \gamma k n$. We construct a randomized[2] decision tree T for $f^{\otimes k}$. With probability γ, T queries all kn variables to compute $f^{\otimes k}$ exactly, and with probability $1-\gamma$, it outputs 0. When it queries all variables, it has no error so its average error is at most $1-\gamma$. Furthermore, its average depth is $\gamma k n$. $\triangleleft$

References

1 Andris Ambainis, Loïck Magnin, Martin Roetteler, and Jérémie Roland. Symmetry-assisted adversaries for quantum state generation. In *2011 IEEE 26th Annual Conference on Computational Complexity (CCC)*, pages 167–177. IEEE, 2011.

2 Andris Ambainis, Robert Špalek, and Ronald de Wolf. A new quantum lower bound method, with applications to direct product theorems and time-space tradeoffs. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, pages 618–633, 2006.

3 Sepehr Assadi and Vishvajeet N. Graph streaming lower bounds for parameter estimation and property testing via a streaming XOR lemma. In Samir Khuller and Virginia Vassilevska Williams, editors, *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 612–625, 2021.

4 Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC)*, pages 67–76, 2010.

5 Shalev Ben-David and Robin Kothari. Randomized query complexity of sabotaged and composed functions. *Theory of Computing*, 14(5):1–27, 2018. `doi:10.4086/toc.2018.v014a005`.

6 Eric Blais and Joshua Brody. Optimal Separation and Strong Direct Sum for Randomized Query Complexity. In *34th Computational Complexity Conference (CCC)*, volume 137, pages 29:1–29:17, 2019. `doi:10.4230/LIPIcs.CCC.2019.29`.

7 Guy Blanc, Caleb Koch, Carmen Strassle, and Li-Yang Tan. A strong composition theorem for junta complexity and the boosting of property testers. In *Proceedings of the 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1757–1777, 2023.

8 Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. In *Proceedings of the 54th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 746–755, 2013.

9 Joshua Brody, Jae Tak Kim, Peem Lerdputtipongporn, and Hariharan Srinivasulu. A strong xor lemma for randomized query complexity. *Theory of Computing*, 19(11):1–14, 2023. `doi:10.4086/toc.2023.v019a011`.

10 Andrew Drucker. Improved direct product theorems for randomized query complexity. *computational complexity*, 21(2):197–244, 2012.

[1] At the cost of increasing expected depth by 1, the tree can be derandomized. To derandomize it, read a single bit of the input and only query the rest if that bit is 1, which occurs with probability $\frac{1}{2}$.

[2] Again, this can be derandomized at the cost of adding ≤ 2 to the expected depth, since any biased coin can be simulated with a fair coin using 2 flips in expectation.

11 Andrew Drucker. Nondeterministic direct product reductions and the success probability of sat solvers. In *Proceedings of the 54th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 736–745, 2013.

12 Oded Goldreich, Noam Nisan, and Avi Wigderson. On yao's xor-lemma. *Studies in Complexity and Cryptography*, 6650:273–301, 2011.

13 William Hoza. A technique for hardness amplification against AC0. *ECCC preprint TR23-176*, 2023.

14 Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proceedings of IEEE 36th Annual Foundations of Computer Science (FOCS)*, pages 538–545, 1995.

15 Russell Impagliazzo, Ragesh Jaiswal, Valentine Kabanets, and Avi Wigderson. Uniform direct product theorems: simplified, optimized, and derandomized. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC)*, pages 579–588, 2008.

16 Russell Impagliazzo, Ran Raz, and Avi Wigderson. A direct product theorem. In *Proceedings of IEEE 9th Annual Conference on Structure in Complexity Theory*, pages 88–96, 1994.

17 Russell Impagliazzo and Avi Wigderson. P = BPP if E requires exponential circuits: Derandomizing the XOR lemma. In *Proceedings of the 27th Annual ACM Symposium on Theory of Computing (STOC)*, pages 220–229, 1997.

18 Rahul Jain. New strong direct product results in communication complexity. *Journal of the ACM (JACM)*, 62(3):1–27, 2015.

19 Rahul Jain, Hartmut Klauck, and Miklos Santha. Optimal direct sum results for deterministic and randomized decision tree complexity. *Information Processing Letters*, 110(20):893–897, 2010.

20 Rahul Jain, Attila Pereszlényi, and Penghui Yao. A direct product theorem for the two-party bounded-round public-coin communication complexity. In *Proceedings of the 53rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 167–176, 2012.

21 Hartmut Klauck. A strong direct product theorem for disjointness. In *Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC)*, pages 77–86, 2010.

22 Hartmut Klauck, Robert Špalek, and Ronald de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. *SIAM Journal on Computing*, 36(5):1472–1493, 2007. Preliminary version in FOCS 2004.

23 Adam R Klivans and Rocco A Servedio. Boosting and hard-core set construction. *Machine Learning*, 51:217–238, 2003.

24 Troy Lee and Jérémie Roland. A strong direct product theorem for quantum query complexity. *computational complexity*, 22:429–462, 2013.

25 Troy Lee, Adi Shraibman, and Robert Špalek. A direct product theorem for discrepancy. In *Proceedings of the 23rd Annual IEEE Conference on Computational Complexity (CCC)*, pages 71–80, 2008.

26 Leonid A Levin. One-way functions and pseudorandom generators. In *Proceedings of the 17th Annual ACM Symposium on Theory of Computing (STOC)*, pages 363–365, 1985.

27 Noam Nisan, Steven Rudich, and Michael Saks. Products and help bits in decision trees. In *Proceedings 35th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 318–329, 1994.

28 Noam Nisan and Avi Wigderson. Hardness vs randomness. *Journal of computer and System Sciences*, 49(2):149–167, 1994.

29 Ryan O'Donnell. Hardness amplification within np. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC)*, pages 751–760, 2002.

30 Ronen Shaltiel. Towards proving strong direct product theorems. *Computational Complexity*, 12(1/2):1–22, 2004.

31 Alexander A Sherstov. Strong direct product theorems for quantum communication and query complexity. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 41–50, 2011.

32 Robert Špalek. The multiplicative quantum adversary. In *23rd Annual IEEE Conference on Computational Complexity (CCC)*, pages 237–248, 2008.
33 Volker Strassen. Vermeidung von divisionen. *Journal für die reine und angewandte Mathematik*, 264:184–202, 1973.
34 Luca Trevisan. The Impagliazzo Hard-Core-Set Theorem. `https://lucatrevisan.wordpress.com/2007/11/06/the-impagliazzo-hard-core-set-theorem/`, 2007.
35 Emanuele Viola and Avi Wigderson. Norms, xor lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4(1):137–168, 2008.
36 Andrew Yao. Theory and application of trapdoor functions. In *Proceedings of the 23rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 80–91, 1982.
37 Huacheng Yu. Strong xor lemma for communication with bounded rounds. In *Proceedings of the 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1186–1192, 2022.

A Figures of stacked and fair trees

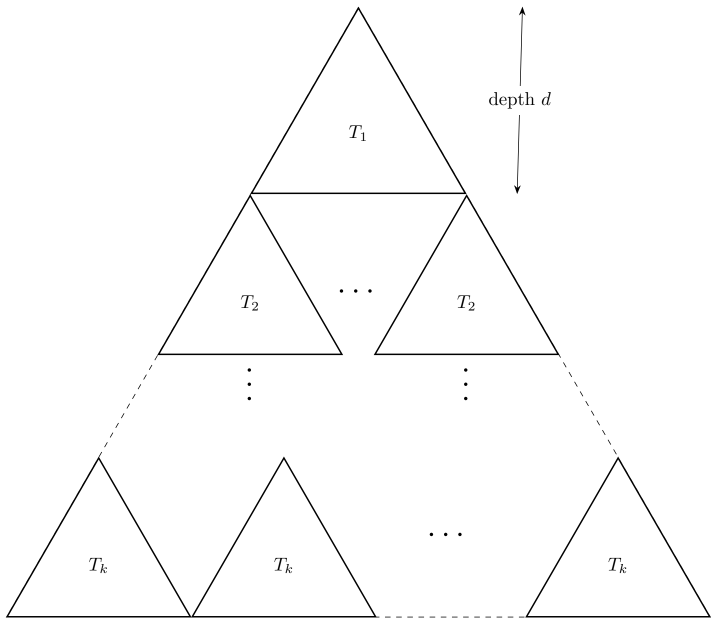


Figure 3 Illustration of a stacked decision tree for a function $f^{\otimes k}$. The decision tree consists of k depth-d decision trees, $T_1, \dots, T_k$, stacked on top of each other. For an input $X \in (\{\pm 1\}^n)^k$, the output $T(X)$ is computed sequentially, first by computing $T_1(X)$, then $T_2(X)$, and so on. The final output is $T(X) \coloneqq (T_1(X), \dots, T_k(X))$.

B Proof of Theorem 6

Let $\mathcal{H}$ denote the set of measures of density $\delta/2$ with respect to μ and let $\mathcal{T}$ denote the set of decision trees T whose expected depth with respect to μ is at most d. Suppose for contradiction that there *does not* exist an $H \in \mathcal{H}$ which is (γ, d)-hardcore. That is, for all $H \in \mathcal{H}$ there is a tree T of expected depth at most d satisfying

$$\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})T(\boldsymbol{x})H(\boldsymbol{x})] > \gamma \mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}[H(\boldsymbol{x})] = \gamma\delta/2. \tag{14}$$

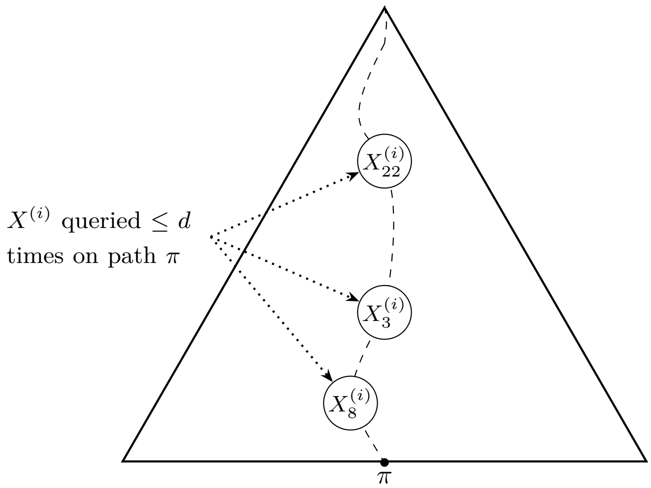


Figure 4 Illustration of a fair decision tree. For every block $i \in [k]$ and path π, the input block $X^{(i)}$ is queried at most d times.

We use the minimax theorem to switch the quantifiers in the above statement. Consider the payoff matrix M whose rows are indexed by distributions from $\mathcal{H}$ and whose columns are indexed by algorithms from $\mathcal{T}$ and whose entries are given by $M_{H,T} \coloneqq \mathbb{E}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})T(\boldsymbol{x})H(\boldsymbol{x})]$. This is the payoff matrix for the zero-sum game where the row player first chooses a row H and the column player then chooses a column T and the payoffs are determined by $M_{H,T}$. Note that once the first player's strategy is fixed, we can assume without loss of generality that the second player's strategy is deterministic. Therefore, the minimax theorem for zero-sum games yields

$$\begin{aligned} \gamma\delta/2 &< \min_{\rho\in\mu(\mathcal{H})} \max_{T\in\mathcal{T}} (\rho^\top M)_T && \text{(Equation (14))} \\ &= \max_{\tau\in\mu(\mathcal{T})} \min_{H\in\mathcal{H}} (M\tau)_H && \text{(minimax theorem)} \end{aligned}$$

where $\mu(\cdot)$ denotes the set of distributions over a given set. Therefore, there is a fixed distribution τ over the set $\mathcal{T}$ such that for all $H \in \mathcal{H}$

$$\mathop{\mathbb{E}}_{\boldsymbol{T}\sim\tau}\left[\mathop{\mathbb{E}}_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x})\boldsymbol{T}(\boldsymbol{x})H(\boldsymbol{x})]\right] > \gamma\delta/2. \tag{15}$$

This shows that

$$\Pr_{\boldsymbol{x}\sim\mu}\left[\mathop{\mathbb{E}}_{\boldsymbol{T}\sim\tau}[\boldsymbol{T}(\boldsymbol{x})]f(\boldsymbol{x}) \geq \gamma\right] \geq 1-\delta/2.$$

In particular, if instead $\Pr_{\boldsymbol{x}\sim\mu}[\mathbb{E}_{\boldsymbol{T}\sim\tau}[\boldsymbol{T}(\boldsymbol{x})]f(\boldsymbol{x}) < \gamma] \geq \delta/2$ then we can contradict Equation (15) by constructing a $\delta/2$-density H such that $H(x) \coloneqq \Pr_{\boldsymbol{x}\sim\mu}[\boldsymbol{x} = x]$ for a $\delta/2$-fraction of x satisfying $\mathbb{E}_{\boldsymbol{T}\sim\tau}[\boldsymbol{T}(\boldsymbol{x})]f(\boldsymbol{x}) < \gamma$. Equation (15) shows that $\mathbb{E}_{\boldsymbol{T}\sim\tau}[\boldsymbol{T}(\boldsymbol{x})]$ has good correlation with f for a large fraction of inputs. We obtain a single strategy from the distribution τ by sampling $\boldsymbol{T}_1, \dots, \boldsymbol{T}_r \sim \tau$ for r sufficiently large (chosen later) and defining $T^\star$ as $\boldsymbol{T}^\star(x) \coloneqq \mathrm{MAJ}(\boldsymbol{T}_1(x), \dots, \boldsymbol{T}_r(x))$. For every x for which $\mathbb{E}_{\boldsymbol{T}\sim\tau}[\boldsymbol{T}(x)]f(x) \geq \gamma$, we have

$$\Pr_{\boldsymbol{T}_1,\dots,\boldsymbol{T}_r\sim\tau}\left[\mathrm{MAJ}(\boldsymbol{T}_1(x), \dots, \boldsymbol{T}_r(x)) \neq f(x)\right] \leq 2^{-\Omega(\gamma^2 r)}$$

by a Chernoff bound. Choosing $r = \Theta(\log(1/\delta)/\gamma^2)$ ensures that the failure probability is at most $\delta/2$. The decision tree $\boldsymbol{T}^\star$ satisfies $\Pr_{\boldsymbol{x}\sim\mu}[\boldsymbol{T}^\star(\boldsymbol{x}) \neq f(\boldsymbol{x})] \leq \delta$. The expected depth of $\boldsymbol{T}^\star$ is less than

$$r \cdot d = \Theta(d \log(1/\delta)/\gamma^2) < \overline{\mathrm{Depth}}^\mu(f, \delta)$$

which is a contradiction.

C The lack of error reduction for distributional error

In Section 2, we showed how error reduction gave a simple proof of a strong direct sum theorem for *randomized* query complexity. The specific statement needed in that proof is the following standard error reduction by repetition theorem.

▶ **Fact 3** (Error reduction for $\overline{\mathrm{R}}$). *For any function $f : \{\pm 1\}^n \to \{\pm 1\}$ and $\delta > 0$,*

$$\overline{\mathrm{R}}(f, \delta) \leq O(\log(\tfrac{1}{\delta})) \cdot \overline{\mathrm{R}}(f, 1/4).$$

Here, we give a short proof that no error reduction holds in the distributional setting, even with substantially weaker parameters.

▷ Claim 42. For any $n \in \mathbb{N}$, let μ be the uniform distribution over $\{\pm 1\}^n$. There is a function $f : \{\pm 1\}^n \to \{\pm 1\}$ satisfying,

$$\overline{\mathrm{Depth}}^\mu(f, 1/4) = 0 \qquad \text{and} \qquad \overline{\mathrm{Depth}}^\mu(f, 1/8) \geq \Omega(n).$$

Since any function on n bits can be computed exactly using n, the f in the above claim requires essentially the maximum number of queries to be computed to error $1/8$ despite requiring no queries to be computed to error $1/4$.

Proof. We first define f: If $x_1 = 0$, then $f(x) = 0$. Otherwise, $f(x)$ is the parity of the remaining $n-1$ bits of x.

Note that,

$$\Pr_{\boldsymbol{x}\sim\mu}[f(\boldsymbol{x}) = 0] = \frac{1}{2} \cdot (\Pr[f(\boldsymbol{x}) = 0 \mid \boldsymbol{x}_1 = 0] + \Pr[f(\boldsymbol{x}) = 0 \mid \boldsymbol{x}_1 = 1]) = \frac{3}{4}.$$

Therefore, the 0 query algorithm that simply outputs 0 has an error of only $1/4$ on f.

It only remains to prove that $\overline{\mathrm{Depth}}^\mu(f, 1/10) \geq \Omega(n)$. Consider any (potentially randomized) $T : \{\pm 1\}^n \to \{\pm 1\}$ and leaf ℓ of T at depth strictly less than $n-1$. By Proposition 41

$$\Pr_{\boldsymbol{x}\sim\mu}[T(\boldsymbol{x}) \neq f(\boldsymbol{x}) \mid \boldsymbol{x} \text{ reaches } \ell, \boldsymbol{x}_1 = 1] = 1/2.$$

Let p be the probability that $T(\boldsymbol{x})$ queries a leaf of depth at least $n-1$ given that $\boldsymbol{x}_1 = 1$. The above allows us to conclude that

$$\Pr_{\boldsymbol{x}\sim\mu}[T(\boldsymbol{x}) \neq f(\boldsymbol{x})] \geq \frac{1}{2} \cdot \Pr[T(\boldsymbol{x}) \neq f(\boldsymbol{x}) \mid \boldsymbol{x}_1 = 1] \geq \frac{1}{4} \cdot p.$$

Therefore, if T has error at most $1/8$, then p must be at least $1/2$, which shows that $\overline{\mathrm{Depth}}^\mu(f, 1/8) \geq \frac{n-1}{4}$. ◁