AC0 unpredictability

Emanuele Viola*

December 7, 2018

Abstract

We prove that for every distribution D on n bits with Shannon entropy $\geq n - a$ at most $O(2^d a \log^{d+1} g)/\gamma^5$ of the bits D_i can be predicted with advantage γ by an AC^0 circuit of size g and depth d that is a function of all the bits of D except D_i. This answers a question by Meir and Wigderson (2017) who proved a corresponding result for decision trees.

We also show that there are distributions D with entropy $\geq n - O(1)$ such that any subset of $O(n/\log n)$ bits of D on can be distinguished from uniform by a circuit of depth 2 and size $\mathrm{poly}(n)$. This separates the notions of predictability and distinguishability in this context.

A line of papers in the literature [EIRS01, Raz98, Unr07, SV10, DGK17, CDGS18, MW17, ST17, GSV18] proves that if a distribution D on n bits has Shannon entropy H close to n then D possesses several properties of the uniform distribution on n bits. For a discussion and comparison of these results we refer the reader to [GSV18]. In this paper we consider two such properties.

Predictability. Meir and Wigderson prove [MW17] that most coordinates cannot be *predicted* by shallow decision trees. We state their result next with a slightly optimized bound given soon after by Smal and Talebanfard [ST17].

Theorem 1. *[MW17, ST17] Let $D = (D_1, D_2, \ldots, D_n)$ be a distribution on n bits with $H(D) \geq n - a$. Let $t_1, t_2, \ldots, t_n$ be n decision trees of depth q, where t_i does not query D_i. Let $B := \{i \in [n] : \mathbb{P}_D[D_i = t_i(D)] \geq 1/2 + \gamma\}$. Then $|B| \leq 2aq/\gamma^2$.*

The bound in [MW17] is $|B| \leq O(aq/\gamma^3)$. Throughout this paper $O(.)$ and $\Omega(.)$ stand for absolute constants. The result in [MW17, ST17] applies to a stronger model that we think of as roughly the intersection of DNF and CNF. But it does not apply to DNF. Meir and Wigderson raised the question of proving a similar result for AC^0. We answer their question affirmatively in this paper.

*Supported by NSF CCF award 1813930. Work partially done while visiting and supported by the Simons institute.

Theorem 2. *Let $D = (D_1, D_2, \ldots, D_n)$ be a distribution on n bits with $H(D) \geq n - a$. Let $C_1, C_2, \ldots, C_n$ be n circuits on n bits, each of size g and depth d, where C_i does not depend on D_i. Let $B := \{i \in [n] : \mathbb{P}_D[D_i = C_i(D)] \geq 1/2 + \gamma\}$. Then $|B| \leq O(2^d a \log^{d+1} g)/\gamma^5$.*

It is noted in [ST17] that Theorem 1 is tight. In a tight example, the decision trees simply compute parities on $q + 1$ bits. Such parities can be computed by circuits of depth $\exp(q^{1/(d-1)})$. Hence the bound on $|B|$ in Theorem 2 is tight up to a factor of $\log^2(g)/\gamma^3$.

The proof of Theorem 2 is in Section 1.

Distinguishability. A result in [GSV18], stated next, shows that if we forbid to query a few bits, the distribution D is *indistinguishable* from uniform by small-depth decision trees. (This is called the *forbidden-set lemma* in [GSV18].)

Theorem 3. *[GSV18] Let D be a distribution on n bits with $H(D) \geq n - a$. For every γ, q there exists a set $B \subseteq [n]$ of size $O(aq^3/\gamma^3)$ such that for every decision tree t of depth q that does not make queries in B,*

$$|\mathbb{P}[t(U) = 1] - \mathbb{P}[t(D) = 1]| \leq \gamma.$$

Theorem 1 can be used to give an alternative proof of Theorem 3, see the discussion in [GSV18]. The other way around is not clear.

In the spirit of the previous result, we ask if Theorem 3 can be extended to constant-depth circuits. We give a negative answer.

Theorem 4. *For infinitely many n:*

There is a distribution D on n bits with $H(D) \geq n - O(1)$ such that for any set B of size $O(n/\log n)$ there is a read-once $O(\log n)$-DNF C with no variable in B such that

$$|\mathbb{P}[C(U) = 1] - \mathbb{P}[C(D) = 1]| \geq \Omega(1).$$

The proof of this theorem is in Section 2.

Whereas for the model of decision trees theorems 1 and 3 give similar bounds for predictability and distinguishability, theorems 2 and 4 give a strong separation between these notions for AC^0.

Given the negative result in Theorem 4 it is natural to ask if Theorem 3 can be extended in other ways. We note that it is possible to extend it to q-DNF, that is DNF with terms of size q. However the size of B now depends exponentially on q.

Theorem 5. *Let D be a distribution on n bits with $H(D) \geq n - a$. For every γ, q there exists a set $B \subseteq [n]$ of size $a2^{O(q)}/\gamma^{O(1)}$ such that for every q-DNF C that does not contain variables in B,*

$$|\mathbb{P}[C(U) = 1] - \mathbb{P}[C(D) = 1]| \leq \gamma.$$

The proof of this theorem is in Section 3.

One can use Theorem 4 to show that the exponential dependence on q in Theorem 5 is necessary. Given n and q, use Theorem 4 to obtain a distribution D' on $n' = 2^{\Theta(q)}$ bits with entropy $\geq n' - O(1)$ so that for any set B of size $O(n'/\log n')$ there is a q-DNF C with no variable in B such that

$$|\mathbb{P}[C(U) = 1] - \mathbb{P}[C(D') = 1]| \geq \Omega(1).$$

Let D be the distribution that equals D' on the first n' bits and is uniform on the other $n - n'$. The entropy of D is $n' - O(1) + n - n' \geq n - O(1)$, but for indistinguishability we have to exclude a set B of size $\geq \Omega(n'/\log n') = 2^{\Omega(q)}$.

The proofs use standard facts about entropy which can be found online or in the book [CT06]. In particular we use extensively the *chain rule* $H(X, Y) = H(X) + H(Y|X)$ for any random variables X and Y. We find it convenient to use the notation X for either the random variable or a fixed sample. The meaning is given by the context. If X is fixed the expression $H(Y|X)$ denotes the entropy of Y conditioned on the fixed outcome X. If X is not fixed it denotes the average over X of the entropy of Y conditioned on the fixed outcome X.

1 Proof of Theorem 2

The high-level idea is to perform some kind of *restriction* so that the circuits collapse to shallow decision trees and also a lot of entropy is preserved. If that happens we can use Theorem 1 to get a bound. However executing this plan is not straightforward.

High-entropy switching lemma. First we recall the switching lemma. It will be important for our results to use the latest analysis [Hås14].

Definition 6. A function $f : \{0,1\}^m \to \{0,1\}^n$ is computable by a *q'-partial common decision tree of depth q* if there is a (standard) decision tree of depth q such that on every input, the function f restricted along a path of this tree has the property that every output bit of f is computable by a decision tree of depth q'.

In other words, we can compute f with a decision tree of depth q that has at its leaves decision forests of depth q'.

A *restriction* on n bits is a subset of $\{0, 1, \star\}^n$ where the symbol $\star$ is called *star*. For an integer s the distribution R_s is obtained by picking uniformly a subset of size s for the stars and setting the other bits uniformly.

Lemma 7. *[Switching lemma] Let $C : \{0,1\}^n \to \{0,1\}^n$ be a circuit of size g and depth d with $g \geq n \geq d$. Let $R = R_s$ be a random restriction with $s = \Theta(n/\log^{d-1} g)$ stars. Except with error probability α over R, the circuit restricted to R can be computed by an $O(\log g)$-partial common decision tree of depth-$O(2^d \log(g/\alpha))$.*

Now we are ready for our switching-lemma for high-entropy distributions.

Definition 8. A D-restriction with s stars is obtained by picking the locations for the stars uniformly at random, and setting the other bits according to D.

Lemma 9. *In the same setting of Theorem 7, let R be a D-restriction, where $H(D) \geq n-a$. Then the error bound is* $(1+a)/\log(1/\alpha)$.

For σ a subset of $[n]$ we write D_σ for the $|\sigma|$ bits of D corresponding to D, and $D_{\bar{\sigma}}$ for the others.

Proof. Let A be the set of all possible restrictions with s $\star$. We have $|A| = \binom{n}{s}2^{n-s}$. Let H be the set of restrictions that don't collapse the circuits in the sense of Lemma 7. By the same lemma, $|H|/|A| \leq \alpha$.

R is a distribution over A. We shall show that it lands in H with small probability. Write R as $(S, D_{\bar{S}})$, where S is the subset of the $\star$, and $D_{\bar{S}}$ is the projection of D outside of S. We have

$$H(R) = H(S, D_{\bar{S}}) = H(S) + H(D_{\bar{S}}|S) \geq \log_2 \binom{n}{s} + n - a - s.$$

In the inequality we use that for every fixed S, the distribution $D_{\bar{S}}$ is over $n-s$ variables and we have $H(D) = H(D_S, D_{\bar{S}}) = H(D_{\bar{S}}) + H(D_S|D_{\bar{S}})$. The latter term is at most s. And so we have $H(D_{\bar{S}}) \geq H(D) - s \geq n - a - s$.

Thus the entropy of R is only a away from the maximum entropy $m := \log_2 \binom{n}{s} + n - s$ of any distribution over A.

Let p be the probability that $R \in H$. Let E be the indicator random variable of the event $R \in H$. We have

$$\begin{aligned} m - a \leq H(R) &= H(R,E) = H(R|E) + H(E) \leq H(R|E) + 1 \\ &= pH(R|E=1) + (1-p)H(R|E=0) + 1 \leq p\log_2|H| + (1-p)m + 1. \\ &\leq p\log\alpha + pm + (1-p)m + 1. \end{aligned}$$

Hence $p\log(1/\alpha) \leq 1 + a$, and the result follows. □

We apply Lemma 9 with $\alpha := 2^{-200a/\gamma}$. This gives an $O(\log g)$-partial common tree of depth $q = O(2^d(\log g + a/\gamma))$ and an error bound of 0.01γ.

High-entropy after restrictions. We need to show that after the restriction the entropy is still large. First note $H(D|R) \geq s - a$, indeed this holds for any fixed choice for the positions S for the stars. To verify this note that, for any fixed S,

$$n - a \leq H(D) = H(D_S, D_{\bar{S}}) \leq H(D_{\bar{S}}) + H(D_S|D_{\bar{S}}) = H(R) + H(D|R) \leq n - s + H(D|R).$$

Applying Markov's inequality to $\mathbb{E}_R[s - H(D|R)] = s - H(D|R) \leq a$, where note the argument inside the expectation is non-negative, we obtain $\mathbb{P}_R[s - H(D|R) \geq a/\epsilon] \leq \epsilon$ for any ϵ. Setting $\epsilon = 0.01\gamma$ we obtain that with probability $\geq 1 - 0.01\gamma$ over R, $H(D|R) \geq s - O(a/\gamma)$.

Intersecting B. We argue that $|S \bigcap B| \geq 0.5(s|B|/n) = \Omega(|B|/\log^{d-1} g)$ with high probability. This quantity is the hypergeometric distribution of the number of red balls sampled without replacement from a set of n balls $|B|$ of which are red. The expected number of red balls is sp where $p := |B|/n$. The probability of sampling less than half of that is at most (see Section 4 in [Hoe63])

$$2^{-\mathcal{D}(0.5p|p)s} \leq 2^{-\Omega(ps)} \leq 2^{\Omega(|B|/\log^{d-1} g)}$$

where $\mathcal{D}$ is divergence. The upper bound is at most 1/1000 (else the theorem is vacuously true).

Fixing restrictions. Call a fixed restriction R *good* if both $H(D|R) \geq s - O(a/\gamma)$ and every circuit collapses to an $O(\log g)$-partial common depth-q tree. By above and a union bound, the probability that R is not good is $\leq 0.01\gamma + 0.01\gamma \leq \gamma/10$. Writing R as $(S, D_{\bar{S}})$ we conclude that

$$\mathbb{P}_S[\mathbb{P}_{D_{\bar{S}}}[R \text{ bad}] \geq \gamma/2] \leq 1/5,$$

because otherwise the probability of being bad is $> (1/5)(\gamma/2) = \gamma/10$, contradicting the previous fact.

Combining this with the bound on intersecting B we obtain that there exists a fixed S such that

(1) $\mathbb{P}_{D_{\bar{S}}}[R \text{ bad}] \leq \gamma/2$,

(2) $|S \bigcap B| \geq \Omega(|B|/\log^{d-1} g)$.

Now, for this fixed S, let $L := S \bigcap B$. Because $L \subseteq B$, we have by assumption

$$1/2 + \gamma \leq \mathbb{P}_{i \in L}[D_i = C_i(D)] \leq \mathbb{P}_{i \in L}[D_i = C_i(D)|R \text{ good}] + \mathbb{P}[R \text{ bad}].$$

So $\mathbb{P}_{i \in L}[D_i = C_i(D)|R \text{ good}] \geq 1/2 + \gamma - \gamma/2 \geq 1/2 + \gamma/2$. Fix a good restriction R for which this holds. (Note S was fixed already, so we are just fixing $D_{\bar{S}}$.) Project the resulting distribution onto S and call it X. We have $H(X) \geq s - O(a/\gamma)$, the circuit is computable by a $O(\log g)$-partial common depth-q tree, and moreover there is a set L of size $\geq \Omega(|B|/\log^{d-1} g)$ such that $\mathbb{P}_{i \in L}[X_i = C_i(X)] \geq 1/2 + \gamma/2$.

Handling the common part. Now we need to handle the common part of the decision tree. We need to fix the variables along a path so that both the entropy and the prediction is preserved. Let t be the common decision tree. We think of sampling X by first sampling the q bits Y along a path, and then sampling the other $s - q$ bits Z, in a fixed order. We want to show that $H(Z|Y)$ is large. Indeed,

$$s - O(a/\gamma) = H(X) = H(Y, Z) = H(Z|Y) + H(Y) \leq H(Z|Y) + q.$$

The second equality can be verified by noting that X is a function of (Y, Z) and (Y, Z) is a function of X. Rearranging and using our bound on q we get $s - H(Z|Y) \leq q + O(a/\gamma) = O(q)$. By a Markov argument, the probability over Y that $s - H(Z|Y) \geq O(q/\gamma)$ is at most $\gamma/4$. Call such a Y *bad.* Like before, we have

$$1/2 + \gamma/2 \leq \mathbb{P}_{i \in L}[X_i = C_i(X)] \leq \mathbb{P}_{i \in L}[X_i = C_i(X)|Y \text{ good}] + \mathbb{P}[Y \text{ bad}].$$

Hence $\mathbb{P}_{i \in L}[X_i = C_i(X)|Y \text{ good}] \geq 1/2 + \gamma/4$. Fix a good Y such that this holds, and call the resulting distribution V. We have $H(V) \geq s - O(q/\gamma)$,

$$\mathbb{P}_{i \in L}[V_i = C_i(V)] \geq 1/2 + \gamma/4, \tag{1}$$

and now each C_i is a decision tree of depth $O(\log g)$.

Finishing up. By Theorem 1 the number of the s coordinates of V that can be $(1/2+\gamma/8)$-predicted is at most twice the entropy deficiency $O(q/\gamma)$ times the depth of the tree $O(\log g)$, divided by $O(1/\gamma)^2$. This equals

$$O(q/\gamma^3) \log g. \tag{2}$$

Hence we have

$$\mathbb{P}_{i \in L}[V_i = C_i(V)] \leq O(q/\gamma^3) \log(g)/|L| + 1/2 + \gamma/8.$$

Combining equations 1 and 2 we obtain

$$O(q/\gamma^3) \log g/|L| \geq \gamma/8.$$

Now recall $q = O(2^d(\log g + a/\gamma))$. Hence we can crudely bound $O(q/\gamma^3) \log g$ above by $O(2^d \log^2(g) a/\gamma^4)$. Also recall $|L| \geq \Omega(|B|/\log^{d-1} g)$. Hence we get

$$O(2^d \log^{d+1}(g) a/|B|\gamma^4) \geq \gamma/8.$$

This concludes the proof.

1.1 Proof of Lemma 7

We denote by R_p the standard distribution on restrictions where the bits are independent and each comes up $1, 0, \star$ with probabilities $(1-p)/2, (1-p)/2, p$.

Lemma 10. *[Lemma 3.8 in [Hås14] with $s := 1 + \log S$] Let $f : \{0,1\}^n \to \{0,1\}^S$ be a function computable by a depth-2 circuit with input fan-in r. Then the probability over R_p that f restricted to R_p cannot be computed by a $(1 + \log S)$-partial common depth-q decision tree is at most $S(24pr)^q$.*

The straightforward corollary we need is not stated anywhere.

Corollary 11. *Let $C : \{0,1\}^n \to \{0,1\}^n$ be a circuit of size g and depth d with $g \geq n \geq d$. Let $p = \Theta(1/\log^{d-1} g)$. With probability $1 - \alpha$ over R_p the circuit restricted to R_p can be computed by a $(1 + \log n)$-partial common depth-$O(2^d \log(g/\alpha))$ decision tree.*

Proof. First we take a restriction with $p = \Omega(1)$, and apply Lemma 10 to the g_1 gates at level 1 (viewed as a DNF or CNF with input fan-in 1). For a parameter q_0, with probability $1 - g_1 2^{-q_0}$ we can compute f by a common decision tree of depth q_0 at the leaves of which we have circuits of depth d whose number of gates at levels ≥ 2 hasn't changed, and whose input fan-in is $O(\log g)$.

Then we take a restriction with $p = \Omega(1/\log g)$, and apply Lemma 10 to the g_2 gates at level 2. We take a union bound over all 2^{q_0} paths of the common decision tree just discussed. For a parameter q_1, with probability $1 - 2^{q_0} g_2 2^{-q_1}$we can compute f by a decision tree of depth $q_0 + q_1$ at the leaves of which we have circuits of depth $d-2$ whose inputs are decision trees of depth $O(\log g)$. We can write the latter trees as CNF or DNF as appropriate and merge them with the next layer of gates. Hence we can compute f by a decision tree of depth $q_0 + q_1$ at the leaves of which we have circuits of depth $d-1$ with input fan-in $O(\log g)$. The number of gates at the higher levels hasn't changed.

We continue in this fashion. In the end, we can compute f by a tree of depth $q_0 + q_1 + \cdots + q_{d-1}$ whose leaves are forests of depth $O(\log g)$. The error probability is $g_1 2^{-q_0} + 2^{q_0} g_2 2^{-q_1} + 2^{q_0+q_1} g_3 2^{-q_2} + \cdots$. Picking $q_i = t \cdot 2^i$ this is at most $g \cdot d \cdot 2^{-t}$.

So for error α we should take $t = \log(1/\alpha) + \log(g) + \log(d) \leq O(\log g/\alpha)$. This gives a common tree of depth $O(\log g/\alpha)2^d$ whose leaves are forests of depth $O(\log g)$. □

To conclude the proof of Lemma 7 we only need to verify that the same result holds if we take a restriction with exactly $s = np \star$. Indeed, the probability that R_p has exactly s stars is $\geq \Omega(1/\sqrt{s}) \geq \Omega(1/g)$. So if we set the error probability to $O(\alpha/g)$ in Corollary 11 we obtain an error probability of α for restrictions with exactly s stars, and the depth of the tree hasn't changed asymptotically.

2 Proof of Theorem 4

Let $n = m(\log_2 m + 1)$ and think of the n bits as divided in m blocks of $(\log_2 m + 1)$ bits each. The distribution D is sampled as follows. First select $I \in \{1, 2, \ldots, m\}$ uniformly. Set the I block to all zero. Then for every other block independently, set the block to a uniform value *excluding* all zero. We can write D as (I, X) where X are non-zero values for $m-1$ blocks.

We have

$$\begin{aligned} H(D) =& H(I,X) = H(I) + H(X) = \log_2 m + (m-1)\log_2(2m-1) \\ =& \log_2 m + (m-1)\log_2(2m) + (m-1)\log_2(1 - 1/2m) \\ \geq& m \log_2(2m) - O(1). \end{aligned}$$

The set B intersects $\leq |B|$ of the blocks. Let G be the other blocks. Consider the function C that outputs 1 if any of the blocks in G is all zero. This function can be written as a read-once DNF with terms of size $\log_2 m + 1$.

Under the uniform distribution, the probability that C equals 1 is at most $m/2^{\log_2 m+1} = 1/2$.

Under D it is at least the probability that $I \in G$, which is $\geq (m-|B|)/m$. So if $|B| \leq m/3$ the DNF C distinguishes. The result follows because $m \geq \Omega(n/\log n)$.

3 Proof of Theorem 5

We rely on a simulation of DNF by *decision trees*, showing that a q-DNF can be written as a tree of depth about 2^q, which may output "?" with small probability. A weaker version of the result was proved by Ajtai and Wigderson [AW89]. The stronger version, stated next, is due to Trevisan [Tre04].

Lemma 12. *For every q-DNF C there exists a decision tree t_C of depth $\le 2q2^q\log(1/\epsilon)$ with range $\{0,1,?\}$ such that*
(1) for every input x, $t_C(x) \neq ? \Rightarrow t_C(x) = C(x)$, and
(2) $\mathbb{P}[t_C(U) = ?] \le \epsilon$.

Proof. A *covering* of the terms is a set of variables such that any term contains a variable from the set, possibly negated. We define $t_C : \{0,1\}^n \to \{0,1,?\}$ recursively as follows. If C is a constant then t_C is the same constant. If C has $\ge 2^q\log(1/\epsilon)$ disjoint terms, then t_C queries the first $2^q\log(1/\epsilon)$ of them. If any term is True, t_C outputs 1, else it outputs ?. Otherwise, there exists a covering of the terms of size $\le q2^q\log(1/\epsilon)$. The tree t_C first queries this covering, and then recursively queries the resulting $(q-1)$-DNF.

The tree t_C has depth $\le q2^q\log(1/\epsilon) + (q-1)2^{q-1}\log(1/\epsilon) + \ldots \le 2q2^q\log(1/\epsilon)$.

Item (1) follows by definition.

To verify Item (2), note that the only case in which t_C outputs ? is that none of $\ge 2^q\log(1/\epsilon)$ disjoint terms is True. This happens with probability at most

$$(1-1/2^q)^{2^q\log(1/\epsilon)} \le (1/e)^{\log(1/\epsilon)} \le \epsilon.$$

□

As a corollary, any distribution which fools decision trees of depth about 2^q also fools q-DNF. We say that a distribution D ϵ-fools a class of functions F if for every $f \in F$ we have $|\mathbb{P}[f(D)=1] - \mathbb{P}[f(U)=1]| \le \epsilon$, where U is the uniform distribution.

Corollary 13. *Let D be a distribution that ϵ-fools decision trees of depth $2q2^q\log(1/\epsilon)$. Then D $O(\epsilon)$-fools q-DNF.*

Proof. For a q-DNF C let t_C be the tree from Lemma 12. By its properties we have, for every distribution X:

$$\mathbb{P}[t_C(X)=1] \le \mathbb{P}[C(X)=1] \le \mathbb{P}[t_C(X)=1] + \mathbb{P}[t_C(X)=?].$$

Writing down this fact for both $X = D$ and $X = U$ we have

$$\mathbb{P}[t_C(U)=1] \le \mathbb{P}[C(U)=1] \le \mathbb{P}[t_C(U)=1] + \mathbb{P}[t_C(U)=?],$$
$$\mathbb{P}[t_C(D)=1] \le \mathbb{P}[C(D)=1] \le \mathbb{P}[t_C(D)=1] + \mathbb{P}[t_C(D)=?].$$

By assumption, the left-hand sides are within ϵ, and so are the rightmost terms. Moreover, $\mathbb{P}[t_C(U) = ?] \le \epsilon$. Hence $\mathbb{P}[C(X) = 1]$ for both $X = D$ and $X = U$ lies in the interval $[\mathbb{P}[t_C(U)=1] - \epsilon, \mathbb{P}[t_C(U)=1] + 3\epsilon]$ and so they are within $O(\epsilon)$. □

Combining Corollary 13 with Theorem 3 we immediately obtain Theorem 5.

Acknowledgment. We thank Chin Ho Lee for pointing out that Lemma 12 appears in [Tre04]. We also thank Or Meir and Li-Yang Tan for helpful discussions.

References

[AW89] Miklos Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constant-depth circuits. *Advances in Computing Research - Randomness and Computation*, 5:199–223, 1989.

[CDGS18] Sandro Coretti, Yevgeniy Dodis, Siyao Guo, and John Steinberger. Random oracles and non-uniformity. In *Int. Conf. on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, 2018.

[CT06] Thomas Cover and Joy Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[DGK17] Yevgeniy Dodis, Siyao Guo, and Jonathan Katz. Fixing cracks in the concrete: Random oracles with auxiliary input, revisited. In *Int. Conf. on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, pages 473–495, 2017.

[EIRS01] Jeff Edmonds, Russell Impagliazzo, Steven Rudich, and Jiri Sgall. Communication complexity towards lower bounds on circuit depth. *Computational Complexity*, 10(3):210–246, 2001.

[GSV18] Aryeh Grinberg, Ronen Shaltiel, and Emanuele Viola. Indistinguishability by adaptive procedures with advice, and lower bounds on hardness amplification proofs. In *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2018. Available at http://www.ccs.neu.edu/home/viola/.

[Hås14] Johan Håstad. On the correlation of parity and small-depth circuits. *SIAM J. on Computing*, 43(5):1699–1708, 2014.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[MW17] Or Meir and Avi Wigderson. Prediction from partial information and hindsight, with application to circuit lower bounds. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:149, 2017.

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM J. on Computing*, 27(3):763–803, 1998.

[ST17] Alexander Smal and Navid Talebanfard. Prediction from partial information and hindsight, an alternative proof. *Electronic Colloquium on Computational Complexity (ECCC)*, 24:191, 2017.

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

[Tre04] Luca Trevisan. A note on approximate counting for k-DNF. In *7th Workshop on Randomization and Computation (RANDOM)*, volume 3122 of *Lecture Notes in Computer Science*, pages 417–426. Springer, 2004.

[Unr07] Dominique Unruh. Random oracles and auxiliary input. In *Int. Cryptology Conf. (CRYPTO)*, pages 205–223, 2007.