Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 87 (2022)

Depth-d Threshold Circuits vs. Depth-$(d+1)$ AND-OR Trees

Pooya Hatami * William M. Hoza † Avishay Tal ‡ Roei Tell §

November 22, 2022

Abstract

For any $n \in \mathbb{N}$ and $d = o(\log \log n)$, we prove that there is a Boolean function F on n bits and a value $\gamma = 2^{-\Theta(d)}$ such that F can be computed by a uniform depth-$(d+1)$ AC^0 circuit with $O(n)$ wires, but F cannot be computed by any depth-d TC^0 circuit with $n^{1+\gamma}$ wires. This bound matches the current state-of-the-art lower bounds for computing explicit functions by threshold circuits of depth $d > 2$, which were previously known only for functions outside AC^0 such as the parity function. Furthermore, in our result, the AC^0 circuit computing F is a monotone *read-once formula* (i.e., an AND-OR tree), and the lower bound holds even in the average-case setting with respect to advantage $n^{-\gamma}$.

At a high level, our proof strategy combines two prominent approaches in circuit complexity from the last decade: The celebrated *random projections* method of Håstad, Rossman, Servedio, and Tan (J. ACM 2017), which was previously used to show a tight average-case depth hierarchy for AC^0; and the line of works analyzing the effect of *random restrictions* on threshold circuits. We show that under a modified version of Håstad, Rossman, Servedio, and Tan's projection procedure, any depth-d threshold circuit with $n^{1+\gamma}$ wires simplifies to a near-trivial function, whereas an appropriately parameterized AND-OR tree of depth $d+1$ maintains structure.

*Department of Computer Science and Engineering, The Ohio State University, OH, USA. Email: pooyahat@gmail.com. Supported by NSF grant CCF-1947546.

†Simons Institute for the Theory of Computing, University of California at Berkeley, CA, USA. Email: williamhoza@berkeley.edu. Part of this work was done while the author was visiting the Simons Institute for the Theory of Computing. Part of this work was done while the author was a graduate student at the University of Texas at Austin, supported by the NSF GRFP under Grant DGE-1610403 and by a Harrington Fellowship from UT Austin.

‡Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, CA, USA. Email: atal@berkeley.edu. Supported by NSF CAREER award CCF-2145474 and by a Sloan Fellowship.

§The Institute for Advanced Study at Princeton NJ and the DIMACS Center at Rutgers University, NJ. Email: roeitell@gmail.com. Part of this work was supported by the National Science Foundation under grant number CCF-1445755 and under grant number CCF-1900460.

ISSN 1433-8092

Contents

1 Introduction

The focus of this paper is *linear threshold circuits* (LTF circuits). These are non-uniform circuits in which each gate can compute an arbitrary *linear threshold function* (LTF), of the form

$$\Phi_{w,\theta}(x_1,\ldots,x_n) = 1 \iff \sum_{i\in[n]} w_i \cdot x_i \geq \theta\,,$$

where $w \in \mathbb{R}^n$ and $\theta \in \mathbb{R}$ and the arithmetic is over the reals. We define the *size* of a circuit to be its number of wires.

Proving explicit lower bounds for LTF circuits is one of the most important current challenges in complexity theory. However, despite more than 50 years of research into this circuit class, the best lower bounds known are only for circuits of slightly super-linear size. Specifically, in the 1990s, Impagliazzo, Paturi, and Saks [IPS97] showed that LTF circuits of depth d and size $n^{1+\gamma_d}$ (where $\gamma_d = 2^{-\Theta(d)}$) cannot compute the parity function. This was recently strengthened by Chen, Santhanam, and Srinivasan [CSS18] to an average-case lower bound for circuits of the same size (up to the constant inside the Θ-notation in the expression for γ_d) computing the Generalized Andreev function. The latter work is part of an influential line of works in the last decade, which introduced new ways of analyzing LTFs and LTF circuits (see, e.g., [Ser07; DGJ+10; CSS18; Tel18; HHT+22]).[1]

Our main result in this paper is a stronger lower bound, where the improvement is that the bound holds for a function that is "even more explicit" than parity (in the sense that it has lower computational complexity). Specifically, we show that LTF circuits fail to compute a simple, read-once, AND-OR tree. In particular, such trees are monotone, read-once AC^0 formulas – arguably one of the weakest complexity classes that has been studied. Moreover, our main result asserts that LTF circuits of depth d and size $n^{1+2^{-\Theta(d)}}$ cannot even compute a read-once AND-OR tree *of depth* $d+1$; that is, the depth difference amounts to a single layer.[2] And lastly, our lower bound also holds on average, rather than just in the worst case.

Theorem 1.1 (LTF circuits cannot compute simple AND-OR trees)**.** *Let $n \in \mathbb{N}$ be sufficiently large, let $d \leq \frac{1}{20}\cdot \mathrm{loglog}(n)$, and let $\gamma_d = 2^{-10\cdot d}$. Then, there exists an explicit depth-$(d+1)$ read-once AC^0 formula $F = F^{(n)}_{d+1}$ on n input bits such that for every depth-d LTF circuit f with at most $n^{1+\gamma_d}$ wires,*

$$\Pr_{\mathbf{x}\in\{0,1\}^n}[f(\mathbf{x}) = F(\mathbf{x})] \leq \frac{1}{2} + n^{-\gamma_d}.$$

The key contribution underlying Theorem 1.1 is a new and more refined way of analyzing LTF circuits, which paves the way to proving our new lower bound, and which we hope may facilitate further progress in proving lower bounds for LTF circuits. At a high level, our proof strategy combines two main approaches in circuit complexity from the last decade that were separate so far: The celebrated *random projections* method of Håstad, Rossman, Servedio, and Tan [HRS+17], which was previously used to show a tight depth hierarchy for AC^0; and the line of works (mentioned above) that introduced new ways of analyzing LTF circuits, and that in particular analyzed the effect of *radom restrictions* on LTF circuits.[3]

[1] Complementing these two lower bounds, Chen and Tell [CT19] showed that to prove lower bounds against LTF circuits of polynomial size, it suffices to prove lower bounds for LTF circuits of depth d size $n^{1+\delta_d}$, where $\delta_d = 2^{-\Theta(d)}$ differs from γ_d only in the constant hidden inside the Θ-notation.

[2] Needless to say, LTF circuits of depth d and super-linear size (or even linear size) can compute read-once AND-OR trees of depth d or less, and thus to get a lower bound it is necessary for the tree to have at least one additional layer. Our result shows that one additional layer is also sufficient.

[3] See Sections 2 and 3.2 for definitions of random restrictions and random projections.

Combining the two lines of work requires significant technical effort, in order to make them "fit together". From a bird's eye view, our paper analyzes the effect of a very specific (and suitably chosen) random projections procedure on LTF circuits. Replicating the analysis from Håstad, Rossman, Servedio, and Tan's work [HRS+17] with a different and careful parameterization, we show that the AND-OR tree F "maintains structure" under this projections procedure; and the crux of our technical contribution is in showing that this projections procedure trivializes every LTF circuit of depth d and size $n^{1+\gamma_d}$, with high probability. We refer the reader to Section 2 for a technical overview of the proof.

1.1 Impossibility of depth-reduction using LTF gates

Theorem 1.1 can also be viewed in the context of *circuit depth reduction*, which is the task of decreasing the depth of a circuit without significantly increasing its size (and without changing the function that it computes). There are classic, strong *lower bounds* regarding depth reduction of AC^0 circuits. Indeed, for a certain AC^0 circuit of depth $d+1$ and size $O(n)$, Håstad showed that every equivalent depth-d AC^0 circuit has size at least $2^{n^{\Omega(1/d)}}$ [Hås87, Chapter 6], improving earlier work by Sipser [Sip83b] and Yao [Yao85]. However, the situation changes if we allow the shallower circuit to use a stronger model. In this case, strong depth reduction *upper bounds* are known for AC^0 circuits and, much more generally, for ACC^0 circuits (i.e., AC^0 circuits augmented with MOD_m gates where m is constant). In particular, building on and improving a long line of work [Tod91; All89; AH94; Yao90; AG93; BT94; Wil14a], Chen and Papakonstantinou showed that for every ACC^0 circuit of depth d and size w, there exists an equivalent $\mathsf{SYM} \circ \mathsf{AND}$ circuit of size $2^{(\log w)^{O(d)}}$ [CP19]. Depth reduction theorems along these lines have found applications in circuit analysis algorithms [Wil14b; ACW16; Wil18b; CP19], circuit lower bounds [Wil14b; Wil16; Wil18b; COS18; CP19; MW20; VW20; Che19; CR22], and even graph algorithms [Wil18a].

If we insist on the shallower circuit having the specific form $\mathsf{SYM} \circ \mathsf{AND}$, then near-matching lower bounds are known. Indeed, there are constructions of constant-depth polynomial size ACC^0 circuits [RW93], and even AC^0 circuits [BH12], for which every equivalent $\mathsf{MAJ} \circ \mathsf{SYM} \circ \mathsf{AND}$ circuit must have super-polynomial size. But what happens if we allow the shallower circuit to use an even stronger model? Observe that a $\mathsf{SYM} \circ \mathsf{AND}$ circuit can be converted into an equivalent depth-three LTF circuit with only a polynomial increase in size, because every symmetric function can be computed by a depth-two polynomial-size LTF circuit. Thus, a special case of Chen and Papakonstantinou's result [CP19] is that for every constant-depth polynomial-size AC^0 circuit, there is an equivalent depth-three quasipolynomial-size LTF circuit. Indeed, this special case was proven already by Allender in the 1980s [All89], who presented the theorem in basically this form. This raises the following question: If we start with an AC^0 circuit – the weakest model that we have discussed – and we wish to convert it to an equivalent LTF circuit – the strongest model that we have discussed – then is depth reduction possible without a significant size blowup?

Theorem 1.1 gives a strong negative answer to the foregoing question, showing that depth-reduction of AC^0 circuits to LTF circuits, even one that attempts to save only a single layer, is impossible without a super-linear increase in size. Thus, although one can achieve a massive depth reduction using LTF circuits with $2^{\mathrm{polylog}(n)}$ wires [Tod91; All89; AH94; Yao90; AG93; BT94; Wil14a; CP19], our lower bound asserts that using only $n^{1+c^{-d}}$ wires does not allow for any depth reduction at all. (In fact, as explained below in Section 1.3, we further prove that the complexity of such depth-reduction for our particular AND-OR tree F is *either* precisely super-linear, *or* it is super-polynomial.) Our theorem can thus be interpreted as saying that some functions have an

intrinsic "depth complexity" that is robust to changes in the gate set (i.e., from AND/OR gates to LTF gates), at least in the near-linear size regime.

1.2 Hard functions in extremely weak complexity classes

Another lens through which to view Theorem 1.1 is via the recent success in proving "super-explicit" lower bounds for circuit classes. Recall that in classical lower bounds (e.g., in [Hås87; Raz87; Smo87; Hås98; IPS97] and in many other works) the hard function is typically computable in NC^1; the two most well-known examples are the parity function and Andreev's function. A long-standing question, dating back to [Sip83a; Yao85; Hås87], is whether one can prove lower bounds for functions that are "even more explicit", such as the AND-OR tree in Theorem 1.1.

Our work follows in the footsteps of several influential works in the last decade, which were able to prove lower bounds in which the hard function is computable in uniform AC^0.[4] Among these works is the celebrated average-case depth-hierarchy theorem for AC^0 by Håstad, Rossman, Servedio, and Tan [HRS+17] mentioned above (which improved several earlier works [Sip83a; Yao85; Hås87; OW07; Vio14]); results asserting that $\mathsf{AC}^0[\oplus]$ circuits of depth d and size s fail to compute a function computable by uniform AC^0 formulas of depth $d+1$ and linear size,[5] and a function computable by uniform AC^0 formulas of depth d and size $\mathrm{poly}(s)$ [LSS+21; LST19]; and a recent work by Filmus, Meir, and Tal [FMT21], who showed a function in uniform AC^0 that cannot be computed by De Morgan formulas of sub-cubic size $n^{3-o(1)}$.

The works above "cover" the most widely-studied classes in circuit complexity, the main exception being LTF circuits (i.e., the class TC^0), which are the focus of the current work. We stress that many prior works have shown that AC^0 is hard for various *subclasses* of LTF circuits, which have particular structural restrictions (such as $\mathsf{LTF} \circ \mathsf{PARITY}$ circuits or $\mathsf{MAJ} \circ \mathsf{LTF}$ circuits or monotone circuits; see [MP69; BS92; Yao89; HG91; KP97; KP98; FKL+01; Cha07; BVW07; She09; OS10; RS10; She11; BH12; BT15; BT16; She18a; She18b; BT21; SW21]).[6] However, Theorem 1.1 is the first result showing that AC^0 is hard for LTF circuits of any constant depth (and super-linear size), and without any particular structural restrictions.

As mentioned above, our techniques use Håstad, Rossman, Servedio, and Tan's work on AC^0 [HRS+17] as a starting point (and the crux of our technical contribution is in analyzing the effect of a procedure similar to theirs on LTF circuits). However, our techniques are completely different than the techniques used in the works [OW07; Ama09; LSS+21; LST19] on $\mathsf{AC}^0[\oplus]$ (the latter works build on the line of research on the "coin problem" [SV10; BV10; Aar10; Ste13; CGR14; LV18; GII+19; LST19; Agr20; BGW20; LSS+21; BGZ22]), and also completely different than the techniques in [FMT21] on De Morgan formulas.

1.3 Tightness of our result

When d is constant, the correlation bound $n^{-\gamma}$ in Theorem 1.1 cannot be significantly improved. The reason is that our hard function is so computationally simple that it can be approximated, to a reasonable extent (i.e., almost matching Theorem 1.1), by shallower circuits:

- Every depth-$(d+1)$ AC^0 circuit with $O(n)$ wires and top fan-in m can be approximated, with success probability $1/2 + \Omega(1/m)$, by a depth-d AC^0 circuit with $O(n)$ wires. (This

[4] Of course, to show that uniform AC^0 is hard for circuits from a certain class C, one needs to give the AC^0 circuit some advantage over C, such as more depth or size (as in Theorem 1.1, in which the AND-OR tree has depth $d+1$).

[5] As explained by Limaye, Sreenivasaiah, Srinivasan, Tripathi, and Venkitesh [LSS+21], this follows from works of O'Donnell and Wimmer [OW07] and Amano [Ama09].

[6] In fact, the works mentioned here managed to prove *super-polynomial* lower bounds for these subclasses.

follows from the discriminator lemma [HMP+93].)

- Every monotone function can be approximated, with success probability $1/2 + \Omega((\log n)/n)$, by a constant or a variable. (This follows from the Kahn-Kalai-Linial theorem [KKL88].)
- Every AC^0 circuit (of any constant depth and any polynomial size) can be approximated, with success probability $1/2 + 2^{-\mathrm{polylog}(n)}$, by a depth-1 AC^0 circuit with $\mathrm{polylog}(n)$ wires, i.e., either a conjunction of $\mathrm{polylog}(n)$ literals or else a disjunction of $\mathrm{polylog}(n)$ literals. (This follows from the Linial-Mansour-Nisan theorem [LMN93].)

For completeness, we include proofs of the three preceding bounds in Appendix A. It is an interesting problem to close the remaining quantitative gaps between our correlation bound and the three preceding bounds, especially the last one.

In contrast, if we focus on exact (worst-case) simulations, then it is unclear whether the *size bound* $n^{1+\gamma}$ in our theorem is tight. We partially address this question by proving a "hardness magnification" result. Recall that such results assert that specific, seemingly-minor improvements to known lower bounds would imply dramatic, strong lower bounds for powerful models of computation. (An intensive recent interest in such results was sparked by the work of Oliveira and Santhanam [OS18], who coined the term, following older results such as those by Srinivasan [Sri03] and Allender and Koucký [AK10].) Regarding TC^0, prior work shows that tiny improvements in the known lower bounds for certain NC^1-complete functions or for MCSP would yield lower bounds for LTF circuits of arbitrarily large polynomial size [AK10; CT19; CJW20; HHT+22]. In the same spirit, we show that even a very small improvement to the size bound in Theorem 1.1 would imply that our AND-OR tree is hard for LTF circuits of arbitrarily large polynomial size. (Recall that $F^{(n)}_{d+1}$ denotes the depth-$(d+1)$ AND-OR tree from Theorem 1.1.)

Theorem 1.2 (hardness magnification for our construction)**.** *Let $d_0 \in \mathbb{N}$ and $k > 1$ be constants. Suppose that for all sufficiently large n, the function $F^{(n)}_{d_0+1}$ can be computed by depth-d_0 LTF circuits with n^k wires. Then for all sufficiently large n and all $2d_0 \le d \le \frac{1}{20}\log\log n$, the function $F^{(n)}_{d+1}$ can be computed by depth-d LTF circuits with $\widetilde{O}(n^{1+k\cdot 10^{-d}})$ wires.*

In particular, suppose that for every sufficiently large constant d and sufficiently large n, the function $F^{(n)}_{d+1}$ cannot be computed by depth-d LTF circuits with $n^{1+2^{-3\cdot d}}$ wires (slightly improving the $n^{1+2^{-10\cdot d}}$ bound from Theorem 1.1). Under that assumption, Theorem 1.2 implies that $F^{(n)}_{d+1}$ cannot even be computed by depth-d LTF circuits with any $\mathrm{poly}(n)$ number of wires. The proof of Theorem 1.2 is simple and relies on the recursive structure of $F^{(n)}_{d+1}$ (see Section 9).

The optimal size complexity of depth-d LTF circuits computing our depth-$(d+1)$ AND-OR tree $F^{(n)}_{d+1}$ remains unclear. However, taken together, our results and prior depth-reduction theorems [Tod91; All89; AH94; Yao90; AG93; BT94; Wil14a; CP19] essentially narrow down the optimal size complexity to *two* relatively small intervals. Either the optimal size complexity is $n^{1+2^{-\Theta(d)}}$ (for all sufficiently large n and d with $d \le \frac{1}{20}\log\log n$), or else the optimal size complexity is between $n^{\omega(1)}$ and $n^{\mathrm{polylog}(n)}$ (for each constant $d \ge 4$ and infinitely many n).

2 Technical overview

We prove our result using the method of *random projections*, which is a generalization of the traditional method of random restrictions that (to the best of our knowledge) was first used by

Impagliazzo and Segerlind [IS01]. A projection maps each variable either to a constant (0 or 1) or else to another variable. The key feature distinguishing projections from traditional restrictions is that a projection might "merge" multiple variables by mapping them to a single variable, thereby keeping the variables alive but "tying them together."

For this technical overview, let us focus on the problem of proving a *worst-case* separation between depth-d LTF circuits and depth-$(d+1)$ AND-OR trees, and let us focus on the case that d is constant. Such a separation follows from the following theorem.

Theorem 2.1 (see Theorems 5.1 and 6.1)**.** *Let $d \in \mathbb{N}$ be a constant, let $\gamma = 2^{-10 \cdot d}$, and let $n \in \mathbb{N}$ be sufficiently large. There exists an explicit depth-$(d+1)$ read-once AC^0 formula $F = F_{d+1}^{(n)}$ on n input bits, a random projection $\boldsymbol{\pi}$, and a distribution $\boldsymbol{\sigma}$ such that:*

1. *(Survival of the AND-OR tree) With probability $1 - o(1)$ over $\pi \sim \boldsymbol{\pi}$, the projected function $F\restriction_\pi$ is $o(1)$-unbiased under $\boldsymbol{\sigma}$, i.e.,*
$$\left| \Pr_{\boldsymbol{\sigma}} [F\restriction_\pi(\boldsymbol{\sigma}) = 1] - \frac{1}{2} \right| = o(1).$$
2. *(Simplification of any LTF circuit) For any depth-d LTF circuit f on n input bits with at most $n^{1+\gamma}$ wires, with probability $1 - o(1)$ over $\pi \sim \boldsymbol{\pi}$, the projected function $f\restriction_\pi$ is $o(1)$-close to a constant under $\boldsymbol{\sigma}$, i.e., there is some $b \in \{0,1\}$ such that*
$$\Pr_{\boldsymbol{\sigma}} [f\restriction_\pi(\boldsymbol{\sigma}) = b] = 1 - o(1).$$

The distribution $\boldsymbol{\sigma}$ is simply an i.i.d. product distribution (with biased marginals). Furthermore, projecting according to $\boldsymbol{\pi}$ and then assigning values according to $\boldsymbol{\sigma}$ yields the uniform distribution over $\{0,1\}^n$, which is why we actually get an *average-case* separation.

Both our hard function F and our projection $\boldsymbol{\pi}$ are based on the work of Håstad, Rossman, Servedio, and Tan [HRS+17]. We do modify the parameters, but still, the fact that the hard function F survives the projection (Item 1 above) follows from a fairly straightforward generalization of their analysis. The main challenge is showing that LTF circuits simplify under the specific random projection procedure $\boldsymbol{\pi}$ (i.e., proving Item 2).

Random projections and LTF circuits. The last couple of decades have seen the development of what is often referred to as the *structural theory of linear threshold functions*, which can be viewed as a special case of the "structure vs. randomness" paradigm. One of the main applications of this body of knowledge is the analysis of the effect of *random and pseudorandom restrictions* on LTF circuits of depth d and size at most $n^{1+2^{-O(d)}}$ [Ser07; DGJ+10; CSS18; Tel18; HHT+22].

The main technical contribution of our work is showing that this body of knowledge can be extended so that it works in an inherently different setting, namely the setting of random projections as discussed above. The projections that we analyze are quite different than traditional random restrictions: Not only are variables sometimes merged, but also the assigned values are heavily biased, and the values assigned to them have significant correlations. Indeed, these projections cannot even be considered "pseudorandom," but we nevertheless show that LTF circuits of depth d simplify under these projections. The underlying technical challenges require extending and refining techniques used in previous works in the last decade.

In Section 2.1, we describe the hard function F, the projection $\boldsymbol{\pi}$, and the distribution $\boldsymbol{\sigma}$ in more detail (the full definitions are given in Section 4), and we briefly explain why F survives the projection (following the analysis in [HRS+17]). Then, in Sections 2.2–2.4, we discuss the main part of the proof, which shows that LTF circuits simplify under $\boldsymbol{\pi}$.

2.1 Setup and high-level plan

Let us describe our random projection procedure, which is a very careful modification of the one in [HRS+17]. One of our contributions is to devise an alternative way of thinking about the projection procedure that hides some complexity, enabling us to analyze its effect on LTF circuits. This abstraction might be useful in other contexts as well. While the new perspective is enough to carry out most of our analysis of LTF circuits, we do still rely on the original perspective for some parts of this analysis. (Readers who are familiar with the work of Håstad *et al.* [HRS+17] are encouraged to refer to Section 6.1, in which we prove the formal connection between the two perspectives.) We now give an overview of the procedure from the new perspective.

The random projections. We denote by $\mathbf{b}_{1-\beta}$ a Bernoulli RV that gets value 1 with probability $1-\beta$. The projection procedure works in d iterations. For the first iteration, consider the gates of F just above the inputs, denoted $g_1, \ldots, g_t$, which are all AND gates. We partition the variables into disjoint blocks $B_1, \ldots, B_t$, where B_j consists of the variables that feed into the gate g_j. Then, in each block B_j independently, for suitable parameters $p_1, \beta_1 > 0$:

1. With probability p_1 the block *survives*, in which case a random subset of its variables of density $\approx \beta_1$ is kept alive,[7] and all the other variables in the block are assigned the value 1 (recall that the g_j is an AND gate).

2. If the block does not survive, the variables are assigned values from $\mathbf{b}_{1-\beta_1}$ independently, *except* that the probability that all variables are assigned 1 is artificially decreased.

Note that the expected fraction of living variables in such an iteration is approximately $\overline{p}_1 = p_1 \cdot \beta_1$. In the end of the iteration, we merge the living variables in each surviving block; that is, we project these variables to a single new variable. We consider this new variable to be a "level-1 variable" whereas the original input variables are level-0 variables. We denote the end-result of this assignment and merging (i.e., projection) process by $\boldsymbol{\pi}^{(1)}$, and we refer to any projection with the above structure as a *corrupted biased block projection*. (See Definitions 6.2 and 6.3.)

After applying $\boldsymbol{\pi}^{(1)}$, we can identify each AND gate g_j either with a constant (in case a variable in B_j was assigned 0, or all variables in B_j were assigned 1), or with the living variable corresponding to the "merged" variables in B_j (in case some variables in B_j were left alive, and all the others were assigned 1). We thus recurse into the next iteration with a circuit of smaller depth on the level-1 variables. However, in our setting (and unlike [HRS+17]), subsequent iterations of the projection, denoted $\boldsymbol{\pi}^{(2)}, \ldots, \boldsymbol{\pi}^{(d)}$, will be parameterized by *different values* of $p_2, \ldots, p_d$ and $\beta_2, \ldots, \beta_d$. The projection $\boldsymbol{\pi}^{(i)}$ maps level-$(i-1)$ variables to level-i variables (or constants). The projection $\boldsymbol{\pi}$ referred to in Theorem 2.1 applies $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$ successively, thereby mapping level-0 variables to level-d variables or constants.

Loosely speaking, our goal is to prove that each iteration of the procedure above reduces the depth of F by exactly one layer, and simultaneously reduces the depth of any LTF circuit by at least one layer; thus, after d iterations, F maintains structure whereas any LTF circuit of depth d trivializes. As articulated in Theorem 2.1, our notions of "maintaining structure" and "trivialization" are defined with respect to a suitable distribution $\boldsymbol{\sigma}$. Similarly, at each intermediate stage, we carry out our analysis with respect to a corresponding intermediate distribution $\boldsymbol{\sigma}^{(i)}$ over the level-$(i-1)$ variables. This distribution is also an i.i.d. product distribution with suitably biased marginals (the bias differs from one i to the next).

[7] Specifically, we include each element with probability β_1 and condition on getting a nonempty set.

Specifying the parameters. Let us now motivate our choice of parameters, informed by prior work on LTF circuits. In [HRS+17], the fraction $\overline{p}_i$ of living variables is essentially the same from one iteration to the next: $\overline{p}_1 \approx \overline{p}_2 \approx \cdots \approx \overline{p}_d$. In contrast, previous restriction procedures for LTF circuits apply d rounds of restrictions where the fraction of living variables decreases from one iteration to the next: $\overline{p}_{i+1} = (\overline{p}_i)^C$ for a large constant $C > 1$. We follow the latter line of works, and adapt the AND-OR tree and the random projections above to use corresponding parameters. Accounting for the required changes to maintain the properties above, the resulting AND-OR tree is such that the fan-ins of gates in the tree grow rapidly as we go up the layers.

Being more specific, recall that we are assuming that the depth $d+1$ is constant for simplicity. For a parameter M (where $M \approx n^\epsilon$ for a small constant $\epsilon > 0$), we define a sequence of parameters $M_1 = M$ and $M_{i+1} = M_i^{100}$ for $i = 1, \ldots, d-1$. We choose the fan-ins in the AND-OR tree such that under a uniform random input, for each gate g at distance $i \le d$ from the inputs, the subformula rooted at g has acceptance probability roughly $1/M_i$ if i is even or $1 - 1/M_i$ if i is odd, and overall, the AND-OR tree has acceptance probability roughly $1/2$. In more detail, we set

$$
\begin{aligned}
f_i &= \text{fan-ins at distance } i \text{ from inputs} && \approx \begin{cases} \log(M_1) & i = 1 \\ M_{i-1} \cdot \ln(M_i) & 2 \le i \le d \\ M_d \cdot \ln(2) & i = d+1 \end{cases} \\
p_i &= \text{probability that a block survives under } \boldsymbol{\pi}^{(i)} && \approx 1/\sqrt{M_i} \\
\beta_i &= \begin{array}{l}\text{bias parameter of } \boldsymbol{\pi}^{(i)} \\ \text{(biased toward 1 if } i \text{ is odd, 0 if } i \text{ is even)}\end{array} && = 1/\sqrt{M_{i-1}} \\
\boldsymbol{\sigma}^{(i)} &= \text{distribution over level-}(i-1)\text{ variables} && = \begin{cases} \mathbf{b}_{1-\beta_i} \text{ to each variable (i.i.d.)} & i \text{ odd} \\ \mathbf{b}_{\beta_i} \text{ to each variable (i.i.d.)} & i \text{ even.} \end{cases}
\end{aligned}
$$

For the precise values, see Sections 4 and 6.1. For intuition, we remark that our AND-OR tree corresponds to a d-fold composition that alternates between the well-known *tribes function* (a read-once DNF) and its Boolean dual (a read-once CNF), with widths approximately $\log(M_1), \ldots, \log(M_d)$. The tribes function and its dual are approximately balanced, so the composition is also approximately balanced. The top fan-ins of these CNFs and DNFs are approximately $M_1 \cdot \ln(2), \ldots, M_d \cdot \ln(2)$, and hence after merging adjacent layers when possible, we get a depth-$(d+1)$ tree with the fan-ins as described above.

Observe that for $i \ge 2$, the values that $\boldsymbol{\pi}^{(i)}$ assigns to fixed variables are highly biased, alternately toward 1 or toward 0.

As mentioned above, the proof that F survives the projection procedure simply generalizes the analysis in Håstad, Rossman, Servedio, and Tan's work [HRS+17] to our different parameter setting. Intuitively, the "blockwise" correlations that are present in the projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$ (e.g., recall that when $\boldsymbol{\pi}^{(1)}$ keeps a variable alive, it assigns 1 to all of the non-surviving variables in that block) are tailored to the AND-OR-tree and designed to keep it alive. See Section 5 for details. The innovative part in our argument is showing that LTF circuits simplify under the projections.

LTF circuits simplify under projections: The high-level plan. Our argument has a high-level structure similar to the ones in prior work [CSS18; Tel18; HHT+22], but instead of arguing about the effects of traditional random restrictions, we now argue about the effects of each random projection $\boldsymbol{\pi}^{(i)}$ above (for any fixed $i \in [d]$). For simplicity, we assume from now on that i is odd,

in which case the assigned values of $\pi^{(i)}$ are biased toward 1 and the corresponding gate in F is an AND gate.

Similarly to the analysis in [HHT+22] (also implicit in the work of Chen, Santhanam, and Srinivasan [CSS18]), we will work with *hybrid computational models*. That is, on the way to proving that any LTF circuit f becomes (close to) a constant, we argue that after intermediate projections $\pi^{(i)}$ the circuit can be computed by a computational model that combines decision trees and LTF circuits; specifically, the tree queries variables to reach a leaf, and the leaf is labeled by an LTF circuit that is then applied to the input.[8]

Our proof has three main steps:

1. As a first step, we prove that applying $\pi^{(i)}$ to any LTF (i.e., any single gate in f) makes the LTF extremely close to a constant, with somewhat high probability. This probability is high, but not enough to allow a union bound on all gates. We will elaborate in Section 2.2.

2. Our second step is to argue that applying $\pi^{(i)}$ to any LTF circuit of depth $d+1-i$ with $n^{1+2^{-O(d)}}$ wires simplifies the circuit, with somewhat high probability, to be very close to a decision tree with LTF circuits of depth $d-i$ at its leaves, where the depth of the tree is significantly smaller than the number of living variables under $\pi^{(i)}$. We will elaborate in Section 2.3.

3. The final step is to show that applying $\pi^{(i)}$ to a decision tree with LTF circuits of depth $d+1-i$ at its leaves simplifies the tree, with high probability, such that it is close to a shallower decision tree in which the leaves are labeled by depth $d-i$ LTF circuits. We will elaborate in Section 2.4.

We stress that in all the statements above, the notion of "approximating a function" (i.e., when saying that a restricted function is close to a simpler function) refers to the distribution $\sigma^{(i+1)}$, rather than to the uniform distribution. Having proved the three steps above, the proof will analyze the applications of $\pi^{(i)}$ for $i = 1, \ldots, d$, arguing at each iteration i that the circuit simplifies with respect to the "next" distribution $\sigma^{(i+1)}$. In the last step the circuit will be a decision tree that queries only a sub-constant fraction of its variables. Such a tree cannot approximate the AND (or OR) function $F\restriction_\pi$ under $\sigma^{(d+1)}$. Indeed, with high probability over π, the tree is close to a constant under $\sigma^{(d+1)}$, whereas $F\restriction_\pi$ is nearly balanced (because $\sigma^{(d+1)}$ is biased).

2.2 Random projections simplify any single LTF

Chen, Santhanam, and Srinivasan showed that a random restriction that keeps a p-fraction of the variables alive simplifies any single LTF to be $\exp(-p^{-\Omega(1)})$-close to a constant, with probability at least $1 - p^{\Omega(1)}$ [CSS18]. A motivating observation for our analysis is that a *biased restriction*, which keeps a p-fraction of variables alive and fixes the other variables independently by $\mathbf{b}_{1-\beta}$, *simplifies any unweighted threshold function* to be $\exp(-p^{-\Omega(1)})$-close to a constant with respect to any product distribution, with probability $1 - (p/\beta)^{\Omega(1)}$.[9] In other words, for any bias $\beta > 0$

[8] This hybrid model is simpler than the one considered in [HHT+22], since the latter also allowed queries to LTF gates at internal nodes of the tree.

[9] To see this, let Φ be an n-bit unweighted LTF, and let $1/n \leq p \leq 1/2$. Consider a random restriction ρ that keeps a random subset S of $p \cdot n$ variables alive and fixes the variables in $[n] \setminus S$ independently according to $\mathbf{b}_{1-\beta}$. Hoeffding's inequality implies that (with respect to any fixed product distribution) the function is ϵ-close to a constant only if the sum of values assigned to variables in $[n] \setminus S$ falls in an interval of length $O(\sqrt{\log(1/\epsilon) \cdot p \cdot n})$ (see Corollary 3.14). By the Berry-Esseen theorem (see Lemma 6.9), the probability of this event is at most $O(\sqrt{\log(1/\epsilon) \cdot (p/\beta)})$.

of assignments to the fixed variables, the probability that an unweighted LTF fails to simplify is $(p/\beta)^{\Omega(1)}$, compared to the $p^{\Omega(1)}$ bound that Chen, Santhanam, and Srinivasan showed in their setting [CSS18].

We extend this statement to hold for an *arbitrary* (weighted) LTF rather than just the unweighted LTFs; to hold when the values assigned to fixed variables are correlated (i.e., within each block, if the block survives then all fixed variables are simultaneously assigned the value 1, and otherwise the probability of the all-ones string is artificially decreased); and to hold also when considering a merging of the variables after applying the restriction. In this more challenging setting, we show a slightly worse error bound of $(p/\beta^2)^{\Omega(1)}$ compared to $(p/\beta)^{\Omega(1)}$:

Theorem 2.2 (LTFs simplify under corrupted biased block projections; informal, see Theorem 6.5)**.** *Let Φ be an* LTF *on n variables, let $\boldsymbol{\pi}$ be a corrupted biased block projection with parameters p and β, and let $\boldsymbol{\sigma}$ be a product distribution (possibly with biased marginals). Assume that each block B_j satisfies $\epsilon/n \leq (1-\beta)^{|B_j|} \leq p$. Then the probability that $\Phi{\restriction_{\boldsymbol{\pi}}}$ is not ϵ-close to a constant under $\boldsymbol{\sigma}$ is*

$$O\left(\left(\frac{p}{\beta^2}\right)^{1/3} \cdot \log\left(\frac{n}{\epsilon}\right)\right).$$

Before explaining the ideas in the proof, let us comment on the subtlety of the parameters obtained in Theorem 2.2. First, in Theorem 2.2, we assume both upper and lower bounds on the quantity $(1-\beta)^{|B_j|}$. That is, we assume that the block size $|B_j|$ is neither too big nor too small. Fortunately, this "Goldilocks" condition is indeed satisfied by our projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$ with high probability.[10] (It is also satisfied by Håstad, Rossman, Servedio, and Tan's projections [HRS+17].) Thus, Theorem 2.2 applies to $\boldsymbol{\pi}^{(i)}$ with failure probability $O((p_i/\beta_i^2)^{1/3} \cdot \log(n/\epsilon))$.

Secondly, the projections in Håstad, Rossman, Servedio, and Tan's original work [HRS+17] satisfy $p \approx \beta$. For such parameters, the bound of Theorem 2.2 would be useless. However, our modified projections have parameters p and β that vary from one iteration to the next, and crucially, the bias parameter β in each iteration is approximately equal to the block survival probability p in the *previous* iteration, i.e., $\beta_i \approx p_{i-1}$. A key property of our projections is that they are "increasingly aggressive" in the sense that $p_i \ll p_{i-1}$ (specifically $p_i \approx p_{i-1}^{100}$). Therefore, the bound of Theorem 2.2 is indeed small when we apply it to our projections.

To prove Theorem 2.2, we generalize the "structure vs. randomness" approach that Chen, Santhanam, and Srinivasan used to show that LTFs simplify under traditional random restrictions [CSS18]. Loosely speaking, their proof first analyzes "regular" LTFs, i.e., LTFs in which the weights are reasonably well-spread (this is the "random" case). Under the assumption of regularity, they argue that the weighted sum of assigned values is *anti-concentrated*, and thus unlikely to land in the small interval that would cause the restricted LTF to be somewhat balanced. To complement this analysis, they analyze LTFs that have a small number of "heavy" variables (this is the "structure" case). If the number of heavy variables is sufficiently small, then it is possible to fix them and reduce to the regular case, and otherwise they show anti-concentration among these "heavy" variables alone.

To make this approach work in our setting, the main challenge is establishing anti-concentration in the regular case.[11] Recall that in the projection $\boldsymbol{\pi}^{(i)}$, after the set of living

[10] Actually, in the rare event that a block is an unacceptable size, our projection assigns values to all variables in that block from an i.i.d. product distribution independently of the other blocks (see Definition 6.3), and we show that this does not affect the rest of the analysis.

[11] By comparison, our analysis of non-regular LTFs is a relatively straightforward adaptation of techniques from

variables has been fixed, non-surviving variables in surviving blocks are always assigned the value 1; this deterministic assignment does not contribute any anti-concentration at all. In non-surviving blocks, the assignment is random, but the assigned values are not independent, because the probability of the all-ones assignment is artificially decreased. The effect of this "corruption" within a single block is limited (because the all-ones assignment would be rare even without corruption). However, there are many blocks, and the *overall* effect is statistically significant; nevertheless, our goal is to show anti-concentration despite these corruptions.

To prove anti-concentration of the weighted sum of the assigned values, we first observe that with high probability, the variables in non-surviving blocks have a constant fraction of the total ℓ_2-weight. (By "ℓ_2-weight," we mean the sum of the squares of the weights.) We may therefore focus on such variables (and ignore the non-surviving variables in surviving blocks). To handle the corruptions in non-surviving blocks, we show that the weighted sum of assigned values to fixed variables in non-surviving blocks can be represented as the weighted sum of truly independent Bernoulli variables, plus an error term. The sum of independent variables is anti-concentrated by the Berry-Esseen theorem (see Lemma 6.9). To handle the error term, we bound its variance (this is where we use the assumption that the blocks are not too small, in which case the all-ones assignment would be rare even without corruption; see Lemma 6.7 for details). We thereby show that the error term is *concentrated*, and therefore it does little harm to the anti-concentration property of the sum of independent variables.

The anti-concentration established by the preceding arguments must be compared to the ℓ_2-weight of the living variables. Here we face another potential pitfall: When variables are merged, their weights effectively add, which typically increases the ℓ_2-weight of the living variables (making the LTF more balanced). This potential pitfall is the reason that we assume that the blocks are not too big. The assumption in Theorem 2.2 implies that with high probability, the number of variables that are merged in each block is small – only $O(\log(n/\epsilon))$ – and therefore the detrimental effect of mergings is limited.

2.3 Random projections simplify LTF circuits (if we allow some queries)

The next step is to argue that for every LTF circuit f of slightly super-linear size, with high probability, $\pi^{(i)}$ "simplifies" the entire bottom layer of f. Ideally, we are hoping that the gates in the bottom layer become close to constants. We cannot simply apply a union bound to claim that they are all close to constants simultaneously, because the failure probability in Theorem 2.2 might be too large. Instead, following prior work, we argue that *after querying a sub-linear number of the remaining variables*, each gate in the bottom layer is either close to a constant (over the distribution $\sigma^{(i+1)}$) or has fan-in one. Thus, the projected circuit $f\restriction_{\pi^{(i)}}$ can be approximated (over $\sigma^{(i+1)}$) by a decision tree whose leaves are labeled by shallower LTF circuits.

Given appropriate techniques from prior work [CSS18; Tel18; KL18; HHT+22; BKK+22], this is the easiest part of our argument, because those techniques do not depend on the assignments to fixed variables, but rather only on concentration properties of the number of living variables inside certain sets. We include a brief explanation of the argument here for completeness (see Section 6.3 for details).

We partition the gates in the bottom layer of f into "heavy gates" and "light gates" based on their fan-in. Most light gates have only one (or zero) living variable feeding into them after the projection, so they can be replaced with a wire (or eliminated), and we query the variables feeding

prior work. Note that both the regular and the non-regular cases contribute to the final error bound in Theorem 2.2, and as discussed, that error bound forces us to use a careful choice of parameters in our construction.

into the remaining light gates (there are few such variables, because these gates are light). Most heavy gates become close to a constant by Theorem 2.2, and we query all the variables feeding into the remaining heavy gates. The total number of such queries is bounded because the total number of wires in the circuit is bounded. (The latter argument is carried out by a standard bucketing technique, looking at all gates with fan-in roughly 2^i for each i and using the simple observation that there can be at most $w/2^i$ such gates for each i.)

2.4 Random projections simplify decision trees with LTF circuits at their leaves

The previous step yielded a decision tree T with LTF circuits at its leaves. The last key piece in our proof is arguing that each such decision tree simplifies, under $\boldsymbol{\pi}^{(i)}$, to a shallower decision tree with shallower LTF circuit at its leaves. (Indeed, we need the tree depth to decrease by a factor of $\approx \overline{p}_i$, and we need the circuits to decrease by one layer.)

First, we show that the tree depth indeed shrinks, with high probability, by a factor of $\approx \overline{p}_i$. This turns out to be not as straightforward as it might seem, due to correlations and mergings in $\boldsymbol{\pi}^{(i)}$; see Lemma 7.1 for details. Nonetheless, the more interesting part of the argument is arguing that we can use shallower LTF circuits at the leaves. The natural strategy to try and prove this is to claim that for each leaf, the corresponding circuit simplifies under $\boldsymbol{\pi}^{(i)}$ with high probability, and thus the fraction of "bad" leaves is small and we can replace those by constants, obtaining a tree with similar functionality.

The problem with this approach is the correlations between variables in the same block under the projection $\boldsymbol{\pi}^{(i)}$. At each *fixed* leaf, simplification occurs with high probability, but we must analyze the *random* leaf reached when we apply $\boldsymbol{\pi}^{(i)}$ to T and then plug in an input sampled from $\boldsymbol{\sigma}^{(i+1)}$. In particular, the leaf is determined in part by $\boldsymbol{\pi}^{(i)}$, and the event of reaching a particular leaf can be *correlated* with the event that simplification fails at that leaf. It is therefore not clear how to show that simplification occurs with high probability at the random leaf that we reach.[12]

Dealing with this issue is the most subtle part of our argument, and it involves a two-step approach.

Preprocessing the tree. As a first step, we "preprocess" the tree T, transforming it into a new tree $\widetilde{T}$. The new tree $\widetilde{T}$ simulates T and in fact refines T in the following way: if T ever queries too many variables in a block, or if T ever queries a variable in some block and observes a 0 (a somewhat unlikely event as bits are biased towards 1), then $\widetilde{T}$ queries all variables in that block. It turns out that these modifications are not too costly, in the sense that after applying the projection $\boldsymbol{\pi}^{(i)}$, the two projected trees $T\restriction_{\boldsymbol{\pi}^{(i)}}$ and $\widetilde{T}\restriction_{\boldsymbol{\pi}^{(i)}}$ have similar query complexities. Briefly, this holds for the following two reasons:

- The event of querying too many variables in a single block can only happen so many times given T's depth bound, and most such blocks don't survive the projection, so these events only cause $\widetilde{T}\restriction_{\boldsymbol{\pi}^{(i)}}$ to perform a few additional queries compared to $T\restriction_{\boldsymbol{\pi}^{(i)}}$.
- If a variable x_j in block B is observed to be 0, we have two cases. If the block B is non-surviving under $\boldsymbol{\pi}^{(i)}$, then $\widetilde{T}\restriction_{\boldsymbol{\pi}^{(i)}}$ does not need to query any variable in B, because they are all assigned. On the other hand, if the block B is surviving, then the individual

[12] In previous work [HHT+22] a similar challenge occurred, since the path to each leaf contained LTFs. However, the challenge there was significantly easier, since the number of LTFs on a path was small and thus it was possible to easily upper bound their effect on the resulting distribution.

variable $x_j \in B$ must survive in $\boldsymbol{\pi}^{(i)}$, because non-surviving variables in surviving blocks are assigned the value 1. Therefore, in this case, both $T\restriction_{\boldsymbol{\pi}^{(i)}}$ and $\widetilde{T}\restriction_{\boldsymbol{\pi}^{(i)}}$ query the single "merged" variable corresponding to the entire block B. Thus, in either case, observing a 0 ultimately does not cause $\widetilde{T}\restriction_{\boldsymbol{\pi}^{(i)}}$ to perform any additional queries compared to $T\restriction_{\boldsymbol{\pi}^{(i)}}$.

Conditional analysis of corrupted biased block projections. For the second step, consider the process of applying $\boldsymbol{\pi}^{(i)}$ to $\widetilde{T}$ and then plugging in an input sampled from $\boldsymbol{\sigma}^{(i+1)}$. We analyze the joint distribution of $\boldsymbol{\pi}^{(i)}$ and $\boldsymbol{\sigma}^{(i+1)}$ *conditioned* on the event of reaching a leaf ℓ. Because of the preprocessing step, we can make a "win-win" argument: for each block, either (a) the tree queries every single variable in the block, or (b) the tree only makes a few queries to the block and observes 1 each time. In case (a), we can assume without loss of generality that the circuit C_ℓ labeling the leaf ℓ ignores all variables in that block, hence we can ignore the block. In case (b), the constraints on the queries help us to bound the extent to which conditioning distorts the distributions of $\boldsymbol{\pi}^{(i)}$ and $\boldsymbol{\sigma}^{(i+1)}$.

For example, we show that the event we are conditioning on in case (b) is a high-probability event regardless of whether the block survives, and hence the conditioning has little effect on the block's survival probability. By analyzing our projection distribution in more detail, we show that instead of applying $\boldsymbol{\pi}^{(i)}$ to the circuit C_ℓ, plugging in an input sampled from $\boldsymbol{\sigma}^{(i+1)}$, and conditioning on the event of reaching ℓ, we can equivalently imagine applying *another corrupted biased block projection* $\widetilde{\boldsymbol{\pi}}$, plugging in an input sampled from *another product distribution* $\widetilde{\boldsymbol{\sigma}}$, and *not conditioning on anything*. The parameters of $\widetilde{\boldsymbol{\pi}}$ and $\widetilde{\mathbf{y}}$ are slightly different than the parameters of $\boldsymbol{\pi}^{(i)}$ and $\boldsymbol{\sigma}^{(i+1)}$, but our analysis of a single circuit is sufficiently robust against these small distortions to conclude that T simplifies with high probability.

2.5 Putting it all together

To summarize our discussion so far, we show that when we apply the random projection $\boldsymbol{\pi}^{(i)}$ to a decision tree $\mathbf{T}_{i-1}$ with LTF circuits at its leaves, we get another decision tree $\mathbf{T}_i$ with LTF circuits at its leaves that is "simpler" in the sense that the circuit-depth decreases by 1. The tree $\mathbf{T}_i$ agrees with the projected function $\mathbf{T}_{i-1}\restriction_{\boldsymbol{\pi}^{(i)}}$ with high probability under the product distribution $\boldsymbol{\sigma}^{(i+1)}$. To finish the proof, we need to apply some type of triangle inequality. For example, we know that $\mathbf{T}_1 \approx \mathbf{T}_0\restriction_{\boldsymbol{\pi}^{(1)}}$ and $\mathbf{T}_2 \approx \mathbf{T}_1\restriction_{\boldsymbol{\pi}^{(2)}}$; we want to conclude that $\mathbf{T}_2 \approx \mathbf{T}_0\restriction_{\boldsymbol{\pi}^{(2)} \circ \boldsymbol{\pi}^{(1)}}$.

We are indeed able to show that $\mathbf{T}_d \approx \mathbf{T}_0\restriction_{\boldsymbol{\pi}^{(d)} \circ \cdots \circ \boldsymbol{\pi}^{(1)}}$ by relying upon a crucial feature of the projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$ and the product distributions $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(d+1)}$. These projections and product distributions are *compatible* with each other, in the sense that applying $\boldsymbol{\pi}^{(i)}$ and then assigning values sampled from $\boldsymbol{\sigma}^{(i+1)}$ yields exactly the distribution $\boldsymbol{\sigma}^{(i)}$. (See Lemma 5.2.)

The same feature (the "completion property") is also crucial in the work of Håstad, Rossman, Servedio, and Tan [HRS+17]. However, the completion property plays a different role in their work than it does in ours. Their work is focused on *average-case* lower bounds; they rely on the fact that applying $\boldsymbol{\pi}^{(d)} \circ \cdots \circ \boldsymbol{\pi}^{(1)}$ and then assigning values sampled from $\boldsymbol{\sigma}^{(d+1)}$ yields the uniform distribution over inputs. The completion property is likewise an essential ingredient of our average-case separation, but the distinction is that in our setting, the completion property would still be crucial even if we were merely aiming for a *worst-case* separation. After all, the simplification we achieve at intermediate stages of our argument is itself only approximate, forcing us to use techniques designed for average-case separations.

The completion property holds trivially in the traditional setting of truly random restrictions, because the values assigned by the restriction are themselves independent and uniform. In both

our work and the work of Håstad *et al.* [HRS+17], there are correlations between the values assigned to different variables, which are essential for ensuring that the AND-OR tree F survives. In both works, the *purpose of merging variables* (i.e., the purpose of using projections rather than restrictions) is to achieve the completion property despite these correlations.

Organization

In Section 4 we formally define our hard AND-OR functions (in Section 4.1), and random projections (in Section 4.2). Section 5 is dedicated to proving that the hard function survives the sequence of random projections. Sections 6 and 7 establish that random projections simplify any LTF circuit with a bounded number of wires to (nearly) a constant (see Theorem 6.1). Finally, we prove Theorem 1.1 by putting things together in Section 8. Hardness magnification for our construction (Theorem 1.2) is proved in Section 9.

3 Preliminaries

3.1 Approximations

Throughout this paper, the notation $a \pm b$ denotes the interval $[a-b, a+b]$ rather than the two-point set $\{a-b, a+b\}$. The following inequalities are standard.

Proposition 3.1 (The approximation $1-c \approx e^{-c}$ for small c)**.** *If $c < 1$, then*

$$\exp\left(-\frac{c}{1-c}\right) \leq 1-c \leq \exp(-c).$$

(In fact, the right-hand inequality holds for all $c \in \mathbb{R}$.) Consequently, if $c \in [0, 1/2]$, then

$$1 \pm c \subseteq \exp(0 \pm 2c) \quad \text{and} \quad \exp(0 \pm c) \subseteq 1 \pm 2c.$$

3.2 Projections

A *projection* is a generalization of a restriction in which a variable can be assigned a value or mapped to a new variable.[13] In the formal definition, we include 0 and 1 in the domain of the projection to facilitate reasoning about compositions.

Definition 3.2 (Projections)**.** *A* projection *is a function $\pi\colon \mathcal{X} \cup \{0,1\} \to \mathcal{Y} \cup \{0,1\}$, where $\mathcal{X}$ and $\mathcal{Y}$ are sets of formal Boolean variables, with the property that $\pi(0) = 0$ and $\pi(1) = 1$.*

If $\pi(x_i) \in \mathcal{Y}$, we say that x_i *survives* the projection. Observe that the composition of two projections, denoted $\pi_2 \circ \pi_1$, is once again a projection.

We will sometimes identify a projection $\pi\colon \mathcal{X} \cup \{0,1\} \to \mathcal{Y} \cup \{0,1\}$ with the corresponding *string* $\pi \in (\mathcal{Y} \cup \{0,1\})^{\mathcal{X}}$. As a special case, we identify each binary string $y \in \{0,1\}^{\mathcal{Y}}$ with a projection $y\colon \mathcal{Y} \cup \{0,1\} \to \{0,1\}$ in the natural way. This identification allows us to *compose strings with projections*. If $\pi\colon \mathcal{X} \cup \{0,1\} \to \mathcal{Y} \cup \{0,1\}$ is a projection and $y \in \{0,1\}^{\mathcal{Y}}$, then $y \circ \pi \in \{0,1\}^{\mathcal{X}}$.

[13] Filmus, Meir, and Tal study a natural, more general model of projections in which a variable might be mapped to the *negation* of another variable [FMT21]. We have no need of this extra generality, so we omit it from our definition for simplicity.

Recall that we will use projections in which the assignments to fixed variables are biased, alternating between being very close to 1 and being very close to 0. The following definition facilitates swapping the roles of 0 and 1 in a projection, in order to present projections with such biases as "complements" of one another.

Definition 3.3 (Complement projection)**.** *Let $\pi\colon \mathcal{X} \cup \{0,1\} \to \mathcal{Y} \cup \{0,1\}$ be a projection. The* complement projection $\overline{\pi}\colon \mathcal{X} \cup \{0,1\} \to \mathcal{Y} \cup \{0,1\}$ *is defined as follows: for $x_i \in \mathcal{X}$, we set*

$$\overline{\pi}(x_i) = \begin{cases} 1 - \pi(x_i) & \text{if } \pi(x_i) \in \{0,1\} \\ \pi(x_i) & \text{if } \pi(x_i) \in \mathcal{Y}. \end{cases}$$

3.3 Functions

Definition 3.4 (Applying a projection to a function)**.** *Let $f\colon \{0,1\}^{\mathcal{X}} \to \Omega$ be a function and let $\pi\colon \mathcal{X} \cup \{0,1\} \to \mathcal{Y} \cup \{0,1\}$ be a projection. The* projected function $f\restriction_\pi\colon \{0,1\}^{\mathcal{Y}} \to \Omega$ *is given by*

$$f\restriction_\pi(y) = f(y \circ \pi).$$

That is, we assign a value to each variable in $\mathcal{X}$ by first applying the projection π and then, if it survives, plugging in the appropriate coordinate of y. The original function f is evaluated on this assignment.

Definition 3.5 (Approximators)**.** *Let $f, \widetilde{f}\colon \{0,1\}^n \to \{0,1\}$, let $\mathbf{X}$ be a distribution over $\{0,1\}^n$, and let $\epsilon > 0$. We say that $\widetilde{f}$* approximates *f* under *$\mathbf{X}$* with error *ϵ if*

$$\Pr_{\mathbf{x} \sim \mathbf{X}}\left[\widetilde{f}(\mathbf{x}) \neq f(\mathbf{x})\right] \leq \epsilon.$$

As a special case, when $\widetilde{f}$ is a constant function, we say that f is ϵ-close to constant under $\mathbf{X}$*. If f is not ϵ-close to constant under $\mathbf{X}$, then we say that f is* ϵ-far from constant under $\mathbf{X}$*. We say that f is* ϵ-unbiased under $\mathbf{X}$ *if $\mathbb{E}[f(\mathbf{X})] \in \frac{1}{2} \pm \epsilon$.*

3.4 The Bernoulli distribution and product distributions

We denote product distributions over $\{0,1\}^t$ using the following notation.

Definition 3.6. *For $\alpha \in [0,1]$, let $\mathbf{b}_\alpha$ denote the distribution of a Bernoulli random variable that takes value 1 with probability α and 0 with probability $1 - \alpha$. For $t \in \mathbb{N}$, let $\mathbf{b}_\alpha^t$ denote the product distribution of t independent bits, each distributed according to $\mathbf{b}_\alpha$. More generally, for a vector $\vec{\alpha} \in [0,1]^t$, let $\mathbf{b}_{\vec{\alpha}}$ denote the product distribution with t coordinates with marginal distributions $\mathbf{b}_{\alpha_1}, \ldots, \mathbf{b}_{\alpha_t}$. If $\mathcal{X}$ is a set of variables, we also write $\mathbf{b}_\alpha^{\mathcal{X}}$ to denote the distribution $\mathbf{b}_\alpha^{|\mathcal{X}|}$, thought of as an assignment to $\mathcal{X}$.*

To facilitate swapping the roles of 0 and 1, we generalize the notation $\mathbf{b}_\alpha$ as follows.

Definition 3.7. *For $b \in \{0,1\}$ and $\alpha \in [0,1]$, let $\mathbf{b}_{\alpha \to b}$ denote the distribution of a Bernoulli random variable that takes value b with probability α and $1 - b$ with probability $1 - \alpha$. We similarly define $\mathbf{b}_{\alpha \to b}^t$ and $\mathbf{b}_{\alpha \to b}^{\mathcal{X}}$.*

3.5 Conditional independence

We record the following elementary fact that we use several times.

Lemma 3.8. *Let $\mathbf{x}_1, \ldots, \mathbf{x}_t$ be independent random variables. Let $\mathcal{E}$ be an event of the form $\mathcal{E} = \mathcal{E}_1 \wedge \cdots \wedge \mathcal{E}_t$, where $\mathcal{E}_j$ depends only on $\mathbf{x}_j$. Then $\mathbf{x}_1, \ldots, \mathbf{x}_t$ are conditionally independent given $\mathcal{E}$, and furthermore, for each $j \in [t]$, the conditional distributions $(\mathbf{x}_j \mid \mathcal{E})$ and $(\mathbf{x}_j \mid \mathcal{E}_j)$ are identical.*

3.6 Concentration bounds

Theorem 3.9 (Multiplicative Chernoff bounds)**.** *Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be independent random variables in $[0,1]$, and let $\mu = \mathbb{E}[\sum_{i=1}^n \mathbf{x}_i]$. Then, for any $\delta > 0$,*

$$\Pr\left[\sum_{i=1}^n \mathbf{x}_i > (1+\delta)\cdot\mu\right] \leq \exp\left(-\frac{\delta^2}{\delta+2}\cdot\mu\right),$$

and for any $\delta \in (0,1)$,

$$\Pr\left[\sum_{i=1}^n \mathbf{x}_i < (1-\delta)\cdot\mu\right] \leq \exp\left(-\frac{\delta^2}{2}\cdot\mu\right).$$

The following three variants of the Chernoff bound all follow readily from Theorem 3.9; see Appendix B for the proofs. The first variant is a convenient "additive error" setting of parameters.

Corollary 3.10. *Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be independent random variables in $[0,1]$, let $\mu \geq 0$, and suppose $\mathbb{E}[\sum_{i=1}^n \mathbf{x}_i] \leq \mu$. Then, for any $\epsilon > 0$,*

$$\Pr\left[\sum_{i=1}^n \mathbf{x}_i > 2\mu + 3\ln(1/\epsilon)\right] \leq \epsilon.$$

The second variant is a version of the Chernoff bound designed for the situation that the expectation μ is itself only approximately equal to some "ideal" value μ_*.

Corollary 3.11. *Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be independent random variables in $[0,1]$. Let $\mu = \mathbb{E}[\sum_{i=1}^n \mathbf{x}_i]$, let $\mu_* > 0$, let $\epsilon \in (0, 1/2)$, and assume that $\mu \in \mu_* \cdot (1 \pm \epsilon)$. Then for any $\delta \in (2\epsilon, 1)$,*

$$\Pr\left[\sum_{i=1}^n \mathbf{x}_i \notin \mu_* \cdot (1 \pm \delta)\right] \leq 2\exp\left(-\frac{\delta^2 \cdot \mu_*}{42}\right).$$

The third variant is a version of the Chernoff bound for random bits that are *not* independent.

Corollary 3.12 (Upper Chernoff bound for correlated random bits)**.** *Let $\mathbf{x}_0, \ldots, \mathbf{x}_n$ and $\mathbf{y}_1, \ldots, \mathbf{y}_n$ be discrete random variables (not necessarily independent) where $\mathbf{y}_i \in \{0,1\}$. Assume that for every $i \in [n]$, the variables $\mathbf{y}_1, \ldots, \mathbf{y}_i$ are determined by $\mathbf{x}_i$. Let $\zeta > 0$, and assume that for every $i \in [n]$ and every $x \in \mathrm{Supp}(\mathbf{x}_{i-1})$, we have*

$$\Pr[\mathbf{y}_i = 1 \mid \mathbf{x}_{i-1} = x] \leq \zeta.$$

Then for any $\epsilon > 0$,

$$\Pr\left[\sum_{i=1}^n \mathbf{y}_i > 2\zeta \cdot n + 3\ln(1/\epsilon)\right] \leq \epsilon.$$

We will use Hoeffding's inequality to prove that certain LTFs are close to constant under product distributions.

Theorem 3.13 (Hoeffding's inequality)**.** *Let $w \in \mathbb{R}^n$, let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be independent random variables with $\mathbf{x}_i \in [0, w_i]$, and let $\mu = \mathbb{E}[\sum_{i=1}^n \mathbf{x}_i]$. Then, for any $R > 0$,*

$$\Pr\left[\left|\mu - \sum_{i=1}^n \mathbf{x}_i\right| \geq R \cdot \|w\|_2\right] \leq 2\exp\left(-2R^2\right).$$

Corollary 3.14 (sufficient condition for being close to constant under a product distribution)**.** *Let Φ be an* LTF *on t input bits, say $\Phi(x) = 1 \iff \sum_{i=1}^{t} w_i \cdot x_i \geq \theta$. Let $\vec{\alpha} \in [0,1]^t$, let $\mu = \sum_i \alpha_i \cdot w_i$, and let $\epsilon > 0$. If*

$$|\theta - \mu| > \frac{1}{2}\sqrt{\ln(2/\epsilon)} \cdot \|w\|_2,$$

then Φ is ϵ-close to constant under $\mathbf{b}_{\vec{\alpha}}$.

4 The setup: AND-OR trees and corresponding random projections

4.1 The AND-OR tree

Let $d \in \mathbb{N}$. Our "hard function" is a depth-$(d+1)$ AND-OR tree (a monotone read-once AC^0 formula). The gates at distance i from the inputs are AND gates if i is odd and OR gates if i is even. Historically, the "grandparent" of our construction is Sipser's function [Sip83b], which is an AND-OR tree in which every gate has the same fan-in $n^{1/(d+1)}$. Håstad, Rossman, Servedio, and Tan studied a variant where the fan-ins are different in the bottom-most and top-most layers [HRS+17]. We will modify the construction still further. In this section, we define a family of trees in which the fan-ins are allowed to vary from one layer to the next (although gates in the same layer always have the same fan-in). Ultimately, we will pick parameters so that the fan-ins increase as we go up the tree, as discussed after the statement of Theorem 1.1 and in Section 2.1.

The fan-ins are governed by a sequence of parameters $\vec{M} = (M_1, \ldots, M_d) \in \mathbb{N}^d$, where M_1 is a power of two. Ideally, when we plug in a uniform random input, we want each gate at distance $i \leq d$ from the inputs to output 1 with probability $1/M_i$ (resp. $1 - 1/M_i$) assuming i is odd (resp. even), and we want the final output gate to output 1 with probability $1/2$. (When it comes time to analyze LTF circuits, we will fix $M_i = M^{100^{i-1}}$. The results of this section and the next hold for a generic choice of $M_1, \ldots, M_d$, so let us leave them unfixed for now.)

In reality, there will be some roundoff errors. Each gate at distance i from the inputs will output 1 with probability c_i (resp. $1 - c_i$) assuming i is odd (resp. even), where c_i will be close to the "ideal" value $1/M_i$ but potentially slightly smaller. The precise definition follows.

Definition 4.1 (The AND-OR tree $F_{d+1,\vec{M}}$)**.** *Let $\vec{M} = (M_1, \ldots, M_d) \in \mathbb{N}^d$, where M_1 is a power of two. Let $M_{d+1} = 2$ and let $c_0 = 1/2$. For $i = 1, \ldots, d+1$, let f_i be the smallest integer such that $(1 - c_{i-1})^{f_i} \leq 1/M_i$ and let $c_i = (1 - c_{i-1})^{f_i}$. The function $F_{d+1,\vec{M}}$ is defined by a depth-$(d+1)$ read-once formula of alternating levels of* AND *gates and* OR *gates, starting with variables feeding into* AND *gates. For $i \in [d+1]$, the fan-in of each gate at distance i from the inputs is f_i. For each gate v, let F_v denote the subformula rooted at v and let* Children(v) *denote the set of gates feeding into v. (See Table 1 and Figure 1.)*

One can show by induction on i that c_i indeed has the interpretation that we claimed prior to Definition 4.1:

Proposition 4.2. *Let $i \in \{0, 1, \ldots, d+1\}$, let $b = i \bmod 2$, and let v be a gate a distance i from the inputs. Then*

$$\Pr_{\mathbf{x} \in \{0,1\}^n}[F_v(\mathbf{x}) = b] = c_i.$$

Let $M = \min\{M_1, \ldots, M_d\}$; we will often assume that M is greater than some sufficiently large constant (independent of d). For comparison, the function studied by Håstad, Rossman,

Parameter	Meaning	Value to have in mind	
d	Depth $= d+1$	$O(1)$	*
c_i	$c_i = \mathbb{E}[F_v]$ where height$(v) = i$ (odd)	$1/M_i$	**
M_i	Tree is constructed so that $c_i \approx 1/M_i$	$M^{100^{i-1}}$	*
M	$\min\{M_1, \ldots, M_d\}$	$n^{2^{-\Theta(d)}}$	**
f_i	Fan-in at distance i from inputs	$\begin{cases} \log(M_1) & i = 1 \\ M_{i-1} \cdot \ln(M_i) & 1 < i \leq d \\ M_d \cdot \ln(2) & i = d+1 \end{cases}$	**

Table 1: The parameters of the AND-OR tree $F_{d+1,\vec{M}}$.

*The definition of $F_{d+1,\vec{M}}$ allows for other values, but these are the values we are most interested in.
**These values are approximate.

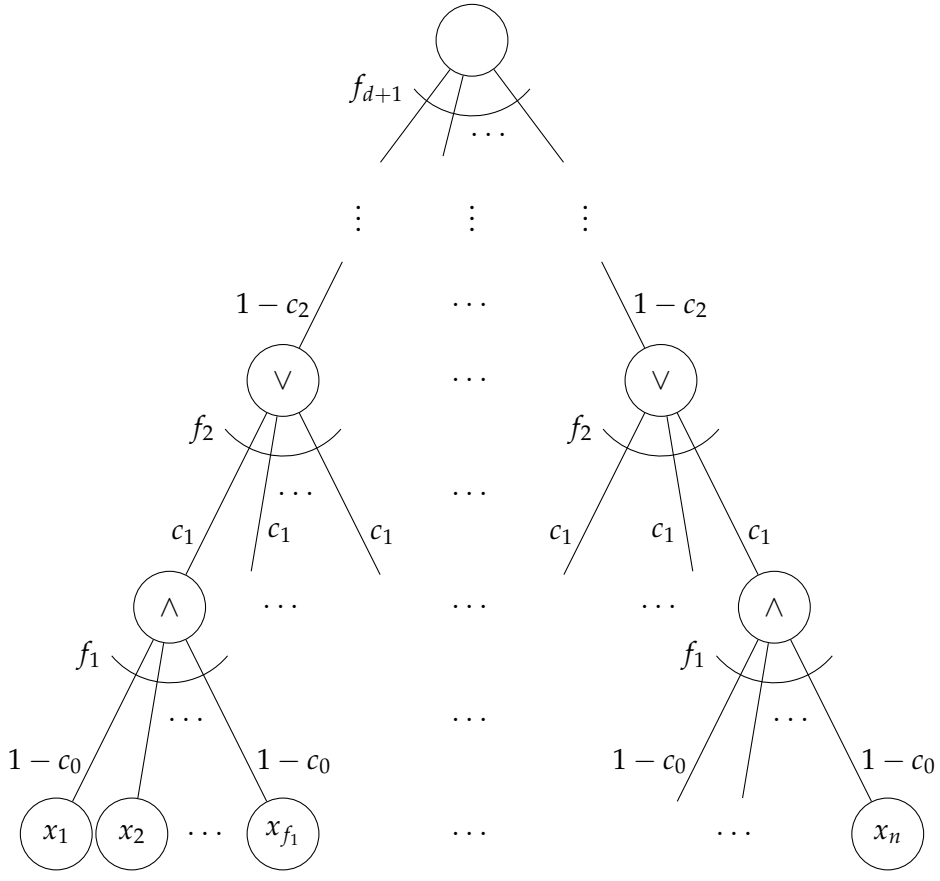

Figure 1: The AND-OR tree $F_{d+1,\vec{M}}$. The depth is $d+1$. The output gate is $\wedge$ if the depth is odd and $\vee$ if the depth is even. The label of a wire is the expectation of that wire when the input to the formula is chosen uniformly at random from $\{0,1\}^n$.

Servedio, and Tan [HRS+17] is parameterized by a single value m and is attained as a special case of Definition 4.1 by taking $M_1 = M_2 = \cdots = M_d = 2^{2m}$. We now estimate the fan-ins $f_1, \ldots, f_{d+1}$ and the probabilities $c_1, \ldots, c_{d+1}$.

Claim 4.3. *Assume M is sufficiently large. Then*

$$\begin{aligned} c_1 &= 1/M_1 \\ c_i &= (1/M_i) \cdot (1 \pm 1/M) && \text{for } i \in \{2, \ldots, d+1\} \\ f_1 &= \log(M_1) \\ f_i &= M_{i-1} \cdot \ln(M_i) \cdot (1 \pm 3/M) && \text{for } i \in \{2, \ldots, d+1\}. \end{aligned}$$

Proof. Since M_1 is a power of two, $f_1 = \log(M_1)$ and $c_1 = 1/M_1$. For all $i \geq 1$ we have $c_i \leq 1/M_i$ by construction, and therefore for $i > 1$ we have

$$c_i > (1/M_i) \cdot (1 - c_{i-1}) \geq (1/M_i) \cdot (1 - 1/M_{i-1}) \geq (1/M_i) \cdot (1 - 1/M), \tag{4.1}$$

completing the proof of the bounds for c_i. Furthermore, for $i > 1$, we have

$$\begin{aligned} \frac{1}{\ln(1/(1-c_{i-1}))} &\leq \frac{1}{c_{i-1}} && \text{by Proposition 3.1} \\ &\leq \frac{M_{i-1}}{1 - 1/M} && \text{by Equation (4.1)} \\ &\leq M_{i-1} \cdot (1 + 2/M), \\ \text{and} \quad \frac{1}{\ln(1/(1-c_{i-1}))} &\geq \frac{1 - c_{i-1}}{c_{i-1}} && \text{by Proposition 3.1} \\ &\geq \frac{1 - 1/M_{i-1}}{1/M_{i-1}} && \text{since } c_{i-1} \leq 1/M_{i-1} \\ &\geq M_{i-1} \cdot (1 - 1/M). \end{aligned}$$

Therefore, for $i > 1$, we have

$$f_i = \left\lceil \frac{\ln(M_i)}{\ln(1/(1-c_{i-1}))} \right\rceil \in M_{i-1} \cdot \ln(M_i) \cdot (1 \pm 3/M). \quad \blacksquare$$

In particular, since $c_{d+1} = \mathbb{E}[F_{d+1,\vec{M}}]$ and $M_{d+1} = 2$, Claim 4.3 implies that $F_{d+1,\vec{M}}$ is approximately balanced.

4.2 The sequence of random projections

We wish to prove that an arbitrary depth-d LTF circuit with $n^{1+\gamma}$ wires cannot compute (or even approximate) the AND-OR tree $F_{d+1,\vec{M}}$. To do so, as discussed in Section 2, we will apply a sequence of random projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$. We will argue that after these projections, the AND-OR tree is still nontrivial, whereas the LTF circuit will be drastically simpler. To get started, in this section, we define the projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$.

We associate a formal variable x_v with any gate v in the tree defining $F_{d+1,\vec{M}}$ including the leaves. We denote by $\mathcal{X}_i$ the set of variables associated with gates at distance i from the inputs. For each internal vertex v, we let $\text{Children}(x_v) = \{x_u : u \in \text{Children}(v)\}$. The projection $\boldsymbol{\pi}^{(i)}$ maps $\mathcal{X}_{i-1} \cup \{0,1\} \to \mathcal{X}_i \cup \{0,1\}$. Moreover, each variable will always be projected to 0, 1, or its parent. Let $\boldsymbol{\pi}^{(1\ldots i)} = \boldsymbol{\pi}^{(i)} \circ \boldsymbol{\pi}^{(i-1)} \circ \cdots \circ \boldsymbol{\pi}^{(1)}$, so $\boldsymbol{\pi}^{(1\ldots i)}$ maps $\mathcal{X}_0 \cup \{0,1\} \to \mathcal{X}_i \cup \{0,1\}$. We

identify $\mathcal{X}_0$ with the original input variables of the AND-OR tree, so it makes sense to apply $\boldsymbol{\pi}^{(1\ldots i)}$ to $F_{d+1,\vec{M}}$.

For $i = 1, \ldots, d+1$ independently, sample a "tentative" assignment $\boldsymbol{\sigma}^{(i)}$ to $\mathcal{X}_{i-1}$ by independently assigning each variable according to the Bernoulli distribution $\mathbf{b}_{1-\beta_i \to i \bmod 2}$, where

$$\beta_i \stackrel{\text{def}}{=} \begin{cases} 1/2 & \text{if } i = 1 \\ M_{i-1}^{-1/2} & \text{if } i \in \{2, \ldots, d+1\}. \end{cases} \tag{4.2}$$

That is,

$$\boldsymbol{\sigma}^{(i)} \sim \mathbf{b}_{1-\beta_i \to i \bmod 2}^{\mathcal{X}_{i-1}}.$$

For each variable $x_v \in \mathcal{X}_i$ and each $x_u \in \text{Children}(x_v)$, the projection $\boldsymbol{\pi}^{(i)}$ will either assign $\boldsymbol{\pi}^{(i)}(x_u) = \boldsymbol{\sigma}_u^{(i)}$ or else it will keep x_u alive, i.e., $\boldsymbol{\pi}^{(i)}(x_u) = x_v$. Now let us explain how $\boldsymbol{\pi}^{(i)}$ decides which variables will stay alive.

Definition 4.4 (The projection $\boldsymbol{\pi}^{(i)}$ for $i = 1, \ldots, d$)**.** *Assume that we have already sampled $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(i-1)}$. Let $b = i \bmod 2$. For each gate v at distance i from the inputs independently: Let $\boldsymbol{B}_v$ be the set of children of x_v such that the corresponding subformulas are not yet determined, i.e.,*

$$\boldsymbol{B}_v = \left\{ x_u \in \text{Children}(x_v) : F_u \restriction_{\boldsymbol{\pi}^{(1\ldots i-1)}} \equiv x_u \right\}.$$

For each $x_u \in \text{Children}(x_v) \setminus \boldsymbol{B}_v$, set $\boldsymbol{\pi}^{(i)}(x_u) = \boldsymbol{\sigma}_u^{(i)}$. Regarding the variables in $\boldsymbol{B}_v$:

1. *If $|\boldsymbol{B}_v| \notin (f_i/\sqrt{M_{i-1}}) \cdot (1 \pm M^{-1/8})$, then set $\boldsymbol{\pi}^{(i)}(x_u) = \boldsymbol{\sigma}_u^{(i)}$ for each $x_u \in \boldsymbol{B}_v$. (Define $M_0 = 1$.)*
2. *Otherwise, if $\boldsymbol{\sigma}_{\boldsymbol{B}_v}^{(i)} = b^{\boldsymbol{B}_v}$, then with probability $M_i^{-1/4}$, set $\boldsymbol{\pi}^{(i)}(x_u) = \boldsymbol{\sigma}_u^{(i)} = b$ for all $x_u \in \boldsymbol{B}_v$. With the remaining probability, sample $\boldsymbol{\varphi} \sim \mathbf{b}_{\beta_i}^{\boldsymbol{B}_v}$ conditioned on $\boldsymbol{\varphi} \neq 0^{\boldsymbol{B}_v}$, and for each $x_u \in \boldsymbol{B}_v$, set*

$$\boldsymbol{\pi}^{(i)}(x_u) = \begin{cases} x_v & \text{if } \boldsymbol{\varphi}_u = 1 \\ \boldsymbol{\sigma}_u^{(i)} = b & \text{if } \boldsymbol{\varphi}_u = 0. \end{cases}$$

3. *Otherwise ($\boldsymbol{\sigma}_{\boldsymbol{B}_v}^{(i)} \neq b^{\boldsymbol{B}_v}$)[14], with probability $1 - q_i(|\boldsymbol{B}_v|)$ (see Equation (1) below), set $\boldsymbol{\pi}^{(i)}(x_u) = \boldsymbol{\sigma}_u^{(i)}$ for each $x_u \in \boldsymbol{B}_v$. With the remaining probability, for each $x_u \in \boldsymbol{B}_v$, set*

$$\boldsymbol{\pi}^{(i)}(x_u) = \begin{cases} x_v & \text{if } \boldsymbol{\sigma}_u^{(i)} = 1 - b \\ \boldsymbol{\sigma}_u^{(i)} & \text{if } \boldsymbol{\sigma}_u^{(i)} = b. \end{cases}$$

(See Table 2.)

As discussed in Section 2, our projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$ are based on the projections studied by Håstad, Rossman, Servedio, and Tan [HRS+17]. The main modification is suitably generalizing the definition to accommodate the varying fan-ins in the AND-OR tree $F_{d+1,\vec{M}}$. (We also made some other minor changes for convenience, but they are not essential.)

Note that when $i = 1$, case 1 does not happen. Also, we always have

$$(\boldsymbol{\pi}^{(i)})_{\boldsymbol{B}_v} \in \{0,1\}^{\boldsymbol{B}_v} \cup \{b, x_v\}^{\boldsymbol{B}_v}.$$

[14] This is the most common case.

Parameter	Meaning	Value
i	$\boldsymbol{\pi}^{(i)}$ maps $\mathcal{X}_{i-1} \cup \{0,1\} \to \mathcal{X}_i \cup \{0,1\}$	$1, 2, \ldots, d$
b	More-likely value for each bit of $\boldsymbol{\sigma}^{(i)}$	$i \bmod 2$
β_i	Bias of each bit of $\boldsymbol{\sigma}^{(i)}$	$\begin{cases} 1/2 & i = 1 \\ 1/\sqrt{M_{i-1}} & 1 < i \le d \end{cases}$
$q_i(\cdot)$	Pr[survival | case 3]	$1/\sqrt{M_i}$ (approximate)

Table 2: The parameters of the projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$. Recall that $\boldsymbol{\sigma}^{(i)}$ is a "tentative assignment" to the variables in $\mathcal{X}_{i-1}$.

That is, if $\boldsymbol{\pi}^{(i)}$ keeps any variables in $\boldsymbol{B}_v$ alive, then it assigns the value b to the remaining variables in $\boldsymbol{B}_v$. The definition of $\boldsymbol{\pi}^{(i)}$ refers to a value $q_i(\cdot)$, which we define as follows:

$$q_i(t) \stackrel{\text{def}}{=} \frac{(1-\beta_i)^t}{1-(1-\beta_i)^t} \cdot \frac{1-\beta_{i+1}}{\beta_{i+1}} \cdot (1 - M_i^{-1/4}). \tag{1}$$

(The reason for this formula will become clear later, in the proof of Lemma 5.2.) Since the definition of $\boldsymbol{\pi}^{(i)}$ uses $q_i(t)$ as a probability, we must verify that $q_i(t) \in [0,1]$, at least for the values of t that actually arise in the definition of $\boldsymbol{\pi}^{(i)}$. For the sake of analysis that will come later, we also give estimates for the values $(1-\beta_i)^t$ and $q_i(t)$.

Lemma 4.5. *Assume M is sufficiently large. Suppose that either (a) $i = 1$ and $t = f_1$, or else (b) $1 < i \le d$ and $t \in \left(f_i/\sqrt{M_{i-1}}\right) \cdot (1 \pm M^{-1/8})$. Then $q_i(t) \in [0,1]$. If additionally $M_i \le \exp(M^{1/16}/2)$, then*

$$(1-\beta_i)^t = (1/M_i) \cdot (1 \pm 2M^{-1/16})$$
$$\text{and } q_i(t) = \left(1/\sqrt{M_i}\right) \cdot (1 \pm 3M^{-1/16}).$$

Proof. If $i = 1$ and $t = f_1 = \log(M_1)$, then we have the exact equality $(1-\beta_1)^t = 1/M_1$. In the other case,

$$\begin{aligned} (1-\beta_i)^t &\le \exp(-t\beta_i) && \text{by Proposition 3.1} \\ &\le \exp(-(f_i/M_{i-1}) \cdot (1 - M^{-1/8})) \\ &\le \exp(-\ln(M_i) \cdot (1 - 3/M) \cdot (1 - M^{-1/8})) && \text{by Claim 4.3} \\ &\le M_i^{-1+o(1)}, \end{aligned} \tag{4.3}$$

where the $o(1)$ term goes to 0 as $M \to \infty$. Consequently, in either case,

$$q_i(t) \le \frac{M_i^{-1+o(1)}}{1 - M_i^{-1+o(1)}} \cdot \frac{1 - 1/\sqrt{M_i}}{1/\sqrt{M_i}} \cdot (1 - M_i^{-1/4}) \le M_i^{-1/2+o(1)} \le 1.$$

Clearly $q_i(t) \ge 0$, so $q_i(t) \in [0,1]$. Now assume $M_i \le \exp(M^{1/16}/2)$. Then in the $i > 1$ case, continuing from Equation (4.3), we have

$$(1-\beta_i)^t \le \exp(-\ln(M_i) \cdot (1 - 2M^{-1/8}))$$

$$
\begin{aligned}
&= \exp(-\ln(M_i) + \ln(M_i) \cdot 2M^{-1/8}) \\
&\le \exp(-\ln(M_i) + (M^{1/16}/2) \cdot 2M^{-1/8}) \\
&= (1/M_i) \cdot \exp(M^{-1/16}), \\
\text{and } (1-\beta_i)^t &\ge \exp\left(-\frac{t\beta_i}{1-\beta_i}\right) && \text{by Proposition 3.1} \\
&\ge \exp\left(-\frac{t}{\sqrt{M_{i-1}}} \cdot \left(1 + 2/\sqrt{M_{i-1}}\right)\right) \\
&\ge \exp\left(-\ln(M_i) \cdot (1 + M^{-1/8}) \cdot (1 + 3/M) \cdot (1 + 2/\sqrt{M})\right) && \text{by Claim 4.3} \\
&\ge \exp\left(-\ln(M_i) \cdot (1 + 2M^{-1/8})\right) \\
&\ge (1/M_i) \cdot \exp\left(-M^{-1/16}\right).
\end{aligned}
$$

Thus, by Proposition 3.1,

$$
(1-\beta_i)^t \in (1/M_i) \cdot \exp\left(0 \pm M^{-1/16}\right) \subseteq (1/M_i) \cdot (1 \pm 2M^{-1/16}).
$$

Consequently,

$$
\begin{aligned}
q_i(t) &\in \frac{(1/M_i) \cdot (1 \pm 2M^{-1/16})}{1 - (1/M_i) \cdot (1 \pm 2M^{-1/16})} \cdot \frac{1 - 1/\sqrt{M_i}}{1/\sqrt{M_i}} \cdot (1 - 1/M_i^{1/4}) \\
&\subseteq (1/\sqrt{M_i}) \cdot (1 \pm 3M^{-1/16}). \quad \blacksquare
\end{aligned}
$$

5 The AND-OR tree survives the projections

In this section, we will show that with high probability, the AND-OR tree $F_{d+1,\vec{M}}$ remains nontrivial after applying the projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$. In particular, the projected function computes an AND or OR of approximately $\sqrt{M_d} \cdot \ln(2)$ many variables.

Theorem 5.1 (The AND-OR tree survives the projections)**.** *Assume M is sufficiently large. With probability at least $1 - O(M^{-1/4})$ over $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$, the projected function $(F_{d+1,\vec{M}})\restriction_{\boldsymbol{\pi}^{(1\ldots d)}}$ is an* AND *gate (if $d+1$ is odd) or an* OR *gate (if $d+1$ is even) with fan-in $\sqrt{M_d} \cdot \ln(2) \cdot (1 \pm 5M^{-1/8})$. Moreover, in this case, the projected function is $(6M^{-1/8})$-unbiased under $\boldsymbol{\sigma}^{(d+1)}$, i.e.,*

$$
\mathop{\mathbb{E}}_{\boldsymbol{\sigma}^{(d+1)}}\left[(F_{d+1,\vec{M}})\restriction_{\boldsymbol{\pi}^{(1\ldots d)}}(\boldsymbol{\sigma}^{(d+1)})\right] \in \frac{1}{2} \pm 6M^{-1/8}.
$$

Theorem 5.1 and its proof mimic the analysis by Håstad, Rossman, Servedio, and Tan [HRS+17].

5.1 The completion property

Following the analysis of Håstad, Rossman, Servedio, and Tan [HRS+17], the first step of the proof of Theorem 5.1 is to prove that the projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$ are "compatible" with the distributions $\boldsymbol{\sigma}^{(1)}, \ldots, \boldsymbol{\sigma}^{(d+1)}$ in a certain sense. This lemma will also be useful later, when we prove that LTF circuits simplify under $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$ (see the proof of Proposition 7.6).

Lemma 5.2 (The completion property)**.** *Let $i \in [d]$ and assume M is sufficiently large. Conditioned on any fixed values for $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(i-1)}$, the random variables $\boldsymbol{\sigma}^{(i)}$ and $\boldsymbol{\sigma}^{(i+1)} \circ \boldsymbol{\pi}^{(i)}$ are identically distributed over $\{0,1\}^{\mathcal{X}_{i-1}}$.*

Proof. It suffices to focus on a single gate v at distance i from the inputs and consider the assignment to its children. Looking at the definition of $\boldsymbol{\pi}^{(i)}$, whether we are in Case 1 depends only on the history $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(i-1)}$, which we have fixed. If we are in Case 1, then the lemma is trivial, so assume that we are not in Case 1. The set $\boldsymbol{B}_v$ is also determined by the history $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(i-1)}$, so it is fixed. Define

$$\boldsymbol{\kappa}_v = \begin{cases} b & \text{if } \boldsymbol{\sigma}_{\boldsymbol{B}_v} = b^{\boldsymbol{B}_v} \\ 1-b & \text{otherwise,} \end{cases}$$

or equivalently, $\boldsymbol{\kappa}_v$ is b if we are in Case 2 and $1-b$ if we are in Case 3.

Let $\mathbf{S}_v \subseteq \boldsymbol{B}_v$ be the set of variables that $\boldsymbol{\pi}^{(i)}$ keeps alive, i.e., $\mathbf{S}_v = (\boldsymbol{\pi}^{(i)})^{-1}(x_v)$. For any nonempty set $S \subseteq \boldsymbol{B}_v$, we have

$$\Pr[\boldsymbol{\kappa}_v = b \wedge \mathbf{S}_v = S] = (1-\beta_i)^{|\boldsymbol{B}_v|} \cdot (1 - M_i^{-1/4}) \cdot \frac{\beta_i^{|S|} \cdot (1-\beta_i)^{|\boldsymbol{B}_v|-|S|}}{1-(1-\beta_i)^{|\boldsymbol{B}_v|}}.$$

Meanwhile,

$$\Pr[\boldsymbol{\kappa}_v = 1-b \wedge \mathbf{S}_v = S] = q_i(|\boldsymbol{B}_v|) \cdot \beta_i^{|S|} \cdot (1-\beta_i)^{|\boldsymbol{B}_v|-|S|}.$$

By our choice of $q_i(\cdot)$ (Equation (1)), we have

$$\frac{\Pr[\boldsymbol{\kappa}_v = b \wedge \mathbf{S}_v = S]}{\Pr[\boldsymbol{\kappa}_v = 1-b \wedge \mathbf{S}_v = S]} = \frac{(1-\beta_i)^{|\boldsymbol{B}_v|} \cdot (1 - M_i^{-1/4})}{(1-(1-\beta_i)^{|\boldsymbol{B}_v|}) \cdot q_i(|\boldsymbol{B}_v|)} = \frac{\beta_{i+1}}{1-\beta_{i+1}},$$

and consequently

$$\Pr[\boldsymbol{\kappa}_v = b \mid \mathbf{S}_v = S] = \beta_{i+1}.$$

Note that the right-hand side has no dependence on S. Furthermore, recall that when $\mathbf{S}_v \neq \emptyset$, the projection $\boldsymbol{\pi}^{(i)}$ assigns b to all variables in $\boldsymbol{B}_v \setminus \mathbf{S}_v$, so in particular, when $\mathbf{S}_v \neq \emptyset$, the action of $\boldsymbol{\pi}^{(i)}$ on $\boldsymbol{B}_v$ is a deterministic function of $\mathbf{S}_v$. Therefore, conditioned on the event $\mathbf{S}_v \neq \emptyset$, the bit $\boldsymbol{\kappa}_v$ is *independent of* $\boldsymbol{\pi}^{(i)}$ and distributed according to $\mathbf{b}_{\beta_{i+1} \to b}$. Consequently, the random variables $\boldsymbol{\sigma}_v^{(i+1)} \circ (\boldsymbol{\pi}^{(i)})_{\text{Children}(v)}$ and $\boldsymbol{\kappa}_v \circ (\boldsymbol{\pi}^{(i)})_{\text{Children}(v)}$ are identically distributed.[15]

Now, recall that when $\boldsymbol{\kappa}_v = b$, we are in Case 2, so for each $x_u \in (\boldsymbol{\pi}^{(i)})^{-1}(x_v)$, we have $\boldsymbol{\sigma}_u^{(i)} = b$. Meanwhile, when $\boldsymbol{\kappa}_v = 1-b$, we are in Case 3, so for each $x_u \in (\boldsymbol{\pi}^{(i)})^{-1}(x_v)$, we have $\boldsymbol{\sigma}_u^{(i)} = 1-b$. Thus, either way,

$$\boldsymbol{\kappa}_v \circ (\boldsymbol{\pi}^{(i)})_{\text{Children}(v)} = (\boldsymbol{\sigma}^{(i)})_{\text{Children}(v)},$$

so $\boldsymbol{\sigma}_v^{(i+1)} \circ (\boldsymbol{\pi}^{(i)})_{\text{Children}(v)}$ and $(\boldsymbol{\sigma}^{(i)})_{\text{Children}(v)}$ are identically distributed, completing the proof. ∎

By induction, Lemma 5.2 immediately implies the following corollary.

Corollary 5.3 (The cumulative completion property)**.** *Assume M is sufficiently large. For any $i \in [d]$, the random variable $\boldsymbol{\sigma}^{(i+1)} \circ \boldsymbol{\pi}^{(1\ldots i)}$ is distributed identically to $\boldsymbol{\sigma}^{(1)}$, i.e., it is distributed uniformly over $\{0,1\}^{\mathcal{X}_0}$.*

[15] Recall that the notation $\kappa \circ \pi$ refers only to the final (composed) projection, regardless of the way each of its components (i.e., κ and π) affected the final projection.

5.2 Subformula collapse probabilities

The next step in the proof of Theorem 5.1 is to analyze the effect of the projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(i)}$ on gates at distance i from the inputs.

Lemma 5.4 (Subformula collapse probabilities)**.** *Assume M is sufficiently large. Let $i \in [d]$ and let $b = i \bmod 2$. For each gate v at distance i from the inputs,*

$$\Pr[F_v\restriction_{\boldsymbol{\pi}^{(1\ldots i)}} \equiv b] \in M_i^{-5/4} \cdot (1 \pm 3/M) \tag{5.1}$$

$$\Pr[F_v\restriction_{\boldsymbol{\pi}^{(1\ldots i)}} \equiv 1-b] \in 1 - M_i^{-1/2} \cdot (1 \pm 3M^{-1/4}) \tag{5.2}$$

$$\Pr[F_v\restriction_{\boldsymbol{\pi}^{(1\ldots i)}} \equiv x_v] \in M_i^{-1/2} \cdot (1 \pm 2M^{-1/4}). \tag{5.3}$$

Moreover, for each gate v at distance $i+1$ from the inputs,

$$\Pr\left[\left|\{u \in \text{Children}(v) : F_u\restriction_{\boldsymbol{\pi}^{(1\ldots i)}} \equiv x_u\}\right| \notin \frac{f_{i+1}}{\sqrt{M_i}} \cdot (1 \pm M^{-1/8})\right] \leq e^{-\frac{M^{1/4}\cdot \ln(M_{i+1})}{50}}. \tag{5.4}$$

Proof. The proof is by induction on i. We start with a base case. When $i = 1$, we have $b = 1$ and F_v is an AND of f_1 variables. Looking at the definition of $\boldsymbol{\pi}^{(1)}$, the AND collapses to 1 with probability $2^{-f_1} \cdot M_1^{-1/4} = M_1^{-5/4}$, showing that Equation (5.1) holds in this case.

Now consider any $i \in [d]$ and assume that Equation (5.1) holds for distance i. Let v be a gate at distance i from the inputs. We have

$$\begin{aligned} c_i &= \Pr[F_v(\boldsymbol{\sigma}^{(1)}) = b] && \text{(Proposition 4.2)} \\ &= \Pr[F_v(\boldsymbol{\sigma}^{(i+1)} \circ \boldsymbol{\pi}^{(1\ldots i)}) = b] && \text{(Corollary 5.3)} \\ &= \Pr[F_v\restriction_{\boldsymbol{\pi}^{(1\ldots i)}} \equiv b] + \Pr[F_v\restriction_{\boldsymbol{\pi}^{(1\ldots i)}} \equiv x_v] \cdot \beta_{i+1}, \end{aligned}$$

and hence

$$\Pr[F_v\restriction_{\boldsymbol{\pi}^{(1\ldots i)}} \equiv x_v] = \frac{(M_i^{-1} - M_i^{-5/4}) \cdot (1 \pm 3/M)}{M_i^{-1/2}} = M_i^{-1/2} \cdot (1 \pm 2M^{-1/4}),$$

completing the proof of Equation (5.3). From the definitions of $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$, it should be clear that $F_v\restriction_{\boldsymbol{\pi}^{(1\ldots i)}}$ is either a constant or else x_v, so Equation (5.2) follows, since the three probabilities must sum to 1.

Next, let v be a gate at distance $i+1$ from the inputs. Like in the definition of $\boldsymbol{\pi}^{(i+1)}$, let $\boldsymbol{B}_v$ be the set of children $x_u \in \text{Children}(x_v)$ such that $F_u\restriction_{\boldsymbol{\pi}^{(1\ldots i)}} \equiv x_u$. Then by Equation (5.3),

$$\mathbb{E}[|\boldsymbol{B}_v|] \in \left(f_{i+1}/\sqrt{M_i}\right) \cdot (1 \pm 2M^{-1/4}),$$

and the events $F_u\restriction_{\boldsymbol{\pi}^{(1\ldots i)}} \equiv x_u$ for $u \in \text{Children}(v)$ are independent, so by a suitable Chernoff bound (Corollary 3.11),

$$\Pr\left[|\boldsymbol{B}_v| \notin \left(f_{i+1}/\sqrt{M_i}\right) \cdot (1 \pm M^{-1/8})\right] \leq 2\exp\left(-\frac{M^{-1/4} \cdot f_{i+1}}{42 \cdot \sqrt{M_i}}\right) \leq \exp\left(-\frac{M^{1/4} \cdot \ln(M_{i+1})}{50}\right),$$

where the last inequality uses Claim 4.3. This completes the proof of Equation (5.4).

Now let us circle back and prove Equation (5.1) when $i > 1$. Let v be a gate at distance $i \in \{2, \ldots, d\}$. We may assume by induction that Equation (5.1) holds for distance $i-1$. Let $\mathcal{E}$ be the event that F_v has already collapsed to a constant *before* $\boldsymbol{\pi}^{(i)}$ is applied, i.e.,

$$F_v\restriction_{\boldsymbol{\pi}^{(1\ldots i-1)}} \equiv 0 \text{ or } F_v\restriction_{\boldsymbol{\pi}^{(1\ldots i-1)}} \equiv 1.$$

Let $p_{0,b}^{(i)}$ be the probability (with respect to the choice of $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(i-1)}$) that $\mathcal{E}$ occurs and the constant is b, i.e.,

$$p_{0,b}^{(i)} = \Pr[F_v\restriction_{\boldsymbol{\pi}^{(1\ldots i-1)}} \equiv b].$$

Next, let $p_{1,b}^{(i)}$ be the probability (this time with respect to the choice of $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(i)}$) that $\mathcal{E}$ does not occur, case 1 of the definition of $\boldsymbol{\pi}^{(i)}$ happens, and $F_v\restriction_{\boldsymbol{\pi}^{(1\ldots i)}} \equiv b$. Furthermore, let $p_2^{(i)}$ be the probability (again with respect to the choice of $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(i)}$) that $\mathcal{E}$ does not occur and case 2 of the definition of $\boldsymbol{\pi}^{(i)}$ happens. Observe that

$$\Pr[F_v\restriction_{\boldsymbol{\pi}^{(1\ldots i)}} \equiv b] = p_{0,b}^{(i)} + p_{1,b}^{(i)} + p_2^{(i)} \cdot M_i^{-1/4}.$$

By Equation (5.1),

$$\begin{aligned} p_{0,b}^{(i)} = \Pr[F_v\restriction_{\boldsymbol{\pi}^{(1,\ldots,i-1)}} \equiv b] &= \prod_{u \in \text{Children}(v)} \Pr[F_u\restriction_{\boldsymbol{\pi}^{(1,\ldots,i-1)}} \equiv b] \\ &\leq (1 - 0.9 \cdot M_{i-1}^{-1/2})^{f_i} \\ &\leq \exp\left(-0.8 \cdot \sqrt{M_{i-1}} \cdot \ln(M_i)\right) \\ &< M_i^{-3}. \end{aligned}$$

Meanwhile, by Equation (5.4),

$$p_{1,b}^{(i)} \leq \exp\left(-\frac{M^{1/4} \cdot \ln(M_i)}{50}\right) < M_i^{-3}.$$

Next, we estimate $p_2^{(i)}$. We have

$$\begin{aligned} c_i &= \Pr[F_v(\boldsymbol{\sigma}^{(1)}) = b] && \text{(Proposition 4.2)} \\ &= \Pr[F_v(\boldsymbol{\sigma}^{(i)} \circ \boldsymbol{\pi}^{(1\ldots i-1)}) = b] && \text{(Corollary 5.3)} \\ &= p_{0,b}^{(i)} + p_{1,b}^{(i)} + p_2^{(i)}, \end{aligned}$$

so

$$p_2^{(i)} = c_i - p_{0,b}^{(i)} - p_{1,b}^{(i)} = (1/M_i) \cdot (1 \pm 1/M) \pm 2M_i^{-3} = (1/M_i) \cdot (1 \pm 2/M).$$

Therefore,

$$\Pr[F_v\restriction_{\boldsymbol{\pi}^{(1\ldots i)}} \equiv b] = p_{0,b}^{(i)} + p_{1,b}^{(i)} + p_2^{(i)} \cdot M_i^{-1/4} = M_i^{-5/4} \cdot (1 \pm 3/M). \quad \blacksquare$$

5.3 Wrapping up the proof that the AND-OR tree survives the projections

Theorem 5.1 follows readily from Lemma 5.4.

Proof of Theorem 5.1. Let $b = d \bmod 2$ and let V_d be the set of gates at distance d from the inputs (i.e., children of the root gate). By Equation (5.1), for any gate $v \in V_d$,

$$\Pr\left[F_v\restriction_{\boldsymbol{\pi}^{(1\ldots d)}} \equiv b\right] = O\left(M_d^{-5/4}\right).$$

By the union bound, when we apply $\boldsymbol{\pi}^{(1\ldots d)}$, the probability that there is any $v \in V_d$ such that F_v collapses to b is at most $O(M_d^{-5/4} \cdot f_{d+1}) = O(M_d^{-1/4})$. Meanwhile, by Equation (5.4), except with

probability $\exp(-\Omega(M^{-1/4})) < O(M^{-1/4})$, the number of gates $v \in V_d$ such that $F_v\restriction_{\boldsymbol{\pi}^{(1\ldots d)}} \equiv x_v$ is $(f_{d+1}/\sqrt{M_d}) \cdot (1 \pm M^{-1/8})$. Assume that this event occurs and that for every $v \in V_d$, we have $F_v\restriction_{\boldsymbol{\pi}^{(1\ldots d)}} \not\equiv b$. When $d+1$ is odd, $b = 0$, so this indeed implies that the projected function $(F_{d+1,\vec{M}})\restriction_{\boldsymbol{\pi}^{(1,\ldots d)}}$ is an AND gate with fan-in

$$\frac{f_{d+1}}{\sqrt{M_d}} \cdot (1 \pm M^{-1/8}) = \sqrt{M_d} \cdot \ln(2) \cdot (1 \pm 5/M^{1/8}).$$

Similarly, when $d+1$ is even, $b = 1$, so $(F_{d+1,\vec{M}})\restriction_{\boldsymbol{\pi}^{(1,\ldots d)}}$ is an OR gate of the same fan-in.

Finally, let us show that in this case, the projected function is approximately balanced under $\boldsymbol{\sigma}^{(d+1)}$. Indeed,

$$\begin{aligned}
\Pr_{\boldsymbol{\sigma}^{(d+1)}}\left[(F_{d+1,\vec{M}})\restriction_{\boldsymbol{\pi}^{(1,\ldots d)}}(\boldsymbol{\sigma}^{(d+1)}) = 1-b\right] &\le \left(1 - 1/\sqrt{M_d}\right)^{\sqrt{M_d}\cdot\ln(2)\cdot(1-5M^{-1/8})} \\
&\le \exp\left(-\ln(2)\cdot(1-5M^{-1/8})\right) \\
&= \frac{1}{2}\cdot\exp\left(5M^{-1/8}\right) \\
&\le \frac{1}{2} + 5M^{-1/8} && \text{by Proposition 3.1,}
\end{aligned}$$

and

$$\begin{aligned}
&\Pr_{\boldsymbol{\sigma}^{(d+1)}}\left[F\restriction_{\boldsymbol{\pi}^{(1\ldots d)}}(\boldsymbol{\sigma}^{(d+1)}) = 1-b\right] \\
&\ge \left(1 - 1/\sqrt{M_d}\right)^{\sqrt{M_d}\cdot\ln(2)\cdot(1+5M^{-1/8})} \\
&\ge \exp\left(-\frac{1/\sqrt{M_d}}{1-1/\sqrt{M_d}}\cdot\sqrt{M_d}\cdot\ln(2)\cdot(1+5M^{-1/8})\right) && \text{by Proposition 3.1} \\
&\ge \exp\left(-\ln(2)\cdot(1+6M^{-1/8})\right) \\
&\ge \frac{1}{2} - 6M^{-1/8} && \text{by Proposition 3.1.} \quad \blacksquare
\end{aligned}$$

6 LTF circuits simplify under the projections

In the previous section, we showed that the depth-$(d+1)$ AND-OR tree $F_{d+1,\vec{M}}$ retains structure under the random projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$. In contrast, we will show that any depth-d LTF circuit with a bounded number of wires simplifies to a constant (approximately) under $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$. This contrast will allow us to conclude that the AND-OR tree cannot be computed (or even approximated) by such a circuit. The theorem below continues using the definitions and notations established in Section 4.

Theorem 6.1 (LTF circuits simplify under the projections)**.** *Let M be a sufficiently large power of two, let $d \in \mathbb{N}$, and use the parameters $M_i = M^{100^{i-1}}$ for $i = 1, \ldots, d$ to define the projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$. Assume $d \le 0.05 \cdot \log_{100} M$. Let $\mathcal{X}_0 \cup \{0,1\}$ be the domain of $\boldsymbol{\pi}^{(1)}$, let $n = |\mathcal{X}_0|$, and let f be a depth-d LTF circuit on $\mathcal{X}_0$ with w wires. The probability that the projected function $f\restriction_{\boldsymbol{\pi}^{(1\ldots d)}}$ is ξ-far from constant under the product distribution $\boldsymbol{\sigma}^{(d+1)}$ is at most ξ, where*

$$\xi \le M^{-1/96} \cdot \lceil w/n \rceil \cdot O(\log n)^{d+O(1)}. \tag{6.1}$$

(Recall that the projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$ are defined in Definition 4.4, and $\mathcal{X}_0$ is the set of input variables to the AND-OR tree $F_{d+1,\vec{M}}$ that is defined in Definition 4.1.) The point is that if w is slightly larger than n, then the leading $M^{-1/96}$ term of Equation (6.1) more than compensates for the w/n term.

6.1 Corrupted biased block projections

In the definition of $\boldsymbol{\pi}^{(i)}$, recall that we first pick a tentative assignment $\boldsymbol{\sigma}^{(i)}$, and then we randomly decide which variables to keep alive (if any) instead of assigning them the value stipulated by $\boldsymbol{\sigma}^{(i)}$. For the proof of Theorem 6.1, it will be helpful to reason in the reverse order: first we randomly choose the set of variables to keep alive, and then we assign values to the other variables. This process is captured by the following definitions.

Definition 6.2 (corrupted biased assignment)**.** *The* $(1-q)$-corrupted $(1-\beta)$-biased assignment *for a set* B *of variables, denoted* $\mathbf{corrb}^{B}_{1-\beta,1-q}$*, is the following distribution over* $\{0,1\}^B$*:*

1. *With probability* q*, we fix each variable independently according to* $\mathbf{b}_{1-\beta}$*.*
2. *With probability* $1-q$*, we fix each variable according to* $\mathbf{b}_{1-\beta}$ conditioned *on* not *fixing all variables in* B *to* 1*.*

Definition 6.3 (corrupted biased block projection)**.** *Let* $\mathcal{X} = \{x_1, \ldots, x_n\}$ *and* $\mathcal{Y} = \{y_1, \ldots, y_t\}$ *be sets of Boolean variables with* $t \leq n$*. Let* $\boldsymbol{\pi} \colon \mathcal{X} \cup \{0,1\} \to \mathcal{Y} \cup \{0,1\}$ *be a random projection. We say that* $\boldsymbol{\pi}$ *is a* p-surviving $(1-\beta)$-biased $(1-q)$-corrupted (ν,ν')-block projection *if for each variable* $y_j \in \mathcal{Y}$ *there is an associated "block"* $B_j \subseteq \mathcal{X}$ *such that* $B_1, \ldots, B_t$ *are disjoint; and for each* j*, either* $B_j = \emptyset$ *or else* $(1-\beta)^{|B_j|} \in [\nu,\nu']$*; and* $\boldsymbol{\pi}$ *behaves as follows:*

1. *Independently for each* $j \in [t]$ *such that* $B_j \neq \emptyset$*:*

   (a) *With probability* $p_j \leq p$*, the block* survives*: Sample* $\boldsymbol{\varphi} \sim \mathbf{b}^{B_j}_{1-\beta}$ *conditioned on* $\boldsymbol{\varphi} \neq 1^{B_j}$*, and for each variable* $x_i \in B_j$*, set*

   $$\boldsymbol{\pi}(x_i) = \begin{cases} 1 & \text{if } \boldsymbol{\varphi}_i = 1 \\ y_j & \text{if } \boldsymbol{\varphi}_i = 0. \end{cases}$$

   (b) *With probability* $1-p_j$*, the block is* non-surviving*, in which case we fix all the variables in* B_j *according to* $\mathbf{corrb}^{B_j}_{1-\beta,1-q_j}$ *for some* $q_j \leq q$*.*

   (We will consider empty blocks to be "non-surviving.")

2. *Independently of all the blocks, we assign values to the remaining variables, i.e., the variables in* $\mathcal{X} \setminus (B_1 \cup \cdots \cup B_t)$*, from the product distribution* $\mathbf{b}^{\mathcal{X}\setminus(B_1\cup\cdots\cup B_t)}_{1-\beta}$*.*

(See Table 3.)

In Definition 6.3, the constants 0 and 1 do not play the same role: each non-surviving variable in a surviving block is assigned 1, and in a non-surviving block, the assignment is biased toward 1 (thinking of β as small). This is just for the sake of convenience; when we want to swap the roles of 0 and 1, we will use the complementation operation given in Definition 3.3.

We allow empty blocks in Definition 6.3. This, too, is just for the sake of convenience. Because we allow empty blocks, the set $\mathcal{Y}$ can include each variable y_j that we "care about," regardless of whether $\boldsymbol{\pi}$ ever actually maps any variables to y_j.

Parameter	Meaning	Approximate value for $\boldsymbol{\pi}^{(i)}$
p	Max survival probability of a block	$1/\sqrt{M_i}$
β	Pr[var not assigned 1]	$\begin{cases} 1/2 & i = 1 \\ 1/\sqrt{M_{i-1}} & 1 < i \leq d \end{cases}$
ν and ν'	Bounds for $(1-\beta)^{\text{block size}}$	$1/M_i$
q	Pr[non-corrupted | non-surviving]	$M_i^{-1/4}$

Table 3: The parameters of a corrupted biased block projection.

6.1.1 Integrating the two definitions of the projection procedure

Continuing with the definitions and notations of Section 4: We now show that the projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$ (given by Definition 4.4) are corrupted biased block projections (up to complementation).

Lemma 6.4 (integration lemma)**.** *Let $d \in \mathbb{N}$, let $\vec{M} \in \mathbb{N}^d$ with M_1 a power of two, let $M = \min\{M_1, \ldots, M_d\}$, and assume M is sufficiently large. Let $i \in [d]$, let $p_i = 2/\sqrt{M_i}$, let $\nu_i = 1/2M_i$, let $q_i = 2M_i^{-1/4}$, and recall that*

$$\beta_i = \begin{cases} 1/2 & \text{if } i = 1 \\ 1/\sqrt{M_{i-1}} & \text{if } i > 1. \end{cases}$$

Assume that $M_i \leq \exp(M^{1/16}/2)$. Conditioned on any fixed values for $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(i-1)}$, either $\boldsymbol{\pi}^{(i)}$ (if i is odd) or the complement projection $\overline{\boldsymbol{\pi}}^{(i)}$ (if i is even) is a (p_i)-surviving $(1-\beta_i)$-biased $(1-q_i)$-corrupted $(\nu_i, 3\nu_i)$-block projection.

Proof. Let v be a vertex at distance i from the inputs in the AND-OR tree $F_{d+1,\vec{M}}$. Recall that $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(i-1)}$ (i.e., the history) determines whether v falls in Case 1 in Definition 4.4 of $\boldsymbol{\pi}^{(i)}$. If we are in Case 1, then the block associated with the variable x_v is $\emptyset$. Assume now that we are in Cases 2 or 3 in the definition of $\boldsymbol{\pi}^{(i)}$ for this vertex v. In this case, we consider the set $\boldsymbol{B}_v$ that appears in the definition of $\boldsymbol{\pi}^{(i)}$ to be the block associated with x_v. (Note that $\boldsymbol{B}_v$ is also determined by $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(i-1)}$.) By Lemma 4.5, we indeed have

$$(1-\beta_i)^{|\boldsymbol{B}_v|} \in (1/M_i) \cdot (1 \pm 2M^{-1/16}) \in [\nu_i, 3\nu_i].$$

Let us now analyze the probability that the block survives. Looking at the definition of $\boldsymbol{\pi}^{(i)}$, the survival probability is

$$(1-\beta_i)^{|\boldsymbol{B}_v|} \cdot (1 - M_i^{-1/4}) + (1 - (1-\beta_i)^{|\boldsymbol{B}_v|}) \cdot q_i(|\boldsymbol{B}_v|).$$

Applying Lemma 4.5 again, we can estimate this survival probability as

$$\begin{aligned} &(1/M_i) \cdot (1 \pm 2M^{-1/16}) \cdot (1 - M_i^{-1/4}) + (1 \pm 2/M_i) \cdot \left(1/\sqrt{M_i}\right) \cdot (1 \pm 3M^{-1/16}) \\ &\quad \subseteq \left(1/\sqrt{M_i}\right) \cdot (1 \pm 4M^{-1/16}) \subseteq [0, p_i]. \end{aligned}$$

Next, let us analyze the case that the block survives. In this case, all of the non-surviving variables are indeed projected to $i \bmod 2$, which is 1 after complementing if i is even. Furthermore, conditioned on the block surviving, the set of surviving variables is indeed distributed correctly.

Finally, we must analyze the case that the block does not survive. Observe that we might think of Cases 2 and 3 as follows. We flip biased coins $\mathbf{c}_{2,v}, \mathbf{c}_{3,v}$ with $\Pr[\mathbf{c}_{2,v} = 1] = M_i^{-1/4}$ and $\Pr[\mathbf{c}_{3,v} = 1] = q_i(|\boldsymbol{B}_v|)$. We decide what to do in Cases 2 and 3 based on the results of $\mathbf{c}_{2,v}$ and $\mathbf{c}_{3,v}$. Furthermore, the coins may be correlated or anti-correlated in any way as long as their marginals are as above since we never "read" both the values of $\mathbf{c}_{2,v}$ and $\mathbf{c}_{3,v}$. Since in our case $\Pr[\mathbf{c}_{2,v} = 1] + \Pr[\mathbf{c}_{3,v} = 1] \leq 1$ we may assume that it never happens that both $\mathbf{c}_{2,v} = 1$ and $\mathbf{c}_{3,v} = 1$.[16] In terms of these coins $\mathbf{c}_{2,v}$ and $\mathbf{c}_{3,v}$, the projection $\boldsymbol{\pi}^{(i)}$ has the following properties.

- Conditioned on $(\mathbf{c}_{2,v}, \mathbf{c}_{3,v}) = (1,0)$, the block does not survive, and the assignment comes from the product distribution $\mathbf{b}_{1-\beta_i}^{\boldsymbol{B}_v}$ (after complementing if i is even).
- Conditioned on $(\mathbf{c}_{2,v}, \mathbf{c}_{3,v}) = (0,1)$, the block survives.
- Conditioned on $(\mathbf{c}_{2,v}, \mathbf{c}_{3,v}) = (0,0)$ and reaching Case 2, the block survives.
- Conditioned on $(\mathbf{c}_{2,v}, \mathbf{c}_{3,v}) = (0,0)$ and reaching Case 3, the block does not survive, and the assignment comes from the conditional distribution $(\boldsymbol{\varphi} \sim \mathbf{b}_{1-\beta_i}^{\boldsymbol{B}_v} \mid \boldsymbol{\varphi} \neq 1^{\boldsymbol{B}_v})$ (again, after complementing if i is even).
- $(\mathbf{c}_{2,v}, \mathbf{c}_{3,v}) = (1,1)$ cannot happen (see discussion above about the choice of the correlation between these coins).

Thus, conditioned on non-survival, the assignment is some convex combination of the product distribution $\mathbf{b}_{1-\beta_i}^{\boldsymbol{B}_v}$ and the conditional distribution $(\boldsymbol{\varphi} \sim \mathbf{b}_{1-\beta_i}^{\boldsymbol{B}_v} \mid \boldsymbol{\varphi} \neq 1^{\boldsymbol{B}_v})$; that is, it is a $(1-q)$-corrupted $(1-\beta_i)$-biased distribution for some parameter q. We conclude the proof by verifying that $q \leq q_i$. We have

$$\begin{aligned} q &= \frac{\Pr[(\mathbf{c}_{2,v}, \mathbf{c}_{3,v}) = (1,0)]}{\Pr[(\mathbf{c}_{2,v}, \mathbf{c}_{3,v}) = (1,0)] + \Pr[\text{Case 3}] \cdot \Pr[(\mathbf{c}_{2,v}, \mathbf{c}_{3,v}) = (0,0)]} \\ &= \frac{M_i^{-1/4}}{M_i^{-1/4} + (1 - o(1)) \cdot (1 - M_i^{-1/4} - q_i(|\boldsymbol{B}_v|))} \\ &= M_i^{-1/4} \cdot (1 \pm o(1)) \leq 2M_i^{-1/4}. \quad \blacksquare \end{aligned}$$

6.2 LTFs simplify under corrupted biased block projections

At this point we have shown that the projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$ are corrupted biased block projections. We will next analyze the effect of a corrupted biased block projection on a single LTF gate. Our goal is to prove the following.

Theorem 6.5 (LTFs simplify under corrupted biased block projections)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets of variables with $|\mathcal{Y}| = t \leq n = |\mathcal{X}|$. Let $\Phi = (w, \theta)$ be an* LTF *on $\mathcal{X}$. Let $\boldsymbol{\pi} \colon \mathcal{X} \cup \{0,1\} \to \mathcal{Y} \cup \{0,1\}$*

[16] More precisely, think of a joint way to sample $\mathbf{c}_{2,v}, \mathbf{c}_{3,v}$ as follows: pick a random real number z uniformly in $[0,1]$, set $\mathbf{c}_{2,v} = 1$ if and only if $z \in [0, M_i^{-1/4}]$ and $\mathbf{c}_{3,v} = 1$ if and only if $z \in [1 - q_i(|\boldsymbol{B}_v|), 1]$. The two intervals are disjoint, since $M_i^{-1/4} + q_i(|\boldsymbol{B}_v|) < 1$.

be a p-surviving $(1-\beta)$-biased $(1-q)$-corrupted (ν,ν')-block projection, where $\beta \le 1/2$. Then for any $\epsilon > 0$ and any $\vec{\alpha} \in [0,1]^t$, the probability that $\Phi\restriction_{\boldsymbol{\pi}}$ is ϵ-far from constant under $\mathbf{b}_{\vec{\alpha}}$ is at most

$$O\left((p^{1/3} + (\nu')^{1/3}) \cdot \beta^{-2/3} \cdot \log\left(\frac{n}{\epsilon \cdot \nu}\right)\right).$$

In particular, when $p^{O(1)} \le \nu \le \nu' \le p$ (this will be the case when we analyze $\boldsymbol{\pi}^{(1)}, \dots, \boldsymbol{\pi}^{(d)}$; see Lemma 6.4), the failure probability bound above simplifies to

$$O\left(\left(\frac{p}{\beta^2}\right)^{1/3} \cdot \log\left(\frac{n}{\epsilon \cdot p}\right)\right).$$

As a reminder, in the theorem statement, $\Phi\restriction_{\boldsymbol{\pi}}$ denotes the result of applying the projection $\boldsymbol{\pi}$, in the sense of both fixing some variables and merging some living variables. (See Definition 3.4.) In particular, the meaning of (being far from constant under) $\mathbf{b}_{\vec{\alpha}}$ is that each surviving block B_j (which corresponds to a merged variable) has bias α_j in this distribution, and biases α_j for non-surviving blocks B_j are ignored.

Recall that we are working toward proving Theorem 6.1, which says that LTF circuits simplify under the projections. In Theorem 6.1, the input distribution to the projected function is $\boldsymbol{\sigma}^{(d+1)}$, which is a product distribution where every variable has the same bias. The reader might therefore wonder why Theorem 6.5 allows a whole vector of potentially-distinct biases $\vec{\alpha}$. Without going into detail, the reason is that at intermediate stages of the analysis, the distributions $\boldsymbol{\sigma}^{(i)}$ will get "distorted" because we will condition on certain events. See Section 2.4 for an informal discussion of this issue, or see Section 7 for details. For now, the reader might wish to focus on the case that the biases α_j are all equal, for simplicity.

6.2.1 Bounding the number of living variables per block

The first step in the proof of Theorem 6.5 is to argue that with high probability, no block has many surviving variables. This follows by a fairly straightforward Chernoff-bound argument:

Lemma 6.6 (each block has at most a few living variables)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets of variables with $|\mathcal{Y}| = t \le n = |\mathcal{X}|$. Let $\boldsymbol{\pi}\colon \mathcal{X} \cup \{0,1\} \to \mathcal{Y} \cup \{0,1\}$ be a p-surviving $(1-\beta)$-biased $(1-q)$-corrupted (ν,ν')-block projection with blocks $B_1, \dots, B_t$. Let $\mathbf{I} \subseteq \mathcal{X}$ be the set of living variables, i.e., $\mathbf{I} = \boldsymbol{\pi}^{-1}(\mathcal{Y})$. Then*

$$\Pr\left[\max_{j \in [t]} |\mathbf{I} \cap B_j| > 3\ln(n/\nu)\right] \le O(p + \nu').$$

Proof. Without loss of generality, we may assume that $\nu' \le 1/2$, because otherwise the conclusion of the lemma is trivial. Fix any nonempty block $B = B_j$. With probability at least $1-p$, the block B does not survive, in which case $\mathbf{I} \cap B = \emptyset$. With probability at most p, the block B does survive. Conditioned on B surviving, to determine which variables within B survive, the projection samples $\boldsymbol{\varphi} \sim \mathbf{b}^B_{1-\beta}$ conditioned on $\boldsymbol{\varphi} \ne 1^B$, and then $\boldsymbol{\varphi}^{-1}(0)$ is the set of living variables in that block. Thus, conditioned on B surviving, we can bound the probability that many variables within B survive by

$$\begin{aligned}\Pr_{\boldsymbol{\varphi} \sim \mathbf{b}^B_{1-\beta}}\left[|\boldsymbol{\varphi}^{-1}(0)| \ge 2\beta \cdot |B| + 3\ln n \mid \boldsymbol{\varphi} \ne 1^B\right] &= \frac{\Pr[|\boldsymbol{\varphi}^{-1}(0)| \ge 2\beta \cdot |B| + 3\ln n]}{\Pr[\boldsymbol{\varphi} \ne 1^B]} \\ &\le \frac{1/n}{1-(1-\beta)^{|B|}}\end{aligned}$$

$$\leq \frac{1/n}{1-\nu'}$$
$$\leq 2/n,$$

where the first inequality follows from the Chernoff bound (see Corollary 3.10). Therefore,

$$\Pr[|\mathbf{I} \cap B| \geq 2\beta \cdot |B| + 3\ln n] \leq 2p/n.$$

The lemma follows by a union bound over the $t \leq n$ blocks, because

$$\nu \leq (1-\beta)^{|B|} \leq \exp(-\beta \cdot |B|)$$

and therefore $\beta \cdot |B| \leq \ln(1/\nu)$. ■

6.2.2 Decomposing corrupted biased block projections

For the second step of the proof of Theorem 6.5, recall that we can think of a corrupted biased block projection as occurring in two stages: first the set of surviving variables is chosen, and then values are assigned to the non-surviving variables. For this section, we focus on the second stage: understanding the distribution over assignments.

Non-surviving variables in surviving blocks are always assigned the value 1. The more interesting case is the non-surviving blocks. We now analyze weighted sums of the assignments to variables in a collection of non-surviving blocks. The following claim decomposes any such sum to a weighted sum under a "non-corrupted" biased assignment minus a low-variance error term. We denote the number of variables here by n' and the number of blocks by t', since we will later on apply this lemma to the set of nonempty non-surviving blocks in a projection with n variables and t blocks.

Lemma 6.7 (decomposing corrupted biased block assignments)**.** *Let $w \in \mathbb{R}^{n'}$, let $[n'] = B_1 \cup \cdots \cup B_{t'}$ be a partition, let $\beta, q \in [0,1]$, and let $q_1, \ldots, q_{t'} \in [0,q]$. Sample $\mathbf{z} \in \{0,1\}^{n'}$ as follows: for each block B_j independently, we sample $\mathbf{z}_{B_j}$ from $\mathbf{corrb}^{B_j}_{1-\beta,1-q_j}$. Sample $\widetilde{\mathbf{z}} \sim \mathbf{b}^{n'}_{1-\beta}$. Let $0 \leq \nu \leq \nu' \leq 1/2$, and assume that for every j, we have $(1-\beta)^{|B_j|} \in [\nu, \nu']$. Then:*

1. *The total variation distance between $\mathbf{z}$ and $\widetilde{\mathbf{z}}$ is at most $t' \cdot \nu'$.*
2. *The sum $\sum_{i=1}^{n'} w_i \cdot \mathbf{z}_i$ is identically distributed to $(\sum_{i=1}^{n'} w_i \cdot \widetilde{\mathbf{z}}_i) - \mathbf{err}$, where the summation is over the reals, and*

$$\mathrm{Var}[\mathbf{err}] \leq O\left(\beta \cdot \nu' \cdot \log(1/\nu) \cdot \|w\|_2^2\right). \tag{6.2}$$

Note that we do not claim that **err** is independent of the $\widetilde{\mathbf{z}}_i$'s. We only argue that **err** satisfies the variance bound given by Eq. (6.2).

Proof of Lemma 6.7. The distribution of $\mathbf{z}$ is equivalent to the following two-step process: For each $j \in [t']$ independently, sample tentative values $\widetilde{\mathbf{z}}_{B_j} \sim \mathbf{b}^{B_j}_{1-\beta}$; and then, if $\widetilde{\mathbf{z}}_{B_j} = 1^{B_j}$, then with probability $1-q_j$, choose a subset of B_j by including each element independently with probability β and conditioning on the set being nonempty, and then flip the assignments of every variable in that set to 0. In other words, to sample the vector $\mathbf{z}$, for each block B_j independently, we set

$$\mathbf{z}_{B_j} = \widetilde{\mathbf{z}}_{B_j} - \mathbf{y}_j \cdot \mathbf{s}^{(j)}. \tag{6.3}$$

Here $\mathbf{y}_j$ is an indicator variable that takes value 1 with probability $1 - q_j$ if $\widetilde{\mathbf{z}}_{B_j} = 1^{B_j}$, and otherwise takes the value 0. Meanwhile, $\mathbf{s}^{(j)}$ is sampled from $\mathbf{b}_\beta^{B_j}$ conditioned on $\mathbf{s}^{(j)} \neq 0^{B_j}$, independently of $\widetilde{\mathbf{z}}_{B_j}$ and $\mathbf{y}_j$. The arithmetic in Eq. (6.3) is over the vector space $\mathbb{R}^{B_j}$.

By the union bound,

$$\Pr[\mathbf{z} \neq \widetilde{\mathbf{z}}] \leq \sum_{j=1}^{t'} \Pr[\mathbf{y}_j = 1] \leq \sum_{j=1}^{t'} (1-\beta)^{|B_j|} \leq t' \cdot \nu',$$

establishing the claimed total variation bound. Now define

$$\mathbf{x}_j = \sum_{i \in B_j} \mathbf{s}_i^{(j)} \cdot w_i$$
$$\mathbf{err} = \sum_{j=1}^{t'} \mathbf{y}_j \cdot \mathbf{x}_j.$$

By construction, the two distributions defined in the statement are identical. Our goal now is to prove the variance bound claimed in Eq. (6.2). Let $j \in [t']$ and let $\nu_j = (1-\beta)^{|B_j|} \in [\nu, \nu']$. For every distinct pair $i, i' \in B_j$, we have

$$\Pr\left[\mathbf{s}_i^{(j)} = 1\right] = \frac{\beta}{1-\nu_j}$$
$$\Pr\left[\mathbf{s}_i^{(j)} = \mathbf{s}_{i'}^{(j)} = 1\right] = \frac{\beta^2}{1-\nu_j}.$$

Consequently,

$$\begin{aligned}
\mathbb{E}[\mathbf{x}_j^2] &= \sum_{i,i' \in B_j} w_i \cdot w_{i'} \cdot \mathbb{E}\left[\mathbf{s}_i^{(j)} \cdot \mathbf{s}_{i'}^{(j)}\right] \\
&= \frac{\beta}{1-\nu_j} \cdot \sum_{i \in B_j} w_i^2 + \frac{\beta^2}{1-\nu_j} \cdot \sum_{i \neq i' \in B_j} w_i \cdot w_{i'} \\
\mathbb{E}[\mathbf{x}_j]^2 &= \left(\sum_{i \in B_j} w_i \cdot \mathbb{E}\left[\mathbf{s}_i^{(j)}\right]\right)^2 \\
&= \frac{\beta^2}{(1-\nu_j)^2} \cdot \left(\sum_{i \in B_j} w_i\right)^2 \\
&\geq \frac{\beta^2}{1-\nu_j} \cdot \left(\sum_{i \in B_j} w_i\right)^2 && \text{because } \frac{1}{1-\nu_j} \geq 1 \\
&\geq \frac{\beta^2}{1-\nu_j} \cdot \sum_{i \neq i' \in B_j} w_i \cdot w_{i'} && \text{because } \sum_{i \in B_j} w_i^2 \geq 0 \\
\implies \mathsf{Var}[\mathbf{x}_j] &\leq \frac{\beta}{1-\nu_j} \cdot \sum_{i \in B_j} w_i^2.
\end{aligned}$$

Since $\mathbf{x}_j$ and $\mathbf{y}_j$ are independent, we have

$$\begin{aligned}
\mathsf{Var}[\mathbf{y}_j \cdot \mathbf{x}_j] &= \mathbb{E}[\mathbf{y}_j^2] \cdot \mathbb{E}[\mathbf{x}_j^2] - \mathbb{E}[\mathbf{y}_j]^2 \cdot \mathbb{E}[\mathbf{x}_j]^2 \\
&= \mathbb{E}[\mathbf{y}_j^2] \cdot \mathsf{Var}[\mathbf{x}_j] + \mathsf{Var}[\mathbf{y}_j] \cdot \mathbb{E}[\mathbf{x}_j]^2 \\
&\leq \nu_j \cdot \left(\frac{\beta}{1-\nu_j} \cdot \sum_{i \in B_j} w_i^2 + \frac{\beta^2}{(1-\nu_j)^2} \cdot \left(\sum_{i \in B_j} w_i \right)^2 \right) \\
&\leq \nu_j \cdot \beta \cdot \left(\frac{1}{1-\nu_j} + \frac{\beta}{(1-\nu_j)^2} \cdot |B_j| \right) \cdot \sum_{i \in B_j} w_i^2 && \text{(Cauchy-Schwarz)} \\
&\leq \nu_j \cdot \beta \cdot \left(\frac{1}{1-\nu_j} + \frac{\ln(1/(1-\beta)^{|B_j|})}{(1-\nu_j)^2} \right) \cdot \sum_{i \in B_j} w_i^2 && \text{by Proposition 3.1} \\
&\leq \nu' \cdot \beta \cdot (2 + 4 \cdot \ln(1/\nu)) \cdot \sum_{i \in B_j} w_i^2 \\
&= O\left(\nu' \cdot \beta \cdot \log(1/\nu) \cdot \sum_{i \in B_j} w_i^2 \right).
\end{aligned}$$

Finally, the random variables $\{\mathbf{x}_j \cdot \mathbf{y}_j\}_{j \in [t']}$ are independent, so

$$\mathsf{Var}[\mathbf{err}] = \sum_{j=1}^{t'} \mathsf{Var}[\mathbf{y}_j \cdot \mathbf{x}_j] \leq O\left(\nu' \cdot \beta \cdot \log(1/\nu) \cdot \|w\|_2^2 \right). \quad \blacksquare$$

6.2.3 The structured case: Regular LTFs

Recall that we are working toward proving that LTFs simplify under corrupted biased block projections (Theorem 6.5). We first analyze the special case of *regular* LTFs.

Definition 6.8 (regular LTF). *Let $\mu > 0$. A vector $w \in \mathbb{R}^n$ is μ-*regular *if $|w_i| \leq \mu \cdot \|w\|_2$ for every $i \in [n]$. An LTF $\Phi = (w, \theta)$ is μ-regular if w is μ-regular.*

The benefit of regularity is that it allows us to use the Berry-Esseen theorem to establish anti-concentration, as we show next.

Lemma 6.9 (anti-concentration for regular linear combinations). *Let $\mu > 0$ and let $w \in \mathbb{R}^n$ be μ-regular. Let $\alpha \in (0,1)$, sample $\mathbf{z} \sim \mathbf{b}_\alpha^n$, and let $\mathbf{Z} = \sum_{i=1}^n \mathbf{z}_i \cdot w_i$. Then for any $\theta \in \mathbb{R}$ and any $R > 0$, we have*

$$\Pr\left[|\mathbf{Z} - \theta| \leq R \cdot \|w\|_2 \right] \leq \frac{\sqrt{2/\pi} \cdot R + 2\mu}{\sqrt{\alpha \cdot (1-\alpha)}}.$$

Proof. Let $\mathbf{x}_i = \mathbf{z}_i \cdot w_i - \alpha \cdot w_i$, so $\mathbf{x}_i$ has mean 0 and variance $\alpha \cdot (1-\alpha) \cdot w_i^2$. Let

$$\sigma = \sqrt{\sum_i \mathsf{Var}[\mathbf{x}_i]} = \sqrt{\sum_i \alpha(1-\alpha) \cdot w_i^2} = \sqrt{\alpha(1-\alpha)} \cdot \|w\|_2.$$

Let $\mathbf{X} = \sum_i \mathbf{x}_i / \sigma$. By the Berry-Esseen theorem, $\mathbf{X}$ behaves like a standard Gaussian (with mean zero and unit variance) up to CDF distance

$$\frac{\sum_i \mathbb{E}[|\mathbf{x}_i|^3]}{\sigma^3} = \frac{\alpha \cdot (1-\alpha) \cdot (1 - 2\alpha + 2\alpha^2) \cdot \|w\|_3^3}{\alpha^{3/2} \cdot (1-\alpha)^{3/2} \cdot \|w\|_2^3} = \frac{1 - 2\alpha + 2\alpha^2}{\sqrt{\alpha \cdot (1-\alpha)}} \cdot \frac{\|w\|_3^3}{\|w\|_2^3} < \frac{\mu}{\sqrt{\alpha \cdot (1-\alpha)}},$$

where we used the fact that $\mathbb{E}[|\mathbf{x}_i|^3] = |w_i|^3 \cdot \alpha \cdot (1-\alpha) \cdot (1 - 2\alpha + 2\alpha^2)$. Now,

$$\begin{aligned} |\mathbf{Z} - \theta| \le R \cdot \|w\|_2 &\iff \left| \sigma \cdot \mathbf{X} + \alpha \sum_{i=1}^{n} w_i - \theta \right| \le R \cdot \|w\|_2 \\ &\iff \left| \mathbf{X} - \frac{\theta - \alpha \sum_{i=1}^{n} w_i}{\sigma} \right| \le \frac{R \cdot \|w\|_2}{\sigma}. \end{aligned}$$

The probability of a standard Gaussian being in any fixed interval of length ℓ is at most $\ell/\sqrt{2\pi}$, since the standard Gaussian PDF has maximum value $1/\sqrt{2\pi}$. Therefore,

$$\begin{aligned} \Pr\left[\left| \mathbf{X} - \frac{\theta - \alpha \sum_{i=1}^{n} w_i}{\sigma} \right| \le \frac{R \cdot \|w\|_2}{\sigma} \right] &\le \frac{2R \cdot \|w\|_2}{\sigma \cdot \sqrt{2\pi}} + \frac{2\mu}{\sqrt{\alpha \cdot (1-\alpha)}} \\ &= \frac{\sqrt{2/\pi} \cdot R}{\sqrt{\alpha \cdot (1-\alpha)}} + \frac{2\mu}{\sqrt{\alpha \cdot (1-\alpha)}}. \quad \blacksquare \end{aligned}$$

Let us now show that under a corrupted biased block projection, a regular LTF is likely to become close to a constant.

Lemma 6.10 (regular LTFs simplify under corrupted biased block projections)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets of variables with $|\mathcal{Y}| = t \le n = |\mathcal{X}|$. Let $\Phi = (w, \theta)$ be a μ-regular* LTF *on $\mathcal{X}$. Let $\boldsymbol{\pi} \colon \mathcal{X} \cup \{0,1\} \to \mathcal{Y} \cup \{0,1\}$ be a p-surviving $(1-\beta)$-biased $(1-q)$-corrupted (ν, ν')-block projection, where $\beta \le 1/2$. Then for any $\epsilon > 0$ and any $\vec{\alpha} \in [0,1]^t$, the probability that $\Phi\restriction_{\boldsymbol{\pi}}$ is ϵ-far from constant under $\mathbf{b}_{\vec{\alpha}}$ is at most*

$$O\left(\frac{\mu}{\sqrt{\beta}} + (p^{1/3} + (\nu')^{1/3}) \cdot \log\left(\frac{n}{\epsilon \cdot \nu} \right) \right).$$

Proof. We may assume without loss of generality that $\nu' \le 1/2$, because otherwise the claimed failure probability is greater than 1. Let $B_1, \ldots, B_t$ be the blocks. We first consider the choice of which variables will survive the projection. We will identify three "good" events and assume by the union bound that all of them occur simultaneously.

Weight of variables in surviving blocks. Let $\mathbf{S} \subseteq [t]$ be the random variable that is the set of surviving blocks. The expected "weight" of variables in surviving blocks is

$$\mathbb{E}\left[\sum_{j \in \mathbf{S}} \sum_{i \in B_j} w_i^2 \right] = \sum_{j=1}^{t} \Pr[j \in \mathbf{S}] \cdot \sum_{i \in B_j} w_i^2 \le p \cdot \|w\|_2^2,$$

and therefore by Markov's inequality,

$$\Pr\left[\sum_{j \in \mathbf{S}} \sum_{i \in B_j} w_i^2 \le \frac{3}{4} \|w\|_2^2 \right] \ge 1 - (4/3)p. \tag{6.4}$$

(The event above is the first "good" event.)

Weight of living variables. Let $\mathbf{I} \subseteq [n]$ be the set of living variables, i.e., $\mathbf{I} = \boldsymbol{\pi}^{-1}(\mathcal{Y})$. The expected weight of variables in $\mathbf{I}$ is

$$\begin{aligned}\mathbb{E}\left[\sum_{i\in\mathbf{I}} w_i^2\right] &= \sum_{j=1}^{t} \Pr[j \in \mathbf{S}] \cdot \sum_{i\in B_j} \Pr[i \in \mathbf{I} \mid j \in \mathbf{S}] \cdot w_i^2 \\ &\leq \sum_{j=1}^{t} p \cdot \sum_{i\in B_j} \frac{\beta}{1-\nu'} \cdot w_i^2 \leq 2 \cdot p \cdot \beta \cdot \|w\|_2^2.\end{aligned}$$

Therefore, by Markov's inequality, for a parameter $R > 0$ that we will choose later,

$$\Pr\left[\sum_{i\in\mathbf{I}} w_i^2 \leq 2 \cdot R \cdot p \cdot \beta \cdot \|w\|_2^2\right] \geq 1 - 1/R. \tag{6.5}$$

(The event above is the second "good" event.)

Number of living variables per block. By Lemma 6.6,

$$\Pr\left[\max_{j\in[t]} |\mathbf{I} \cap B_j| \leq 3\ln(n/\nu)\right] \geq 1 - O(p + \nu'). \tag{6.6}$$

(The event above is the third "good" event.)

Condition on $\mathbf{S} = S$ and $\mathbf{I} = I$, where S and I are arbitrary sets for which the three good events above all occur. Recall that in surviving blocks, variables that are not alive are assigned the value 1. We now consider the choice of assignment to the variables in the non-surviving blocks.

Let $K = [n] \setminus (I \cup J)$ where J is the set of fixed variables in surviving blocks. We consider the choice of assignment to variables in K, and for each $i \in K$ let $\mathbf{z}_i = \boldsymbol{\pi}(x_i) \in \{0,1\}$. By Lemma 6.7, the random variable $\sum_{i\in K} w_i \cdot \mathbf{z}_i$ is distributed identically to the random variable

$$\left(\sum_{i\in K} w_i \cdot \widetilde{\mathbf{z}}_i\right) - \mathbf{err}\,,$$

where $\widetilde{\mathbf{z}} \sim \mathbf{b}_{1-\beta}^K$ and $\mathbf{err}$ has bounded variance as asserted in Lemma 6.7. (Recall that some variables in K might not belong in blocks; for such variables we simply have that $\mathbf{z}_i = \widetilde{\mathbf{z}}_i$, whereas the error term $\mathbf{err}$ is only due to the variables in blocks.)

Without loss of generality, we may assume that the two random variables are coupled in such a way that they are always equal (not just identically distributed). Let $\eta = \mathbb{E}[\mathbf{err}]$. We will identify two more "good" events and assume by the union bound that both occur, and then we will argue that in this case, the projected function is unbalanced.

Anti-concentration of the product distribution $\widetilde{\mathbf{z}}$. Define $w' \in \mathbb{R}^S$ by $w'_j = \sum_{i\in I\cap B_j} w_i$. Unpacking all the definitions, the projected function $\Phi\restriction_{\boldsymbol{\pi}} : \{0,1\}^{\mathcal{Y}} \to \{0,1\}$ is given by

$$\begin{aligned}\Phi\restriction_{\boldsymbol{\pi}}(y) = 1 &\iff \sum_{j\in S} y_j \cdot w'_j + \sum_{i\in J} w_i + \sum_{i\in K} \mathbf{z}_i \cdot w_i \geq \theta \\ &\iff \sum_{i\in K} \mathbf{z}_i \cdot w_i \geq \theta - \sum_{j\in S} y_j \cdot w'_j - \sum_{i\in J} w_i\,.\end{aligned}$$

Recall that we are ultimately interested in plugging in a value y sampled from the product distribution $\mathbf{b}_{\vec{\alpha}}$. With respect to such a choice, the expected value for the RHS above is

$$\tau \stackrel{\text{def}}{=\!=} \theta - \sum_{j \in S} \alpha_j \cdot w'_j - \sum_{i \in J} w_i \ .$$

The bad case is that $\sum_{i \in K} \mathbf{z}_i \cdot w_i \approx \tau$; in such a case, the projected function is somewhat balanced. Let us argue that this bad event has low probability, and as a first step we argue that the bad event $\sum_{i \in K} \widetilde{\mathbf{z}}_i \cdot w_i \approx \tau + \eta$ has low probability. (Recall that $\eta = \mathbb{E}[\mathbf{err}]$.)

Since the good event in Eq. (6.4) has occurred, we have $\|w_K\|_2 \geq \frac{1}{2} \cdot \|w\|_2$, and therefore the vector w_K is (2μ)-regular (because for each $i \in K$ we have that $|w_i| \leq \mu \|w\|_2 \leq 2\mu \|w_K\|_2$). Therefore, by Lemma 6.9, for a parameter R' that we will choose later,

$$\Pr\left[\left|\sum_{i \in K} \widetilde{\mathbf{z}}_i \cdot w_i - (\tau + \eta)\right| \leq R' \cdot \|w_K\|_2\right] \leq \frac{\sqrt{2/\pi} \cdot R' + 4\mu}{\sqrt{\beta(1-\beta)}} \ ,$$

and using the facts that $\|w_K\|_2 \geq \frac{1}{2} \cdot \|w\|_2$ and $\beta \leq 1/2$ we get that

$$\Pr\left[\left|\sum_{i \in K} \widetilde{\mathbf{z}}_i \cdot w_i - (\tau + \eta)\right| > (R'/2) \cdot \|w\|_2\right] \geq 1 - \frac{2/\sqrt{\pi} \cdot R' + 4\sqrt{2}\mu}{\sqrt{\beta}} \ . \tag{6.7}$$

(The event above is another good event.)

Concentration of the error term. By Lemma 6.7 and Chebyshev's inequality,

$$\begin{aligned} \Pr[|\mathbf{err} - \eta| \leq (R'/4) \cdot \|w\|_2] &\geq 1 - \frac{\mathsf{Var}[\mathbf{err}]}{(R'/4)^2 \cdot \|w\|_2^2} \\ &\geq 1 - O\left(\frac{\nu' \cdot \log(1/\nu) \cdot \beta}{(R')^2}\right) \ . \end{aligned} \tag{6.8}$$

(The above is our last good event.) Condition on $\mathbf{z} = z$, $\widetilde{\mathbf{z}} = \widetilde{z}$, and $\mathbf{err} = \mathrm{err}$, where $\widetilde{z}$ and err are such that the two preceding good events occur.

Imbalance of the projected function. Let

$$\theta' = \theta - \sum_{i \in K} z_i \cdot w_i - \sum_{i \in J} w_i$$

such that

$$\Phi\restriction_{\pi}(y) = 1 \iff \sum_{j \in S} y_j \cdot w'_j \geq \theta' \ .$$

Then, unpacking the definitions and relying on the fact that the good events in Eq. (6.7) and Eq. (6.8) have occurred, we have

$$\begin{aligned} \left|\theta' - \left(\sum_{j \in S} \alpha_j \cdot w'_j\right)\right| = \left|\sum_{i \in K} z_i \cdot w_i - \tau\right| &= \left|\left(\sum_{i \in K} w_i \cdot \widetilde{z}_i\right) - \mathrm{err} - \tau\right| \\ &\geq (R'/4) \cdot \|w\|_2 \ . \end{aligned}$$

Furthermore, since the good events of Eq. (6.5) and Eq. (6.6) occurred, we have that

$$\begin{aligned}\|w'\|_2^2 &= \sum_{j\in S}\left(\sum_{i\in I\cap S} w_i\right)^2 \\ &\le \sum_{j\in S}|I\cap S|\cdot\sum_{i\in I\cap S} w_i^2 && \text{(Cauchy-Schwartz)} \\ &\le 3\ln(n/\nu))\cdot\sum_{i\in I} w_i^2 \\ &\le 6\ln(n/\nu)\cdot R\cdot p\cdot\beta\cdot\|w\|_2^2 \ .\end{aligned}$$

Therefore,

$$\left|\theta' - \left(\sum_{j\in S}\alpha_j\cdot w'_j\right)\right| \ge \frac{R'}{\sqrt{6\cdot\ln(n/\nu)\cdot R\cdot p\cdot\beta}}\cdot\|w'\|_2 \ .$$

Let

$$R' = \frac{1}{2}\cdot\sqrt{6\cdot R\cdot p\cdot\beta\cdot\ln(n/\nu)\cdot\ln(2/\epsilon)} + (\nu'\cdot\log(1/\nu))^{1/3}\cdot\sqrt{\beta}\ ,$$

and note that the first term ensures that R' is large enough so that we have

$$\left|\theta' - \left(\sum_{j\in S}\alpha_j\cdot w'_j\right)\right| \ge \frac{1}{2}\cdot\sqrt{\ln(2/\epsilon)}\cdot\|w'\|_2 \ ,$$

and therefore, by Corollary 3.14, the projected function $\Phi\restriction_\pi$ is ϵ-close to constant under $\vec{\alpha}$, as claimed.

Failure probability. Summing up over all five good events, the total failure probability is bounded by

$$O\left(p + 1/R + \nu' + \frac{R'+\mu}{\sqrt{\beta}} + \frac{\nu'\cdot\log(1/\nu)\cdot\beta}{(R')^2}\right) \ ,$$

and plugging in our value for R' this becomes

$$O\left(p + 1/R + \nu' + \sqrt{R\cdot p\cdot\log(n/\nu)\cdot\log(1/\epsilon)} + (\nu'\cdot\log(1/\nu))^{1/3} + \frac{\mu}{\sqrt{\beta}}\right) \ .$$

Now, choosing

$$R = (p\cdot\log(n/\nu)\cdot\log(1/\epsilon))^{-1/3} \ ,$$

the total failure probability is

$$O\left(p^{1/3}\cdot\log^{1/3}(n/\nu)\cdot\log^{1/3}(1/\epsilon) + (\nu'\cdot\log(1/\nu))^{1/3} + \frac{\mu}{\sqrt{\beta}}\right) \ ,$$

and the simpler bound in the lemma's statement follows immediately. ∎

6.2.4 The general case: Arbitrary LTFs

At this point, we are ready to analyze an arbitrary (not-necessarily-regular) LTF. The analysis relies on the concept of a *critical index*.

Definition 6.11. *Let Φ be an* LTF*, say $\Phi(x) = 1 \iff \sum_{i=1}^{n} w_i \cdot x_i \geq \theta$, where the variables are ordered in descending order of weight ($|w_1| \geq |w_2| \geq \cdots \geq |w_n|$). For $\mu > 0$, the* μ-critical index *of Φ is the smallest $i \in [n]$ such that $w_{>i}$ is μ-regular, or ∞ if no such i exists.*

Proof of Theorem 6.5. Assume without loss of generality that the variables are ordered in descending order of weight ($|w_1| \geq |w_2| \geq \cdots \geq |w_n|$). Let the blocks be $B_1, \ldots, B_t$. Let k be a parameter that we will choose later, and let $\mathsf{H} \subseteq [t]$ be the set of indices j such that $B_j \cap [k] \neq \emptyset$, so in particular $|\mathsf{H}| \leq k$.

We consider a modified projection $\boldsymbol{\pi}'$ in which all the variables in blocks in H are outside blocks (and hence get assigned uniform values from $\mathbf{b}_{1-\beta}$). We claim that the total variation distance between $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ is at most $(p + \nu') \cdot k$. To see this, recall that each of the $\leq k$ blocks in H survives with probability at most p under $\boldsymbol{\pi}$, and assuming that none of them survives, Lemma 6.7 guarantees that the total variation distance is at most $k \cdot \nu'$. Thus, for the remainder of the proof we will analyze $\boldsymbol{\pi}'$ instead of $\boldsymbol{\pi}$.

Let $\mu > 0$ be another parameter that we will choose later. Denoting the μ-critical index of Φ by κ, the proof proceeds by a case analysis.

Case 1 (large head): $\kappa > k$. Let $\mathbf{I} \subseteq [n]$ be the set of surviving variables. By Lemma 6.6, except with probability $O(p + \nu')$, for every block B, we have $|\mathbf{I} \cap B| \leq 3\ln(n/\nu)$. Condition on $\mathbf{I} = I$, where I is an arbitrary set such that this good event occurs, and let $J = [n] \setminus I$. For $i \in J$, let $\mathbf{z}_i$ denote the assignment to the i^{th} variable. Let $R \subseteq [k]$ be a set of indices that we will choose later. By the definitions of H and $\boldsymbol{\pi}'$, we have $R \subseteq J$. Condition on any assignment $z_{J \setminus R} = \mathbf{z}_{J \setminus R}$ to the variables outside R.

Like in the proof of Lemma 6.10, let $S \subseteq [t]$ be the set of surviving blocks, and define $w' \in \mathbb{R}^S$ by $w'_j = \sum_{i \in I \cap B_j} w_i$. That way, the projected LTF $\Phi\restriction_{\boldsymbol{\pi}}$ is given by

$$\Phi\restriction_{\boldsymbol{\pi}}(y) = 1 \iff \sum_{j \in S} y_j \cdot w'_j + \sum_{i \in J \setminus R} z_i \cdot w_i + \sum_{i \in R} \mathbf{z}_i \cdot w_i \geq \theta.$$

As a reminder, we are ultimately interested in plugging in a value y sampled from the product distribution $\mathbf{b}_{\tilde{\alpha}}$. With respect to that choice, the expected value for the linear combination above is

$$\sum_{j \in S} \alpha_j \cdot w'_j + \sum_{i \in J \setminus R} z_i \cdot w_i + \sum_{i \in R} \mathbf{z}_i \cdot w_i.$$

Therefore, define

$$\tau = \theta - \sum_{j \in S} \alpha_j \cdot w'_j - \sum_{i \in J \setminus R} z_i \cdot w_i.$$

The bad case is that $\sum_{i \in R} \mathbf{z}_i \cdot w_i \approx \tau$; in such a case, the projected function is somewhat balanced. Let us show that this bad case occurs with low probability.

A lemma by Servedio [Ser07, Lemma 4.5] (together with the bound $1 - \mu^2 \leq \exp(-\mu^2)$) guarantees that for any $1 \leq i < j \leq \kappa + 1$,

$$|w_j| \leq \|w_{\geq j}\|_2 \leq \exp(-\mu^2 \cdot (j-i)/2) \cdot \|w_{\geq i}\|_2 \leq \mu^{-1} \cdot \exp(-\mu^2 \cdot (j-i)/2) \cdot |w_i|. \tag{6.9}$$

Define

$$\Delta = \lceil 2 \cdot \ln(3/\mu) \cdot \mu^{-2} \rceil,$$

and let $i_1 = 1$ and $i_{q+1} = i_q + \Delta$. That way, if $i_{q+1} \leq \kappa + 1$, then Equation (6.9) implies that $|w_{i_{q+1}}| \leq |w_{i_q}|/3$. Define

$$\begin{aligned}
r &= \lceil (1/\beta) \cdot \ln(1/\mu) \rceil \\
\ell &= \frac{1}{2}\sqrt{3\ln(n/\nu) \cdot \ln(2/\epsilon)} \\
k &= i_r + \lceil 2\ln(4\ell/\mu) \cdot \mu^{-2} \rceil = O\left(\frac{\beta^{-1} \cdot \log^2(1/\mu) + \log\log(\frac{n}{\epsilon \cdot \nu})}{\mu^2} \right) \\
R &= \{i_1, \ldots, i_r\} \subseteq [k].
\end{aligned}$$

That way, since $I \cap [k] = \emptyset$ and $\kappa > k$, Equation (6.9) implies that

$$\|w_I\|_2 \leq \|w_{\geq k}\|_2 \leq \mu^{-1} \exp(-\mu^2 \cdot (k - i_r)/2) \cdot |w_{i_r}| \leq |w_{i_r}|/(4\ell).$$

Furthermore, by the Cauchy-Schwarz inequality,

$$\|w'\|_2^2 \leq \sum_{j=1}^{t} |I \cap B_j| \cdot \sum_{i \in I \cap B_j} w_i^2 \leq 3\ln(n/\nu) \cdot \|w_I\|_2^2$$

and thus

$$\|w'\|_2 \leq \frac{2\ell}{\sqrt{\ln(2/\epsilon)}} \cdot \|w_I\|_2.$$

Therefore, for any two distinct assignments $z_R, z'_R \in \{0,1\}^r$, denoting by $i_* \in R$ the smallest index such that $z_{i_*} \neq z'_{i_*}$, we have

$$\begin{aligned}
\left| \langle z_R, w_R \rangle - \langle z'_R, w_R \rangle \right| &= \left| \sum_{\substack{i \in R \\ i \geq i_*}} w_i \cdot (z_i - z'_i) \right| \\
&\geq |w_{i_*} \cdot (z_{i_*} - z'_{i_*})| - \sum_{i \in R \setminus [i_*]} |w_i \cdot (z_i - z'_i)| \\
&\geq |w_{i_*}| - \sum_{i \in R \setminus [i_*]} |w_i| \\
&\geq |w_{i_*}| - \sum_{q=1}^{\infty} |w_{i_*}|/3^q \\
&= |w_{i_*}|/2 \\
&\geq |w_{i_r}|/2 \\
&\geq 2\ell \cdot \|w_I\|_2 \\
&\geq \sqrt{\ln(2/\epsilon)} \cdot \|w'\|_2.
\end{aligned}$$

It follows that there is at most a single assignment $z_R \in \{0,1\}^r$ such that

$$\left| \sum_{i \in R} z_i \cdot w_i - \tau \right| \leq \frac{1}{2}\sqrt{\ln(2/\epsilon)} \cdot \|w'\|_2.$$

Consequently, since we assumed that the variables in R are all outside the blocks,

$$\Pr_{\mathbf{z}_R}\left[\left|\sum_{i\in R}\mathbf{z}_i\cdot w_i-\tau\right|\geq\frac{1}{2}\sqrt{\ln(2/\epsilon)}\cdot\|w'\|_2\right]\geq 1-(1-\beta)^r\geq 1-\mu.$$

Assume that the above good event occurs. In this case, by Corollary 3.14, the projected function is indeed ϵ-close to constant.

Case 2 (small head): $\kappa\leq k$**.** Condition on any assignment to the first κ variables. We claim that the action of the projection $\boldsymbol{\pi}'$ on the remaining $n-\kappa$ variables is another p-surviving $(1-\beta)$-biased $(1-q)$-corrupted (ν,ν')-block projection. Indeed, recall that we defined $\boldsymbol{\pi}'$ in such a way that that the first $k\geq\kappa$ variables are outside blocks. In the definition of a p-surviving $(1-\beta)$-biased corrupted (ν,ν')-block projection, variables outside blocks are assigned values from $\mathbf{b}_{1-\beta}$ *independently* of the action of the projection on the other variables. By the definition of κ, the residual LTF to which we are applying this projection is μ-regular. By Lemma 6.10, the probability that the residual LTF is ϵ-far from constant under $\mathbf{b}_{\vec{\alpha}}$ is at most

$$O\left(\frac{\mu}{\sqrt{\beta}}+(p^{1/3}+(\nu'')^{1/3})\cdot\log\left(\frac{n-\kappa}{\epsilon\cdot\nu}\right)\right).$$

The overall failure probability. Including the total variation distance between $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ and the failure probabilities in the two cases, the total failure probability is at most

$$O\left((p+\nu')\cdot k+\frac{\mu}{\sqrt{\beta}}+(p^{1/3}+(\nu')^{1/3})\cdot\log\left(\frac{n}{\epsilon\cdot\nu}\right)\right).$$

Choose

$$\mu=\frac{(p^{1/3}+(\nu')^{1/3})\cdot\log(1/\nu)}{\beta^{1/6}}.$$

Then plugging in our choice of k and using $\log^2(1/\mu)\leq O(\log(1/\nu))$, the total failure probability is bounded by

$$\begin{aligned}&O\left(\frac{(p+\nu')\cdot(\beta^{-1}\cdot\log^2(1/\mu)+\log\log(\frac{n}{\epsilon\cdot\nu}))}{\mu^2}+\frac{\mu}{\sqrt{\beta}}+(p^{1/3}+(\nu')^{1/3})\cdot\log\left(\frac{n}{\epsilon\cdot\nu}\right)\right)\\&\leq O\left((p^{1/3}+(\nu')^{1/3})\cdot\beta^{-2/3}\cdot\log\left(\frac{n}{\epsilon\cdot\nu}\right)\right)\end{aligned}$$

as claimed. ∎

6.3 LTF circuits simplify under corrupted biased block projections

Next, we would like to show that corrupted biased block projections decrease the depth of an LTF circuit. More accurately, we will show that under such a projection, a depth-Δ LTF circuit can be approximated by a decision tree with depth-$(\Delta-1)$ LTF circuits at the leaves. In the following proposition, we assume that $\nu'\leq p$; this assumptions can be removed, but it is helpful for simplifying the bounds.

Proposition 6.12 (LTF circuits simplify under $\boldsymbol{\pi}$)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets of variables with $|\mathcal{X}| = n \geq t = |\mathcal{Y}|$. Let $\Delta \geq 1$, and let f be a depth-Δ LTF circuit on $\mathcal{X}$ with w wires. Let $\boldsymbol{\pi} \colon \mathcal{X} \cup \{0,1\} \to \mathcal{Y} \cup \{0,1\}$ be a p-surviving $(1-\beta)$-biased $(1-q)$-corrupted (ν,ν')-block projection. Assume $\beta \leq 1/2$ and $\nu' \leq p$. Then for every $\vec{\alpha} \in [0,1]^t$ and every $\epsilon > 0$, except with probability*

$$p^{1/12} \cdot \beta^{-1/6} \cdot \operatorname{polylog}\left(\frac{n \cdot w}{\epsilon \cdot \nu}\right),$$

the projected function $f\restriction_{\boldsymbol{\pi}}$ can be approximated under $\mathbf{b}_{\vec{\alpha}}$ with error ϵ by a depth-D decision tree whose leaves are labeled by depth-$(\Delta-1)$ LTF circuits with w wires, where

$$D \leq p^{13/12} \cdot w \cdot \beta^{-1/6} \cdot \operatorname{polylog}\left(\frac{n \cdot w}{\epsilon \cdot \nu}\right).$$

The key point is that in the bound on D, the exponent of p is greater than 1. For comparison, the number of remaining variables after applying $\boldsymbol{\pi}$ is roughly $p \cdot \beta \cdot n$. When we eventually analyze the projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$ that we defined in Section 4.2, we will have $p = \beta^C$ for a large constant C. Thus, when w is only slightly larger than n, the $p^{13/12}$ term in the bound on D will ensure that D is significantly smaller than $p \cdot \beta \cdot n$ and hence the tree is nontrivial.

Proof. Let ζ be the failure probability from Theorem 6.5 with approximation error ϵ/w, so (recalling $\nu \leq \nu' \leq p$) we have

$$\zeta = O\left(p^{1/3} \cdot \beta^{-2/3} \cdot \log\left(\frac{n \cdot w}{\epsilon \cdot \nu}\right)\right).$$

Let G be the set of gates Φ in f such that every input to Φ is a variable. (Informally, G is the "bottom layer" of f, but formally we don't require f to be layered.) Partition $G = G_H \cup G_L$, where G_L ("light gates") is the set of gates with fan-in at most $\sqrt{\zeta} \cdot p^{-1}$ and $G_H = G \setminus G_L$ ("heavy gates"). Let $\mathbf{A}$ be the set of gates $\Phi \in G_H$ such that $\Phi\restriction_{\boldsymbol{\pi}}$ is (ϵ/w)-far from constant under $\mathbf{b}_{\vec{\alpha}}$. For each gate $\Phi \in G_H \setminus \mathbf{A}$, let $\mathbf{c}_\Phi \in \{0,1\}$ be the constant toward which $\Phi\restriction_{\boldsymbol{\pi}}$ is biased under $\mathbf{b}_{\vec{\alpha}}$.

Let the blocks be $B_1, \ldots, B_t$. We say that two distinct blocks B_i, B_j are a *connected pair* if there exists a light gate $\Phi \in G_L$ such that a variable from B_i feeds into Φ and a variable from B_j feeds into Φ. We say that the connected pair *survives* if B_i and B_j are both surviving blocks.

The tree queries the surviving connected pairs and the surviving blocks feeding into $\mathbf{A}$. The node reached at that point is a leaf, labeled by the circuit obtained from $f\restriction_{\boldsymbol{\pi}}$ by (a) plugging in the value of each queried variable, and (b) replacing each gate $\Phi \in G_H \setminus \mathbf{A}$ with the corresponding constant $\mathbf{c}_\Phi$. This means that every gate in G has zero or one variables feeding into it, so (possibly after some simplification) the circuit has depth $\Delta - 1$.

For any input $y \in \{0,1\}^t$ to this tree, if for every $\Phi \in G_H \setminus \mathbf{A}$ we have $\Phi(y) = \mathbf{c}_\Phi$, then the tree correctly computes $f\restriction_{\boldsymbol{\pi}}(y)$. Therefore, by the union bound, the tree correctly computes $f\restriction_{\boldsymbol{\pi}}(\mathbf{y})$ with probability $1 - \epsilon$ over the choice of $\mathbf{y} \sim \mathbf{b}_{\vec{\alpha}}$. Now let us bound the depth of the tree (with high probability with respect to the random choice of projection $\boldsymbol{\pi}$).

Light gates. The probability that any particular connected pair survives is at most p^2. The number of connected pairs is at most $\sqrt{\zeta} \cdot p^{-1} \cdot w$, since each wire participates in at most $\sqrt{\zeta} \cdot p^{-1}$ connected pairs. Therefore, in expectation, the number of connected pairs that survive is at most $p^2 \cdot \sqrt{\zeta} \cdot p^{-1} \cdot w = O(\sqrt{\zeta} \cdot p \cdot w)$. By Markov's inequality, except with probability $\zeta^{1/4}$, the total number of connected pairs that survive is at most $O(\zeta^{1/4} \cdot p \cdot w)$.

Heavy gates. By the Chernoff bound (see Corollary 3.10), for any particular gate Φ, the probability that more than $2p \cdot \text{fan-in}(\Phi) + 3\ln(w/\zeta)$ blocks with a variable feeding into Φ survive the projection is at most ζ/w. Therefore, by the union bound, with probability $1-\zeta$, every gate Φ has at most $O(p \cdot \text{fan-in}(\Phi) + \log(w/\zeta))$ surviving blocks feeding into it. Assume that this is the case.

For each gate Φ, we have $\Pr[\Phi \in \mathbf{A}] \leq \zeta$. Partition $G_H = G_H^{(1)} \cup \cdots \cup G_H^{(\log w)}$, where $G_H^{(i)} = \{\Phi \in G_H : \text{fan-in}(\Phi) \in [2^{i-1}, 2^i)\}$. Fix some $i \in [\log w]$. We have $\mathbb{E}\left[\left|\mathbf{A} \cap G_H^{(i)}\right|\right] \leq \zeta \cdot \left|G_H^{(i)}\right|$. Therefore, by Markov's inequality,

$$\Pr\left[\left|\mathbf{A} \cap G_H^{(i)}\right| > \zeta^{3/4} \cdot \left|G_H^{(i)}\right|\right] \leq \zeta^{1/4}.$$

By the union bound, with probability $1 - \zeta^{1/4} \cdot \log w$, for every i, we have $\left|\mathbf{A} \cap G_H^{(i)}\right| \leq \zeta^{3/4} \cdot \left|G_H^{(i)}\right|$. Assume that this is the case. Then the number of surviving blocks feeding into heavy gates in $\mathbf{A}$ is bounded by

$$\begin{aligned} O\left(\sum_{i=1}^{\log w} \left|\mathbf{A} \cap G_H^{(i)}\right| \cdot (p \cdot 2^i + \log(w/\zeta))\right) &\leq O\left(\sum_{i=1}^{\log w} \zeta^{3/4} \cdot \left|G_H^{(i)}\right| \cdot (p \cdot 2^i + \log(w/\zeta))\right) \\ &\leq O\left(\zeta^{3/4} \cdot p \cdot w + \zeta^{3/4} \cdot |G_H| \cdot \log(w/\zeta)\right) \\ &= O(\zeta^{1/4} \cdot p \cdot w \cdot \log(w/\zeta)), \end{aligned}$$

where the last line uses the fact that $|G_H| \leq w \cdot p/\sqrt{\zeta}$. Summing up, the total failure probability is $\zeta^{1/4} + \zeta + \zeta^{1/4} \cdot \log w = O(\zeta^{1/4} \cdot \log w)$, and the total number of queries is $O(\zeta^{1/4} \cdot p \cdot w + \zeta^{1/4} \cdot p \cdot w \cdot \log(w/\zeta)) = O(\zeta^{1/4} \cdot p \cdot w \cdot \log(w/\zeta))$. ■

7 Decision trees with LTF circuits at their leaves simplify under the projections

7.1 Bounding the number of survivors that a decision tree can find

In the previous section, we showed that under a corrupted biased block projection, a depth-d LTF circuit becomes a decision tree T with depth-$(d-1)$ LTF circuits at the leaves. To make further progress, we would like to analyze the effect of an additional corrupted biased block projection $\boldsymbol{\pi}$ on such a decision tree. We will eventually argue that the tree simplifies further, being approximated by a decision tree T' with depth-$(d-2)$ LTF circuits at the leaves.

As discussed in Section 2.4, T' will operate in two phases. In the first phase, it will make queries to simulate the tree portion of T until it reaches a leaf, labeled by a depth-$(d-1)$ circuit C. In the second phase, T' will simulate C using a decision tree with leaves labeled by depth-$(d-2)$ circuits, just like in the previous section. This strategy is natural enough, but there are some challenges, because variables are not all independent, so the distributions get distorted when we condition on reaching some vertex in T. In particular, the main challenge in the first phase will be to bound the number of queries. The main challenge in the second phase will be to bound the probability of error.

Our goal in this section is to prove the following lemma, which is the key to analyzing the first phase. In the first phase, T' must query y_j if the simulation of T queries some surviving

variable in the block B_j.[17] Therefore, we are interested in bounding the number of blocks in which a decision tree T can find a surviving variable. The following lemma accomplishes that (and a little bit more due to the second bullet point).

Lemma 7.1 (Shallow decision trees cannot find many survivors)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets of variables with $|\mathcal{X}| = n \geq t = |\mathcal{Y}|$. Let $\boldsymbol{\pi} \colon \mathcal{X} \cup \{0,1\} \to \mathcal{Y} \cup \{0,1\}$ be a p-surviving $(1-\beta)$-biased $(1-q)$-corrupted (ν,ν')-block projection with blocks $B_1, \ldots, B_t$, where $p \leq 1/2$, $\beta \leq 1/2$, and $\nu' \leq 1/8$. Let T be a depth-D decision tree on $\mathcal{X}$, let $y \in \{0,1\}^{\mathcal{Y}}$, and let[18] $\mathbf{S}$ be the number of blocks $j \in [t]$ such that either*

- *$T(y \circ \boldsymbol{\pi})$ queries some surviving variable $x_i \in B_j$, or*
- *$T(y \circ \boldsymbol{\pi})$ makes at least $1/\beta$ queries to variables in B_j and B_j survives.*

Then for every $\zeta \in (0, 1/2)$,

$$\Pr[\mathbf{S} \leq O(p \cdot \beta \cdot D + \log(1/\zeta))] \geq 1 - \zeta.$$

Note that the standard model of a decision tree has Boolean output labels on its leaves. However, the output of T plays no role in Lemma 7.1. (In particular, the expression "$T(y \circ \boldsymbol{\pi})$" should be interpreted to mean "the computation of T on $y \circ \boldsymbol{\pi}$" rather than "the output of T on $y \circ \boldsymbol{\pi}$.") We can therefore consider the leaves of T to be unlabeled; it might be helpful to think of T as outputting the identity of the leaf that it reaches.

The intuition behind Lemma 7.1 is as follows.

- Suppose that T makes fewer than $1/\beta$ queries to some blocks. Each of these blocks survives with probability p. Within each surviving block, each queried variable survives with probability $O(\beta)$, and these events are nearly independent. (They are not perfectly independent because we condition on a non-empty set of surviving variables in the block, but the block has considerably more than $1/\beta$ variables, so the conditioning should have little impact.) These queries should therefore contribute $O(p \cdot \beta \cdot D)$ to $\mathbf{S}$.
- Meanwhile, suppose that T makes more than $1/\beta$ queries to a block. Given its budget of D total queries, T can only afford to make this many queries to $\beta \cdot D$ distinct blocks. Each of these blocks survives with probability at most p, so once again, these events should only contribute $O(p \cdot \beta \cdot D)$ to $\mathbf{S}$.

Formalizing this intuition is not completely straightforward, in part because T is allowed to be adaptive. Nevertheless, Lemma 7.1 is true; the rigorous proof follows.

Proof. For each vertex v in T, let $\mathrm{var}(v) \in \mathcal{X}$ be the variable that v queries. Furthermore, let $\mathrm{prev}(v)$ be the set of variables $x_i \in \mathcal{X}$ such that (a) there is some proper ancestor u of v with $\mathrm{var}(u) = x_i$, and (b) there is some block B_j that contains both x_i and $\mathrm{var}(v)$. Let $\mathbf{v}_1, \ldots, \mathbf{v}_D$ be the random sequence of internal vertices that $T(y \circ \boldsymbol{\pi})$ visits.[19]

Let $\boldsymbol{\rho} \in \{0,1,\star\}^{\mathcal{X}}$ be the restriction corresponding to $\boldsymbol{\pi}$, i.e., $\boldsymbol{\rho}_i = \star$ if x_i survives and $\boldsymbol{\rho}_i = \boldsymbol{\pi}(x_i) \in \{0,1\}$ otherwise. Furthermore, define $\mathbf{r} \in \{0,1,\star\}^D$ by letting $\mathbf{r}_k = \boldsymbol{\rho}_{\mathrm{var}(\mathbf{v}_k)}$. Since y is fixed, the bit $(y \circ \boldsymbol{\pi})_i$ is determined by $\boldsymbol{\rho}_i$. Therefore, the values $\mathbf{r}_1, \ldots, \mathbf{r}_{k-1}$ determine the vertices $\mathbf{v}_1, \ldots, \mathbf{v}_k$.

Let $R = \lceil 1/\beta \rceil$. Write $\mathbf{S} = |\mathbf{Q}_1| + |\mathbf{Q}_2|$, where

[17] Actually, the construction of T' involves some mild preprocessing of T, leading to a few extra queries, but they are handled by the second bullet point in Lemma 7.1.

[18] "$\mathbf{S}$" for "survivors."

[19] We may assume that every root-to-leaf path in T has length precisely D without loss of generality.

- $\mathbf{Q}_1$ is the set of $k \in [D]$ such that query k is to a surviving variable ($\mathbf{r}_k = \star$), query k is among the first $R-1$ queries to its block (i.e., $|\mathrm{prev}(\mathbf{v}_k)| \le R-2$), and each previous query to that block was to a non-surviving variable (i.e., $\boldsymbol{\rho}_{\mathrm{prev}(\mathbf{v}_k)} \in \{0,1\}^{\mathrm{prev}(\mathbf{v}_k)}$).
- $\mathbf{Q}_2$ is the set of $k \in [D]$ such that query k is to a variable in a surviving block, query k is the R^{th} query to its block (i.e., $|\mathrm{prev}(\mathbf{v}_k)| = R-1$), and each previous query to that block was to a non-surviving variable (i.e., $\boldsymbol{\rho}_{\mathrm{prev}(\mathbf{v}_k)} \in \{0,1\}^{\mathrm{prev}(\mathbf{v}_k)}$).

Each of the above definitions requires $\boldsymbol{\rho}_{\mathrm{prev}(\mathbf{v}_k)} \in \{0,1\}^{\mathrm{prev}(\mathbf{v}_k)}$. In fact, for each $k \in \mathbf{Q}_1 \cup \mathbf{Q}_2$, we have $\boldsymbol{\rho}_{\mathrm{prev}(\mathbf{v}_k)} = 1^{\mathrm{prev}(\mathbf{v}_k)}$, because in a surviving block, all non-surviving variables are assigned the value 1.

We will show that

$$\Pr[|\mathbf{Q}_1| \le O(p \cdot \beta \cdot D + \log(1/\zeta))] \ge 1 - \zeta/2 \tag{7.1}$$

and similarly

$$\Pr[|\mathbf{Q}_2| \le O(p \cdot \beta \cdot D + \log(1/\zeta))] \ge 1 - \zeta/2, \tag{7.2}$$

which will complete the proof.

Claim 7.1.1. *Equation (7.1) holds.*

Proof. For brevity, we write $\mathbf{r}_{1\ldots k-1}$ to denote $(\mathbf{r}_1, \ldots, \mathbf{r}_{k-1})$. Fix some $k \in [D]$ and a string $r \in \mathrm{Supp}(\mathbf{r}_{1\ldots k-1}) \subseteq \{0,1,\star\}^{k-1}$. Let us bound the conditional probability $\Pr[k \in \mathbf{Q}_1 \mid \mathbf{r}_{1\ldots k-1} = r]$. As mentioned previously, the string $\mathbf{r}_{1\ldots k-1}$ determines $\mathbf{v}_1, \ldots, \mathbf{v}_k$; let $v_1, \ldots, v_k$ be the vertices such that

$$\mathbf{r}_{1\ldots k-1} = r \implies \mathbf{v}_1 = v_1 \wedge \cdots \wedge \mathbf{v}_k = v_k.$$

Then

$$\Pr[k \in \mathbf{Q}_1 \mid \mathbf{r}_{1\ldots k-1} = r] = \Pr[k \in \mathbf{Q}_1 \mid \boldsymbol{\rho}_{\mathrm{var}(v_1)} = r_1 \wedge \cdots \wedge \boldsymbol{\rho}_{\mathrm{var}(v_{k-1})} = r_{k-1}].$$

Assume that the conditional probability above is nonzero. Then $\mathrm{var}(v_k)$ is in some block B_j, and for each $k' < k$, if $\mathrm{var}(v_{k'}) \in B_j$, then $r_{k'} = 1$ (as discussed after the definitions of $\mathbf{Q}_1$ and $\mathbf{Q}_2$). Furthermore, $|\mathrm{prev}(v_k)| \le R-2$. The only remaining requirement in the definition of $\mathbf{Q}_1$ is that query k is to a surviving variable, so

$$\Pr[k \in \mathbf{Q}_1 \mid \mathbf{r}_{1\ldots k-1} = r] = \Pr[\mathrm{var}(v_k) \text{ survives} \mid \boldsymbol{\rho}_{\mathrm{var}(v_1)} = r_1 \wedge \cdots \wedge \boldsymbol{\rho}_{\mathrm{var}(v_{k-1})} = r_{k-1}].$$

Now, for each $k' < k$, if $\mathrm{var}(v_{k'}) \notin B_j$, then the predicate $\boldsymbol{\rho}_{\mathrm{var}(v_{k'})} = r_{k'}$ has no effect on the above conditional probability by Lemma 3.8. Therefore,

$$\Pr[k \in \mathbf{Q}_1 \mid \mathbf{r}_{1\ldots k-1} = r] = \Pr\left[\mathrm{var}(v_k) \text{ survives} \mid \boldsymbol{\rho}_{\mathrm{prev}(v_k)} = 1^{\mathrm{prev}(v_k)}\right] \le \frac{\Pr[\mathrm{var}(v_k) \text{ survives}]}{\Pr[\boldsymbol{\rho}_{\mathrm{prev}(v_k)} = 1^{\mathrm{prev}(v_k)}]}.$$

For some $q_j \in [0,1]$, we have

$$\begin{aligned}
\Pr\left[\boldsymbol{\rho}_{\mathrm{prev}(v_k)} = 1^{\mathrm{prev}(v_k)}\right] &\ge (1-p) \cdot \Pr\left[(\mathbf{corrb}^{B_j}_{1-\beta,1-q_j})_{\mathrm{prev}(v_k)} = 1^{\mathrm{prev}(v_k)}\right] \\
&\ge (1-p) \cdot \Pr_{\boldsymbol{\varphi} \sim \mathbf{b}^{B_j}_{1-\beta}}[\boldsymbol{\varphi}_{\mathrm{prev}(v_k)} = 1^{\mathrm{prev}(v_k)} \wedge \boldsymbol{\varphi} \ne 1^{B_j}] \\
&= (1-p) \cdot (1-\beta)^{|\mathrm{prev}(v_k)|} \cdot (1 - (1-\beta)^{|B_j| - |\mathrm{prev}(v_k)|}).
\end{aligned}$$

Since $|\mathrm{prev}(v_k)| \leq 1/\beta$ and $\beta \leq 1/2$, we have $(1-\beta)^{|\mathrm{prev}(v_k)|} \geq (1/2)^2 = 1/4$, and therefore

$$\begin{aligned}\Pr\left[\boldsymbol{\rho}_{\mathrm{prev}(v_k)} = 1^{\mathrm{prev}(v_k)}\right] &\geq (1-p)\cdot(1/4)\cdot(1-4\cdot(1-\beta)^{|B_j|}) \\ &\geq (1-p)\cdot(1/4)\cdot(1-4\cdot\nu') \\ &\geq \frac{1}{16}. \end{aligned} \tag{7.3}$$

Therefore,

$$\Pr[k \in \mathbf{Q}_1 \mid \mathbf{r}_{1\ldots k-1} = r] \leq 16\cdot\Pr[\mathrm{var}(v_k) \text{ survives}] \leq \frac{16\cdot p\cdot\beta}{1-\nu'} \leq 19\cdot p\cdot\beta.$$

The random variable $\mathbf{r}_{1\ldots k-1}$ determines $\mathbf{Q}_1 \cap [k-1]$, so we may now apply the upper Chernoff bound for correlated random bits (Corollary 3.12), completing the proof of Equation (7.1). □

Claim 7.1.2. *Equation (7.2) holds.*

Proof. Let $\mathbf{K} = \{k \in [D] : |\mathrm{prev}(\mathbf{v}_k)| = R-1\}$, and let $\mathbf{k}_i$ be the i^{th} element in $\mathbf{K}$; that is, $\mathbf{k}_1 < \mathbf{k}_2 < \ldots$ and $\mathbf{K} = \{\mathbf{k}_1, \mathbf{k}_2, \ldots\}$. Note that $|\mathbf{K}| \leq D/R$, because each $\mathbf{k}_s \in \mathbf{K}$ corresponds to R distinct queries.

Fix an integer $s \leq D/R$, fix a string $r \in \mathrm{Supp}(\mathbf{r}_{1\ldots\mathbf{k}_s-1})$, and denote $k = |r|+1$. [20] We will condition on the event $\mathbf{r}_{1\ldots\mathbf{k}_s-1} = r$, and towards doing so we first argue that it is identical to the event $\mathbf{r}_{1\ldots k-1} = r$. The direction $\Rightarrow$ follows since conditioning on the former, we have $\mathbf{k}_s = k$. To see the direction $\Leftarrow$, note that the condition $\mathbf{r}_{1\ldots k-1} = r$ implies that $\mathbf{v}_1 = v_1, \ldots, \mathbf{v}_k = v_k$ for some fixed vertices $v_1, \ldots, v_k$; then, there are precisely s values $1 \leq k_1 < k_2 < \cdots < k_s = k$ such that for every $s' \leq s$ have $|\mathrm{prev}(v_{k_{s'}})| = R-1$ (since $r \in \mathrm{Supp}(\mathbf{r}_{1\ldots\mathbf{k}_s-1})$ for our fixed choice of s).

We bound the conditional probability $\Pr[\mathbf{k}_s \in \mathbf{Q}_2 \mid \mathbf{r}_{1\ldots\mathbf{k}_s-1} = r]$, as follows:

$$\begin{aligned}\Pr[\mathbf{k}_s \in \mathbf{Q}_2 \mid \mathbf{r}_{1\ldots\mathbf{k}_s-1} = r] &= \Pr[k \in \mathbf{Q}_2 \mid \mathbf{r}_{1\ldots k-1} = r] \\ &= \Pr[k \in \mathbf{Q}_2 \mid \boldsymbol{\rho}_{\mathrm{var}(v_1)} = r_1 \wedge \cdots \wedge \boldsymbol{\rho}_{\mathrm{var}(v_{k-1})} = r_{k-1}].\end{aligned}$$

Assume that the conditional probability above is nonzero. Then $\mathrm{var}(v_k)$ is in some block B_j, and for each $k' < k$, if $\mathrm{var}(v_{k'}) \in B_j$, then $r_{k'} = 1$. As already discussed, $|\mathrm{prev}(v_k)| = R-1$. The only remaining requirement in the definition of $\mathbf{Q}_2$ is that query k is to a variable in a surviving block, so

$$\Pr[\mathbf{k}_s \in \mathbf{Q}_2 \mid \mathbf{r}_{1\ldots\mathbf{k}_s-1} = r] = \Pr[B_j \text{ survives} \mid \boldsymbol{\rho}_{\mathrm{var}(v_1)} = r_1 \wedge \cdots \wedge \boldsymbol{\rho}_{\mathrm{var}(v_{k-1})} = r_{k-1}].$$

Once again, for each $k' < k$, if $\mathrm{var}(v_{k'}) \notin B_j$, then the predicate $\boldsymbol{\rho}_{\mathrm{var}(v_{k'})} = r_{k'}$ has no effect on the above conditional probability by Lemma 3.8. Therefore,

$$\begin{aligned}\Pr[\mathbf{k}_s \in \mathbf{Q}_2 \mid \mathbf{r}_{1\ldots\mathbf{k}_s-1} = r] &= \Pr\left[B_j \text{ survives} \mid \boldsymbol{\rho}_{\mathrm{prev}(v_k)} = 1^{\mathrm{prev}(v_k)}\right] \\ &\leq \frac{\Pr[B_j \text{ survives}]}{\Pr[\boldsymbol{\rho}_{\mathrm{prev}(v_k)} = 1^{\mathrm{prev}(v_k)}]} \\ &\leq 16\cdot p,\end{aligned}$$

where the last step follows by the same calculation as Equation (7.3). The random variable $\mathbf{r}_{1\ldots\mathbf{k}_s-1}$ determines $\mathbf{Q}_2 \cap [\mathbf{k}_{s-1}]$, so we may once again apply the upper Chernoff bound for correlated random bits (Corollary 3.12), completing the proof of Equation (7.2). □

[20] Indeed, the random variable $\mathbf{r}_{1,\ldots,\mathbf{k}_s-1}$ is supported over strings of different lengths, and we condition on a fixed r from its support of length that we denote by $k-1$.

The lemma follows by combining the two latter claims. ■

7.2 Conditional corrupted biased block projections

As discussed in Section 2.4 and the previous section, we would like to analyze the effect of a corrupted biased block projection $\boldsymbol{\pi}$ on a decision tree T that has LTF circuits at the leaves. We would like to prove that under the projection, the tree becomes a decision tree T' with shallower LTF circuits at the leaves. The tree T' operates in two phases: the first phase simulates the tree portion of T until it reaches a leaf labeled by a circuit C, and the second phase simulates C. The purpose of this section is to help with the analysis of the second phase.

In our analysis, we will consider sampling an input $\mathbf{y}$ to the projected function $T\restriction_{\boldsymbol{\pi}}$ from the product distribution $\mathbf{b}_\alpha^t$. To bound the error, we will be interested in the distribution of $(\boldsymbol{\pi}, \mathbf{y})$ *conditioned* on reaching a particular leaf. The following lemma will help us to reason about this conditional probability distribution. We stress that our condition asserts what happens after the *composition* $\mathbf{y} \circ \boldsymbol{\pi}$ is applied, and we ask how this condition affects the distributions of $\boldsymbol{\pi}$ and of $\mathbf{y}$. The following lemma asserts that the condition has little effect on these distributions (other than the obvious effect of fixing the relevant values), as long as we only condition on a small number of variables in each block. We begin by focusing on a single block.

Lemma 7.2 (conditioning on a few variables being fixed to 1 by $\mathbf{y} \circ \boldsymbol{\pi}$ doesn't change the general structure of $(\boldsymbol{\pi}, \mathbf{y} \circ \boldsymbol{\pi})$**).** *Let $B = \{x_1, \ldots, x_m\}$ be a nonempty set of variables and let y be one more variable. Let $\boldsymbol{\pi} \colon B \cup \{0,1\} \to \{y, 0, 1\}$ be a p-surviving $(1-\beta)$-biased $(1-q)$-corrupted (ν, ν')-block projection, where all of B is a single block, $p, \nu', q \leq 0.01$, and $\beta \leq 1/2$. Sample $\mathbf{y} \sim \mathbf{b}_\alpha$ independently of $\boldsymbol{\pi}$. Let $Q \subseteq B$ with $|Q| \leq 1/\beta$, let $\overline{Q} = B \setminus Q$, and let $\mathcal{E}$ be the event $(\mathbf{y} \circ \boldsymbol{\pi})_Q = 1^Q$. Then the conditional joint distribution*

$$\left(\boldsymbol{\pi}_{\overline{Q}}, (\mathbf{y} \circ \boldsymbol{\pi})_{\overline{Q}}\right) \mid \mathcal{E} \tag{7.4}$$

is identical to a joint distribution $(\widetilde{\boldsymbol{\pi}}, \widetilde{\mathbf{y}} \circ \widetilde{\boldsymbol{\pi}})$, where:

1. *The projection $\widetilde{\boldsymbol{\pi}}$ is a $\widetilde{p}$-surviving $(1-\beta)$-biased $(1-\widetilde{q})$-corrupted $(\nu, \widetilde{\nu'})$-block projection for some $\widetilde{p} \leq O(p)$, $\widetilde{\nu'} \leq O(\nu')$, and $\widetilde{q} \in [0,1]$ with $\widetilde{q} \leq O(q+p)$, where the big-O notation hides universal constants.*
2. *The bit $\widetilde{\mathbf{y}}$ is distributed according to $\mathbf{b}_{\widetilde{\alpha}}$ for some $\widetilde{\alpha} \in [0,1]$.*
3. *The random variables $\widetilde{\boldsymbol{\pi}}$ and $\widetilde{\mathbf{y}}$ are independent.*

To be clear, in Equation (7.4), the expression $\boldsymbol{\pi}_{\overline{Q}}$ refers to the action of the projection $\boldsymbol{\pi}$ on variables outside Q. Similarly, $(\mathbf{y} \circ \boldsymbol{\pi})_{\overline{Q}}$ denotes the substring of the composition $\mathbf{y} \circ \boldsymbol{\pi}$ that is obtained by deleting the coordinates in Q. The domain of $\widetilde{\boldsymbol{\pi}}$ is $\overline{Q} \cup \{0,1\}$.

Proof. For any set $\mathcal{X}$ of variables, let $\overline{\mathbf{corrb}}^{\mathcal{X}}_{1-\beta,1-q'}$ be the following distribution over $\{0,1\}^{\mathcal{X}}$:

1. With probability q', we assign 1 to every variable in $\mathcal{X}$.
2. With probability $1-q'$, we sample an assignment from the conditional distribution $(\boldsymbol{\varphi} \sim \mathbf{b}^{\mathcal{X}}_{1-\beta} \mid \boldsymbol{\varphi} \neq 1^{\mathcal{X}})$.

Observe that for any $q \in [0,1]$, if we let $q' = q \cdot (1-\beta)^{|\mathcal{X}|}$, then the two distributions $\mathbf{corrb}^{\mathcal{X}}_{1-\beta,1-q}$ and $\overline{\mathbf{corrb}}^{\mathcal{X}}_{1-\beta,1-q'}$ are identical (see Definition 6.2). For this proof, it will be more convenient to reason about $\overline{\mathbf{corrb}}^{\mathcal{X}}_{1-\beta,1-q'}$.

Let $\mathbf{y}' \sim \mathbf{b}_{\widetilde{\alpha}}$ be independent of $(\boldsymbol{\pi}, \mathbf{y})$; the value $\widetilde{\alpha}$ will be specified later (in particular, it will be carefully chosen to ensure that $\widetilde{\boldsymbol{\pi}}$ and $\widetilde{\mathbf{y}}$ are independent). Define

$$\mathbf{y}'' = \begin{cases} \mathbf{y} & \text{if some variable in } \overline{Q} \text{ survives, i.e., } \boldsymbol{\pi}^{-1}(y) \cap \overline{Q} \neq \emptyset \\ \mathbf{y}' & \text{otherwise.} \end{cases}$$

The joint distribution $(\widetilde{\boldsymbol{\pi}}, \widetilde{\mathbf{y}})$ is defined to be the conditional joint distribution

$$(\boldsymbol{\pi}_{\overline{Q}}, \mathbf{y}'') \mid \mathcal{E}.$$

Observe that when $\boldsymbol{\pi}$ does not keep any variables in $\overline{Q}$ alive, the bit $\mathbf{y}$ has no effect on $(\mathbf{y} \circ \boldsymbol{\pi})_{\overline{Q}}$ anyway; indeed, $(\mathbf{y} \circ \boldsymbol{\pi})_{\overline{Q}} = \boldsymbol{\pi}_{\overline{Q}}$ in this case. Thus, although $\mathbf{y}''$ and $\mathbf{y}$ are not necessarily equal, we nevertheless always have the equality

$$(\mathbf{y}'' \circ \boldsymbol{\pi})_{\overline{Q}} = (\mathbf{y} \circ \boldsymbol{\pi})_{\overline{Q}}.$$

Therefore, the joint distribution $(\widetilde{\boldsymbol{\pi}}, \widetilde{\mathbf{y}} \circ \widetilde{\boldsymbol{\pi}})$ is indeed identical to the conditional distribution of Equation (7.4).

Our remaining task is to prove Items 1-3. In the proof that $\widetilde{\boldsymbol{\pi}}$ is a biased corrupted block projection, we will view $\overline{Q}$ as a single block. We partition the event $\mathcal{E}$ into four cases.

1. Let $\mathcal{E}_1$ be the event that some variable in $\overline{Q}$ survives and $\mathcal{E}$ occurs, i.e.,
$$\boldsymbol{\pi}^{-1}(y) \cap \overline{Q} \neq \emptyset \text{ and } (\mathbf{y} \circ \boldsymbol{\pi})_Q = 1^Q.$$
2. Let $\mathcal{E}_2$ be the event that $\emptyset \neq \boldsymbol{\pi}^{-1}(y) \subseteq Q$ and $\mathbf{y} = 1$. (This implies that $(\mathbf{y} \circ \boldsymbol{\pi})_Q = 1^Q$.)
3. Let $\mathcal{E}_3$ be the event that $\boldsymbol{\pi} = 1^B$.
4. Let $\mathcal{E}_4$ be the event that $\boldsymbol{\pi} \in \{0,1\}^B \setminus \{1^B\}$ and $\boldsymbol{\pi}_Q = 1^Q$.

Note that the above four cases are mutually exclusive and $\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3 \cup \mathcal{E}_4$. Let us first analyze $\boldsymbol{\pi}$ and $\mathbf{y}$ conditioned on $\mathcal{E}_i$ for each $i \in [4]$. Then, later, we will analyze the probability of $\mathcal{E}_i$ conditioned on $\mathcal{E}$ for each i. Together, these analyses will reveal the distribution of $(\widetilde{\boldsymbol{\pi}}, \widetilde{\mathbf{y}} \circ \widetilde{\boldsymbol{\pi}})$.

The case $\mathcal{E}_1$. When $\mathcal{E}_1$ occurs, the new block $\overline{Q}$ survives. Each variable in $\overline{Q}$ either is assigned 1 or else survives. For any $\psi \in \{0,1\}^{\overline{Q}} \setminus \{1^{\overline{Q}}\}$, looking at the definition of a corrupted biased block projection, we have

$$\begin{aligned} \Pr[\boldsymbol{\pi}^{-1}(y) \cap \overline{Q} = \psi^{-1}(0) \mid \mathcal{E}_1] &= \Pr_{\substack{\boldsymbol{\varphi} \sim \mathbf{b}^B_{1-\beta} \\ \mathbf{y} \sim \mathbf{b}_\alpha}} [\boldsymbol{\varphi}_{\overline{Q}} = \psi \mid \boldsymbol{\varphi}_{\overline{Q}} \neq 1^{\overline{Q}} \wedge (\boldsymbol{\varphi}_Q = 1^Q \vee \mathbf{y} = 1)] \\ &= \Pr_{\boldsymbol{\varphi} \sim \mathbf{b}^{\overline{Q}}_{1-\beta}} [\boldsymbol{\varphi} = \psi \mid \boldsymbol{\varphi} \neq 1^{\overline{Q}}], \end{aligned}$$

which is precisely the distribution on living variables stipulated by the definition of a corrupted biased block projection in the case when the block survives. Meanwhile,

$$\begin{aligned}
&\Pr[\mathbf{y}'' = 1 \mid \mathcal{E}_1 \wedge \boldsymbol{\pi}^{-1}(y) \cap \overline{Q} = \psi^{-1}(0)] \\
&\quad = \Pr[\mathbf{y} = 1 \mid \mathcal{E}_1 \wedge \boldsymbol{\pi}^{-1}(y) \cap \overline{Q} = \psi^{-1}(0)] \\
&\quad = \Pr[\mathbf{y} = 1 \mid \mathcal{E} \wedge \boldsymbol{\pi}^{-1}(y) \cap \overline{Q} = \psi^{-1}(0)] \\
&\quad = \Pr_{\substack{\boldsymbol{\varphi} \sim \mathbf{b}^B_{1-\beta} \\ \mathbf{y} \sim \mathbf{b}_\alpha}} [\mathbf{y} = 1 \mid \boldsymbol{\varphi}_{\overline{Q}} = \psi \wedge (\boldsymbol{\varphi}_Q = 1^Q \vee \mathbf{y} = 1)] \\
&\quad = \frac{\Pr_{\mathbf{y},\boldsymbol{\varphi}}[\mathbf{y} = 1 \wedge \boldsymbol{\varphi}_{\overline{Q}} = \psi]}{\Pr_{\mathbf{y},\boldsymbol{\varphi}}[\boldsymbol{\varphi}_{\overline{Q}} = \psi \wedge (\boldsymbol{\varphi}_Q = 1^Q \vee \mathbf{y} = 1)]} \\
&\quad = \frac{\Pr_{\mathbf{y}}[\mathbf{y} = 1]}{\Pr_{\mathbf{y},\boldsymbol{\varphi}}[\boldsymbol{\varphi}_Q = 1^Q \vee \mathbf{y} = 1]} && \text{(independence)} \\
&\quad = \frac{\Pr_{\mathbf{y}}[\mathbf{y} = 1]}{1 - \Pr_{\boldsymbol{\varphi}}[\boldsymbol{\varphi}_Q \neq 1^Q] \Pr_{\mathbf{y}}[\mathbf{y} \neq 1]} && \text{(independence)} \\
&\quad = \frac{\alpha}{1 - (1-\alpha) \cdot (1 - (1-\beta)^{|Q|})}.
\end{aligned}$$

Crucially, the last expression has no dependence on ψ. We can therefore define

$$\widetilde{\alpha} = \frac{\alpha}{1 - (1-\alpha) \cdot (1 - (1-\beta)^{|Q|})},$$

ensuring that when we condition on $\mathcal{E}_1$ and any particular value for $\boldsymbol{\pi}_{\overline{Q}}$, the bit $\mathbf{y}''$ has bias $\widetilde{\alpha}$.

The case $\mathcal{E}_2$. When $\mathcal{E}_2$ occurs, the full block B survives, but the new block $\overline{Q}$ does not survive. This means that under this case $\boldsymbol{\pi}_{\overline{Q}} = 1_{\overline{Q}}$ surely.

The case $\mathcal{E}_3$. When $\mathcal{E}_3$ occurs, the block B does not survive, hence the new block $\overline{Q}$ does not survive either and is assigned the all 1s string.

The case $\mathcal{E}_4$. When $\mathcal{E}_4$ occurs, the block B does not survive (hence the new block $\overline{Q}$ does not survive either). For any $\psi \in \{0,1\}^{\overline{Q}}$,

$$\begin{aligned}
\Pr[\boldsymbol{\pi}_{\overline{Q}} = \psi \mid \mathcal{E}_4] &= \Pr_{\boldsymbol{\varphi} \sim \mathbf{b}^B_{1-\beta}} [\boldsymbol{\varphi}_{\overline{Q}} = \psi \mid \boldsymbol{\varphi}_Q = 1^Q, \boldsymbol{\varphi} \neq 1^B] \\
&= \Pr_{\boldsymbol{\varphi} \sim \mathbf{b}^B_{1-\beta}} [\boldsymbol{\varphi}_{\overline{Q}} = \psi \mid \boldsymbol{\varphi}_{\overline{Q}} \neq 1^{\overline{Q}}],
\end{aligned}$$

which is precisely the distribution on assignments in the corrupted case 2 of the definition of a biased corrupted block projection.

The overall distribution. To summarize, we have shown that conditioned on $\mathcal{E}_1$, the new block $\overline{Q}$ survives and $\boldsymbol{\pi}_{\overline{Q}}$ follows precisely the distribution of the "surviving block" case of the definition of a $(1-\beta)$-biased corrupted block projection;[21] conditioned on $\mathcal{E}_2 \cup \mathcal{E}_3$, the new

[21] To be more precise, the distribution is that of the "surviving block" case of a p_0-surviving $(1-\beta)$-biased $(1-q_0)$-corrupted (ν_0, ν_0')-block restriction, for any values of p_0, q_0, ν_0, ν_0' (since the four latter parameter values do not affect the distribution in the "surviving block" case).

block does not survive and $\boldsymbol{\pi}_{\overline{Q}}$ is the all 1s assignment; and conditioned on $\mathcal{E}_4$, the new block does not survive and $\boldsymbol{\pi}_{\overline{Q}}$ is sampled from the product distribution $\mathbf{b}^{\overline{Q}}_{1-\beta}$ conditioned on not being the all 1s string.

Consequently, conditioned on the new block not surviving, the assignment is sampled from $\overline{\mathbf{corrb}}^{\overline{Q}}_{1-\beta,1-\widetilde{q}'}$ for some $\widetilde{q}'$. We would like to show that the distribution is $\mathbf{corrb}^{\overline{Q}}_{1-\beta,1-\widetilde{q}}$ for some $\widetilde{q} \in [0,1]$. To do so, we bound $\widetilde{q}'$, the probability of sampling the all-ones string in the non-surviving case, as follows:

Claim 7.2.1. *We have that* $\widetilde{q}' \leq (1-\beta)^{|\overline{Q}|} \cdot (1.1q + 6p) \leq (1-\beta)^{|\overline{Q}|}$.

Proof. We have

$$\widetilde{q}' = \Pr[\mathcal{E}_2 \cup \mathcal{E}_3 \mid \mathcal{E}_2 \cup \mathcal{E}_3 \cup \mathcal{E}_4] \leq \frac{\Pr[\mathcal{E}_2]}{\Pr[\mathcal{E}_3 \cup \mathcal{E}_4]} + \frac{\Pr[\mathcal{E}_3]}{\Pr[\mathcal{E}_3 \cup \mathcal{E}_4]}. \tag{2}$$

First let us analyze the denominator $\Pr[\mathcal{E}_3 \cup \mathcal{E}_4]$. Let q' be the probability that $\boldsymbol{\pi}$ is all ones conditioned on non-survival, i.e., $q' = q \cdot (1-\beta)^{|B|}$. We have

$$\begin{aligned}
\Pr[\mathcal{E}_3 \cup \mathcal{E}_4] &= \Pr[B \text{ non-surviving} \wedge \boldsymbol{\pi}_Q = 1^Q] \\
&= \Pr[B \text{ non-surviving}] \cdot \left(q' \cdot 1 + (1-q') \cdot \frac{(1-\beta)^{|Q|} - (1-\beta)^{|B|}}{1-(1-\beta)^{|B|}} \right) \\
&\geq \Pr[B \text{ non-surviving}] \cdot ((1-\beta)^{|Q|} - (1-\beta)^{|B|}) \qquad (7.5) \\
&\geq (1-p) \cdot (1/4 - \nu') \geq 0.99 \cdot 0.24 \geq 1/5
\end{aligned}$$

because $p, \nu' \leq 0.01$, $|Q| \leq 1/\beta$, and $\beta \leq 1/2$, thus $(1-\beta)^{|Q|} \geq (1-\beta)^{1/\beta} \geq 0.5^2 = 1/4$. Now we analyze the two terms on the right-hand side of Equation (2) separately. For the first term, we have

$$\frac{\Pr[\mathcal{E}_2]}{\Pr[\mathcal{E}_3 \cup \mathcal{E}_4]} \leq \frac{\Pr[\mathcal{E}_2]}{1/5} \leq 5p \cdot \frac{(1-\beta)^{|\overline{Q}|}}{1-(1-\beta)^{|B|}} \leq 5p \cdot \frac{(1-\beta)^{|\overline{Q}|}}{1-\nu'} \leq 6p(1-\beta)^{|\overline{Q}|}$$

where we used $\nu' \leq 0.01$. Meanwhile, the second term is bounded by

$$\begin{aligned}
\frac{\Pr[\mathcal{E}_3]}{\Pr[\mathcal{E}_3 \cup \mathcal{E}_4]} &\leq \frac{\Pr[B \text{ non-surviving}] \cdot q'}{\Pr[B \text{ non-surviving}]((1-\beta)^{|Q|} - (1-\beta)^{|B|})} \qquad \text{by Equation (7.5)} \\
&= q \cdot \frac{(1-\beta)^{|B|-|Q|}}{1-(1-\beta)^{|B|-|Q|}}
\end{aligned}$$

and since

$$(1-\beta)^{|\overline{Q}|} \leq (1-\beta)^{|B|-1/\beta} \leq e \cdot \nu' \leq 0.03$$

we have $\frac{\Pr[\mathcal{E}_3]}{\Pr[\mathcal{E}_3 \cup \mathcal{E}_4]} \leq 1.1q \cdot (1-\beta)^{|B|-|Q|}$. Overall, we get that $\widetilde{q}' \leq (1-\beta)^{|\overline{Q}|} \cdot (1.1q + 6p)$. The claim follows since $q, p \leq 0.01$. □

Let $\widetilde{q} = \widetilde{q}'/(1-\beta)^{|\overline{Q}|} \leq 1$. The distribution $\overline{\mathbf{corrb}}^{\overline{Q}}_{1-\beta,1-\widetilde{q}'}$ is identical to $\mathbf{corrb}^{\overline{Q}}_{1-\beta,1-\widetilde{q}}$. Therefore, conditioned on non-survival, $\widetilde{\boldsymbol{\pi}}$ is indeed a corrupted biased assignment. This shows that $\widetilde{\boldsymbol{\pi}}$ is indeed a $\widetilde{p}$-surviving $(1-\beta)$-biased $(1-\widetilde{q})$-corrupted $(\widetilde{\nu}, \widetilde{\nu}')$-block projection for some parameters $\widetilde{p}$, $\widetilde{\nu}$, $\widetilde{\nu}'$, and $\widetilde{q} \leq 1.1q + 6p$. Let us now bound the remaining parameters $\widetilde{p}$, $\widetilde{\nu}$, and $\widetilde{\nu}'$.

We have $\Pr[\mathcal{E}] \geq \Pr[\mathcal{E}_3 \cup \mathcal{E}_4] \geq 1/5$. Therefore, the survival probability is given by

$$\widetilde{p} = \Pr[\mathcal{E}_1 \mid \mathcal{E}] = \frac{\Pr[\boldsymbol{\pi}^{-1}(y) \cap \overline{Q} \neq \emptyset \wedge \mathcal{E}]}{\Pr[\mathcal{E}]} \leq \frac{\Pr[B \text{ surviving}]}{\Pr[\mathcal{E}]} \leq 5 \cdot p.$$

Next, observe that

$$(1-\beta)^{|\overline{Q}|} \geq (1-\beta)^{|B|} \geq \nu,$$

so we can take $\widetilde{\nu} = \nu$, and $(1-\beta)^{|\overline{Q}|} \leq e \cdot \nu'$ so indeed, we can take $\widetilde{\nu'} = e \cdot \nu'$.

Finally, consider any value $\pi \in \{0,1\}^{\overline{Q}} \cup \{1,y\}^{\overline{Q}}$. Conditioned on $\widetilde{\boldsymbol{\pi}} = \pi$, what is the bias of $\widetilde{\mathbf{y}}$? Equivalently, conditioned on $\boldsymbol{\pi}_{\overline{Q}} = \pi$ and $\mathcal{E}$, what is the bias of $\mathbf{y}''$? If $\pi^{-1}(y) \neq \emptyset$, then conditioning on $\boldsymbol{\pi}_{\overline{Q}} = \pi$ and $\mathcal{E}$ is equivalent to conditioning on $\boldsymbol{\pi}_{\overline{Q}} = \pi$ and $\mathcal{E}_1$. Under the latter conditioning, we showed that $\mathbf{y}''$ has bias $\widetilde{\alpha}$, with no dependence on π. Meanwhile, if $\pi^{-1}(y) = \emptyset$, then conditioned on $\boldsymbol{\pi}_{\overline{Q}} = \pi$ and $\mathcal{E}$, we have $\mathbf{y}'' = \mathbf{y}'$, and hence once again, $\mathbf{y}''$ has bias $\widetilde{\alpha}$, with no dependence on π. Therefore, overall, $\widetilde{\mathbf{y}}$ is indeed independent of $\widetilde{\boldsymbol{\pi}}$ and distributed according to $\mathbf{b}_{\widetilde{\alpha}}$. ■

We now extend Lemma 7.2 to the case of multiple blocks. This lemma is the reason that we have had to consider product distributions with a whole vector $\vec{\alpha}$ of potentially-different biases, instead of assuming that every coordinate has the same bias.

Lemma 7.3 (extending Lemma 7.2 to multiple blocks)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets of variables with $|\mathcal{X}| = n \geq t = |\mathcal{Y}|$. Let $\boldsymbol{\pi} \colon \mathcal{X} \cup \{0,1\} \to \mathcal{Y} \cup \{0,1\}$ be a p-surviving $(1-\beta)$-biased $(1-q)$-corrupted (ν,ν')-block projection with blocks $B_1, \ldots, B_t$, where $p, \nu', q \leq 0.01$ and $\beta \leq 1/2$. Sample $\mathbf{y} \sim \mathbf{b}_\alpha^{\mathcal{Y}}$ independently of $\boldsymbol{\pi}$. Let $Q \in \{0,1,\star\}^{\mathcal{X}}$, and assume that for every block $j \in [t]$, either (a) $Q_{B_j} \in \{0,1\}^{B_j}$, or else (b) $Q_{B_j} \in \{1,\star\}^{B_j}$ and $|Q^{-1}(1) \cap B_j| \leq 1/\beta$. Let $\mathcal{E}$ be the event that $\mathbf{y} \circ \boldsymbol{\pi}$ agrees with Q, i.e.,*

$$\mathcal{E} \iff \forall x_i \in \mathcal{X},\ Q_i \in \{(\mathbf{y} \circ \boldsymbol{\pi})_i, \star\}.$$

Then the conditional joint distribution

$$\left(\boldsymbol{\pi}_{Q^{-1}(\star)}, (\mathbf{y} \circ \boldsymbol{\pi})_{Q^{-1}(\star)}\right) \mid \mathcal{E}$$

is identical to a joint distribution $(\widetilde{\boldsymbol{\pi}}, \widetilde{\mathbf{y}} \circ \widetilde{\boldsymbol{\pi}})$, where:

1. *The projection $\widetilde{\boldsymbol{\pi}} \colon Q^{-1}(\star) \cup \{0,1\} \to \mathcal{Y} \cup \{0,1\}$ is a $\widetilde{p}$-surviving $(1-\beta)$-biased $(1-\widetilde{q})$-corrupted $(\nu,\widetilde{\nu'})$-block projection for some $\widetilde{p} \leq O(p)$, $\widetilde{\nu'} \leq O(\nu')$, and $\widetilde{q} \in [0,1]$ with $\widetilde{q} \leq O(q+p)$.*
2. *The vector $\widetilde{\mathbf{y}}$ is distributed over $\{0,1\}^{\mathcal{Y}}$ according to $\mathbf{b}_{\vec{\alpha}'}$ for some $\vec{\alpha}' \in [0,1]^t$.*
3. *The random variables $\widetilde{\boldsymbol{\pi}}$ and $\widetilde{\mathbf{y}}$ are independent.*

Proof. For convenience, let B_0 be the set of variables that are not in blocks, i.e., $B_0 = \mathcal{X} \setminus (B_1 \cup \cdots \cup B_t)$. For each $j \in \{0, \ldots, t\}$, let $\mathcal{E}_j$ be the event that $(\mathbf{y} \circ \boldsymbol{\pi})_{B_j}$ agrees with Q_{B_j}, i.e.,

$$\mathcal{E}_j \iff \forall x_i \in B_j,\ Q_i \in \{(\mathbf{y} \circ \boldsymbol{\pi})_i, \star\}.$$

By Lemma 3.8, the random variables

$$\left(\boldsymbol{\pi}_{Q^{-1}(\star)\cap B_0}, (\mathbf{y} \circ \boldsymbol{\pi})_{Q^{-1}(\star)\cap B_0}\right), \ldots, \left(\boldsymbol{\pi}_{Q^{-1}(\star)\cap B_t}, (\mathbf{y} \circ \boldsymbol{\pi})_{Q^{-1}(\star)\cap B_t}\right)$$

are conditionally independent given $\mathcal{E}$, and the distribution of $\left(\boldsymbol{\pi}_{Q^{-1}(\star)\cap B_j}, (\mathbf{y}\circ\boldsymbol{\pi})_{Q^{-1}(\star)\cap B_j}\right)$ is the same whether we condition on $\mathcal{E}$ or on $\mathcal{E}_j$.

Now consider a fixed j. If $j = 0$, then

$$(\boldsymbol{\pi}_{Q^{-1}(\star)\cap B_0} \mid \mathcal{E}_0)$$

is distributed according to $\mathbf{b}_{1-\beta}^{Q^{-1}(\star)\cap B_0}$, and $(\mathbf{y}\circ\boldsymbol{\pi})_{Q^{-1}(\star)\cap B_0} = \boldsymbol{\pi}_{Q^{-1}(\star)\cap B_0}$. We will consider these variables to still be outside the blocks. Now suppose $j \in [t]$. In $\widetilde{\boldsymbol{\pi}}$, the j^{th} block is $\widetilde{B}_j \stackrel{\text{def}}{=} Q^{-1}(\star)\cap B_j$. If $\widetilde{B}_j = \emptyset$, then we can set $\vec{\alpha}'_j$ arbitrarily; say $\vec{\alpha}'_j = 0$. Finally, if $\widetilde{B}_j \neq \emptyset$, then we are in the situation of Lemma 7.2, completing the proof. ∎

7.3 Simplification of decision trees with LTF circuits at the leaves

In this section, we will show that corrupted biased block projections simplify decision trees that have LTF circuits at the leaves. Recall that we already analyzed the case that the initial function is simply an LTF circuit f (Proposition 6.12). In that case, we argued that with high probability over the projection $\boldsymbol{\pi}$, we have a low approximation error with respect to the input $\mathbf{y}$ to the projected function $f\restriction_{\boldsymbol{\pi}}$. Going forward, it will be more convenient to bound the approximation error *on average* over $\boldsymbol{\pi}$. In other words, we will bound the probability of error with respect to the random choice of $\boldsymbol{\pi}$ *and* the independent random choice of $\mathbf{y}$. The following definition will help us to reason about this type of average-case approximation.

Definition 7.4 (Approximation with low average error)**.** *Let $\mathbf{f}$ and $\widetilde{\mathbf{f}}$ be jointly distributed random functions mapping $\{0,1\}^n \to \{0,1\}$, let $\mathbf{X}$ be a distribution over $\{0,1\}^n$, and let $\epsilon > 0$. We say that $\widetilde{\mathbf{f}}$* approximates *$\mathbf{f}$ under $\mathbf{X}$ with average error ϵ if*

$$\Pr_{\substack{(\mathbf{f},\widetilde{\mathbf{f}})\\ \mathbf{x}\sim\mathbf{X}}}\left[\widetilde{\mathbf{f}}(\mathbf{x}) \neq \mathbf{f}(\mathbf{x})\right] \leq \epsilon.$$

Here $\mathbf{x}$ is sampled independently of $\left(\mathbf{f},\widetilde{\mathbf{f}}\right)$ whereas $\mathbf{f}$ and $\widetilde{\mathbf{f}}$ may be correlated.

In our application, $\mathbf{f}$ will be the projection of a function under a random projection $\boldsymbol{\pi}$, and $\widetilde{\mathbf{f}}$ will be an approximator that is designed based on $\boldsymbol{\pi}$. Definition 3.5 is the special case of Definition 7.4 where $\mathbf{f}$ and $\widetilde{\mathbf{f}}$ are deterministic functions.

Proposition 7.5 (Simplification of decision trees with LTF circuits at their leaves)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be sets of variables with $|\mathcal{X}| = n \geq t = |\mathcal{Y}|$. Let T be a depth-D decision tree whose leaves are labeled by depth-Δ LTF circuits with at most w wires on input variables $\mathcal{X}$, where $\Delta \geq 1$. Let $\boldsymbol{\pi}\colon \mathcal{X}\cup\{0,1\} \to \mathcal{Y}\cup\{0,1\}$ be a random projection, and assume that either $\boldsymbol{\pi}$ or the complement projection $\overline{\boldsymbol{\pi}}$ is a p-surviving $(1-\beta)$-biased 0.99-corrupted (ν,ν')-block projection. Assume $\beta \leq 1/2$ and $\nu' \leq p \leq 0.01$. Then for every $\alpha \in [0,1]$, the projected function $T\restriction_{\boldsymbol{\pi}}$ can be approximated under $\mathbf{b}_\alpha^{\mathcal{Y}}$ with average error ϵ by a depth-D' decision tree $\mathbf{T}'$ whose leaves are labeled by depth-$(\Delta-1)$ LTF circuits with w wires, where*

$$D' \leq O(p\cdot\beta\cdot D) + \lceil p^{13/12}\cdot w\cdot\beta^{-1/6}\rceil \cdot \mathrm{polylog}(nw/\nu)$$
$$\epsilon \leq p^{1/12}\cdot\beta^{-1/6}\cdot\mathrm{polylog}(nw/\nu).$$

We emphasize that the tree $\mathbf{T}'$ is a random variable (determined by $\boldsymbol{\pi}$).

Proof. Assume first that it is $\boldsymbol{\pi}$ rather than $\overline{\boldsymbol{\pi}}$ that is a p-surviving $(1-\beta)$-biased 0.99-corrupted (ν,ν')-block projection. We will define a decision tree $\widetilde{T}$ on the variable set $\mathcal{X}$ that computes the same function as T with some additional convenient properties. For an input $z \in \{0,1\}^{\mathcal{X}}$, the new tree $\widetilde{T}(z)$ simulates $T(z)$ until it reaches a leaf ℓ_0 labeled by a circuit C_{ℓ_0}. Then, for each block $j \in [t]$, if either

- the simulation made at least $1/\beta$ queries to B_j, or
- the simulation queried a variable x_i where $i \in B_j$ and found that $z_i = 0$,

then our tree $\widetilde{T}$ queries all remaining unqueried variables in the block B_j. The vertex ℓ of $\widetilde{T}$ that is reached at the end of this process is a leaf, and we label it with the circuit C_ℓ obtained from C_{ℓ_0} by plugging in all the values learned by all the queries that have occurred. Thus, C_ℓ computes a function of only those variables that are not queried on the path from the root to ℓ.

Let L be the set of leaves of $\widetilde{T}$. For a leaf $\ell \in L$ and a string $z \in \{0,1\}^{\mathcal{X}}$, we let $\ell(z)$ indicate whether $\widetilde{T}(z)$ reaches ℓ. Furthermore, let $Q_\ell \in \{0,1,\star\}^{\mathcal{X}}$ be the string describing the path from the root to ℓ. That is, if ℓ has a proper ancestor v that queries some variable x_i and the outgoing "$x_i = b$" edge is on the path from v to ℓ, then we set $(Q_\ell)_i = b$; if none of ℓ's proper ancestors query x_i, then we set $(Q_\ell)_i = \star$. Observe that

$$\ell(z) = 1 \iff \forall x_i \in \mathcal{X},\ (Q_\ell)_i \in \{z_i, \star\}.$$

By the construction of $\widetilde{T}$, for every block $j \in [t]$, either (a) $(Q_\ell)_{B_j} \in \{0,1\}^{B_j}$, or else (b) $(Q_\ell)_{B_j} \in \{1,\star\}^{B_j}$ and $|Q_\ell^{-1}(1) \cap B_j| \le 1/\beta$. This matches the hypothesis on Q of Lemma 7.3. Sample $\mathbf{y} \sim \mathbf{b}_\alpha^t$ independently of $\boldsymbol{\pi}$. Then Lemma 7.3 guarantees that the conditional joint distribution

$$\left(\boldsymbol{\pi}_{Q_\ell^{-1}(\star)}, (\mathbf{y} \circ \boldsymbol{\pi})_{Q_\ell^{-1}(\star)}\right) \mid \ell(\mathbf{y} \circ \boldsymbol{\pi}) = 1$$

is identical to a joint distribution $(\widetilde{\boldsymbol{\pi}}^{(\ell)}, \widetilde{\mathbf{y}}^{(\ell)} \circ \widetilde{\boldsymbol{\pi}}^{(\ell)})$ such that

- The projection $\widetilde{\boldsymbol{\pi}}^{(\ell)} \colon Q_\ell^{-1}(\star) \cup \{0,1\} \to \mathcal{Y} \cup \{0,1\}$ is a $\widetilde{p}$-surviving $(1-\beta)$-biased $(1-\widetilde{q})$-corrupted $(\nu,\widetilde{\nu'})$-block projection for some $\widetilde{p} \le O(p)$, $\widetilde{\nu'} \le O(\nu') \le O(p)$, and $\widetilde{q} \le 1$. By increasing $\widetilde{p}$ by a constant factor if necessary, we can ensure that $\widetilde{\nu'} \le \widetilde{p}$.
- The vector $\widetilde{\mathbf{y}}^{(\ell)}$ is distributed according to $\mathbf{b}_{\vec{\alpha}'}$ for some $\vec{\alpha}' \in [0,1]^t$.
- The random variables $\widetilde{\boldsymbol{\pi}}^{(\ell)}$ and $\widetilde{\mathbf{y}}$ are independent.

Let us apply Proposition 6.12 with approximation error ν. Proposition 6.12 tells us that with high probability (say $1-\zeta$) over the choice of $\widetilde{\boldsymbol{\pi}}^{(\ell)}$, the function $(C_\ell)\restriction_{\widetilde{\boldsymbol{\pi}}^{(\ell)}}$ can be approximated under $\widetilde{\mathbf{y}}^{(\ell)}$ with error ν by a depth-D_0 decision tree $\mathbf{T}^{(\ell)}$ with leaves labeled by depth-$(\Delta-1)$ LTF circuits with at most w wires, where

$$\begin{aligned} D_0 &= (\widetilde{p})^{13/12} \cdot w \cdot \beta^{-1/6} \cdot \mathrm{polylog}(nw/\nu) \\ &= p^{13/12} \cdot w \cdot \beta^{-1/6} \cdot \mathrm{polylog}(nw/\nu) \\ \zeta &= (\widetilde{p})^{1/12} \cdot \beta^{-1/6} \cdot \mathrm{polylog}(nw/\nu) \\ &= p^{1/12} \cdot \beta^{-1/6} \cdot \mathrm{polylog}(nw/\nu). \end{aligned}$$

(Recall that C_ℓ is the circuit labeling the leaf ℓ.) When the preceding good event does not occur, set $\mathbf{T}^{(\ell)} \equiv 0$.

The tree $\mathbf{T}^{(\ell)}$ is a random variable that is determined by $\widetilde{\boldsymbol{\pi}}^{(\ell)}$. Furthermore, we may assume without loss of generality that $\mathbf{T}^{(\ell)}$ ignores variables y_j for which $(\widetilde{\boldsymbol{\pi}}^{(\ell)})^{-1}(y_j) = \emptyset$ (i.e., it does not query such variables, and the circuits at its leaves do not look at such variables). Therefore, if a variable y_j is *not* ignored by $\mathbf{T}^{(\ell)}$, then the value of y_j can be deduced by looking at $\widetilde{\boldsymbol{\pi}}^{(\ell)}$ and querying a suitable coordinate of $y \circ \widetilde{\boldsymbol{\pi}}^{(\ell)}$. Consequently, for each projection $\pi \in \mathrm{Supp}(\widetilde{\boldsymbol{\pi}}^{(\ell)})$, there is a depth-$D_0$ decision tree $T[\ell, \pi] \colon \{0,1\}^{Q_\ell^{-1}(\star)} \to \{0,1\}$ with leaves labeled by depth-$(\Delta - 1)$ LTF circuits with at most w wires such that for every $y \in \{0,1\}^{\mathcal{Y}}$, we have

$$\mathbf{T}^{(\ell)}(y) = T[\ell, \widetilde{\boldsymbol{\pi}}^{(\ell)}](y \circ \widetilde{\boldsymbol{\pi}}^{(\ell)}).$$

We stress that for fixed ℓ and π, the tree $T[\ell, \pi]$ is not a random variable.

Now we will define a tree $\mathbf{T}'_0$ based on $\boldsymbol{\pi}$. (Our final tree $\mathbf{T}'$ will be a slightly-modified version of $\mathbf{T}'_0$.) On input $y \in \{0,1\}^{\mathcal{Y}}$, the tree $\mathbf{T}'_0$ simulates $\widetilde{T}(y \circ \boldsymbol{\pi})$ until it reaches a leaf $\boldsymbol{\ell}$. Then, the tree $\mathbf{T}'_0$ simulates the tree portion of $T\left[\boldsymbol{\ell}, \boldsymbol{\pi}_{Q_\ell^{-1}(\star)}\right]\left((y \circ \boldsymbol{\pi})_{Q_\ell^{-1}(\star)}\right)$ until it reaches a leaf $\boldsymbol{\ell}'$, labeled by some circuit $C_{\boldsymbol{\ell}'}$. The corresponding vertex of $\mathbf{T}'_0$ is also a leaf labeled by $C_{\boldsymbol{\ell}'}$.

Let us prove correctness, i.e., let us bound the failure probability $\Pr[\mathbf{T}'_0(\mathbf{y}) \neq T\restriction_{\boldsymbol{\pi}}(\mathbf{y})]$, where the probability is taken with respect to both the choice of projection $\boldsymbol{\pi}$ (which determines the truth table of $\mathbf{T}'_0$) and the choice of input $\mathbf{y}$. We can write

$$\begin{aligned}\Pr\left[\mathbf{T}'_0(\mathbf{y}) \neq T\restriction_{\boldsymbol{\pi}}(\mathbf{y})\right] &= \Pr\left[\mathbf{T}'_0(\mathbf{y}) \neq \widetilde{T}\restriction_{\boldsymbol{\pi}}(\mathbf{y})\right] \\ &= \sum_{\ell \in L} \Pr[\ell(\mathbf{y} \circ \boldsymbol{\pi}) = 1] \cdot \Pr\left[\mathbf{T}'_0(\mathbf{y}) \neq \widetilde{T}\restriction_{\boldsymbol{\pi}}(\mathbf{y}) \mid \ell(\mathbf{y} \circ \boldsymbol{\pi}) = 1\right].\end{aligned}$$

For a fixed leaf $\ell \in L$, we have

$$\begin{aligned}&\Pr\left[\mathbf{T}'_0(\mathbf{y}) \neq \widetilde{T}\restriction_{\boldsymbol{\pi}}(\mathbf{y}) \mid \ell(\mathbf{y} \circ \boldsymbol{\pi}) = 1\right] \\ &= \Pr\left[T\left[\ell, \boldsymbol{\pi}_{Q_\ell^{-1}(\star)}\right]\left((\mathbf{y} \circ \boldsymbol{\pi})_{Q_\ell^{-1}(\star)}\right) \neq C_\ell\left((\mathbf{y} \circ \boldsymbol{\pi})_{Q_\ell^{-1}(\star)}\right) \mid \ell(\mathbf{y} \circ \boldsymbol{\pi}) = 1\right] \\ &= \Pr\left[T\left[\ell, \widetilde{\boldsymbol{\pi}}^{(\ell)}\right]\left(\widetilde{\mathbf{y}}^{(\ell)} \circ \widetilde{\boldsymbol{\pi}}^{(\ell)}\right) \neq C_\ell\left(\widetilde{\mathbf{y}}^{(\ell)} \circ \widetilde{\boldsymbol{\pi}}^{(\ell)}\right)\right] \\ &= \Pr\left[\mathbf{T}^{(\ell)}\left(\widetilde{\mathbf{y}}^{(\ell)}\right) \neq C_\ell\restriction_{\widetilde{\boldsymbol{\pi}}^{(\ell)}}\left(\widetilde{\mathbf{y}}^{(\ell)}\right)\right] \\ &\leq \zeta + \nu.\end{aligned}$$

Therefore, overall,

$$\Pr[\mathbf{T}'_0(\mathbf{y}) \neq T\restriction_{\boldsymbol{\pi}}(\mathbf{y})] \leq \sum_{\ell \in L} \Pr[\ell(\mathbf{y} \circ \boldsymbol{\pi}) = 1] \cdot (\zeta + \nu) = \zeta + \nu.$$

Now let us bound the complexity of $\mathbf{T}'_0$. Let us argue that for each fixed y, with high probability over $\boldsymbol{\pi}$, the tree $\mathbf{T}'_0(y)$ does not make too many queries to its input y. (Ultimately we want a depth bound that holds for all y simultaneously; that will come later.) Recall that in the first phase, $\mathbf{T}'_0(y)$ simulates $\widetilde{T}(y \circ \boldsymbol{\pi})$ until it reaches a leaf. By the definition of $\widetilde{T}$, the number of queries to y that this simulation requires is equal to the number of blocks $j \in [t]$ such that one of the following three conditions holds:

- $T(y \circ \boldsymbol{\pi})$ queries some surviving variable $x_i \in B_j$, or
- B_j survives and $T(y \circ \boldsymbol{\pi})$ makes at least $1/\beta$ queries to variables in B_j, or

- B_j survives and $T(y \circ \boldsymbol{\pi})$ queries a variable $x_i \in B_j$ such that $(y \circ \boldsymbol{\pi})_i = 0$.

The last condition actually implies the first, because in a surviving block B_j, every non-surviving variable is assigned 1. Meanwhile, by Lemma 7.1, with probability $1 - \nu$, the number of blocks satisfiying one of the first two conditions is at most $O(p \cdot \beta \cdot D + \log(1/\nu))$.

In the second phase, $\mathbf{T}'_0(y)$ simulates $T[\ell, \pi]$ for a particular ℓ and π. This tree has depth D_0, so overall, with probability $1 - \nu$, the tree $\mathbf{T}'_0(y)$ makes $D' = O(p \cdot \beta \cdot D + D_0 + \log(1/\nu))$ queries. Our final tree $\mathbf{T}'$ simulates $\mathbf{T}'_0$, except that if the simulation tries to make $D' + 1$ queries, then the corresponding vertex of $\mathbf{T}'$ is a leaf labeled with the constant 0 function. This only increases the average error by ν, so overall, we get a depth-D' tree with leaves labeled by depth-$(\Delta - 1)$ circuits and with average error $\zeta + 2\nu$. Summing up and using $\nu \leq p$ yields the desired bounds.

Now, finally, let us consider the case that it is $\overline{\boldsymbol{\pi}}$ rather than $\boldsymbol{\pi}$ that is a p-surviving $(1 - \beta)$-biased p-corrupted ν-block projection. We reduce to the previous case using the fact that the class of decision trees with leaves labeled by LTF circuits is closed under negation of inputs. That is, define

$$\overline{T}(x_1, \ldots, x_n) = T(1 - x_1, \ldots, 1 - x_n).$$

Then $\overline{T}$ can also be computed by a depth-D decision tree with leaves labeled by depth-Δ LTF circuits with at most w wires. Therefore, applying what we have already proven, we get an approximator $\overline{\mathbf{T}}'$ for $\overline{T}_{\overline{\boldsymbol{\pi}}}$ under $\mathbf{b}^t_{1-\alpha}$. Now define

$$\mathbf{T}'(y_1, \ldots, y_t) = \overline{\mathbf{T}}'(1 - y_1, \ldots, 1 - y_t).$$

Then $\mathbf{T}'$, like $\overline{\mathbf{T}}'$, can be computed by a depth-D' decision tree with leaves labeled by depth-$(\Delta - 1)$ LTF circuits with at most w wires. Furthermore, if we sample $\mathbf{y} \sim \mathbf{b}^t_\alpha$ independently of $\boldsymbol{\pi}$ and define $\overline{\mathbf{y}} = (1 - \mathbf{y}_1, \ldots, 1 - \mathbf{y}_t)$, then

$$\Pr\left[\mathbf{T}'(\mathbf{y}) \neq T\restriction_{\boldsymbol{\pi}}(\mathbf{y})\right] = \Pr\left[\overline{\mathbf{T}}'(\overline{\mathbf{y}}) \neq \overline{T}(\overline{\mathbf{y}} \circ \overline{\boldsymbol{\pi}})\right] \leq \epsilon. \quad \blacksquare$$

7.4 Iterative analysis of the sequence of projections

In this section, our goal is to apply Proposition 7.5 successively for each projection $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$, arguing that an initial LTF circuit gets simpler and simpler with each projection, ultimately proving Theorem 6.1. We begin by arguing that after applying $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(i)}$, the circuit becomes a shallow decision tree with depth-$(d - i)$ circuits at the leaves. (As a reminder, the notion of approximation in the statement below is specified in Definition 7.4.)

Proposition 7.6 ($\boldsymbol{\pi} = \boldsymbol{\pi}^{(d)} \circ \cdots \circ \boldsymbol{\pi}^{(1)}$ simplifies any LTF circuit of depth d)**.** *Let M be a sufficiently large power of two, let $C \geq 97$ be an integer, assume $d \leq 0.05 \cdot \log_C M$, let $M_i = M^{C^{i-1}}$ for $i = 1, \ldots, d$, and let $\vec{M} = (M_1, \ldots, M_d)$. Let $F_{d+1,\vec{M}}$ be the corresponding* AND-OR *tree defined in Section 4.1, let $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$ be the corresponding projections defined in Section 4.2, let $\mathcal{X}_0$ be the set of input variables of $F_{d+1,\vec{M}}$, and let $n = |\mathcal{X}_0|$. Let f be a depth-d* LTF *circuit on $\mathcal{X}_0$ with w wires. Let $i \leq d$ and let $b = i + 1 \bmod 2$. The projected function $f\restriction_{\boldsymbol{\pi}^{(1\ldots i)}}$ can be approximated under the product distribution $\boldsymbol{\sigma}^{(i+1)}$ with average error ϵ by a decision tree of depth D with leaves labeled by depth-$(d - i)$* LTF *circuits with at most w wires, where*

$$D \leq \frac{M^{-1/96} \cdot w \cdot K^i \cdot \log^K(nw)}{M_1 \cdot M_2 \cdots M_{i-1} \cdot \sqrt{M_i}} \qquad \text{and} \qquad \epsilon \leq M^{-1/48} \cdot i \cdot \log^K(nw)$$

for a suitably large universal constant K.

Proof. The proof is by induction on i. The base case $i = 0$ is trivial, so consider $i > 0$. Let $\mathbf{T}_{i-1}$ be the tree that approximates $f\restriction_{\boldsymbol{\pi}^{(1\ldots i-1)}}$ by the induction hypothesis. Using the assumption $d \le 0.05 \cdot \log_C M$, we have

$$M_i \le M^{C^d} \le M^{M^{0.05}} \le \exp(M^{1/16}/2),$$

for sufficiently large M. Therefore, we may apply Lemma 6.4: conditioned on any fixed values for $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(i-1)}$, either $\boldsymbol{\pi}^{(i)}$ or the complement projection $\overline{\boldsymbol{\pi}}^{(i)}$ is a p_i-surviving $(1-\beta_i)$-biased $(1-q_i)$-corrupted $(\nu_i, 3\nu_i)$-block projection, where $p_i = 2/\sqrt{M_i}$, $\nu_i = 0.5/M_i$, $q_i = 2M_i^{-1/4}$, and β_i is given by Equation (4.2). Therefore, Proposition 7.5 gives a tree $\mathbf{T}_i$ that approximates $\mathbf{T}_{i-1}\restriction_{\boldsymbol{\pi}^{(i)}}$. Let us bound the average error of this approximation:

$$\begin{aligned}
&\Pr\left[f\restriction_{\boldsymbol{\pi}^{(1\ldots i)}}(\boldsymbol{\sigma}^{(i+1)}) \neq \mathbf{T}_i(\boldsymbol{\sigma}^{(i+1)})\right] \\
&\quad\le \Pr\left[f\restriction_{\boldsymbol{\pi}^{(1\ldots i)}}(\boldsymbol{\sigma}^{(i+1)}) \neq \mathbf{T}_{i-1}\restriction_{\boldsymbol{\pi}^{(i)}}(\boldsymbol{\sigma}^{(i+1)})\right] + \Pr\left[\mathbf{T}_{i-1}\restriction_{\boldsymbol{\pi}^{(i)}}(\boldsymbol{\sigma}^{(i+1)}) \neq \mathbf{T}_i(\boldsymbol{\sigma}^{(i+1)})\right].
\end{aligned}$$

We can bound the first term using the completion property (Lemma 5.2) and the induction hypothesis:

$$\begin{aligned}
\Pr\left[f\restriction_{\boldsymbol{\pi}^{(1\ldots i)}}(\boldsymbol{\sigma}^{(i+1)}) \neq \mathbf{T}_{i-1}\restriction_{\boldsymbol{\pi}^{(i)}}(\boldsymbol{\sigma}^{(i+1)})\right] &= \Pr\left[f\restriction_{\boldsymbol{\pi}^{(1\ldots i-1)}}(\boldsymbol{\sigma}^{(i+1)} \circ \boldsymbol{\pi}^{(i)}) \neq \mathbf{T}_{i-1}(\boldsymbol{\sigma}^{(i+1)} \circ \boldsymbol{\pi}^{(i)})\right] \\
&= \Pr\left[f\restriction_{\boldsymbol{\pi}^{(1\ldots i-1)}}(\boldsymbol{\sigma}^{(i)}) \neq \mathbf{T}_{i-1}(\boldsymbol{\sigma}^{(i)})\right] \\
&\le (i-1) \cdot M^{-1/48} \cdot \log^K(nw).
\end{aligned}$$

Meanwhile, the second term is bounded by Proposition 7.5:

$$\begin{aligned}
\Pr\left[\mathbf{T}_{i-1}\restriction_{\boldsymbol{\pi}^{(i)}}(\boldsymbol{\sigma}^{(i+1)}) \neq \mathbf{T}_i(\boldsymbol{\sigma}^{(i+1)})\right] &\le p_i^{1/12} \cdot \beta_i^{-1/6} \cdot \mathrm{polylog}(nw/\nu_i) \\
&= \begin{cases} M_i^{-1/24} \cdot M_{i-1}^{1/12} \cdot \mathrm{polylog}(nw) & \text{if } i > 1 \\ M_i^{-1/24} \cdot 2^{1/6} \cdot \mathrm{polylog}(nw) & \text{if } i = 1 \end{cases} \\
&\le M^{-C^{i-1}/24 + C^{i-2}/12} \cdot \log^K(nw) && \text{for a suitable } K \\
&\le M^{-C^{i-1}/48} \cdot \log^K(nw) && (C \ge 4) \\
&\le M^{-1/48} \cdot \log^K(nw).
\end{aligned}$$

Adding up completes the bound on ϵ.

Now let us bound the depth of $\mathbf{T}_i$, denoted by D_i. Let D_{i-1} be the depth of $\mathbf{T}_{i-1}$ and note that $D_0 = 0$. By Proposition 7.5, we get

$$\begin{aligned}
D_i &\le O(p_i \cdot \beta_i \cdot D_{i-1}) + p_i^{13/12} \cdot w \cdot \beta_i^{-1/6} \cdot \mathrm{polylog}(nw/\nu_i) \\
&\le \begin{cases} \frac{(K/2) \cdot D_{i-1}}{\sqrt{M_i} \cdot \sqrt{M_{i-1}}} + M_i^{-13/24} \cdot M_{i-1}^{1/12} \cdot w \cdot \mathrm{polylog}(nw) & \text{if } i > 1 \\ M_i^{-13/24} \cdot 2^{1/6} \cdot w \cdot \mathrm{polylog}(nw) & \text{if } i = 1 \end{cases}
\end{aligned}$$

for a suitable choice of K. Let us first bound the first term of the $i > 1$ expression. By induction, we have

$$D_{i-1} \le \frac{M^{-1/96} \cdot w \cdot K^{i-1} \cdot \log^K(nw)}{M_1 \cdot M_2 \cdots M_{i-2} \cdot \sqrt{M_{i-1}}}.$$

Therefore, the first term of the $i > 1$ expression is at most

$$\frac{M^{-1/96} \cdot w \cdot K^i \cdot \log^K n}{2 \cdot M_1 \cdot M_2 \cdots M_{i-1} \cdot \sqrt{M_i}}.$$

Now let us bound the $i = 1$ expression and (simultaneously) the second term of the $i > 1$ expression. By a suitable choice of K, each of these terms is bounded by

$$M^{-C^{i-1} \cdot 13/24 + C^{i-2}/12} \cdot w \cdot \log^K(nw),$$

which is at most

$$M^{-C^{i-1} \cdot 25/48} \cdot w \cdot \log^K(nw)$$

since $C \geq 4$. For comparison,

$$\begin{aligned} M_1 \cdot M_2 \cdots M_{i-1} \cdot \sqrt{M_i} &= M^{1+C+C^2+\cdots+C^{i-2}+C^{i-1}/2} \\ &\leq M^{C^{i-1} \cdot (1/2 + 1/(C-1))} \\ &\leq M^{C^{i-1} \cdot 49/96} \qquad (C \geq 97.) \end{aligned}$$

Therefore, each of the terms in question is at most

$$\frac{M^{-C^{i-1}/96} \cdot w \cdot \log^K(nw)}{M_1 \cdot M_2 \cdots M_{i-1} \cdot \sqrt{M_i}}.$$

Overall,

$$D_i \leq \frac{\left(\frac{1}{2} \cdot M^{-1/96} \cdot K^i + M^{-C^{i-1}/96}\right) \cdot w \cdot \log^K(nw)}{M_1 \cdot M_2 \cdots M_{i-1} \cdot \sqrt{M_i}} \leq \frac{M^{-1/96} \cdot K^i \cdot w \cdot \log^K(nw)}{M_1 \cdot M_2 \cdots M_{i-1} \cdot \sqrt{M_i}}. \quad \blacksquare$$

We are now ready to prove Theorem 6.1, restated below for convenience.

Theorem 6.1 (LTF circuits simplify under the projections, restated)**.** *Let M be a sufficiently large power of two, let $d \in \mathbb{N}$, and use the parameters $M_i = M^{100^{i-1}}$ for $i = 1, \ldots, d$ to define the projections $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$. Assume $d \leq 0.05 \cdot \log_{100} M$. Let $\mathcal{X}_0 \cup \{0,1\}$ be the domain of $\boldsymbol{\pi}^{(1)}$, let $n = |\mathcal{X}_0|$, and let f be a depth-d LTF circuit on $\mathcal{X}_0$ with w wires. The probability that the projected function $f\restriction_{\boldsymbol{\pi}^{(1\ldots d)}}$ is ξ-far from constant under the product distribution $\boldsymbol{\sigma}^{(d+1)}$ is at most ξ, where*

$$\xi \leq M^{-1/96} \cdot \lceil w/n \rceil \cdot O(\log n)^{d+O(1)}. \tag{7.6}$$

Proof. We may assume $w \leq n^2$, since otherwise the theorem is trivial. Therefore, $O(\log w) = O(\log n)$, and taking $i = d$, Proposition 7.6 gives a tree $\mathbf{T}$ of depth

$$D \leq \frac{M^{-1/96} \cdot w \cdot 2^{O(d)} \cdot \operatorname{polylog}(n)}{M_1 \cdot M_2 \cdots M_{d-1} \cdot \sqrt{M_d}}$$

that approximates $f\restriction_{\boldsymbol{\pi}^{(1\ldots d)}}$ under $\boldsymbol{\sigma}^{(d+1)}$ with average error

$$\epsilon \leq M^{-1/48} \cdot \operatorname{polylog}(n).$$

By Markov's inequality,

$$\Pr_{\boldsymbol{\pi}^{(1)},\dots,\boldsymbol{\pi}^{(d)}}\left[\Pr_{\boldsymbol{\sigma}^{(d+1)}}[\mathbf{T}(\boldsymbol{\sigma}^{(d+1)}) \neq f\restriction_{\boldsymbol{\pi}^{(1\dots d)}}(\boldsymbol{\sigma}^{(d+1)})] \leq \sqrt{\epsilon}\right] \geq 1-\sqrt{\epsilon}. \tag{7.7}$$

Condition on $\boldsymbol{\pi}^{(1)} = \pi^{(1)},\dots,\boldsymbol{\pi}^{(d)} = \pi^{(d)}$, where $\pi^{(1)},\dots,\pi^{(d)}$ are any projections such that the above good event occurs. Recall that the projections determine the tree $\mathbf{T}$; let T be the tree such that this conditioning implies $\mathbf{T} = T$. Let $\pi^{(1\dots d)} = \pi^{(d)} \circ \cdots \circ \pi^{(1)}$.

Since we are considering $i = d$, we have a decision tree with leaves labeled by "depth-0" LTF circuits, i.e., the leaves are labeled by constants. Let $b = d+1 \bmod 2$, let ℓ be the leaf that T reaches on input $b^{\mathcal{X}_d}$, and let $z \in \{0,1\}$ be the label of ℓ. We will argue that $T(\boldsymbol{\sigma}^{(d+1)})$ reaches ℓ with high probability, and therefore $T(\boldsymbol{\sigma}^{(d+1)})$ outputs z with high probability. Indeed, let $S \subseteq \mathcal{X}_d$ be the set of variables that are queried on the path from the root to ℓ, and note that $|S| \leq D$. Recall that $\boldsymbol{\sigma}^{(d+1)}$ is distributed according to $\mathbf{b}^{\mathcal{X}_d}_{1-\beta_d\to b}$. Therefore,

$$\begin{aligned}
\Pr\left[T(\boldsymbol{\sigma}^{(d+1)}) \neq z\right] &\leq \Pr\left[\boldsymbol{\sigma}^{(d+1)}_S \neq b^S\right] \\
&\leq D \cdot \beta_d && \text{(Union bound)} \\
&\leq M^{-1/96} \cdot \frac{w}{M_1 \cdot M_2 \cdots M_d} \cdot 2^{O(d)} \cdot \operatorname{polylog}(n).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Pr_{\boldsymbol{\sigma}^{(d+1)}}\left[f\restriction_{\pi^{(1\dots d)}}(\boldsymbol{\sigma}^{(d+1)}) \neq z\right] &\leq \sqrt{\epsilon} + M^{-1/96} \cdot \frac{w}{M_1 \cdot M_2 \cdots M_d} \cdot 2^{O(d)} \cdot \operatorname{polylog}(n) \\
&\leq M^{-1/96} \cdot \left\lceil \frac{w}{M_1 \cdot M_2 \cdots M_d} \right\rceil \cdot 2^{O(d)} \cdot \operatorname{polylog}(n) \\
&\leq M^{-1/96} \cdot \lceil w/n \rceil \cdot \frac{f_1 \cdot f_2 \cdots f_{d+1}}{M_1 \cdot M_2 \cdots M_d} \cdot 2^{O(d)} \cdot \operatorname{polylog}(n) \\
&\leq M^{-1/96} \cdot \lceil w/n \rceil \cdot O(\log n)^{d+O(1)},
\end{aligned}$$

where the last line uses Claim 4.3 and the fact that $M_i \leq n$ for every i.

To summarize, we have shown that with probability $1-\sqrt{\epsilon}$ over the choice of $\boldsymbol{\pi}^{(1\dots d)} = \pi^{(1\dots d)}$, the function $f\restriction_{\pi^{(1\dots d)}}$ is ξ-close to a constant under the distribution $\boldsymbol{\sigma}^{(d+1)}$, where $\xi = M^{-1/96} \cdot \lceil w/n \rceil \cdot O(\log n)^{d+O(1)}$. Since $\xi > \sqrt{\epsilon}$, we are done. ■

8 Putting everything together: LTF circuits vs. AND-OR trees

In this section, we will complete the proof of our main theorem (Theorem 1.1), namely, the existence of a depth-$(d+1)$ AND-OR tree F that is average-case hard for depth-d LTF circuits with a super-linear number of wires. We begin with a crude bound on the number of variables in our AND-OR tree $F_{d+1,\vec{M}}$.

Claim 8.1. *Let M be a sufficiently large power of two, let $d \in \mathbb{N}$, and let $M_i = M^{100^{i-1}}$ for $i = 1,\dots,d$. Let $n = |\mathcal{X}_0|$ where $\mathcal{X}_0$ is the set of input variables to the AND-OR tree $F_{d+1,\vec{M}}$ defined in Definition 4.1. Then $n < M^{2\cdot 100^{d-1}}$.*

Proof. We have $n = f_1 \cdot f_2 \cdots f_{d+1}$, where $f_1, \ldots, f_{d+1}$ are the fan-ins given in Definition 4.1. By Claim 4.3,

$$\begin{aligned}
\prod_{i=1}^{d+1} f_i &\leq \log(M) \cdot \left(\prod_{i=2}^{d} 2 \cdot M^{100^{i-2}} \cdot \ln(M^{100^{i-1}}) \right) \cdot 2 \cdot M^{100^{d-1}} \cdot \ln(2) \\
&< \prod_{i=1}^{d} M^{1.1 \cdot 100^{i-1}} && (M \text{ is sufficiently large}) \\
&= M^{1.1 \cdot (100^d - 1)/99} \\
&< M^{2 \cdot 100^{d-1}}. \quad \blacksquare
\end{aligned}$$

Proof of Theorem 1.1. For each $M \in \mathbb{N}$, let $\text{powers}(M) = (M, M^{100}, \ldots, M^{100^{d-1}})$. Let n_M be the number of variables $|\mathcal{X}_0|$ in the AND-OR tree $F_{d+1,\text{powers}(M)}$. Let M be the largest power of two such that $n_M \leq n$, let $M_i = M^{100^{i-1}}$, let $\vec{M} = \text{powers}(M)$, and let $F = F_{d+1,\vec{M}}$. Then we can consider F to be a function on n variables that ignores the last $n - n_M$ input variables.

We must show that M is "sufficiently large," i.e., larger than some universal constant. To do so, we rely on the assumption that n is sufficiently large. Let M_* be the largest power of two that is smaller than (say) $\log n$. By Claim 8.1,

$$n_{M_*} \leq M_*^{2 \cdot 100^{d-1}} < 2^{\sqrt{\log n}} < n,$$

where the second inequality uses the assumption $d \leq \frac{1}{20} \log\log n$. Consequently, $M \geq M_*$, so indeed, M is sufficiently large.

Let $w = n^{1+2^{-9d}}$ and suppose f is a depth-d LTF circuit on n variables with at most w wires. We may assume without loss of generality that f ignores all variables outside $\mathcal{X}_0$. Therefore, by the cumulative completion property (Corollary 5.3), we have

$$\begin{aligned}
\Pr_{\mathbf{x} \in \{0,1\}^n}[f(\mathbf{x}) = F(\mathbf{x})] &= \Pr_{\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}, \boldsymbol{\sigma}^{(d+1)}}[f(\boldsymbol{\sigma}^{(d+1)} \circ \boldsymbol{\pi}^{(1\ldots d)}) = F(\boldsymbol{\sigma}^{(d+1)} \circ \boldsymbol{\pi}^{(1\ldots d)})] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}} \left[\Pr_{\boldsymbol{\sigma}^{(d+1)}} \left[f\restriction_{\boldsymbol{\pi}^{(1\ldots d)}}(\boldsymbol{\sigma}^{(d+1)}) = F\restriction_{\boldsymbol{\pi}^{(1\ldots d)}}(\boldsymbol{\sigma}^{(d+1)}) \right] \right].
\end{aligned}$$

By Theorem 5.1, except with probability $O(M^{-1/4})$ over the choice of $\boldsymbol{\pi}^{(1)}, \ldots, \boldsymbol{\pi}^{(d)}$, the function $F\restriction_{\boldsymbol{\pi}^{(1\ldots d)}}$ is $(6M^{-1/8})$-unbiased under the product distribution $\boldsymbol{\sigma}^{(d+1)}$. Meanwhile, Theorem 6.1 gives a bound on a value ξ such that except with probability ξ, the function $f\restriction_{\boldsymbol{\pi}^{(1\ldots d)}}$ can be approximated under the product distribution $\boldsymbol{\sigma}^{(d+1)}$ with error ξ by a constant function. Fix any $\pi = \boldsymbol{\pi}^{(1\ldots d)}$ such that both of these good events occur. Let $z \in \{0,1\}$ be the constant that approximates $f\restriction_{\pi}$. Then

$$\begin{aligned}
\Pr_{\boldsymbol{\sigma}^{(d+1)}} \left[f\restriction_{\pi}(\boldsymbol{\sigma}^{(d+1)}) = F\restriction_{\pi}(\boldsymbol{\sigma}^{(d+1)}) \right] &\leq \xi + \Pr_{\boldsymbol{\sigma}^{(d+1)}} \left[F\restriction_{\pi}(\boldsymbol{\sigma}^{(d+1)}) = z \right] \\
&\leq \frac{1}{2} + \xi + 6M^{-1/8}.
\end{aligned}$$

Now let us bound ξ. Theorem 6.1 gives the bound

$$\xi \leq M^{-1/96} \cdot \lceil w/n_M \rceil \cdot O(\log n)^{d+O(1)},$$

so we need to bound w/n_M. Recall that $w = n^{1+2^{-9d}}$. By Claim 4.3,

$$\begin{aligned} n < n_{2M} &\le \log(2M) \cdot \left(\prod_{i=2}^{d} 2 \cdot (2M)^{100^{i-2}} \cdot \ln((2M)^{100^{i-1}}) \right) \cdot 2 \cdot (2M)^{100^{d-1}} \cdot \ln(2) \\ &\le 2^d \cdot \left(\prod_{i=1}^{d} 4^{100^{i-1}} \right) \cdot \log(M_1) \cdot \left(\prod_{i=2}^{d} M_{i-1} \cdot \ln(M_i) \right) \cdot M_d \cdot \ln(2) \\ &\le 2^d \cdot 4^{2 \cdot 100^{d-1}} \cdot \log(M_1) \cdot \left(\prod_{i=2}^{d} M_{i-1} \cdot \ln(M_i) \right) \cdot M_d \cdot \ln(2) \qquad (8.1) \\ &\le 4^d \cdot 4^{2 \cdot 100^{d-1}} \cdot n_M. \end{aligned}$$

Therefore,

$$\xi \le M^{-1/96} \cdot n^{2^{-9d}} \cdot 4^{2 \cdot 100^{d-1}} \cdot O(\log n)^{d+O(1)}.$$

Incorporating the chance of getting a bad $\boldsymbol{\pi}^{(1\ldots d)}$, overall, we get

$$\begin{aligned} \Pr_{\mathbf{x} \in \{0,1\}^n}[f(\mathbf{x}) = F(\mathbf{x})] &\le \frac{1}{2} + O(M^{-1/4}) + 2\xi + 6M^{-1/8} \\ &\le \frac{1}{2} + M^{-1/96} \cdot n^{2^{-9d}} \cdot 4^{2 \cdot 100^{d-1}} \cdot O(\log n)^{d+O(1)}. \end{aligned}$$

Now we need to argue that the $M^{-1/96}$ term is small enough to overpower the other terms. By Claim 8.1, $n < n_{2M} \le (2M)^{2 \cdot 100^{d-1}}$, so

$$M \ge \frac{1}{2} n^{0.5 \cdot 100^{-(d-1)}},$$

and hence

$$M^{-1/96} \le O\left(n^{-(0.5/96) \cdot 100^{-(d-1)}} \right) \le O\left(n^{-2^{-8d}} \right).$$

Thus,

$$\begin{aligned} \Pr_{\mathbf{x} \in \{0,1\}^n}[f(\mathbf{x}) = F(\mathbf{x})] &\le \frac{1}{2} + n^{-2^{-8d}} \cdot n^{2^{-9d}} \cdot 4^{2 \cdot 100^{d-1}} \cdot O(\log n)^{d+O(1)} \\ &\le \frac{1}{2} + n^{-2^{-9d}} \cdot 4^{2 \cdot 100^{d-1}} \cdot O(\log n)^{d+O(1)} \\ &\le \frac{1}{2} + n^{-2^{-9d}} \cdot n^{2^{-10d}} \qquad \text{because } d \le \frac{1}{20} \log\log n \\ &\le \frac{1}{2} + n^{-2^{-10d}}. \end{aligned}$$

■

9 Hardness magnification for our construction

We first give an informal overview of the proof of Theorem 1.2. We exploit the recursive structure of $F_{d+1}^{(n)}$. The bottom $d_0 + 1$ layers of $F_{d+1}^{(n)}$ are a collection of subformulas. Each subformula applies the same function (say $\widetilde{F}_{d_0+1}$) to some subset of the input variables. The function $\widetilde{F}_{d_0+1}$ is an AND-OR tree of depth $d_0 + 1$ with fan-ins approximately $\log(M_1), M_1 \ln(M_2), M_2 \ln(M_3), \ldots, M_{d_0} \ln(M_{d_0+1})$, where $M_i = M^{100^{i-1}}$. Thus, $\widetilde{F}_{d_0+1}$ is quite

similar to $F^{(n')}_{d_0+1}$ for some n' that is much smaller than n. The only difference is that the fan-in of the output gate is different in $\widetilde{F}_{d_0+1}$ than it is in $F^{(n')}_{d_0+1}$. In particular, the function computed by $\widetilde{F}_{d_0+1}$ is significantly biased, whereas the function computed by $F^{(n')}_{d_0+1}$ is nearly balanced.

To address this difference, we view $\widetilde{F}_{d_0+1}$ as a *restriction* of $F^{(m)}_{d_0+1}$ for a value m that is slightly larger than n' (but still much smaller than n). By assumption, $F^{(m)}_{d_0+1}$ has depth-d_0 LTF circuits with only m^k wires, hence so does $\widetilde{F}_{d_0+1}$, and hence we can use those LTF circuits to decrease the depth of $F^{(n)}_d$ while barely increasing the number of wires. The rigorous proof follows.

Proof of Theorem 1.2. Like in the proof of Theorem 1.1, let $\text{powers}(M) = (M, M^{100}, \ldots, M^{100^{d-1}})$. Let M be such that $F^{(n)}_{d+1} = F_{d+1,\text{powers}(M)}$ (applied to the first n_M variables where possibly $n_M < n$). Let $\vec{M} = (M_1, \ldots, M_d) = \text{powers}(M)$, and let $f_1, \ldots, f_{d+1}$ be the sequence of fan-ins in $F_{d+1,\vec{M}}$. Like in the proof of Theorem 1.1, we may assume that M is sufficiently large.

Let $K = 1000 \cdot \ln(M)$, let $M' = MK$, let $M'_i = (M')^{100^{i-1}}$, let $\vec{M}' = (M'_1, \ldots, M'_{d_0})$, and let $f'_1, \ldots, f'_{d_0+1}$ be the sequence of fan-ins in $F_{d_0+1,\vec{M}'}$. We claim that for every $i \in [d_0+1]$, we have $f'_i \geq f_i$. Indeed, by Claim 4.3, we have $f'_1 = \log(M'_1) > \log(M_1) = f_1$, and for $i = 2, \ldots, d_0$, we have

$$f'_i \geq \frac{1}{2} \cdot M'_{i-1} \cdot \ln(M'_i) = \frac{1}{2} \cdot K^{100^{i-2}} \cdot M_{i-1} \cdot \ln(M'_i) > 2 \cdot M_{i-1} \cdot \ln(M_i) \geq f_i,$$

and finally

$$\begin{aligned} f'_{d_0+1} \geq \frac{1}{2} \cdot M'_{d_0} \cdot \ln(2) &= \frac{\ln(2)}{2} \cdot M_{d_0} \cdot K^{100^{d_0-1}} \\ &= \frac{\ln(2)}{2} \cdot M_{d_0} \cdot (1000 \ln M)^{100^{d_0-1}} \\ &> 2 \cdot M_{d_0} \cdot 100^{d_0} \cdot \ln(M) \cdot M_{d_0} \\ &= 2 \cdot M_{d_0} \cdot \ln(M_{d_0+1}) \geq f_{d_0+1}. \end{aligned}$$

Therefore, for each gate v of $F_{d+1,\vec{M}}$ at distance d_0+1 from the input, there is a restriction ρ_v such that $F_v = F_{d_0+1,\vec{M}'}|_{\rho_v}$. (The restriction ρ_v fixes some variables to 1 and keeps all other variables alive.) Let m be the number of input variables of $F_{d_0+1,\vec{M}'}$. Then $F_{d_0+1,\vec{M}'} = F^{(m)}_{d_0+1}$, so by assumption, $F_{d_0+1,\vec{M}'}$ can be computed by a depth-d_0 LTF circuit with at most m^k wires. Consequently, $F^{(n)}_{d+1}$ can be computed by a depth-d LTF circuit with at most $O(n \cdot m^k)$ wires.

Now let us bound m, the number of variables in $F_{d_0+1,\vec{M}'}$. We claim that for every $i \in [d_0+1]$, we have

$$f'_i \leq M_{i-1} \cdot 2K^{100^{i-2}} \cdot 100^{i-1} \cdot \ln(KM),$$

defining $M_0 = 1$ for convenience. Indeed, by Claim 4.3, $f'_1 = \log(M'_1) \leq 2\ln(KM)$; for $i = 2, \ldots, d_0$, we have

$$f'_i \leq 2 \cdot M'_{i-1} \cdot \ln(M'_i) = 2 \cdot K^{100^{i-2}} \cdot M_{i-1} \cdot 100^{i-1} \cdot \ln(KM);$$

and finally

$$f'_{d_0+1} \leq 2 \cdot M'_{d_0} \cdot \ln(2) = 2 \cdot K^{100^{d_0-1}} \cdot M_{d_0} \cdot \ln(2).$$

Therefore,

$$\begin{aligned} m = \prod_{i=1}^{d_0+1} f_i' &\le \prod_{i=1}^{d_0+1} M_{i-1} \cdot 2K^{100^{i-2}} \cdot 100^{i-1} \cdot \ln(KM) \\ &= \left(\prod_{i=2}^{d_0+1} M_{i-1} \right) \cdot O(\log n)^{2 \cdot 100^{d_0}} \\ &= \left(\prod_{i=d-d_0+2}^{d+1} M_{i-1} \right)^{100^{-(d-d_0)}} \cdot O(\log n)^{2 \cdot 100^{d_0}} \\ &< \left(\prod_{i=d-d_0+2}^{d+1} f_i \right)^{100^{-(d-d_0)}} \cdot O(\log n)^{2 \cdot 100^{d_0}} \\ &< n^{100^{-(d-d_0)}} \cdot O(\log n)^{2+100^{d_0}}. \end{aligned}$$

Since d_0 and k are constants, it follows that $F_{d+1}^{(n)}$ can be computed by a depth-d LTF circuit with at most $\widetilde{O}(n^{1+k \cdot 100^{-(d-d_0)}})$ wires. Assuming $d \ge 2d_0$, the number of wires is indeed at most $\widetilde{O}(n^{1+k \cdot 100^{-d/2}}) = \widetilde{O}(n^{1+k \cdot 10^{-d}})$. ■

References

[Aar10] Scott Aaronson. "BQP and the polynomial hierarchy". In: *Proc. 42nd Annual ACM Symposium on Theory of Computing (STOC)*. ACM, New York, 2010, pp. 141–150.

[ACW16] Josh Alman, Timothy M. Chan, and Ryan Williams. "Polynomial representations of threshold functions and algorithmic applications". In: *Proc. 57th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2016, pp. 467–476. DOI: 10.1109/FOCS.2016.57.

[AG93] Eric Allender and Vivek Gore. "On strong separations from AC^0". In: *Advances in computational complexity theory (New Brunswick, NJ, 1990)*. Vol. 13. DIMACS Ser. Discrete Math. Theoret. Comput. Sci. Amer. Math. Soc., Providence, RI, 1993, pp. 21–37.

[Agr20] Rohit Agrawal. "Coin theorems and the Fourier expansion". In: *Chicago Journal of Theoretical Computer Science* (2020), Art. 4, 15.

[AH94] Eric Allender and Ulrich Hertrampf. "Depth reduction for circuits of unbounded fan-in". In: *Inform. and Comput.* 112.2 (1994), pp. 217–238. ISSN: 0890-5401.

[AK10] Eric Allender and Michal Koucký. "Amplifying lower bounds by means of self-reducibility". In: *Journal of the ACM* 57.3 (2010), pp. 14, 36.

[All89] Eric Allender. "A note on the power of threshold circuits". In: *Proc. 30th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 1989, pp. 580–584.

[Ama09] Kazuyuki Amano. "Bounds on the size of small depth circuits for approximating majority". In: *Proc. 36th International Colloquium on Automata, Languages and Programming (ICALP)*. 2009, pp. 59–70.

[BGW20] Mark Braverman, Sumegha Garg, and David P. Woodruff. "The coin problem with applications to data streams". In: *Proc. 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Soc., Los Alamitos, CA, 2020, pp. 318–329.

[BGZ22] Mark Braverman, Sumegha Garg, and Or Zamir. "Tight space complexity of the coin problem". In: *Proc. 62nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Soc., Los Alamitos, CA, 2022, pp. 1068–1079.

[BH12] Paul Beame and Trinh Huynh. "Multiparty Communication Complexity and Threshold Circuit Size of AC^0". In: *SIAM Journal of Computing* 41.3 (2012), pp. 484–518.

[BKK+22] Swapnam Bajpai, Vaibhav Krishan, Deepanshu Kush, Nutan Limaye, and Srikanth Srinivasan. "A #SAT algorithm for small constant-depth circuits with PTF gates". In: *Algorithmica* 84.4 (2022), pp. 1132–1162. ISSN: 0178-4617.

[Bop97] Ravi B. Boppana. "The average sensitivity of bounded-depth circuits". In: *Information Processing Letters* 63.5 (1997), pp. 257–261. ISSN: 0020-0190.

[BS92] Jehoshua Bruck and Roman Smolensky. "Polynomial threshold functions, AC^0 functions, and spectral norms". In: *SIAM J. Comput.* 21.1 (1992), pp. 33–42. ISSN: 0097-5397.

[BT15] Mark Bun and Justin Thaler. "Hardness amplification and the approximate degree of constant-depth circuits". In: *Proc. 42nd International Colloquium on Automata, Languages and Programming (ICALP)*. Vol. 9134. Lecture Notes in Comput. Sci. Springer, Heidelberg, 2015, pp. 268–280.

[BT16] Mark Bun and Justin Thaler. "Improved bounds on the sign-rank of AC^0". In: *Proc. 43rd International Colloquium on Automata, Languages and Programming (ICALP)*. Vol. 55. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016, Art. No. 37, 14.

[BT21] Mark Bun and Justin Thaler. "The large-error approximate degree of AC^0". In: *Theory of Computing* 17 (2021), Paper No. 7, 46.

[BT94] Richard Beigel and Jun Tarui. "On ACC". In: *Computational Complexity* 4.4 (1994), pp. 350–366.

[BV10] Joshua Brody and Elad Verbin. "The coin problem, and pseudorandomness for branching programs". In: *Proc. 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2010, pp. 30–39.

[BVW07] Harry Buhrman, Nikolay Vereshchagin, and Ronald de Wolf. "On Computation and Communication with Small Bias". In: *Proc. 22nd Annual IEEE Conference on Computational Complexity (CCC)*. 2007, pp. 24–32.

[CGR14] Gil Cohen, Anat Ganor, and Ran Raz. "Two sides of the coin problem". In: *Proc. 18th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*. Vol. 28. LIPIcs. Leibniz Int. Proc. Inform. 2014, pp. 618–629.

[Cha07] Arkadev Chattopadhyay. "Discrepancy and the Power of Bottom Fan-in in Depth-three Circuits". In: *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2007, pp. 449–458.

[Che19] Lijie Chen. "Non-deterministic Quasi-Polynomial Time is Average-case Hard for ACC Circuits". In: *Proc. 60th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2019. DOI: 10.1109/FOCS.2019.00079.

[CJW20] Lijie Chen, Ce Jin, and Richard Ryan Williams. "Sharp threshold results for computational complexity". In: *Proc. 52nd Annual ACM Symposium on Theory of Computing (STOC)*. 2020, 1335–1348.

[COS18] Ruiwen Chen, Igor C. Oliveira, and Rahul Santhanam. "An average-case lower bound against ACC^0". In: *LATIN 2018: Theoretical informatics*. Vol. 10807. Lecture Notes in Comput. Sci. Springer, Cham, 2018, pp. 317–330. DOI: 10.1007/978-3-319-77404-6_2.

[CP19] Shiteng Chen and Periklis A. Papakonstantinou. "Depth reduction for composites". In: *SIAM J. Comput.* 48.2 (2019), pp. 668–686. ISSN: 0097-5397. DOI: 10.1137/17M1129672. URL: https://doi.org/10.1137/17M1129672.

[CR22] Lijie Chen and Hanlin Ren. "Strong Average-Case Circuit Lower Bounds from Nontrivial Derandomization". In: *SIAM Journal on Computing* 51.3 (2022), STOC20–115–STOC20–173. DOI: 10.1137/20M1364886.

[CSS18] Ruiwen Chen, Rahul Santhanam, and Srikanth Srinivasan. "Average-case lower bounds and satisfiability algorithms for small threshold circuits". In: *Theory of Computing* 14 (2018), Paper No. 9, 55.

[CT19] Lijie Chen and Roei Tell. "Bootstrapping results for threshold circuits 'just beyond' known lower bounds". In: *Proc. 51st Annual ACM Symposium on Theory of Computing (STOC)*. 2019, pp. 34–41.

[DGJ+10] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A. Servedio, and Emanuele Viola. "Bounded independence fools halfspaces". In: *SIAM Journal of Computing* 39.8 (2010), pp. 3441–3462.

[FKL+01] Jürgen Forster, Matthias Krause, Satyanarayana V. Lokam, Rustam Mubarakzjanov, Niels Schmitt, and Hans Ulrich Simon. "Relations Between Communication Complexity, Linear Arrangements, and Computational Complexity". In: *Proc. 21st Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*. 2001, pp. 171–182.

[FMT21] Yuval Filmus, Or Meir, and Avishay Tal. "Shrinkage under random projections, and cubic formula lower bounds for *AC*0". In: *Proc. 12th Conference on Innovations in Theoretical Computer Science (ITCS)*. Vol. 185. 2021, Art. No. 89, 7.

[GII+19] Alexander Golovnev, Rahul Ilango, Russell Impagliazzo, Valentine Kabanets, Antonina Kolokolova, and Avishay Tal. "$AC^0[p]$ lower bounds against MCSP via the coin problem". In: *Proc. 46th International Colloquium on Automata, Languages and Programming (ICALP)*. Vol. 132. LIPIcs. Leibniz Int. Proc. Inform. 2019, Art. No. 66, 15.

[Hås01] Johan Håstad. "A slight sharpening of LMN". In: *Journal of Computer and System Sciences* 63.3 (2001), pp. 498–508. ISSN: 0022-0000.

[Hås87] Johan Håstad. *Computational Limitations for Small-Depth Circuits*. MIT Press, 1987.

[Hås98] Johan Håstad. "The shrinkage exponent of De Morgan formulas is 2". In: *SIAM J. Comput.* 27.1 (1998), pp. 48–64. ISSN: 0097-5397.

[HG91] Johan Håstad and Mikael Goldmann. "On the power of small-depth threshold circuits". In: *Computational Complexity* 1.2 (1991), pp. 113–129.

[HHT+22] Pooya Hatami, William M. Hoza, Avishay Tal, and Roei Tell. "Fooling constant-depth threshold circuits". In: *Proc. 62nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2022 (albeit "FOCS 2021"), pp. 104–115.

[HMP+93] András Hajnal, Wolfgang Maass, Pavel Pudlák, Mario Szegedy, and György Turán. "Threshold Circuits of Bounded Depth". In: *Journal of Computer and System Sciences* 46.2 (1993), pp. 129–154.

[HRS+17] Johan Håstad, Benjamin Rossman, Rocco A Servedio, and Li-Yang Tan. "An average-case depth hierarchy theorem for boolean circuits". In: *Journal of the ACM (JACM)* 64.5 (2017), pp. 1–27.

[IPS97] Russell Impagliazzo, Ramamohan Paturi, and Michael E. Saks. "Size-depth tradeoffs for threshold circuits". In: *SIAM Journal of Computing* 26.3 (1997), pp. 693–707.

[IS01] Russell Impagliazzo and Nathan Segerlind. "Counting axioms do not polynomially simulate counting gates". In: *Proc. 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 2001.

[KKL88] J. Kahn, G. Kalai, and N. Linial. "The influence of variables on Boolean functions". In: *Proc. 29th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 1988, pp. 68–80.

[KL18] Valentine Kabanets and Zhenjian Lu. "Satisfiability and derandomization for small polynomial threshold circuits". In: *Proc. 22nd International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*. 2018, Art. No. 46, 19.

[KP97] Matthias Krause and Pavel Pudlák. "On the computational power of depth-2 circuits with threshold and modulo gates". In: *Theoretical Computer Science* 174.1-2 (1997), pp. 137–156. ISSN: 0304-3975.

[KP98] Matthias Krause and Pavel Pudlák. "Computing Boolean functions by polynomials and threshold circuits". In: *Computational Complexity* 7.4 (1998), pp. 346–370. ISSN: 1016-3328.

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. "Constant depth circuits, Fourier transform, and learnability". In: *Journal of the ACM* 40.3 (1993), pp. 607–620.

[LSS+21] Nutan Limaye, Karteek Sreenivasaiah, Srikanth Srinivasan, Utkarsh Tripathi, and S. Venkitesh. "A fixed-depth size-hierarchy theorem for $\mathrm{AC}^0[\oplus]$ via the coin problem". In: *SIAM Journal of Computing* 50.4 (2021), pp. 1461–1499.

[LST19] Nutan Limaye, Srikanth Srinivasan, and Utkarsh Tripathi. "More on $AC^0[\oplus]$ and Variants of the Majority Function". In: *Proc. 39th Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*. 2019, 22:1–22:14.

[LV18] Chin Ho Lee and Emanuele Viola. "The coin problem for product tests". In: *ACM Transactions on Computation Theory* 10.3 (2018), Art. 14, 10. ISSN: 1942-3454.

[MP69] Marvin Minsky and Seymour Papert. *Perceptrons: an Introduction to Computational Geometry*. MIT Press, 1969. ISBN: 9780262630221.

[MW20] Cody D. Murray and R. Ryan Williams. "Circuit lower bounds for nondeterministic quasi-polytime from a new easy witness lemma". In: *SIAM J. Comput.* 49.5 (2020), STOC18–300–STOC18–322. ISSN: 0097-5397. DOI: 10.1137/18M1195887. URL: https://doi.org/10.1137/18M1195887.

[OS10] Ryan O'Donnell and Rocco A. Servedio. "New degree bounds for polynomial threshold functions". In: *Combinatorica* 30.3 (2010), pp. 327–358. ISSN: 0209-9683.

[OS18] Igor C. Oliveira and Rahul Santhanam. "Hardness magnification for natural problems". In: *Proc. 59th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE Computer Soc., Los Alamitos, CA, 2018, pp. 65–76.

[OW07] Ryan O'Donnell and Karl Wimmer. "Approximation by DNF: examples and counterexamples". In: *Proc. 34th International Colloquium on Automata, Languages and Programming (ICALP)*. Vol. 4596. Lecture Notes in Comput. Sci. Springer, Berlin, 2007, pp. 195–206.

[Raz87] Alexander A. Razborov. "Lower bounds on the size of constant-depth networks over a complete basis with logical addition". In: *Mathematical Notes of the Academy of Science of the USSR* 41.4 (1987), pp. 333–338.

[RS10] Alexander A. Razborov and Alexander A. Sherstov. "The Sign-Rank of AC^0". In: *SIAM Journal of Computing* 39.5 (2010), pp. 1833–1855.

[RW93] Alexander Razborov and Avi Wigderson. "$n^{\Omega(\log n)}$ lower bounds on the size of depth-3 threshold circuits with AND gates at the bottom". In: *Inform. Process. Lett.* 45.6 (1993), pp. 303–307. ISSN: 0020-0190. DOI: 10.1016/0020-0190(93)90041-7.

[Ser07] Rocco A. Servedio. "Every linear threshold function has a low-weight approximator". In: *Computational Complexity* 16.2 (2007), pp. 180–209.

[She09] Alexander A. Sherstov. "Separating AC^0 from Depth-2 Majority Circuits". In: *SIAM Journal of Computing* 38.6 (2009), pp. 2113–2129.

[She11] Alexander A. Sherstov. "The pattern matrix method". In: *SIAM Journal of Computing* 40.6 (2011), pp. 1969–2000. ISSN: 0097-5397.

[She18a] Alexander A. Sherstov. "Breaking the Minsky-Papert barrier for constant-depth circuits". In: *SIAM Journal of Computing* 47.5 (2018), pp. 1809–1857. ISSN: 0097-5397.

[She18b] Alexander A. Sherstov. "The power of asymmetry in constant-depth circuits". In: *SIAM Journal of Computing* 47.6 (2018), pp. 2362–2434. ISSN: 0097-5397.

[Sip83a] Michael Sipser. "A complexity theoretic approach to randomness". In: *Proc. 15th Annual ACM Symposium on Theory of Computing (STOC)*. 1983, pp. 330–335.

[Sip83b] Michael Sipser. "Borel Sets and Circuit Complexity". In: *Proc. 15th Annual ACM Symposium on Theory of Computing (STOC)*. 1983, pp. 61–69.

[Smo87] Roman Smolensky. "Algebraic methods in the theory of lower bounds for Boolean circuit complexity". In: *Proc. 19th Annual ACM Symposium on Theory of Computing (STOC)*. 1987, pp. 77–82.

[Sri03] Aravind Srinivasan. "On the approximability of clique and related maximization problems". In: *Journal of Computer and System Sciences* 67.3 (2003), pp. 633–651.

[Ste13] John Steinberger. "The distinguishability of product distributions by read-once branching programs". In: *Proc. 28th Annual IEEE Conference on Computational Complexity (CCC)*. 2013, pp. 248–254.

[SV10] Ronen Shaltiel and Emanuele Viola. "Hardness amplification proofs require majority". In: *SIAM Journal on Computing* 39.7 (2010), pp. 3122–3154.

[SW21] Alexander A. Sherstov and Pei Wu. "Near-Optimal Lower Bounds on the Threshold Degree and Sign-Rank of AC^0". In: *SIAM Journal of Computing* online ahead of print (2021), STOC19–1–STOC19–86.

[Tal17] Avishay Tal. "Tight Bounds on the Fourier Spectrum of AC0". In: *Proc. 32nd Annual IEEE Conference on Computational Complexity (CCC)*. 2017, 15:1–15:31.

[Tel18] Roei Tell. "Quantified Derandomization of Linear Threshold Circuits". In: *Proc. 50th Annual ACM Symposium on Theory of Computing (STOC)*. 2018, pp. 855–865.

[Tod91] Seinosuke Toda. "PP is as hard as the polynomial-time hierarchy". In: *SIAM J. Comput.* 20.5 (1991), pp. 865–877. ISSN: 0097-5397. DOI: 10.1137/0220053.

[Vio14] Emanuele Viola. "Randomness Buys Depth for Approximate Counting". In: *Computational Complexity* 23.3 (2014), pp. 479–508.

[VW20] Nikhil Vyas and R. Ryan Williams. "Lower bounds against sparse symmetric functions of ACC circuits: expanding the reach of #SAT algorithms". In: *Proc. 37th Symposium on Theoretical Aspects of Computer Science (STACS)*. Vol. 154. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2020, Art. No. 59, 17. DOI: 10.4230/LIPIcs.STACS.2020.59.

[Wil14a] Ryan Williams. "Algorithms for circuits and circuits for algorithms: Connecting the tractable and intractable". In: *Proc. International Congress of Mathematicians (ICM)*. 2014, pp. 659–682.

[Wil14b] Ryan Williams. "Nonuniform ACC circuit lower bounds". In: *J. ACM* 61.1 (2014), Art. 2, 32. ISSN: 0004-5411. DOI: 10.1145/2559903.

[Wil16] R. Ryan Williams. "Natural proofs versus derandomization". In: *SIAM J. Comput.* 45.2 (2016), pp. 497–529. ISSN: 0097-5397. DOI: 10.1137/130938219. URL: https://doi.org/10.1137/130938219.

[Wil18a] R. Ryan Williams. "Faster All-Pairs Shortest Paths via Circuit Complexity". In: *SIAM Journal on Computing* 47.5 (2018), pp. 1965–1985. DOI: 10.1137/15M1024524.

[Wil18b] Richard Ryan Williams. "New algorithms and lower bounds for circuits with linear threshold gates". In: *Theory of Computing* 14 (2018), Paper No. 17, 25. DOI: 10.4086/toc.2018.v014a017.

[Yao85] Andrew C-C. Yao. "Separating the Polynomial-time Hierarchy by Oracles". In: *Proc. 26th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 1985, pp. 1–10.

[Yao89] Andrew Chi-Chih Yao. "Circuits and Local Computation". In: *Proc. 21st Annual ACM Symposium on Theory of Computing (STOC)*. 1989, pp. 186–196.

[Yao90] Andrew Chi-Chih Yao. "On ACC and threshold circuits". In: *Proc. 31st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE Comput. Soc. Press, Los Alamitos, CA, 1990, pp. 619–627.

A Positive results for average-case depth reduction

In this section, we record proofs for the three positive results mentioned in Section 1.3, showing that the correlation bound in our main result (Theorem 1.1) cannot be significantly improved. First we consider the problem of decreasing the depth of a linear-size AC^0 circuit by a single layer.

Proposition A.1. *Let $d \geq 0$ and let F be a depth-$(d+1)$ AC^0 circuit over n input bits with w wires and top fan-in $m \leq w$. There exists a depth-d AC^0 circuit f with at most w wires such that*

$$\Pr_{\mathbf{x}\in\{0,1\}^n}[f(\mathbf{x}) = F(\mathbf{x})] \geq \frac{1}{2} + \frac{1}{4m}.$$

Proposition A.1 is well-known and follows easily from the discriminator lemma of Hajnal, Maass, Pudlak, Szegedy, and Turan [HMP+93]. We include a proof only for completeness. The proof below does not explicitly invoke the discriminator lemma, but it amounts to essentially the same argument.

Proof. Assume without loss of generality that the output gate of F is an OR gate. If $\mathbb{E}[F] \geq \frac{1}{2} + \frac{1}{4m}$, then we can take $f \equiv 1$. Assume now that $\mathbb{E}[F] < \frac{1}{2} + \frac{1}{4m}$. Let $F_1, \ldots, F_m$ be the children of the output gate of F, so $F(x) = \bigvee_{i=1}^{m} F_i(x)$. By the union bound, $\mathbb{E}[F] \leq \mathbb{E}[F_1] + \cdots + \mathbb{E}[F_m]$, so there is some i such that $\mathbb{E}[F_i] \geq \frac{1}{m}\,\mathbb{E}[F]$. Furthermore, $F_i \leq F$, so

$$\begin{aligned}
\Pr_{\mathbf{x}\in\{0,1\}^n}[F_i(\mathbf{x}) = F(\mathbf{x})] &= \Pr_{\mathbf{x}\in\{0,1\}^n}[F(\mathbf{x}) = 0] + \Pr_{\mathbf{x}\in\{0,1\}^n}[F_i(\mathbf{x}) = 1] \\
&\geq 1 - \left(1 - \frac{1}{m}\right) \cdot \mathbb{E}[F] \\
&\geq 1 - \left(1 - \frac{1}{m}\right) \cdot \left(\frac{1}{2} + \frac{1}{4m}\right) \\
&\geq \frac{1}{2} + \frac{1}{4m}.
\end{aligned}$$

Therefore, we can take $f = F_i$. ∎

Next, we consider the problem of approximating an arbitrary monotone function in low depth.

Proposition A.2. *Let F be a monotone Boolean function on n input bits. There exists a Boolean function f that depends on at most one variable such that*

$$\Pr_{\mathbf{x}\in\{0,1\}^n}[f(\mathbf{x}) = F(\mathbf{x})] \geq \frac{1}{2} + \Omega\left(\frac{\log n}{n}\right).$$

Proposition A.2 follows easily from the Kahn-Kalai-Linial theorem [KKL88]. Once again, the implication is well-known; we include a proof only for completeness.

Proof. If $\mathbb{E}[F] > 3/4$ or $\mathbb{E}[F] < 1/4$, then we can take f to be a constant function. Assume now that $\mathbb{E}[F] \in [1/4, 3/4]$. By the Kahn-Kalai-Linial theorem [KKL88], there exists a variable with influence $\Omega((\log n)/n)$, i.e., there exists $i \in [n]$ such that

$$\Pr_{\mathbf{x}\in\{0,1\}^n}[F(\mathbf{x}^{(i\leftarrow 0)}) \neq F(\mathbf{x}^{(i\leftarrow 1)})] \geq \Omega\left(\frac{\log n}{n}\right),$$

where $x^{(i\leftarrow b)}$ is the string obtained from x by setting coordinate i to the value b. We can write

$$\Pr_{\mathbf{x}\in\{0,1\}^n}[F(\mathbf{x}) = \mathbf{x}_i] = \mathop{\mathbb{E}}_{\mathbf{x}\in\{0,1\}^n}\Bigg[\underbrace{\Pr_{\mathbf{b}\in\{0,1\}}[F(\mathbf{x}^{(i\leftarrow \mathbf{b})}) = \mathbf{b}]}_{(*)}\Bigg].$$

For any fixing of $\mathbf{x} \in \{0,1\}^n$, if $F(\mathbf{x}^{(i\leftarrow 0)}) = F(\mathbf{x}^{(i\leftarrow 1)})$, then quantity (*) is $1/2$. On the other hand, if $F(\mathbf{x}^{(i\leftarrow 0)}) \neq F(\mathbf{x}^{(i\leftarrow 1)})$, then because F is monotone, $F(\mathbf{x}^{(i\leftarrow 0)})$ must be 0 and $F(\mathbf{x}^{(i\leftarrow 1)})$ must be 1, so quantity (*) is 1. Therefore,

$$\Pr_{\mathbf{x}\in\{0,1\}^n}[F(\mathbf{x}) = \mathbf{x}_i] = \frac{1}{2} + \frac{1}{2}\Pr_{\mathbf{x}\in\{0,1\}^n}[F(\mathbf{x}^{(i\leftarrow 0)}) \neq F(\mathbf{x}^{(i\leftarrow 1)})] \geq \frac{1}{2} + \Omega\left(\frac{\log n}{n}\right). \quad ∎$$

Finally, we consider the problem of approximating an arbitrary AC^0 circuit (of any constant depth and any polynomial size) by circuits with depth bounded by a universal constant.

Proposition A.3. *Let $d \geq 1$ and let F be a depth-d AC^0 circuit over n input bits with w wires. There exists a depth-1 AC^0 circuit f with at most $O(\log^{d-1} w)$ wires such that*

$$\Pr_{\mathbf{x}\in\{0,1\}^n}[f(\mathbf{x}) = F(\mathbf{x})] \geq \frac{1}{2} + \frac{1}{n^{O(\log^{d-1} w)}}.$$

Proposition A.3 follows easily from the Linial-Mansour-Nisan theorem [LMN93]. (Actually, to get the specific bounds in Proposition A.3, one should use the later work by Boppana [Bop97]; see also later improvements by Håstad [Hås01] and Tal [Tal17].) We include the proof for completeness.

Proof. Let $g(x) = (-1)^{F(x)}$. There is a value $k = O(\log^{d-1} w)$ such that [LMN93; Bop97; Hås01; Tal17]

$$\sum_{\substack{S\subseteq[n]\\|S|>k}} \hat{g}(S)^2 \leq 1/2.$$

By Parseval's theorem, $\sum_{S\subseteq[n]} \hat{g}(S)^2 = 1$, so $\sum_{S\subseteq[n],|S|\leq k} \hat{g}(S)^2 \geq 1/2$. Therefore, there is some $S \subseteq [n]$ with $|S| \leq k$ such that

$$|\hat{g}(S)| \geq \Omega\left(\frac{1}{(n+1)^{k/2}}\right).$$

For each $a \in \{0,1\}^{|S|}$, define $f_a \colon \{0,1\}^n \to \{0,1\}$ by $f_a(x) = 1 \iff x_S = a$. Furthermore, define $h_a = (-1)^{f_a}$. Finally, let $\chi_S(x) = (-1)^{\sum_{i\in S} x_i}$. Then

$$\chi_S = \sum_{a\in\{0,1\}^{|S|}} (-1)^{\sum_i a_i} \cdot f_a = \sum_{a\in\{0,1\}^{|S|}} (-1)^{\sum_i a_i} \cdot \left(\frac{1}{2} - \frac{1}{2}h_a\right) = -\frac{1}{2} \sum_{a\in\{0,1\}^{|S|}} (-1)^{\sum_i a_i} \cdot h_a.$$

Consequently,

$$\begin{aligned}|\hat{g}(S)| = \left|\mathop{\mathbb{E}}_{\mathbf{x}\in\{0,1\}^n}[g(\mathbf{x}) \cdot \chi_S(\mathbf{x})]\right| &= \left|-\frac{1}{2} \sum_{a\in\{0,1\}^{|S|}} (-1)^{\sum_i a_i} \cdot \mathop{\mathbb{E}}_{\mathbf{x}\in\{0,1\}^n}[g(\mathbf{x}) \cdot h_a(\mathbf{x})]\right| \\ &\leq \frac{1}{2} \sum_{a\in\{0,1\}^{|S|}} \left|\mathop{\mathbb{E}}_{\mathbf{x}\in\{0,1\}^n}[g(\mathbf{x}) \cdot h_a(\mathbf{x})]\right|.\end{aligned}$$

Therefore, there is some $a \in \{0,1\}^{|S|}$ and some $\sigma \in \{\pm 1\}$ such that

$$\sigma \cdot \mathop{\mathbb{E}}_{\mathbf{x}\in\{0,1\}^n}[g(\mathbf{x}) \cdot h_a(\mathbf{x})] \geq \frac{2|\hat{g}(S)|}{2^{|S|}} \geq \Omega\left(\frac{1}{2^k \cdot (n+1)^{k/2}}\right).$$

If $\sigma = 1$, define $f = f_a$; otherwise define $f = 1 - f_a$. In either case, $(-1)^f = \sigma \cdot h_a$, so

$$\begin{aligned}\Pr_{\mathbf{x}\in\{0,1\}^n}[f(\mathbf{x}) = F(\mathbf{x})] = \frac{1}{2} + \frac{1}{2} \mathop{\mathbb{E}}_{\mathbf{x}\in\{0,1\}^n}[(-1)^{f(\mathbf{x})} \cdot (-1)^{F(\mathbf{x})}] &= \frac{1}{2} + \frac{1}{2} \mathop{\mathbb{E}}_{\mathbf{x}\in\{0,1\}^n}[\sigma \cdot h_a(\mathbf{x}) \cdot g(\mathbf{x})] \\ &\geq \frac{1}{2} + \frac{1}{n^{O(\log^{d-1} w)}}.\end{aligned}$$

Finally, f is either a conjunction of at most k literals (if $\sigma = 1$) or else a disjunction of at most k literals (if $\sigma = -1$). ∎

B Proofs of concentration bounds

In this section, for completeness, we record the proofs of the variants of the Chernoff bound that we use.

Proof of Corollary 3.10. Assume without loss of generality that $\mu = \mathbb{E}[\sum_{i=1}^{n} \mathbf{x}_i]$ and $\mu > 0$. Taking $\delta = 1 + \frac{3\ln(1/\epsilon)}{\mu} \geq 1$ in Theorem 3.9, we get

$$\begin{aligned}
\Pr\left[\sum_{i=1}^{n} \mathbf{x}_i > 2\mu + 3\ln(1/\epsilon)\right] &\leq \exp\left(-\frac{\delta}{\delta+2} \cdot (\mu + 3\ln(1/\epsilon))\right) \\
&= \exp\left(-\left(1 - \frac{2}{\delta+2}\right) \cdot (\mu + 3\ln(1/\epsilon))\right) \\
&\leq \exp\left(-\frac{1}{3} \cdot 3\ln(1/\epsilon)\right) \\
&= \epsilon. \quad \blacksquare
\end{aligned}$$

Proof of Corollary 3.11. We can calculate

$$\mu_* \cdot (1+\delta) \geq \mu \cdot \frac{1+\delta}{1+\epsilon} = \mu \cdot \left(1 + \frac{\delta - \epsilon}{1+\epsilon}\right) \geq \mu \cdot (1 + \delta/3).$$

By Theorem 3.9, therefore,

$$\begin{aligned}
\Pr\left[\sum_{i=1}^{n} \mathbf{x}_i > \mu_* \cdot (1+\delta)\right] \leq \Pr\left[\sum_{i=1}^{n} \mathbf{x}_i > \mu \cdot (1+\delta/3)\right] &\leq \exp\left(-\frac{(\delta/3)^2}{2+\delta/3} \cdot \mu\right) \\
&\leq \exp\left(-\frac{\delta^2 \cdot \mu_* \cdot (1-\epsilon)}{3^2 \cdot (2+1/3)}\right) \\
&\leq \exp\left(-\frac{\delta^2 \cdot \mu_*}{42}\right).
\end{aligned}$$

Similarly,

$$\mu_* \cdot (1-\delta) \leq \mu \cdot \frac{1-\delta}{1-\epsilon} = \mu \cdot \left(1 - \frac{\delta - \epsilon}{1-\epsilon}\right) \leq \mu \cdot (1 - \delta/2)\,,$$

and hence

$$\begin{aligned}
\Pr\left[\sum_{i=1}^{n} \mathbf{x}_i < \mu_* \cdot (1-\delta)\right] \leq \Pr\left[\sum_{i=1}^{n} \mathbf{x}_i < \mu \cdot (1-\delta/2)\right] &\leq \exp\left(-\frac{(\delta/2)^2}{2} \cdot \mu\right) \\
&\leq \exp\left(-\frac{\delta^2 \cdot \mu_* \cdot (1-\epsilon)}{2^2 \cdot 2}\right) \\
&\leq \exp\left(-\frac{\delta^2 \cdot \mu_*}{16}\right).
\end{aligned}$$

The union bound completes the proof. $\blacksquare$

Proof of Corollary 3.12. Sample $\mathbf{r}_1, \ldots, \mathbf{r}_n \in [0,1]$ independently and uniformly at random. Define $\widetilde{\mathbf{y}}_1, \ldots, \widetilde{\mathbf{y}}_n \in \{0,1\}$ by the following iterative procedure. Having already picked $\widetilde{\mathbf{y}}_1 = y_1, \ldots, \widetilde{\mathbf{y}}_{i-1} = y_{i-1}$, let

$$\widetilde{\mathbf{y}}_i = 1 \iff \mathbf{r}_i \leq \Pr[\mathbf{y}_i = 1 \mid \mathbf{y}_1 = y_1, \ldots, \mathbf{y}_{i-1} = y_{i-1}].$$

Observe that $(\widetilde{\mathbf{y}}_1, \ldots, \widetilde{\mathbf{y}}_n)$ is distributed identically to $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$.

Next, for each $i \in [n]$, define $\mathbf{z}_i \in \{0,1\}$ by

$$\mathbf{z}_i = 1 \iff \mathbf{r}_i \leq \zeta.$$

Observe that $\mathbf{z}_1, \ldots, \mathbf{z}_n$ are independent and $\Pr[\mathbf{z}_i = 1] = \zeta$.

We claim that (with probability 1) for every $i \in [n]$, we have $\widetilde{\mathbf{y}}_i \leq \mathbf{z}_i$. To see it, let $y_1, \ldots, y_{i-1}$ be arbitrary, and let $\mathcal{X}$ be the set of $x \in \mathrm{Supp}(\mathbf{x}_{i-1})$ such that

$$\mathbf{x}_{i-1} = x \implies \mathbf{y}_1 = y_1, \ldots, \mathbf{y}_{i-1} = y_{i-1}.$$

(Recall that $\mathbf{x}_{i-1}$ determines $\mathbf{y}_1, \ldots, \mathbf{y}_{i-1}$.) Then

$$\begin{aligned}
\Pr[\mathbf{y}_i = 1 \mid \mathbf{y}_1 = y_1, \ldots, \mathbf{y}_{i-1} = y_{i-1}] &= \Pr[\mathbf{y}_i = 1 \mid \mathbf{x}_{i-1} \in \mathcal{X}] \\
&= \frac{\sum_{x \in \mathcal{X}} \Pr[\mathbf{x}_{i-1} = x] \cdot \Pr[\mathbf{y}_i = 1 \mid \mathbf{x}_{i-1} = x]}{\sum_{x \in \mathcal{X}} \Pr[\mathbf{x}_{i-1} = x]} \\
&\leq \frac{\zeta \cdot \sum_{x \in \mathcal{X}} \Pr[\mathbf{x}_{i-1} = x]}{\sum_{x \in \mathcal{X}} \Pr[\mathbf{x}_{i-1} = x]} \\
&= \zeta.
\end{aligned}$$

Thus, indeed, $\widetilde{\mathbf{y}}_i = 1 \implies \mathbf{z}_i = 1$.

Therefore,

$$\Pr\left[\sum_{i=1}^{n} \mathbf{y}_i > 2\zeta \cdot n + 3\ln(1/\epsilon)\right] \leq \Pr\left[\sum_{i=1}^{n} \mathbf{z}_i > 2\zeta \cdot n + 3\ln(1/\epsilon)\right] \leq \epsilon,$$

where the last step follows from Corollary 3.10. ∎

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il