Tight Correlation Bounds for Circuits Between AC0 and TC0

Vinayak M. Kumar
Department of Computer Science, University of Texas at Austin, TX, USA

Abstract

We initiate the study of generalized AC^0 circuits comprised of arbitrary unbounded fan-in gates which only need to be constant over inputs of Hamming weight $\geq k$ (up to negations of the input bits), which we denote $\mathsf{GC}^0(k)$. The gate set of this class includes biased LTFs like the k-OR (outputs 1 iff $\geq k$ bits are 1) and k-AND (outputs 0 iff $\geq k$ bits are 0), and thus can be seen as an interpolation between AC^0 and TC^0.

We establish a tight multi-switching lemma for $\mathsf{GC}^0(k)$ circuits, which bounds the probability that several depth-2 $\mathsf{GC}^0(k)$ circuits do not simultaneously simplify under a random restriction. We also establish a new depth reduction lemma such that coupled with our multi-switching lemma, we can show many results obtained from the multi-switching lemma for depth-d size-s AC^0 circuits lifts to depth-d size-$s^{.99}$ $\mathsf{GC}^0(.01 \log s)$ circuits with *no loss in parameters* (other than hidden constants).

Our result has the following applications:

- Size-$2^{\Omega(n^{1/d})}$ depth-d $\mathsf{GC}^0(\Omega(n^{1/d}))$ circuits do not correlate with parity (extending a result of Håstad (SICOMP, 2014)).
- Size-$n^{\Omega(\log n)}$ $\mathsf{GC}^0(\Omega(\log^2 n))$ circuits with $n^{.249}$ arbitrary threshold gates or $n^{.499}$ arbitrary symmetric gates exhibit exponentially small correlation against an explicit function (extending a result of Tan and Servedio (RANDOM, 2019)).
- There is a seed length $O((\log m)^{d-1} \log(m/\varepsilon) \log\log(m))$ pseudorandom generator against size-m depth-d $\mathsf{GC}^0(\log m)$ circuits, matching the AC^0 lower bound of Håstad up to a $\log\log m$ factor (extending a result of Lyu (CCC, 2022)).
- Size-m $\mathsf{GC}^0(\log m)$ circuits have exponentially small Fourier tails (extending a result of Tal (CCC, 2017)).

2012 ACM Subject Classification Theory of computation → Circuit complexity; Theory of computation → Pseudorandomness and derandomization

Keywords and phrases AC^0, TC^0, Switching Lemma, Lower Bounds, Correlation Bounds, Circuit Complexity

Digital Object Identifier 10.4230/LIPIcs.CCC.2023.18

Related Version *Full Version*: https://eccc.weizmann.ac.il/report/2023/045/

Funding *Vinayak M. Kumar*: Supported by NSF Grant CCF-2008076 and a Simons Investigator Award (#409864, David Zuckerman).

Acknowledgements The author thanks David Zuckerman and Chin Ho Lee for many valuable discussions, Xin Lyu for explaining his work in [18], anonymous reviewers for valuable feedback and for pointing us to the construction presented in Theorem 54, and Jeffrey Champion, Shivam Gupta, Michael Jaber and Jiawei Li for helpful comments.

1 Introduction

Proving superpolynomial circuit lower bounds against explicit functions is one of the most central questions in complexity theory. However, after the initial flurry of work resulting in Blum's lower bound of $3n - o(n)$ [5], followed by a recent revival 30 years later leading to the state of the art $3.1n - o(n)$ size lower bound by Li and Yang [15], this problem has proven to be extremely difficult. Furthermore, there are various proof barriers that give strong evidence that our current intuition is not developed enough to tackle this problem [4, 25, 1].

In order to gain more understanding on this problem, researchers considered circuits with constant depth whose gates are AND, OR, or NOT with unbounded fanin. To this end there has been a fruitful line of work culminating in the state of the art average case hardness of depth d size $2^{\Omega(n^{\frac{1}{d-1}})}$ AC^0 circuits computing the parity and majority functions [7, 9] (with this result being tight for parity). A natural followup question to ask is how powerful AC^0 would then be if $\oplus$ (parity) or MAJ (majority) gates were added, corresponding to the circuit classes $\mathsf{AC}^0[\oplus]$ and TC^0. With regard to $\mathsf{AC}^0[\oplus]$, Oliveira, Santhanam, and Srinivasan [21] proved that it is average case hard for any size-$2^{\Omega(n^{1/(2d-4)})}$ $\mathsf{AC}^0_d[\oplus]$ circuit to compute MAJ, improving earlier work by Razborov [24]. Smolensky [29] proved exponential size lower bounds even if one replaces the $\oplus$ gate with a MOD_p gate for prime p (MOD_p is the gate that outputs 0 iff p divides the sum of the input bits.).

As we see, MAJ is a hard function demonstrating exponential circuit lower bounds for almost all the circuit classes mentioned thus far, and so we would guess TC^0 is extremely powerful and thus challenging to show circuit lower bounds for. This is evident in the current state of the art for TC^0, which is stark in contrast with the landscape of $\mathsf{AC}^0[\oplus]$. In 1993, Impagliazzo, Paturi, and Saks [10] showed that parity is hard for depth-d size-$\Omega(n^{1+\varepsilon_{IPS}^{-d}})$ circuits for some constant $C_{IPS} > 1$, which remains as the current state of the art modulo the case of $d = 2$, where Kane and Williams established a $n^{2.49}$-size lower bound [11]. In fact, a bootstrapping result by Chen and Tell [6] shows that if one slightly improves (e.g. decreases C_{IPS}) this superlinear lower bound against certain NC^1-complete problems (NC^1 is the class of $O(\log n)$-depth, polysized, constant fan-in circuits), we would immediately get superpolynomial lower bounds and a separation of TC^0 and NC^1, attesting to the hardness of this task.

Due to the halted state of affairs for TC^0 circuits, we study a circuit class not as strong as TC^0, but still captures the motivation of analyzing "AC^0 with the power of majority." After it had been shown AC^0 circuits cannot efficiently compute the majority of n bits, it seemed natural that the next step would be to add unbounded MAJ gates to AC^0 to create TC^0. However, due to having unbounded fan-in, TC^0 gives a size-s circuit the power to calculate the majority of up to s bits. Hence, one could argue the reason why size s TC^0 circuits are much harder to analyze than AC^0 is because they are getting much more power than simply calculating the majority of n bits when $s \gg n$. In order to maintain the unbounded fan-in property of the circuit but also ration the computational power we give AC^0 to be "just sufficient" to compute the majority of n bits, one can consider the following circuit class.

▶ **Definition 1** ($\mathsf{AC}^0(k)$ Circuits)**.** *Define the unbounded fan-in gates k-OR to output 1 iff there are at least k ones in the input string, and k-AND to output 0 iff there are at least k zeros in the input string. Define the class of constant depth circuits created by negations and $\{k'\text{-AND}, k'\text{-OR}\}$s for $k' \leq k$ to be $\mathsf{AC}^0(k)$.*

One can observe that $\mathsf{AC}^0(n/2)$ is a natural circuit class that contains the majority of n bits and doesn't add "extra power" like the majority of a much larger quantity of bits. Therefore, analyzing $\mathsf{AC}^0(n/2)$ will give us a better understanding on how much power majority gives to circuits. More generally, $\mathsf{AC}^0(k)$ also allows us to nicely interpolate between AC^0 and TC^0, since a size-s AC^0 circuit is characterized by $\mathsf{AC}^0(1)$, while a size-s TC^0 circuit is characterized by $\mathsf{AC}^0(s/2)$. Hence, studying $\mathsf{AC}^0(k)$ for increasing k is a necessary step and a compelling intermediary model that can help us understand the power of TC^0.

For how large of a k will $\mathsf{AC}^0(k)$ trivially collapse to AC^0? An immediate observation is that $\mathsf{AC}^0(k)$ contains the majority gate over $2k$ bits, for which we know $2^{\Omega(k^{1/2d})}$-size AC^0 lower bounds. Hence, for $k = \mathsf{polylog}(n)$, we have a superpolynomial size seperation between

AC^0 and $\mathsf{AC}^0(k)$. Even for any $k = \omega(1)$, it is unknown whether $\mathsf{AC}^0(k)$ is equivalent to AC^0. A standard argument would be to represent a k-OR with fan-in m as a width-k DNF with $\binom{m}{k}$ clauses (check over all size k subsets of input bits to see if some subset are all 1s) or a width-$(m-k)$ CNF with $\binom{m}{k}$ clauses (check over all size $m-k$ subsets of input bits to see if all subsets contain some 1). Therefore, if we have a size s circuit made from k-OR and k-AND gates, we can turn this into an AC^0 circuit with size $s \cdot \binom{s}{k} \approx s^k$ (since a gate in the original circuit can have fan in size up to s) and depth $d+1$ (one can naively get depth $2d$, but by alternating CNFs and DNFs, we can collapse the depth to $d+1$). Hence, we see that a size s lower bound for AC^0_d translates to a size $s^{1/k}$ lower bound for $\mathsf{AC}^0_{d-1}(k)$. This reduction is not an equivalence, as we pay with a reduction in depth, as well as an asymptotically weaker size lower bound for any $k = \omega(1)$. For example, a polynomial size bound for AC^0 cannot be converted to a polynomial size lower bound for $\mathsf{AC}^0(k)$ for any superconstant k. Consequently, the relationship between AC^0 and $\mathsf{AC}^0(k)$ already becomes nontrivial in the mild regime of $k = \omega(1)$.

In this paper, we study an *even more* general class of circuits, which we denote as $\mathsf{GC}^0(k)$.

▶ **Definition 2** ($\mathsf{G}(k)$ gates/$\mathsf{GC}^0(k)$ circuits)**.** *Let $\mathsf{G}(k)$ be the set of all unbounded fan-in gates that are constant over all input bits with $\geq k$ ones, or over all input bits with $\geq k$ zeros (notice for $k \geq 1$ this includes negations by definition). Define $\mathsf{GC}^0(k)$ to be the class of constant depth circuits created from $\mathsf{G}(k)$ gates.*

Some concrete examples of $\mathsf{G}(k)$ gates are arbitrary gates of fan-in k, majority of $2k$ bits, the k-OR, and functions that compute parity if the input has $< k$ ones, and is 0 otherwise. Notice that this is indeed a generalization of $\mathsf{AC}^0(k)$.

On top of being an alternative generalization of AND/OR gates which may be of independent interest, one nice property about $\mathsf{G}(k)$ is that it includes a generalized notion of k-AND and k-OR gates to arbitrary LTFs (functions of the form $\operatorname{sgn}(\sum w_i x_i - \theta)$).

▶ **Definition 3** (k-balanced LTFs)**.** *Let $f(x) = sgn(\sum_{i=1}^n w_i x_i - \theta)$ be an* LTF*, and let $\sigma : [n] \to [n]$ be a permutation sorting the w_i in increasing magnitude (i.e. $|w_{\sigma(1)}| \leq \cdots \leq |w_{\sigma(n)}|$). We say f is k-*balanced *if k is the smallest index j such that $-\sum_{i \leq j} |w_{\sigma(i)}| + \sum_{i > j} |w_{\sigma(i)}| < |\theta|$.*

One can verify that k-balanced LTFs are indeed in $\mathsf{G}(k)$ (see Theorem 51). Therefore, our results can also be seen as a study of arbitrary LTFs that are biased.

Various notions of *balancedness* (or *regularity* in some literature) for LTFs has been defined in previous work about threshold functions [26, 22, 8], but are all distinct from the combinatorial definition we have proposed. In light of being able to show lower bounds for this characterization of *balanced*, it may be of interest to explore this class of balanced LTFs in other contexts regarding LTF circuit complexity.

1.1 Our Results

We outline all the results we obtain regarding $\mathsf{AC}^0(k)$ (or more generally $\mathsf{GC}^0(k)$) circuits. The core result from which all the other results are derived from is an optimal multi-switching lemma for $\mathsf{GC}^0(k)$ circuits. We state the result without getting into the fine-grained definitions.

▶ **Theorem 4** (Multi-Switching Lemma for $\mathsf{GC}(k)$ Circuits (Informal))**.** *Let $\mathcal{F} = \{F_1, \ldots, F_m\}$ be a list of $\mathsf{G}(k) \circ \mathsf{AND}_w$ circuits on $\{0,1\}^n$. Then*

$$\Pr_{\rho \sim R_p}[\mathcal{F}|_\rho \text{ do not all simultaneously "simplify"}] \leq (2^k m)^{t/r} (O(pw))^t$$

The theorem statement and proof is formally written in Theorem 20. This bound can be proven to be optimal in the regime of large t. See the appendix (Theorem 52) for the proofs of this claim.

It is illuminating to compare this result to the multi-switching lemma for AC^0 circuits, which bounds the probability by the very similar expression of $m^{t/r}(O(pw))^t$. The only difference is that in the new lemma, the m and 2^k are coupled in the base of the exponent. This seems to hint that as long as $2^k = O(m)$, one gets the same probability bound when using either the AC^0 or $\mathsf{GC}^0(k)$ version of the multi-switching lemma. In practice, the parameter m is upper bounded by s, the size of the circuit. Hence, we would intuit that any result obtained from the multi-switching lemma for depth d size s AC^0 circuits can then be lifted to size s $\mathsf{GC}^0_d(\log s)$ circuits. This indeed turns out to be the case as we demonstrate through four different results. We obtain a surprising lifting theorem: any depth d size s AC^0 lower bound obtained by the multi-switching lemma immediately lifts to depth d size $s^{.99}$ $\mathsf{GC}^0(.01 \log s)$-circuits *with no loss in parameters*. We demonstrate three different results exhibiting this phenomenon.

For the first result, denote PAR to be the parity gate.

▶ **Theorem 5** (Optimal Correlation Bounds Against Parity). *Let C be a size m depth d $\mathsf{GC}^0(k)$-circuit. Then the correlation of C against parity is*

$$|\mathbb{E}_x[(-1)^{C(x)+\mathsf{PAR}(x)}]| \leq 2^{-\Omega_d(n/(k+\log m)^{d-1})+k}.$$

In particular, we get a $2^{\Omega(n^{1/d})}$-size lower bound for $\mathsf{GC}^0_d(\Omega(n^{1/d}))$ circuits almost matching the lower bound of $2^{\Omega(n^{1/(d-1)})}$ we know for AC^0! This is especially surprising in light of the fact that $\mathsf{GC}^0_d(\Omega(n^{1/d}))$ is a much stronger class than AC^0; there exist singleton $\mathsf{G}(n^{1/d})$ gates that cannot be computed by size $O(2^{n^{1/2d}})$ AC^0_d circuits. This can be seen as a limited dual result to [24], who showed AC^0_d augmented with parity gates requires size $2^{\Omega(n^{1/2d})}$ to compute majority, whereas we show AC^0_d augmented with $n^{1/d}$-biased majority gates requires size $2^{\Omega(n^{1/d})}$ to compute parity. It also contrasts with [21], who surprisingly showed that adding parity gates to AC^0 improved optimal circuit constructions of majority. Here, we show that majority gates whose threshold value is shifted to $\Omega(n^{1/d})$ has no effect on AC^0's ability to calculate parity, even though such gates adds a lot of power to AC^0. (majority gates whose threshold has been biased to $n^{1/d}$ cannot be computed by size $2^{\Omega(n^{1/2d^2})}$ AC^0_d circuits).

Notice that this result is tight in an extremely sensitive way. Letting PAR_n denote the parity gate over n bits, we see $\mathsf{PAR}_{n^{1/d}} \in \mathsf{G}(n^{1/d})$, and we can calculate the parity of n bits by creating a depth d $n^{1/d}$-ary tree of $\mathsf{PAR}_{n^{1/d}}$ gates, where the ith layer from the bottom has $n^{1-i/d}$ $\mathsf{PAR}_{n^{1/d}}$ gates that take the parity of all the bits fed below it in blocks of $n^{1/d}$. This is a depth d size $O(n^{1-1/d})$ circuit computing parity. Therefore, we have a simple counterexample of a $\mathsf{GC}^0_d(n^{1/d})$ circuit computing parity (which is sublinear in size!). This demonstrates a sharp threshold behavior where the exponential lower bound of $2^{\Omega(n^{1/d})}$ is tight up to the hidden constant factor of the $\Omega(\cdot)$ in $\mathsf{GC}^0_d(\Omega(n^{1/d}))$, and if the constant is too large, we suddenly go from requiring exponentially large circuits to only needing sublinear size ones.

This theorem is tight in all other parameters as well. We show that this result is tight in the size parameter by giving a size-$2^{\Omega(n^{1/d})}$ $\mathsf{GC}^0(.1n^{1/d})$ circuit computing parity. Furthermore, we show that the correlation bound is tight by giving a size-m $\mathsf{GC}^0(k)$ circuit that approximates parity.

For what k will analyzing $\mathsf{AC}^0(k)$ give implications for TC^0? A result by Allender and Koucký ([2], Theorem 3.8) states that there exists an absolute constant C_{AK} such that MAJ_n can be written as an $\mathsf{AC}^0(n^\varepsilon)$ circuit with depth $\leq C_{AK}/\varepsilon$ and size $O(n^{1+\varepsilon})$. Therefore, beating the current state of the art depth d size $\Omega(n^{1+C_{IPS}^{-d}})$ lower bound for TC^0 reduces to beating depth $C_{AK}d/\varepsilon$ size $n^{(1+50^{-d})(2+\varepsilon)}$ lower bounds for $\mathsf{GC}^0(n^\varepsilon)$ circuits for any choice of ε. In our paper, we show exponential size lower bounds against parity when $\varepsilon = 1/d$ but at depth d rather than $C_{AK}d^2$. It would be interesting to see whether with some NC^1-complete problem can display strong lower bounds for $\mathsf{AC}^0(n^{1/d})$ for depth larger than d, even if it may be less than $C_{AK}d^2$ (but a function other than parity would need to be considered).

Another angle researchers have taken towards understanding the power of threshold circuits has been to start with an AC^0 circuit and augment some of the gates to arbitrary threshold gates [33, 17, 27]. Our multi-switching lemma shows that we can instead start with a base $\mathsf{GC}^0(\log s)$ circuit and obtain the same state of the art parameters as [27] if we started from an AC^0 circuit.

▶ **Theorem 6.** *There exists a function* $\mathsf{RW} \in \mathsf{P}$ *(introduced by Razborov and Widgerson [23]) and absolute constant* τ *such that for* C*, a size* $n^{\Omega(\log n)}$ $\mathsf{GC}^0(\Omega(\log^2 n))$*-circuit with* $n^{.249}$ THR *gates, we have*

$$|\mathbb{E}_x[(-1)^{\mathsf{RW}(x)+C(x)}]| \leq 2^{-\tilde{\Omega}(n^{.249})}.$$

The original motivation to study AC^0 with a small number of THR gates was to use this to gradually convert circuits gate by gate from AC^0 to TC^0. This result "speeds up" this process by augmenting *all* AC^0 gates to $\mathsf{G}(\log^2 n)$ gates (which contain unbalanced LTFs as discussed above). If one tried proving this theorem by expanding the $\mathsf{GC}^0(\log^2 n)$ circuits into an AC^0 circuit, completing the proof would require solving a longstanding open problem regarding correlation bounds against $\omega(\log n)$-party NOF protocols! In Section 4.2, we point out this observation explicitly along with the proof.

As another application, we can create PRGs for $\mathsf{GC}^0(\log m)$ circuits whose seed length matches that of size m AC^0 circuits. This is accomplished by fully derandomizing Theorem 4 and using the partition-based PRG approach in [18]. The resulting PRG for $\mathsf{GC}^0_d(\log m)$ has identical seed length as Lyu's PRG, thereby also matching Håstad's AC^0 lower bound barrier up to a $\log\log m$ factor (see [31, 28, 12] for a discussion on why an $o(\log^d(m/\varepsilon))$ seed length implies breakthrough circuit lower bounds).

▶ **Theorem 7.** *For every* $m, n, d \geq 3$ *and* $\varepsilon > 0$*, there is an* ε*-PRG for size-*m $\mathsf{GC}^0_d(\log m)$ *with seed length* $O(\log^{d-1}(m)\log(m/\varepsilon)\log\log m)$

The proof is covered in Section 4.3. Notice that if we had simply expanded out all gates as width $\log m$ CNF/DNFs, we would have a size $\approx m^{\log m}$ AC^0_{d+1} circuit, and plugging in Lyu's near-optimal PRG would yield us a suboptimal seed length of $O((\log^2 m + \log(1/\varepsilon))\log^{2d} m \log\log m)$.

Finally, we establish results on the Fourier spectrum of $\mathsf{GC}^0(k)$ circuits. It can be shown that every Boolean function, when written as a map $\{\pm1\}^n \to \{\pm1\}$, can be uniquely expressed as a multivariate polynomial $f(x) = \sum_{S\subset[n]} \widehat{f}(S)\prod_{i\in S} x_i$. We show exponentially small Fourier tail bounds for any $C \in \mathsf{GC}^0(k)$. More concretely,

▶ **Theorem 8.** *For arbitrary* $C \in \mathsf{GC}^0_d(k)$ *of size* m*, the following is true for any* $0 \leq \ell \leq n$*.*

$$\sum_{|S|\geq\ell} \widehat{C}(S)^2 \leq 2^{-\Omega\left(\frac{\ell}{(k+\log m)^{d-1}}\right)+k}.$$

Linial, Kushilevitz, Mansour, and Tal [13, 16, 19, 30] showed that with small Fourier tails, one can get a variety of Fourier structure results, efficient learning algorithms, and correlation bounds. We demonstrate applications of such techniques to $\mathsf{GC}^0(k)$ in detail in Section 4.4.

1.2 The Switching Lemma

To grasp this section, an understanding of Lyu's witness/transcript proof of the switching lemma [18] would be helpful. We still give an overview of the proof here and provide intuition from an information-theoretic lens, which differs in certain places than the intuition presented in [18]. After the overview, we will highlight the necessary changes needed to prove the more general lemma for $\mathsf{GC}^0(k)$ circuits.

Say we have a k-$\mathsf{OR} \circ \mathsf{AND}$ circuit F (in the formal proof, we consider general $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}$ circuits). For $\Lambda \subset [n]$ and $z \in \{0,1\}^n$, denote $\rho(\Lambda, z)$ to be the restriction/partial assignment where all variables whose indices are in Λ are kept alive/unfixed, and all remaining variables x_i with $i \notin \Lambda$ are fixed to the corresponding bit in z, z_i. Consider a random restriction $\rho(\Lambda, z)$, where $z \sim \{0,1\}^n$ is a uniformly random ground assignment, and Λ is a random subset of $[n]$ such that each element is added with probability p. To show that with low probability, $F|_\rho$ has decision tree depth $\geq t$, it suffices to create a specific *canonical decision tree* (CDT) for each ρ and argue that this tree has depth $\geq t$ with low probability (because if the decision tree depth of $F|_\rho$ is $\geq t$, then surely the canonical decision tree has depth $\geq t$). We consider the following CDT, where we first initialize a counter $ctr \leftarrow 0$, and then scan the bottom layer clauses from left to right.

- If the clause is fixed to 1 and $ctr = k$, terminate since we know that F evaluates to 1. Otherwise increment ctr and move to the next clause.
- If the clause is fixed to 0, move to the next clause.
- If the clause is ambiguous, query all variables in the clause, and behave accordingly as above.

If we think of our CDT as an algorithm that queries certain bits of the input, then *bad* ρ that creates a depth $\geq t$ CDT will produce a unique "transcript" of large size recording the behavior of CDT (i.e. the clauses and variables the CDT queries from). Like [18], we consider transcripts that store (ℓ_i), the indices of the clauses queried, along with a set P that further elucidates which variables in the clauses were queried in an information-efficient manner. We get the following inequality

$$\begin{aligned}\Pr_\rho[\mathsf{DT}(F|_\rho) \geq t] &\leq \Pr_\rho[\mathsf{CDT}(F|_\rho) \geq t] \\ &\leq \sum_{\substack{\text{large transcripts} \\ (\ell_i, P)}} \Pr_\rho[(\ell_i, P) \text{ is a large transcript for } \rho] \qquad (1)\end{aligned}$$

via the union bound. A natural thought is to then bound each term in the sum. Unfortunately, it turns out that the number of transcripts (ℓ_i, P), when counted naively by multiplying the total possible lists (ℓ_i) by the total possible sets P, is far too large to get our switching lemma due to the vast amount of possible (ℓ_i).

However, it turns out that (ℓ_i) contains *redundant* information. Say P is a partial transcript for ρ if it can be completed with a suitable (ℓ_i) to form a transcript for ρ. We can show that given ρ and P that is a partial transcript for it, there is a unique list (ℓ_i) that completes P to a full transcript. Hence

$$\sum_{\substack{\text{large transcripts} \\ (\ell_i, P)}} \Pr_{\rho}[(\ell_i, P) \text{ is a large transcript for } \rho]$$
$$= \sum_{\text{partial transcripts } P} \Pr_{\rho}[P \text{ is a partial transcript for } \rho] \tag{2}$$

which is a sum of far fewer terms, making the union bound feasible. It remains to bound each individual term in the sum.

We want to bound the probability a particular P is a partial transcript for ρ. If we were given the complementary (ℓ_i)'s, this would be easy. The (ℓ_i) along with P would give a transcript of the specific set of $\geq t$ variables that the CDT queried, which ρ must keep alive in order to have any hope of (ℓ_i, P) being a transcript for ρ. This would happen with probability $\leq p^t$, which is a sufficiently small probability to apply the union bound. However, the trickiness arises due to ℓ_i not being specified. It turns out different (ℓ_i) might couple with the same P to form transcripts for different ρ! Therefore if we use no information about ρ, then we have no hope of recovering a fixed (ℓ_i).

On the other hand, if we were given complete information about ρ, then we can recover a unique (ℓ_i) or deduce none exists. However, this eliminates all randomness of ρ and we get the trivial large upper bound of 1 for each term. Therefore, for such an approach to work, we need to condition on partial information about ρ and hope that it is enough information to recover (ℓ_i) but not too much information to the point where we get a weak bound on the probability due to the lack of randomness.

This motivates us to think of a restriction by first assigning a uniform random string z to x and then covering up a p-subset Λ with stars to create a restriction $\rho(\Lambda, z)$. The intuition for this is that hopefully the random string z, combined with P, will be enough information from ρ to fix (ℓ_i), from which we can use the remaining randomness in ρ (namely Λ) to obtain the p^t bound. In particular, we hope that there is a "transcript searcher" $\mathcal{S}$, which on input (z, P), can recover a completed transcript (ℓ_i, P) such that all ρ designed by initially assigning $x = z$ will have partial transcript P only if (ℓ_i, P) is its transcript. If such a function exists, then we could say

$$\begin{aligned}
\Pr_{\rho}[P \text{ is a partial transcript for } \rho] &= \mathbb{E}_{z \sim U_n} \Pr_{\Lambda}[P \text{ is a partial transcript for } \rho(\Lambda, z)] \\
&= \mathbb{E}_{z \sim U_n} \Pr_{\Lambda}[\mathcal{S}(z, P) \text{ is a transcript for } \rho(\Lambda, z)] \\
&\leq p^t
\end{aligned}$$

where the last inequality follows since ρ must keep the variables in the transcript alive. Alas, such an $\mathcal{S}$ cannot exist. There can exist different restrictions created from the same ground assignment z that are witnessed by different completions of P (this ambiguity is an unavoidable side effect of not being able to condition on all information about ρ). We cannot hope for a unique completion, but what if our $\mathcal{S}$ output all of these potential completions with decent probability over the randomness in z? Say ρ is *good* if P is a partial transcript for it. In formal terms, say we can construct $\mathcal{S}$ such that for any good ρ, $\Pr_z[\mathcal{S}(z, P) \text{ is a partial transcript for } \rho] \geq \gamma$ (earlier we were demanding $\gamma = 1$, which turned out to be impossible). Then we can deduce

$$\begin{aligned}
&\Pr_{\rho}[P \text{ is a partial transcript for } \rho] && (3)\\
&\quad = \mathbb{E}_{\Lambda} \Pr_{z}[\rho(\Lambda, z) \text{ is good}]\\
&\quad = \mathbb{E}_{\Lambda} \frac{\Pr_z[\mathcal{S}(z, P) \text{ is a transcript for } \rho(\Lambda, z)]}{\Pr_z[\mathcal{S}(z, P) \text{ is a transcript for } \rho(\Lambda, z) | \rho(\Lambda, z) \text{ is good}]}\\
&\quad \leq \frac{1}{\gamma} \mathbb{E}_{\Lambda} \Pr_{z}[\mathcal{S}(z, P) \text{ is a transcript for } \rho(\Lambda, z)\\
&\quad = \frac{1}{\gamma} \mathbb{E}_{z} \Pr_{\Lambda}[\mathcal{S}(z, P) \text{ is a transcript for } \rho(\Lambda, z)]\\
&\quad \leq p^t/\gamma. && (4)
\end{aligned}$$

Stringing Equations (1),(2), and (4) lets us bound

$$\Pr_{\rho}[\mathsf{DT}(F|_{\rho}) \geq t] \leq (p^t/\gamma) \cdot \#\{\text{partial transcripts } P\}$$

It turns out we can define our partial transcripts P and construct a transcript searcher such that the above term is small enough to give us the desired switching lemma. See Theorem 20 for the technical details.

1.2.1 Comparison to Lyu [18]

Although the proof structure for proving our switching lemma is similar to Lyu's [18] proof of the AC^0 switching lemma, some changes are necessary to accommodate the general structure of $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}$ circuits.

- We need to create a more complex CDT that can compute $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}$ circuits, and a corresponding new definition of witnesses/partial witnesses that records the transcript of the complex CDT so that our witness searcher can effectively reconstruct a transcript given information about ρ and a partial witness.
- Because our CDT contains more steps, there will naturally be more possible transcripts/witnesses. As the switching lemma hinges on a low quantity of possible partial witnesses to union bound over, we need to argue with our new CDT, the number of partial witnesses can be controlled by the parameter k. This makes designing the CDT and partial witness to be an act of balancing contrasting parameters

  - For example, the more complicated a CDT procedure is, the closer to the true decision tree depth it will reach (and hence a tighter bound on $\Pr_{\rho}[\mathsf{CDT}(F|_{\rho}) \geq t]$ can be expected), but the larger the possible number of transcripts it will have (thereby increasing the number of terms we union bound over). Therefore, this approach demands the designed CDT to be complicated enough to give a small depth decision tree with high probability, but simple enough to be tractable to analyze with a union bound.
  - Similarly, the more that a partial witness keeps track of, the larger amount of possible partial witnesses we will need to union bound over. However, if we keep track of too little, there will not exist an effective witness searcher that can use the information from the partial witness to construct the whole witness. Hence we need to keep track of just the right amount of information.

- In the argument for AC^0 circuits, one would show a multi-switching lemma on depth 2 AC^0 circuits. In other words, one would argue that a collection of AC^0_2 circuits simultaneously simplify after a one random restriction is applied to all of them. Rather than the natural idea of proving a switching lemma for the analogous $\mathsf{GC}^0_2(k)$ circuits, we consider the hybrid class of $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}$ circuits. It turns out a switching lemma on these simpler circuit classes suffice to depth reduce and prove bounds on $\mathsf{GC}^0_d(k)$ as we will see below.

1.3 The Depth Reduction Lemma

The multi-switching lemma gives a simplification lemma for depth 2 circuits. To extend this to constant depth circuits, we would like to iteratively decrease the depth of the circuit and induct. The argument for AC^0 circuits was quite simple. Say we have a depth 3 $\mathsf{OR} \circ \mathsf{AND} \circ \mathsf{OR}$ circuit F. Using the switching lemma, we can say with high probability, $F|_\rho$ is an $\mathsf{OR} \circ \mathsf{DT}_t$ circuit. We now expand each bottom layer decision tree into an $\mathsf{OR} \circ \mathsf{AND}_t$ circuit by enumerating over all 1-paths. Consequently this simplifies $F|_\rho$ to a $\mathsf{OR} \circ (\mathsf{OR} \circ \mathsf{AND}_t) = \mathsf{OR} \circ \mathsf{AND}_t$ circuit, since an OR of OR of variables is simply a single OR over all variables involved, getting us a depth reduction from depth 3 to 2.

What happens when we try the same argument for a k-$\mathsf{OR} \circ k$-$\mathsf{AND} \circ \mathsf{OR}$ circuit F? By our switching lemma, $F|_\rho$, with high probability will simplify to a k-$\mathsf{OR} \circ \mathsf{DT}_t$ circuit. We can then unravel the decision trees into $\mathsf{OR}_{2^t} \circ \mathsf{AND}_t$ CNFs, resulting in a k-$\mathsf{OR} \circ \mathsf{OR}_{2^t} \circ \mathsf{AND}_t$ circuit. Here, we reach an issue: a k-$\mathsf{OR} \circ \mathsf{OR}$ circuit is not necessarily itself a k-OR function! We could have up to $(k-1)2^t$ input bits of a k-$\mathsf{OR} \circ \mathsf{OR}_{2^t}$ be 1 while still evaluating to 0 (set all 2^t bits of $k-1$ of the bottom depth ORs to be 1). The best we can do is say the function is in $\mathsf{G}((k-1)2^t+1)$, which is too large of a blowup in the "k" parameter for our switching lemma to handle.

We rewind a bit to our k-$\mathsf{OR} \circ \mathsf{DT}_t$ circuit and unravel the bottom-layer trees to $\mathsf{OR}_{2^t} \circ \mathsf{AND}_t$ DNFs by enumerating over 1-paths. But we now make the key observation about each DNF which follows from the fact any assignment uniquely defines a path on a decision tree: *any assignment of x makes at most* 1 AND_t *clause true.* To use more standard terminology, the DNF created from decision trees is *unambiguous.* This means the pathological case above of all clauses under $k-1$ OR_{2^t} gates being satisfied cannot happen. In fact, we can prove something stronger. Since at most one clause under each OR_{2^t} gate can be satisfied in the unraveled k-$\mathsf{OR} \circ \mathsf{OR}_{2^t} \circ \mathsf{AND}_t$ circuit, the number of middle layer OR_{2^t} clauses that are satisfied will be precisely the total number of bottom layer AND_t clauses that are satisfied. Hence, a k-OR over the OR gates is exactly the same as a k-OR over the AND_t clauses themselves, and we can indeed collapse to a k-$\mathsf{OR} \circ \mathsf{AND}_t$ circuit! This gets us our depth reduction. A slightly more involved argument is carried out to show the more general $\mathsf{G}(k) \circ \mathsf{DT}_t$ circuit can be calculated by a $\mathsf{G}(k) \circ \mathsf{AND}_t$ circuit, but the heart of the argument is captured in the k-OR case itself.

1.4 Putting It All Together

We now have all the ingredients to simplify $\mathsf{GC}^0_d(k)$ circuits. The argument will be the following inductive process, where we are effectively inducting on circuits of the form $\mathsf{GC}^0_d(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_w$ rather than GC^0_d directly. Given a $\mathsf{GC}^0_d(k)$ circuit,

1. Add a trivial $(d+1)$-st layer at the bottom that is simply the identity gate (think of it as an AND_1 gate)
2. By the multi-switching lemma, we know the depth 2 $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}$ subcircuits simplify to DT_t trees with high probability, resulting in a $\mathsf{GC}^0_{d-1}(k) \circ \mathsf{DT}_t$ circuit.

3. By the depth reduction lemma, each of the bottom depth 2 $\mathsf{G}(k) \circ \mathsf{DT}_t$ subcircuits can be calculated by a $\mathsf{G}(k) \circ \{\mathsf{AND}_t, \mathsf{OR}_t\}$ circuit, resulting in a $\mathsf{GC}^0_{d-1}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}$ circuit.
4. The depth has reduced by 1, so we go back to Step 2 and induct.

This argument allows us to use the multi-switching lemma along with the depth reduction lemma to establish size lower bounds for $\mathsf{GC}^0_d(k)$ bounds. We show a formal argument of this outline in Section 3.

2 Preliminaries

2.1 Notation

$[n] = \{1, 2, \ldots, n\}$ denotes the set of the first n positive integers. $\binom{[n]}{k}$ denotes the set of all size k subsets of $[n]$. log is assumed to be in base 2. This paper concerns *constant*-depth circuits, and so the depth variable, d, should be treated as a constant. In particular hidden constants in $O(\cdot)$ or $\Omega(\cdot)$ may depend on d. For $S \subset [n]$, we denote $x^S = \prod_{i \in S} x_i$.

2.2 Random Restrictions and Partial Assignments

A *partial assignment* or *restriction* is a string $\rho \in \{0, 1, \star\}^n$. Intuitively, a $\star$ represents an index that is still "alive" and hasn't been fixed to a value yet.

An alternative way of defining a restriction is by the set of alive variables and a "ground assignment" string. Given a "$\star$ set" Λ and a ground assignment $z \in \{0,1\}^n$, we define $\rho(\Lambda, z)$ to be the partial assignment where we assign

$$\rho(\Lambda, z)_i = \begin{cases} \star & i \in \Lambda \\ z_i & i \notin \Lambda \end{cases}$$

Sometimes, Λ may be in the form of an indicator $\{0,1\}^n$ string, where the set is defined to be the set of indices containing a 1.

We also define a composition operation on partial assignments. For two restrictions ρ^1, ρ^2, define $\rho^1 \circ \rho^2$ so that

$$(\rho^1 \circ \rho^2)_i = \begin{cases} \rho^1_i & \rho^1_i \neq \star \\ \rho^2_i & \rho^1_i = \star. \end{cases}$$

Intuitively, one can see this as fixing bits determined by ρ^1 first, and then out of the remaining alive positions, fix them according to ρ^2.

A *random restriction* is simply a distribution over restrictions. A common random restriction we will use is R_p, the distribution where each index will be assigned $\star$ with probability p, and $0, 1$ each with probability $\frac{1-p}{2}$.

The main reason for defining restrictions is to observe their action on functions. Given a restriction ρ and function $f : \{0,1\}^n \to \{0,1\}$, we define $f|_\rho : \{0,1\}^n \to \{0,1\}$ to be the function mapping $f|_\rho(x) := f(\rho \circ x)$.

2.3 Models of Computation

Circuits

We measure the size of a circuit by the total number of wires (including input wires) in it. We define the width of a DNF or CNF to be the maximum number of variables in any of its clauses. We also use k-DNF (resp. k-CNF) to denote DNF (resp. CNF) of width at most

k. AC^0_d are depth d circuits with unbounded fan-in whose gate set is $\{\mathsf{AND}, \mathsf{OR}, \mathsf{NOT}\}$. In general, if we have a gate G, a subscript G_k will refer to its fan-in (in this case, G is fixed to have fan-in k). We now define more general circuit classes that we analyze in this work.

▶ **Definition 9** (k-OR/k-AND/$\mathsf{AC}^0_d(k)$)**.** *Define k-$\mathsf{OR}_m : \{0,1\}^m \to \{0,1\}$ to be a function that evaluates to 1 iff x contains $\geq k$ ones. Analogously define k-AND_m to be 0 iff x contains $\geq k$ zeros. Define $\mathsf{AC}^0_d(k)$ to be the class of depth d circuits with unbounded fan-in whose gate set is $\{k'$-AND, k'-$\mathsf{OR}, \mathsf{NOT}\}$ for all $k' \leq k$.*

In more generality, we define $\mathsf{G}(k)$ gates and $\mathsf{GC}^0_d(k)$ circuits.

▶ **Definition 10** ($\mathsf{G}(k)/\mathsf{GC}^0(k)$)**.** *Define a gate set $\mathsf{G}(k)$ to be the set of all arbitrary fan-in gates such that they are constant on inputs with $\geq k$ ones (we call such gates* orlike*) or are constant on inputs with $\geq k$ zeros (we call such gates* andlike*). $\mathsf{GC}^0(k)$ is the class of constant depth circuits made by $\mathsf{G}(k)$ gates.*

In the rest of the paper, we may write circuit classes $\mathsf{GC}^0_d(k) \circ \{\mathsf{AND}, \mathsf{OR}\}$ or $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}$. In the literature, this usually refers to the circuit class whose gates above the bottom layer are in $\mathsf{G}(k)$, and whose bottom layer gates can either be AND or OR with no restriction on the choice. However, in this paper, assume this notation implicitly restricts AND gates to only be under orlike $\mathsf{G}(k)$ gates and OR gates to only be under andlike $\mathsf{G}(k)$ gates.

On top of being an alternate generalization of AND/OR gates, $\mathsf{G}(k)$ gates capture arbitrary LTFs that are "unbalanced" in some sense. We will use the $\{\pm 1\}$ bits to define these, but one can convert between $\{0,1\}$ and $\{\pm 1\}$ via the map $b \to (-1)^b$.

▶ **Definition 11** (Balance of an $\mathsf{LTF}/\mathsf{TC}^0(k)$)**.** *Consider an arbitrary LTF $f : \{\pm 1\} \to \{\pm 1\}$ with $f(x) = sgn(\sum w_i x_i - \theta)$. Let $\sigma : [n] \to [n]$ be a permutation ordering (w_i) such that $|w_{\sigma(1)}| \leq \cdots \leq |w_{\sigma(n)}|$. Define the* balance *of f (denoted as $bal(f)$) to be the smallest integer k such that $-\sum_{i \leq k} |w_i| + \sum_{i>k} |w_i| < |\theta|$. Now denote $\mathsf{TC}^0(k)$ to be the class of constant depth circuits made out of THR gates with balance $\leq k$.*

We prove that up to negations in the inputs and output, THR gates with balance k are in $\mathsf{G}(k)$ in the appendix (Theorem 51). All results in this paper hold for $\mathsf{TC}^0(k)$, but from now on, we will only refer to $\mathsf{GC}^0(k)$ as it is the more general class.

Decision Trees

We assume knowledge of decision trees (see Definition 3.13 in [20] for a reference). We will be using slightly more complex models of decision trees in this work.

▶ **Definition 12** (Partial Decision Trees)**.** *For a collection of functions $\mathcal{F} = \{F_1, \ldots, F_m\}$, we say $\mathcal{F}$ can be computed by an r-partial depth-t DT if there exists a singe depth r tree such that for all F_i and paths π of T, $F_i|_\pi$ can be computed by a depth t decision tree (here, $F|_\pi$ is F acted on by the restriction induced by taking path π down T).*

▶ **Definition 13** ($(d, \mathcal{C})$-tree)**.** *Let d be an integer and $\mathcal{C}$ a computational model (e.g. a circuit class). A function is computable by a $(d, \mathcal{C})$-tree if it is computable by a depth t decision tree with $\mathcal{C}$ functions as its leaves. That is, there exists a depth d decision tree T such that for every path π in T, $F|_\pi \in \mathcal{C}$.*

2.4 Pseudorandomness and Probability

We will use various pseudorandom primitives and terminology. We will use U_n to denote the uniform distribution over n bits unless specified otherwise.

▶ **Definition 14** (ε-error PRG/Seed Length). *A distribution D over $\{0,1\}^n$ is called an ε-error PRG for a computational model $\mathcal{C}$ if for all $C \in \mathcal{C}$,*

$$|\mathbb{E}_{x\sim U_n}[C(x)] - \mathbb{E}_{x\sim D}[C(x)]| \leq \varepsilon$$

The seed length s of D is defined to be the minimal quantity s such that the following is true: there exists a polytime computable function $G : \{0,1\}^s \to \{0,1\}^n$ such that the distribution of $G(z)$ over $z \sim U_s$ is exactly D.

▶ **Definition 15** ((ε, k)-wise independent source). *A distribution D over $\{0,1\}^n$ is an (ε, k)-wise independent source if for all $1 \leq i_1 < \cdots < i_k \leq n$ and $\alpha \in \{0,1\}^k$,*

$$|\Pr_{x\sim D}[x_{i_1}x_{i_2}\ldots x_{i_k} = \alpha] - 2^{-k}| < \varepsilon.$$

There exists constructions of these sources with seed length $O(\log\log n + k + \log(1/\varepsilon))$ [3].

▶ **Definition 16** (k-wise Independent Hash Family). *Let $\mathcal{H}$ be a distribution over hash functions mapping $\{0,1\}^n \to \{0,1\}^m$. We say that $\mathcal{H}$ is k-wise independent if for any k input-output pairs $(x_1, y_1), \ldots, (x_k, y_k) \in \{0,1\}^n \times \{0,1\}^m$ where $x_1, \ldots, x_t$ are distinct, it holds that*

$$\Pr_{h\sim\mathcal{H}}[\forall i \in [k], h(x_i) = y_i] = 2^{-km}.$$

Such functions can be sampled using $O(k(n+m))$ bits (Chapter 3.5.5 of [32]).

▶ **Definition 17** (k-wise p-bounded Subset). *Let Λ be a random subset of $[n]$. Λ is a k-wise p-bounded subset iff for all subsets $S \subset [n]$ of size $\leq k$, $\Pr_\Lambda[S \subset \Lambda] \leq p^{|S|}$.*

For example, R_p is n-wise p-bounded.

2.5 Fourier Analysis

Every Boolean function $f : \{\pm 1\}^n \to \{\pm 1\}$ has a unique representation as a multilinear real polynomial

$$f(x) = \sum_{S\subset[n]} c_S x^S.$$

Given f, we can think of the Fourier transform of f, $\widehat{f}$ to be a function mapping $2^{[n]} \to \mathbb{R}$ such that $\widehat{f}(S) = c_S$. This is well defined by the uniqueness of the polynomial representation of S. One can explicitly compute $\hat{f}(S) = \mathbb{E}_x[f(x)x^S]$. By Parseval's, one can derive $\sum_{S\subset[n]} \widehat{f}(S)^2 = 1$. There are various quantities involving the Fourier coefficients that we will work with.

▶ **Definition 18** (Fourier Tails). *For a Boolean function f, define*

$$W^{\geq k}[f] := \sum_{|S|\geq k} \widehat{f}(S)^2.$$

▶ **Definition 19** (Discrete Derivative/Influence). *For Boolean f and $i \in [n]$, define the discrete derivative*

$$D_i f(x) = \frac{f(x^{(i\to 1)}) - f(x^{(i\to -1)})}{2}$$

where $x^{(i\to b)} = (x_1, \ldots, x_{i-1}, b, x_{i+1}, \ldots, x_n)$. Now for $S \subset [n]$ with $S = \{i_1, \ldots, i_k\}$, define

$$D_S f = D_{i_1} D_{i_2} \ldots D_{i_k}.$$

Now for $S \subset [n]$, define the influence

$$Inf_S(f) = \mathbb{E}_{x\sim\{\pm 1\}^n}[D_S f(x)^2].$$

Finally, define the degree k influence

$$Inf^k(f) = \sum_{|S|=k} Inf_S(f).$$

3 Simplification Theorem of $\mathsf{GC}^0_d(k)$ Circuits

▶ **Theorem 20.** *Let F be computable by a depth-2 $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_w$ circuit. Let Λ be a $(t+w)$-wise p-bounded subset of $[n]$, and x a uniform string. Then*

$$\Pr_{\Lambda,x}[\mathsf{DT}(F|_{\rho(\Lambda,x)}) \geq t] \leq (20pw)^t 2^k.$$

Proof. The proof will follow that of Section 5 in [18]. We urge the reader to first read the overview given in Section 1.2. As discussed there, the main differences between this proof and the one presented there are present in the constructions of the canonical decision tree and witness searchers, the definition of witnesses, and the counting of partial witnesses. These are altered to support the more general $\mathsf{G}(k)$ gates. Besides this, the general proof strategy remains the same. Let m be the fan-in of the F. We present a procedure that constructs a decision tree (which we deem the "Canonical Decision Tree") in Algorithm 1.

The difference between the CDT defined in [18] and the one presented here is the use of ctr. Intuitively, this is added in to keep track of the number of satisfied clauses we see before we reach our limit of k. We rigorously prove this in the following claim.

▷ Claim 21. The CDT correctly outputs $F(\alpha)$.

Proof. The CDT scans the clauses in order to find the first one not fixed to zero. There are only two return statements in the algorithm, so we consider the two cases of terminating on each one. Suppose we terminate at the first return statement and output $F(1^m)$. Notice ctr is incremented each time the CDT encounters a satisfied clause. Therefore, when $ctr = k$, at least k C_i evaluate to 1, and therefore $F(C_1, \ldots, C_m) = F(1^m)$ by virtue of $F \in \mathsf{G}(k)$ being orlike, proving correctness. Now suppose we terminate after the while loop and output $C(x \circ 0)$. If CDT finishes the while loop without terminating, that must mean all clauses must be determined by the partial assignment x. This is because for any clause $C_{j'}$, if the clause wasn't already determined in the algorithm when $j^* = j'$, all unknowns of $C_{j'}$ would have been queried and fixed in the partial assignment x, thereby determining it. Therefore, in this case, $C(x)$ is determined, and in particular is equal to $C(x \circ 0^n)$. ◁

Algorithm 1 Canonical Decision Tree.

Input: (orlike $G(k)) \circ \mathsf{AND}_w$ circuit $F = G(C_1, \ldots, C_m)$, black-box access to a string $\alpha \in \{0,1\}^n$.
initialize:
  $j^* \leftarrow 0$
  $x \leftarrow (\star)^n$
  $ctr \leftarrow 0$
while $j^* < m$ **do**
  Find the first $j > j^*$ such that $C_j(x) \not\equiv 0$. If no such j exists, exit the loop.
  $B_j \leftarrow$ the set of unknown variables in $C_j(x)$ (may be empty).
  Query α_{B_j}.
  Set $x_{B_j} \leftarrow \alpha_{B_j}$.
  if $C_j(x) = 1$ **then**
    $ctr \leftarrow ctr + 1$;
    if $ctr = k$ **then**
      return $G(1^m)$
    end
  end
  $j^* \leftarrow j$
end
return $F(x \circ 0^n)$.

Therefore, CDT is indeed a decision tree computing F. Define $\mathsf{CDT}(F)$ to be the depth of the canonical decision tree T_F.If ρ is bad (i.e. $\mathsf{DT}(F|_\rho) \geq t$), then clearly $\mathsf{CDT}(F|_\rho) \geq \mathsf{DT}(F|_\rho) \geq t$ and so $T_{F|_\rho}$ will result in at least t queries for *some* choice of α (this is equivalent to saying that *some* path of $T_{F|_\rho}$ must have length $\geq t$). We define a witness that will effectively be the transcript of the algorithm on this particular α.

▶ **Definition 22.** *Let F be the circuit described above and ρ a restriction. Let $t \geq 1$. Consider the tuple $(r, \ell_i, s_i, B_i, \alpha_i)$ where*

- *$r \in [1, t+k]$ is an integer*
- *$(\ell_1, \ldots, \ell_r) \in [m]^r$ is an increasing list of indices*
- *$(s_1, \ldots, s_r)$ is a list of non-negative integers, at most k of which are allowed to be 0, such that $s := \sum_{i=1}^r s_i \in [t, t+w-1]$*
- *$(B_1, \ldots, B_r)$ is a list of (potentially empty) subsets of $[w]$ satisfying $|B_i| = s_i$.*
- *$(\alpha_1, \ldots, \alpha_r)$ is a list of (potentially empty) bit strings satisfying $|\alpha_i| = s_i$.*

$(r, \ell_i, s_i, B_i, \alpha_i)$ is called a t-witness for ρ if there exists an $\alpha \in \{0,1\}^n$ such that

- *When we run $T_{F|_\rho}$ on α, C_{ℓ_i} is the i-th term queried by $T_{F|_\rho}$.*
- *$T_{F|_\rho}$ queries s_i variables in C_{ℓ_i}, and the relative location of those variables within C_{ℓ_i} are specified by set B_i.*
- *$T_{F|_\rho}$ receives α_i in response to its i-th batch query.*

The size *of the witness $(r, \ell_i, s_i, B_i, \alpha_i)$ is defined to be $s := \sum_{i=1}^r s_i$. We may denote the size of a witness W as $\mathsf{size}(W)$.*

▷ Claim 23. For every ρ such that $\mathsf{DT}(F|_\rho) \geq t$, there exists a t-witness for ρ

Proof. We simply run $T_{F|_\rho}$ on α that causes at least t queries to be issued. We then record the transcript *until the number of variables queried exceeds* t, after which we halt. To be more explicit, we let r be the number of times the CDT stops at a clause before either outputting a bit or exeeding t queries. At the ith stop, say on clause C_j, we set $\ell_i = j$, s_i to be the number of unknown variables queried (which may be 0), B_i to be the subset of $[w]$ indicating the relative positions of the variables in the clause, and α_i being the query replies received from the black-boxed α. We can first verify this creates a valid tuple.

- $r \in [1, t+k]$. Every time we stop at a clause, either it evaluates to 1 and we increment ctr, or we have to query at least 1 variable. Since ctr can incremented at most k times and we can query variables from at most t clauses before reaching our quota of t queried variables, it follows we stop at most $t+k$ times.
- $(\ell_1, \ldots, \ell_r) \in [m]^r$ is an increasing list of indices since the CDT linearly sweeps the clauses in increasing index order.
- $(s_1, \ldots, s_r)$ is a list of non-negative integers, at most c of which are allowed to be 0, such that $s := \sum_{i=1}^{r} s_i \in [t, t+w-1]$. A particular s_i being zero implies that the CDT's ith stop was at a clause C_j that was already determined to be 1, and hence ctr was incremented. Since ctr can be incremented at most k times, at most k of the s_i's are zero. s is the total number of variables queried before halting. Notice after the penultimate clause is queried, there are $< t$ clauses queried. Consequently when the ultimate clause is queried, there clearly will be $< t + w$ variables queried, since k is the width of the clause. Of course, since the transcript halted after this clause, $\geq t$ variables had to be queried.
- $(B_1, \ldots, B_r)$ is a list of (potentially empty) subsets of $[w]$ satisfying $|B_i| = s_i$ trivially by construction.
- $(\alpha_1, \ldots, \alpha_r)$ is a list of (potentially empty) bit strings satisfying $|\alpha_i| = s_i$ trivially by construction.

We can then easily see by construction of the tuple, it is indeed a t-witness for ρ. ◁

We note that the difference between Definition 4 in [18] and the one here is the relaxation to allow $(s_1, \ldots, s_r)$ to contain up to k zeros, rather than to all be positive. As evident in the proof, this is to handle cases the CDT encounters a clause that was already fixed to 1, which causes the corresponding s_i value to be 0. This wasn't recorded in Lyu's witness definition because in the case of CNFs, one satisfied clause determines the value of the circuit, and the CDT immediately halts. Why do we still record that the CDT didn't query any variables at a clause instead of just ignoring this behavior and moving on? It turns out if we don't include this piece of information, the witness searcher we create will not have enough information to reconstruct the whole witness (see the "balancing act" discussion in Section 1.2.1).

We now move on to define partial witnesses.

▶ **Definition 24.** *Let F be a circuit and ρ a restriction. We call (r, s_i, B_i, α_i) a partial t-witness for ρ if there exists $(\ell_1, \ldots, \ell_r)$ such that $(r, \ell_i, s_i, B_i, \alpha_i)$ is a t-witness for ρ.*

We note the following important claim.

▷ Claim 25. If P is a partial witness for ρ, then there exists exactly one list of integers (ℓ_i) such that (ℓ_i, P) is a witness for ρ.

Proof. By construction of Algorithm 1, ℓ_1 must be be the index of the first clause not fixed to 0 by ρ. But now, we notice ℓ_2 must be the index of the first clause after C_{ℓ_1} not fixed to 0 by $\rho \circ \alpha_1$. We then continue this induction to get our unique list (ℓ_i), Where ℓ_j will be forced to be the index of the first clause after $C_{\ell_{j-1}}$ that is not fixed to 0. ◁

Therefore, by Claims 23 and 25

$$\Pr_{\rho}[\mathsf{DT}(F|_\rho) \geq t] \leq \sum_{(\ell_i,P)} \Pr_{\rho}[(\ell_i, P) \text{ is a } t\text{-witness for } \rho] \leq \sum_{P} \Pr_{\rho}[P \text{ is a partial } t\text{-witness for } \rho]] \tag{5}$$

where P ranges over all partial t-witness tuples.

Going back to our proof, we now define our *witness searcher* $\mathcal{S}$ as Algorithm 2.

Algorithm 2 Witness Searcher $\mathcal{S}$.

Input: (orlike $\mathsf{G}(k)) \circ \mathsf{AND}_w$ circuit $F(C_1, \ldots, C_m)$, ground assignment $z \in \{0,1\}^n$, partial witness $W = (r, s_i, B_i, \alpha_i)$.

initialize:
- $j^* \leftarrow 0$
- $x \leftarrow (\star)^n$
- $ctr \leftarrow 1$

while $ctr \leq r$ **do**
- **while** $j^* < m$ **do**
  - Find the first $j > j^*$ such that $C_j(z) \equiv 1$. If no such j exists, exit the inner while loop.
  - $\ell_{ctr} \leftarrow j$
  - Query α_{B_j}.
  - Set the B_{ctr} portion of z to be α_{ctr}.
  - $ctr \leftarrow ctr + 1$
  - $j^* \leftarrow j$
- **end**

end

return (ℓ_i, W)

We now prove the following essential property about $\mathcal{S}$, stating in a probabilistic way, it can use a partial witness to reconstruct a total witness for ρ.

▶ **Lemma 26.** *Let P be a partial witness, and let s be its size. Define a restriction ρ to be* good *for P if P is a partial t-witness for it.*

$$\Pr_{z}[\mathcal{S}(z, P) \text{ is a } t\text{-witness for } \rho(\Lambda, z) | \rho(\Lambda, z) \text{ is good for } P] = 2^{-s}$$

Furthermore this event is solely dependent on z_I, where I is the set of variable indices referred to by the unique completion of P with respect to ρ.

Proof. If $\rho = \rho(\Lambda, z)$ is good, then by Claim 25 we know there exists unique ℓ_i such that (ℓ_i, P) witnesses ρ. In particular, we know that Λ must contain all the indices I that (ℓ_i, B_i) identify. Let I_j be the index set identified by ℓ_j and B_j (so $I = I_1 \sqcup \cdots \sqcup I_r$). Now condition on a fixed ρ. This means all bits in z not covered by Λ are fixed. In particular, only source of randomness left are the bits covered by Λ, which is a superstring of z_I. We now claim every z_{I_j} is assigned the unique bit string such that $C_{\ell_i}|_{z_{I_j}} \not\equiv 0$ (not forced to be unsatisfied) iff $\mathcal{S}$ successfully outputs (ℓ_i, P). This consequently proves the lemma, since this has probability $2^{-|I|} = 2^{-s}$ of happening.

By construction of $T_{F|_\rho}$ we know that all clauses before C_{ℓ_1} was falsified by ρ. Upon inspection, we see $\mathcal{S}$ correctly skips past these clauses (as z is a completion of ρ). Now we note C_{ℓ_1} was not fixed to 0 ρ, causing $T_{F|_\rho}$ to query all unknowns in C_{ℓ_1} at the time (which

might be nothing if C_{ℓ_1} was fixed to 1), which is x_{I_1}. Inspecting $\mathcal{S}$, we see $\mathcal{S}$ will set the correct ℓ_1 iff $C_{\ell_1}(z)$ is satisfied iff z_{I_1} is assigned the unique string such that $C_{\ell_i}|_{z_{I_j}} \not\equiv 0$ (since all variables outside I_1 ocurring in C_{ℓ_1} is fixed by ρ). $\mathcal{S}$ then (importantly) replace z_{I_1} with α_{I_1} so that all variables encountered thus far are assigned exactly as $T_{F|_\rho}$ did.

We then repeat this argument r times, noting that due to $z \circ \alpha_{I_1} \circ \cdots \circ \alpha_{I_j}$ being a completion of $\rho \circ \alpha_{I_1} \circ \cdots \alpha_{I_j}$, $\mathcal{S}$ rightfully skips all clauses between C_{ℓ_j} and $C_{\ell_{j+1}}$. We then similarly argue that $\mathcal{S}$ will set ℓ_{j+1} to be the $j+1$st clause $T_{F|_\rho}$ queries iff $z_{I_{j+1}}$ is the unique string such that $C_{\ell_{j+1}}|_{z_{I_{j+1}}} \not\equiv 0$. ◀

Combining Equation (5) and Lemma 26, it follows that

$$\begin{aligned}\Pr_\rho[\mathsf{DT}(F|_\rho) \geq t] &\leq \sum_P \Pr_\rho[P \text{ is a partial } t\text{-witness for } \rho] \\ &\leq \sum_P \mathbb{E}_\Lambda \frac{\Pr_z[\mathcal{S}(z,P) \text{ is a } t\text{-witness for } \rho(\Lambda,z)]}{\Pr_z[\mathcal{S}(z,P) \text{ is a } t\text{-witness for } \rho(\Lambda,z) | \rho(\Lambda,z) \text{ is good}]} \\ &\leq \sum_P 2^{\mathsf{size}(P)} \mathbb{E}_z \Pr_\Lambda[\mathcal{S}(z,P) \text{ is a } t\text{-witness for } \rho(\Lambda,z)] \qquad (6)\end{aligned}$$

Notice a necessary condition for a restriction $\rho(\Lambda, z)$ to be t-witnessed by a size-s W is for Λ to cover the s variables that W recorded as the CDT needing to query, which happens with probability $\leq p^s$ (as $s \leq t + w$ and Λ is $(t+w)$-wise p-bounded). Hence, every term in the sum in (6) can be bounded by $p^{\mathsf{size}(P)}$ and it remains to find the number of partial t-witness tuples P.

For a fixed s, we can bound the number of potential partial witnesses naively by noting

- the number of choices of (r, s_i) can be bounded by the number of ways to write s as the sum of at most $s+k$ nonnegative integers, which is $\sum_{r=1}^{s+k} \binom{s+r}{r} \leq \sum_{r=1}^{s+k} \binom{2s+k}{r} \leq 2^{2s+k}$ (notice that we get a larger count here than the analogous quantity of 2^{2s} in Lyu's proof [18], which is a side effect of looking at a more complicated circuit class),
- the choices for (B_i) can be bounded by $\prod_i \binom{w}{s_i} \leq w^s$,
- and the choices for (α_i) can be bounded by 2^s,

giving a total count of $(8w)^s 2^k$. Combining this count with the previous paragraph's observation and (6), while remembering to sum over all sizes, we derive

$$\Pr_\rho[\mathsf{DT}(F|_\rho) \geq t] \leq \sum_{s=t}^{t+w-1} (2p)^s (8w)^s 2^k = \sum_{s=t}^{t+w-1} (16pw)^s 2^k \leq (20pw)^s 2^k. \quad ◀$$

▶ Remark 27. One may ask whether the failure probability of $(20pw)^t 2^k$ tight. We show that PAR_{kw} can be expressed as a $\mathsf{G}(k) \circ \mathsf{AND}_w$ cirucit and prove this saturates the above bound in the Appendix (Theorem 52).

After defining witnesses, partial witnesses, and witness searchers for $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}$ circuits, we notice that Lyu's proof of the multi-switching lemma directly goes through with these definitions with zero changes (even down to the exact algorithm of the canonical partial decision tree and global witness searcher). Due to this, we defer the proof of the multi-switching lemma to the appendix. However, we do highlight here what properties about the circuit class is needed in order to invoke Lyu's lift from a switching lemma to a multi-switching lemma. The key properties needed were that

- the number of partial witnesses for a depth t canonical decision tree needed to be small, and
- there needed to exist a witness searcher function $\mathcal{S}$ such that for all ρ and a partial witness for ρ, $\mathcal{S}$ recovers the full witness with decent probability over an advice string.

- given a complete witness, there needed to be a small chance that a random restriction witnessed it

▶ **Theorem 28.** *Let $\mathcal{F} = \{F_1, \ldots, F_m\}$ be a list of $\mathsf{G}(k) \circ \mathsf{AND}_w$ circuits on $\{0,1\}^m$. Then*

$$\Pr_{\rho \sim R_p}[\mathcal{F}|_\rho \text{ does not have } r\text{-partial depth-}t\ \mathsf{DT}] \leq 4(64(2^k m)^{1/r} pw)^t$$

Proof. See Theorem 56. ◀

▶ Remark 29. At this point, we can observe an aspect of the expression that illuminates an important unifying flavor of the rest of the results. Notice that the only difference in the failure probability expression between the standard AC^0 multi-switching lemma in (lyu) and the above one for $\mathsf{G}(k)$ is that every occurrence of m is multiplied by a factor of 2^k. This means if constants in the exponent can be ignored, the multi-switching lemma asymptotically gives the same result as the AC^0 version if $2^k \approx m$. In particular, we can expect any result for size s AC^0 circuits to immediately extend to $\mathsf{GC}^0(\log s)$ circuits with no loss in parameters! This will be demonstrated in various settings in future sections.

With our multi-switching lemma in hand, we can simplify depth 2 circuits with high probability. To extend this to constant depth circuits, we also require a depth reduction lemma. In the case of AC^0, this was trivial enough to embed in the main proof, but in the case of $\mathsf{GC}^0(k)$, we need to be more delicate and use more specific properties of decision trees.

▶ **Lemma 30.** *Any depth 2 circuit of the form $\mathsf{G}(k) \circ \mathsf{DT}_w$ with top gate fan-in m can be expressed as a circuit in $G(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_w$ of size $m2^w$.*

Proof. Say the circuit we start with is $F(D_1, \ldots, D_m)$, where D_i are the bottom layer depth w decision trees. Assume F is orlike (the andlike case is analogous). By enumerating over all 1-paths, expand out each D_i as an OR of ANDs, namely $C_1^i \vee C_2^i \vee \cdots \vee C_{2^w}^i$. Now define a function F' over $m2^w$ bits, where

$$F'(x_1^1, \ldots x_{2^w}^1, x_1^2, \ldots, x_{2^w}^m) = \begin{cases} F(1^m) & \sum_{i,j} x_j^i \geq k \\ F(\bigvee_{i=1}^{2^w} x_j^1, \ldots, \bigvee_{i=1}^{2^w} x_j^m) & \text{otherwise} \end{cases}$$

Clearly by construction, $F' \in \mathsf{G}(k)$. Therefore to prove the lemma, it suffices to show that over all input assignments, $F(D_1, \ldots, D_m) = F'(C_1^1, \ldots C_{2^w}^1, C_1^2, \ldots, C_{2^w}^m)$.

If $\geq k$ of the D_i are satisfied, we know since F is an orlike $\mathsf{G}(k)$ function, $F(D_1, \ldots, D_m) = F(1^m)$. This also clearly implies $\geq k$ of the C_j^i are satisfied. Therefore by construction of F', $F'(C_1^1, \ldots, C_{2^w}^m)$ also evaluates to $F(1^m)$.

If $< k$ of the D_i are satisfied, then we need to use the following observation. For any assignment of inputs, at most one of the clauses $C_1^i, \ldots, C_{2^m}^i$ can be satisfied for each i, since each assignment uniquely defines a path in a decision tree. In more conventional terms, the DNF created by the decision tree D_i is *unambiguous*. Therefore the amount of D_i satisfied is exactly equal to the number of C_j^i satisfied, and so $< k$ clauses C_i^j are satisfied. This forces us into the second case of the piecewise definition of F', and so

$$F'(C_1^1, \ldots, C_{2^w}^m) = F(\bigvee_{i=1}^{2^w} C_j^1, \ldots, \bigvee_{i=1}^{2^w} C_j^m) = F(D_1, \ldots, D_m)$$

as desired. ◀

We have finally built up the tools to prove our main result: a constant depth simplification lemma.

▶ **Theorem 31.** *Let G be any gate, and let F be a $G \circ \mathsf{GC}^0_d(k)$ circuit of size m. Then for $p = \frac{1}{128(m2^k)^{1/w}}(128w(m2^k)^{1/w})^{-d+1}$ and any $t \geq 1$,*

$$\Pr_{\rho \sim R_p}[F|_\rho \text{ is not computed by a } ((2^d-1)t, G \circ \mathsf{DT}_w)\text{-decision tree}] \leq 4d \cdot 2^{-t}$$

Proof. WLOG assume the circuit is layered (all paths down the circuit are of length exactly $d+1$). We first append an extra layer of $\{\mathsf{AND}, \mathsf{OR}\}_1$ gates to the bottom of the circuit so that the input level fan-in is 1. We then apply a random restriction $\rho_0 \sim R_{p_0}$ with $p_0 = \frac{1}{128(m2^k)^{1/w}}$ and use Theorem 28 on all the depth-2 subcircuits to deduce that

$$\Pr_{\rho_0 \sim R_{p_0}}[F|_\rho \text{ is not computed by } (t, G \circ \mathsf{GC}^0_{d-1}(k) \circ \mathsf{DT}_w)\text{-decision tree}] \leq 4 \cdot 2^{-t}.$$

Letting $F^{(0)}$ be a good tree from above which does simplify, we see that there are at most 2^t leaves of the partial decision tree, with each leaf containing a $G \circ \mathsf{GC}^0(k)_{d-1} \circ \mathsf{DT}_w$ circuit (which we will refer to as "leaf-circuits"). By Lemma 30, these circuits can be simplified to $G \circ \mathsf{GC}^0_{d-1}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_w$ circuits. We apply Theorem 28 on the depth-2 subcircuits of a particular leaf-circuit with $p_1 = \frac{1}{128w(m2^k)^{1/w}}$, using $2t$ instead of t, union bound over all 2^t leaves, and then apply Lemma 30 to get that

$$\begin{aligned}\Pr_{\rho_1 \sim R_{p_1}}[F^{(0)}|_{\rho_1} \text{ is not a } (t+2t, G \circ \mathsf{GC}^0_{d-2}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_w\text{-decision tree}] &\leq 4 \cdot 2^{-2t} \cdot 2^t \\ &= 4 \cdot 2^{-t}.\end{aligned}$$

Iterating this argument $d-2$ more times, where we apply Theorem 28 on the depth 2 subcircuits using $p_i = \frac{1}{128w(m2^k)^{1/w}}$ and $2^i t$ instead of t on the ith iteration, and then union bound over all $2^{(2^i-1)t}$ leaves, we get that on the ith iteration, our desired single depth simplification happens with probability $2 \cdot 2^{-t}$. If the desired simplifications happen on all iterations, we result in a $((2^d-1)t, G \circ \mathsf{DT}_w)$-decision tree with probability at most $\sum_{i=0}^{d-1} 4 \cdot 2^{-t} = 4d \cdot 2^{-t}$ (via a union bound over the d iterations) and with a restriction from R_p where $p = \prod p_i = \frac{1}{128(m2^k)^{1/w}} \cdot (128w(m2^k)^{1/w})^{-d+1}$. The conclusion follows. ◀

With this theorem, we can let G be a $\mathsf{G}(k)$ gate to get the following corollary.

▶ **Corollary 32.** *Let C be a $\mathsf{GC}^0_d(k)$ circuit of size m and let $p = \frac{1}{40(128(k+\log m))^{d-1}}$. Then*

$$\Pr_{\rho \sim R_p}[\mathsf{DT}(C|_\rho) \geq t] \leq 2 \cdot 2^{-\frac{t}{2^{d-1}}+k}$$

Proof. Applying Theorem 31 to C with $w = k + \log m$, it follows for $p_1 = \frac{1}{128(128(k+\log m))^{d-2}}$,

$$\Pr_{\rho \sim R_{p_1}}[C|_\rho \text{ is not computed by a } ((2^{d-1}-1)t, \mathsf{G}(k) \circ \mathsf{DT}_{k+\log m})\text{-decision tree}] \leq 4d \cdot 2^{-t}.$$

Fix a ρ such that C simplifies to such a tree, T. By Lemma 30, the leaf circuits simplify to $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_{k+\log m}$ circuits. Let ℓ be a leaf, and let C_ℓ be the associated leaf-circuit. By Theorem 20, we know that for $p_2 = 1/40w$, $\Pr_{\tau \sim R_{p_2}}[\mathsf{DT}(C_\ell|_\tau) \geq 2^{d-1}t] \leq 2^{-2^{d-1}t}2^k$. Union bounding over all $\leq 2^{(2^d-1)t}$ leaves ℓ, it follows that

$$\begin{aligned}\Pr_{\rho \sim R_{p_1}, \tau \sim R_{p_2}}&[C|_{\rho\circ\tau} \text{ is not computed by a } ((2^{d-1}-1)t, \mathsf{DT}_{2^{d-1}t})\text{-decision tree}] \\ &\leq 2^{(2^d-1)t} \cdot 2^{-2^{d-1}t}2^k + 4d \cdot 2^{-t} \\ &\leq 2 \cdot 2^{-t+k}. \qquad (7)\end{aligned}$$

Because $\rho \circ \tau \sim R_{p_1 p_2}$, $p := p_1 p_2 = \frac{1}{40(128(k+\log m))^{d-1}}$, and a $((2^{d-1}-1)t, \mathsf{DT}_{2^{d-1}t})$-tree is simply a DT_{2^d-1}, (7) implies

$$\Pr_{\rho \sim R_p}[\mathsf{DT}(C|_\rho) \geq (2^d-1)t] \leq 2 \cdot 2^{-t+k}.$$

The desired result then follows after a change of variables from $t \to t/(2^d-1)$. ◀

4 Applications of The $\mathsf{GC}^0(k)$ Simplification Theorem

4.1 Exponential Lower Bounds Against Parity

Given Theorem 20 and Corollary 32, we can establish correlation bounds of $\mathsf{GC}^0_d(k)$ circuits against PAR (parity).

▶ **Theorem 33.** *Let $C \in \mathsf{GC}^0_d(k)$ have size m and let PAR be the parity function. Then the correlation of C and PAR is*

$$\mathbb{E}_{x \sim \{0,1\}^n}[(-1)^{C(x)+\mathsf{PAR}(x)}] \leq 2^{-\Omega(n/(k+\log m)^{d-1})+k}.$$

Proof. The uniform distribution is equivalent to performing a fair random restriction (a random restriction where non-star variables are set to a uniform bit), and then filling in the $\star$s with uniform bits. We will show that under a fair random restriction, C will become constant with high probability while PAR becomes a parity over the live variables. Averaging over these live variables then gives a correlation of zero. The total correlation is then the probability C doesn't become constant.

Let $p = \frac{1}{40(128(k+\log m))^{d-1}}$. Applying Corollary 32, we see that

$$\Pr_{\rho \sim R_p}[\mathsf{DT}(C|_\rho) \geq pn/4] \geq 2 \cdot 2^{-\frac{pn}{4(2^d-1)}+k}.$$

By a Chernoff bound, we know ρ will have $\geq pn/2$ stars with $\geq 1 - 2^{-pn/8}$ probability. Let $\mathcal{E}$ be the event both of these events happen, and fix such a ρ. Consider performing a random walk down the depth $\leq pn/4$ decision tree (start at the root and iteratively pick which of the 2 children to travel to uniformly, effectively filling in $\leq pn/4$ of the variables with uniform bits), which induces a random restriction τ. No matter which path restriction τ was taken, $C|_{\rho\circ\tau}$ becomes constant, while $\mathsf{PAR}|_{\rho\circ\tau}$ becomes a parity over $\geq pn/2 - pn/4 = pn/4$ variables. The correlation of these two functions is trivially 0. Therefore,

$$\begin{aligned}
&\mathbb{E}_{x \sim \{0,1\}^n}[(-1)^{C(x)+\mathsf{PAR}(x)}] \\
&\qquad = |\mathbb{E}_\rho \mathbb{E}_x[(-1)^{C|_\rho(x)+\mathsf{PAR}|_\rho(x)}]| \\
&\qquad \leq \Pr[\neg\mathcal{E}] + \mathbb{E}_\rho[|\mathbb{E}_x[(-1)^{C|_\rho(x)+\mathsf{PAR}|_\rho(x)}]|\big|\mathcal{E}] \\
&\qquad \leq 2 \cdot 2^{-\frac{pn}{4(2^d-1)}+k} + 2^{-pn/8}\mathbb{E}_\rho[\mathbb{E}_\tau|\mathbb{E}_x[(-1)^{C|_{\rho\circ\tau}(x)+\mathsf{PAR}|_{\rho\circ\tau}(x)}]|\big|\mathcal{E}] \\
&\qquad \leq 2^{-\Omega(n/(k+\log m)^{d-1})+k}.
\end{aligned}$$

◀

As an application, we can observe that for $0 \leq k \leq .1n^{1/d}$ one can set $m = 2^{\Theta((n/k)^{\frac{1}{d-1}})}$ in the above lemma such that the correlation is $< 1/2$, yielding us the following corollary.

▶ **Corollary 34.** *For some absolute constant C, integer d, and $0 \leq k \leq .1n^{1/d}$, $\mathsf{GC}^0_d(k)$ circuits computing PAR_n requires size $2^{\Omega((n/k)^{\frac{1}{d-1}})}$.*

This is an interesting result in multiple ways. First, notice the dependence of the lower bound on the "k" parameter is extremely tight and the corollary becomes absurdly false if $k = n^{1/d}$. This is seen by the fact $\mathsf{PAR}_{n^{1/d}} \in \mathsf{G}(n^{1/d})$, and so one can create a size $O(n^{1-1/d})$ $\mathsf{GC}^0_d(n^{1/d})$ formula computing PAR_n simply by having the ith depth from the bottom have $n^{1-i/d}$ $\mathsf{PAR}_{n^{1/d}}$ gates, each of which takes in inputs from $n^{1/d}$ gates below it. Hence we observe a "sharp threshold" behavior where a difference in constants can change an exponential lower bound to a sublinear one.

We also observe that this lower bound almost matches the classic $2^{\Omega(n^{\frac{1}{d-1}})}$ construction for AC^0_d circuits calculating parity. Thus, augmenting AC^0 with unbounded fan-in gates which have the power to calculate the majority of polynomially many bits has *no effect* on its ability to calculate parity, even thought we know such gates require exponentially sized AC^0_d circuits. In fact, by an argument resembling Shannon's classic circuit lower bound, there exists gates in $\mathsf{G}(k)$ which require size $2^{\Omega(n^{1/2d})}$ AC^0_d circuits.

We can also show that the size lower bound in Corollary 34 and the correlation bound in Theorem 33 is tight. In particular, the gap between the $2^{\Omega(n^{\frac{1}{d-1}})}$ lower bound for AC^0 and $2^{\Omega(n^{\frac{1}{d}})}$ bound established for $\mathsf{GC}^0(.1n^{1/d})$ cannot be bridged. We defer the formal proofs to Appendix A.2.

4.2 Correlation Bounds for $\mathsf{GC}^0(k)$ Circuits With Few Arbitrary Threshold Gates

In this section, we prove that state of the art correlation bounds against AC^0 circuits [27] with a small number of threshold gates extends to if we instead start with $\mathsf{GC}^0(\log^2 n)$ circuits. We first give an overview of their proof. As in previous works studying this correlation [23, 33, 17, 27], the hard function we uncorrelate with is

$$\mathsf{RW}_{m,k,r}(x) = \bigoplus_{i=1}^{m} \bigwedge_{j=1}^{k} \bigoplus_{\ell=1}^{r} x_{ijk}$$

A uniform string can be sampled by performing a random restriction and then filling the $\star$s with uniform bits. Driven by this, the overlying strategy is to apply a random restriction, and show the circuit collapses while the RW function maintains integrity. It turns out because our multi-switching lemma gives no loss in parameters (up to constants), we can apply the exact same argument in [27], except replace the AC^0 simplification lemma (Corollary 3.2 of [27]) with our more general Theorem 31.

▶ **Theorem 35.** *Fix u. Let $v = .005 \log n$ and $q = \sqrt{n/(v+1)}$ There exists a function $\mathsf{RW}_{q,v,q} \in \mathsf{P}$ and small enough constant τ such that for all circuits $\mathsf{ANY}_u \circ \mathsf{THR} \circ \mathsf{GC}^0_d(\Omega(\log^2 n))$ circuits F where each of the u $\mathsf{THR} \circ \mathsf{GC}^0_d$ subcircuits of F has size at most $s = n^{\tau \log n}$, we have*

$$|\mathbb{E}_{x \sim U_n}[(-1)^{\mathsf{RW}(x)+C(x)}] \leq 2^{-\Omega(n^{.499}/u)}$$

Proof. We immediately apply Theorem 31 to F with G being the top $\mathsf{ANY}_u \circ \mathsf{THR}$ circuit, $m = u \cdot 2^{(\varepsilon/100d)\log^2 n}, w = \varepsilon \log m, t = q/2$ and $k = (\varepsilon/100d)\log^2 n$ to get that for $\rho' \sim R_p$, where $p = n^{-\varepsilon/50}$,

$$\Pr[F|_{\rho'} \text{ is an } (m/2, \mathsf{ANY}_u \circ \mathsf{THR} \circ \mathsf{DT}_w)\text{-decison tree}] \leq 1 - 4d \cdot 2^{-q/2}.$$

This computational model is now void of $\mathsf{G}(k)$ gates, and we can essentially port in the rest of [27] to finish. By the "Second Step" and "Third Step" under Section 2 of [27], one can compose ρ' with another restriction to get a final random restriction ρ that simplifies the tree further and prunes the fan-in to an $\mathsf{ANY}_u \circ \mathsf{THR} \circ \mathsf{AND}_v$ circuit.

It was shown in Lemma 4.3 of [27] that the same random restriction ρ will have $\mathsf{RW}|_\rho$ equal(after restricting additional bits and negating input bits and/or the output) $\mathsf{GIP}_{q/2,v+1}$ except with probability $2^{-\tilde{\Omega}(pq)}$. Theorem 21 in [27] then states that an $\mathsf{ANY}_u \circ \mathsf{THR} \circ \mathsf{AND}_v$ circuits can be calculated by a randomized NOF $(v+1)$-party protocol with error $\gamma = 2^{-q^{.99}/u}$ using $O(uv^3 \log n \log(n/\gamma)) = O(q^{.99}v^3 \log n)$ bits. Finally, by Theorem 14 in [27], we can conclude the correlation between $\mathsf{GIP}_{q/2,v+1}$ and $\mathsf{ANY}_u \circ \mathsf{THR} \circ \mathsf{AND}_v$ is at most $2^{-\Omega(q^{.99}/u)}$. Hence the overall correlation can be bounded, via union bound, by the sum of the error probabilities and the correlation of $\mathsf{GIP}_{q/2,v+1}$ and $\mathsf{ANY}_u \circ \mathsf{THR} \circ \mathsf{AND}_v$, yielding

$$|\mathbb{E}_{x\sim U_n}[(-1)^{\mathsf{RW}(x)+C(x)}] \leq 4d \cdot 2^{-q/2} + 2^{-\tilde{\Omega}(pq)} + 2^{-\Omega(q^{.99}/u)} = 2^{-\Omega(n^{.49}/u)}$$

as desired. ◀

With this theorem, we can prove the actual correlation bound for $\mathsf{GC}^0(k)$ circuits with arbitrary gates.

▶ **Theorem 36.** *Let C be a $\mathsf{GC}^0(\Omega(\log^2 n))$ circuit, g of whose gates are arbitrary THR gates. Then*

$$\mathbb{E}[(-1)^{C(x)+\mathsf{RW}(x)}] \leq 2^{-\Omega(\frac{n^{.499}}{g}-g)}.$$

In particular, plugging in $g = \Theta(n^{.249})$ tells us

$$\mathbb{E}[(-1)^{C(x)+\mathsf{RW}(x)}] \leq 2^{-\Omega(n^{.249})}$$

Proof. This follows from Theorem 35 exactly like how Theorem 3 follows from Lemma 6 in [17]. ◀

▶ Remark 37. We note that an argument analogous to the above can be used to show $2^{-\Omega(n^{.499})}$ correlation bounds against $\mathsf{GC}^0(\Omega(\log^2 n))$ circuits with $n^{.499}$ gates, via the same argument presented in [27].

It is worth noting that if we had tried performing this argument by expanding the size $n^{\Omega(\log n)}$ $\mathsf{GC}^0(\log^2 n)$ circuit naively into an AC^0 circuit, not only would we get a loss in parameters, but the argument *will not go through*. The proof crucially relied on correlation bounds against $v = .005 \log n$ party protocols. Had we asymptotically increased the size of our circuit by writing it as an AC^0 circuit, then after applying random restrictions to prune the fan-in of our circuit, we will be left with trying to uncorrelate against arbitrary $\omega(\log n)$-party protocols, a longstanding open problem (Problem 6.21 in [14]).

4.3 Derandomizing the Multi-Switching Lemma and PRGs for $\mathsf{GC}^0(k)$

Using the same techniques appearing in [12, 18], we can completely derandomize our switching and multi-switching lemma. We defer the proof to the appendix (Theorem 62).

▶ **Theorem 38.** *Let $\mathcal{F} = \{F_1, \ldots, F_m\}$ be a list of size m $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_w$ circuits. Let (Λ, z) be a joint random variable such that*

- *Λ is a $(t+w)$-wise p-bounded subset of $[n]$*
- *Conditioned on any instance of Λ, z ε-fools CNF of size $\leq m^2$.*

Then

$$\Pr_{\Lambda,z}[\mathcal{F}|_{\rho(\Lambda,x)} \text{ has no } r\text{-partial depth-}t\ \mathsf{DT}] \leq 4(m2^k)^{t/r}(64pw)^t + (64wm)^{t+w}(2m)^{2kt/r} \cdot \varepsilon$$

Proof. See the Theorem 62 in the appendix. ◀

Using the derandomized multi-switching lemma, we can use the partition-based template in order to create PRGs for $\mathsf{GC}^0(k)$. The arument to reduce any constant depth to depth 2 will be a very similar argument. [18] simply uses a CNF PRG to tackle the base case of AC^0_2 circuits, but we cannot do so with a $\mathsf{G}(k) \circ \mathsf{AND}_w$ circuit unless we want to expand it out as a CNF and incur a multiplicative k loss in our seed length. We instead use the derandomized switching lemma one more time to simplify $\mathsf{GC}(k) \circ \mathsf{AND}_w$ to a $\mathsf{DT}_{\log m}$ and then fool this with an $(\varepsilon/m, \log m)$-wise independent seed. We quickly prove this latter statement in the following lemma.

▶ **Lemma 39** (PRG for Depth t Decision Trees). *There exists a ε-error PRG with $O(\log\log n + t + \log(1/\varepsilon))$ seed length for DT_t.*

Proof. Let D be a $(\varepsilon/2^t, t)$-wise independent distribution, samplable using $O(\log\log n + t + \log(2^t/\varepsilon)) = O(\log n + t + \log(1/\varepsilon))$ bits. For arbitrary $T \in \mathsf{DT}_t$, label the leaves $L_1, \ldots, L_{2^t}$, and let the value of leaf L_i be ℓ_i. Then

$$T(x) = \sum_{j=1}^{2^t} \ell_j \cdot \mathbb{1}(T(x) \text{ reaches } L_j).$$

Note that $\ell_j \cdot \mathbb{1}(T(x) \text{ reaches } L_j)$ depends on at most t bits, and so D will $\varepsilon/2^t$-fool it. Therefore,

$$\begin{aligned}
&|\mathbb{E}_{x\sim U_n}[T(x)] - \mathbb{E}_{x\sim D}[T(x)]| \\
&\quad\leq \sum_{j=1}^{2^t} |\mathbb{E}_{x\sim U}[\ell_j \cdot \mathbb{1}(T(x) \text{ reaches } L_j)] - \mathbb{E}_{x\sim D}[\ell_j \cdot \mathbb{1}(T(x) \text{ reaches } L_j)]| \\
&\quad\leq \sum_{j=1}^{2^t} \varepsilon/2^t \\
&\quad= \varepsilon.
\end{aligned}$$

◀

With this, we are now ready to prove our final PRG for $\mathsf{GC}^0(\log m)$.

▶ **Theorem 40.** *For $m, n \in \mathbb{N}$ and $w \leq \log m$, there is an ε-error PRG with $O((w\log^{d-1}(m) + \log^2(m))\log(m/\varepsilon)\log\log m)$ seed length for $\mathsf{GC}^0_d(\log m) \circ \mathsf{AND}_w$ circuits.*

Proof. Let $\ell = 512w$ and $t = 10\log(m/\varepsilon)$.

- Let $H : [n] \to [\ell]$ be a $2t$-wise independent hash function which needs $O(t\log n) = O(\log n \log(m/\varepsilon))$ bits. We will let H_i be an n-bit string such that $(H_i)_j = 1$ iff $H(j) = i$.
- Let $\varepsilon' = \varepsilon/(\ell \cdot 2^{t+1})$ and set $X_1, \ldots, X_\ell$ to be strings that ε'-fool $\mathsf{GC}^0_{d-1}(\log m) \circ \mathsf{AND}_{\log m}$ circuits of size $4m^2$ if $d \geq 2$, which by the inductive hypothesis uses

  $$O(\log^{d-1}(m)\log(m/\varepsilon)\log\log m)$$

  random bits per X_i. If $d = 1$, use the PRG from Lemma 39 giving a seed length, which needs $O(\log(m/\varepsilon))$ seed per X_i.

- Let Y be a string that $\varepsilon/((64mw)^{t+w+1}(2m)^{2t})$-fools CNF of size m^2, samplable using

  $$O(\log m \log((mw)^{t+w} m^t/\varepsilon) \log\log m) = O(\log(m/\varepsilon) \log^2 m \log\log m)$$

  bits.

  The PRG will sample the above strings and output the following computation

  $$Y \oplus (X_1 \wedge H_1) \oplus \cdots \oplus (X_\ell \wedge H_\ell)$$

where $\wedge$ and $\oplus$ are the bitwise AND and XOR operations, respectively. Therefore, we get a total seed length of

$$\begin{aligned} O(\log n \log(m/\varepsilon) + \ell \log(m/\varepsilon) + \log(m/\varepsilon) \log^2 m \log\log m) \\ = O(\log(m/\varepsilon) \log^2 m \log\log m) \end{aligned}$$

if $d = 1$ and

$$\begin{aligned} O(\log n \log(m/\varepsilon) + \ell \log^{d-1}(m) \log(m/\varepsilon) \log\log m + \log(m/\varepsilon) \log^2 m \log\log m) \\ = O((w \log^{d-1}(m) + \log^2(m)) \log(m/\varepsilon) \log\log m). \end{aligned}$$

Let C be an arbitrary $\mathsf{GC}^0_d(\log m) \circ \mathsf{AND}_w$ circuit, and let $U_1, \ldots U_\ell$ be independent and uniform n-bit strings. Like in [18] we use a hybrid argument to prove the theorem using the hybrid distributions

$$D_i = Y \oplus \bigoplus_{1 \le j \le i} (U_i \wedge H_i) \oplus \bigoplus_{i < j \le \ell} (X_i \wedge H_i)$$

for $0 \le i \le \ell$. Noting D_0 is the PRG output, while D_ℓ is a uniform string, it suffices to show

$$|\mathbb{E}_{x \sim D_{i-1}}[C(x)] - E_{x \sim D_i}[C(x)]| \le \varepsilon/\ell \tag{8}$$

for all $1 \le i \le \ell$, from which summing over all i and applying the triangle inequality gets the desired result.

Notice each H_i is $2t$-wise $\frac{1}{\ell}$-bounded. Conditioned on H, note that $Z_i := Y \oplus \bigoplus_{1 \le j < i}(U_i \wedge H_i) \oplus \bigoplus_{i < j \le \ell}(X_i \wedge H_i)$ $\varepsilon/((64mw)^{t+w+1}(2m)^{2t})$-fools CNF of size m^2 since Y does. Let $\mathcal{F}$ be the collection of all bottom depth-2 $\mathsf{G}(k) \circ \{\mathsf{AND}_w, \mathsf{OR}_w\}$ subcircuits of C. Therefore, if we let $\mathcal{E}$ be the event that $\mathcal{F}|_{\rho(H_i, Z_i)}$ has no $\log m$-partial depth-t DT, by Theorem 38 it follows

$$\begin{aligned} &\Pr_{H, Y, U_{<i}, X_{>i}}[\mathcal{E}] \\ &\quad \le 4(m 2^{\log m})^{t/\log m}(64w/\ell)^t + (64mw)^{t+w}(2m)^{2t \log m/\log m} \cdot \frac{\varepsilon}{(64mw)^{2t}(2m)^{2t}} \\ &\quad \le 4 \cdot 2^{2t}(1/16)^t + \frac{\varepsilon}{(64mw)^{40\log(m/\varepsilon) - \log m}} \\ &\quad \le 4(1/4)^t + \frac{\varepsilon}{4\ell} \\ &\quad \le \frac{\varepsilon}{2\ell} \end{aligned}$$

Conditioning on $\neg\mathcal{E}, H, Y, U_{<i}, X_{>i}$, we see upon replacing all depth 2 subcircuits with DT_ts and applying Lemma 30, $C|_{\rho(H_i, Z_i)}$ is computable by a depth-t DT where each leaf L_j is an $\mathsf{GC}^0_{d-1}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_w$ circuit of size $\le m \cdot 2^{\log m} + m \le 2m^2$, and will become a

$\mathsf{DT}_{\log m}$ tree if $d = 1$. In either case, we see by construction that X_i will fool it. Formally, we note that conditioned on the good events above, we have $C_{\rho(H_i,Z_i)}(y) = \sum_{j=1}^{2^t} L_j(y) \cdot \mathbb{1}\{T(y) \text{ reaches } L_j\}$. By construction of X_i,

$$|\mathbb{E}_{X_i}[L_j(X_i + Z_i) \cdot \mathbb{1}\{T(X_i + Z_i) \text{ reaches } L_j\}] - \mathbb{E}_{U_i}[L_j(U_i + Z_i) \cdot \mathbb{1}\{T(U_i + Z_i) \text{ reaches } L_j\}]| \leq \frac{\varepsilon}{\ell \cdot 2^{t+1}}$$

Summing over all $1 \leq j \leq 2^t$, applying the Triangle Inequality, and using linearity of expectation, we see

$$\mathbb{E}_{X_i}[C_{\rho(H_i,Z_i)}(X_i + Z_i)] - \mathbb{E}_{U_i}[C_{\rho(H_i,Z_i)}(U_i + Z_i)]| \leq \frac{\varepsilon}{2\ell}.$$

Therefore,

$$|\mathbb{E}_{x\sim D_{i-1}}[C(x)] - \mathbb{E}_{x\sim D_i}[C(x)]| \leq \Pr[\mathcal{E}] + \Pr[\neg\mathcal{E}] \cdot \frac{\varepsilon}{2\ell} \leq \frac{\varepsilon}{\ell}$$

and (8) is proven. ◀

From this, we immediately get PRGs for size-m $\mathsf{GC}^0_d(\log m)$ circuits.

▶ **Theorem 41.** *For every m, n, d and $\varepsilon > 0$, there is an ε-PRG for size-m $\mathsf{GC}^0_d(\log m)$ with seed length $O((\log^{d-1}(m) + \log^2(m)) \log(m/\varepsilon) \log\log m)$*

Proof. Add trivial fan-in 1 gates to the bottom so that we effectively have a $\mathsf{GC}^0_d(\log m)\circ\mathsf{AND}_1$ circuit. By Theorem 40, we can fool this with seed length

$$O((\log^{d-1}(m) + \log^2(m)) \log(m/\varepsilon) \log\log m).$$ ◀

4.4 Fourier Spectrum Bounds for $\mathsf{GC}^0(k)$

Linial, Mansour, Nisan, and Tal showed that many notions of the Fourier spectrum of a function class is intimately related [16, 19, 30]. [30] writes out four key properties and conveniently describes the implications existing between them. We report a slightly altered version here.

▶ **Theorem 42** ([16, 19, 30])**.** *Say for a class of functions, $\mathcal{C}$ we have the following property.*

- ***ESFT**: Exponentially small Fourier tails. For all $f \in \mathcal{C}$,*

  $$W^{\geq k}[f] \leq Ce^{-\Omega(k/t)}.$$

  for some constant C.

Then, $\mathcal{C}$ also satisfies the following for some constant C'.

- ***SLPT**: Switching lemma type property. For all $f \in \mathcal{C}, d, p$,*

  $$\Pr_{\rho\sim R_p}[\deg(C|_\rho) \geq d] \leq C' \cdot O(pt)^d.$$

- ***InfK**: Bounded total degree-k influence. For all $f \in \mathcal{C}$,$0 \leq k \leq n$,*

  $$Inf^k[f] \leq C' \cdot O(t)^k.$$

- ***L1**: Bounded L_1 norm at the kth level. For all $f \in \mathcal{C}$, $0 \leq k \leq n$,*

  $$\sum_{|S|=k} |\widehat{f}(x)| \leq C' \cdot O(t)^k$$

- ***FMC**: Fourier mass concentration. For all $f \in \mathcal{C}$, f is ε-concentrated on $t^{O(t\log(1/\varepsilon))}$ coefficients.*

Due to the above unification result, it appears like we can bootstrap Corollary 32 to give us a plethora of information about the Fourier spectrum of $\mathsf{GC}^0(k)$. Unfortunately, upon closer inspection, the Corollary doesn't quite give the exact property of SLPT. We instead show that $\mathsf{GC}^0(k)$ has ESFT. Our proofs will use the following lemma.

▶ **Lemma 43** ([16]). *For $f : \{\pm1\}^n \to \{\pm1\}$, $0 \le \ell \le n$, and $p \in [0,1]$,*

$$W^{\ge \ell}[f] \le 2\mathbb{E}_{\rho \sim R_p} W^{\ge kp}[f|_\rho]$$

We first start off with depth 2 circuits.

▶ **Lemma 44.** *Let f be a $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_w$. Then*

$$W^{\ge \ell}[f] \le 2 \cdot 2^{-\ell/80w+k}$$

Proof. Let $p = 1/40w$ and $t = \ell/80w$. By Theorem 20, if $\rho \sim R_p$, $f|_\rho$ becomes a depth-t DT with $\ge 1 - (20w/40w)^t 2^k = 1 - 2^{-t+k}$ probability. Such trees have no Fourier mass above level t. Say ρ is good if $f|_\rho$ does indeed become a DT. Using Lemma 43 it follows

$$\begin{aligned} W^{\ge \ell}[f] &\le 2\mathbb{E}_{\rho \sim R_p}[W^{\ge p\ell}[f|_\rho]] \\ &\le 2\mathbb{E}_{\rho \sim R_p}[W^{\ge \ell/40w}[f|_\rho] | \rho \text{ is good}] + 2 \cdot 2^{-t+k} \\ &\le 2 \cdot 2^{-\ell/80w+k}. \end{aligned}$$

◀

We can now use this as a base case to prove ESFT for GC^0. We will need to utilize the following lemma.

▶ **Lemma 45** ([30]). *Let $f : \{\pm1\}^n \to \{\pm1\}$, $0 \le \ell \le n$, and let T be a depth d decision tree such that for any leaf ℓ and the corresponding restriction ρ_ℓ induced by the root-to-leaf path, we have $W^{\ge \ell}[f|_{\rho_\ell}] \le \varepsilon$. Then $W^{\ge \ell+d}[f] \le \varepsilon$.*

We now state and prove the theorem. Define the *effective size* of a Boolean circuit to be the number of gates in the circuit at distance 2 or more from the inputs.

▶ **Theorem 46.** *Let f be a $\mathsf{GC}^0_d(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_w$ circuit with effective size m. Then*

$$W^{\ge \ell}[f] \le 4^d \cdot 2^{-\frac{\ell}{80w(128(k+\log m))^{d-1}}+k}$$

Proof. We apply induction. The base case of $d = 1$ is taken care of by Lemma 44.

We now prove the inductive step for depth d. Sample $\rho \sim R_p$ with

$$p = \frac{1}{128w(m2^k)^{1/(k+\log m)}} = \frac{1}{128w},$$

and let $t = p\ell/2 = \ell/256w$. By Theorem 28, all the bottom depth-2 $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_w$ subcircuits of $f|_\rho$ can be calculated by a $(k + \log m)$-partial depth-t decision tree with probability $\ge 1 - 4 \cdot 2^{-t}$. By Lemma 30, this implies $f|_\rho$ becomes a $(t, \mathsf{GC}^0_{d-1}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_{k+\log m})$-tree T. Furthermore, each leaf circuit has effective size $\le m$. Call ρ *good* if $f|_\rho$ simplifies to such a tree. Then

$$W^{\ge \ell}[f] \le 2\mathbb{E}_{\rho \sim R_p}[W^{\ge p\ell}[f|_\rho]] \le 2\mathbb{E}_{\rho \sim R_p}[W^{\ge p\ell}[f|_\rho] | \rho \text{ is good}] + 8 \cdot 2^{-t}.$$

Fix a good ρ. For a leaf L of T, let τ_L be the restriction induced by the path to L in T. We know by Lemma (cite) that

$$W^{\ge p\ell}[f|_\rho] \le \max_{\text{leaf } L} W^{\ge p\ell - t}[f|_{\rho \circ \tau_L}] \le \max_{\text{leaf } L} W^{\ge p\ell/2}[f|_{\rho \circ \tau_L}].$$

As $f|_{\rho\circ\tau_L}$ is a $\mathsf{GC}^0_{d-1}(k)\circ\{\mathsf{AND},\mathsf{OR}\}_{k+\log m}$ circuit for every L we can then use the inductive hypothesis to bound

$$\max_{\text{leaf } L} W^{\geq p\ell/2}[f|_{\rho\circ\tau_L}] \leq 4^{d-1}\cdot 2^{-\frac{p\ell}{80(k+\log m)(128(k+\log m))^{d-2}}+k} = 4^{d-1}\cdot 2^{-\frac{\ell}{80w(128(k+\log m))^{d-1}}+k}.$$

Putting this all together, we get

$$\begin{aligned} W^{\geq\ell}[f] &\leq 2\mathbb{E}_{\rho\sim R_p}[W^{\geq p\ell}[f|_\rho]|\rho \text{ is good}] + 8\cdot 2^{-t} \\ &\leq 2\cdot 4^{d-1}\cdot 2^{-\frac{\ell}{80w(128(k+\log m))^{d-1}}+k} + 8\cdot 2^{-\ell/256w} \\ &\leq 4^d\cdot 2^{-\frac{\ell}{80w(128(k+\log m))^{d-1}}+k} \end{aligned}$$

◀

We can bootstrap Theorem 46 with Theorem 42 to yield the following properties about $\mathsf{GC}^0(k)$

▶ **Theorem 47.** *Let f be a size-m $\mathsf{GC}^0_d(k)$ circuit and define $t := (k+\log m)^{d-1}$. Then the following is true for some C*

1. *ESFT: $W^{\geq\ell}[f] \leq C\cdot 2^k\cdot 2^{-\Omega\left(\frac{\ell}{t}\right)}$.*
2. *SLTP: For all $0<p<1$, $\Pr_{\rho\sim R_p}[\deg(f|_\rho)\geq\ell] \leq C\cdot O(pkt)^\ell$.*
3. *InfK: $\mathrm{Inf}^\ell[f] \leq C\cdot O(kt)^\ell$.*
4. *L1: $\sum_{|S|=\ell}|\widehat{f}(x)| \leq C\cdot O(kt)^\ell$.*
5. *FMC: f is ε-concentrated on $2^{O((k+\log(1/\varepsilon))t\log t)}$ coefficients.*

where f and any hidden constants only depend on d.

Proof. Add a trivial $(d+1)$-st layer of AND_1 gates at the base of f and apply Theorem 46 to deduce that

$$W^{\geq\ell}[f] \leq 4^d\cdot 2^{-\frac{\ell}{80(128(k+\log m))^{d-1}}+k},$$

proving the first item. Now since we know $W^{\geq\ell}[C]\leq 1$ (by Parseval's) and $k\geq 1$, it follows that

$$W^{\geq\ell}[f] \leq (W^{\geq\ell}[C])^{1/k} \leq C_d\cdot 2^{-\Omega\left(\frac{\ell}{kt}\right)}.$$

Therefore, the second, third, and fourth items follow by applying Theorem 42 (as well as a version of the fifth item with weaker parameters). We now prove Item 5.

Notice that for $w := t\cdot O(k+\log(1/\varepsilon))$, we have by Item 1 that $W^{\geq w}[f]\leq\varepsilon/2$. Now by Item 4,

$$\sum_{|S|<w}|\widehat{f}(S)| \leq \sum_{i=0}^{w-1} O(kt)^i \leq (C'kt)^w. \tag{9}$$

Now let $\mathcal{F} = \{S : |S|<w \text{ and } |\widehat{f}(S)| \geq \frac{\varepsilon/2}{(C'kt)^w}\}$. Notice that

$$\begin{aligned} \sum_{S\in\mathcal{F}}\widehat{f}(S)^2 &= 1 - \sum_{|S|\geq w}\widehat{f}(S)^2 - \sum_{|S|<w, S\notin\mathcal{F}}\widehat{f}(S)^2 \\ &\geq 1-\varepsilon/2 - \frac{\varepsilon/2}{(C'kt)^w}\sum_{|S|<w}|\widehat{f}(S)| \\ &\geq 1-\varepsilon. \end{aligned}$$

By Equation (9), the maximum number of terms in $\mathcal{F}$ can be at most

$$(C'kt)^w / \left(\frac{\varepsilon/2}{(C'kt)^w} \right) = 2(C'kt)^{2w}/\varepsilon = 2^{O(w \log(kt) + \log(1/\varepsilon))} = 2^{O((k+\log(1/\varepsilon))t \log t)}$$

and thus Item 5 is proved. ◀

As a first application, the work of Kushilevitz and Mansour [13] allows us to translate FMC to learnability results.

▶ **Lemma 48** ([13]). *Let f be a Boolean function such that there exists a t-sparse multivariate polynomial g (over the Fourier basis) such that $\mathbb{E}_{x \sim U_n}[(f(x) - g(x))^2] \leq \varepsilon$. There exists a randomized algorithm, whose running time is polynomial in $t, n, 1/\varepsilon, \log(1/\delta)$ such that given blackbox access to f and $\delta > 0$, outputs a function h such that over the randomness of the algorithm,*

$$\Pr[\mathbb{E}_{x \sim U_n}[(f(x) - h(x))^2] \leq O(\varepsilon)] \geq 1 - \delta.$$

Using this lemma, we can derive a learning algorithm for $\mathsf{GC}^0(k)$.

▶ **Theorem 49.** *There exists an algorithm such that given blackbox access to any $C \in \mathsf{GC}^0_d(k)$ of size m and $\delta > 0$, outputs a function h such that over the randomness of the algorithm,*

$$\Pr[\mathbb{E}_{x \sim U_n}[(f - h)^2] \leq O(\varepsilon)] \geq 1 - \delta.$$

Furthermore, this algorithm runs in poly$(n, 2^{\tilde{O}((k+\log(1/\varepsilon))(k+\log m)^{d-1})}, 1/\varepsilon, \log(1/\delta))$

Proof. From Theorem 47, for any $C \in \mathsf{GC}^0_d(k)$, there exists g of sparsity

$$t = 2^{\tilde{O}((k+\log(1/\varepsilon))(k+\log m)^{d-1})},$$

created by taking the Fourier expansion of C and only keeping the ε-concentrated coefficients $\mathcal{S} \subset 2^{[n]}$, such that

$$\mathbb{E}_{x \sim \{\pm 1\}^n}[(C(x) - g(x))^2] \leq \mathbb{E}_{x \sim \{\pm 1\}^n}\left[\left(\sum_{S \notin \mathcal{S}} \widehat{C}(S) x^S \right)^2 \right] = \sum_{S \notin \mathcal{S}} \widehat{C}(S)^2 \leq \varepsilon.$$

The result then follows by Lemma 48. ◀

We also can prove a new correlation bound result with this Fourier spectrum. It is known that MAJ is a symmetric function that has $O_d(\log^{d-1}(m)/\sqrt{n})$ correlation against size-m $\mathsf{AC}^0_d[\oplus]$ circuits. A natural question to ask is whether Majority is special in this regard, or if a random symmetric function (use $n+1$ coin tosses to assign a bit to each Hamming level) will display $O_d(\log^{d-1}(m)/n^\alpha)$ correlation against size-m $\mathsf{AC}^0_d[\oplus]$ circuits for some α. Tal ([30], Theorem 6.1) used ESFT and L1 of AC^0 to prove that random symmetric functions (or more specifically balanced symmetric functions) display $O_d(\log^{d-1}(m)/\sqrt{n})$ correlation against size-m AC^0_d circuits, so it is natural to believe that this should similarly be true against $\mathsf{AC}^0[\oplus]$. Unfortunately, since PAR has all its Fourier weight at level n, this proof approach is doomed to fail for $\mathsf{AC}^0[\oplus]$ circuits, as the class doesn't demonstrate ESFT. However, we can now give partial progress towards this goal by showing a random symmetric function uncorrelates with $\mathsf{GC}^0(k)$ circuits, as this class contains gates which calculate parity as long as the Hamming weight of the input is at most k. This result can be seen as finding out how general of a circuit class we can stretch the Fourier argument before we reach the roadblock on this approach demonstrated by PAR.

▶ **Theorem 50.** *Let $f \in \mathsf{GC}^0_d(k)$, and let g be a symmetric function, both mapping $\{\pm 1\}^n \to \{\pm 1\}$, and let $(k+\log m)^{d-1} \le O\left(\frac{n}{k+\log n}\right)^{1/3}$. Then*

$$\mathsf{corr}(f,g) := \mathbb{E}_x[f(x)g(x)] \le |\widehat{g}(\emptyset)| + \frac{C_d k(k+\log m)^{d-1}}{\sqrt{n}}$$

Proof. We note for ℓ' to be picked later, we can decompose

$$\begin{aligned}\mathsf{corr}(f,g) &= |\mathbb{E}_x[f(x)g(x)]| \\ &= \left|\sum_{S\subset[n]} \widehat{f}(S)\widehat{g}(S)\right| \\ &\le |\widehat{g}(\emptyset)| + \sum_{|S|<\ell'} |\widehat{f}(S)\widehat{g}(S)| + \sum_{|S|\ge\ell} |\widehat{f}(S)\widehat{g}(S)|. \end{aligned} \tag{10}$$

We will bound the first summation using L1, and the second summation by ESFT. The second summation can be bounded as follows using Cauchy-Schwarz.

$$\sum_{|S|\ge\ell} |\widehat{f}(S)\widehat{g}(S)| \le \sqrt{W^{\ge\ell'}[f]\cdot W^{\ge\ell'}[g]} \le \sqrt{4^d \cdot 2^{-\frac{\ell'}{80(128(k+\log m))^{d-1}}+k}} \le 1/\sqrt{n} \tag{11}$$

if we set $\ell' = c_d(k+\log n)(k+\log m)^{d-1}$ for some constant c_d only depending on d. Now to bound the first summation, note since g is symmetric, $\widehat{g}(S)$ is constant over all S of same cardinality. Therefore,

$$|\widehat{g}(S)| = \sqrt{\widehat{g}(S)^2} = \sqrt{\frac{1}{\binom{n}{|S|}}\sum_{S';|S'|=|S|}\widehat{g}(S')^2} \le \sqrt{\frac{1}{\binom{n}{|S|}}}.$$

Hence using L1 from Theorem 47, we can bound

$$\sum_{|S|<\ell'} |\widehat{f}(S)\widehat{g}(S)| \le \sum_{1\le\ell<\ell'} \sqrt{\frac{1}{\binom{n}{\ell}}}\sum_{|S|=\ell}|\widehat{f}(S)| \le \sum_{1\le\ell<\ell'}\left(\frac{C_d k(k+\log m)^{d-1}}{\sqrt{n/\ell}}\right)^{\ell} \tag{12}$$

where C_d is some constant depending on d. We bound this sum by a geometric series of the same first term and with common ratio $1/2$. Indeed, we see that the ratio of consecutive terms will be

$$\frac{\left(\frac{C_d k(k+\log m)^{d-1}}{\sqrt{n/(\ell+1)}}\right)^{\ell+1}}{\left(\frac{C_d k(k+\log m)^{d-1}}{\sqrt{n/\ell}}\right)^{\ell}} = \frac{C_d k(k+\log m)^{d-1}}{\sqrt{n}}\sqrt{\frac{(\ell+1)^{\ell+1}}{\ell^\ell}} \le \frac{C_d k(k+\log m)^{d-1}}{\sqrt{n}}\sqrt{e\ell'} \le 1/2$$

where the last inequality follows from the assumption $(k+\log m)^{d-1} \le O\left(\frac{n}{k+\log n}\right)^{1/3}$. Hence the quantity in Equation (12) can be upper bounded by twice the first term, so

$$\sum_{|S|<\ell'} |\widehat{f}(S)\widehat{g}(S)| \le \frac{C_d k(k+\log m)^{d-1}}{\sqrt{n}}.$$

Hence from (10),

$$\mathsf{corr}(f,g) \le |\widehat{g}(\emptyset)| + \frac{1}{\sqrt{n}} + \frac{C_d k(k+\log m)^{d-1}}{\sqrt{n}}.$$

◀

5 Open Problems

We conclude with some directions for future research.

- Our tightness result in Theorem 52 uses a function in $\mathsf{G}(k) \circ \mathsf{AND}_w$, but it is not known whether a k-$\mathsf{OR} \circ \mathsf{AND}_w$ circuit can saturate the bound. In particular, is Theorem 20 tight for $\mathsf{AC}^0(k)$ or $\mathsf{TC}^0(k)$ circuits?
- It was already noted that Corollary 34 is tight in essentially every way possible for $\mathsf{GC}^0(k)$ circuits. However, all our tightness results (Theorem 52) use constructions that abuse the generality of $\mathsf{GC}(k)$. Are there constructions exhibiting tightness which are in $\mathsf{TC}^0(k)$ or $\mathsf{AC}^0(k)$? Alternatively, can we obtain stronger size lower bounds if we were only concerned with $\mathsf{AC}^0(k)$ or even $\mathsf{TC}^0(k)$ circuits? Either finding a pathological construction in $\mathsf{AC}^0(k)/\mathsf{TC}^0(k)$ or proving stronger lower bounds for these weaker circuit class would be interesting.
- We touched up on how a result of Allender and Koucký ([2], Theorem 3.8) states that there exists an absolute constant C_{AK} such that MAJ_n can be written as an $\mathsf{AC}^0(n^\varepsilon)$ circuit with depth $\leq C_{AK}/\varepsilon$ and size $O(n^{1+\varepsilon})$. [2] actually only use $\mathsf{AND}_2, \mathsf{OR}_2$, and $\mathsf{MAJ}_{n^\varepsilon}$ gates. If we were allowed all the gate classes in $\mathsf{AC}^0(n^\varepsilon)$, could we find a better construction (more specifically a lower depth blowup)? That way, we would get a stronger reduction from proving bounds on TC to $\mathsf{AC}^0(n^\varepsilon)$ where we only need to show size lower bounds on smaller depth $\mathsf{AC}^0(n^\varepsilon)$ circuits.
- Are there any other applications of the generalized switching lemma? Due to the versatility of this theorem, it can essentially be plugged in wherever the classical switching lemma was used to get more general results. Perhaps this can generalize other results or even push a switching lemma argument that initially wouldn't go through (the remark after Theorem 52 gives an example of the general switching lemma giving stronger bounds than the classical one).

References

1 Scott Aaronson and Avi Wigderson. Algebrization: A new barrier in complexity theory. *ACM Trans. Comput. Theory*, 1(1), February 2009. doi:10.1145/1490270.1490272.

2 Eric Allender and Michal Koucký. Amplifying lower bounds by means of self-reducibility. *J. ACM*, 57(3), March 2010. doi:10.1145/1706591.1706594.

3 N. Alon, O. Goldreich, J. Hastad, and R. Peralta. Simple construction of almost k-wise independent random variables. In *Proceedings [1990] 31st Annual Symposium on Foundations of Computer Science*, pages 544–553 vol.2, 1990. doi:10.1109/FSCS.1990.89575.

4 Theodore Baker, John Gill, and Robert Solovay. Relativizations of the p?=np question. *SIAM Journal on Computing*, 4(4):431–442, 1975. doi:10.1137/0204037.

5 Norbert Blum. A boolean function requiring 3n network size. *Theor. Comput. Sci.*, 28:337–345, 1984. doi:10.1016/0304-3975(83)90029-4.

6 Lijie Chen and Roei Tell. Bootstrapping results for threshold circuits "just beyond" known lower bounds. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, pages 34–41, New York, NY, USA, 2019. Association for Computing Machinery. doi:10.1145/3313276.3316333.

7 Johan Håstad. On the correlation of parity and small-depth circuits. *SIAM Journal on Computing*, 43(5):1699–1708, 2014. doi:10.1137/120897432.

8 Pooya Hatami, William M. Hoza, Avishay Tal, and Roei Tell. Fooling constant-depth threshold circuits (extended abstract). In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 104–115, 2022. doi:10.1109/FOCS52979.2021.00019.

9 Russell Impagliazzo, William Matthews, and Ramamohan Paturi. A satisfiability algorithm for ac0. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '12, pages 961–972, USA, 2012. Society for Industrial and Applied Mathematics.

10 Russell Impagliazzo, Ramamohan Paturi, and Michael E. Saks. Size–depth tradeoffs for threshold circuits. *SIAM Journal on Computing*, 26(3):693–707, 1997. doi:10.1137/S0097539792282965.

11 Daniel M. Kane and Ryan Williams. Super-linear gate and super-quadratic wire lower bounds for depth-two and depth-three threshold circuits. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16, pages 633–643, New York, NY, USA, 2016. Association for Computing Machinery. doi:10.1145/2897518.2897636.

12 Zander Kelley. An improved derandomization of the switching lemma. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 272–282. ACM, 2021. doi:10.1145/3406325.3451054.

13 Eyal Kushilevitz and Yishay Mansour. Learning decision trees using the fourier spectrum. *SIAM Journal on Computing*, 22(6):1331–1348, 1993. doi:10.1137/0222080.

14 Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1996. doi:10.1017/CBO9780511574948.

15 Jiatu Li and Tianqi Yang. 3.1n - o(n) circuit lower bounds for explicit functions. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, pages 1180–1193, New York, NY, USA, 2022. Association for Computing Machinery. doi:10.1145/3519935.3519976.

16 Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *J. ACM*, 40(3):607–620, July 1993. doi:10.1145/174130.174138.

17 Shachar Lovett and Srikanth Srinivasan. Correlation bounds for poly-size ac0 circuits with n(1-o(1)) symmetric gates. In Leslie Ann Goldberg, Klaus Jansen, R. Ravi, and José D. P. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 640–651, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.

18 Xin Lyu. Improved pseudorandom generators for ac0 circuits. In *Proceedings of the 37th Computational Complexity Conference*, CCC '22, Dagstuhl, DEU, 2022. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.CCC.2022.34.

19 Yishay Mansour. An o(nlog log n) learning algorithm for dnf under the uniform distribution. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT '92, pages 53–61, New York, NY, USA, 1992. Association for Computing Machinery. doi:10.1145/130385.130391.

20 Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. doi:10.1017/CBO9781139814782.

21 Igor Carboni Oliveira, Rahul Santhanam, and Srikanth Srinivasan. Parity Helps to Compute Majority. In Amir Shpilka, editor, *34th Computational Complexity Conference (CCC 2019)*, volume 137 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 23:1–23:17, Dagstuhl, Germany, 2019. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.CCC.2019.23.

22 Aaron Potechin. On the approximation resistance of balanced linear threshold functions. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2019, pages 430–441, New York, NY, USA, 2019. Association for Computing Machinery. doi:10.1145/3313276.3316374.

23 Alexander Razborov and Avi Wigderson. n(log n) lower bounds on the size of depth-3 threshold cicuits with and gates at the bottom. *Information Processing Letters*, 45(6):303–307, 1993. doi:10.1016/0020-0190(93)90041-7.

24 Alexander A. Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical notes of the Academy of Sciences of the USSR*, 41:333–338, 1987.

25 Alexander A Razborov and Steven Rudich. Natural proofs. *Journal of Computer and System Sciences*, 55(1):24–35, 1997. doi:10.1006/jcss.1997.1494.

26 Rocco A. Servedio. Every linear threshold function has a low-weight approximator. In *Proceedings of the 21st Annual IEEE Conference on Computational Complexity*, CCC '06, pages 18–32, USA, 2006. IEEE Computer Society. doi:10.1109/CCC.2006.18.

27 Rocco A. Servedio and Li-Yang Tan. Luby-Velickovic-Wigderson Revisited: Improved Correlation Bounds and Pseudorandom Generators for Depth-Two Circuits. In Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2018)*, volume 116 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 56:1–56:20, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.APPROX-RANDOM.2018.56.

28 Rocco A. Servedio and Li-Yang Tan. Improved pseudorandom generators from pseudorandom multi-switching lemmas. *Theory of Computing*, 18(4):1–46, 2022. doi:10.4086/toc.2022.v018a004.

29 R. Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, STOC '87, pages 77–82, New York, NY, USA, 1987. Association for Computing Machinery. doi:10.1145/28395.28404.

30 Avishay Tal. Tight Bounds on the Fourier Spectrum of AC0. In Ryan O'Donnell, editor, *32nd Computational Complexity Conference (CCC 2017)*, volume 79 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 15:1–15:31, Dagstuhl, Germany, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.CCC.2017.15.

31 Luca Trevisan and Tongke Xue. A derandomized switching lemma and an improved derandomization of ac0. In *2013 IEEE Conference on Computational Complexity*, pages 242–247, 2013. doi:10.1109/CCC.2013.32.

32 Salil P. Vadhan. Pseudorandomness. *Foundations and Trends® in Theoretical Computer Science*, 7(1–3):1–336, 2012. doi:10.1561/0400000010.

33 Emanuele Viola. Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM Journal on Computing*, 36(5):1387–1403, 2007. doi:10.1137/050640941.

A Deferred Proofs

A.1 Showing $\mathsf{TC}^0(k) \subset \mathsf{GC}^0(k)$

Here, we prove that circuits created by biased LTF gates are indeed contained in $\mathsf{GC}^0(k)$.

▶ **Theorem 51.** *Any* THR *gate* f *with balance* $\leq k$ *(see Definition 3) is, upon negating certain input bits, in* $\mathsf{G}(k)$.

Proof. Let $f : \{0,1\}^n \to \{\pm 1\}$ be defined as $f(x) = \operatorname{sgn}(\sum_{i=1}^{n} w_i(-1)^{x_i} - \theta)$. By negating input bits, we can assume each $w_i \geq 0$. Furthermore, since the definition of $\mathsf{G}(k)$ is symmetric (solely depends on the sum of input bits), we can assume WLOG that $0 \leq w_1 \leq \cdots \leq w_n$. Since f has balance $\leq k$, we know that $-\sum_{i\leq k} w_i + \sum_{i>k} w_i < \theta$. Assuming $-\sum_{i\leq k} w_i + \sum_{i>k} w_i < |\theta|$, we will show f is an orlike $\mathsf{G}(k)$ gate (an analogous proof will show the case for $-\theta$ and andlike).

Consider x such that $\sum x_i \geq k$. Let $S_x \subset [n]$ denote the set where $i \in S_x$ iff $x_i = 1$. It follows that

$$f(x) = \text{sgn}\left(-\sum_{i\in S_x} w_i + \sum_{i\notin S_x} w_i - \theta\right)$$
$$= \text{sgn}\left(-\sum_{i\leq k} w_i + \sum_{i>k} w_i - \theta + 2\sum_{i\in[k]\setminus S_x} w_i - 2\sum_{i\in[k+1,n]\cap S_x} w_i\right) \tag{13}$$

We know by assumption that $-\sum_{i\leq k} w_i + \sum_{i>k} w_i - \theta \leq 0$. Since $|S_x| \geq k$,

$$|[k] \setminus S_x| = k - |k \cap S_x| \leq |[k+1, n] \cap S_x|,$$

and each element in $[k] \setminus S_x$ is strictly smaller than each element in $[k+1, n] \cap S_x$. Combining these observations with the fact $w_1 \leq \cdots \leq w_n$, it follows

$$2\sum_{i\in[k]\setminus S_x} w_i - 2\sum_{i\in[k+1,n]\cap S_x} w_i \leq 0.$$

Therefore,

$$-\sum_{i\leq k} w_i + \sum_{i>k} w_i - \theta + 2\sum_{i\in[k]\setminus S_x} w_i - 2\sum_{i\in[k+1,n]\cap S_x} w_i \leq 0.$$

Combining this with (13) it follows $f(x) = -1$ for any x with $\sum x_i \geq k$. Hence f is an orlike $\mathsf{G}(k)$ gate as desired. ◀

A.2 Tightness of the $\mathsf{GC}^0(k)$ Switching Lemma and Correlation Bounds

In this section, we give constructions which show that various bounds we establish are indeed tight. We first show that the switching lemma we established is tight.

▶ **Theorem 52.** *Let p, w, t, k be parameters such that $pkw < 1/2$ and $t > k/2$. There exists a $\mathsf{G}(k) \circ \mathsf{AND}_w$ circuit C such that*

$$\Pr_{\rho\sim R_p}[\mathsf{DT}(C|_\rho) \geq t] \geq 2^{k/2}(.5pw)^t$$

Proof. We will take $C = \mathsf{PAR}_{kw}$. To see that this is computable in $\mathsf{G}(k) \circ \mathsf{AND}_w$, write

$$\mathsf{PAR}_{kw}(x) = \mathsf{PAR}_k(\mathsf{PAR}_w(x_1, \ldots, x_w), \ldots, \mathsf{PAR}_w(x_{(k-1)w+1}, \ldots, x_{kw})).$$

Now, write out each bottom layer PAR_w as a size $w2^{w-1}$ CNF which takes the ORs of the 2^{w-1} AND clauses corresponding to w-bit inputs with an odd number of ones. Notice that for any assignment,

- at most *one* of the 2^{w-1} clauses under each OR can simultaneously be satisfied,
- which implies at most k of the bottom layer clauses can be simultaneously satisfied.

By the first bullet point, we can turn the OR gates into PAR gates, turning C into a $\mathsf{PAR}_{k2^{w-1}} \circ \mathsf{AND}_w$ circuit. By the second bullet point, we can replace the top $\mathsf{PAR}_{k2^{w-1}}$ gate with the gate $\mathsf{G} \in \mathsf{G}(k)$ which calculates parity if at most k input bits are one, and outputs 0 otherwise. Consequently, C can be calculated by a $\mathsf{G}(k) \circ \mathsf{AND}_w$ circuit.

We can now directly calculate the simplification probability. Notice that if $\geq t$ variables are alive in ρ, then $C|_\rho$ must have decision tree depth $\geq t$ (since even if $d-1$ input bits are known, parity remains ambiguous). Hence we calculate

$$\begin{aligned}\Pr_{\rho\sim R_p}[\mathsf{DT}(C|_\rho)\geq t] &\geq \sum_{i=t}^{kw}\binom{kw}{i}p^i(1-p)^{kw-i}\\ &\geq (1-p)^{kw}\left(\frac{kw}{t}\right)^t\left(\frac{p}{1-p}\right)^t\\ &\geq (1-pkw)2^{k/2}\left(\frac{pw}{1-p}\right)^t\\ &\geq 2^{k/2}(.5pw)^t\end{aligned}$$

where the third inequality follows from the fact $(k/t)^t$ is increasing in t for $t > k/2$ ◀

▶ Remark 53. Notice that if PAR_n was written as a width n-CNF and the classical switching lemma was applied, we would get a "failure to simplify" probability bound of $\leq (5pn)^t$. However, if we rewrite PAR_n as a $\mathsf{G}(t\log(n/t))\circ\mathsf{AND}_{n/t\log(n/t)}$ circuit in a similar manner as above, and use Theorem 20, we get a bound of $\leq (20pn/t\log(n/t))^t 2^{t\log(n/t)} \leq (20pn^2/t^2)^t$, which is asymptotically stronger when $t=\omega(\sqrt{n})$. This shows that we can potentially obtain tighter parameters by expressing functions as a more compact $\mathsf{GC}^0(k)$ circuit and applying Theorem 20 rather than using the classical switching lemma on a larger AC^0 circuit computing the same function.

We will now show that the circuit lower bound established in Corollary 34 is tight. We thank an anonymous reviewer for pointing us to this construction.

▶ **Theorem 54.** *There exists a size-*$2^{O((n/k)^{\frac{1}{d-1}})}$ $\mathsf{GC}^0_d(k)$ *circuit which computes* PAR_n.

Proof. Split the input string into k blocks of size n/k bits each. We can compute the parity of each of the n/k-size blocks straightforwardly using a depth $d-1$ circuit made out of $\mathsf{PAR}_{(n/k)^{\frac{1}{d-1}}}$ gates; iteratively group the bits into blocks of $(n/k)^{\frac{1}{d-1}}$, and use a gate to take the parity of each block, thereby creating a depth $d-1$ tree of $\mathsf{PAR}_{(n/k)^{\frac{1}{d-1}}}$ gates. Finally, we can take a PAR_k of these k depth-$(d-1)$ circuits to get a depth d circuit which computes the parity of all n bits. We now focus on converting this parity-riddled circuit to a $\mathsf{GC}^0(k)$ one.

Consider the top depth-2 subcircuit, which is a $\mathsf{PAR}_k\circ\mathsf{PAR}_{(n/k)^{\frac{1}{d-1}}}$ circuit. Notice that $\mathsf{PAR}_k\in\mathsf{G}(k)$ and $\mathsf{PAR}_{(n/k)^{\frac{1}{d-1}}}$ is trivially computable by a decision tree in $\mathsf{DT}_{(n/k)^{\frac{1}{d-1}}}$. Therefore by Lemma 30, this top subcircuit can be replaced by a size $2^{O((n/k)^{\frac{1}{d-1}})}$ $\mathsf{G}(k)\circ\mathsf{OR}_{(n/k)^{\frac{1}{d-1}}}$. Consequently, we have converted our original depth-d circuit into a new one where the first 2 layers are made from gates in $\mathsf{G}(k)$.

We now use the fact that any $\mathsf{PAR}_{(n/k)^{\frac{1}{d-1}}}$ can be expressed as a size $2^{O((n/k)^{\frac{1}{d-1}})}$ CNF or DNF to convert the remaining $\mathsf{PAR}_{(n/k)^{\frac{1}{d-1}}}$ gates to AND/OR gates while preserving the depth. Convert the third layer $\mathsf{PAR}_{(n/k)^{\frac{1}{d-1}}}$ gates to a DNF (OR of ANDs), and then collapse the 2nd and 3rd layer as they both consist solely of OR gates. The third layer now consist of AND gates, so replace the fourth layer $\mathsf{PAR}_{(n/k)^{\frac{1}{d-1}}}$ gates with a CNF (AND of ORs) to again induce a collapse of the consecutive AND layers. Repeat this procedure down to the bottom of the circuit.

Clearly after this procedure, all gates are in $\mathsf{G}(k)$ (in fact, all but the top gate are AND or OR). Notice that at each stage, we increased the circuit depth by 1 when plugging in the DNF/CNF, but then reduced the circuit depth when collapsing layers of the same gate type. Hence the final circuit is still of depth d. We already calculated the top depth-2 subcircuit is of size $2^{O((n/k)^{\frac{1}{d-1}})}$. Assuming we didn't collapse any gates, we would have replaced each of the $o(n^{1-\varepsilon})$ parity gates with a size-$2^{O((n/k)^{\frac{1}{d-1}})}$ circuit, so clearly the final circuit size is at most

$$2^{O((n/k)^{\frac{1}{d-1}})} + o(n/k) \cdot 2^{O((n/k)^{\frac{1}{d-1}})} \leq 2^{O((n/k)^{\frac{1}{d-1}})}.$$

Therefore at the end of this procedure, we get our desired circuit. ◀

We now show that not only is the size lower bound tight, but the average case correlation bound established in Theorem 33 as well.

▶ **Theorem 55.** *Assume $m \geq k^d$. There is a $\mathsf{GC}^0(k)$ circuit C of size $\leq m$ such that*

$$\mathbb{E}_x[(-1)^{C(x)+\mathsf{PAR}(x)}] \geq 2^{\Omega(k)} \cdot 2^{-O(n/(k+\log m)^{d-1})}$$

Proof. Let $M = \max\{k, c_d \log m\}$, where c_d is a constant such that the parity over $(c \log m)^{d-1}$ bits can be computed by an AC^0_d circuit of size $\leq m$. Split the input into $\lceil n/M^{d-1} \rceil$ blocks of size $\leq M^{d-1}$. If $M = c_d \log m$, each block can be calculated by an AC^0_d circuit of size m, and if $M = k$, each block can be calculated by a $\leq k^d \leq m$-size $\mathsf{GC}^0_{d-1}(k)$ circuit using a tree of PAR_ks. Now if $\lceil n/M^{d-1} \rceil = 1$, this circuit computes the parity of all n bits and we are done, so assume $\lceil n/M^{d-1} \rceil \geq 2$.

Join all the $\lceil n/M^{d-1} \rceil$ subcircuits by the gate G defined to compute parity if the Hamming weight of the input is at most k, and to equal 0 otherwise. Clearly $G \in \mathsf{G}(k)$. Therefore, if the subcircuits were constructed to be in AC^0_d, we can collapse the top two layers into one using Lemma 30 (similar to Theorem 54), giving us a $\mathsf{GC}^0_d(k)$ circuit. If the subcircuits were $\mathsf{GC}^0_{d-1}(k)$, we trivially get a $\mathsf{GC}^0_d(k)$ gate after adding G. In the case more than k of the $\lceil n/M^{d-1} \rceil$ input blocks have parity 1, our circuit will be constant, and thus will agree with parity about half the time. Let us crudely lower bound the correlation in this case to be 0. When $\leq k$ have parity 1, the top gate computes parity exactly. Therefore our correlation is simply the probability at most k of the input blocks have parity 1, which is simply

$$2^{-\lceil n/M^{d-1} \rceil} \sum_{i \leq k} \binom{\lceil n/M^{d-1} \rceil}{k} \geq 2^{-\lceil n/M^{d-1} \rceil} \cdot 2^{\Omega(k)} \geq 2^{\Omega(k)} \cdot 2^{-O(n/(k+\log m)^{d-1})}.$$

This gives our correlation lower bound as desired. ◀

A.3 Proof of the $\mathsf{GC}^0(k)$ Multi-Switching Lemma

We prove the multi-switching lemma here.

▶ **Theorem 56** (Proof of Theorem 28). *Let $\mathcal{F} = \{F_1, \ldots, F_m\}$ be a list of $\mathsf{G}(k) \circ \mathsf{AND}_w$ circuits on $\{0,1\}^m$. Then*

$$\Pr_{\rho \sim R_p}[\mathcal{F}|_\rho \text{ does not have } r\text{-partial depth-}t \text{ } DT] \leq 4(64(2^k m)^{1/r} pw)^t$$

Proof. We follow the proof in [18] exactly, where the only difference is that the "Canonical Partial Decision Tree" (CPDT) will use the modified CDT we created in Algorithm 1, the definition of global witnesses (resp. global partial witnesses) will now use the witnesses (resp. partial witnesses) that we defined in Definition 22 (resp. Definition 24), and the "Global Witness Searcher" will run the modified witness searcher we created in Algorithm 2.

Consider the following CPDT procedure.

Algorithm 3 Canonical Partial Decision Tree.

Input: A list of $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_w$ circuits $\mathcal{F} = \{F_1, \ldots, F_m\}$, black-box access to a string $\beta \in \{0,1\}^n$, and an auxiliary string $z \in \{0,1\}^n$.

initialize:
- $x \leftarrow (\star)^n$.
- $j \leftarrow 1$.
- counter $\leftarrow 0$.

while counter $< t$ **do**
- Find the smallest $i \geq j$ such that $\mathsf{DT}(F_i|_x) > w$. If no such i exists, exit the loop.
- $y \leftarrow (\star)^n$.
- $I \leftarrow \emptyset$.
- **while** $F_i|_{x \circ y}(\star)$ *is not constant and* counter $< t$ **do**
  - $C_{i,q} \leftarrow$ the term that $T_{F_i|_{x\circ y}}$ from Algorithm 1 will query.
  - $B_{i,q} \leftarrow$ the set of unknown variables in $C_{i,q}|_{x\circ y}$.
  - $y_{B_{i,q}} \leftarrow z_{B_{i,q}}$.
  - $I \leftarrow I \cup B_{i,q}$.
  - counter $\leftarrow$ counter $+ |B_{i,q}|$.
- **end**
- Query β_I, and set $x_I \leftarrow \beta_I$.
- $j \leftarrow i$.

end

return x

With this, we can define the following notion of a "global witness" to intuitively be a transcript on adversarially chosen inputs.

▶ **Definition 57.** *Let t, w be two integers. Consider a list of $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_w$ circuits $\mathcal{F} = \{F_1, \ldots, F_m\}$. Suppose $\rho \in \{0,1,\star\}^n$ is a restriction. Let $(R, L_i, S_i, W_i, \beta_i)$ be a tuple, where*

- *$1 \leq R \leq \frac{t}{r}$ is an integer;*
- *$1 \leq L_1 \leq L_2 \leq \cdots \leq L_R \leq m$ is a list of R non-decreasing indices;*
- *$S_1 \ldots, S_R$ is a list of R integers such that $\sum_{i=1}^{R} S_i \in [t, t+w]$;*
- *$W_1, \ldots, W_R$ is a list of witnesses (as per Definition 22). For every $i \in [R]$, W_i has size S_i;*
- *$\beta_1, \ldots, \beta_R$ are R strings where $|\beta_i| = S_i$ for every $i \in [R]$.*

We call the tuple a (r,t)-global witness for ρ, if it satisfies the following.

1. *Set $\rho_1 = \rho$. W_1 is a S_1-witness for $F_{L_1}|_{\rho_1}$.*
2. *For every $i \geq 2$, let $I_{i-1} \subseteq [n]$ be the set of variables involved in W_{i-1}. Note that $|I_{i-1}| = S_{i-1}$ since the size of W_{i-1} is S_{i-1}. Identify β_{i-1} as a partial assignment in $\{0,1,\star\}^n$ where only the part $\beta_{i-1,I_{i-1}}$ is set and other coordinates are filled in with $\star$. Construct $\rho_i = \rho_{i-1} \circ \beta_{i-1}$. Then W_i is a S_i-witness for $F_{L_i}|_{\rho_i}$.*

The size of the global witness is defined as $\sum_{i=1}^{R} S_i$.

▶ **Lemma 58.** *Consider a list of* $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_w$ *circuits* $\mathcal{F} = \{F_1, \ldots, F_m\}$*. Suppose* $\rho \in \{0,1,\star\}^n$ *is a restriction such that* $\mathcal{F}|_\rho$ *does not have* w*-partial depth-*t *decision tree. Then there exists an* (r,t)*-global witness for* ρ*.*

Proof. Same as the proof of Corollary 1 in [18]. ◀

We now define *partial global witnesses*, as there are far too many global witnesses to union bound over.

▶ **Definition 59.** *Let* t, w *be two integers. Consider a list of* $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_w$ *circuits* $\mathcal{F} = \{F_1, \ldots, F_m\}$*. Suppose* $\rho \in \{0,1,\star\}^n$ *is a restriction. Let* $(R, L_i, S_i, P_i, \beta_i)$ *be a tuple, where*

- $1 \leq R \leq \frac{t}{r}$ *is an integer;*
- $1 \leq L_1 \leq L_2 \leq \cdots \leq L_R \leq m$ *is a list of* R *non-decreasing indices;*
- $S_1 \ldots, S_R$ *is a list of* R *integers such that* $\sum_{i=1}^{R} S_i \in [t, t+w]$*;*
- $P_1, \ldots, P_R$ *is a list of partial witnesses. For every* $i \in [R]$*,* P_i *has size* S_i*.*
- $\beta_1, \ldots, \beta_R$ *are* R *strings where* $|\beta_i| = S_i$ *for every* $i \in [R]$*.*

We call $(R, L_i, S_i, P_i, \beta_i)$ *a* (r,t)*-global partial witness for* ρ*, if we can complete* P_i *to get a witness* W_i *for every* $i \in [R]$*, such that* $(R, L_i, S_i, W_i, \beta_i)$ *is a global witness for* ρ*.*

By a simple induction, one can show the following claim.

▷ Claim 60. Given a global partial witness for ρ, there is exactly one way to complete it and get a global witness for ρ.

We now construct a global witness searcher that will reconstruct a global witness from a partial one using advice, and present it as Algorithm 4.

Algorithm 4 Global Witness Searcher.

Input: A list of DNFs $\mathcal{F} = \{F_1, \ldots, F_m\}$, a global partial witness $(R, L_i, S_i, P_i, \beta_i)$, and an advice $z \in \{0,1\}^n$.

initialize:

  $c \leftarrow 1$.

  $\rho^{(1)} \leftarrow \rho$.

while $c \leq R$ **do**

  Run Algorithm 2 on $(F_{L_c}, \rho^{(c)}, P_c, y)$. If it reports ERROR, report ERROR and terminate the procedure. Otherwise let W_c be the witness returned.

  $I_c \leftarrow$ the set of variables involved in W_c.

  Identify β_c as a partial assignment, where only β_{I_c} is fixed.

  $\rho^{(c+1)} \leftarrow \rho^{(c)} \circ \beta_c$.

  $c \leftarrow c + 1$.

end

return x

We note the following important lemma about the searcher, which we denote $\mathcal{S}$.

▶ **Lemma 61.** *Let* P *be a size* S *global partial witness. Say* ρ *is* good *if* P *is a global partial witness for* ρ*, then*

$$\Pr_z[\mathcal{S}(z, P) \text{ is a global witness for } \rho(\Lambda, z) | \rho \text{ is good}] = 2^{-S}$$

Proof. Same as proof of Lemma 8 in [18], but we instead appeal to Lemma 26 whenever Lyu's proof refers to Lemma 6. ◀

By this lemma, we have

$$\begin{aligned}\Pr_{\Lambda,z}[\rho(\Lambda,z) \text{ is good for } P] &= \mathbb{E}_\Lambda \frac{\Pr_z[\mathcal{S}(z,P) \text{ is a global witness for } \rho(\Lambda,z)]}{\Pr_z[\mathcal{S}(z,P) \text{ is a global witness for } \rho(\Lambda,z) | \rho \text{ is good}]} \\ &= 2^S \mathbb{E}_\Lambda \Pr_z[\mathcal{S}(z,P) \text{ is a global witness for } \rho(\Lambda,z)] \\ &\leq 2^S \mathbb{E}_z \Pr_\Lambda[\mathcal{S}(z,P) \text{ is a global witness for } \rho(\Lambda,z)] \\ &\leq (2p)^s \end{aligned} \tag{14}$$

where the last inequality follows from the fact that ρ needs to keep the S variables specified by the global witness alive in order to have any hope of being witnessed by it. By the $(t+w)$-wise p-boundedness of Λ, this happens with probability $\leq p^S$.

Finally, if we let N_S be the number of global witnesses of size S, we can see by using Lemma 58 and Claim 60, along with (14) that

$$\begin{aligned}\Pr_\rho[\mathcal{F}|_\rho \text{ has no } r\text{-partial depth-}t \ \mathsf{DT}] &\leq \sum_P \Pr_\rho[P \text{ is a global partial witness for } \rho] \\ &\leq \sum_{S=t}^{t+w} N_S (2p)^S \end{aligned} \tag{15}$$

We can upper bound N_S as follows.

- There are $\leq \frac{t}{r} \cdot \binom{m}{t/r} \leq 2m^{t/r}$ ways to pick (R, L_i)
- There are $\leq 2^S$ choices for (S_i) (since there are $\leq 2^{n-1}$ ways to write n as an ordered partition)
- From Theorem 20, we know that there are $(8w)^S 2^k$ partial witnesses of size S, giving a total amount of $\leq \prod_i (8w)^{S_i} 2^k = (8w)^S 2^{kR} \leq (8w)^S 2^{kt/r}$
- There are clearly $2^{\sum S_i} = 2^S$ possibilities for (β_i).

Combining all this tells us that $N_S \leq 2m^{t/r}(32w)^S 2^{kt/r}$. Hence, from (15), we deduce

$$\begin{aligned}\Pr_\rho[\mathcal{F}|_\rho \text{ has no } r\text{-partial depth-}t \ \mathsf{DT}] &\leq \sum_{S=t}^{t+w} N_S (2p)^S \\ &\leq m^{t/r} 2^{kt/r} \sum_{S=t}^{t+w} (32pw)^S \\ &\leq 4(64(2^k m)^{1/r} pw)^t \end{aligned}$$ ◀

▶ **Theorem 62** (Proof of Theorem 38). *Let $\mathcal{F} = \{F_1, \ldots, F_m\}$ be a list of size m $\mathsf{G}(k) \circ \{\mathsf{AND}, \mathsf{OR}\}_w$ circuits. Let (Λ, z) be a joint random variable such that*

- *Λ is a $(t+w)$-wise p-bounded subset of $[n]$*
- *Conditioned on any instance of Λ, z ε-fools CNF of size $\leq m^2$.*

Then

$$\Pr_{\Lambda,z}[\mathcal{F}|_{\rho(\Lambda,x)} \text{ has no } r\text{-partial depth-}t \ \mathsf{DT}] \leq 4(m2^k)^{t/r}(64pw)^t + (64wm)^{t+w}(2m)^{2kt/r} \cdot \varepsilon$$

Proof. All steps of the proof of Theorem 28 follow identically until we reach the step

$$\Pr_{\Lambda,z}[F|_{\rho(\Lambda,x)} \text{ has no } w\text{-partial depth-}t \text{ DT}]$$
$$= \Pr \sum_{(R,L,S_i,P_i,\beta_i)} \Pr_{\Lambda,z}[(R,L,S_i,P_i,\beta_i) \text{ is a global partial witness for } \rho(\Lambda,z)]$$

From Claim 60, we can deduce the event decomposes

$$\mathbb{1}\{(R,L_i,S_i,P_i,\beta_i) \text{ is a global partial witness for } \rho(\Lambda,z)\}$$
$$= \sum_{W_i:\text{completion of } P_i} \mathbb{1}\{(R,L_i,S_i,W_i,\beta_i) \text{ is a global witness for } \rho(\Lambda,z)\}.$$

For a fixed Λ and (R,L_i,S_i,W_i,β_i), we can let

$$h_\Lambda^{(R,L_i,S_i,W_i,\beta_i)}(z) = \mathbb{1}\{(R,L_i,S_i,W_i,\beta_i) \text{ is a global witness for } \rho(\Lambda,z)\}.$$

We will now show that h is a predicate computable by a small size CNF, so that z will fool it. h is true iff W_i is a witness for $F_{L_i}|_{\rho(\Lambda,z)\circ\beta_1\ldots\beta_{i-1}}$ for all $1 \le i \le R$. Now for each i, one can verify $W_i = (r', \ell_i, s_i, B_i, \alpha_i)$ is a witness by a size m CNF as follows.

- For all $j < \ell_1$, C_j is falsified by $\rho(\Lambda,z)$. This is true iff $(\neg C_1) \wedge (\neg C_2) \wedge \cdots \wedge (\neg C_{\ell_1})$. Notice $\neg C_i$ becomes an OR clause by De Morgan's Law
- C_{j_1} is satisfied, which is an AND of variables.
- C_j is falsified by $\rho(\Lambda,z)\circ\alpha_1$ for $\ell_1 < j < \ell_2$, which is true iff $(\neg C_{\ell_1+1}) \wedge \cdots \wedge (\neg C_{\ell_2-1})$. Each $\neg C_i$ is an OR clause
- and so on and so forth until we verify $C_{j_{r'}}$.

Each bullet points gives a disjunction of (maybe trivial) conjunctions, so for all bullet points to hold, we simply take the AND of all of them, resulting in a CNF whose size is bounded by m (since our CNF is essentially F_{L_i} but with some gates and negations changed. Hence, if we want to verify W_i simultaneously over all i, we take the AND of all $R \le m$ CNFs to get a size m^2 CNF. Hence, z ε-fools h. From Theorem 20, we know over a uniform string x,

$$\sum_{(R,L_i,S_i,W_i,\beta_i)} \mathbb{E}_\Lambda \mathbb{E}_x h_\Lambda^{(R,L_i,S_i,W_i,\beta_i)}(x) \le 4(m2^k)^{t/r}(64pw)^t$$

Therefore, over z, we have

$$\begin{aligned}\Pr_{\Lambda,z}[F|_{\rho(\Lambda,x)} \text{ has no } w\text{-partial depth-}t \text{ DT}] &= \sum_{(R,L_i,S_i,W_i,\beta_i)} \mathbb{E}_\Lambda \mathbb{E}_z h_\Lambda^{(R,L_i,S_i,W_i,\beta_i)}(z) \\ &\le \sum_{(R,L_i,S_i,W_i,\beta_i)} \mathbb{E}_\Lambda(\varepsilon + \mathbb{E}_x h_\Lambda^{(R,L_i,S_i,W_i,\beta_i)}(x)) \\ &\le 4(m2^k)^{t/r}(64pw)^t + \sum_{(R,L_i,S_i,W_i,\beta_i)} \varepsilon.\end{aligned}$$

The number of tuples (R,L_i,S_i,W_i,β_i) can be bounded as follows.

- From the proof of Theorem 56, we know there are $2m^{t/r}2^{kt/r}(32w)^S$ many global partial witnesses of size S.
- We now multiply by the number of (ℓ_i) possible for each partial P_i of size S_i, which is at most $\binom{m}{t+k} < m^{S_i+k}$. Hence, the total number of (W_i) over all $1 \le i \le R$ is

$$\prod_i m^{S_i+k} = m^{S+kR} \le m^{S+kt/r}$$

Hence the total number of size-S tuples is upper bounded by $2m^{t/r}2^{kt/r}(32w)^S m^{S+kt/r} \leq (32mw)^S(2m)^{2kt/r}$. Summing over all $t \leq S \leq t+w$ gives us a grand total of

$$\sum_{S=t}^{t+w}(32mw)^S(2m)^{2kt/r} \leq (64mw)^S(2m)^{2kt/r}.$$

Therefore,

$$\Pr_{\Lambda,z}[F|_{\rho(\Lambda,x)} \text{ has no } r\text{-partial depth-}t\ \mathsf{DT}] \leq 4(m2^k)^{t/r}(64pw)^t + (64mw)^{t+w}(2m)^{2kt/r} \cdot \varepsilon.\blacktriangleleft$$