Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 157 (2018)

A Fixed-Depth Size-Hierarchy Theorem for $\mathrm{AC}^0[\oplus]$ via the Coin Problem

Nutan Limaye * Karteek Sreenivasaiah † Srikanth Srinivasan ‡
Utkarsh Tripathi § S. Venkitesh ¶

February 20, 2019

Abstract

In this work we prove the first *Fixed-depth Size-Hierarchy Theorem* for uniform $\mathrm{AC}^0[\oplus]$. In particular, we show that for any fixed d, the class $\mathcal{C}_{d,k}$ of functions that have uniform $\mathrm{AC}^0[\oplus]$ formulas of depth d and size n^k form an infinite hierarchy. We show this by exibiting the first class of *explicit* functions where we have nearly (up to a polynomial factor) matching upper and lower bounds for the class of $\mathrm{AC}^0[\oplus]$ formulas.

The explicit functions are derived from the *δ-Coin Problem*, which is the computational problem of distinguishing between coins that are heads with probability $(1+\delta)/2$ or $(1-\delta)/2$, where δ is a parameter that is going to 0. We study the complexity of this problem and make progress on both upper bound and lower bound fronts.

- **Upper bounds.** For any constant $d \geq 2$, we show that there are *explicit* monotone AC^0 formulas (i.e. made up of AND and OR gates only) solving the δ-coin problem that have depth d, size $\exp(O(d(1/\delta)^{1/(d-1)}))$, and sample complexity (i.e. number of inputs) $\mathrm{poly}(1/\delta)$. This matches previous upper bounds of O'Donnell and Wimmer (ICALP 2007) and Amano (ICALP 2009) in terms of size (which is optimal) and improves the sample complexity from $\exp(O(d(1/\delta)^{1/(d-1)}))$ to $\mathrm{poly}(1/\delta)$.
- **Lower bounds.** We show that the above upper bounds are nearly tight (in terms of size) even for the significantly stronger model of $\mathrm{AC}^0[\oplus]$ formulas (which are also allowed NOT and Parity gates): formally, we show that any $\mathrm{AC}^0[\oplus]$ formula solving the δ-coin problem must have size $\exp(\Omega(d(1/\delta)^{1/(d-1)}))$. This strengthens a result of Shaltiel and Viola (SICOMP 2010), who prove a $\exp(\Omega((1/\delta)^{1/(d+2)}))$ lower bound for $\mathrm{AC}^0[\oplus]$, and a result of Cohen, Ganor and Raz (APPROX-RANDOM 2014), who show a $\exp(\Omega((1/\delta)^{1/(d-1)}))$ lower bound for the smaller class AC^0.

The upper bound is a derandomization involving a use of Janson's inequality (from probabilistic combinatorics) and classical combinatorial designs; as far as we know, this is the first such use of Janson's inequality. For the lower bound, we prove an optimal (up to a constant factor) degree lower bound for multivariate polynomials over $\mathbb{F}_2$ solving the δ-coin problem, which may be of independent interest.

*Indian Institute of Technology, Bombay. Email: nutan@cse.iitb.ac.in

†Indian Institute of Technology Hyderabad. Email: karteek@iith.ac.in

‡Indian Institute of Technology, Bombay. Email: srikanth@math.iitb.ac.in

§Indian Institute of Technology, Bombay. Supported by the Ph.D. Scholarship of NBHM, DAE, Government of India. Email: utkarshtripathi.math@gmail.com

¶Indian Institute of Technology, Bombay. Supported by the Senior Research Fellowship of HRDG, CSIR, Government of India. Email: venkitesh.mail@gmail.com

ISSN 1433-8092

1 Size-Hierarchy theorems for $\mathrm{AC}^0[\oplus]$

Given any natural computational resource, an intuitive conjecture one might make is that access to more of that resource results in more computational power. *Hierarchy theorems* make this intuition precise in various settings. Classical theorems in Computational complexity theory such as the time and space hierarchy theorems [HS65, SHI65, Coo73, FS04] show that Turing Machine-based computational models do become strictly more powerful with more access to time or space respectively.

The analogous questions in the setting of Boolean circuit complexity deal with the complexity measures of depth and size of Boolean circuits. Both of these have been intensively studied in the case of AC^0 circuits and by now, we have near-optimal *Depth* and *Size-hierarchy* theorems for AC^0 circuits [Has89, Ros08a, Ama10, HRST17]. Our focus in this paper is on size-hierarchy theorems for $\mathrm{AC}^0[\oplus]$ circuits.

Essentially, a size-hierarchy theorem for a class of circuits says that there are Boolean functions $f : \{0,1\}^n \to \{0,1\}$ that can be computed by circuits of some size $s = s(n)$ but not by circuits of size significantly smaller than s, e.g. $\sqrt{s}$. However, stated in this way, such a statement is trivial to prove, since we can easily show by counting arguments that there are more functions computed by circuits of size s than by circuits of size $\sqrt{s}$ and hence there must be a function that witnesses this separation. As is standard in the setting of circuits, what is interesting is an *explicit* separation. (Equivalently, we could consider the question of separating uniform versions of these circuit classes.)

Strong results in this direction are known in the setting of AC^0 circuits (i.e. constant-depth Boolean circuits made up of AND, OR and NOT gates).

Size hierarchy theorem for AC^0. In order to prove a size-hierarchy theorem for AC^0, we need an explicit function that has circuits of size s but no circuits of size less than $\sqrt{s}$. If we fix the depth of the circuits under consideration, a result of this form follows immediately from the *tight* exponential AC^0 circuit lower bound of Håstad [Has89] from the 80s. Håstad shows that any depth-d AC^0 circuit for the Parity function on n variables must have size $\exp(\Omega(n^{1/(d-1)}))$; further, this result is tight, as demonstrated by a folklore depth-d AC^0 upper bound of $\exp(O(n^{1/(d-1)}))$. Using both the lower and upper bounds for Parity, we get a separation between circuits of size $s_0 = \exp(O(n^{1/(d-1)}))$ and s_0^{ε} for some fixed $\varepsilon > 0$. The same separation also holds between s and s^{ε} for any s such that $s \leq s_0$, since we can always take the Parity function on some $m \leq n$ variables so that the above argument goes through. We thus get a *Fixed-depth Size-Hierarchy theorem* for AC^0 for any $s(n) \leq \exp(n^{o(1)})$.

Even stronger results are known for AC^0. Note that the above results do not separate, for example, size s circuits of depth 2 from size s^{ε} circuits of depth 3. However, recent results of Rossman [Ros08b] and Amano [Ama10] imply the existence of explicit functions[1] that have AC^0 circuits of depth 2 and size n^k (for any constant k) but not AC^0 circuits of *any* constant depth and size $n^{k-\varepsilon}$.

$\mathrm{AC}^0[\oplus]$ setting. Our aim is to prove size-hierarchy theorems for $\mathrm{AC}^0[\oplus]$ circuits (i.e. constant-depth Boolean circuits made up of AND, OR, NOT and $\oplus$ gates).[2] As for AC^0, we have known expo-

[1] The explicit functions are the k-clique problem and variants.

[2] Our results also extend straightforwardly to $\mathrm{AC}^0[\mathrm{MOD}_p]$ gates for any constant prime p (here, a MOD_p gate accepts if the sum of its input bits is non-zero modulo p).

nential lower bounds for this circuit class from the 80s, this time using the results of Razborov [Raz89] and Smolensky [Smo87]. Unfortunately, however, most of these circuit lower bounds are *not* tight. For instance, we know that the Majority function on n variables does not have $\mathrm{AC}^0[\oplus]$ circuits of depth d and size $\exp(\Omega(n^{1/2(d-1)}))$, but the best upper bounds are worse than $\exp(O(n^{1/(d-1)}))$ (in fact, the best known upper bound [KPPY84] is an AC^0 circuit of size $\exp(O(n^{1/(d-1)})(\log n)^{1-1/(d-1)}))$.[3] As a direct consequence of this fact, we do not even have *fixed-depth* size-hierarchy theorems of the above form for $\mathrm{AC}^0[\oplus]$: the known results only yield a separation between circuits of size s and circuits of size $\exp(O(\sqrt{\log s}))$, which is a considerably worse result.

In this work we present the first fixed-depth size-hierarchy theorem for $\mathrm{AC}^0[\oplus]$ circuits. Formally, we prove the following.

Theorem 1 (Fixed-depth size-hierarchy theorem)**.** *There is an absolute constant $\varepsilon_0 \in (0,1)$ such that the following holds. For any fixed depth $d \geq 2$, and for infinitely many n and any $s = s(n) = \exp(n^{o(1)})$, there is an explicit monotone depth-d AC^0 formula F_n on n variables of size at most s such that any $\mathrm{AC}^0[\oplus]$ formula computing the same function has size at least s^{ε_0}.*

In particular, if $\mathcal{C}_{d,k}$ denotes the family of languages that have uniform $\mathrm{AC}^0[\oplus]$ formulas of depth d and size n^k, then the hierarchy $\mathcal{C}_{d,1} \subseteq \mathcal{C}_{d,2} \cdots$ is infinite.

We can also get a similar result for $\mathrm{AC}^0[\oplus]$ *circuits* of fixed depth d by using the fact that circuits of depth d and size s_1 can be converted to formulas of depth d and size s_1^d. Using this idea, we can get a separation between circuits (in fact formulas) of depth d and size s and circuits of depth d and size $s^{\varepsilon_0/d}$.

To get this (almost) optimal fixed-depth size-hierarchy theorem we design an explicit function f and obtain tight upper and lower bounds for it for each fixed depth d. The explicit function is based on the *Coin Problem*, which we define below.

2 The Coin Problem

The Coin Problem is the following natural computational problem. Given a two-sided coin that is heads with probability either $(1+\delta)/2$ or $(1-\delta)/2$, decide which of these is the case. The algorithm is allowed many independent tosses of the coin and has to accept in the former case with probability at least 0.9 (say) and accept in the latter case with probability at most 0.1. The formal statement of the problem is given below.

Definition 2 (The δ-Coin Problem)**.** *For any $\alpha \in [0,1]$ and integer $N \geq 1$, let D_α^N be the product distribution over $\{0,1\}^N$ obtained by setting each bit to 1 independently with probability α.*

Let $\delta \in (0,1)$ be a parameter. Given an $N \in \mathbb{N}$, we define the probability distributions $\mu_{\delta,0}^N$ and $\mu_{\delta,1}^N$ to be the distributions $D_{(1-\delta)/2}^N$ and $D_{(1+\delta)/2}^N$ respectively. We omit the δ in the subscript when it is clear from context.

Given a function $g : \{0,1\}^N \to \{0,1\}$, we say that g solves the δ-coin problem with error ε *if*

$$\Pr_{\boldsymbol{x} \sim \mu_0^N}[g(\boldsymbol{x}) = 1] \leq \varepsilon \text{ and } \Pr_{\boldsymbol{x} \sim \mu_1^N}[g(\boldsymbol{x}) = 1] \geq 1 - \varepsilon. \tag{1}$$

In the case that g solves the coin problem with error 0.1, we simply say that g solves the δ-coin problem (and omit mention of the error).

We say that the sample complexity *of g is N.*

[3] A similar fact is also true for the MOD_p functions, for p an odd prime.

We think of δ as a parameter that is going to 0. We are interested in both the sample complexity and *computational complexity* of functions solving the coin problem. Both these complexities are measured as a function of the parameter δ.

The problem is folklore, and has also been studied (implicitly and explicitly) in many papers in the Computational complexity literature [AB84, OW07, Ama09, SV10, BV10, Vio14, Ste13, CGR14, LV18, RS17]. It was formally introduced in the work of Brody and Verbin [BV10], who studied it with a view to devising pseudorandom generators for Read-once Branching Programs (ROBPs).

It is a standard fact that $\Omega(1/\delta^2)$ samples are necessary to solve the δ-coin problem (irrespective of the computational complexity of the underlying function). Further, the algorithm that takes $O(1/\delta^2)$ many independent samples and accepts if and only if the majority of the coin tosses are heads, does indeed solve the δ-coin problem. We call this the "trivial solution" to the coin problem.

It is not clear, however, if this is the most computationally "simple" method of solving the coin problem. Specifically, one can ask if the δ-coin problem can be solved in computational models that cannot compute the Boolean Majority function on $O(1/\delta^2)$ many input bits. (Recall that the Majority function on n bits accepts inputs of Hamming weight greater than $n/2$ and rejects other inputs.)

Such questions have received quite a bit of attention in the computational complexity literature. Our focus in this paper is on the complexity of this problem in the setting of AC^0 and $\mathrm{AC}^0[\oplus]$ circuits.

Perhaps surprisingly, the Boolean circuit complexity of the coin problem in the above models is *not* the same as the circuit complexity of the Boolean Majority function. We describe below some of the interesting upper as well as lower bounds known for the coin problem in the setting of constant-depth Boolean circuits.

Known upper bounds. It is implicit in the results of O'Donnell and Wimmer [OW07] and Amano [Ama09] (and explicitly noted in the paper of Cohen, Ganor and Raz [CGR14]) that the complexity of the coin problem is closely related to the complexity of *Promise* and *Approximate* variants of the Majority function. Here, a *promise majority* is a function that accepts inputs of relative Hamming weight at least $(1+\delta)/2$ and rejects inputs of relative Hamming weight at most $(1-\delta)/2$; and an *approximate majority* is a function that agrees with the Majority function on 90% of its inputs.[4]

O'Donnell and Wimmer [OW07] and Amano [Ama09] show that the AC^0 circuit complexity of some approximate majorities is superpolynomially *smaller* than the complexity of the Majority function. More specifically, the results in [OW07, Ama09] imply that there are approximate majorities that are computed by monotone AC^0 *formulas* of depth d and size $\exp(O(dn^{1/2(d-1)}))$, while a well-known result of Håstad [Has89] implies that any AC^0 circuit of depth d for computing the Majority function must have size $\exp(\Omega(n^{1/(d-1)}))$. For example, when $d=2$, there are approximate majorities that have formulas of size $\exp(O(\sqrt{n}))$ while any circuit for the Majority function must have size $\exp(\Omega(n))$. These upper bounds were slightly improved to AC^0 *circuits* of size $\exp(O(n^{1/2(d-1)}))$ in a recent result of Rossman and Srinivasan [RS17].

The key step in the results of [OW07, Ama09] is to show that the δ-coin problem can be solved

[4]Unfortunately, both these variants of the Majority function go by the name of "approximate majority" in the literature.

by explicit read-once[5] monotone AC^0 formulas of depth d and size $\exp(O(d(1/\delta)^{1/(d-1)}))$. (This is improved to circuits of size $\exp(O((1/\delta)^{1/(d-1)}))$ in [RS17]. However, these circuits are *not* explicit.) Compare this to the trivial solution (i.e. computing Majority on $\Theta(1/\delta^2)$ inputs) that requires AC^0 circuit size $\exp(\Omega((1/\delta)^{2/(d-1)}))$, which is superpolynomially worse.

Known lower bounds. Lower bounds for the coin problem have also been subject to a good deal of investigation. Shaltiel and Viola [SV10] show that if the δ-coin problem can be solved by a circuit C of size s and depth d, then there is a circuit C' of size $\mathrm{poly}(s)$ and depth $d+3$ that computes the Majority function on $n = \Omega(1/\delta)$ inputs. Using Håstad's lower bound for Majority [Has89], this implies that any depth-d AC^0 circuit C solving the δ-coin problem must have size $\exp(\Omega((1/\delta)^{1/(d+2)}))$. Using Razborov and Smolensky's [Raz89, Smo93] lower bounds, this also yields the same lower bound for the more powerful circuit class $\mathrm{AC}^0[\oplus]$.[6]

In a later result, Aaronson [Aar10] observed that a stronger lower bound can be deduced for AC^0 by constructing a circuit C'' of depth $d+2$ that only distinguishes inputs of Hamming weight $n/2$ from inputs of Hamming weight $(n/2)-1$. By Håstad's results, this suffices to recover a lower bound of $\exp(\Omega((1/\delta)^{1/(d+1)}))$ for AC^0, but does not imply anything for $\mathrm{AC}^0[\oplus]$ (since the parity function can distinguish between inputs of weight $n/2$ and $(n/2)-1$).

Note that these lower bounds for the δ-coin problem, while exponential, do not meet the upper bounds described above. In fact, they are quasipolynomially weaker.

Lower bounds for the closely related promise and approximate majorities were proved by Viola [Vio14] and O'Donnell and Wimmer [OW07] respectively. Viola [Vio14] shows that any $\mathrm{poly}(n)$-sized depth-d AC^0 circuit cannot compute a promise majority for $\delta = o(1/(\log n)^{d-2})$. O'Donnell and Wimmer [OW07] show that any depth-d AC^0 circuit that approximates the Majority function on 90% of its inputs must have size $\exp(\Omega(n^{1/2(d-1)})$. Using the connection between the coin problem and approximate majority, it follows that any *monotone* depth-d AC^0 circuit solving the δ-coin problem must have size $\exp(\Omega((1/\delta)^{1/(d-1)}))$ matching the upper bounds above. The lower bound of [OW07] is based on the Fourier-analytic notion of the *Total Influence* of a Boolean function (see [O'D14, Chapter 2]) and standard upper bounds on the total influence of a Boolean function computed by a small AC^0 circuit [LMN93, Bop97].

Using more Fourier analytic ideas [LMN93], Cohen, Ganor and Raz [CGR14] proved near-optimal AC^0 circuit lower bounds for the δ-coin problem (with no assumptions on monotonicity). They show that any depth-d AC^0 circuit for the δ-coin problem must have size $\exp(\Omega((1/\delta)^{1/(d-1)}))$, nearly matching the upper bound constructions above.

The Coin Problem and Size Hierarchy Theorems for $\mathrm{AC}^0[\oplus]$. Recall that to prove size-hierarchy theorems for $\mathrm{AC}^0[\oplus]$, we need to come up with explicit functions for which we can prove near-tight lower bounds. One class of functions for which the Razborov-Smolensky proof technique does yield such a lower bound is the class of approximate majorities defined above. Unfortunately, however, this does not yield an explicit separation, since the functions constructed in [OW07, Ama09, RS17] are randomized and not explicit. These circuits are obtained by starting

[5] i.e. each variable in the formula appears exactly once

[6] Using [SV10] as a black box will yield a weaker lower bound of $\exp(\Omega((1/\delta)^{1/2(d+2)}))$. However, slightly modifying the proof of Shaltiel and Viola, we can obtain circuits of depth $d+3$ that *approximate* the Majority function on $1/\delta^2$ inputs instead of computing the Majority function on $(1/\delta)$ inputs *exactly*. This yields the stronger lower bound given here.

with explicit large monotone read-once formulas for the coin problem from [OW07, Ama09] and replacing each variable with one of the n variables of the approximate majority; one can show that this probabilistic procedure produces an approximate majority with high probability. However, explicitness is then lost.

Our starting point is that instead of working with approximate majorities, we can directly work with the explicit formulas solving the coin problem. As shown in [OW07, Ama09], there are explicit formulas of size $\exp(O(d(1/\delta)^{1/(d-1)}))$ solving the δ-coin problem. Since these yield optimal-sized circuits for computing approximate majorities, it follows that the functions computed by these formulas cannot be computed by much smaller circuits.

While this is true, nevertheless the explicit formulas of [OW07, Ama09] do not yield anything non-trivial by way of size-hierarchy theorems. This is because, as noted above, these formulas are *read-once*. Hence, showing that the underlying functions cannot be computed in smaller size would prove a separation between size $s = O(n)$ circuits and circuits of size much smaller than n, which is trivial.

The way we circumvent this obstacle is to construct explicit circuits solving the δ-coin problem with optimal size and much smaller sample complexity. In fact, we are able to bring down the sample complexity from exponential to polynomial in $(1/\delta)$. This allows us to prove a size hierarchy theorem for all s up to $\exp(n^{o(1)})$.

2.1 Our results for the Coin Problem

We make progress on the complexity of the coin problem on both the upper bound and lower bound fronts.

Upper bounds. Note that the upper bound results known so far only yield circuit size and depth upper bounds for the coin problem, and do not say anything about the sample complexity of the solution. In fact, the explicit AC^0 formulas of O'Donnell and Wimmer [OW07] and Amano [Ama09] are *read-once* in the sense that each input corresponds to a distinct input variable. Hence, these results imply explicit formulas of size $s = \exp(O(d(1/\delta)^{1/(d-1)}))$ and sample size $\Theta(s)$ for the δ-coin problem. (Recall that, in contrast, the trivial solution has sample complexity only $O(1/\delta^2)$.) The sample complexity of these formulas can be reduced to $O(1/\delta^2)$ by a probabilistic argument (as essentially shown by [OW07, Ama09]; more on this below), but then we no longer have explicit formulas. The circuit construction of Rossman and Srinivasan [RS17] can be seen to use $O(1/\delta^2)$ samples, but is again not explicit.

We show that the number of samples can be reduced to $\mathrm{poly}(1/\delta)$ (where the degree of the polynomial depends on the depth d of the circuit), which is the in same ballpark as the trivial solution, while retaining both the size and the explicitness of the formulas. The result is as follows.

Theorem 3 (Explicit formulas for the coin problem with small sample complexity)**.** *Let $\delta \in (0,1)$ be a parameter and $d \geq 2$ any fixed constant. There is an explicit depth-d monotone AC^0 formula Γ_d that solves the δ-coin problem, where Γ_d has size $\exp(O(d(1/\delta)^{1/(d-1)}))$ and sample complexity $(1/\delta)^{2^{O(d)}}$. (All the constants implicit in the $O(\cdot)$ notation are absolute constants.)*

Approximate majority and the coin problem. This result may be interpreted as a "partial derandomization" of the approximate majority construction of [OW07, Ama09] in the following sense. It is implicit in [OW07, Ama09] that an approximate majority on n variables can be obtained by starting with a *monotone* circuit C solving the δ-coin problem for $\delta = \Theta(1/\sqrt{n})$, and

replacing each input of C with a random input among the n input bits on which we want an approximate majority. While, as noted above, the coin-problem-solving circuits of [OW07, Ama09] have exponential sample-complexity, our circuits only have polynomial sample-complexity, leading to a much more randomness-efficient way of constructing such an approximate majority.

Indeed, this feature of our construction is crucial for proving the Size-Hierarchy theorem for $\mathrm{AC}^0[\oplus]$ circuits (Theorem 1).

Lower bounds. As noted above, Shaltiel and Viola [SV10] prove that any $\mathrm{AC}^0[\oplus]$ circuit of depth d solving the δ-coin problem must have size at least $\exp(\Omega((1/\delta)^{1/(d+2)}))$. For the weaker class of AC^0 circuits, Cohen, Ganor and Raz [CGR14] proved an optimal bound of $\exp(\Omega((1/\delta)^{1/(d-1)}))$. We are also able to strengthen these incomparable results by proving an optimal lower bound for $\mathrm{AC}^0[\oplus]$ circuits. More formally, we prove the following.

Theorem 4 (Lower bounds for the coin problem)**.** *Say g is a Boolean function solving the δ-coin problem, then any $\mathrm{AC}^0[\oplus]$ formula of depth d for g must have size $\exp(\Omega(d(1/\delta)^{1/(d-1)}))$. (The $\Omega(\cdot)$ hides an absolute constant.)*

While the above result is stated for $\mathrm{AC}^0[\oplus]$ *formulas*, it easily implies a $\exp(\Omega((1/\delta)^{1/(d-1)}))$ lower bound for depth-d *circuits*, since any $\mathrm{AC}^0[\oplus]$ circuit of size s and depth d can be converted to an $\mathrm{AC}^0[\oplus]$ formula of size s^d and depth d. We thus get a direct extension of the results of Shaltiel and Viola [SV10] and Cohen, Ganor and Raz [CGR14].

The proof of this result is closely related to the results of Razborov [Raz89] and Smolensky [Smo87] (also see [Sze89, Smo93]) that prove lower bounds for $\mathrm{AC}^0[\oplus]$ circuits computing the Majority function. For *monotone* functions[7], the lower bound immediately follows from the standard lower bounds of [Raz89] and [Smo87] for approximate majorities[8] (actually, we need a slightly stronger lower bound for $\mathrm{AC}^0[\oplus]$ *formulas*) and the reduction [OW07, Ama09] from computing approximate majorities to the coin problem outlined above. This special case is already enough for the size-hierarchy theorem stated in Section 1.

However, to prove the result in the non-monotone setting, it is not clear how to use the lower bounds of [Raz89, Smo87] directly. Instead, we use the ideas behind these results, specifically the connections between $\mathrm{AC}^0[\oplus]$ circuits and low-degree polynomials. We show that if a function $g(x_1, \ldots, x_N)$ solves the δ-coin problem, then its degree, as a polynomial from $\mathbb{F}_2[x_1, \ldots, x_N]$, must be at least $\Omega(1/\delta)$ (independent of its sample complexity). From this statement and Razborov's [Raz89] low-degree polynomial approximations for $\mathrm{AC}^0[\oplus]$, it is easy to infer the lower bound. Further, we think that the statement about polynomials is interesting in its own right.

Note that Theorems 3 and 4 immediately imply the Fixed-depth Size Hierarchy Theorem for $\mathrm{AC}^0[\oplus]$ (Theorem 1).

[7]Recall that a function $g : \{0,1\}^m \to \{0,1\}$ is monotone if it is non-decreasing w.r.t. the standard partial order on the hypercube.

[8]The standard lower bounds of Razborov and Smolensky are usually stated for computing the hard function (e.g. Majority) *exactly*. However, it is easily seen that the proofs only use the fact that the circuit computes the function on most (say 90%) of its inputs (see, e.g. [RS17]). In particular, this yields lower bounds even for approximate majorities, which, moreover, turn out to be tight. This can be seen as an alternate proof of the (later) lower bound of O'Donnell and Wimmer [OW07, Theorem 4] for a stronger class of circuits. (The lower bound of [OW07] only holds for AC^0.)

Independent work of Chattopadhyay, Hatami, Lovett and Tal [CHLT18]. A beautiful recent paper of Chattopadhyay et al. [CHLT18] proves a result on the Fourier spectrum of low-degree polynomials over $\mathbb{F}_2$ (Theorem 3.1 in [CHLT18]) which is equivalent[9] to the degree lower bound on the coin problem mentioned above. Indeed, the main observation, which is an extension of the Smolensky [Smo87, Smo93] lower bound for the Majority function, is common to our proof as well as that of [CHLT18].

2.2 Proof Outline

Upper bounds. We start with a description of the read-once formula construction of [OW07, Ama09]. In these papers, it is shown that for every $d \geq 2$, there is an explicit read-once formula F_d that solves the δ-coin problem. This formula F_d is defined as follows. We fix a $d \geq 2$ and let $m = \Theta((1/\delta)^{1/(d-1)})$, a large positive integer. We define positive integer parameters $f_1, \ldots, f_d \approx \exp(m)$,[10] and define the formula F_d to be a read-once formula with alternating AND and OR input gates where the gates at height i in the formula all have fan-in f_i. (It does not matter if we start with AND gates or OR gates, but for definiteness, let us assume that the bottom layer of gates in the formula is made up of AND gates.) Each leaf of the formula is labelled by a distinct variable (or equivalently, the formula is read-once).

The formula F_d is shown to solve the coin problem (for suitable values of $f_1, \ldots, f_d$). Note that the size of the formula, as well as its sample complexity, is $\Theta(f_1 \cdots f_d)$, which turns out be $\exp(\Omega(md)) = \exp(\Omega(d(1/\delta)^{1/(d-1)}))$.

To show that the formula F_d solves the coin problem, we proceed as follows. For each $i \in \{1, \ldots, d\}$, let us define $\mathrm{Acc}_i^{(0)}$ (resp. $\mathrm{Rej}_i^{(0)}$) to be the probability that some subformula F_i of height i accepts (resp. rejects) an input from the distribution $\mu_0^{N_i}$ (where N_i is the sample complexity of F_i). Similarly, also define $\mathrm{Acc}_i^{(1)}$ and $\mathrm{Rej}_i^{(1)}$ w.r.t. the distribution μ_1^i. Define

$$p_i^{(b)} = \min\{\mathrm{Acc}_i^{(b)}, \mathrm{Rej}_i^{(b)}\}$$

for each $b \in \{0, 1\}$. Note that these definitions are independent of the exact subformula F_i of height i that we choose.

It can be shown via a careful analysis that for each odd $i < d$ and each $b \in \{0, 1\}$, $p_i^{(b)} = \mathrm{Acc}_i^{(b)} = \Theta(1/2^m)$ (i.e. the acceptance probability is smaller than the rejection probability and is roughly $1/2^m$) and we have

$$\frac{p_i^{(1)}}{p_i^{(0)}} = (1 + \Theta(m^i \delta)). \tag{2}$$

(Note that when $i < d - 1$, the quantity $m^i \delta = o(1)$ and so $p_i^{(0)}$ and $p_i^{(1)}$ are actually quite close to each other.) An analogous fact holds for even i and rejection probabilities, where we now measure $p_i^{(0)}/p_i^{(1)}$ instead. At $i = d - 1$, we get that the ratio is in fact a large constant. From here, it is easy to argue that F_d accepts an input from $\mu_1^{N_d}$ w.h.p., and rejects an input from $\mu_0^{N_d}$ w.h.p.. This concludes the proof of the fact that F_d solves the δ-coin problem.

[9] We thank Avishay Tal (personal communication) for pointing out to us that the results of [CHLT18] imply the degree lower bounds for the coin problem using an observation of [CHHL18]. This direction in fact works for any class of functions closed under *restrictions* (i.e. setting inputs to constants from $\{0, 1\}$).

[10] These numbers have to be chosen carefully for the proof, but we do not need to know them exactly here.

We now describe the ideas behind our derandomization. The precise calculations that are required for the analysis of F_d use crucially the fact that the formulas are read-once. In particular, this implies that we are considering the AND or OR of distinct subformulas, it is easy to compute (using independence) the probability that these formulas accept an input from the distributions $\mu_0^{N_d}$ or $\mu_1^{N_d}$. In the derandomized formulas that we construct, we can no longer afford read-once formulas, since the size of our formulas is (necessarily) exponential in $(1/\delta)$, but the number of distinct variables (i.e. the sample complexity) is required to be poly$(1/\delta)$. Thus, we need to be able to carry out the same kinds of precise computations for ANDs or ORs of formulas that share many variables.

For this, we use a tool from probabilistic combinatorics named *Janson's inequality* [Jan90, AS92]. Roughly speaking, this inequality says the following. Say we have a *monotone* formula F over n Boolean variables that is the OR of M subformulas $F_1, \ldots, F_M$, and we want to analyze the probability that F rejects a random input $\mathbf{x}$ from some product distribution over $\{0,1\}^n$. Let p_i denote the probability that F_i rejects a random input. If the F_is are variable disjoint, we immediately have that F rejects $\mathbf{x}$ with probability $\prod_i p_i$. However, when the F_is are not variable disjoint but *most pairs* of these subformulas are variable disjoint, then Janson's inequality allows us to infer that this probability is *upper bounded* by $(\prod_i p_i) \cdot (1+\alpha)$ where α is quite small. Furthermore, by the monotonicity of F and the resulting positive correlation between the distinct F_i, we immediately see that the probability that F rejects is always *lower bounded* by $\prod_i p_i$ and hence we get

$$\prod_i p_i \leq \Pr_{\mathbf{x}}[F \text{ rejects } \mathbf{x}] \leq \left(\prod_i p_i\right) \cdot (1+\alpha).$$

In other words, the estimate $\prod_i p_i$, which is an exact estimate of the rejection probability of F in the disjoint case, is a good *multiplicative* approximation to the same quantity in the correlated case. Note that this is *exactly* the kind of approximation that would allow us to recover an inequality of the form in (2) and allow an analysis similar to that of [OW07, Ama09] to go through even in the correlated setting.

Remark 5. *While Janson's inequality has been used earlier in the context of Boolean circuit complexity (for example in the work of Rossman [Ros08b, Ros14]), as far as we know, this is the first application in the area of the fact that the inequality actually yields a* multiplicative approximation *to the probability being analyzed.*

This observation motivates the construction of our derandomized formulas (with only poly$(1/\delta)$ variables). At each depth d, we construct the derandomized formula Γ_d as follows. The structure (i.e. fan-ins) of the formula Γ_d is exactly the same as that of F_d. However, the subformulas of Γ_d are not variable disjoint. Instead, we use the n_d variables of Γ_d to obtain a family $\mathcal{F}$ of f_d many sets of size n_{d-1} (one for each subformula of depth $d-1$) in a way that ensures that Janson's inequality can be used to analyze the acceptance or rejection probability of Γ_d.

As mentioned above, to apply Janson's inequality, this family $\mathcal{F}$ must be chosen in a way that ensures that most pairs of sets in $\mathcal{F}$ are disjoint. It turns out that we also need other properties of this family to ensure that the multiplicative approximation $(1+\alpha)$ is suitably close to 1. However, we show that standard designs due to Nisan and Wigderson [Nis91, NW94] used in the construction of pseudorandom generators already have these properties (though these properties were not needed in these earlier applications, as far as we know).

With these properties in place, we can analyze the derandomized formula Γ_d. For each subformula Γ of depth $i \leq d$, we can define $p_\Gamma^{(b)}$ analogously to above. Using a careful analysis, we show that $p_\Gamma^{(b)} \in [p_i^{(b)}(1-\alpha_i), p_i^{(b)}(1+\alpha_i)]$ for a suitably small α_i. This allows to infer an analogue of (2) for $p_\Gamma^{(1)}$ and $p_\Gamma^{(0)}$, which in turn can be used to show (as in the case of F_d) that Γ_d solves the δ-coin problem.

Lower bounds. We now describe the ideas behind the proof of Theorem 4. It follows from the result of O'Donnell and Wimmer [OW07] that there is a close connection between the δ-coin problem and computing an approximate majority on n Boolean inputs. In particular, it follows from this connection that if there is an $\mathrm{AC}^0[\oplus]$ formula F of size s and depth d solving the δ-coin problem for $\delta = \Theta(1/\sqrt{n})$ that *additionally computes a monotone function*,[11] then we also have a formula F' of size s and depth d computing an approximate majority on n inputs. (The formula F' is obtained by substituting each input of F with a uniformly random input among the n inputs to the approximate majority.) Since standard lower bounds for $\mathrm{AC}^0[\oplus]$ formulas [Raz89, Smo87, RS17] imply lower bounds for computing approximate majorities, we immediately get a lower bound of $\exp(\Omega(d(1/\delta)^{1/(d-1)}))$ for $\mathrm{AC}^0[\oplus]$ formulas F that solve the δ-coin problem *by computing a monotone function.*

For the general case, the above reduction from approximate majorities to the coin problem no longer works and we have to do something different. Our strategy is to look inside the proof of the $\mathrm{AC}^0[\oplus]$ formula lower bounds and use these ideas to prove the general lower bound for the coin problem. In particular, by the polynomial-approximation method due to Razborov [Raz89] (and a quantitative improvement from [RS17]), it suffices to prove degree lower bounds on polynomials from $\mathbb{F}_2[x_1, \ldots, x_N]$ that solve the δ-coin problem.

We are able to prove the following theorem in this direction, which we believe is independently interesting.

Theorem 6. *Let $g \in \mathbb{F}_2[x_1, \ldots, x_N]$ solve the δ-coin problem. Then,* $\deg(g) = \Omega(1/\delta)$.

Remark 7.

1. *Note that the degree lower bound in Theorem 6 is independent of the sample complexity N of the underlying function g.*

2. *The lower bound obtained is tight up to a constant factor. This can be seen by using the fact that this yields tight lower bounds for the coin problem (which we show), or by directly approximating the Majority function on $1/\delta^2$ bits suitably [BGL06] to obtain a degree $O(1/\delta)$ polynomial that solves the δ-coin problem.*

3. *A weaker degree lower bound of $\Omega(1/(\delta \cdot (\log^2(1/\delta))))$ can be obtained by using an idea of Shaltiel and Viola [SV10], who show how to use any solution to the coin problem and some additional Boolean circuitry to approximate the Majority function on $1/\delta^2$ inputs. Unfortunately, this weaker degree lower bound only implies a formula lower bound that is superpolynomially weaker than the upper bound.*

4. *As mentioned above, an independent recent paper of Chattopadhyay et al. [CHLT18] proves a result on the Fourier spectrum of low-degree polynomials which can be used to recover the*

[11] Note that we are not restricting the formula F itself to be monotone. We only require that it computes a monotone function.

degree bound in Theorem 6 (Avishay Tal (personal communication)). Conversely, Theorem 6 can be used to recover the corresponding result of Chattopadhyay et al. [CHLT18].

The proof of Theorem 6 is inspired by a standard result in circuit complexity that says that any polynomial $P \in \mathbb{F}_2[x_1, \ldots, x_n]$ that computes an approximate majority must have degree $\Omega(\sqrt{n})$. The basic ideas of this proof go back to Smolensky [Smo87],[12] though the result itself was proved in Szegedy's PhD thesis [Sze89] and a later paper of Smolensky [Smo93]. Here, we modify a slightly different "dual" proof of this result which appears in the work of Kopparty and Srinivasan [KS12], which itself builds on ideas of Aspnes, Beigel, Furst and Rudich [ABFR94] and Green [Gre00]. (The proof idea of Smolensky [Smo87] can also be made to work.)

The first idea is to note that the proof in [KS12] can be modified to prove a lower bound of $\Omega(\sqrt{n})$ on the degree of any $P \in \mathbb{F}_2[x_1, \ldots, x_n]$ that satisfies the following condition: there exist constants $a > b$ such that P agrees with the Majority function on n bits on all but an ε fraction of inputs of Hamming weight in $[(n/2) - a\sqrt{n}, (n/2) - b\sqrt{n}] \cup [(n/2) + b\sqrt{n}, (n/2) + a\sqrt{n}]$ (where ε is suitably small depending on a, b).

Using the sampling argument of O'Donnell and Wimmer [OW07] and the above degree lower bound, it follows that if g satisfies the property that it accepts w.h.p. inputs from any product distribution D_α^N for $\alpha \in [(1/2) - a\delta, (1/2) - b\delta]$ and rejects w.h.p. inputs from any product distribution D_β^N for $\beta \in [(1/2) + b\delta, (1/2) + a\delta]$, then the degree of g must be $\Omega(1/\delta)$.

But g might not satisfy this hypothesis. Informally, solving the δ-coin problem only means that the acceptance probability of g is small on inputs from $D^N_{(1-\delta)/2}$ and large on inputs from $D^N_{(1+\delta)/2}$. It is not clear that these probabilities will remain small for α, β in some intervals of length $\Omega(\delta)$. For example, it may be that the acceptance probability of the polynomial g on distribution D_α^N oscillates rapidly for $\alpha \in [(1/2) - a\delta, (1/2) - b\delta]$ even for a, b that are quite close to each other. In this case, however, we observe that g can be used to distinguish $D^{N'}_{\alpha'}$ and $D^{N'}_{\alpha''}$ for α', α'' quite close to each other. In other words, we are solving a 'harder' coin problem (since $|\alpha' - \alpha''|$ is small). Further, we can show that this new distinguisher, say g', has not much larger degree and sample complexity than the old one. We can thus try to prove the degree lower bound for g' instead.

We repeat this argument until we can prove a degree lower bound on the new distinguisher g' (which implies a degree lower bound on g). We can show that since the sample complexities of successive distinguishers are not increasing too quickly, but the coin problems that they solve are getting much harder, this iteration cannot continue for more than finitely many steps. Hence, after finitely many steps, we will be able to obtain a degree lower bound.

2.3 Other related work

The coin problem has also been investigated in other computational models. Brody and Verbin [BV10], who formally defined the coin problem, studied the complexity of this model in read-once branching programs. Their lower bounds were strengthened by Steinberger [Ste13] and Cohen, Ganor and Raz [CGR14]. Lee and Viola [LV18] studied the problem which has also been studied in the model of "product tests." Both these models are incomparable in strength to the constant-depth circuits we study here.

[12] Though Razborov [Raz89] was the first to prove an exponential $\mathrm{AC}^0[\oplus]$ circuit lower bound for the Majority function, he did not explicitly prove a lower bound on the degree of approximating polynomials for the Majority function. Instead, he worked with a different symmetric function for the polynomial question.

3 Preliminaries

Throughout this section, let $d \geq 2$ be a fixed constant and $\delta \in (0,1)$ be a parameter. For any $N \geq 1$, let μ_0^N and μ_1^N denote the product distributions over $\{0,1\}^N$ where each bit is set to 1 with probability $(1-\delta)/2$ and $(1+\delta)/2$ respectively.

3.1 Some technical preliminaries

Throughout, we use $\log(\cdot)$ to denote logarithm to the base 2 and $\ln(\cdot)$ for the natural logarithm. We use $\exp(x)$ to denote e^x.

Fact 8. *Assume that $x \in [-1/2, 1/2]$. Then we have the following chain of inequalities.*

$$\exp(x - (|x|/2)) \underset{\text{(a)}}{\leq} \exp(x - x^2) \underset{\text{(b)}}{\leq} 1 + x \underset{\text{(c)}}{\leq} \exp(x) \underset{\text{(d)}}{\leq} 1 + x + x^2 \underset{\text{(e)}}{\leq} 1 + x + (|x|/2) \tag{3}$$

The following is an easy consequence of the Chernoff bound.

Fact 9 (Error reduction)**.** *Say g solves the coin problem with error $(1/2) - \eta$ for some $\eta > 0$ and let N denote the sample complexity of g. Let $G_t : \{0,1\}^{N \cdot t} \to \{0,1\}$ be defined as follows. On input $x \in \{0,1\}^{N \cdot t}$,*

$$G_t(x) = \mathrm{Maj}_t(g(x_1, \ldots, x_N), g(x_{N+1}, \ldots, x_{2N}), \ldots, g(x_{(t-1)N+1}, \ldots, x_{t \cdot N})).$$

Then, for $t = O(\log(1/\varepsilon)/\eta^2)$, G_t solves the δ-coin problem with error at most ε.

3.2 Boolean formulas

We assume standard definitions regarding Boolean circuits. The size of a circuit always refers to the total number of gates (including input gates) in the circuit.

We abuse notation and use AC^0 *formulas of size s and depth d* (even for superpolynomial s) to denote depth d formulas of size s made up of AND, OR and NOT gates. Similar notation is also used for $\mathrm{AC}^0[\oplus]$ formulas.

Given a Boolean formula F, we use $\mathrm{Vars}(F)$ to denote the set of variables that appear as labels of input gates of F.

We say that a Boolean formula family $\{F_n\}_{n \geq 1}$ is explicit if there is a deterministic polynomial-time algorithm which when given as input n (in binary) and the description of two gates g_1, g_2 of F_n is able to compute whether there is a wire from g_1 to g_2 or not. Such a notion of explicitness has been described as *uniformity* in [Vol99](see Chapter 2 and definition 2.24).

3.3 Amano's formula construction

In this section we present the construction of a depth d AC^0 formula that solves the δ-coin problem. The construction presented here is by Amano [Ama09], which works for $d \geq 3$. For $d = 2$, a construction was presented by O'Donnell and Wimmer [OW07]. We describe their construction in Section 5.

Define $m = \lceil (1/\delta)^{1/(d-1)} \cdot (1/\ln 2) \rceil$. For $i \in [d-2]$, define δ_i inductively by $\delta_1 = m\delta$ and $\delta_i = \delta_{i-1} \cdot (m \ln 2)$.

Define fan-in parameters $f_1 = m$, $f_2 = f_3 = \cdots = f_{d-2} = \lceil m \cdot 2^m \cdot \ln 2 \rceil$, $f_{d-1} = C_1 \cdot m 2^m$ and $f_d = \lceil \exp(C_1 \cdot m) \rceil$, where $C_1 = 50$.

Define the formula F_d to be an alternating formula with AND and OR gates such that

- Each gate at level i above the variables has fan-in f_i.
- The gates at level 1 (just above the variables) are AND gates.
- Each leaf is labelled by a distinct variable.

Note that F_d is a formula on $N = \prod_{i\in[d]} f_i \le \exp(O(dm))$ variables of size $O(N)$.

Amano [Ama09] showed that F_d solves the δ-coin problem. We state a more detailed version of his result below. Since this statement does not exactly match the statement in his paper, we give a proof in the appendix.

For each $i \le d$, let F_i denote any subformula of F_d of depth i. Let N_i denote $|\mathrm{Vars}(F_i)|$ and let $p_i^{(b)}$ denote the probability

$$p_i^{(b)} = \min\{\Pr_{\boldsymbol{x}\sim\mu_b^{N_i}}[F_i(\boldsymbol{x}) = 0], \Pr_{\boldsymbol{x}\sim\mu_b^{N_i}}[F_i(\boldsymbol{x}) = 1]\}. \tag{4}$$

Note that the definition of $p_i^{(b)}$ is independent of the exact subformula F_i chosen: any subformula of depth i yields the same value.

Theorem 10. *Assume $d \ge 3$ and F_d is defined as above. Then, for small enough δ, we have the following.*

1. *For $b, \beta \in \{0,1\}$ and each $i \in [d-1]$ such that $i \equiv \beta \pmod 2$, we have*

$$p_i^{(b)} = \Pr_{\boldsymbol{x}\sim\mu_b^{N_i}}[F_i(\boldsymbol{x}) = \beta].$$

   In particular, for any $i \in \{2,\dots,d-2\}$ and any $b \in \{0,1\}$

$$p_i^{(b)} = (1 - p_{i-1}^{(b)})^{f_i}. \tag{5}$$

2. *For $\beta \in \{0,1\}$ and $i \in [d-2]$ such that $i \equiv \beta \pmod 2$, we have*

$$\frac{1}{2^m}(1 + \delta_i \exp(-3\delta_i)) \le p_i^{(\beta)} \le \frac{1}{2^m}(1 + \delta_i \exp(3\delta_i))$$
$$\frac{1}{2^m}(1 - \delta_i \exp(3\delta_i)) \le p_i^{(1-\beta)} \le \frac{1}{2^m}(1 - \delta_i \exp(-3\delta_i))$$

3. *Say $d - 1 \equiv \beta \pmod 2$. Then*

$$p_{d-1}^{(\beta)} \ge \exp(-C_1 m + C_2) \text{ and } p_{d-1}^{(1-\beta)} \le \exp(-C_1 m - C_2)$$

   where $C_2 = C_1/10$.

4. *For each $b \in \{0,1\}$, $\Pr_{\boldsymbol{x}\sim\mu_b^N}[F_d(\boldsymbol{x}) = 1-b] \le 0.05$. In particular, F_d solves the δ-coin problem.*

Observation 11. *For any $i \in \{2,\dots,d\}$ and $b \in \{0,1\}$, $p_{i-1}^{(b)} \cdot f_i \le 50m$.*

Remark 12. *A similar construction to Amano's formula above was used by Rossman, Servedio and Tan [RST15] to prove an average-case* Depth-hierarchy *theorem for* AC^0 *circuits. Their construction was motivated by the* Sipser functions *used in the work of Sipser [Sip83] and Håstad [Has89] to prove worst-case Depth-hierarchy theorems.*

3.4 Janson's inequality

We state Janson's inequality [Jan90] in the language of Boolean circuits. The standard proof due to Boppana and Spencer (see, e.g. [AS92, Chapter 8]) easily yields this statement. Since Janson's inequality is not normally presented in this language, we include a proof in the appendix for completeness.

Theorem 13 (Janson's inequality)**.** *Let $C_1, \ldots, C_M$ be any monotone Boolean circuits over inputs $x_1, \ldots, x_N$, and let C denote $\bigvee_{i\in[M]} C_i$. For each distinct $i, j \in [M]$, we use $i \sim j$ to denote the fact that $\mathrm{Vars}(C_i) \cap \mathrm{Vars}(C_j) \neq \emptyset$. Assume each $\boldsymbol{x}_j$ ($j \in [n]$) is chosen independently to be 1 with probability $p_i \in [0,1]$, and that under this distribution, we have $\max_{i\in[M]} \Pr_{\boldsymbol{x}}[C_i(\boldsymbol{x}) = 1] \leq 1/2$. Then, we have*

$$\prod_{i\in[M]} \Pr_{\boldsymbol{x}}[C_i(\boldsymbol{x}) = 0] \leq \Pr_{\boldsymbol{x}}[C(\boldsymbol{x}) = 0] \leq \left(\prod_{i\in[M]} \Pr_{\boldsymbol{x}}[C_i(\boldsymbol{x}) = 0]\right) \cdot \exp(2\Delta) \tag{6}$$

where $\Delta := \sum_{i<j:i\sim j} \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 1) \wedge (C_j(\boldsymbol{x}) = 1)]$.

Remark 14. *By using DeMorgan's law, a similar statement also holds for the probability that the conjunction $C' = \bigwedge_{i\in[M]} C_i$ takes the value 1. More precisely, if $\max_{i\in[M]} \Pr_{\boldsymbol{x}}[C_i(\boldsymbol{x}) = 0] \leq 1/2$, we have*

$$\prod_{i\in[M]} \Pr_{\boldsymbol{x}}[C_i(\boldsymbol{x}) = 1] \leq \Pr_{\boldsymbol{x}}[C'(\boldsymbol{x}) = 1] \leq \left(\prod_{i\in[M]} \Pr_{\boldsymbol{x}}[C_i(\boldsymbol{x}) = 1]\right) \cdot \exp(2\Delta) \tag{7}$$

where $\Delta := \sum_{i<j:i\sim j} \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 0) \wedge (C_j(\boldsymbol{x}) = 0)]$.

4 Design construction

In order to define a derandomized version of the formulas in Section 3.3, we will need a suitable notion of a combinatorial design. The following definition of a combinatorial design refines the well-known notion of a Nisan-Wigderson design from the work of [Nis91, NW94]. We give a construction of our combinatorial design by using a construction of Nisan-Wigderson design from [NW94] and showing that this construction in fact satisfies the additional properties we need.

Definition 15 (Combinatorial Designs)**.** *For positive integers N_1, N_2, M, ℓ and $\gamma, \eta \in (0,1)$, an* $(N_1, M, N_2, \ell, \gamma, \eta)$-Combinatorial Design *is a family $\mathcal{F}$ of subsets of $[N_1]$ such that*

1. *$|\mathcal{F}| \geq M$,*
2. *$\mathcal{F} \subseteq \binom{[N_1]}{N_2}$ (i.e. every set in $\mathcal{F}$ has size N_2),*
3. *Given any distinct $S, T \in \mathcal{F}$ we have $|S \cap T| \leq \ell$,*
4. *For any $a \in [N_2]$, we have $|\{S \in \mathcal{F} \mid S \ni a\}| \leq \gamma \cdot M$,*
5. *For any $i \in [\ell]$, we have $|\{\{S, T\} \subseteq \mathcal{F} \mid S \neq T, |S \cap T| = i\}| \leq \eta^i \cdot M^2$.*

The main result of this section is the following.

Lemma 16 (Construction of Combinatorial design)**.** *Given positive integers N_2 and M and real parameters $\gamma, \eta \in (0,1)$ satisfying $N_2 \geq (\log M)/10$, $M \geq 10 \cdot N_2/\eta$, and $\gamma \geq \eta/N_2$, there exist positive integers $\ell = \Theta(\log M/\log(N_2/\eta))$ and $Q = O((N_2/\eta)^{1+1/\ell})$ and an $(N_1 = Q \cdot N_2, M, N_2, \ell, \gamma, \eta)$-combinatorial design.*

Further, the design is explicit in the following sense. Identify $[N_1]$ with $[N_2] \times [Q]$ via the bijection $\rho : [N_1] \to [N_2] \times [Q]$ such that $\rho(i) = (j,k)$ where $i = (k-1)N_2 + j$. Then, each set in $\mathcal{F}$ is of the form $\{(1,k_1), \ldots, (N_2, k_{N_2})\}$ for some $k_1, \ldots, k_{N_2} \in [Q]$. Finally, there is a deterministic algorithm $\mathcal{A}$, which when given as input an $i \in [|\mathcal{F}|]$ and a $j \in [N_2]$, produces $k_j \in [Q]$ in $\mathrm{poly}(\log M)$ time.

Proof. Define ℓ to be the largest integer such that $M^{1/\ell} \geq 10 \cdot N_2/\eta$: note that $\ell \geq 1$ by our assumption that $M \geq 10 \cdot N_2/\eta$. Thus, we have

$$\ell \leq \frac{\log M}{\log(10 \cdot N_2/\eta)} \leq \frac{\log M}{\log \log M} \tag{8}$$

and also

$$M^{1/(\ell+1)} < \frac{10 \cdot N_2}{\eta}. \tag{9}$$

Define the parameter $Q_1 = \lceil M^{1/\ell} \rceil$. We have

$$\frac{10 \cdot N_2}{\eta} \leq M^{1/\ell} \leq Q_1 \leq 2M^{1/\ell} \leq O\left(\left(\frac{N_2}{\eta}\right)^{1+\frac{1}{\ell}}\right) \tag{10}$$

where we used (9) for the last inequality.

Let Q be the smallest power of 2 greater than or equal to Q_1 and let $\mathbb{F}_Q$ be a finite field of size Q. By a result of Shoup [Sho90], we can construct in time $\mathrm{poly}(\log Q) = \mathrm{poly}(\log M)$ time an implicit representation of $\mathbb{F}_Q$ where each element of $\mathbb{F}_Q$ is identified with an element of $\{0,1\}^{\log Q}$ and arithmetic can be performed in time $\mathrm{poly}(\log Q)$. Fix such a representation of $\mathbb{F}_Q$.

Let $A \subseteq \mathbb{F}_Q$ be any subset of size N_2 (note that by (10) we have $N_2 \leq Q_1$ which is at most Q) and let $B \subseteq \mathbb{F}_Q$ be any fixed subset of size Q_1. Let $A_1 \subseteq A$ be a set of size ℓ (note that by (8) $\ell \leq (\log M)/10$ which is at most N_2 by assumption).

Fix $N_1 = Q \cdot N_2$ and identify $[N_1]$ with the set $A \times \mathbb{F}_Q$ in an arbitrary way. Assume that $A = \{a_1, \ldots, a_{N_2}\}$ and $A_1 = \{a_1, \ldots, a_\ell\}$. We define $\mathcal{P}$ to be the set of all polynomials $P \in \mathbb{F}_Q[x]$ of degree at most $\ell - 1$ such that $P(a) \in B$ for each $a \in A_1$.

We are now ready to define the family $\mathcal{F}$. For each $\mathbf{b} = (b_1, \ldots, b_\ell) \in A^\ell$, we define the polynomial $P_{\mathbf{b}}(x)$ to be the unique polynomial in $\mathcal{P}$ such that $P(a_i) = b_i$ for each $i \in [\ell]$ (note that P is uniquely defined since any polynomial of degree at most $\ell - 1$ can be specified by its evaluations at any ℓ distinct points). We add the set $S_{\mathbf{b}} \subseteq A \times \mathbb{F}_Q$ to $\mathcal{F}$, where $S_{\mathbf{b}}$ is defined by

$$S_{\mathbf{b}} = \{(a_i, P_{\mathbf{b}}(a_i)) \mid i \in [N_2]\}. \tag{11}$$

In words, $S_{\mathbf{b}}$ is the graph of the polynomial $P_{\mathbf{b}}$ restricted to the domain A.

We now show that $\mathcal{F}$ is indeed a $(N_1, M, N_2, \ell, \gamma, \eta)$-combinatorial design.

1. For distinct $\mathbf{b}, \mathbf{b}' \in B^\ell$, the sets $|S_{\mathbf{b}} \cap S_{\mathbf{b}'}| \leq \ell$ since the graphs of the distinct polynomials $P_{\mathbf{b}}$ and $P_{\mathbf{b}'}$ can intersect at at most $\ell - 1$ points. In particular, we have $|\mathcal{F}| = |B|^\ell = Q_1^\ell \geq M$ (by (10)). Further, we also have that any pair of distinct sets in $\mathcal{F}$ have an intersection of at most ℓ. This proves properties 1 and 3 in Definition 15 above.

2. Each set in $\mathcal{F}$ has size N_2, since it is of the form $\{(a_i, b_i) \mid i \in [N_2]\}$ for some choice of $b_1, \ldots, b_{N_2} \in \mathbb{F}_Q$. This proves property 2.

3. We now consider property 4. Fix any $(a, b) \in A \times \mathbb{F}_Q$. If $(a, b) \in S \in \mathcal{F}$, then S is the graph of a polynomial $P \in \mathcal{P}$ such that $P(a) = b$. To uniquely specify such a polynomial, it suffices to provide its evaluations at any $\ell - 1$ other points. We choose the evaluation points to be a fixed set $A_1' \subseteq A_1 \setminus \{a\}$ of size $\ell - 1$. Since $P(a') \in B$ for each $a' \in A_1'$, there are at most $|B|^{\ell-1} = Q_1^{\ell-1}$ many choices for these evaluations, which yields the same bound for the number of sets $S \in \mathcal{F}$ such that $(a, b) \in S$.

   Hence, we have

$$|\{S \in \mathcal{F} \mid S \ni (a, b)\}| \leq \frac{Q_1^\ell}{Q_1} = \frac{\left(\lceil M^{1/\ell} \rceil\right)^\ell}{Q_1} \leq \frac{\eta}{N_2} \cdot M \leq \gamma \cdot M$$

   where the final inequality follows from our assumption that $\gamma \geq \eta/N_2$, and the second last inequality uses the fact that $Q_1 \geq 10 \cdot N_2/\eta$ and

$$\left(\lceil M^{1/\ell} \rceil\right)^\ell \leq \left(M^{1/\ell} + 1\right)^\ell = M \cdot \left(1 + \frac{1}{M^{1/\ell}}\right)^\ell \leq M \cdot \left(1 + \frac{1}{\ell}\right)^\ell \leq 3M \tag{12}$$

   (using $\ell^\ell \leq M$, which follows from (8), for the second-last inequality).

4. For property 5, we use a similar argument to property 4. Fix distinct sets $S, T \in \mathcal{F}$ such that $|S \cap T| = i$. The sets S and T are graphs of distinct polynomials $P_1, P_2 \in \mathcal{P}$ respectively that agree in i places. We bound the number of such pairs of polynomials.

   The number of choices for S, and hence P_1, is exactly $|\mathcal{F}| = Q_1^\ell$. Given P_1, we can specify P_2 as follows.

   - Specify a set $A' \subseteq A$ of size i such that P_1 and P_2 agree on A'. This gives the evaluation of P_2 at i points. Further, the number of such A' is $\binom{N_2}{i} \leq N_2^i$.
   - Specify the evaluation of P_2 at the first $\ell - i$ points from $A_1 \setminus A'$. This gives the evaluation of P_2 at $\ell - i$ points outside A' and hence specifies P_2 exactly. The number of possible evaluations is $|B|^{\ell-i} = Q_1^{\ell-i}$.

   Hence, the number of pairs of polynomials (P_1, P_2) whose graphs agree at i points is at most

$$Q_1^\ell \cdot N_2^i \cdot Q_1^{\ell-i} = Q_1^{2\ell} \cdot \left(\frac{N_2}{Q_1}\right)^i = \left(\lceil M^{1/\ell} \rceil\right)^{2\ell} \cdot \left(\frac{N_2}{Q_1}\right)^i \leq 9M^2 \cdot \left(\frac{N_2}{Q_1}\right)^i \leq 9M^2 \cdot \left(\frac{\eta}{10}\right)^i \leq \eta^i \cdot M^2$$

   where for the first inequality we have used (12) and the second inequality follows from the fact that $Q_1 \geq 10 \cdot N_2/\eta$.

We have thus shown that $\mathcal{F}$ is indeed a $(N_1, M, N_2, \ell, \gamma, \eta)$-combinatorial design as required. The explicitness of the design follows easily from its definition. □

5 Proof of Theorem 3 for $d = 2$

In this section we will present the proof of Theorem 3 for the special case when $d = 2$. The proof is quite similar to the case for general d, but is somewhat simpler (as the construction of the AC^0 formulas is simpler) and illustrates many of the ideas of the general proof.

Throughout this section, let δ be a parameter going to 0.

We start by stating a result of O'Donnell and Wimmer [OW07], who gave a depth-2 AC^0 formula for solving the δ-coin problem. Formally, they defined a depth-2 circuit as follows.

Let $C_0 \geq 10$. Let $m = 1/\delta \cdot C_0$, $f_1 = m$, $f_2 = 2^m$, where $C = 2^{C_0}$. The formula F_2 is defined as follows:

- At layer 1 we have AND gates and the fan-in of each AND gate is f_1.
- At layer 2 we have a single OR gate with fan-in f_2.
- Each leaf is labelled with distinct variables.

For F_2 defined as above [OW07] proved the following theorem.

Theorem 17 ([OW07])**.** *Let $N = f_1 \cdot f_2$. For each $b \in \{0, 1\}$*

$$\Pr_{\boldsymbol{x} \sim \mu_b^N}[F_2(\boldsymbol{x}) = 1 - b] \leq 0.05,$$

i.e. specifically F_2 solves the δ-coin problem.

Here, the number of inputs is N and the size of F_2 is also $O(N)$. We now give a construction of an explicit depth 2 formula of the same size as in the theorem above which solves the δ-coin problem, but using far fewer inputs. We achieve this by an application of the Janson's inequality coupled with our combinatorial design.

We now describe the construction of such a depth 2 formula Γ_2. Fix m, f_1, f_2 as above. Define parameters $\gamma = 1$, $\eta = 1/(16 \cdot (\frac{1+\delta}{2})^m \cdot f_2) = 1/(16 \cdot (1+\delta)^m)$. Let $\mathcal{F}$ be an $(n, f_2, f_1, \ell, \gamma, \eta)$-design obtained using Lemma 16. We are now ready to define Γ_2.

- Let $S_1, S_2, \ldots, S_{f_2} \in \binom{[n]}{f_1}$ be the first f_2 sets in the $(n, f_2, f_1, \ell, \gamma, \eta)$-design $\mathcal{F}$. At layer 1 we have f_2 many AND gates, say $\Gamma_1^1, \ldots, \Gamma_1^{f_2}$, with fan-in f_1 each. For each $i \in [f_2]$, the inputs of the gate Γ_1^i are the variables indexed by the set S_i.
- At layer 2 we have a single gate, which is an OR of $\Gamma_1^1, \ldots, \Gamma_1^{f_2}$.

With this definition of Γ_2, we now prove Theorem 3. From the definition of the parameters, it can be checked that $\eta = \Theta(1)$. Therefore, we get $\ell = \Theta(m/\log m)$ and $Q = O(f_1/\eta)^{1+1/\ell} = O((1/\delta))$. Therefore, the number of inputs in the formula is $N = O(Q \cdot f_1) = O(1/\delta^2)$ and the size of the formula is $O(f_1 \cdot f_2) = \exp(O(1/\delta))$.

The only thing we need to prove now is that for any $b \in \{0, 1\}$, $\Pr_{\boldsymbol{x} \sim \mu_b^n}[\Gamma_2(\boldsymbol{x}) = 1 - b] \leq 0.1$. Let $q^{(0)} = (1 - \delta)/2$ and $q^{(1)} = (1 + \delta)/2$. Let $p_1^{(0)} = \left(\frac{1-\delta}{2}\right)^m$ and $p_1^{(1)} = \left(\frac{1+\delta}{2}\right)^m$. Note that $p_1^{(b)}$ ($b \in \{0, 1\}$) is the probability that each subformula Γ_1^i accepts on a random input $\boldsymbol{x}$ chosen from the distribution μ_b^n.

Let $b = 0$. In this case

$$\begin{aligned}\Pr_{\boldsymbol{x}\sim\mu_0^n}[\Gamma_2(\boldsymbol{x}) = 1] &= \Pr_{\boldsymbol{x}\sim\mu_0^n}[\exists i \in [f_2] : \Gamma_1^i(\boldsymbol{x}) = 1] \\ &\leq f_2 \cdot p_1^{(0)} = (1-\delta)^m \leq \exp(-C_0) \leq 0.1.\end{aligned}$$

Here the first inequality is due to a union bound. The other inequalities are obtained by simple substitutions of the parameters and using (3).

Now consider the $b = 1$ case. Here $\Pr_{\boldsymbol{x}\sim\mu_1^n}[\Gamma_2(\boldsymbol{x}) = 0] = \Pr_{\boldsymbol{x}\sim\mu_1^n}[\forall i \in [f_2] : \Gamma_1^i(\boldsymbol{x}) = 0]$. Now we would like to bound this using Janson's inequality (Theorem 13). Applying Janson's inequality, we get

$$\begin{aligned}\Pr_{\boldsymbol{x}\sim\mu_1^n}[\Gamma_2(\boldsymbol{x}) = 0] &= \Pr_{\boldsymbol{x}\sim\mu_1^n}[\forall i \in [f_2] : \Gamma_1^i(\boldsymbol{x}) = 0] \\ &\leq \prod_{i\in[f_2]} \Pr_{\boldsymbol{x}\sim\mu_1^n}[\Gamma_1^i(\boldsymbol{x}) = 0] \cdot \exp(2\Delta) \\ &\leq (1-p_1^{(1)})^{f_2} \cdot \exp(2\Delta) \\ &\leq \exp(-p_1^{(1)} f_2 + 2\Delta), \end{aligned} \tag{13}$$

where

$$\Delta = \sum_{\substack{j<k: \\ \mathrm{Vars}(\Gamma_1^j)\cap\mathrm{Vars}(\Gamma_1^k)\neq\emptyset}} \Pr_{\boldsymbol{x}\sim\mu_1^n}[(\Gamma_1^j(\boldsymbol{x}) = 1) \wedge (\Gamma_1^k(\boldsymbol{x}) = 1)].$$

We will now obtain a bound on Δ.

$$\begin{aligned}\Delta &= \sum_{\substack{j<k: \\ \mathrm{Vars}(\Gamma_1^j)\cap\mathrm{Vars}(\Gamma_1^k)\neq\emptyset}} \Pr_{\boldsymbol{x}\sim\mu_1^n}[(\Gamma_1^j(\boldsymbol{x}) = 1) \wedge (\Gamma_1^k(\boldsymbol{x}) = 1)]. \\ &= \sum_{r=1}^{\ell} \sum_{\substack{j<k: \\ |\mathrm{Vars}(\Gamma_1^j)\cap\mathrm{Vars}(\Gamma_1^k)|=r}} \Pr_{\boldsymbol{x}\sim\mu_1^n}[(\Gamma_1^j(\boldsymbol{x}) = 1) \wedge (\Gamma_1^k(\boldsymbol{x}) = 1)]\end{aligned}$$

As Γ_1^j and Γ_1^k are both ANDs of size m, $\Gamma_1^j \wedge \Gamma_1^k$ is an AND of size $(2m - |\mathrm{Vars}(\Gamma_1^j) \cap \mathrm{Vars}(\Gamma_1^k)|)$. Therefore, we get

$$\begin{aligned}
\Delta &= \sum_{r=1}^{\ell} \sum_{\substack{j<k: \\ |\mathrm{Vars}(\Gamma_1^j)\cap\mathrm{Vars}(\Gamma_1^k)|=r}} \Pr_{\boldsymbol{x}\sim\mu_1^n}[((\Gamma_1^j \wedge \Gamma_1^k)(\boldsymbol{x}) = 1)] \\
&= \sum_{r=1}^{\ell} \sum_{\substack{j<k: \\ |\mathrm{Vars}(\Gamma_1^j)\cap\mathrm{Vars}(\Gamma_1^k)|=r}} \left(\frac{1+\delta}{2}\right)^{2m-r} \\
&= \sum_{r=1}^{\ell} \sum_{\substack{j<k: \\ |\mathrm{Vars}(\Gamma_1^j)\cap\mathrm{Vars}(\Gamma_1^k)|=r}} \frac{(p_1^{(1)})^2}{((1+\delta)/2)^r} \\
&= \sum_{r=1}^{\ell} \frac{(p_1^{(1)})^2}{(q^{(1)})^r} \cdot |\{(j,k) \mid j<k \text{ and } |\mathrm{Vars}(\Gamma_1^j)\cap\mathrm{Vars}(\Gamma_1^k)| = r\}|
\end{aligned}$$

From the construction of the formula and the combinatorial design $\mathcal{F}$, we know that $|\{(j,k) \mid j<k \text{ and } |\mathrm{Vars}(\Gamma_1^j)\cap\mathrm{Vars}(\Gamma_1^k)| = r\}| \le \eta^r f_2^2$. We can also bound $1/q^{(1)}$ by a small constant, say 3.

Therefore, we can simplify the above equation as follows:

$$\begin{aligned}
\Delta &\le \sum_{r=1}^{\ell} (p_1^{(1)})^2 \cdot 3^r \cdot \eta^r f_2^2 \\
&= (p_1^{(1)})^2 \cdot f_2^2 \sum_{r=1}^{\ell} 3^r \cdot \eta^r \\
&\le (p_1^{(1)})^2 \cdot f_2^2 \cdot 4 \cdot \eta \qquad (14)
\end{aligned}$$

using the fact that $3\eta \le 1/4$ as $\eta \le 1/16$.

Now, by using our setting of $\eta = 1/(16 \cdot p_1^{(1)} \cdot f_2)$ in (14), we get $\Delta \le p_1^{(1)} f_2/4$. Using this value of Δ in (13), we get $\Pr_{\boldsymbol{x}\sim\mu_1^n}[\Gamma_2(\boldsymbol{x}) = 0] \le \exp(-\frac{p_1^{(1)}\cdot f_2}{2}) \le 0.1$, by our choice of parameters. This completes the proof of Theorem 3 for $d = 2$.

6 Proof of Theorem 3 for $d \ge 3$

Throughout this section, fix a constant depth $d \ge 3$ and a parameter $\delta \in (0,1)$. The parameter δ is assumed to be asymptotically converging to 0.

We also assume the notation from Section 3.3.

6.1 Definition of the formula Γ_d

The formula Γ_d is an alternating monotone depth-d formula made up of AND and OR gates. The structure of the formula and the labels of the gates are the same as in the formula F_d defined in Section 3.3. However, the leaves are labelled with only poly(m) distinct variables.

We now proceed to the formal definition. We iteratively define a sequence of formulas $\Gamma_1, \ldots, \Gamma_d$ (where Γ_i has depth i) as follows. Define the parameters γ and η by

$$\gamma = \frac{1}{m^3} \text{ and } \eta = \frac{1}{m^{10d}}. \tag{15}$$

- Γ_1 is just an AND of $n_1 = m$ distinct variables.
- Recall that for $i \geq 2$, any gate at level i in the formula F_d has fan-in f_i for $f_i = \exp(\Theta(m))$. For each $i \in \{2, \ldots, d\}$, define n_i so that by Lemma 16, we have an explicit $(n_i, f_i, n_{i-1}, \ell, \gamma, \eta)$-combinatorial design $\mathcal{F}_i$ where $\ell = \Theta(\log f_i / \log(n_{i-1}/\eta))$. Note that $n_i = O((n_{i-1})^{2+1/\ell}/\eta^{1+1/\ell}) \leq n_{i-1}^3/\eta^2$.

  The formula Γ_i is defined on a set X of n_i variables by taking the OR/AND (depending on whether i is even or odd respectively) of f_i copies of Γ_{i-1}, each defined on a distinct subset $Y \subseteq X$ of n_{i-1} variables obtained from the combinatorial design $\mathcal{F}_i$.

  Formally, let $S_1, \ldots, S_{f_i} \in \binom{[n_i]}{n_{i-1}}$ be the first f_i many sets in the design $\mathcal{F}_i$ (in lexicographic order, say). Identifying $[n_i]$ with the variable set X of Γ_i, we obtain corresponding subsets $Y_1, \ldots, Y_{f_i}$ of X. The formula Γ_i is an OR/AND of f_i many subformulas $\Gamma_i^1, \ldots, \Gamma_i^{f_i}$ where the jth subformula Γ_i^j is a copy of Γ_{i-1} with variable set Y_j.

Observation 18. *The size of Γ_d is $\exp(O(dm))$. The number of variables appearing in Γ_d is $n_d = m^{2^{O(d)}}$.*

Explicitness of the formula Γ_d. The structure of the formula is determined completely by the parameter δ. Thus to argue that the formula Γ_d is explicit, it suffices to show that the labels of the input gates can be computed efficiently. Note that the inputs are in 1-1 correspondence with the set $[f_d] \times [f_{d-1}] \times \cdots [f_2] \times [f_1]$.

Let Γ_i be any subformula of Γ_d of depth i. If $i = 1$, then Γ_i is simply an AND of $m = f_1$ variables and we identify its variable set with $[f_1]$. When $i > 1$, by the properties of the design constructed in Lemma 16, we see that the set $\mathrm{Vars}(\Gamma_i)$ is in a natural 1-1 correspondence with the set $\mathrm{Vars}(\Gamma_{i-1}) \times [Q_i]$ where Γ_{i-1} is any subformula of depth $i-1$ and $Q_i = n_i/n_{i-1}$. Each subformula Γ_i^j ($j \in [f_i]$) of depth $i-1$ in Γ_i has as its variable set a set of the form $\{(x, k_x) \mid x \in \mathrm{Vars}(\Gamma_{i-1}), k_x \in [Q_i]\}$.

Further, by the explicitness properties of the design constructed in Lemma 16, we see that given any $x \in \mathrm{Vars}(\Gamma_{i-1})$ and $j \in [f_i]$, we can find in $\mathrm{poly}(\log(f_i)) \leq \mathrm{poly}(m)$ time the variable $(x, k) \in \mathrm{Vars}(\Gamma_i)$ that belongs to $\mathrm{Vars}(\Gamma_i^j)$. Equivalently, given a leaf $\ell = (j_i, \ldots, j_1) \in [f_i] \times \cdots \times [f_1]$ of Γ_i and the variable $x \in \mathrm{Vars}(\Gamma_{i-1})$ corresponding to the leaf $(j_{i-1}, \ldots, j_1)$ in Γ_{i-1}, we can find the variable labelling ℓ in $\mathrm{poly}(m)$ time. Using this algorithm and a recursive procedure to find the variable x, we see that the variable labelling the leaf ℓ can be found in $\mathrm{poly}(m)$ time. In particular, given a leaf of Γ_d, the variable labelling it can be found in $\mathrm{poly}(m)$ time.

Thus, the formula Γ_d is explicit.

6.2 Analysis of Γ_d

In this section, we will show that Γ_d distinguishes between the distributions $\mu_0^{n_d}$ and $\mu_1^{n_d}$ as defined in Definition 2. For brevity, we use n to denote n_d.

Fix any subformula Γ of Γ_d and $b \in \{0,1\}$. Assume Γ has depth $i \in [d]$ and $\beta \in \{0,1\}$ is such that $i \equiv \beta \pmod 2$. We define $p_\Gamma^{(b)} = \Pr_{\boldsymbol{x}\sim\mu_b^n}[\Gamma(\boldsymbol{x}) = \beta]$. Assume that Γ is an OR/AND of depth-$(i-1)$ subformulas $\Gamma^1, \ldots, \Gamma^f$. We define

$$\Delta_\Gamma^{(b)} = \sum_{\substack{j<k: \\ \mathrm{Vars}(\Gamma^j)\cap \\ \mathrm{Vars}(\Gamma^k)\neq\emptyset}} \Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Gamma^j(\boldsymbol{x}) = 1-\beta) \wedge (\Gamma^k(\boldsymbol{x}) = 1-\beta)]. \tag{16}$$

The following lemma is the main technical lemma of this section. Along with Theorem 10, it easily implies Theorem 3 (as we show below).

Lemma 19. *Let Γ_d be as constructed above. Then for each $i \in \{2,\ldots,d\}$, each $b \in \{0,1\}$, and any subformula Γ of depth i, we have the following.*

1. $p_\Gamma^{(b)} \in [p_i^{(b)}(1-\eta\cdot(C_3 m)^i), p_i^{(b)}(1+\eta\cdot(C_3 m)^i)]$ *where* $C_3 = 1000$.
2. $\Delta_\Gamma^{(b)} \le (C_4 m)^2 \cdot \eta$ *where* $C_4 = 100$.

Assuming the above lemma, we first prove Theorem 3.

Proof of Theorem 3. We use the explicit formula Γ_d described above. By Lemma 19 applied in the case that $i = d$, it follows that for each $b \in \{0,1\}$

$$|\Pr_{\boldsymbol{x}\sim\mu_b^n}[\Gamma_d(\boldsymbol{x}) = 1-b] - \Pr_{\boldsymbol{x}\sim\mu_b^n}[F_d(\boldsymbol{x}) = 1-b]| = |p_{\Gamma_d}^{(b)} - p_d^{(b)}| \le p_d^{(b)} \cdot \eta(C_3 m)^d = o(1).$$

In particular, using Theorem 10, it follows that $\Pr_{\boldsymbol{x}\sim\mu_b^n}[\Gamma_d(\boldsymbol{x}) = 1-b] \le 0.1$ and hence Γ_d solves the δ-coin problem. The sample complexity of Γ_d is $m^{2^{O(d)}} = (1/\delta)^{2^{O(d)}}$ by construction. □

Proof of Lemma 19. We prove the lemma by induction on i. The base case is when $i = 2$. This proof is quite similar to the proof of the $d = 2$ case from Section 5.

Base case, i.e. $i = 2$: Recall that for $i = 2$, Γ is an OR of f_2-many subformulas $\Gamma^1, \Gamma^2, \ldots, \Gamma^{f_2}$, where each Γ^j is an AND of distinct set of variables. Therefore, we have that $p_{\Gamma^j}^{(b)}$ is the same as in the case of Amano's proof, i.e. $p_{\Gamma^j}^{(b)} = p_1^{(b)}$. Recall that $p_1^{(b)}$ is equal to $(\frac{1-\delta}{2})^m$ if $b = 0$ and it is equal to $(\frac{1+\delta}{2})^m$ if $b = 1$. Let $q^{(0)}$ ($q^{(1)}$) denote $\frac{1-\delta}{2}$ (respectively, $\frac{1+\delta}{2}$).

$$\begin{aligned}\Delta_\Gamma^{(b)} &= \sum_{\substack{j<k: \\ \mathrm{Vars}(\Gamma^j)\cap\mathrm{Vars}(\Gamma^k)\neq\emptyset}} \Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Gamma^j(\boldsymbol{x}) = 1) \wedge (\Gamma^k(\boldsymbol{x}) = 1)]. \\ &= \sum_{r=1}^{\ell} \sum_{\substack{j<k: \\ |\mathrm{Vars}(\Gamma^j)\cap\mathrm{Vars}(\Gamma^k)|=r}} \Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Gamma^j(\boldsymbol{x}) = 1) \wedge (\Gamma^k(\boldsymbol{x}) = 1)]\end{aligned}$$

As Γ^j and Γ^k are both ANDs of size m, $\Gamma^j \wedge \Gamma^k$ is an AND of size $(2m - |\mathrm{Vars}(\Gamma^j) \cap \mathrm{Vars}(\Gamma^k)|)$. Therefore, we get

$$\begin{aligned}
\Delta_\Gamma^{(b)} &= \sum_{r=1}^{\ell} \sum_{\substack{j<k: \\ |\mathrm{Vars}(\Gamma^j)\cap\mathrm{Vars}(\Gamma^k)|=r}} \Pr_{\boldsymbol{x}\sim\mu_b^n}[((\Gamma^j \wedge \Gamma^k)(\boldsymbol{x}) = 1)] \\
&= \sum_{r=1}^{\ell} \sum_{\substack{j<k: \\ |\mathrm{Vars}(\Gamma^j)\cap\mathrm{Vars}(\Gamma^k)|=r}} (q^{(b)})^{2m-r} \\
&= \sum_{r=1}^{\ell} \sum_{\substack{j<k: \\ |\mathrm{Vars}(\Gamma^j)\cap\mathrm{Vars}(\Gamma^k)|=r}} \frac{(p_1^{(b)})^2}{(q^{(b)})^r} \\
&= \sum_{r=1}^{\ell} \frac{(p_1^{(b)})^2}{(q^{(b)})^r} \cdot |\{(j,k) \mid j<k \text{ and } |\mathrm{Vars}(\Gamma^j)\cap\mathrm{Vars}(\Gamma^k)| = r\}|
\end{aligned}$$

From the construction of the formula, we know that $|\{(j,k) \mid j<k \text{ and } |\mathrm{Vars}(\Gamma^j)\cap\mathrm{Vars}(\Gamma^k)| = r\}| \leq \eta^r f_2^2$. We can also bound $1/q^{(b)}$ by a small constant, say 3.

Therefore, we can simplify the above equation as follows:

$$\begin{aligned}
\Delta_\Gamma^{(b)} &\leq \sum_{r=1}^{\ell} (p_1^{(b)})^2 \cdot 3^r \cdot \eta^r f_2^2 \\
&= (p_1^{(b)})^2 \cdot f_2^2 \sum_{r=1}^{\ell} 3^r \cdot \eta^r \\
&\leq (p_1^{(b)})^2 \cdot f_2^2 \cdot 4 \cdot \eta
\end{aligned}$$

The last inequality comes from summing up a geometric series. Now using Observation 11 we get that $p_1^{(b)} \cdot f_2 \leq 50m$. Hence, we get $\Delta_\Gamma^{(b)} \leq (p_1^{(b)})^2 \cdot f_2^2 \cdot 4 \cdot \eta \leq (50m)^2 \cdot 4\eta = (100m)^2 \cdot \eta$. This proves the bound on $\Delta_\Gamma^{(b)}$ in the base case.

We now prove the bounds claimed for $p_\Gamma^{(b)}$ in the base case. When $i = 2$, $\beta = 0$, hence $p_\Gamma^{(b)} = \Pr_{\boldsymbol{x}\sim\mu_b^n}[\Gamma(\boldsymbol{x}) = 0]$. By Janson's inequality (Theorem 13), we get the following bounds on the value of $p_\Gamma^{(b)}$.

$$\prod_{j=1}^{f_2}(1 - p_{\Gamma^j}^{(b)}) \leq p_\Gamma^{(b)} \leq \prod_{j=1}^{f_2}(1 - p_{\Gamma^j}^{(b)}) \cdot \exp(2 \cdot \Delta_\Gamma^{(b)}).$$

Recall that $p_{\Gamma^j}^{(b)} = p_1^{(b)}$ as we are in the base case. Also, from Equation (5) we have that

$(1-p_1^{(b)})^{f_2} = p_2^{(b)}$. Therefore, we get

$$\begin{aligned} p_2^{(b)} \le p_\Gamma^{(b)} &\le p_2^{(b)} \cdot \exp(2\Delta_\Gamma^{(b)}) \\ &\le p_2^{(b)} \cdot (1 + 4 \cdot \Delta_\Gamma^{(b)}) && \text{Using (3) (d)} \\ &\le p_2^{(b)} \cdot (1 + 4 \cdot (C_4 m)^2 \cdot \eta) \\ &\le p_2^{(b)} \cdot (1 + (C_3 m)^2 \cdot \eta) \end{aligned}$$

This finishes the proof of the base case.

Inductive case, i.e. $i \ge 3$**:** We now proceed to proving the inductive case. Assume that the statement holds for $(i-1)$. Let Γ be a subformula at depth i which is OR/AND of subformulas $\Gamma^1, \Gamma^2, \ldots, \Gamma^{f_i}$ each of depth $(i-1)$. From the definition of $\Delta_\Gamma^{(b)}$, we get the following:

$$\begin{aligned} \Delta_\Gamma^{(b)} &= \sum_{\substack{j<k: \\ \mathrm{Vars}(\Gamma^j)\cap\mathrm{Vars}(\Gamma^k)\neq\emptyset}} \Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Gamma^j(\boldsymbol{x}) = 1-\beta) \wedge (\Gamma^k(\boldsymbol{x}) = 1-\beta)]. \\ &= \sum_{r=1}^{\ell} \sum_{\substack{j<k: \\ |\mathrm{Vars}(\Gamma^j)\cap\mathrm{Vars}(\Gamma^k)|=r}} \Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Gamma^j(\boldsymbol{x}) = 1-\beta) \wedge (\Gamma^k(\boldsymbol{x}) = 1-\beta)] \end{aligned}$$

Let t_r denote the maximum value of $\Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Gamma^j(\boldsymbol{x}) = 1-\beta) \wedge (\Gamma^k(\boldsymbol{x}) = 1-\beta)]$, where the maximum is taken over $j<k$ such that $|\mathrm{Vars}(\Gamma^j) \cap \mathrm{Vars}(\Gamma^k)| = r$. Then we get

$$\begin{aligned} \Delta_\Gamma^{(b)} &\le \sum_{r=1}^{\ell} t_r \cdot |\{(j,k) \mid j<k \text{ and } |\mathrm{Vars}(\Gamma^j) \cap \mathrm{Vars}(\Gamma^k)| = r\}| \\ &\le \sum_{r=1}^{\ell} t_r \cdot \eta^r \cdot f_i^2 \end{aligned} \tag{17}$$

Let us now bound t_r, which we will do by using the construction parameters and the inductive hypothesis. Fix any $j<k$. We have

$$\Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Gamma^j(\boldsymbol{x}) = 1-\beta)\wedge(\Gamma^k(\boldsymbol{x}) = 1-\beta)] = \Pr_{\boldsymbol{x}\sim\mu_b^n}[\Gamma^j(\boldsymbol{x}) = 1-\beta] \cdot \Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Gamma^k(\boldsymbol{x}) = 1-\beta) | (\Gamma^j(\boldsymbol{x}) = 1-\beta)]. \tag{18}$$

As Γ^j is a formula of depth $i-1$ and $i-1 \equiv (1-\beta) \pmod 2$, using the induction hypothesis, we can upper bound the quantity $\Pr_{\boldsymbol{x}\sim\mu_b^n}[\Gamma^j(\boldsymbol{x}) = 1-\beta]$. We get

$$\Pr_{\boldsymbol{x}\sim\mu_b^n}[\Gamma^j(\boldsymbol{x}) = 1-\beta] = p_{\Gamma^j}^{(b)} \le p_{i-1}^{(b)} \cdot (1 + \eta \cdot (C_3 m)^{i-1}) = p_{i-1}^{(b)}(1 + o(1)). \tag{19}$$

We now analyse the second term on the right hand side of Equation (18). From the construction of the formula, we know that for any $y \in \mathrm{Vars}(\Gamma^j)$, the variable y appears in at most $\gamma \cdot f_{i-1}$ many

depth-$(i-2)$ subformulas of Γ^k. Since $|\mathrm{Vars}(\Gamma^j) \cap \mathrm{Vars}(\Gamma^k)| = r$, the number of depth-$(i-2)$ subformulas T of Γ^k that contain some variable from Γ^j is at most $\gamma \cdot f_{i-1} \cdot r$ which is at most $\gamma \cdot f_{i-1} \cdot \ell$, as $r \leq \ell$.

Let us construct a formula Φ^k from Γ^k by deleting all the depth-$(i-2)$ subformulas containing some variable from Γ^j. Then we get

$$\begin{aligned}\Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Gamma^k(\boldsymbol{x}) = 1-\beta)|(\Gamma^j(\boldsymbol{x}) = 1-\beta)] &\leq \Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Phi^k(\boldsymbol{x}) = 1-\beta)|(\Gamma^j(\boldsymbol{x}) = 1-\beta)] \\ &= \Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Phi^k(\boldsymbol{x}) = 1-\beta)] \qquad (20)\end{aligned}$$

The first inequality follows from the fact that Φ^k was constructed by removing some subformulas of depth-$(i-2)$ from Γ^k, and this can only increase the probability of taking value $1-\beta$. The equality follows from the fact that Φ^k and Γ^j share no variables in common and hence the events $(\Phi^k(\boldsymbol{x}) = 1-\beta)$ and $(\Gamma^j(\boldsymbol{x}) = 1-\beta)$ are independent.

Let $\Gamma^{k,1}, \Gamma^{k,2}, \ldots, \Gamma^{k,f_{i-1}}$ be the depth-$(i-2)$ subformulas of Γ^k. By ordering the variables if necessary, let $\Gamma^{k,1}, \Gamma^{k,2}, \ldots, \Gamma^{k,f_{i-1}-T}$ be the depth-$(i-2)$ subformulas of Φ^k.

We will show below that

$$\Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Phi^k(\boldsymbol{x}) = 1-\beta)] \leq \Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Gamma^k(\boldsymbol{x}) = 1-\beta)] \cdot (1+o(1)). \qquad (21)$$

Suppose we have this then we will proceed as follows.

$$\Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Phi^k(\boldsymbol{x}) = 1-\beta)] \leq \Pr_{\boldsymbol{x}\sim\mu_b^n}[\Gamma^k(\boldsymbol{x}) = 1-\beta] \cdot (1+o(1)) \leq p_{i-1}^{(b)} \cdot (1+o(1)) \qquad (22)$$

Here the last inequality is obtained by using the induction hypothesis for Γ^k. Now using (19), (20), and (22) in (18) we get

$$\begin{aligned}\Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Gamma^j(\boldsymbol{x}) = 1-\beta) \wedge (\Gamma^k(\boldsymbol{x}) = 1-\beta)] &\leq (p_{i-1}^{(b)}(1+o(1))) \cdot (p_{i-1}^{(b)}(1+o(1))) \\ &\leq (p_{i-1}^{(b)})^2(1+o(1))\end{aligned}$$

Since the above holds for all $j < k$ such that $|\mathrm{Vars}(\Gamma^j) \cap \mathrm{Vars}(\Gamma^k)| = r$, this gives us a bound on t_r. Using this in (17), we get

$$\begin{aligned}\Delta_\Gamma^{(b)} &\leq \sum_{r=1}^{\ell} (p_{i-1}^{(b)})^2 \cdot (1+o(1)) \cdot \eta^r \cdot f_i^2 \\ &= (p_{i-1}^{(b)})^2 \cdot f_i^2 \cdot (1+o(1)) \cdot \sum_{r=1}^{\ell} \eta^r \\ &\leq (50m)^2 \cdot 2\eta\end{aligned}$$

Here the last inequality is by applying Observation 11 and by summing a geometric series. This therefore proves the inductive bound on $\Delta_\Gamma^{(b)}$ assuming (21).

In order to prove (21), we note that by using Janson's inequality (Theorem 13) for Φ^k, we get that

$$\Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Phi^k(\boldsymbol{x}) = 1-\beta)] \leq \prod_{u\leq f_{i-1}-T}(1-p^{(b)}_{\Gamma^{k,u}})\cdot\exp(2\Delta_{\Phi^k})$$

Also observe (Theorem 13) that $\Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Gamma^k(\boldsymbol{x}) = 1-\beta)]$ is lower bounded by $\prod_{u\leq f_{i-1}}(1-p^{(b)}_{\Gamma^{k,u}})$. Therefore, we get

$$\prod_{u\leq f_{i-1}-T}(1-p^{(b)}_{\Gamma^{k,u}}) \leq \frac{\Pr_{\boldsymbol{x}\sim\mu_b^n}[\Gamma^k(\boldsymbol{x}) = 1-\beta]}{\prod_{u>f_{i-1}-T}(1-p^{(b)}_{\Gamma^{k,u}})}.$$

Now, we have $\Delta^{(b)}_{\Phi^k} \leq \Delta^{(b)}_{\Gamma^k}$ by the definitions of these quantities and the fact that Φ^k is obtained from Γ^k by removing some depth-$(i-2)$ subformulas. Also, by the induction hypothesis, we have $\Delta^{(b)}_{\Gamma^k} \leq (C_4 m)^2\eta$. As $\eta = 1/m^{10d}$, we get that $\Delta_{\Phi^k} = o(1)$. Hence, $\exp(2\Delta_{\Phi^k}) = \exp(o(1)) \leq (1+o(1))$. Putting these together, we obtain the following inequality.

$$\Pr_{\boldsymbol{x}\sim\mu_b^n}[(\Phi^k(\boldsymbol{x}) = 1-\beta)] \leq \frac{\Pr_{\boldsymbol{x}\sim\mu_b^n}[\Gamma^k(\boldsymbol{x}) = 1-\beta]}{\prod_{u>f_{i-1}-T}(1-p^{(b)}_{\Gamma^{k,u}})}\cdot(1+o(1)) \tag{23}$$

Now using the induction hypothesis for $p^{(b)}_{\Gamma^{k,u}}$, we get $p^{(b)}_{\Gamma^{k,u}} \leq (1+o(1))\cdot p^{(b)}_{i-2} \leq 2\cdot p^{(b)}_{i-1}$. Using this bound on the value of $p^{(b)}_{\Gamma^{k,u}}$, we get the following lower bound on $\prod_{u>f_{i-1}-T}(1-p^{(b)}_{\Gamma^{k,u}})$.

$$\begin{aligned}\prod_{u>f_{i-1}-T}(1-p^{(b)}_{\Gamma^{k,u}}) &\geq (1-2p^{(b)}_{i-2})^T \\ &\geq 1-2\cdot T\cdot p^{(b)}_{i-2} \\ &\geq 1-2\cdot\gamma\cdot f_{i-1}\cdot\ell\cdot p^{(b)}_{i-2} \\ &\geq (1-o(1))\end{aligned}$$

The third inequality comes from the upper bound on the value of T argued above. Using Observation 11 we get that $f_{i-1}\cdot p^{(b)}_{i-2} \leq 50m$. From our choice of parameters, $\gamma = 1/m^3$ and $\ell \leq m$. Therefore, we get $\gamma\cdot\ell\cdot f_{i-1}p^{(b)}_{i-2} \leq (1/m^3)\cdot m\cdot 50m = o(1)$. This gives the last inequality above. Putting it together, this gives is (21). This finishes the proof of part 2 in Lemma 19.

We now proceed to proving the inductive step for part 1 of Lemma 19. The proof is very similar to the proof of the analogous statement in the base case. We give the details for the sake of completeness. Using Janson's inequality, we get

$$\prod_{j=1}^{f_i}(1-p^{(b)}_{\Gamma^j}) \leq p^{(b)}_{\Gamma} \leq \prod_{j=1}^{f_i}(1-p^{(b)}_{\Gamma^j})\cdot\exp(2\cdot\Delta^{(b)}_{\Gamma}) \tag{24}$$

Using (3) we get $p^{(b)}_{\Gamma} \geq \prod_{j=1}^{f_i}(1-p^{(b)}_{\Gamma^j}) \geq \exp(-\sum_{j\leq f_i}p^{(b)}_{\Gamma^j} - \sum_{j\leq f_i}(p^{(b)}_{\Gamma^j})^2)$. To lower bound this quantity, we will first upper bound $p^{(b)}_{\Gamma^j}$. By using the induction hypothesis, we get $p^{(b)}_{\Gamma^j} \leq p^{(b)}_{i-1}(1+\eta\cdot(C_3 m)^{i-1})$. Using this, we get $\sum_{j\leq f_i}p^{(b)}_{\Gamma^j} \leq f_i\cdot p^{(b)}_{i-1}(1+\eta\cdot(C_3 m)^{i-1})$.

We will also show that $\sum_{j\le f_i}(p_{\Gamma^j}^{(b)})^2$ is negligible. For that observe the following:

$$\begin{aligned}\sum_{j\le f_i}(p_{\Gamma^j}^{(b)})^2 &\le f_i\cdot(p_{i-1}^{(b)}(1+\eta\cdot(C_3m)^{i-1}))^2\\ &\le 4\frac{(f_i\cdot p_{i-1}^{(b)})^2}{f_i}\le\frac{O(m^2)}{\lceil m\cdot 2^m\cdot\ln 2\rceil}\le\eta\cdot(C_3m)^{i-1}\end{aligned}$$

Here the second inequality comes from the fact that $(1+\eta\cdot(C_3m)^{i-1}))\le 2$. The other inequalities easily follow from our choice of parameters and Observation 11.

$$\begin{aligned}p_\Gamma^{(b)} &\ge \exp(-\sum_{j\le f_i}p_{\Gamma^j}^{(b)}-\sum_{j\le f_i}(p_{\Gamma^j}^{(b)})^2)\\ &\ge \exp(-f_i\cdot p_{i-1}^{(b)}(1+\eta\cdot(C_3m)^{i-1})-\eta\cdot(C_3m)^{i-1})\\ &= \exp(-f_i\cdot p_{i-1}^{(b)}-\eta\cdot(C_3m)^{i-1}\cdot(f_ip_{i-1}^{(b)}+1))\\ &\ge (1-p_{i-1}^{(b)})^{f_i}(1-(\eta\cdot(C_3m)^{i-1}(f_ip_{i-1}^{(b)}+1)) \qquad (25)\\ &\ge p_i^{(b)}(1-(\eta\cdot(C_3m)^{i-1}(50m+1)) \qquad (26)\\ &\ge p_i^{(b)}(1-(\eta\cdot(C_3m)^{i-1}C_3m))\\ &= p_i^{(b)}(1-(\eta\cdot(C_3m)^i))\end{aligned}$$

Here, the above inequalities can be obtained primarily by simple rearrangement of terms. The inequality (25) uses (3), while inequality (26) uses the induction hypothesis and Observation 11. This proves the desired lower bound on $p_\Gamma^{(b)}$. Now we prove the upper bound.

$$\begin{aligned}p_\Gamma^{(b)} &\le \prod_{j\le f_i}(1-p_{\Gamma^j}^{(b)})\exp(\Delta_\Gamma^{(b)})\\ &\le \exp(-\sum_{j\le f_i}p_{\Gamma^j}^{(b)}+2\Delta_\Gamma^{(b)})\\ &\le \exp(-p_{i-1}^{(b)}(1-\eta\cdot(C_3m)^{i-1})\cdot f_i+2\Delta_\Gamma^{(b)})\\ &\le \exp(-p_{i-1}^{(b)}f_i)\exp\left(p_{i-1}^{(b)}\cdot\eta\cdot(C_3m)^{i-1}\cdot f_i+2\Delta_\Gamma^{(b)}\right)\\ &= \exp(-p_{i-1}^{(b)})^{f_i}\exp\left(p_{i-1}^{(b)}\cdot\eta\cdot(C_3m)^{i-1}\cdot f_i+2\Delta_\Gamma^{(b)}\right)\\ &\le \left((1-p_{i-1}^{(b)})\cdot\exp((p_{i-1}^{(b)})^2)\right)^{f_i}\cdot\exp\left(p_{i-1}^{(b)}\cdot\eta\cdot(C_3m)^{i-1}\cdot f_i+2\Delta_\Gamma^{(b)}\right) \qquad (27)\\ &= (1-p_{i-1}^{(b)})^{f_i}\cdot\exp\left((p_{i-1}^{(b)})^2\cdot f_i+p_{i-1}^{(b)}\cdot\eta\cdot(C_3m)^{i-1}\cdot f_i+2\Delta_\Gamma^{(b)}\right)\\ &\le p_i^{(b)}\cdot\exp\left(\eta\cdot(C_3m)^{i-1}+50m\cdot\eta\cdot(C_3m)^{i-1}+2\eta\cdot(C_3m)^{i-1}\right) \qquad (28)\\ &\le p_i^{(b)}\cdot\exp\left(\eta\cdot(C_3m)^{i-1}\cdot(50m+3)\right)\\ &\le p_i^{(b)}\cdot(1+2\cdot\eta\cdot(C_3m)^{i-1}\cdot(50m+3)) \qquad (29)\\ &\le p_i^{(b)}\cdot(1+\eta\cdot(C_3m)^i)\end{aligned}$$

Most inequalities above are obtained by simple rearrangement of terms. Inequality (27) is obtained by applying the inequality (b) from (3). Inequality (28) is obtained by applying (5), by using Observation 11, and by using the fact that $f_i \cdot (p_{i-1}^{(b)})^2 \leq \eta \cdot (C_3 m)^{i-1}$. Finally, (29) is obtained by using inequalities (d) and (e) of (3). This completes the proof of part 1 of Lemma 19.

□

7 Lower bounds for the Coin Problem

In this section, we prove Theorem 4. We start with a special case of the theorem (that we call the monotone case) the proof of which is shorter and which suffices for the application to the Fixed-Depth Size-Hierarchy theorem (Theorem 1). We then move on to the general case.

The special case is implicit in the results of O'Donnell and Wimmer [OW07] and Amano [Ama09], but we prove it below for completeness.

7.1 The monotone case

In this section, we prove a near-optimal size lower bound (i.e. matching the upper bound construction from Theorem 3) on the size of any $\mathrm{AC}^0[\oplus]$ formula computing any *monotone* Boolean function solving the δ-coin problem. Observe that this already implies Theorem 1, since the formula F_n from the statement of Theorem 1 computes a monotone function.

Let $g : \{0,1\}^N \to \{0,1\}$ be any *monotone* Boolean function solving the δ-coin problem. Note that the monotonicity of g implies that for all $\alpha \in [0, (1-\delta)/2]$ and $\beta \in [(1+\delta)/2, 1]$, we have

$$\Pr_{\boldsymbol{x} \sim D_\alpha^N}[g(\boldsymbol{x}) = 1] \leq 0.1 \text{ and } \Pr_{\boldsymbol{x} \sim D_\beta^N}[g(\boldsymbol{x}) = 1] \geq 0.9. \tag{30}$$

Let F be any $\mathrm{AC}^0[\oplus]$ formula of size s and depth d computing g. We will show that $s \geq \exp(d \cdot \Omega(1/\delta)^{1/(d-1)})$.

Our main tool is the following implication of the results of Razborov [Raz89], Smolensky [Smo93], and Rossman and Srinivasan [RS17].

Theorem 20. *Let F' be any $\mathrm{AC}^0[\oplus]$ formula of size s' and depth d with n input bits that agrees with the n-bit Majority function in at least a 0.75 fraction of its inputs. Then, $s' \geq \exp(d \cdot \Omega(n)^{1/2(d-1)})$.*

We will use the above theorem to lower bound s (the size of F) by using F to construct a formula F' of size at most s that agrees with the Majority function on $n = \Theta(1/\delta^2)$ bits at a 0.8 fraction of its inputs. Theorem 20 then implies the result.

We now describe the construction of F'. Let $n = \lfloor (1/100\delta^2) \rfloor$. We start by defining a *random* formula $\boldsymbol{F''}$ on n inputs as follows. On input $x = (x_1, \ldots, x_n) \in \{0,1\}^n$, define $\boldsymbol{F''}(x)$ to be $F(x_{\boldsymbol{i}_1}, \ldots, x_{\boldsymbol{i}_N})$ where $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_N$ are chosen i.u.a.r. from $[n]$.

We make the following easy observation. For any $x \in \{0,1\}^n$ and for $\alpha = |x|/n$,

$$\Pr_{\boldsymbol{F''}}[\boldsymbol{F''}(x) = 1] = \Pr_{\boldsymbol{y} \sim D_\alpha^N}[F(\boldsymbol{y}) = 1] = \Pr_{\boldsymbol{y} \sim D_\alpha^N}[g(\boldsymbol{y}) = 1]. \tag{31}$$

In particular, from (30), we see that if $\alpha \le (1-\delta)/2$ or $\alpha \ge (1+\delta)/2$, we have $\Pr_{\boldsymbol{F}''}[\boldsymbol{F}''(x) \neq \mathrm{Maj}_n(x)] \le 0.1$. As a result we get

$$\begin{aligned}\Pr_{\boldsymbol{x}\sim\{0,1\}^n,\boldsymbol{F}''}[\boldsymbol{F}''(\boldsymbol{x}) \neq \mathrm{Maj}_n(\boldsymbol{x})] &= \mathop{\mathbf{E}}_{\boldsymbol{x}}\left[\Pr_{\boldsymbol{F}''}[\boldsymbol{F}''(\boldsymbol{x}) \neq \mathrm{Maj}_n(\boldsymbol{x})]\right] \\ &\le \Pr_{\boldsymbol{x}}[|\boldsymbol{x}|/n \in ((1-\delta)/2,(1+\delta)/2)] \\ &\quad + \max_{\alpha\notin[(1-\delta)/2,(1+\delta)/2)]} \Pr_{\boldsymbol{F}''}[\boldsymbol{F}''(\boldsymbol{x}) \neq \mathrm{Maj}_n(\boldsymbol{x}) \mid |\boldsymbol{x}| = \alpha n] \\ &\le \Pr_{\boldsymbol{x}}[|\boldsymbol{x}|/n \in ((1-\delta)/2,(1+\delta)/2)] + 0.1.\end{aligned}$$

By Stirling's approximation, it follows that for any $i \in [n]$, $\Pr_{\boldsymbol{x}}[|\boldsymbol{x}| = i] \le \binom{n}{\lfloor n/2 \rfloor}/2^n \le 1/\sqrt{n}$. Hence, by a union bound, we have $\Pr_{\boldsymbol{x}}[|\boldsymbol{x}|/n \in ((1-\delta)/2,(1+\delta)/2)] \le (\delta n)\cdot 1/\sqrt{n} \le \delta\sqrt{n} \le 0.1$. Plugging this in above, we obtain

$$\Pr_{\boldsymbol{x}\sim\{0,1\}^n,\boldsymbol{F}''}[\boldsymbol{F}''(\boldsymbol{x}) \neq \mathrm{Maj}_n(\boldsymbol{x})] \le 0.2.$$

By an averaging argument, there is a fixed choice of $\boldsymbol{F}''$, which we denote by F', that agrees with the Majority function Maj_n on a 0.8 fraction of all inputs. Note that $F' = F(x_{i_1},\ldots,x_{i_N})$ for some choices of $i_1,\ldots,i_N \in [n]$. Hence, F' is a circuit of depth d and size at most s.

Theorem 20 now implies the lower bound on s.

7.2 The general case

In this section, we prove a general lower bound on the size of any $\mathrm{AC}^0[\oplus]$ formula that solves the coin problem (not necessarily by computing a monotone function). The main technical result is the following theorem about polynomials that solve the coin problem.

Theorem 21. *Let $g \in \mathbb{F}_2[x_1,\ldots,x_N]$ solve the δ-coin problem. Then,* $\deg(g) = \Omega(1/\delta)$.

Given the above result, it is easy to prove Theorem 4 in its general form.

Proof of Theorem 4. Assume that F is an $\mathrm{AC}^0[\oplus]$ formula F of size s and depth d on N inputs that solves the δ-coin problem.

Building on Razborov [Raz89], Rossman and Srinivasan [RS17] show that for any such $\mathrm{AC}^0[\oplus]$ formula F of size s and depth d and any probability distribution μ on $\{0,1\}^N$, there exists a polynomial $P \in \mathbb{F}[x_1,\ldots,x_N]$ of degree $O((\log s)/d)^{d-1}$ such that

$$\Pr_{\boldsymbol{x}\sim\mu}[P(\boldsymbol{x}) \neq F(\boldsymbol{x})] \le 0.05.$$

Taking $\mu = (\mu_0^N + \mu_1^N)/2$, we have for each $b \in \{0,1\}$, the above polynomial P satisfies

$$\Pr_{\boldsymbol{x}\sim\mu_b^N}[P(\boldsymbol{x}) \neq F(\boldsymbol{x})] \le 2\Pr_{\boldsymbol{x}\sim\mu}[P(\boldsymbol{x}) \neq F(\boldsymbol{x})] \le 0.1. \tag{32}$$

In particular, if F solves the δ-coin problem, then P solves the δ-coin problem with error at most 0.2. By Fact 9 applied with t being a suitably large constant, it follows that there is a polynomial $Q \in \mathbb{F}[x_1,\ldots,x_N]$ that solves the δ-coin problem (with error at most 0.1) and satisfies $\deg(Q) \le t\cdot\deg(P) = O(\deg(P))$. By Theorem 21, it follows that $\deg(Q) = \Omega(1/\delta)$ and hence we have $\deg(P) = \Omega(1/\delta)$ as well.

Since $\deg(P) = O((\log s)/d)^{(d-1)}$, we get $s \ge \exp(\Omega(d\cdot(1/\delta)^{1/(d-1)}))$. □

We now turn to the proof of Theorem 21.

7.2.1 Proof of Theorem 21

We define a *probabilistic function* to be a random function $\boldsymbol{g} : \{0,1\}^N \to \{0,1\}$, chosen according to some distribution. We say that $\deg(\boldsymbol{g}) \leq D$ if this distribution is supported over polynomials (from $\mathbb{F}_2[x_1, \ldots, x_N]$) of degree at most D. A probabilistic function solves the δ-coin problem with error at most ε if it satisfies (1), where the probability is additionally taken over the randomness used to sample $\boldsymbol{g}$. If no mention is made of the error, we assume that it is 0.1. Note that a standard (i.e. non-random) function is also a probabilistic function of the same degree.

We will prove the stronger statement that any probabilistic function $\boldsymbol{g}$ solving the δ-coin problem must have degree $\Omega(1/\delta)$. [13]

Given a probabilistic function $\boldsymbol{g} : \{0,1\}^N \to \{0,1\}$, we define the *profile of* $\boldsymbol{g}$, denoted $\pi_{\boldsymbol{g}}$, to be a function $\pi_{\boldsymbol{g}} : [0,1] \to [0,1]$ where

$$\pi_{\boldsymbol{g}}(\alpha) = \Pr_{\substack{\boldsymbol{g}, \\ \boldsymbol{x} \sim D_\alpha^N}} [\boldsymbol{g}(\boldsymbol{x}) = 1].$$

Note that since $\boldsymbol{g}$ solves the δ-coin problem, we have

$$\pi_{\boldsymbol{g}}((1-\delta)/2) \leq 0.1 \text{ and } \pi_{\boldsymbol{g}}((1+\delta)/2) \geq 0.9. \tag{33}$$

The proof of the lower bound on $\deg(\boldsymbol{g})$ proceeds in two phases. In the first phase, we use $\boldsymbol{g}$ to obtain a probabilistic function $\boldsymbol{h}$ (of related degree) which satisfies a stronger criterion than (33): namely that the profile of $\boldsymbol{h}$ is small in an *interval* close to $(1-\delta')/2$ and large in an interval close to $(1+\delta')/2$ (for some $\delta' \leq \delta$). In the second phase, we use algebraic arguments [Smo87] to lower bound $\deg(\boldsymbol{h})$, which leads to a lower bound on $\deg(\boldsymbol{g})$.

Let $r, t \in \mathbb{N}$ and $\zeta \in (0,1)$ denote absolute constants that we will fix later on in the proof.

We start the first phase of the proof as outlined above. We iteratively define a sequence of probabilistic functions $(\boldsymbol{g}_k)_{k \geq 0}$ where $\boldsymbol{g}_k : \{0,1\}^{N_k} \to \{0,1\}$ solves the δ_k-coin problem where N_k, δ_k are parameters that are defined below.

- The function $\boldsymbol{g}_0$ is simply the function $\boldsymbol{g}$. Hence, $N_0 = N$ and we can take $\delta_0 = \delta$.
- Having defined $\boldsymbol{g}_k$, we consider which of the following 3 cases occur.
  - Case 1: There is an $\alpha \in ((1-\delta_k)/2, (1-\delta_k)/2 + \delta_k/r]$ such that $\pi_{\boldsymbol{g}_k}(\alpha) \geq 0.4$.
  - Case 2: There is an $\beta \in [(1+\delta_k)/2 - \delta_k/r, (1+\delta_k)/2)$ such that $\pi_{\boldsymbol{g}_k}(\beta) \leq 0.6$.
  - Case 3: Neither Case 1 nor Case 2 occur. In this case, the sequence of probabilistic functions ends with $\boldsymbol{g}_k$.
- If Case 1 or Case 2 occurs, we extend the sequence by defining $\boldsymbol{g}_{k+1}$. For simplicity, we assume Case 1 occurs (Case 2 is handled similarly). Note that in this case we have

$$\pi_{\boldsymbol{g}_k}((1-\delta)/2) \leq 0.1 \text{ and } \pi_{\boldsymbol{g}_k}(\alpha) \geq 0.4 \tag{34}$$

[13] While formally stronger, this statement is more or less equivalent, since given such a probabilistic function $\boldsymbol{g}$, one can always extract a deterministic function of the same degree that solves the coin problem with error at most 0.21 by an averaging argument. Then using error reduction (Fact 9), we can obtain a deterministic function with a slightly larger degree that solves the coin problem with error 0.1.

for some $\alpha \in ((1-\delta_k)/2, (1-\delta_k)/2 + \delta_k/r]$.

We will need the following technical claim.

Claim 22. *Let $\delta', \delta'' \in (0,1)$ be such that $\delta'' \geq 4\delta'$. Assume we have $\alpha_1, \alpha_2, \beta_1, \beta_2 \in (0,1)$ such that $(1/4) \leq \alpha_1 \leq (1/2), \alpha_2 = \alpha_1 + \delta', \beta_1 = (1-\delta'')/2, \beta_2 = (1+\delta'')/2$. Then, there exist $\gamma, \eta \in [0,1]$ such that for each $i \in [2]$, $\alpha_i = \gamma \cdot \beta_i + (1-\gamma) \cdot \eta$.*

Proof. We immediately get

$$(\alpha_2 - \alpha_1) = \gamma(\beta_2 - \beta_1) \quad \Longrightarrow \quad \gamma = \frac{\alpha_2 - \alpha_1}{\beta_2 - \beta_1} = \frac{\delta'}{\delta''} \in (0, 1/4].$$

Then

$$\eta = \frac{\alpha_1 - \beta_1\gamma}{1-\gamma} > \alpha_1 - \beta_1\gamma \geq \frac{1-\beta_1}{4} > 0.$$

Further

$$\eta = \frac{\alpha_1 - \beta_1\gamma}{1-\gamma} \leq \frac{4\alpha_1}{3} \leq \frac{2}{3}.$$

So $\eta \in (0, 2/3]$. □

Applying the above claim to $\alpha_1 = (1-\delta_k)/2, \alpha_2 = \alpha, \delta' = \alpha_2 - \alpha_1$ and $\delta'' = 4\delta_k/r$, we see that there exist $\gamma, \eta \in [0,1]$ such that

$$(1-\delta_k)/2 = \gamma \cdot (1-\delta'')/2 + (1-\gamma) \cdot \eta \text{ and } \alpha = \gamma \cdot (1+\delta'')/2 + (1-\gamma) \cdot \eta \tag{35}$$

To define the function $\boldsymbol{g}_{k+1}$, we start with an intermediate probabilistic function $\boldsymbol{h}_k$ on N_k inputs. On any input $x \in \{0,1\}^{N_k}$, the function $\boldsymbol{h}_k$ is defined as follows.

$\boldsymbol{h}_k(x)$:

- Sample a random $\boldsymbol{b}$ from the distribution $D_\gamma^{N_k}$ and $\boldsymbol{y}$ from the distribution $D_\eta^{N_k}$.
- Define $\boldsymbol{z} \in \{0,1\}^{N_k}$ by $\boldsymbol{z}_i = \boldsymbol{b}_i \cdot x_i + (1 - \boldsymbol{b}_i) \cdot \boldsymbol{y}_i$.
- $\boldsymbol{h}_k(x)$ is defined to be $\boldsymbol{g}_k(\boldsymbol{z})$.

Note that as each $\boldsymbol{z}_i$ is a (random) degree-1 polynomial in x, the probabilistic function $\boldsymbol{h}_k(x)$ satisfies $\deg(\boldsymbol{h}_k) \leq \deg(\boldsymbol{g}_k)$.

Also, note that when $\boldsymbol{x}$ is sampled from the $D^{N_k}_{(1-\delta'')/2}$ or $D^{N_k}_{(1+\delta'')/2}$, then by (35), $\boldsymbol{z}$ has the distribution $D^{N_k}_{(1-\delta_k)/2}$ or $D^{N_k}_{\alpha}$ respectively. Hence, we get

$$\pi_{\boldsymbol{h}_k}((1-\delta'')/2) = \pi_{\boldsymbol{g}_k}((1-\delta_k)/2) \leq 0.1 \text{ and } \pi_{\boldsymbol{h}_k}((1+\delta'')/2) = \pi_{\boldsymbol{g}_k}(\alpha) \geq 0.4.$$

We are now ready to define $\boldsymbol{g}_{k+1}$. Let $\mathrm{Thr}^t_{t/4} : \{0,1\}^t \to \{0,1\}$ be the Boolean function that accepts inputs of Hamming weight at least $t/4$. We set $N_{k+1} = N_k \cdot t$ and define $\boldsymbol{g}_{k+1} : \{0,1\}^{N_{k+1}} \to \{0,1\}$ by

$$\boldsymbol{g}_{k+1}(x) = \mathrm{Thr}^t_{t/4}(\boldsymbol{h}_k^{(1)}(x^{(1)}), \ldots, \boldsymbol{h}_k^{(t)}(x^{(t)}))$$

where $\boldsymbol{h}_k^{(1)}, \ldots, \boldsymbol{h}_k^{(t)}$ are *independent* copies of the probabilistic function $\boldsymbol{h}_k$ and $x^{(i)} \in \{0,1\}^{N_k}$ is defined by $x_j^{(i)} = x_{(i-1)N_k+j}$ for each $j \in [N_k]$. Clearly, $\deg(\boldsymbol{g}_{k+1}) \leq \deg(\mathrm{Thr}_{t/4}^t) \cdot \deg(\boldsymbol{h}_k) \leq \deg(\boldsymbol{g}_k) \cdot t$.

Note that if $\boldsymbol{x}$ is chosen according to the distribution $D^{N_{k+1}}_{(1-\delta'')/2}$, each $\boldsymbol{h}_k^{(i)}(\boldsymbol{x}^{(i)})$ is a Boolean random variable that is 1 with probability at most 0.1. Similarly, if $\boldsymbol{x}$ is chosen according to the distribution $D^{N_{k+1}}_{(1+\delta'')/2}$, each $\boldsymbol{h}_k^{(i)}(\boldsymbol{x}^{(i)})$ is a Boolean random variable that is 1 with probability at least 0.4. By a standard Chernoff bound (see e.g. [DP09, Theorem 1.1]), we see that for a large enough absolute constant t,

$$\pi_{\boldsymbol{g}_{k+1}}((1-\delta'')/2) = \exp(-\Omega(t)) \leq 0.1 \text{ and } \pi_{\boldsymbol{g}_{k+1}}((1+\delta'')/2) = 1 - \exp(-\Omega(t)) \geq 0.9. \quad (36)$$

We now fix the value of t so that the above inequalities hold. Note that we have shown the following.

Observation 23. *$\boldsymbol{g}_{k+1} : \{0,1\}^{N_{k+1}} \to \{0,1\}$ is a probabilistic function that solves the δ_k-coin problem where $N_{k+1} = N_k \cdot t, \deg(\boldsymbol{g}_{k+1}) \leq \deg(\boldsymbol{g}_k) \cdot t$ and $\delta_{k+1} \leq 4\delta_k/r$.*

We now argue that, for $r = 10t$, the above process cannot produce an infinite sequence of probabilistic functions. In other words, there is a fixed k such that $\boldsymbol{g}_k$ is in neither Case 1 nor Case 2 mentioned above.

Assume to the contrary that the above process produces an infinite sequence of probabilistic functions. By Observation 23 and induction, we see that $\boldsymbol{g}_k$ is a probabilistic function on at most $N \cdot t^k$ variables solving the δ_k-coin problem for $\delta_k \leq \delta \cdot (4/r)^k$. We now appeal to the following standard fact.

Fact 24 (Folklore)**.** *Let $\delta' \in (0,1)$ and $N' \in \mathbb{N}$ be arbitrary. Then, the statistical distance between $D^{N'}_{(1-\delta')/2}$ and $D^{N'}_{(1+\delta')/2}$ is at most $O(\sqrt{N'} \cdot \delta')$.*

Thus, for $\boldsymbol{g}_k$ to able to solve the δ_k-coin problem with N_k samples, we must have $\sqrt{N_k} \cdot \delta_k \geq \alpha_0$ for some absolute positive constant α_0. On the other hand, this cannot be true for large enough k, since $\sqrt{N_k}\delta_k \leq N_k\delta_k \leq N\delta \cdot (4t/r)^k$ and $r \geq 10t$. This yields a contradiction.

Thus, we have shown that for large enough k, the function $\boldsymbol{g}_k$ is in neither Case 1 nor Case 2. Equivalently, for any $\alpha \in [1/2 - \delta_k, 1/2 - \delta_k + \delta_k/r]$ and any $\beta \in [1/2 + \delta_k - \delta_k/r, 1/2 + \delta_k]$, we have

$$\pi_{\boldsymbol{g}_k}(\alpha) \leq 0.4 \text{ and } \pi_{\boldsymbol{g}_k}(\beta) \geq 0.6.$$

Using error reduction as above, we can obtain a probabilistic function that satisfies the above inequalities with parameters $\zeta := \exp(-10r^2)$ and $1 - \zeta$ respectively. Set $\ell = 10\lceil \log(1/\zeta) \rceil$ and define $\boldsymbol{h} : \{0,1\}^{N_k \cdot \ell} \to \{0,1\}$ by

$$\boldsymbol{h}(x) = \mathrm{Maj}_\ell(\boldsymbol{g}_k^{(1)}(x^{(1)}), \ldots, \boldsymbol{g}_k^{(t)}(x^{(t)}))$$

where $\boldsymbol{g}_k^{(1)}, \ldots, \boldsymbol{g}_k^{(t)}$ are *independent* copies of the probabilistic function $\boldsymbol{g}_k$ and $x^{(i)} \in \{0,1\}^{N_k}$ is defined by $x_j^{(i)} = x_{(i-1)N_k+j}$ for each $j \in [N_k]$.

Clearly, $\deg(\boldsymbol{h}) \leq \ell \cdot \deg(\boldsymbol{g}_k) = O(\deg(\boldsymbol{g}_k))$ as ℓ is an absolute constant. Further, by the Chernoff bound, $\boldsymbol{h}$ satisfies

$$\pi_{\boldsymbol{h}}(\alpha) \leq \zeta \text{ and } \pi_{\boldsymbol{h}}(\beta) \geq 1 - \zeta. \tag{37}$$

This concludes the first phase of the proof. In the second phase, we will show the following lower bound on $\deg(\boldsymbol{h})$.

Claim 25. $\deg(\boldsymbol{h}) \geq \Omega(1/\delta_k)$.

Note that this immediately implies the result since we have

$$\deg(\boldsymbol{g}) = \deg(\boldsymbol{g}_0) \geq \frac{\deg(\boldsymbol{g}_k)}{t^k} = \Omega\left(\frac{\deg(\boldsymbol{h})}{t^k}\right) = \Omega\left(\frac{1}{\delta_k t^k}\right) \geq \Omega\left(\frac{1}{\delta \cdot (4t/r)^k}\right) \geq \Omega(1/\delta)$$

where we have used Observation 23 and the fact that $r \geq 10t$.

It therefore suffices to prove Claim 25. We prove this in two steps.

We start with an extension of a lower bound of Smolensky [Smo87] (see also the earlier results of Razborov [Raz89] and Szegedy [Sze89]) on the degrees of polynomials approximating the Majority function.[14] Our method is a slightly different phrasing of this bound following the results of Aspnes, Beigel, Furst and Rudich [ABFR94], Green [Gre00] and Kopparty and Srinivasan [KS12].[15]

Lemma 26 (A slight extension of Smolensky's bound)**.** *Let $h : \{0,1\}^n \to \{0,1\}$ be a (deterministic) function satisfying the following. There exist integers $D < R < n/2$ such that E_h^R defined by*

$$E_h^R = \{x \in \{0,1\}^n \mid h(x) \neq \mathrm{Maj}_n(x), |x| \notin (R, n-R)\} \tag{38}$$

satisfies $|E_h^R| < \binom{n}{\leq (R-D)}$.[16] Then, $\deg(h) > D$.

Proof. Consider the vector space V_{R-D} of all multilinear polynomials of degree $\leq (R-D)$. A generic polynomial $g \in V_{R-D}$ is given by

$$g(x_1, \ldots, x_n) = \sum_{|S| \leq R-D} a_S \cdot \prod_{i \in S} x_i$$

where $a_S \in \mathbb{F}_2$ for each S. We claim that there is a *non-zero* g as above that vanishes at *all points* in E_h^R. To see this, note that finding such a g is equivalent to finding a non-zero assignment to the a_S so that the resulting g vanishes at E_h^R. Vanishing at any point of $\{0,1\}^n$ gives a linear constraint on the coefficients a_S. Since we have $|E_h^R| < \binom{n}{\leq (R-D)}$, we have a homogeneous linear system with more variables than constraints and hence, there exists a non-zero multilinear polynomial g of degree $\leq (R-D)$ which vanishes on E_h^R. Thus, there is a non-zero g as claimed.

Let $B_1 = \{x \in \{0,1\}^n \mid |x| \leq R\}$ and $B_2 = \{x \in \{0,1\}^n \mid |x| \geq n - R\}$. Note that B_1 and B_2 are both Hamming balls of radius R. Let f be the pointwise product of the functions g and h. Note that f can be represented as a *multilinear* polynomial of degree at most $\deg(g) + \deg(h)$ (by replacing x_i^2 by x_i as necessary in the polynomial $g \cdot h$).

We will need the following standard fact (see e.g. [KS12] for a proof).

[14] In [Smo87], Smolensky proves a lower bound for MOD_p functions. However, the same idea also can be used to prove lower bounds for the Majority function, as observed by Szegedy [Sze89].

[15] It can be viewed as a 'dual' version of Smolensky's proof. Smolensky's standard proof can also be modified to yield this.

[16] Recall that $\binom{n}{\leq i}$ denotes $\sum_{j=0}^{i} \binom{n}{j}$.

Fact 27. *Let P be a non-zero multilinear polynomial of degree $\leq d$ in n variables. Then P cannot vanish on a Hamming ball of radius d.*

On B_1, g vanishes wherever h does not (by definition of E_h^R) and therefore f vanishes everywhere.

On B_2, h is non-vanishing wherever g is non-vanishing. But since B_2 is a Hamming ball of radius $R > R - D \geq \deg(g)$, Fact 27 implies that $g(x_0) \neq 0$ for some point x_0 in B_2. Therefore $h(x_0) \neq 0$ and so $f(x_0) \neq 0$. In particular, f is a non-zero multilinear polynomial of degree at most $\deg(g) + \deg(h)$.

Since f is non-zero and vanishes on B_1 which is a Hamming ball of radius R, Fact 27 implies that $\deg(f) > R$. Therefore $(R - D) + \deg(h) \geq \deg(g) + \deg(h) \geq \deg(f) > R \Rightarrow \deg(h) > D$. $\square$

We now prove Claim 25.

Proof of Claim 25. The idea is to use $\boldsymbol{h}$ to produce a deterministic function h of the same degree to which Lemma 26 is applicable. Let M denote $N_k \cdot \ell$, the sample complexity (i.e. number of inputs) of $\boldsymbol{h}$.

Let $n = \lceil r^2/\delta_k^2 \rceil$. Define a probabilistic function $\tilde{\boldsymbol{h}} : \{0,1\}^n \to \{0,1\}$ as follows. On any input $x \in \{0,1\}^n$, we choose $\boldsymbol{i}_1, \ldots, \boldsymbol{i}_M \in [n]$ i.u.a.r. and set

$$\tilde{\boldsymbol{h}}(x) = \boldsymbol{h}(x_{\boldsymbol{i}_1}, \ldots, x_{\boldsymbol{i}_M}).$$

Clearly, we have $\deg(\tilde{\boldsymbol{h}}) \leq \deg(\boldsymbol{h})$. Also note that for any $x \in \{0,1\}^n$, we have

$$\Pr_{\tilde{\boldsymbol{h}}}[\tilde{\boldsymbol{h}}(x) = 1] = \pi_{\boldsymbol{h}}(|x|/n)$$

since in this case $(x_{\boldsymbol{i}_1}, \ldots, x_{\boldsymbol{i}_M})$ is drawn from the distribution $D^M_{|x|/n}$. By (37), we have for any x such that $|x|/n \in [1/2 - \delta_k, 1/2 - \delta_k + \delta_k/r] \cup [1/2 + \delta_k - \delta_k/r, 1/2 + \delta_k]$,

$$\Pr_{\tilde{\boldsymbol{h}}}[\tilde{\boldsymbol{h}}(x) \neq \mathrm{Maj}_n(x)] \leq \zeta.$$

In particular, for $\boldsymbol{x}$ chosen uniformly at random from $\{0,1\}^n$, we have

$$\Pr_{\boldsymbol{x},\tilde{\boldsymbol{h}}}[\tilde{\boldsymbol{h}}(\boldsymbol{x}) \neq \mathrm{Maj}_n(\boldsymbol{x}) \mid |\boldsymbol{x}|/n \in [1/2 - \delta_k + \delta_k/r, 1/2 - \delta_k] \cup [1/2 + \delta_k - \delta_k/r, 1/2 + \delta_k]] \leq \zeta.$$

We will apply Lemma 26 below with n unchanged, $R = \lfloor (1/2 - \delta_k + \delta_k/r) \cdot n \rfloor$, and $D = \lfloor \delta_k n/(2r) \rfloor$. For these parameters, we have

$$\begin{aligned}
\mathop{\mathbf{E}}_{\tilde{\boldsymbol{h}}}\left[\frac{|E^R_{\tilde{\boldsymbol{h}}}|}{2\binom{n}{\leq R}}\right] &= \mathop{\mathbf{E}}_{\tilde{\boldsymbol{h}}}\left[\Pr_{\boldsymbol{x}}[\tilde{\boldsymbol{h}}(\boldsymbol{x}) \neq \mathrm{Maj}_n(\boldsymbol{x}) \mid |\boldsymbol{x}| \notin (R, n-R)]\right] \\
&= \Pr_{\boldsymbol{x},\tilde{\boldsymbol{h}}}[\tilde{\boldsymbol{h}}(\boldsymbol{x}) \neq \mathrm{Maj}_n(\boldsymbol{x}) \mid |\boldsymbol{x}| \notin (R, n-R)] \\
&\leq \Pr_{\boldsymbol{x},\tilde{\boldsymbol{h}}}[\tilde{\boldsymbol{h}}(\boldsymbol{x}) \neq \mathrm{Maj}_n(\boldsymbol{x}) \mid |\boldsymbol{x}|/n \in [1/2 - \delta_k, 1/2 - \delta_k + \delta_k/r] \cup [1/2 + \delta_k - \delta_k/r, 1/2 + \delta_k]] \\
&+ \Pr_{\boldsymbol{x},\tilde{\boldsymbol{h}}}[|\boldsymbol{x}|/n \notin [1/2 - \delta_k, 1/2 + \delta_k] \mid |\boldsymbol{x}| \notin (R, n-R)] \\
&\leq \zeta + \Pr_{\boldsymbol{x}}[|\boldsymbol{x}|/n \notin [1/2 - \delta_k, 1/2 + \delta_k] \mid |\boldsymbol{x}| \notin (R, n-R)] \leq \zeta + \frac{2\binom{n}{\leq R'}}{2\binom{n}{\leq R}}
\end{aligned}$$

where $R' = \lfloor (1/2 - \delta_k)n \rfloor$. Hence, we have

$$\mathop{\mathbf{E}}_{\tilde{\boldsymbol{h}}}\left[|E^R_{\tilde{\boldsymbol{h}}}|\right] \le 2\zeta\binom{n}{\le R} + 2\binom{n}{\le R'}.$$

By averaging, we can fix a deterministic $h : \{0,1\}^n \to \{0,1\}$ so that $\deg(h) \le \deg(\tilde{\boldsymbol{h}})$ and we have the same bound for $|E^R_h|$. By a computation, we show below that the above bound is strictly smaller than $\binom{n}{\le R-D}$. Lemma 26 then implies that $\deg(h) \ge D = \Omega(1/\delta_k)$, completing the proof of Claim 25.

It remains to prove only that

$$2\zeta\binom{n}{\le R} + 2\binom{n}{\le R'} \le \binom{n}{\le R - D}. \tag{39}$$

Recall that $\zeta = e^{-10r^2}$, where $r = 10t$ and $t \ge 1$.

Now let $V_n(\alpha) = \frac{1}{2^n}\binom{n}{\le n/2 - \alpha\sqrt{n}}$ and $\Phi(x) = \int_x^\infty \frac{e^{-t^2/2}}{\sqrt{2\pi}}\,dt$. Then by the Central Limit Theorem (see, e.g., [Fel71, Chapter VIII, vol II]), for any fixed α, we have

$$\lim_{n\to\infty}(V_n(\alpha) - \Phi(2\alpha)) = 0. \tag{40}$$

Also, we have the estimates [Fel71, Lemma 2, Chapter VII, vol I]

$$\frac{1}{\sqrt{2\pi}} \cdot \left(\frac{1}{x} - \frac{1}{x^3}\right) e^{-x^2/2} \le \Phi(x) \le \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{x} e^{-x^2/2}, \quad x > 0. \tag{41}$$

Note that we have $\delta_k n \ge \delta_k \cdot (r/\delta_k) \cdot \sqrt{n} = r\sqrt{n}$. So we get

$$\begin{aligned}
\frac{1}{2^n} \cdot 2\zeta\binom{n}{\le R} + \frac{1}{2^n}\binom{n}{\le R'} &\le \frac{1}{2^n} \cdot 2\zeta\binom{n}{\le n/2 - (1-1/r)\delta_k n} + \frac{1}{2^n} \cdot 2\binom{n}{\le n/2 - \delta_k n} \\
&\le \frac{1}{2^n} \cdot 2\zeta\binom{n}{\le n/2 - (1-1/r)r\sqrt{n}} + \frac{1}{2^n} \cdot 2\binom{n}{\le n/2 - r\sqrt{n}} \\
&= 2\zeta \cdot V_n(r-1) + V_n(r) \\
&\le 2\zeta + V_n(r) \\
&\le 2e^{-10r^2} + \frac{1}{\sqrt{2\pi}r} \cdot e^{-r^2/2} + o(1)
\end{aligned} \tag{42}$$

where the second-last inequality uses the fact $V_n(\alpha) \le 1$ for any α, and the final inequality follows from the definition of ζ and (40) and (41) above. Note that the $o(1)$ above goes to 0 as $n \to \infty$ (which happens as $\delta \to 0$).

Further, we have $\delta_k n \le \delta_k \cdot ((r/\delta_k)+1) \cdot \sqrt{n} = (r+o(1))\sqrt{n}$, which yields

$$\begin{aligned}
\frac{1}{2^n}\binom{n}{\le R-D} &= \frac{1}{2^n}\binom{n}{\le \lfloor n/2-(1-1/r)\delta_k n\rfloor - \lfloor (1/2r)\delta_k n\rfloor} \\
&\ge \frac{1}{2^n}\binom{n}{\le n/2-(1-1/2r)\delta_k n - 1} \\
&= \frac{1}{2^n}\binom{n}{\le n/2-(1-1/2r+o(1))\cdot \delta_k n} \\
&\ge \frac{1}{2^n}\binom{n}{\le n/2-(1-1/2r+o(1))\cdot (r+o(1))\sqrt{n}} \\
&= \frac{1}{2^n}\binom{n}{\le (n/2)-(r-(1/2)+o(1))\cdot \sqrt{n}} \ge V_n(r-(1/4)) \qquad \text{(for large enough } n\text{)} \\
&\ge \frac{1}{\sqrt{2\pi}} e^{-(r-(1/4))^2/2} \cdot \left(\frac{1}{r-(1/4)} - \frac{1}{(r-1/4)^3}\right) - o(1) \\
&\ge \frac{1}{\sqrt{2\pi}} e^{-(r^2/2)+2} \cdot \frac{1}{2r} - o(1) \qquad (43)
\end{aligned}$$

where the second-last inequality uses (40) and (41) and the final inequality uses the fact that $r \ge 10t \ge 10$. From (42) and (43), the inequality follows for all large n. □

8 Open Problems

We close with some open problems.

- We get almost optimal upper and lower bounds on the complexity of the δ-coin problem. In particular, as mentioned in Theorem 3 and Theorem 4, the upper and lower bounds on the size of $\mathrm{AC}^0[\oplus]$ formulas computing the δ-coin problem are $\exp(O(d(1/\delta)^{1/(d-1)}))$ and $\exp(\Omega(d(1/\delta)^{1/(d-1)}))$, respectively. It may be possible to get even tighter bounds by exactly matching the constants in the exponents in these bounds. The strongest result in this direction would be to give explicit separations between $\mathrm{AC}^0[\oplus]$ formulas of size s and size $s^{1+\varepsilon}$ for any fixed $\varepsilon > 0$ (for $s = n^{O(1)}$, for example).

  For circuits, we have a *non-explicit* upper bound of $\exp(O((1/\delta)^{1/(d-1)}))$ due to Rossman and Srinivasan [RS17]. It would be interesting to achieve this upper bound with an explicit family of circuits.

- In Theorem 3 we get a $(1/\delta)^{2^{O(d)}}$ upper bound on the sample complexity of the δ-coin problem. We believe that this can be improved to $O((1/\delta)^2)$. (We get this for depth-2 formulas, but not for larger depths.)

- Finally, can we match the AC^0 size-hierarchy theorem of Rossman [Ros08b] by separating $\mathrm{AC}^0[\oplus]$ circuits of size s and some fixed depth (say 2) and $\mathrm{AC}^0[\oplus]$ circuits of size s^{ε} (for some absolute constant $\varepsilon > 0$) and *any* constant depth?

Acknowledgements. The authors thank Paul Beame, Prahladh Harsha, Ryan O'Donnell, Ben Rossman, Rahul Santhanam, Ramprasad Saptharishi, Madhu Sudan, Avishay Tal, Emanuele Viola,

and Avi Wigderson for helpful discussions and comments. The authors also thank the anonymous referees of STOC 2018 for their comments.

References

[Aar10] Scott Aaronson. BQP and the polynomial hierarchy. In *STOC*, pages 141–150. ACM, 2010.

[AB84] Miklós Ajtai and Michael Ben-Or. A theorem on probabilistic constant depth computations. In *STOC*, pages 471–474. ACM, 1984.

[ABFR94] James Aspnes, Richard Beigel, Merrick L. Furst, and Steven Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):135–148, 1994.

[Ama09] Kazuyuki Amano. Bounds on the size of small depth circuits for approximating majority. In *ICALP (1)*, Lecture Notes in Computer Science, pages 59–70. Springer, 2009.

[Ama10] Kazuyuki Amano. k-subgraph isomorphism on AC^0 Circuits. *Computational Complexity*, 19(2):183–210, 2010. URL: `https://doi.org/10.1007/s00037-010-0288-y`, `doi:10.1007/s00037-010-0288-y`.

[AS92] Noga Alon and Joel Spencer. *The Probabilistic Method*. John Wiley & Sons Inc., 1992.

[BGL06] Nayantara Bhatnagar, Parikshit Gopalan, and Richard J. Lipton. Symmetric polynomials over z_m and simultaneous communication protocols. *J. Comput. Syst. Sci.*, 72(2):252–285, 2006. URL: `https://doi.org/10.1016/j.jcss.2005.06.007`, `doi:10.1016/j.jcss.2005.06.007`.

[Bop97] Ravi B. Boppana. The average sensitivity of bounded-depth circuits. *Inf. Process. Lett.*, 63(5):257–261, 1997.

[BV10] Joshua Brody and Elad Verbin. The coin problem and pseudorandomness for branching programs. In *FOCS*, pages 30–39. IEEE Computer Society, 2010.

[CGR14] Gil Cohen, Anat Ganor, and Ran Raz. Two sides of the coin problem. In *APPROX-RANDOM*, volume 28 of *LIPIcs*, pages 618–629. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2014.

[CHHL18] Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom generators from polarizing random walks. In *33rd Computational Complexity Conference, CCC 2018, June 22-24, 2018, San Diego, CA, USA*, pages 1:1–1:21, 2018. URL: `https://doi.org/10.4230/LIPIcs.CCC.2018.1`, `doi:10.4230/LIPIcs.CCC.2018.1`.

[CHLT18] Eshan Chattopadhyay, Pooya Hatami, Shachar Lovett, and Avishay Tal. Pseudorandom generators from the second fourier level and applications to AC0 with parity gates. *Electronic Colloquium on Computational Complexity (ECCC)*, 25:155, 2018. URL: `https://eccc.weizmann.ac.il/report/2018/155`.

[Coo73] Stephen A. Cook. A hierarchy for nondeterministic time complexity. *J. Comput. Syst. Sci.*, 7(4):343–353, 1973. URL: https://doi.org/10.1016/S0022-0000(73)80028-5, doi:10.1016/S0022-0000(73)80028-5.

[DP09] Devdatt Dubhashi and Alessandro Panconesi. *Concentration of Measure for the Analysis of Randomized Algorithms*. Cambridge University Press, New York, NY, USA, 1st edition, 2009.

[Fel71] William Feller. *An introduction to probability theory and its applications*. John Wiley & Sons Inc., New York, NY, USA, 2nd edition, 1971.

[FS04] Lance Fortnow and Rahul Santhanam. Hierarchy theorems for probabilistic polynomial time. In *FOCS*, pages 316–324. IEEE Computer Society, 2004.

[Gre00] F. Green. A complex-number fourier technique for lower bounds on the mod-m degree. *computational complexity*, 9(1):16–38, Nov 2000. URL: https://doi.org/10.1007/PL00001599, doi:10.1007/PL00001599.

[Has89] John Hastad. Almost optimal lower bounds for small depth circuits. *Advances in Computing Research*, 5:143–170, 1989.

[HRST17] Johan Håstad, Benjamin Rossman, Rocco A. Servedio, and Li-Yang Tan. An average-case depth hierarchy theorem for boolean circuits. *J. ACM*, 64(5):35:1–35:27, 2017. URL: http://doi.acm.org/10.1145/3095799, doi:10.1145/3095799.

[HS65] J. Hartmanis and R. E. Stearns. On the computational complexity of algorithms. *Transactions of the American Mathematical Society*, 117:285306, 1965.

[Jan90] Svante Janson. Poisson approximation for large deviations. *Random Struct. Algorithms*, 1(2):221–230, 1990.

[KPPY84] Maria Klawe, Wolfgang J. Paul, Nicholas Pippenger, and Mihalis Yannakakis. On monotone formulae with restricted depth. In *Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing*, STOC '84, pages 480–487, New York, NY, USA, 1984. ACM. URL: http://doi.acm.org/10.1145/800057.808717, doi:10.1145/800057.808717.

[KS12] Swastik Kopparty and Srikanth Srinivasan. Certifying polynomials for ac^0(parity) circuits, with applications. In *FSTTCS*, volume 18 of *LIPIcs*, pages 36–47. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2012.

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *J. ACM*, 40(3):607–620, July 1993. URL: http://doi.acm.org/10.1145/174130.174138, doi:10.1145/174130.174138.

[LV18] Chin Ho Lee and Emanuele Viola. The coin problem for product tests. *TOCT*, 10(3):14:1–14:10, 2018.

[Nis91] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, Mar 1991. URL: https://doi.org/10.1007/BF01375474, doi:10.1007/BF01375474.

[NW94] Noam Nisan and Avi Wigderson. Hardness vs randomness. *J. Comput. Syst. Sci.*, 49(2):149–167, 1994.

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[OW07] Ryan O'Donnell and Karl Wimmer. Approximation by DNF: examples and counterexamples. In *ICALP*, volume 4596 of *Lecture Notes in Computer Science*, pages 195–206. Springer, 2007.

[Raz89] Alexander A. Razborov. On the method of approximations. In *STOC*, pages 167–176. ACM, 1989.

[Ros08a] Benjamin Rossman. On the constant-depth complexity of k-clique. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 721–730, 2008. URL: `http://doi.acm.org/10.1145/1374376.1374480`, `doi:10.1145/1374376.1374480`.

[Ros08b] Benjamin Rossman. On the constant-depth complexity of k-clique. In *STOC*, pages 721–730. ACM, 2008.

[Ros14] Benjamin Rossman. The monotone complexity of k-clique on random graphs. *SIAM J. Comput.*, 43(1):256–279, 2014.

[RS17] Benjamin Rossman and Srikanth Srinivasan. Separation of $\mathrm{AC}^0[\oplus]$ formulas and circuits. In *ICALP*, volume 80 of *LIPIcs*, pages 50:1–50:13. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017.

[RST15] Benjamin Rossman, Rocco A. Servedio, and Li-Yang Tan. An average-case depth hierarchy theorem for boolean circuits. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 1030–1048, 2015. URL: `https://doi.org/10.1109/FOCS.2015.67`, `doi:10.1109/FOCS.2015.67`.

[SHI65] Richard Edwin Stearns, Juris Hartmanis, and Philip M. Lewis II. Hierarchies of memory limited computations. In *6th Annual Symposium on Switching Circuit Theory and Logical Design, (FOCS) Ann Arbor, Michigan, USA, October 6-8, 1965*, pages 179–190, 1965. URL: `https://doi.org/10.1109/FOCS.1965.11`, `doi:10.1109/FOCS.1965.11`.

[Sho90] Victor Shoup. New algorithms for finding irreducible polynomials over finite fields. *Mathematics of Computation*, 54(189):435–447, 1990.

[Sip83] Michael Sipser. Borel sets and circuit complexity. In *Proceedings of the 15th Annual ACM Symposium on Theory of Computing, 25-27 April, 1983, Boston, Massachusetts, USA*, pages 61–69, 1983. URL: `https://doi.org/10.1145/800061.808733`, `doi:10.1145/800061.808733`.

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *STOC*, pages 77–82. ACM, 1987.

[Smo93] Roman Smolensky. On representations by low-degree polynomials. In *FOCS*, pages 130–138. IEEE Computer Society, 1993.

[Ste13] John P. Steinberger. The distinguishability of product distributions by read-once branching programs. In *IEEE Conference on Computational Complexity*, pages 248–254. IEEE Computer Society, 2013.

[SV10] Ronen Shaltiel and Emanuele Viola. Hardness amplification proofs require majority. *SIAM J. Comput.*, 39(7):3122–3154, 2010. URL: `https://doi.org/10.1137/080735096`, `doi:10.1137/080735096`.

[Sze89] M. Szegedy. Algebraic methods in lower bounds for computational models with limited communication. *PhD thesis*, The University of Chicago, 1989.

[Vio14] Emanuele Viola. Randomness buys depth for approximate counting. *Computational Complexity*, 23(3):479–508, 2014. URL: `https://doi.org/10.1007/s00037-013-0076-6`, `doi:10.1007/s00037-013-0076-6`.

[Vol99] Heribert Vollmer. *Introduction to Circuit Complexity - A Uniform Approach*. Texts in Theoretical Computer Science. An EATCS Series. Springer, 1999.

A Omitted Proofs from Section 3.3

Theorem 10. *Assume $d \geq 3$ and F_d is defined as in section 3.3. Then, for small enough δ, we have the following.*

1. *For $b, \beta \in \{0,1\}$ and each $i \in [d-1]$ such that $i \equiv \beta \pmod 2$, we have*

$$p_i^{(b)} = \Pr_{\boldsymbol{x} \sim \mu_b^{N_i}}[F_i(\boldsymbol{x}) = \beta].$$

   In particular, for any $i \in \{2, \ldots, d-2\}$ and any $b \in \{0,1\}$

$$p_i^{(b)} = (1 - p_{i-1}^{(b)})^{f_i}. \tag{44}$$

2. *For $\beta \in \{0,1\}$ and $i \in [d-2]$ such that $i \equiv \beta \pmod 2$, we have*

$$\frac{1}{2^m}(1 + \delta_i \exp(-3\delta_i)) \leq p_i^{(\beta)} \leq \frac{1}{2^m}(1 + \delta_i \exp(3\delta_i))$$
$$\frac{1}{2^m}(1 - \delta_i \exp(3\delta_i)) \leq p_i^{(1-\beta)} \leq \frac{1}{2^m}(1 - \delta_i \exp(-3\delta_i))$$

3. *Say $d - 1 \equiv \beta \pmod 2$. Then*

$$p_{d-1}^{(\beta)} \geq \exp(-C_1 m + C_2) \text{ and } p_{d-1}^{(1-\beta)} \leq \exp(-C_1 m - C_2)$$

   where $C_2 = C_1/10$.

4. *For each $b \in \{0,1\}$, $\Pr_{\boldsymbol{x} \sim \mu_b^N}[F_d(\boldsymbol{x}) = 1 - b] \leq 0.05$. In particular, F_d solves the δ-coin problem.*

Proof. **Proof of (1) (for $i \in [d-2]$) and (2):** We show these by induction on i. We start with the base case $i = 1$. Each formula at level 1 computes an AND of $N_1 = f_1 = m$ many variables. Hence we have:

$$\Pr_{D_0^{N_1}}[F_1(\boldsymbol{x}) = 1] = \left(\frac{1-\delta}{2}\right)^{f_1} = \left(\frac{1-\delta}{2}\right)^{m} \leq \frac{1}{2^m} \leq 0.5$$

$$\Pr_{D_1^{N_1}}[F_1(\boldsymbol{x}) = 1] = \left(\frac{1+\delta}{2}\right)^{f_1} = \left(\frac{1+\delta}{2}\right)^{m} \leq 0.5 \qquad \text{(for small enough } \delta\text{)}$$

This implies $p_1^{(b)} = \Pr_{D_{(b)}^{N_1}}[F(\boldsymbol{x}) = 1]$. For part (2):

$$\begin{aligned}
p_1^{(1)} &= \Pr_{D_1^m}[F_1(\boldsymbol{x}) = 1] = \left(\frac{1+\delta}{2}\right)^m \\
&= \frac{1}{2^m}(1+\delta)^m \leq \frac{1}{2^m}\exp(\delta m) && \text{By Fact 8 (c) and } \delta m = o(1) \\
&\leq \frac{1}{2^m}(1+\delta m + \delta^2 m^2) && \text{By Fact 8 (d)} \\
&= \frac{1}{2^m}(1+\delta m(1+\delta m)) \\
&\leq \frac{1}{2^m}(1+\delta m \exp(\delta m)) && \text{By Fact 8 (c)} \\
&= \frac{1}{2^m}(1+\delta_1 \exp(\delta_1)) \\
&\leq \frac{1}{2^m}(1+\delta_1 \exp(3\delta_1)) \\
p_1^{(1)} &= \frac{1}{2^m}(1+\delta)^m \geq \frac{1}{2^m}\exp(m(\delta-\delta^2)) && \text{By Fact 8 (b) with } x = \delta \\
&\geq \frac{1}{2^m}(1+(m(\delta-\delta^2))) && \text{By Fact 8 (c)} \\
&= \frac{1}{2^m}(1+(m\delta(1-\delta))) \geq \frac{1}{2^m}(1+(m\delta\exp(-2\delta))) && \text{By Fact 8 (a),(b)} \\
&\geq \frac{1}{2^m}(1+(\delta_1\exp(-3\delta_1)))
\end{aligned}$$

Similarly, we bound $p_1^{(0)}$:

$$\begin{aligned}
p_1^{(0)} &= \frac{1}{2^m}(1-\delta)^m \leq \frac{1}{2^m}\exp(-m\delta) && \text{By Fact 8 (c)}\\
&\leq \frac{1}{2^m}(1-m\delta(1-m\delta)) && \text{By Fact 8 (d)}\\
&\leq \frac{1}{2^m}(1-m\delta\exp(-3m\delta)) && \text{By Fact 8 (a),(b)}\\
&= \frac{1}{2^m}(1-\delta_1\exp(-3\delta_1))\\
p_1^{(0)} &= \frac{1}{2^m}(1-\delta)^m\\
&\geq \frac{1}{2^m}\exp(m(-\delta-\delta^2)) && \text{By Fact 8 (b)}\\
&\geq \frac{1}{2^m}(1-m\delta(1+\delta)) && \text{By Fact 8 (c)}\\
&\geq \frac{1}{2^m}(1-m\delta(1+3m\delta))\\
&\geq \frac{1}{2^m}(1-\delta_1\exp(3\delta_1)) && \text{By Fact 8 (c)}
\end{aligned}$$

We now show the inductive step of parts (1) and (2) for $p_{i+1}^{(b)}$. Since the circuit consists of alternating layers of OR gates and AND gates, we obtain $\forall i \in [d-2]$:

$$\Pr_{D_b^{N_i}}[F_i(\boldsymbol{x}) = \beta] = \left(1 - \Pr_{D_b^{N_{i-1}}}[F_{i-1}(\boldsymbol{x}) = (1-\beta)]\right)^{f_i}$$

Without loss of generality, assume $i \equiv 0 \pmod 2$. Then we have:

$$\begin{aligned}
\Pr_{\boldsymbol{x}\sim\mu_0^{N_{i+1}}}[F_{i+1}(\boldsymbol{x})=1] &= \left(1-\Pr_{\boldsymbol{x}\sim\mu_0^{N_i}}[F_i(\boldsymbol{x})=0]\right)^{f_{i+1}} \\
&= \left(1-p_i^{(0)}\right)^{f_{i+1}} && \text{From induction part (1)} \\
&\le \exp(-p_i^{(0)}\cdot f_{i+1}) && \text{By Fact 8 (c)} \\
&= \exp(-p_i^{(0)}\lceil 2^m m\ln 2\rceil) \\
&\le \exp(-p_i^{(0)}(2^m m\ln 2)) \\
&= \exp\left(-\left(\frac{1}{2^m}(1+\delta_i\exp(-3\delta_i))\right)2^m m\ln 2\right) && \text{From induction part(2)} \\
&\le \exp\left(-(m\ln 2)(1+\delta_i\exp(-3\delta_i))\right) \\
&\le \exp(-m\ln 2) = \frac{1}{2^m} \le 0.5
\end{aligned}$$

$$\begin{aligned}
\text{Similarly, } \Pr_{\boldsymbol{x}\sim\mu_1^{N_{i+1}}}[F_{i+1}(\boldsymbol{x})=1] &= (1-p_i^{(1)})^{f_{i+1}} \\
&\le \exp(-p_i^{(1)}\cdot f_{i+1}) = \exp(-p_i^{(1)}\lceil 2^m m\ln 2\rceil) && \text{By Fact 8 (c)} \\
&\le \exp\left(-\left(\frac{1}{2^m}(1-\delta_i\exp(3\delta_i))\right)(2^m m\ln 2)\right) \\
&\le \exp((-m\ln 2)(1-\delta_i(1+3\delta_i+9\delta_i^2))) && \text{By Fact 8 (d)} \\
&= 2^{-m(1-\delta_i(1+3\delta_i+9\delta_i^2))} \le 0.5 && \text{Since } \delta_i = o(1)
\end{aligned}$$

This completes the induction step for part (1). Now we show the bounds for part (2):

$$
\begin{aligned}
p_{i+1}^{(1)} &= (1-p_i^{(1)})^{\lceil 2^m m \ln 2 \rceil} \leq (1-p_i^{(1)})^{m2^m \ln 2} \\
&\leq \left(1 - \frac{1}{2^m}(1-\delta_i \exp(3\delta_i))\right)^{m2^m \ln 2} && \text{From induction hypothesis} \\
&\leq \exp\left(-\frac{1}{2^m}(1-\delta_i \exp(3\delta_i)) m 2^m \ln 2\right) && \text{By Fact 8 (c)} \\
&= \exp\left((-m \ln 2)(1-\delta_i \exp(3\delta_i))\right) \\
&= \exp\left(-m \ln 2 + \delta_i \exp(3\delta_i) m \ln 2\right) \\
&= \frac{1}{2^m} \exp(\delta_i \exp(3\delta_i) m \ln 2) \\
&= \frac{1}{2^m} \exp(\delta_{i+1} \exp(3\delta_i)) \\
&\leq \frac{1}{2^m}\left(1 + \delta_{i+1} \exp(3\delta_i)(1 + \delta_{i+1}\exp(3\delta_i))\right) && \text{By Fact 8 (d)} \\
&\leq \frac{1}{2^m}(1 + \delta_{i+1}\exp(3\delta_i)(1+2\delta_{i+1})) && \text{Since } \delta_i = o(1) \\
&\leq \frac{1}{2^m}(1 + \delta_{i+1}\exp(3\delta_i + \delta_{i+1})) && \text{By Fact 8 (c)} \\
&\leq \frac{1}{2^m}(1 + \delta_{i+1}\exp(3\delta_{i+1})) \\
p_{i+1}^{(1)} &= (1-p_i^{(1)})^{\lceil 2^m m \ln 2 \rceil} \\
&\geq (1-p_i^{(1)})^{2^m m \ln 2 + 1} \\
&\geq \left(1 - \frac{1}{2^m}(1-\delta_i \exp(-3\delta_i))\right)^{m2^m \ln 2 + 1} && \text{From induction} \\
&\geq \exp\left(\left(\frac{-1}{2^m}(1-\delta_i\exp(-3\delta_i)) - \frac{1}{2^{2m}}(1-\delta_i\exp(-3\delta_i))^2\right)(m2^m \ln 2 + 1))\right) && \text{By Fact 8 (b)} \\
&\geq \exp\left(\left(\frac{-1}{2^m}(1-\delta_i\exp(-3\delta_i)) - \frac{1}{2^{2m}}\right)(m2^m \ln 2 + 1)\right) \\
&\geq \frac{1}{2^m}\exp\left(\delta_i \exp(-3\delta_i) m \ln 2 - \frac{m}{2^m}\right) \\
&= \frac{1}{2^m}\exp\left(\delta_{i+1}\exp(-3\delta_i) - \frac{m}{2^m}\right) \\
&\geq \frac{1}{2^m}\exp\left(\delta_{i+1}\exp(-3\delta_i) - \delta_{i+1}^2\right) && \text{Using } \delta_{i+1} \geq \frac{1}{m} \\
&\geq \frac{1}{2^m}\exp\left(\delta_{i+1}(1-3\delta_i) - \delta_{i+1}^2\right) && \text{By Fact 8 (c)} \\
&\geq \frac{1}{2^m}\exp\left(\delta_{i+1}(1-2\delta_{i+1})\right) \\
&\geq \frac{1}{2^m}\exp\left(\delta_{i+1}\exp(-3\delta_{i+1})\right) && \text{By Fact 8 (a)(b)} \\
&\geq \frac{1}{2^m}\left(1 + \delta_{i+1}\exp(-3\delta_{i+1})\right) && \text{By Fact 8 (c)}
\end{aligned}
$$

The case of $p_{i+1}^{(0)}$ is similar and hence omitted.

Proof of (3): Assume, without loss of generality, $d-1 \equiv 1 \pmod 2$. Let $i = d-1$. Then we have:

$$\begin{aligned}
\delta_{i-1} &= \delta_{d-2} = \delta m(m(\ln 2))^{d-3} \\
&= \delta m^{d-2}(\ln 2)^{d-3} \\
\implies \delta_{i-1} m &= \delta m^{d-1}(\ln 2)^{d-3} \\
&= \delta \left(\left\lceil \left(\frac{1}{\delta}\right)^{1/(d-1)} \frac{1}{\ln 2} \right\rceil \right)^{d-1} (\ln 2)^{d-3} \\
&= \frac{1}{\ln 2^{d-1}} (\ln 2)^{d-3} \epsilon && \text{for some } \epsilon \in [1,2] \\
&= \frac{1}{(\ln 2)^2} \epsilon
\end{aligned}$$

With the above estimate for $\delta_{i-1}m$, we show the required bounds as follows. It follows exactly as in the proof of Part (1) for $i \in [d-2]$ that

$$p_i^{(b)} = \Pr_{\boldsymbol{x} \in \mu_b^{N_i}} [F_i(\boldsymbol{x}) = 1].$$

Hence, we have

$$\begin{aligned}
p_i^{(1)} &= (1-p_{i-1}^{(1)})^{f_{d-1}} \\
&= (1-p_{i-1}^{(1)})^{C_1 \cdot m2^m} \\
&\geq \left(1 - \frac{1}{2^m}(1-\delta_{i-1}\exp(-3\delta_{i-1}))\right)^{C_1 m 2^m} && \text{From part (2)} \\
&\geq \exp\left(\left(\frac{-1}{2^m}(1-\delta_{i-1}\exp(-3\delta_{i-1}))\right)\left(1+\frac{1}{2^m}(1-\delta_{i-1}\exp(-3\delta_{i-1}))\right)C_1 m 2^m\right) && \text{By Fact 8 (b)} \\
&\geq \exp\left(\frac{-1}{2^m}\left(1-\delta_{i-1}\exp(-3\delta_{i-1})+\frac{1}{2^m}\right)C_1 m 2^m\right) \\
&= \exp(-C_1 m)\exp\left(C_1\left(\delta_{i-1}m\exp(-3\delta_{i-1})-\frac{m}{2^m}\right)\right) \\
&\geq \exp(-C_1 m)\exp\left(\frac{C_1}{4(\ln 2)^2}\right) && \delta_{i-1}m \geq \frac{1}{(\ln 2)^2} \\
&\geq \exp(-C_1 m)\exp\left(C_2\right) && C_2 = C_1/10
\end{aligned}$$

The upper bound for $p_i^{(0)}$ is as follows:

$$\begin{aligned}
p_i^{(0)} &= (1-p_{i-1}^{(0)})^{C_1 m 2^m} \\
&\leq \left(1 - \frac{1}{2^m}(1+\delta_{i-1}\exp(-3\delta_{i-1}))\right)^{C_1 m 2^m} && \text{From part (2)} \\
&\leq \exp\left(\frac{-1}{2^m}(1+\delta_{i-1}\exp(-3\delta_{i-1}))C_1 m 2^m\right) && \text{From Fact 8 (c)} \\
&\leq \exp(-C_1 m)\exp(-C_1\delta_{i-1}m\exp(-3\delta_{i-1})) \\
&\leq \exp(-C_1 m)\exp(-C_1/4(\ln 2)^2) \\
&\leq \exp(-C_1 m)\exp(-C_2)
\end{aligned}$$

Proof of (4): Without loss of generality, assume $d \equiv 0 \pmod 2$. Then, the output gate of the circuit is an OR gate. Thus:

$$\begin{aligned}
\Pr_{\mu_1^{N_d}}[F_d(\boldsymbol{x}) = 0] &= (1-p_{d-1}^{(1)})^{f_d} \\
&\leq \exp(-p_{d-1}^1 \cdot f_d) && \text{From Fact 8 (c)} \\
&\leq \exp(-\exp(-C_1 m + C_2)\cdot\exp(C_1 m)) && \text{From part (3)} \\
&= \exp(-\exp(C_2)) \\
&= \frac{1}{e^{e^5}} \leq 0.05 \\
\text{Similarly, } \Pr_{D_0}[F_d(\boldsymbol{x}) = 1] &\leq f_d \cdot \Pr_{D_n}[F_{d-1}(\boldsymbol{x}) = 1] && \text{by union bound} \\
&\leq 2\exp(C_1 m)\exp(-C_1 m \cdot C_2) \\
&\leq 2\exp(-C_2) = \frac{2}{e^5} \\
&\leq 0.05
\end{aligned}$$

□

B Omitted Proofs from Section 3.4

Theorem 13 (Janson's inequality)**.** *Let $C_1, \ldots, C_M$ be any monotone Boolean circuits over inputs $x_1, \ldots, x_N$, and let C denote $\bigvee_{i\in[M]} C_i$. For each distinct $i, j \in [M]$, we use $i \sim j$ to denote the fact that $\mathrm{Vars}(C_i) \cap \mathrm{Vars}(C_j) \neq \emptyset$. Assume each $\boldsymbol{x}_j$ ($j \in [n]$) is chosen independently to be 1 with probability $p_i \in [0,1]$, and that under this distribution, we have $\max_{i\in[M]} \Pr_{\boldsymbol{x}}[C_i(\boldsymbol{x}) = 1] \leq 1/2$. Then, we have*

$$\prod_{i\in[M]} \Pr_{\boldsymbol{x}}[C_i(\boldsymbol{x}) = 0] \leq \Pr_{\boldsymbol{x}}[C(\boldsymbol{x}) = 0] \leq \left(\prod_{i\in[M]} \Pr_{\boldsymbol{x}}[C_i(\boldsymbol{x}) = 0] \right) \cdot \exp(2\Delta) \tag{45}$$

where $\Delta := \sum_{i<j:i\sim j} \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 1) \wedge (C_j(\boldsymbol{x}) = 1)]$.

We will use the following inequality in the proof of the above theorem.

Lemma 28 (Kleitman's inequality)**.** *Let $F, G : \{0,1\}^n \to \{0,1\}$ be two monotonically increasing Boolean functions or monotonically decreasing Boolean functions. Then,*

$$\Pr_{\boldsymbol{x}}[F(\boldsymbol{x}) = 1 | G(\boldsymbol{x}) = 0] \underset{\text{(i)}}{\leq} \Pr_{\boldsymbol{x}}[F(\boldsymbol{x}) = 1] \underset{\text{(ii)}}{\leq} \Pr_{\boldsymbol{x}}[F(\boldsymbol{x}) = 1 | G(\boldsymbol{x}) = 1]$$

Proof of Theorem 13. As $C(\boldsymbol{x})$ is an OR over $C_i(\boldsymbol{x})$ for $i \in [M]$, $\Pr_{\boldsymbol{x}}[C(\boldsymbol{x}) = 0] = \Pr_x[\forall i \in [M], (C_i(\boldsymbol{x}) = 0)]$. The lower bound on $\Pr_{\boldsymbol{x}}[C(\boldsymbol{x}) = 0]$ follows easily from Kleitman's inequality (Lemma 28) and induction on M.

Now we prove the upper bound on $\Pr_{\boldsymbol{x}}[C(\boldsymbol{x}) = 0]$. In order to prove the intended upper bound, we use the following intermediate lemma.

Lemma 29. *For all $i \in [M]$,*

$$\Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 1) \mid \forall j < i, (C_j(\boldsymbol{x}) = 0)] \geq \Pr_{\boldsymbol{x}}[C_i(\boldsymbol{x}) = 1] - \sum_{j:j<i, j\sim i} \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 1) \text{ AND } (C_j(\boldsymbol{x}) = 1)]$$

Assuming the above lemma, we will complete the proof of Theorem 13.

$$\begin{aligned}
\Pr_{\boldsymbol{x}}[C(\boldsymbol{x})=0] &= \Pr_{\boldsymbol{x}}[\forall i \in [M], (C_i(\boldsymbol{x})=0)] \\
&= \prod_{i\in[M]} \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x})=0) \mid \forall j<i, (C_j(\boldsymbol{x})=0)] \\
&= \prod_{i\in[M]} (1-\Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x})=1) \mid \forall j<i, (C_j(\boldsymbol{x})=0)]) \\
&\leq \prod_{i\in[M]} (1-\Pr_{x}[C_i(\boldsymbol{x})=1] - \sum_{j:j<i,j\sim i} \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x})=1) \text{ AND } (C_j(\boldsymbol{x})=1)]) \qquad (46) \\
&= \prod_{i\in[M]} (\Pr_{\boldsymbol{x}}[C_i(\boldsymbol{x})=0] - \sum_{j:j<i,j\sim i} \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x})=1) \text{ AND } (C_j(\boldsymbol{x})=1)]) \\
&= \prod_{i\in[M]} \left(\Pr_{\boldsymbol{x}}[C_i(\boldsymbol{x})=0] \cdot (1 + \frac{1}{\Pr_x[(C_i(\boldsymbol{x})=0)]} \sum_{j:j<i,j\sim i} \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x})=1) \text{ AND } (C_j(\boldsymbol{x})=1)] \right) \\
&\leq \prod_{i\in[M]} \left(\Pr_{\boldsymbol{x}}[C_i(\boldsymbol{x})=0] \cdot (1 + 2 \sum_{j:j<i,j\sim i} \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x})=1) \text{ AND } (C_j(\boldsymbol{x})=1)] \right) \qquad (47) \\
&\leq \prod_{i\in[M]} \left(\Pr_{\boldsymbol{x}}[C_i(\boldsymbol{x})=0] \cdot \exp(2 \sum_{j:j<i,j\sim i} \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x})=1) \text{ AND } (C_j(\boldsymbol{x})=1)]) \right) \qquad (48) \\
&= \left(\prod_{i\in[M]} \Pr_{\boldsymbol{x}}[C_i(\boldsymbol{x})=0] \right) \cdot \exp(2\Delta)
\end{aligned}$$

Inequality (46) follows from Lemma 29. Inequality (47) follows from the fact that $\Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 0)] \leq 1/2$ for each $i \in [M]$. Finally, (48) follows from (3). Therefore, assuming Lemma 29, we are done. We now prove this lemma.

Proof of Lemma 29. By reordindering the indices if required, assume that d is an index such that $d < i$ and for $1 \leq j \leq d$, $i \sim j$ and for $d < j < i$, $i \not\sim j$. Let $\mathcal{E}$ be the event that $[\forall j \leq d, (C_j(\boldsymbol{x}) = 0)]$

and $\mathcal{F}$ be the event that $[\forall d < j < i, (C_j(\boldsymbol{x}) = 0)]$.

$$
\begin{aligned}
&\Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 1) \mid \forall j < i, (C_j(\boldsymbol{x}) = 0)] \\
&= \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 1) \mid \mathcal{E} \text{ AND } \mathcal{F}] \\
&\geq \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 1) \text{ AND } \mathcal{E} \mid \mathcal{F}] && \text{(Bayes' Rule)} \\
&= \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 1) \mid \mathcal{F}] - \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 1) \text{ AND } \overline{\mathcal{E}} \mid \mathcal{F}] \\
&= \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 1)] - \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 1) \text{ AND } \exists j \leq d, (C_j(\boldsymbol{x}) = 1) \mid \mathcal{F}] && (C_i(\boldsymbol{x}) \text{ and } \mathcal{F} \text{ are independent}) \\
&= \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 1)] - \Pr_{\boldsymbol{x}}[\exists j \leq d, [(C_i(\boldsymbol{x}) = 1) \text{ AND } (C_j(\boldsymbol{x}) = 1)] \mid \mathcal{F}] \\
&\geq \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 1)] - \sum_{j \leq d} \Pr_{\boldsymbol{x}}[[(C_i(\boldsymbol{x}) = 1) \text{ AND } (C_j(\boldsymbol{x}) = 1)] \mid \mathcal{F}] && \text{(Union bound)} \\
&\geq \Pr_{\boldsymbol{x}}[(C_i(\boldsymbol{x}) = 1)] - \sum_{j \leq d} \Pr_{\boldsymbol{x}}[[(C_i(\boldsymbol{x}) = 1) \text{ AND } (C_j(\boldsymbol{x}) = 1)]] && \text{(Kleitman's inequality)}
\end{aligned}
$$

□

□

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il