Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 23 (2014)

Two Sides of the Coin Problem

Gil Cohen * Anat Ganor * Ran Raz †

February 25, 2014

Abstract

In the *coin problem*, one is given n independent flips of a coin that has bias $\beta > 0$ towards either Head or Tail. The goal is to decide which side the coin is biased towards, with high confidence. An optimal strategy for solving the coin problem is to apply the majority function on the n samples. This simple strategy works as long as $\beta > \Omega(1/\sqrt{n})$. However, computing majority is an impossible task for several natural computational models, such as bounded width read once branching programs and $\mathbf{AC^0}$ circuits.

Brody and Verbin [FOCS 2010] proved that a length n, width w read once branching program cannot solve the coin problem for $\beta < O(1/(\log n)^{3w})$. This result was tightened by Steinberger [CCC 2013] to $O(1/(\log n)^{w-2})$. The coin problem in the model of $\mathbf{AC^0}$ circuits was first studied by Shaltiel and Viola [STOC 2008], and later by Aaronson [STOC 2010] who proved that a depth d size s Boolean circuit cannot solve the coin problem for $\beta < O(1/(\log s)^{d+2})$.

This work has two contributions:

- We strengthen Steinberger result and show that any Santha-Vazirani source with bias $\beta < O(1/(\log n)^{w-2})$ fools length n, width w read once branching programs. In other words, the strong independence assumption in the coin problem is completely redundant in the model of read once branching programs. That is, the exact same result holds for a much more general class of sources.
- We tighten Aaronson result and show that a depth d, size s Boolean circuit cannot solve the coin problem for $\beta < O(1/(\log s)^{d-1})$. Moreover, our proof technique is different and we believe that it is simpler and more natural.

*Weizmann Institute of Science, Rehovot, Israel. {gil.cohen, anat.ganor}@weizmann.ac.il. Supported by an ISF grant and by the I-CORE Program of the Planning and Budgeting Committee.

†Weizmann Institute of Science, Rehovot, Israel, and the Institute for Advanced Study, Princeton, NJ. ran.raz@weizmann.ac.il. Supported by an ISF grant, by the I-CORE Program of the Planning and Budgeting Committee and by NSF grant numbers CCF-0832797, DMS-0835373.

ISSN 1433-8092

1 Introduction

In the *Coin Problem*, defined by Brody and Verbin [BV10], one is given n independent flips of a coin that has bias $\beta > 0$ towards either Head or Tail. The goal is to decide which side the coin is biased towards, with high confidence (say, 2/3). It is not hard to see that the best strategy for solving the coin problem is to apply the majority function on the n outcomes. By Chernoff bound, this strategy works as long as $\beta > c/\sqrt{n}$, for some large enough constant c. However, taking the majority on n bits is provably an impossible task for several natural computational models, such as bounded width read once branching programs (henceforth, ROBP) and $\mathbf{AC^0}$ circuits.

The coin problem is related to two other well-studied notions of approximating the majority function. The first notion is the "promise" problem of computing majority. Namely, it asks for upper and lower bounds, in different computational models, for computing the majority function correctly only on inputs that have bias at least ε. This central problem received a considerable attention in the literature (see [Ajt83], [ABO84], [Sto85], [Ajt93], [CR96], [Ama09], [Vio09], [Vio11], [KS12], [CDI+13] and references therein). In the second notion (see, e.g., [OW07], [Ama09]) one considers functions that agree with the majority function on all but δ fraction of the inputs (regardless of their Hamming weight). The coin problem can be seen as a combination of these two notions. Intuitively, it is an easier problem to solve as it allows both types of slackness, and thus poses a greater challenge for proving lower bounds.

Motivated by the construction of pseudorandom generators for ROBP, Brody and Verbin [BV10] considered the coin problem for the model of ROBP with bounded width. Informally speaking, a width w ROBP is a non-uniform model of computation that gets the flip outcomes one by one in a stream, and can "remember" at most $\log_2 w$ bits of information at each point in time, concerning the past outcomes. For, say, constant w, such model cannot compute majority.[1]

In [BV10] it is shown that width w length n ROBP cannot solve the coin problem for $\beta < 1/(\log n)^{cw}$, where c is some constant. This result was later tightened by Steinberger [Ste11] to $\beta < c/(2\log n)^{w-2}$, where c is an appropriate constant.

A different, yet essentially equivalent formulation of the coin problem, is where given a coin that is either unbiased or has bias at least $\beta > 0$ towards Head, the goal is to distinguish between the two cases, with high confidence. As mentioned above, Brody and Verbin [BV10], and later on Steinberger [Ste11], proved that a product distribution of bits, with small enough bias each, cannot be distinguished from the uniform distribution by ROBPs with bounded width. In other words, a product distribution with small enough bias *fools* ROBPs with bounded width.

Our first result shows that the independence assumption, which is necessary in many settings, is in fact completely redundant. More formally, we show that any Santha-Vazirani source, with small enough bias, fools ROBPs with bounded width. Recall that a Santha-

[1] In this context, it is interesting to note that the classical result of Barrington [Bar89] states that without the read-once requirement, width 5 is sufficient for computing majority.

Vazirani source [SV86] on n bits with bias β is a distribution $\mathbf{X}_1, \ldots, \mathbf{X}_n$, where each bit $\mathbf{X}_i$ is some adversarially chosen (probabilistic) function of $\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}$, under the promise that $\text{bias}(\mathbf{X}_i \mid \mathbf{X}_1 = x_1, \ldots, \mathbf{X}_{i-1} = x_{i-1}) \leq \beta$, for all prefixes $x_1, \ldots, x_{i-1}$.

Santha-Vazirani sources have been extensively studied in the pseudorandomness literature and form a much richer class of sources than product distributions. As mentioned, the original motivation of Brody and Verbin for studying the coin problem came from their approach of constructing pseudorandom generators for ROBPs. This approach yields pseudorandom generators for a natural subclass of ROBPs called *regular* ROBPs (see also, [BRRY10]). It is not clear how to construct pseudorandom generators for the non-regular case, and it is plausible that Santha-Vazirani sources are a much better starting point for such constructions. In fact, one can view the proof of Braverman *et al.* [BRRY10] as approximating Santha-Vazirani sources by recursively applying the pseudorandom generator of Impagliazzo *et al.* [INW94].

Theorem 1.1 (Santha-Vazirani sources fool ROBPs, informally stated)**.** *There exists a universal constant $c > 0$ such that the following holds. Any Santha-Vazirani source on n bits with bias $\beta < c/(2 \log n)^{w-2}$ fools length n, width w ROBPs.*

The proof of the above theorem is based on a reduction to the result of Steinberger [Ste11]. We move on to present our second result. As a matter of fact, the coin problem was studied by Shaltiel and Viola [SV10] and later by Aaronson [Aar10] even prior to the work of Brody and Verbin. The motivation for studying the coin problem in each of these papers was completely different. While Brody and Verbin were motivated by the study of pseudorandom generators for ROBPs, Shaltiel and Viola considered the problem of hardness amplification, and Aaronson studied the seemingly unrelated problem of obtaining an exponential oracle separation between **BQP** and **PH**. Aaronson ([Aar10], Corollary 12) proved that any depth d Boolean circuit on n inputs that distinguishes a fair coin from an ε-biased coin, with constant confidence, must have size exponential in $(1/\varepsilon)^{1/(d+2)}$. A similar result is implicit in [SV10].

The proof strategy of Shaltiel and Viola was to transform any circuit that distinguishes a fair coin from a coin with bias ε, to a circuit that computes majority on $\Omega(1/\varepsilon)$ inputs, to which standard lower bounds apply [Hås86, LMN93]. Taking a similar strategy, to prove his lower bound, Aaronson shows that a circuit that solves the coin problem can be transformed into a circuit that accepts all n bit strings with Hamming weight $n/2 + 1$ while rejecting all strings with Hamming weight $n/2$. Again, by [Hås86, LMN93], the latter task is known to require large bounded depth circuits. In the reduction, both papers make use of depth 3 circuits for the problem of approximate majority [Ajt83], [Vio09].

Our second contribution is an improvement over Aaronson's result. We give a tight lower bound for the size of a depth d circuit that solves the coin problem.

Theorem 1.2 (Coin Problem for $\mathbf{AC^0}$, informally stated)**.** *There exists a universal constant $c > 0$ such that the following holds. A depth d, size s Boolean circuit on n inputs cannot solve the coin problem for*

$$\beta < \frac{1}{(c \log s)^{d-1}}.$$

This is tight up to the multiplicative constant c [Ama09].

Moreover, our proof technique is different and we believe that it is simpler and more natural, and it gives the tight bound. The intuition is the following: suppose one applies, say, 10β random restriction to the input (see Section 2.3 for the precise definition). Then, one expects that a function f that solves the coin problem for bias β, applied to the resulting restricted input, should not be constant with high probability, whereas by known results [Hås86], [LMN93] such random restriction typically collapses a bounded depth circuit to some constant. The formal proof formalizes this by expressing the confidence of f in terms of the Fourier spectrum of a random restriction applied to f (see Lemma 4.2), where the restriction parameter is related to the bias of the coin.

In this context, it is interesting to mention the work of Viola [Vio11], who showed that a depth d, size s circuit can compute majority under the promise that the input has bias $\Omega(1/(\log s)^{d-3})$, and proved that this is tight. Moreover, Viola proved that a *randomized* circuit with depth d and size s can compute majority under the promise that the input has bias $\Omega(1/(\log s)^{d-1})$, and again, he proved that this bound is tight. We note however that the coin problem is an easier problem to solve, since the circuit may err on many inputs, and thus proving lower bounds is potentially more challenging.

2 Preliminaries

It will be convenient for us to think about coins with sides $\{\pm 1\}$. The bias of a $\{\pm 1\}$ random variable X, denoted by $\mathrm{bias}(X)$, is defined as $\frac{1}{2} \cdot |\mathbf{Pr}[X = 1] - \mathbf{Pr}[X = -1]|$.

Definition 2.1. *Let $\varepsilon \in [0, \frac{1}{2}]$. Define the product distribution $\mathbf{X}_\varepsilon^n$ supported on $\{\pm 1\}^n$ as follows. For $x \sim \mathbf{X}_\varepsilon^n$ it holds that $\mathbf{Pr}[x_i = 1] = \frac{1}{2} + \varepsilon$ (and thus $\mathbf{Pr}[x_i = -1] = \frac{1}{2} - \varepsilon$) for all $i \in [n]$.*

We note that the uniform distribution over $\{\pm 1\}^n$, denoted by $\mathbf{U}^n$, is the same as $\mathbf{X}_0^n$. When n is clear from context we omit the superscript and write $\mathbf{X}_\varepsilon$ and $\mathbf{U}$.

Definition 2.2 (Santha-Vazirani Sources)**.** *A distribution $\mathbf{X}$ supported on $\{\pm 1\}^n$ is called a* Santha-Vazirani source *with bias ε, if for every $i \in [n]$ and every $x_1, \ldots, x_n \in \{\pm 1\}$, it holds that*

$$\mathrm{bias}\left(\mathbf{X}_i \mid \mathbf{X}_1 = x_1, \mathbf{X}_2 = x_2, \ldots, \mathbf{X}_{i-1} = x_{i-1}\right) \leq \varepsilon.$$

Definition 2.3. *For a function $f \colon \{\pm 1\}^n \to \{\pm 1\}$ and a distribution $\mathbf{D}$ supported on $\{\pm 1\}^n$, the distinguishability of $\mathbf{D}$ from $\mathbf{U}$ by f is given by*

$$\mathsf{Distinguishability}(f, \mathbf{D}) = \left|\mathbb{E}[f(\mathbf{D})] - \mathbb{E}[f(\mathbf{U})]\right|.$$

For $\varepsilon \in [0, \frac{1}{2}]$, let $\mathsf{Distinguishability}(f, \varepsilon)$ denote $\mathsf{Distinguishability}(f, \mathbf{X}_\varepsilon)$.

2.1 Read Once Branching Programs

A *branching program* of length n and width w is a directed (multi-) graph with n layers $V_0, \ldots, V_{n-1}$ of w nodes each, called states, and a final layer V_n with two nodes, accept and reject. The branching program has a designated start node on layer V_0. For every internal node (that is, nodes in layers $V_0, \ldots, V_{n-1}$), there are exactly 2 edges going out of it and both these edges go to nodes on the next layer of the branching program. One of these edges is labeled by 1 and the other is labeled by -1. There are no edges going out of the accept and reject nodes. The computation of a branching program of length n on a string $x = x_1, \ldots, x_n \in \{\pm 1\}^n$ is defined in the natural way, by following the edge labeled x_i at step i, starting from the start node. The computation accepts x if it reaches the accept state and rejects otherwise. This branching program just described is a read-once branching program, since each character of x is examined exactly once.

For a branching program f, an internal state s and $b \in \{\pm 1\}$, let $s_f(b)$ denote the state reached by following the edge labeled b going out of s in f. When f is clear from context we write $s(b)$ instead of $s_f(b)$. For a string $x \in \{\pm 1\}^n$ we define the output of f to be $f(x) := -1$ if f accepts x and $f(x) := 1$ otherwise. For any state s, let $\mathcal{R}_{f,x}(s)$ denote the event that the computation of f on input x reaches s. When f and x are clear from context we write $\mathcal{R}(s)$ instead of $\mathcal{R}_{f,x}(s)$. Hence, for a distribution $\mathbf{X}$ over $\{\pm 1\}^n$, $\mathbb{E}[f(\mathbf{X}) \mid \mathcal{R}(s)]$ is the expected output of f on input $\mathbf{X}$, conditioned on the event that the computation reaches s, and $\mathbf{Pr}[\mathbf{X}_i = b \mid \mathcal{R}(s)]$ for $b \in \{\pm 1\}$ and $i \in [n]$, is the probability that $\mathbf{X}_i = b$ conditioned on the event that f on input $\mathbf{X}$ reaches s. To simplify notation, for any two states s_1, s_2, let $\mathcal{R}(s_1, s_2)$ denote the event that the computation of f on input x reaches both s_1 and s_2.

2.2 Bounded Depth Circuits

We consider circuits consisting of unbounded fan-in AND, OR gates applied to input variables and their negation. We only consider circuits with one output. The size of a circuit is the number of gates it contains. The depth is defined as the length of the longest path (in edges) from any input to the output. The depth of a gate g in a circuit is the depth of the sub-circuit with output gate g.

A circuit is called *layered* if for every $d \geq 1$, the inputs of any depth d gate in the circuit are the outputs of depth $d-1$ gates. A layered circuit is called *alternating* if the inputs to an AND gate (OR gate) with depth greater than 1 are the outputs of OR gates (AND gates). By standard arguments, for any size s, depth d Boolean circuit C there exists an alternating circuit C' with size at most $d \cdot s$ and depth d that computes the same function as C. The width of a layered circuit is the maximum fan-in of the gates in the bottom layer.

2.3 Fourier Analysis

Definition 2.4. *For a parameter $\rho \in [0,1]$ and $x \in \{\pm 1\}^n$, define the distribution $N_\rho(x)$ supported on $\{\pm 1\}^n$ as follows. For $y \sim N_\rho(x)$, for all $i \in [n]$ independently, with probability*

ρ the variable y_i is being set to x_i, and with probability $1-\rho$ the variable y_i is sampled uniformly at random from $\{\pm 1\}$.

Definition 2.5. *Let $\rho \in [0,1]$. The* noise operator T_ρ, *acting on the set of functions $\{f\colon \{\pm 1\}^n \to \mathbb{R}\}$, is defined as*

$$T_\rho f(x) = \mathbb{E}[f(N_\rho(x))].$$

The following well-known claim relates the Fourier representation of $T_\rho f$ to that of f.

Claim 2.6. *For any $f : \{\pm 1\}^n \to \mathbb{R}$ and $\rho \in [0,1]$,*

$$T_\rho f(x) = \sum_{S \subseteq [n]} \rho^{|S|} \widehat{f}(S) \chi_S(x).$$

For $\rho \in [0,1]$, a ρ *random restriction* is the following probabilistic process. For each $i \in [n]$, independently, leave it unset with probability ρ, and with probability $1-\rho$ set it to ± 1 uniformly and independently at random. We denote a restriction by $(J|z)$ where $J \subseteq [n]$ is the set of indices of unset variables and $z \in \{\pm 1, *\}^n$ is the values assigned to the variables, where the variables in J are assigned the symbol $*$. More precisely, $z_i = *$ if and only if $i \in J$, and otherwise z_i is the value assigned to the i^{th} variable.

Let $f\colon \{\pm 1\}^n \to \mathbb{R}$, and let $(J|z)$ be a restriction. We define the restricted function $f_{(J|z)}\colon \{\pm 1\}^n \to \mathbb{R}$ as follows: For $x \in \{\pm 1\}^n$, $f_{(J|z)}(x) = f(y)$, where $y \in \{\pm 1\}^n$ is defined as follows:

$$y_i = \begin{cases} x_i, & i \in J; \\ z_i, & i \notin J. \end{cases}$$

The following claim can be found in [O'D].

Claim 2.7. *Let $f : \{\pm 1\}^n \to \mathbb{R}$. Let $(J|z)$ be a ρ random restriction. Then for $S \subseteq [n]$,*

$$\mathbb{E}_{(J|z)}\left[\widehat{f_{(J|z)}}(S)\right] = \rho^{|S|} \cdot \widehat{f}(S).$$

3 Santha-Vazirani Sources Fool Read Once Branching Programs

The following is the main theorem of this section, which is a formal restatement of Theorem 1.1.

Theorem 3.1. *For any n, w such that $2 \le w \le \frac{\log n}{\log\log n}$ the following holds. Let f be a width w, length n ROBP. Then, for any $\varepsilon \in [0,1]$ and any Santha-Vazirani source $\mathbf{X}$ with bias ε,*

$$\mathsf{Distinguishability}\,(f, \mathbf{X}) \le \varepsilon \cdot (2\log n)^{w-2} \cdot (1 + o(1)).$$

The proof of Theorem 3.1 is via a reduction to the lower bound for ROBP solving the coin problem given by Brody and Verbin [BV10] and later improved by Steinberger [Ste11]. The following theorem is an adjustment of the lower bound of [Ste11] (see Theorem 1 therein) to our notation.

Theorem 3.2 ([Ste11]). *For any n, w such that $2 \leq w \leq \frac{\log n}{\log\log n}$ the following holds. Let f be a width w, length n ROBP. Then, for any $\varepsilon \in [0,1]$,*

$$\mathsf{Distinguishability}(f, \varepsilon) \leq \varepsilon \cdot (2\log n)^{w-2} \cdot (1 + o(1)).$$

The following lemma, which formalizes the reduction, together with Theorem 3.2 complete the proof of Theorem 3.1.

Lemma 3.3. *Let f be a width w, length n ROBP. Let $\mathbf{X}$ be a Santha-Vazirani source with bias ε. Then, there exists a width w, length n ROBP g such that*

$$\mathsf{Distinguishability}\,(f, \mathbf{X}) \leq \mathsf{Distinguishability}(g, \varepsilon). \tag{3.1}$$

Proof: Assume, without loss of generality, that $\mathbb{E}[f(\mathbf{U})] \leq \mathbb{E}[f(\mathbf{X})]$. Note that we may assume that for every layer $i \in [n]$ and every internal state s on V_{i-1}, it holds that

$$\mathbf{Pr}[\mathbf{X}_i = -1 \mid \mathcal{R}(s)] \leq \mathbf{Pr}[\mathbf{X}_i = 1 \mid \mathcal{R}(s)]. \tag{3.2}$$

If this is not the case, we flip the i^{th} coordinate of $\mathbf{X}$ in the event that f reaches s on input $\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}$. Note that $\mathbf{X}$ remains a Santha-Vazirani source with bias ε. We change f accordingly, by switching the edges going out of s in f. Doing so, the expected output of (the resulted) f, both under the uniform distribution and under (the resulted distribution) $\mathbf{X}$, does not change.

We define hybrid distributions $\mathbf{X}^{(n)}, \mathbf{X}^{(n-1)}, \ldots, \mathbf{X}^{(0)}$ as follows. For every $i \in [n+1]$, let $\mathbf{X}^{(i-1)}$ be a distribution where the first $i-1$ bits are distributed according to $\mathbf{X}$ and the rest of the bits are distributed according to $\mathbf{X}_\varepsilon$, independently of all other bits. Note that $\mathbf{X}^{(0)}$ is exactly $\mathbf{X}_\varepsilon$ and $\mathbf{X}^{(n)}$ is exactly $\mathbf{X}$. We define $f^{(n)} = f$ and given $f^{(i)}$ for some $i \in [n]$, we define $f^{(i-1)}$ as follows. Let $t_1, t_2, \ldots, t_{|V_i|}$ be an order of the states on layer V_i such that for every $1 \leq j < |V_i|$,

$$\mathbb{E}[f^{(i)}(\mathbf{X}_\varepsilon) \mid \mathcal{R}(t_j)] \geq \mathbb{E}[f^{(i)}(\mathbf{X}_\varepsilon) \mid \mathcal{R}(t_{j+1})].$$

We start with $f^{(i-1)} = f^{(i)}$. Fix some state s on layer V_{i-1} and let j_1, j_{-1} be the indices such that $s_{f^{(i-1)}}(1) = t_{j_1}$ and $s_{f^{(i-1)}}(-1) = t_{j_{-1}}$. If $j_1 > j_{-1}$ then we switch the edges going out of s in $f^{(i-1)}$. Clearly, the expected output of $f^{(i-1)}$ under the uniform distribution does not change. Moreover, since we change only edges that are going out of layer $i-1$, we get that

$$\mathbb{E}[f^{(i-1)}(\mathbf{X}_\varepsilon) \mid \mathcal{R}(s_{f^{(i-1)}}(-1))] \leq \mathbb{E}[f^{(i-1)}(\mathbf{X}_\varepsilon) \mid \mathcal{R}(s_{f^{(i-1)}}(1))]. \tag{3.3}$$

First, we analyze how the expectation under the distribution $\mathbf{X}^{(i)}$ changes when we switch from $f^{(i)}$ to $f^{(i-1)}$. By the definition of $\mathbf{X}^{(i)}$, for every $b \in \{\pm 1\}$ it holds that

$$\mathbb{E}[f^{(i-1)}(\mathbf{X}^{(i)}) \mid \mathcal{R}(s, s_{f^{(i-1)}}(b))] = \mathbb{E}[f^{(i-1)}(\mathbf{X}_\varepsilon) \mid \mathcal{R}(s_{f^{(i-1)}}(b))]$$

and

$$\mathbf{Pr}[\mathbf{X}_i^{(i)} = b \mid \mathcal{R}(s)] = \mathbf{Pr}[\mathbf{X}_i = b \mid \mathcal{R}(s)].$$

Therefore,

$$\begin{aligned}\mathbb{E}[f^{(i-1)}(\mathbf{X}^{(i)}) \mid \mathcal{R}(s)] &= \sum_{b\in\{\pm 1\}} \mathbb{E}[f^{(i-1)}(\mathbf{X}^{(i)}) \mid \mathcal{R}(s, s_{f^{(i-1)}}(b))] \cdot \mathbf{Pr}[\mathbf{X}_i^{(i)} = b \mid \mathcal{R}(s)] \\ &= \sum_{b\in\{\pm 1\}} \mathbb{E}[f^{(i-1)}(\mathbf{X}_\varepsilon) \mid \mathcal{R}(s_{f^{(i-1)}}(b))] \cdot \mathbf{Pr}[\mathbf{X}_i = b \mid \mathcal{R}(s)].\end{aligned}$$

In the same way,

$$\begin{aligned}\mathbb{E}[f^{(i)}(\mathbf{X}^{(i)}) \mid \mathcal{R}(s)] &= \sum_{b\in\{\pm 1\}} \mathbb{E}[f^{(i)}(\mathbf{X}_\varepsilon) \mid \mathcal{R}(s_{f^{(i)}}(b))] \cdot \mathbf{Pr}[\mathbf{X}_i = b \mid \mathcal{R}(s)] \\ &= \sum_{b\in\{\pm 1\}} \mathbb{E}[f^{(i-1)}(\mathbf{X}_\varepsilon) \mid \mathcal{R}(s_{f^{(i)}}(b))] \cdot \mathbf{Pr}[\mathbf{X}_i = b \mid \mathcal{R}(s)].\end{aligned}$$

When we switch from $f^{(i)}$ to $f^{(i-1)}$, we ensure that Equation (3.3) holds, and thus, assuming that Equation (3.2) also holds, the expectation can only increase. That is,

$$\mathbb{E}[f^{(i)}(\mathbf{X}^{(i)}) \mid \mathcal{R}(s)] \leq \mathbb{E}[f^{(i-1)}(\mathbf{X}^{(i)}) \mid \mathcal{R}(s)]. \tag{3.4}$$

Next, we analyze how the expectation of $f^{(i-1)}$ changes when we switch from $\mathbf{X}^{(i)}$ to $\mathbf{X}^{(i-1)}$. By the definition of $\mathbf{X}^{(i-1)}$, for every $b \in \{\pm 1\}$ it holds that

$$\mathbb{E}[f^{(i-1)}(\mathbf{X}^{(i-1)}) \mid \mathcal{R}(s, s_{f^{(i-1)}}(b))] = \mathbb{E}[f^{(i-1)}(\mathbf{X}_\varepsilon) \mid \mathcal{R}(s_{f^{(i-1)}}(b))]$$

and

$$\mathbf{Pr}[\mathbf{X}_i^{(i-1)} = b \mid \mathcal{R}(s)] = \mathbf{Pr}[(\mathbf{X}_\varepsilon)_i = b \mid \mathcal{R}(s)].$$

Therefore,

$$\mathbb{E}[f^{(i-1)}(\mathbf{X}^{(i-1)}) \mid \mathcal{R}(s)] = \sum_{b\in\{\pm 1\}} \mathbb{E}[f^{(i-1)}(\mathbf{X}_\varepsilon) \mid \mathcal{R}(s_{f^{(i-1)}}(b))] \cdot \mathbf{Pr}[(\mathbf{X}_\varepsilon)_i = b \mid \mathcal{R}(s)].$$

Since $\mathbf{Pr}[\mathbf{X}_i = 1 \mid \mathcal{R}(s)] \leq \frac{1}{2} + \varepsilon = \mathbf{Pr}[(\mathbf{X}_\varepsilon)_i = 1 \mid \mathcal{R}(s)]$, and since Equation (3.3) holds, when we switch from $\mathbf{X}^{(i)}$ to $\mathbf{X}^{(i-1)}$, the expectation can only increase. That is,

$$\mathbb{E}[f^{(i-1)}(\mathbf{X}^{(i)}) \mid \mathcal{R}(s)] \leq \mathbb{E}[f^{(i-1)}(\mathbf{X}^{(i-1)}) \mid \mathcal{R}(s)]. \tag{3.5}$$

Combining Equations (3.4) and (3.5), we get that

$$\mathbb{E}[f^{(i)}(\mathbf{X}^{(i)}) \mid \mathcal{R}(s)] \leq \mathbb{E}[f^{(i-1)}(\mathbf{X}^{(i-1)}) \mid \mathcal{R}(s)].$$

Finally, note that $\mathbf{Pr}_{\mathbf{X}^{(i-1)}}[\mathcal{R}(s)] = \mathbf{Pr}_{\mathbf{X}^{(i)}}[\mathcal{R}(s)]$. Therefore, by repeating the above arguments for every $s \in V_{i-1}$, and summing over them, we get that $\mathbb{E}[f^{(i)}(\mathbf{U})] = \mathbb{E}[f^{(i-1)}(\mathbf{U})]$ and $\mathbb{E}[f^{(i)}(\mathbf{X}^{(i)})] \leq \mathbb{E}[f^{(i-1)}(\mathbf{X}^{(i-1)})]$. Since this holds for every $i \in [n]$, we get that

$$\mathsf{Distinguishability}(f^{(n)}, \mathbf{X}^{(n)}) \leq \mathsf{Distinguishability}(f^{(0)}, \mathbf{X}^{(0)}),$$

as stated. □

4 The Coin Problem for $\mathbf{AC^0}$

The following is the main theorem of this section, which is a formal restatement of Theorem 1.2.

Theorem 4.1. *Let f be a function computable by a size s, depth d Boolean circuit. Then, for all $\delta \in (0, \frac{1}{2}]$*

$$\mathsf{Distinguishability}\left(f, \frac{\delta}{(120 \cdot \log(12s/\delta))^{d-1}}\right) \leq \delta.$$

We note that this result is tight [Ama09]. To prove Theorem 4.1, we start by proving a lemma that expresses the distinguishability of a function in terms of the behavior of the function under random restrictions.

Lemma 4.2. *Let $f : \{\pm 1\}^n \to \{\pm 1\}$ and let $\varepsilon \in [0, \frac{1}{2}]$. If $(J|z)$ is a 2ε random restriction then*

$$\mathsf{Distinguishability}(f, \varepsilon) = \left| \mathbb{E}_{(J|z)} \left[\sum_{\emptyset \neq S \subseteq [n]} \widehat{f_{(J|z)}}(S) \right] \right|.$$

Proof: We first note that the distributions $\mathbf{X}^n_\varepsilon$ and $N_{2\varepsilon}(1^n)$ are the same. Indeed, both are product distributions, and for any $x \sim \mathbf{X}^n_\varepsilon$ and $i \in [n]$, $\mathbf{Pr}[x_i = 1] = \frac{1}{2} + \varepsilon$ by definition. On the other hand, if $x \sim N_{2\varepsilon}(1^n)$ then

$$\mathbf{Pr}[x_i = 1] = 2\varepsilon \cdot 1 + (1 - 2\varepsilon) \cdot \frac{1}{2} = \frac{1}{2} + \varepsilon.$$

Thus

$$\mathbb{E}[f(\mathbf{X}^n_\varepsilon)] = \mathbb{E}[f(N_{2\varepsilon}(1^n))].$$

According to Definition 2.5, we can write the RHS of the above equation as

$$T_{2\varepsilon} f(1^n) = \sum_{S \subseteq [n]} (2\varepsilon)^{|S|} \widehat{f}(S) \chi_S(1^n) = \sum_{S \subseteq [n]} (2\varepsilon)^{|S|} \widehat{f}(S),$$

where the first equality follows by Claim 2.6. This, together with Claim 2.7, implies that for $(J|z)$, a 2ε random restriction, we have that

$$\mathbb{E}[f(\mathbf{X}^n_\varepsilon)] = \sum_{S \subseteq [n]} \mathbb{E}_{(J|z)}\left[\widehat{f_{(J|z)}}(S)\right] = \mathbb{E}_{(J|z)}\left[\sum_{S \subseteq [n]} \widehat{f_{(J|z)}}(S)\right].$$

On the other hand,

$$\mathbb{E}[f(\mathbf{U})] = \widehat{f}(\emptyset) = \mathbb{E}_{(J|z)}\left[\widehat{f_{(J|z)}}(\emptyset)\right],$$

where the last inequality follows by Claim 2.7. Thus,

$$\mathsf{Distinguishability}(f,\varepsilon) = |\mathbb{E}[f(\mathbf{X}_\varepsilon)] - \mathbb{E}[f(\mathbf{U})]| = \left|\mathbb{E}_{(J|z)}\left[\sum_{\emptyset\neq S\subseteq[n]} \widehat{f_{(J|z)}}(S)\right]\right|$$

as claimed. □

We also need the following well-known theorem, which is implicit in the result of [LMN93] (see also [O'D], Chapter 4). For completeness, we give a proof of this theorem in Appendix A.

Theorem 4.3. *For any $\delta \in (0,1)$ the following holds. Let $f\colon \{\pm1\}^n \to \{\pm1\}$ be a function computable by an alternating circuit with size s, depth $d \geq 3$ and width w. Let $\ell = \log(\frac{2s}{\delta})$ and let $\rho = \frac{1}{10w} \cdot \left(\frac{1}{10\ell}\right)^{d-3} \cdot \frac{\delta}{10\ell}$. If $(J|z)$ is a ρ random restriction then*

$$\Pr_{(J|z)}\left[f_{(J|z)} \text{ is non-constant}\right] \leq \delta.$$

Lastly, the proof of Theorem 4.1 makes use of the following lemma.

Lemma 4.4. *Let $\delta > 0$. Let f be a function computable by an alternating circuit with size s and depth d. Assume further that the bottom layer is an* AND *layer. Then, there exists a function g computable by an alternating circuit with size s, depth d and width $3\log(2s/\delta)$ such that for every $\varepsilon \leq 1/4$*

$$|\mathbb{E}_{x\sim\mathbf{X}_\varepsilon}[(f-g)(x)]| \leq \delta.$$

Proof of Lemma 4.4: Consider an alternating circuit with size s and depth d that computes f, with bottom layer consists of AND gates. By cutting all AND gates in the bottom layer with fan-in larger than $3\log(2s/\delta)$ we get a function g computable by an alternating circuit, consists of AND gates at the bottom layer, with size s, depth d and width $3\log(2s/\delta)$.

We note that $g^{-1}(1) \subseteq f^{-1}(1)$. On the other hand, consider a fan-in k AND gate that we cut. The probability, under $\mathbf{X}_\varepsilon$, that this AND gate outputs 1 is at most $(\frac{1}{2}+\varepsilon)^k$. Since we only cut AND gates with fan-in at least $3\log(2s/\delta)$

$$\left(\frac{1}{2}+\varepsilon\right)^k \leq \left(\frac{1}{2}+\varepsilon\right)^{3\log(2s/\delta)} \leq \frac{\delta}{2s},$$

where the last inequality follows by our assumption that $\varepsilon \leq \frac{1}{4}$ (which yields $\left(\frac{1}{2}+\frac{1}{4}\right)^3 < \frac{1}{2}$). Thus, by taking a union bound over all, at most s, AND gates with fan-in at least $3\log(2s/\delta)$ we get that

$$\Pr_{x\sim\mathbf{X}_\varepsilon}[f(x) \neq g(x)] \leq \frac{\delta}{2}.$$

Since f, g have range $\{\pm1\}$ the above equation implies that $|\mathbb{E}_{x\sim\mathbf{X}_\varepsilon}[(f-g)(x)]| \leq \delta$ as stated. □

Proof of Theorem 4.1: By the assumption of the theorem, there exists a size s, depth d circuit C that computes f. By standard arguments (see Section 2.2), there exists a size $d \cdot s$, depth d alternating circuit C' that computes f. We may assume, without loss of generality, that the bottom layer of C' consists of AND gates. If this is not the case then we can replace every OR gate at the bottom layer with an AND gate applied to the negation of the literals which are wired to the original OR gate. By DeMorgen, it follows that the output of this new AND gate is the negation of the output of the original OR gate. We can thus continue with this process, layer by layer from bottom to top, switching the type of gates in each layer. At the end of the process we get an alternating circuit, with size $d \cdot s$ and depth d, that computes the negation of f. Clearly, a function and its negation have the same distinguishability.

By Lemma 4.4, there exists a function g, computable by an alternating circuit with size $d \cdot s$, depth d and width $w = 3\log(12ds/\delta)$ such that

$$|\mathbb{E}_{x \sim \mathbf{X}_\varepsilon}[(f-g)(x)]| \leq \frac{\delta}{6} \tag{4.1}$$

for all $\varepsilon \leq 1/4$. Let $\ell = \log(12ds/\delta)$ and let $(J|z)$ be a $\rho = \frac{1}{30\ell} \cdot \left(\frac{1}{10\ell}\right)^{d-3} \cdot \frac{\delta}{60\ell}$ random restriction. Since g is computable by a width $w = 3\ell$ alternating circuit, Theorem 4.3 implies that

$$\mathop{\mathbf{Pr}}_{(J|z)}\left[g_{(J|z)} \text{ is non-constant}\right] \leq \frac{\delta}{6}. \tag{4.2}$$

By Lemma 4.2,

$$\textsf{Distinguishability}\left(g, \frac{\rho}{2}\right) = \left|\mathbb{E}_{(J|z)}\left[\sum_{\emptyset \neq S \subseteq [n]} \widehat{g_{(J|z)}}(S)\right]\right|.$$

In the event that $g_{(J|z)}$ is a constant function, the entire Fourier mass of $g_{(J|z)}$ lies in the empty coefficient, and in such case, the sum within the expectation in the above equation is 0. On the other hand, by Equation (4.2), $g_{(J|z)}$ is non-constant with probability at most $\delta/6$ and so,

$$\textsf{Distinguishability}\left(g, \frac{\rho}{2}\right) \leq \frac{\delta}{6} \cdot \left|\mathbb{E}_{(J|z)}\left[\sum_{\emptyset \neq S \subseteq [n]} \widehat{g_{(J|z)}}(S) \;\middle|\; g_{(J|z)} \text{ is non-constant}\right]\right|.$$

Note that

$$\sum_{\emptyset \neq S \subseteq [n]} \widehat{g_{(J|z)}}(S) = g_{(J|z)}(1^n) - \widehat{g_{(J|z)}}(\emptyset),$$

which is some number in $[-2, +2]$ as $g_{(J|z)}$ has range $\{\pm 1\}$. Thus,

$$\textsf{Distinguishability}\left(g, \frac{\rho}{2}\right) \leq 2 \cdot \frac{\delta}{6} = \frac{\delta}{3}.$$

The above equation together with Equation (4.1) implies that

$$\begin{aligned}\mathsf{Distinguishability}\left(f, \frac{\rho}{2}\right) &= \left|\mathbb{E}[f(\mathbf{X}_{\rho/2})] - \mathbb{E}[f(\mathbf{U})]\right| \\ &\leq \left|\mathbb{E}[f(\mathbf{X}_{\rho/2})] - \mathbb{E}[g(\mathbf{X}_{\rho/2})]\right| + \\ &\quad \left|\mathbb{E}[g(\mathbf{X}_{\rho/2})] - \mathbb{E}[g(\mathbf{U})]\right| + \\ &\quad |\mathbb{E}[g(\mathbf{U})] - \mathbb{E}[f(\mathbf{U})]| \\ &\leq \frac{\delta}{6} + \frac{\delta}{3} + \frac{\delta}{6} < \delta.\end{aligned}$$

Thus, by the choice of ρ we have

$$\mathsf{Distinguishability}\left(f, \frac{\delta}{\left(60 \cdot \log\left(12ds/\delta\right)\right)^{d-1}}\right) \leq \delta.$$

The proof then follows since $d \leq s$. □

References

[Aar10] S. Aaronson. BQP and the Polynomial Hierarchy. In *Proceedings of the 42nd ACM symposium on Theory of computing*, pages 141–150. ACM, 2010.

[ABO84] M. Ajtai and M. Ben-Or. A theorem on probabilistic constant depth computations. In *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, pages 471–474. ACM, 1984.

[Ajt83] M. Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24:1–48, 1983.

[Ajt93] M. Ajtai. *Approximate counting with uniform constant-depth circuits*, volume 13. Amer. Math. Soc. Providence, RI, 1993.

[Ama09] K. Amano. Bounds on the size of small depth circuits for approximating majority. In *Automata, Languages and Programming*, pages 59–70. Springer, 2009.

[Bar89] D. A. Barrington. Bounded-width polynomial-size branching programs recognize exactly those languages in NC^1. *Journal of Computer and System Sciences*, 38(1):150–164, 1989.

[BRRY10] M. Braverman, A. Rao, R. Raz, and A. Yehudayoff. Pseudorandom generators for regular branching programs. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 40–47. IEEE, 2010.

[BV10] J. Brody and E. Verbin. The coin problem and pseudorandomness for branching programs. In *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*, pages 30–39. IEEE, 2010.

[CDI+13] G. Cohen, I. B. Damgård, Y. Ishai, J. Kölker, P. B. Miltersen, R. Raz, and R. D. Rothblum. Efficient multiparty protocols via log-depth threshold formulae. In *Advances in Cryptology–CRYPTO 2013*, pages 185–202. Springer, 2013.

[CR96] S. Chaudhuri and J. Radhakrishnan. Deterministic restrictions in circuit complexity. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 30–36. ACM, 1996.

[Hås86] J. Håstad. Almost optimal lower bounds for small depth circuits. In *Proceedings of the 18th Annual STOC*, pages 6–20, 1986.

[INW94] R. Impagliazzo, N. Nisan, and A. Wigderson. Pseudorandomness for network algorithms. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 356–364. ACM, 1994.

[KS12] S. Kopparty and S. Srinivasan. Certifying polynomials for AC0 (parity) circuits, with applications. In *32nd International Conference on Foundations of Software Technology and Theoretical Computer Science*, page 36, 2012.

[LMN93] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform and learnability. *J. ACM*, 40(3):607–620, 1993.

[O'D] R. O'Donnell. Analyis of Boolean functions. http://analysisofbooleanfunctions.org/.

[OW07] R. O'Donnell and K. Wimmer. Approximation by DNF: examples and counterexamples. In *Automata, Languages and Programming*, pages 195–206. Springer, 2007.

[Ste11] J. Steinberger. The distinguishability of product distributions by read-once branching programs. 2011.

[Sto85] L. Stockmeyer. On approximation algorithms for #P. *SIAM Journal on Computing*, 14(4):849–861, 1985.

[SV86] M. Santha and U. V. Vazirani. Generating quasi-random sequences from semi-random sources. *Journal of Computer and System Sciences*, 33(1):75–87, 1986.

[SV10] R. Shaltiel and E. Viola. Hardness amplification proofs require majority. *SIAM Journal on Computing*, 39(7):3122–3154, 2010.

[Vio09] E. Viola. On approximate majority and probabilistic time. *computational complexity*, 18(3):337–375, 2009.

[Vio11] E. Viola. Randomness buys depth for approximate counting. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 230–239. IEEE, 2011.

A Proof of Theorem 4.3

The proof of Theorem 4.3 relies on Håstad Switching Lemma [Hås86] (see also [O'D], Chapter 4).

Lemma A.1 (Håstad switching lemma)**.** *Let $f : \{\pm 1\}^n \to \{\pm 1\}$ be a function computable by a width w DNF or width w CNF. Let $(J|z)$ be a ρ random restriction. Then, for every $k \in \mathbb{N}$ it holds that*

$$\mathbf{Pr}\left[\mathsf{DTdepth}(f_{(J|z)}) \geq k\right] \leq (5\rho w)^k.$$

Proof of Theorem 4.3: Let C be an alternating circuit with size s, depth d and width w that computes f. For $i = 1, \ldots, d$ denote the number of gates at level i by s_i (and so $s_1 + \cdots + s_d = s$ and $s_d = 1$). For $j = 1, \ldots, s_2$ denote by g_j the j^{th} gate at level 2, and denote by C_j the circuit with top gate g_j. Note that all circuits $\{C_j\}_j$ are either CNF or DNF with width w. Assume without loss of generality that they are DNF.

Let $\rho_1 = \frac{1}{10w}$. Consider a ρ_1 random restriction $(J_1|z_1)$. By Håstad switching lemma (Lemma A.1), for all $j \in [s_2]$,

$$\mathop{\mathbf{Pr}}_{(J_1|z_1)}\left[\mathsf{DTdepth}(C_j|_{(J_1|z_1)}) \geq \ell\right] \leq (5\rho_1 w)^\ell = 2^{-\ell},$$

and so, by union bound

$$\mathop{\mathbf{Pr}}_{(J_1|z_1)}\left[\exists j \in [s_2] \text{ such that } \mathsf{DTdepth}(C_j|_{(J_1|z_1)}) \geq \ell\right] \leq s_2 \cdot 2^{-\ell}.$$

Consider the event in which $\forall j \in [s_2]\ \mathsf{DTdepth}(C_j|_{(J_1|z_1)}) \leq \ell$. It is well known that if a function can be computed by a depth ℓ decision tree then it can be computed both by a width ℓ CNF and by a width ℓ DNF. We can therefore replace each C_j with a width ℓ CNF. Both the second and third layers in the resulted circuit consisting of AND gates. We can therefore collapse these two layers into one layer consists of s_3 AND gates. Denote by $g'_1, \ldots, g'_{s_3}$ the AND gates in the second layer of this new circuit. For $j = 1, \ldots, s_3$ denote by C'_j the width ℓ CNF with top gate g'_j.

Let $\rho_2 = \frac{1}{10\ell}$ and let $(J_2|z_2)$ be a ρ_2 random restriction. By Håstad switching lemma (Lemma A.1), for $j_3 = 1, \ldots, s_3$,

$$\mathop{\mathbf{Pr}}_{(J_2|z_2)}\left[\mathsf{DTdepth}(C'_j|_{(J_2|z_2)}) \geq \ell\right] \leq (5\rho_2 \ell)^\ell = 2^{-\ell},$$

and so, by union bound,

$$\mathop{\mathbf{Pr}}_{(J_2|z_2)}\left[\exists j \in [s_3] \text{ such that } \mathsf{DTdepth}(C'_j|_{(J_2|z_2)}) \geq \ell\right] \leq s_3 \cdot 2^{-\ell}.$$

We restrict ourselves again to the event in which $\forall j \in [s_3]\ \mathsf{DTdepth}(C'_j|_{(J_2|z_2)}) \leq \ell$, use this fact to replace all C'_j with a width ℓ DNF and collapse the new second and third layer. We continue performing ρ_2 random restrictions until we are left with a depth 2 circuit, that is, either with a CNF or a DNF. Note that we perform a total of $d - 3$ ρ_2 random restrictions

(on top of the first ρ_1 random restriction). Denote the composed random restriction by $(J'|z')$. Then, except with probability $s \cdot 2^{-\ell}$ we end up with a width ℓ depth 2 circuit C''. We restrict ourselves to the event in which C'' has width at most ℓ.

Let $\rho_3 = \frac{\delta}{10\ell}$. Consider a ρ_3 random restriction $(J_3|z_3)$. By Håstad switching lemma,

$$\mathop{\mathbf{Pr}}_{(J_3|z_3)}\left[\mathsf{DTdepth}(C''|_{(J_3|z_3)}) \geq 1\right] \leq 5\rho_3\ell = \frac{\delta}{2}.$$

Thus, if we denote by $(J|z)$ the $\rho_1\rho_2^{d-3}\rho_3$ composed random restriction over all the random process described above, then

$$\mathop{\mathbf{Pr}}_{(J|z)}\left[f_{(J|z)} \text{ is non-constant}\right] \leq s \cdot 2^{-\ell} + \frac{\delta}{2} \leq \delta,$$

where the last inequality follows by the choice of ℓ. □

ECCC ISSN 1433-8092

http://eccc.hpi-web.de