2018 IEEE 59th Annual Symposium on Foundations of Computer Science

Indistinguishability by adaptive procedures with advice, and lower bounds on hardness amplification proofs

Aryeh Grinberg
Dept. of computer science
University of Haifa
Haifa, Israel
akir94@gmail.com

Ronen Shaltiel
Dept. of Computer Science
University of Haifa
Haifa, Israel
ronen@cs.haifa.ac.il

Emanuele Viola
College of Computer and Information Science
Northeastern University
Boston, MA, USA
viola@ccs.neu.edu

***Abstract*—We study how well can q-query decision trees distinguish between the following two distributions: (i) $R = (R_1, \ldots, R_N)$ that are i.i.d. indicator random variables, (ii) $X = (R|R \in A)$ where A is an event s.t. $\Pr[R \in A] \geq 2^{-a}$. We prove two lemmas:**

- **Forbidden-set lemma: There exists $B \subseteq [N]$ of size poly$(a, q, \frac{1}{\eta})$ such that q-query trees that do not query variables in B cannot distinguish X from R with advantage η.**
- **Fixed-set lemma: There exists $B \subseteq [N]$ of size poly$(a, q, \frac{1}{\eta})$ and $v \in \{0,1\}^B$ such that q-query trees do not distinguish $(X|X_B = v)$ from $(R|R_B = v)$ with advantage η.**

The first can be seen as an extension of past work by Edmonds, Impagliazzo, Rudich and Sgall (Computational Complexity 2001), Raz (SICOMP 1998), and Shaltiel and Viola (SICOMP 2010) to *adaptive* decision trees. It is independent of recent work by Meir and Wigderson (ECCC 2017) bounding the number of $i \in [N]$ for which there exists a q-query tree that predicts X_i from the other bits.

We use the second, fixed-set lemma to prove lower bounds on black-box proofs for hardness amplification that amplify hardness from δ to $\frac{1}{2} - \epsilon$. Specifically:

- **Reductions must make $q = \Omega(\log(1/\delta)/\epsilon^2)$ queries, implying a "size loss factor" of q. We also prove the lower bound $q = \Omega(\log(1/\delta)/\epsilon)$ for "error-less" hardness amplification proofs, and for direct-product lemmas. These bounds are tight.**
- **Reductions can be used to compute Majority on $\Omega(1/\epsilon)$ bits, implying that black box proofs cannot amplify hardness of functions that are hard against constant depth circuits (unless they are allowed to use Majority gates).**

Both items extend to pseudorandom-generator constructions.

These results prove 15-year-old conjectures by Viola, and improve on three incomparable previous works (Shaltiel and Viola, SICOMP 2010; Gutfreund and Rothblum, RANDOM 2008; Artemenko and Shaltiel, Computational Complexity 2014).

***Keywords*-Hardness amplification; Decision trees; Black-box impossibility**

I. Introduction

In this paper we develop tools to bound the ability of adaptive procedures that make few queries to distinguish between an i.i.d. distribution on which the procedure "receives small advice" and the original i.i.d. distribution. We then use these tools to prove tight lower bounds on hardness amplification proofs.

A. Adaptive procedures that receive small advice

Let $R = (R_1, \ldots, R_N)$ be a collection of i.i.d. indicator random variables. Suppose that "a bits of advice" are given about R. That is, let $A \subseteq \{0,1\}^N$ be an event with $\Pr[R \in A] \geq 2^{-a}$, and let $X = (R|R \in A)$. We can think of X as "the way that R appears to an adversary that received a bits of information on R". We are interested in the following question:

How well can a decision tree making q queries distinguish between X and R?

For simplicity, let us focus on the case where each bit R_i is uniformly distributed. (In some sense, made precise later, this is w.l.o.g.). It is instructive to consider the following two examples:

- *Bad bits:* Consider $A = \{r : r_1 = 0\}$, then $\Pr[R \in A] = 2^{-a}$ for $a = 1$, and R_1 is a random coin, whereas X_1 is the constant zero. In other words, R_1 and X_1 are far in statistical distance. Consequently, there exists a 1-query decision tree that distinguishes R from X with large advantage. (By advantage, we refer to the difference between the probabilities that the decision tree accepts in the two experiments).
- *Pointer chasing:* Let $N = \ell + 2^\ell$, and for $r \in \{0,1\}^N$, we write $r = (r^1, r^2)$ where $|r^1| = \ell$ and $|r^2| = 2^\ell$. We can interpret $r^1 \in \{0,1\}^\ell$ as a number $r^1 \in [2^\ell]$ and consider $A = \left\{r : r^2_{r^1} = 0\right\}$. Namely, that the bit that r^1 "points to" in r^2 is fixed to zero. Once again, $\Pr[R \in A] = 2^{-a}$ for

2575-8454/18/$31.00 ©2018 IEEE
DOI 10.1109/FOCS.2018.00094

$a = 1$. Note that an $(\ell+1)$-query decision tree P, that queries the bits of r^1 and "follows the pointer" to decide which query to ask in r^2, distinguishes R from X with large advantage.

The two examples point out a distinction between two types of "local procedures": Adaptive procedures are general decision trees, that can decide on their next query based on the answers to past queries. We say that a q-query decision tree P is *nonadaptive*, if there exists a set $Q \subseteq [N]$ of size q, and a function f_P such that $P(x) = f_P(x_Q)$ (namely, if all "computation paths" of P query the variables in S in some order).

It is interesting to note that a q-query nonadaptive decision tree cannot substantially distinguish R from X in the case of "pointer chasing" even if q approaches 2^ℓ, whereas an adaptive $(\ell+1)$-query decision tree can.

1) Settings in which X and R are indistinguishable by shallow decision trees: We would like to identify settings in which we can argue that adaptive/nonadaptive q-query decision trees cannot substantially distinguish X from R. Given the two examples above, we need additional constraints. We discuss two settings, the second of which is introduced in this paper.

- *Forbidden sets:* Here one forbids decision trees from querying variables in a certain *small* "forbidden set" $B \subseteq [N]$. One shows that decision trees that do not make queries in B cannot substantially distinguish R from $X = (R|R \in A)$.
  Loosely speaking, this means that except for a few "damaged variables" one can assume that X is composed of i.i.d. variables (at least from the point of view of a shallow decision tree).
- *Fixed sets:* Here one fixes the variables in a certain *small* "fixed set" $B \subseteq [N]$ to some value v, and considers the conditional distributions $R' = (R|R_B = v)$ and $X' = (X|X_B = v) = (R|R_B = v, R \in A)$. One shows that no decision tree can substantially distinguish X' from R', even if the tree queries variables in B.
  Loosely speaking, this means that we can "get rid" of correlations between bits of X (at least from the point of view of a shallow decision tree) if we are willing to fix few "damaged variables".

In both cases, given integer parameters N, q, a, and an event A such that $\Pr[R \in A] \geq 2^{-a}$, we will want that B is of size $b = \text{poly}(a, q, \frac{1}{\eta})$ where $\eta > 0$ is measuring the required statistical distance. (It can be easily observed by extending the "bad bits example" that we cannot expect $b = o(\frac{a \cdot q}{\eta})$).

2) Past work on nonadaptive procedures: A well-known lemma states that if we have N i.i.d. random variables and we condition on an event that has not too small probability, then most variables are still close to uniform.

Lemma I.1. *Let N, a be integers. Let $R = (R_1, \ldots, R_N)$ be i.i.d. indicator random variables, let $A \subseteq \{0,1\}^N$ be an event such that $\Pr[R \in A] \geq 2^{-a}$, and let $X = (R|R \in A)$. For every $\eta > 0$, there exists a set $B \subseteq [N]$ of size $O(a/\eta^2)$, such that for every $i \in [N] \setminus B$, R_i and X_i are η-close.*[1]

Lemma I.1 has found many applications in a wide variety of contexts, see for example the 12 references in [1]. In our terminology, Lemma I.1 is a forbidden set lemma for $q = 1$. An extension of Lemma I.1 to $q > 1$ was given by Shaltiel and Viola [2], and is stated next.

Lemma I.2 ([2])**.** *Let N, a, q be integers. Let $R = (R_1, \ldots, R_N)$ be i.i.d. indicator random variables, let $A \subseteq \{0,1\}^N$ be an event such that $\Pr[R \in A] \geq 2^{-a}$, and let $X = (R|R \in A)$. For every $\eta > 0$, there exists a set $B \subseteq [N]$ of size $O(a \cdot q/\eta^2)$, such that for every $Q \subseteq [N] \setminus B$ of size q, R_Q and X_Q are η-close.*

Lemma I.2 and so in particular Lemma I.1 can already be obtained from the techniques in the 1991 paper [3]. A proof will also be given later in this paper. In our terminology, Lemma I.2 is a forbidden set lemma for *nonadaptive* q-query decision trees. Namely, it says that for every *nonadaptive* q-query decision tree P that does not make queries in B, P does not distinguish X and R with advantage η.

The extension of Lemma I.1 to larger q given by Lemma I.2 has also found several applications. The application in [2] concerns the complexity of *hardness amplification proofs*. This application is also a main motivation for this work and is discussed in detail below in Section I-B. Lemma I.2 has also found application in *data-structure lower bounds*, see [4, 5].

A significant shortcoming of Lemma I.2 is that it only applies to *non-adaptive* decision trees. As a consequence, several applications of this result are also proved only in the non-adaptive setting. For example, the bounds in [2] on the complexity of hardness amplification proofs only apply to non-adaptive procedures. Although some of the available hardness amplification proofs are indeed non-adaptive, others are not. This point is discussed further in Section I-B2.

In the area of data structures, some of the lower bounds obtained using Lemma I.2 were later generalized

[1] Here, distance is statistical distance, namely, two distributions are η-close if the probabilities they assign to every event differ by at most η.

to the adaptive setting [4, 6]. However, in some cases this generalization is not yet available. For example, [5] proves a non-adaptive lower bound for the *matching brackets* problem, and this is not known in the *adaptive* setting.

Therefore, it will be desirable to extend Lemma I.2 to handle general, *adaptive* decision trees. In this paper we obtain such an extension. In fact, we shall prove two extensions. The first is closest to the setting of Lemma I.2 and gives a "forbidden set". However when applying this version, several technical difficulties arise because of adaptive procedures that may query the forbidden set in complicated ways. So we prove a "fixed-set" extension, where these difficulties disappear. We then discuss our main application to hardness amplification proofs. We believe that the results in this paper will also be useful in the study of data structures.

3) A forbidden set lemma for adaptive decision trees: In this paper we prove a forbidden set lemma for *adaptive* decision trees.

Lemma I.3 (Forbidden set lemma for adaptive decision trees)**.** *Let N, a, q be integers. Let $R = (R_1, \ldots, R_N)$ be i.i.d. indicator random variables, let $A \subseteq \{0,1\}^N$ be an event such that $\Pr[R \in A] \geq 2^{-a}$, and let $X = (R|R \in A)$. For every $\eta > 0$, there exists a set $B \subseteq [N]$ of size $O(\frac{a \cdot q^3}{\eta^3}) = poly(a, q, \frac{1}{\eta})$, such that for every q-query decision tree P that does not make queries in B, $P(R)$ and $P(X)$ are η-close.*

In the full versionwe also prove an extension of this lemma where the tree may query variables in B with small probability.

It is illustrative to note that in the case of the "pointer chasing" example, the forbidden set lemma must put (many of) the "pointer bits" in B (as otherwise an adaptive decision tree can distinguish by querying the "pointer bits").

Lemma I.3 is independent of recent work by Meir and Wigderson, see Corollary 1.5 in [1], and the follow-up work [7]. They prove a result similar to Lemma I.3 but for unpredictability rather than indistinguishability. Specifically, they show that every variable X_i except those in a set B_{MW} of $O(aq/\epsilon^3)$ variables has the following property: no adaptive decision tree making q queries to other variables can predict X_i with advantage more than ϵ over random guessing. We elaborate more on the difference between the works in the full version.

4) A fixed set lemma for adaptive decision trees: Forbidden set lemmas have the drawback that they guarantee nothing in case the decision tree does make queries in B. This is unavoidable, as can be seen by the "bad bits" and "pointer chasing" examples. In this paper we propose the idea of *fixed set lemmas* where X is further conditioned to the distribution X' by fixing the variables in some set B. A corresponding conditioning is applied to R to obtain R', whose bits are independent, $|B|$ of them are fixed, and the rest are unaltered. Then, even decision trees that make queries in B cannot distinguish R' and X'. We prove the following fixed set lemma for *adaptive* decision trees.

Lemma I.4 (Fixed set lemma for adaptive decision trees)**.** *Let N, a, q be integers. Let $R = (R_1, \ldots, R_N)$ be i.i.d. indicator random variables, let $A \subseteq \{0,1\}^N$ be an event such that $\Pr[R \in A] \geq 2^{-a}$, and let $X = (R|R \in A)$. For every $\eta > 0$, there exists a set $B \subseteq [N]$ of size $\leq O(a \cdot q/\eta^2)$, and $v \in \{0,1\}^B$, such that for $R' = (R|R_B = v)$ and $X' = (X|X_B = v) = (R|R_B = v, R \in A)$, and every q-query decision tree P, $P(R')$ and $P(X')$ are η-close.*

In the full version we explain how to reduce the size bound on B in Lemma I.4 from $O(a \cdot q/\eta^2)$ to $O(a \cdot q/\eta)$.

This loosely means that if the application allows us to fix few bits of X, then we need not worry about trees that make queries in B.

We note that one can't derive a fixed set lemma from a forbidden set lemma by fixing the bits in B. For example, consider the "pointer chasing" example. A forbidden set lemma may choose B to be the "pointer bits" (namely $B = \{1, 2, \ldots, \ell\}$). However, if these bits get fixed, then the bit they point to is also fixed and not in B, and thus a 1-query decision tree can distinguish. By using pointer chasing with several layers, this example can be extended to show that even after many applications of a nonadaptive lemma, and fixing the bad bits, one does not obtain a fixed set lemma.

Remark I.5. *Stefano Tessaro pointed out the similarity between Lemma I.4 and results in cryptography related to* random oracles with auxiliary input*: Theorem 2 in [8] (cf. Theorem 1 in [9]) and Lemma 1 in [10]. These results seem incomparable to Lemma I.4. Lemma 1 in [10] roughly proves that the distribution X in Lemma I.4 is indistinguishable by shallow decision trees from a* convex combination *of distributions of the form $R|R_B = v$. Instead, in Lemma I.4 we show that the distribution $X|X_B = v$ is indistinguishable from* the single *distribution $R|R_B = v$. Another difference is that actually Lemma 1 in [10] is not stated for X but rather for an "average conditioning" of R, modeled as an adversary who is given $f(R)$, for a function f with bounded output length, and can make few queries to R.*

B. Hardness amplification

Hardness amplification is a technique to transform "hard functions" into "harder functions". It is closely related to error-correcting codes and plays a fundamental role in complexity theory, derandomization, and cryptography. We give a brief survey of the aspects that are most relevant to this work. For additional background we refer the reader to Chapter 17 "Hardness Amplification and Error Correcting Codes" in the textbook [11], and to the discussion in [2].

Hardness amplification results in the complexity theoretic literature have the following black-box form: Given a function f, there is a "construction map" that produces a function Con_f. The proof provides a "reduction" Red, that converts an "adversary" D that "breaks" the (strong hardness) of Con_f, into an "adversary" C that "breaks" the (weaker hardness) of the original function. A precise definition follows:

Definition I.6 (Black-box hardness amplification)**.** [2] *We say that two functions h_1, h_2 on the same finite domain W, p-agree if $\Pr_{X \leftarrow W}[h_1(X) = h_2(X)] \geq p$.*

A $\delta \rightarrow (\frac{1}{2} - \epsilon)$ black-box hardness amplification with input lengths k and n, and list size 2^a is a pair $(\mathsf{Con}, \mathsf{Red})$ such that:

- *A* construction Con *is a map from functions $f : \{0,1\}^k \rightarrow \{0,1\}$ to functions $\mathsf{Con}_f : \{0,1\}^n \rightarrow \{0,1\}$.*
- *A* reduction Red *is an oracle circuit $\mathsf{Red}^{(\cdot)}(x, \alpha)$ that accepts two inputs: $x \in \{0,1\}^k$ and $\alpha \in \{0,1\}^a$ (we call α a "nonuniform advice string"). Red also receives oracle access to a function $D : \{0,1\}^n \rightarrow \{0,1\}$.*

We require that for all functions $f : \{0,1\}^k \rightarrow \{0,1\}$ and $D : \{0,1\}^n \rightarrow \{0,1\}$ such that D $(\frac{1}{2} + \epsilon)$-agrees with Con_f, there exists $\alpha \in \{0,1\}^a$ such that $\mathsf{Red}^D(., \alpha)$ $(1 - \delta)$-agrees with f.

Note that for if $\delta < 2^{-k}$ then it follows that $\mathsf{Red}^D(., \alpha)$ 1-agrees with f, which means $\mathsf{Red}^D(., \alpha) = f$.

A reduction (or proof) is nonadaptive *if the queries Red makes to the oracle are nonadaptive.*

The following specific "construction maps" are extensively studied in the literature:

- *Yao's XOR Lemma:* Let $n = t \cdot k$ for a parameter t and $\mathsf{Con}_f(x_1, \ldots, x_t) = f(x_1) \oplus \ldots \oplus f(x_t)$. This is the "mother" of all hardness amplifications, dating back to oral presentations by Yao in the 80's, cf. [12]. To give an example setting of parameters, one can start with δ constant and then the hardness parameter ϵ decays exponentially with t.
- *Direct product Lemma:* Let $n = t \cdot k$ and $\mathsf{Con}_f(x_1, \ldots, x_t) = (f(x_1) \circ \ldots \circ f(x_t))$. Such results are called "Concatenation Lemma" or "Direct product lemma". Note that here, Con_f is nonboolean and outputs t bits, and consequently $\frac{1}{2} + \epsilon$ should be replaced with $2^{-t} + \epsilon$.
- *Local list-decodable codes:* Let $K = 2^k$, $N = 2^n$ and $\mathsf{Con}_f(y) = \mathsf{Enc}(f)_y$ where $\mathsf{Enc} : \{0,1\}^K \rightarrow \{0,1\}^N$ is an encoding map for a binary error-correcting code, and we view the function $f : \{0,1\}^k \rightarrow \{0,1\}$ as a K-bit string that is the truth table of f. In fact, for $\delta = 0$, it is necessary that Enc is a "locally decodable, list-decodable code", and Red is a "local list decoding algorithm". The connection between such codes and hardness amplification is established in [13], where it is also said that is was observed independently by other researchers (see [13]).

We note that black-box hardness amplification indeed lets us amplify hardness. In short, this is because if there exists a small circuit D that $(\frac{1}{2} + \epsilon)$-agrees with Con_f then the reduction gives a procedure of the form $\mathsf{Red}^D(., \alpha)$ which $(1 - \delta)$-agrees with f. Now, if $\mathsf{Red}^D(., \alpha)$ is also a small circuit then this gives a larger circuit $C(.) = \mathsf{Red}^D(., \alpha)$ that $(1 - \delta)$-agrees with f. If the latter is impossible, because we started with a function f that is sufficiently hard, we have reached a contradiction, which means that D does not exist.

It is evident from this argument that to obtain hardness against a circuit class $\mathcal{D}$ one needs to start from hardness against the larger circuit class $\mathsf{Red}^{\mathcal{D}}$. The complexity of this class depends on the complexity of Red and on the number of queries that Red is allowed to make. Thus a natural question, and a main focus of this paper, is what is the complexity of Red, and how many queries are required.

Red requires TC^0: Fifteen years ago Viola [14, 15] made several conjectures regarding the complexity of hardness amplification. Informally, he conjectured that the smallest circuit class that can prove hardness amplification with $\epsilon = 1/n$ is the class TC^0 of constant-depth circuits with majority gates. This is problematic because (1) lower bounds against TC^0 are a long-standing open problem in circuit complexity, (2) the "Natural Proofs" barrier [16, 17, 18] indicates that achieving such bounds will be very difficult, and (3) average-case hardness with

[2] In several papers, including [2], instead of a single reduction Red that receives an a bit long advice string α, they define a set of reduction circuits of size 2^a. The two definitions are equivalent. We also remark that we allow the map Con to be arbitrary (with no complexity restrictions) whereas some previous work placed a bound on its complexity. This only makes our results stronger.

$\epsilon = 1/n$ is required for several important applications such as the construction of pseudorandom generators a la Nisan [19], and deriving further lower bounds via the so-called "discriminator lemma" [20].

This is especially frustrating because for several important circuit classes, such as constant-depth circuits with mod p gates for prime p, or constant-depth circuits with a limited number of Majority gates, explicit hard functions are known, and in fact even functions with hardness $1/2 - o(1)$. However, we can't use hardness amplification to amplify the hardness to the point where we could use it to construct pseudorandom generators or infer additional lower bounds. For classes such as constant-depth circuits with parity gates the best known pseudorandom generators have very poor seed length of the form $m(1-o(1))$ where m is the output length [21].

More specifically, [15] conjectures that every circuit class that can prove hardness amplification to hardness $1/2 - 1/\epsilon$ can compute majority on $\Omega(1/\epsilon)$ bits, and must use $\Omega(1/\epsilon)$ queries. A more precise conjecture on the number of queries is $\Omega(\log(1/\delta)/\epsilon^2)$. Special cases of these conjectures are proved in [15] and in subsequent works [22, 2, 23, 24], the last three of which are incomparable as they restrict the hardness amplification in different ways. Previous work is discussed in more detail in Section I-B2 below.

In this paper we prove the conjectures, thus in particular improving on three incomparable works [2, 23, 24].

1) Our results on hardness amplification proofs: First, we prove a tight query lower bound. Showing that reductions in black-box hardness amplification must make at least q queries translates to saying that when transforming a function f that is hard against circuits of size s to a function Con_f that is harder against circuits of size s', then $s' \leq s/q$. This means that a "size loss" is unavoidable in black-box hardness amplification.

Theorem I.7 (Lower bound on the number of queries)**.** *There exist constants $\delta_0, \nu > 0$ such that: Let $(\mathsf{Con}, \mathsf{Red})$ be a $\delta \to (\frac{1}{2} - \epsilon)$ black-box hardness amplification with input lengths k and n, and list size 2^a. Assume that:*

- *$\mathsf{Red}^{(\cdot)}$ is a size r oracle circuit, that makes at most q queries.*
- *$n, a, \frac{1}{\epsilon} \leq r \leq 2^{\nu \cdot k}$ and $2^{-2k} \leq \delta \leq \delta_0$.*

Then $q = \Omega(\log(1/\delta)/\epsilon^2)$.

The parameters achieved by Theorem I.7 are tight up to constants, matching the parameters obtained by Klivans and Servedio [25] for the XOR lemma, using Impagliazzo's hard-core lemma [26]. Note that the special case of $\delta = 2^{-2k}$ (which is the same as $\delta = 0$) captures worst-case to average-case hardness amplification, and then the lower bound is $q = \Omega(k/\epsilon^2)$.

Remark I.8 (A strengthening of Theorem I.7)**.** *The assumption that $\mathsf{Red}^{(\cdot)}$ is a size r oracle circuit is not necessary in Theorem I.7, and can be omitted. We explain how to omit it in the full version. Making this additional assumption simplifies the notation in the proof, and note that reductions of exponential size are not useful for proving hardness amplification theorems, and so the additional condition doesn't matter.*

We then prove that hardness amplification proofs require majority.

Theorem I.9 (Hardness amplification proofs require majority)**.** *There exist constants $\delta_0, \nu > 0$ such that: Let $(\mathsf{Con}, \mathsf{Red})$ be a $\delta \to (\frac{1}{2} - \epsilon)$ black-box hardness amplification with input lengths k and n, and list size 2^a. Assume that:*

- *$\mathsf{Red}^{(\cdot)}$ is a size r oracle circuit of depth $d = O(\log(1/\epsilon))$ (over a set of gates that includes the standard boolean gates with unbounded fan-in)*
- *$n, a, \frac{1}{\epsilon} \leq r \leq 2^{\nu \cdot k}$ and $\delta \leq \delta_0$.*

Then there exists a circuit C of size poly(r), and depth $O(d)$ that uses the gates allowed to Red, and computes the majority function on inputs of length $\Omega(1/\epsilon)$.

In particular, if $\epsilon = 1/n$ and $\mathsf{Red}(., \alpha)$ are constant-depth circuits with And, Or, and Parity gates of unbounded fan-in, and Not gates, then its size must be $2^{n^{\Omega(1)}}$ by the known lower bounds [27].

Theorem I.9 is tight in the sense that Klivans [28] observed that there are reductions that can be implemented by a constant-depth circuit with only one Majority gate. For an exposition of a simplification of Klivans' argument, due to Klivans and Vadhan, see Lectures 6 and 7 in [29].

Lower bounds on local decoding algorithms for list-decodable codes: The aforementioned results of [13] show that a $0 \to (\frac{1}{2} - \epsilon)$ black-box hardness amplification $(\mathsf{Con}, \mathsf{Red})$ with list-size 2^a yields $(\frac{1}{2} - \epsilon, 2^a)$-list decodable code, with an encoding map $\mathsf{Enc} : \{0,1\}^{2^k} \to \{0,1\}^{2^n}$, given by $\mathsf{Enc}(f) = \mathsf{Con}_f$ (here functions are identified with their truth tables) and Red is a "local list-decoding algorithm".[3] In this terminology, our results give lower bounds on the complexity, and on the number of queries of local list decoding algorithms for list decodable codes.

[3]A function $\mathsf{Enc} : \{0,1\}^K \to \{0,1\}^N$ is (ρ, ℓ)-list decodable if for every $y \in \{0,1\}^N$, there are at most ℓ elements $x \in \{0,1\}^K$ such that $\mathsf{Enc}(x)$ ρ-agrees with y.

Limitations on direct-product lemmas and decoding from erasures: Another extensively studied construction map is the nonboolean map $\mathsf{Con}_f(x_1, \ldots, x_t) = (f(x_1), \ldots, f(x_t))$, and proofs for this map are known as *concatenation lemmas* or *direct product lemmas*. Our techniques also apply in this setting, and give the lower bound of $q = \Omega(\log(1/\delta)/\epsilon)$ in this case. This is tight [30]. We also consider hardness amplification in the "error-less" setting, a study initiated by Bogdanov and Safra [31]. We prove a query lower bound of $q = \Omega(\log(1/\delta)/\epsilon)$, which matches a construction by Watson [32]. These results follow by proving lower bounds on the more general setup of "basic hardness amplification" considered by Artemenko and Shaltiel [24] and Watson [32]. It should be noted that these proofs can be implemented by constant-depth circuits, and so computing majority is not required. This was observed in [15] for the direct product lemma, and can be verified by inspection for example of [32] for error-less hardness amplification, see [24] for discussion.

Limitations on black-box PRG constructions: Our techniques also apply to "hardness versus randomness", that is to constructing pseudorandom generators from hard functions. Loosely speaking, we show the same lower bounds for constructing pseudorandom generators with error ϵ as for amplifying hardness to $1/2 - \epsilon$.

2) More on related work, and why previous negative results do not capture all available techniques: As mentioned earlier, previous works proved special cases of the conjectures by restricting the hardness amplification in various ways. One way to restrict was requiring that the proof Red is *non-adaptive*, that is, only makes non-adaptive query to the oracle. The restriction to non-adaptive reductions is severe because there exist hardness amplification proofs that do exploit adaptivity, such as [33, 34, 35, 23]. We note that the proofs of [33, 34] use adaptive reductions for a slightly different task: converting hard functions to pseudorandom generators. As mentioned earlier, our lower bounds also apply to pseudorandom-generator constructions.

Another way to restrict was to limit the length of the advice string a, that is, considering reductions Red that are uniform. Some reductions for hardness amplification in the literature use as little as $a = O(\log(1/\epsilon))$ bits of nonunifomity [30], but most reductions use $a = \text{poly}(1/\epsilon)$ (or more) bits of nonuniformity.

Viola [15] proves that majority is necessary for hardness amplification proofs based on the Hadamard code, or on the Reed-Muller code concatenated with Hadamard. This result applies to adaptive proofs, however only if the list size a is small ($a = O(\log(1/\epsilon))$). Viola [15] also proves the query lower bound $q \geq \Omega(\min\{1/\epsilon, k/\log k\})$. This result can handle large a, even $a = 2^{\Omega(k)}$, but only applied to *non-adaptive reductions*. These results were subsumed by the next three works.

Shaltiel and Viola [2] proved limitations on *non-adaptive reductions*. Their results are identical to Theorem I.7 and Theorem I.9, except that it only handles *nonadaptive* reductions. Our results extend theirs by allowing for adaptive reductions.

Gutfreund and Rothblum [23] extend the above result about Reed-Muller code concatenated with Hadamard to any code. That is, they can handle any hardness amplification but they require a to be small. Our results extend theirs by allowing large a.

Artemenko and Shaltiel [24] proved a lower bound of $q = \Omega(1/\epsilon)$ for the number of queries (allowing both adaptivity and large advice). Their approach is to consider a hardness amplification variant that corresponds to codes that are locally list-decodable from *erasures*. They called this setup "basic hardness amplification." Our results extend their work and give a (tight) lower bound of $q = \Omega(\log(1/\delta)/\epsilon)$ in that setup.

Watson [32] considered an intermediate setup of "errorless amplification" and obtained a tight lower bound of $q = \Omega(\log(1/\delta)/\epsilon)$ on *nonadaptive* reductions (by reducing to the lower bound of [2]). Our results extend this lower bound to *adaptive* reductions.

Applebaum, Artemenko, Shaltiel and Yang [36] considered a more powerful class of reductions that are allowed to be *nondeterministic* oracle circuits. They prove limitations on such reductions for the case that $\epsilon \ll 1/r$. These results are incomparable to ours.

Organization

This version is an extended abstract. Due to space limitations it only contains a high level description of our results and techniques. The reader is referred to the full version [37] for precise details and full proofs.

II. Techniques

In this section we aim to give an informal overview of our arguments, trying to sum up the main ideas. The technical section of this paper includes full proofs, and does not build on this informal overview.

A. The forbidden and fixed set lemma

Our proofs of the new forbidden set lemma and fixed set lemma use basic information theory (as is the case for the proofs of Lemma I.1 and Lemma I.2). For simplicity let us focus on the case where $R_1, \ldots, R_N$

are i.i.d. and uniformly distributed. In the full version we observe that it is indeed sufficient to study the case of the uniform distribution to obtain results on arbitrary i.i.d. variables).

The setting for Lemmas I.1, I.2, I.3 and I.4 is the following: Let $X = (R|R \in A)$ for A such that $\Pr[R \in A] \geq 2^{-a}$. We aim to show that X and R cannot be distinguished from uniform by shallow decision trees. It is immediate that $H(X) \geq N - a$, where H is the Shannon entropy function. All previous works in this area eventually connect entropy and statistical distance using Pinsker's inequality, which guarantees that a distribution Y over n bits is η-close to uniform if $H(Y) \geq n - \eta^2$.

1) Previous proofs for the nonadaptive case: It is instructive to first explain the argument used in Lemma I.1 and Lemma I.2, and point out where this approach fails for adaptive decision trees. The proof of Raz [38] for Lemma I.1 works by first using the chain rule for entropy:

$$N - a \leq H(X) = \sum_{i \in [N]} H(X_i | X_1, \ldots, X_{i-1}).$$

Then, choose $\alpha = \eta^2$, and let B be the set of indices i such that $H(X_i|X_1, \ldots, X_{i-1}) < 1 - \alpha$. By a Markov argument, there are at most $a/\alpha = a/\eta^2$ such "weak" i. For every $i \in [N] \setminus B$,

$$H(X_i) \geq H(X_i|X_1, \ldots, X_{i-1}) \geq 1 - \alpha = 1 - \eta^2,$$

which by Pinsker's inequality gives that X_i is η-close to uniform.

An extension of this argument was used in [2], where they choose $\alpha = \eta^2/q$ and for $i_1, \ldots, i_q \notin B$,

$$H(X_{i_1}, \ldots, X_{i_q}) = \sum_{j \in [q]} H(X_{i_j} | X_{i_1}, \ldots, X_{i_{j-1}})$$

$$\geq \sum_{j \in [q]} H(X_{i_j} | X_1, X_2, \ldots, X_{i_j - 1}) \geq q \cdot (1 - \alpha) = q - \eta^2,$$

which by Pinsker's inequality gives that $(X_{i_1}, \ldots, X_{i_q})$ is η-close to uniform.

It is instructive to consider that in the pointer chasing example, this argument produces $B = \emptyset$ and does not identify the indices of the pointer. Loosely speaking, this means that the criteria used by the proofs above for finding "bad indices" and placing them in B isn't suited for adaptive decision trees.

2) The forbidden set lemma: We explain the argument for Lemma I.3. We need to come up with a criteria for selecting indices that does identify the "pointer bits" in the pointer chasing example. We use the following idea (inspired by a similar argument from [3]). We say that an $i \in [N]$ is α-*weak* if $H(X_i|X_{[N]\setminus\{i\}}) < 1 - \alpha$. A key observation is that this criteria (which is less stringent than the one used above) does identify the bits of the pointer in the pointer chasing example. Moreover, we can bound the number of α-weak bits, by the following iterative process: while there is an α-weak bit i', remove it, place it in B and continue. In an iteration of this process, by the chain rule:

$$H(X_{[N]\setminus\{i'\}}) = H(X) - H(X_{i'}|X_{[N]\setminus\{i'\}})$$
$$\geq N - a - (1 - \alpha) = (N - 1) - (a - \alpha),$$

and so in each iteration the gap between the bit-length of X and its entropy decreases by α. The initial gap is a, and so, we can have at most a/α iterations, as entropy is bounded above by bit-length. We will choose $\alpha = \text{poly}(\eta/q)$, so that after the last iteration we have a "forbidden set" B of size $b \leq a/\alpha = \text{poly}(a, q, 1/\eta)$.

To conclude we need to show that if an adaptive q-query decision tree P that does not make queries in B distinguishes X from R with advantage η, then there exists some $i' \in [N] \setminus B$ that is not α-weak (which gives a contradiction). Essentially, we use the "distinguisher to predictor hybrid argument" [39, 40, 41] to obtain an index i' such that the bit $X_{i'}$ can be predicted from $X_{[N]\setminus\{i'\}}$ with large advantage over random guessing, showing that i' is α-weak.

3) The fixed set lemma: We explain the argument for Lemma I.4. The proof will follow the same overall approach of the forbidden set lemma. However, this time, we will show that if a q-query decision tree distinguishes X from uniform, then we can fix a few bits of X, and reduce the gap between its bit-length and its entropy.

More precisely, we will show that if there exists a decision tree P that distinguishes X from R with advantage η, then for $\alpha = \eta^2$, there exist $i_1, \ldots, i_q \in [N]$, and $v_1, \ldots, v_q \in \{0, 1\}$ such that:

$$H(X | X_{i_1} = v_1, \ldots, X_{i_q} = v_q) \geq (N - q) - (a - \alpha). \quad (1)$$

For this purpose, let $I = (I_1, \ldots, I_q)$ be the indices of the variables queried by P on input X. Note I is a random variable that is a function of X. By the chain rule we have that:

$$H(X) = H(X, X_{I_1}, \ldots, X_{I_q})$$
$$= H(X_{I_1}, \ldots, X_{I_q}) + H(X | X_{I_1}, \ldots, X_{I_q})$$

This gives that:

$$H(X | X_{I_1}, \ldots, X_{I_q}) = H(X) - H(X_{I_1}, \ldots, X_{I_q})$$
$$\geq N - a - (q - \alpha) = (N - q) - (a - \alpha),$$

where the inequality follows by Pinsker's because the answers $(X_{I_1}, \ldots, X_{I_q})$ are not η-close to uniform. Equation (1) now follows by an averaging argument that fixes X_I and hence I.

Equation (1) gives that we are able to fix q variables, and decrease the gap between the bit-length of X and its entropy by α. Once again, this gap is initially less than a, and so, this can happen at most a/α times. Consequently, by iteratively applying this process, we end up with a distribution where we fixed at most $q \cdot a/\alpha = \text{poly}(a, q, 1/\eta)$ bits to some specific values. The final distribution has that the bits that we did not fix cannot be distinguished from uniform by a q-query decision tree. The precise argument appears in the full version. In that version we also include and discuss a slightly different argument suggested to us by an anonymous referee. It only easily applies to distributions uniform on their support, which are sufficient for our application, but it avoids entropy and Pinsker's inequality, and gives a better dependence on η.

B. Lower bounds on hardness amplification

Past work of Viola [15] and Shaltiel and Viola [2] provides a framework that can be used in conjunction with lemmas of the type discussed in the previous section to obtain lower bounds on black-box hardness amplification. Theorem I.7 (on the number of queries) directly follows by extending the approach of [2] using Lemma I.4 (instead of Lemma I.2). Proving Theorem I.9 (on "hardness amplification implies majority") raises additional difficulties that do not come up in the non-adaptive case. We start by a high-level explanation of the approach of [2].

1) The Zoom lemma: Let Noise_p^N denote the distribution of N i.i.d. indicator random variables, where each is one with probability p. The approach of [2] is to show a "Zoom lemma" saying that: under certain conditions, a black-box hardness amplification $(\mathsf{Con}, \mathsf{Red})$ implies a procedure P on $N = 2^n$ variables that distinguishes $\mathsf{Noise}_{\frac{1}{2}}^N$ from $\mathsf{Noise}_{\frac{1}{2}-\epsilon}^N$ with probability roughly $1 - \delta$. The complexity of this procedure is inherited by the complexity of Con and Red specifically:

- If Red makes at most q queries, then P can be implemented by a q-query decision tree.
- We can view an element in $\{0,1\}^N$ as a function $D : \{0,1\}^n \to \{0,1\}$. In this notation, the distinguishing procedure P is an oracle procedure $P^{(\cdot)}$ that receives oracle access to D that is chosen from either $\mathsf{Noise}_{\frac{1}{2}}^N$ or $\mathsf{Noise}_{\frac{1}{2}-\epsilon}^N$. Using this terminology, $P^D = \mathsf{Red}^{\mathsf{Con}_{f \oplus D}}(x, \alpha)$ for some specific function f, input x and advice string α that are chosen in the proof.

Our first step is to use our new tools to prove a zoom lemma for adaptive reductions. The lemma appears in the full version. A side benefit of our new approach is that our new tools simplify the proof of the zoom lemma (even in the nonadaptive case).

The argument for the zoom lemma: The high level idea of the proof of the zoom lemma is to fix some function $f : \{0,1\}^k \to \{0,1\}$ and consider the behavior of the reduction when given oracle access to $\mathsf{Con}_f \oplus \mathsf{Noise}_{\frac{1}{2}}^N$ and $\mathsf{Con}_f \oplus \mathsf{Noise}_{\frac{1}{2}-2\epsilon}^N$. In the first case, the noise completely masks Con_f and the reduction receives no information about f. This means that for a random f, Red is unlikely to 0.51-agree with f. In the second case, for any f, $\mathsf{Con}_f \oplus \mathsf{Noise}_{\frac{1}{2}-2\epsilon}^N$ is (w.h.p.) an oracle on which there exists α (which can be an arbitrary function of $\mathsf{Noise}_{\frac{1}{2}-2\epsilon}^N$) on which $\mathsf{Red}(\cdot, \alpha)$ needs to $(1-\delta)$-agree with f. This intuitively means that the procedure Red can distinguish $\mathsf{Noise}_{\frac{1}{2}-2\epsilon}^N$ from uniform when it receives a "advice bits". Thus, in order to prove the zoom lemma, it is sufficient to show that Red cannot distinguish $R = \mathsf{Noise}_{\frac{1}{2}-2\epsilon}^N$ from $X = (R | R \in A)$ where A is an event of probability 2^{-a}. This is the setup considered in Section I-A.

Lower bound on the number of queries: Theorem I.7 immediately follows from the first item in the zoom lemma, as a q query decision tree that distinguishes $\mathsf{Noise}_{\frac{1}{2}}^N$ from $\mathsf{Noise}_{\frac{1}{2}-\epsilon}^N$ with probability $1 - \delta$, must make $\Omega(\log(1/\delta)/\epsilon^2)$ queries, cf. [2]. The precise statement appears in the full version.

Hardness amplification proofs require majority: Viola [15] and Shaltiel and Viola [2], used an idea of Sudan (see discussion in [15]) to give a reduction from the task of computing Majority on $\frac{1}{\epsilon}$ bits, to the task of distinguishing $\mathsf{Noise}_{\frac{1}{2}}^N$ from $\mathsf{Noise}_{\frac{1}{2}-\epsilon}^N$. For constant distinguishing advantage, this reduction transforms a constant depth distinguisher into a constant depth circuit (of polynomially related size) that computes Majority. Thus, in order to obtain Theorem I.9 we need to simulate the computation $\mathsf{Red}^{\mathsf{Con}_{f \oplus D}}(x, \alpha)$ (from the second item of the zoom lemma) by a small constant depth circuit. In this computation, x, α, and f are fixed, while D is not.

Simulating this computation raises another difficulty: it is not clear how to compute Con_f. Although in some cases this can be done in general it is not clear. This problem is even more pronounced in our extensions to pseudorandom-generator construction, where the relevant oracle is the NP function which checks if the string is in the support of the generator.

To overcome this difficulty we adapt an idea of [23] which removes the need to compute Con_f. Roughly, using a hybrid argument we can arrange so that there exists a depth i in the computation of $\mathsf{Red}^{\mathsf{Con}_f \oplus D}(x, \alpha)$ so that the queries at depth $< i$ have a fixed answer (where Con_f can be hardwired), and the queries at depth $> i$ have a completely uniform answer (independent of Con_f) . This holds however only for queries not in the fixed set B: we note that even with this idea we need a fixed-set lemma, as otherwise we do not know how to control the query answers at depth larger than i.

Pseudorandom-generator constructions: The above ideas can be extended in a rather straightforward way to pseudorandom-generator constructions. Roughly, the oracle $\mathsf{Con}_f \oplus D$ is replaced by the oracle $\mathsf{Dist}_f \oplus D$ where Dist_f is the indicator function of the support of the pseudorandom generator.

III. Conclusion

This paper concludes a line of research initiated in [14, 15] by establishing that hardness amplification requires majority and many queries. Recall that a function f can be written as the majority of a polynomial number of functions from a class C if and only if for any distribution on the inputs there exists a function in C that computes f correctly with probability $\geq 1/2+1/\text{poly}$. (One direction can be proved via boosting [42, Section 2.2] or min-max/linear-programming duality [43, Section 5]. The other direction follows from the "discriminator lemma" of [20].) Hence we offer the following alternative interpretation of hardness amplification:

> Black-box hardness amplification is a process that takes a function f that is *already* $1/2 - \epsilon$ *hard under some distribution*, and produces another function f' that has roughly the same hardness *under the uniform distribution*.

A fundamental question remains: is hardness amplification *false*? In particular, is the XOR lemma true for restricted circuit classes, such as constant-depth circuits with parity gates? One can show that the XOR lemma is false for the class of constant-depth circuits with one majority gate. This follows by the bounds in [44], and the fact that the XOR lemma applied to parity is again just parity. But that is essentially the only counterexample that we know.

Our results apply to black-box techniques. We remark that one possible way to break the black-box barrier is to come up with proofs that use the fact that D is a small circuit (bypassing the black-box limitations). A potential non-black-box approach was presented by Gutfreund, Shaltiel and Ta-Shma [45] (see also [46, 47, 48]) in a very specific scenario that has some similarity to "worst-case to average case reductions in NP". The techniques of these papers provably break black-box limitations in a related setting. See discussion by Gutfreund and Ta-Shma [48].

Another question, already highlighted in [2], is whether the construction of pseudorandom generators *with constant error* requires majority. The techniques in this paper have recently enabled the first progress on this question, see [49].

Acknowledgment

Ronen Shaltiel was supported by ISF grant 1628/17. Emnanuele Viola was supported by NSF CCF award 1813930.

We are grateful to Iftach Haitner for very helpful discussions. We also thank the anonymous referees for detailed and helpful feedback.

References

[1] O. Meir and A. Wigderson, "Prediction from partial information and hindsight, with application to circuit lower bounds," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 24, p. 149, 2017. [Online]. Available: https://eccc.weizmann.ac.il/report/2017/149

[2] R. Shaltiel and E. Viola, "Hardness amplification proofs require majority," *SIAM J. on Computing*, vol. 39, no. 7, pp. 3122–3154, 2010.

[3] J. Edmonds, R. Impagliazzo, S. Rudich, and J. Sgall, "Communication complexity towards lower bounds on circuit depth," *Computational Complexity*, vol. 10, no. 3, pp. 210–246, 2001.

[4] E. Viola, "Bit-probe lower bounds for succinct data structures," *SIAM J. on Computing*, vol. 41, no. 6, pp. 1593–1604, 2012.

[5] ——, "Cell-probe lower bounds for prefix sums," 2009, arXiv:0906.1370v1.

[6] M. Pătraşcu and E. Viola, "Cell-probe lower bounds for succinct partial sums," in *21th ACM-SIAM Symp. on Discrete Algorithms (SODA)*, 2010, pp. 117–122.

[7] A. Smal and N. Talebanfard, "Prediction from partial information and hindsight, an alternative proof," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 24, p. 191, 2017. [Online]. Available: https://eccc.weizmann.ac.il/report/2017/191

[8] D. Unruh, "Random oracles and auxiliary input," in *Int. Cryptology Conf. (CRYPTO)*, 2007, pp.

205–223. [Online]. Available: https://doi.org/10.1007/978-3-540-74143-5_12

[9] Y. Dodis, S. Guo, and J. Katz, "Fixing cracks in the concrete: Random oracles with auxiliary input, revisited," in *Int. Conf. on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, 2017, pp. 473–495. [Online]. Available: https://doi.org/10.1007/978-3-319-56614-6_16

[10] S. Coretti, Y. Dodis, S. Guo, and J. Steinberger, "Random oracles and non-uniformity," in *Int. Conf. on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, 2018.

[11] S. Arora and B. Barak, *Computational Complexity*. Cambridge University Press, 2009, a modern approach.

[12] O. Goldreich, N. Nisan, and A. Wigderson, "On Yao's XOR lemma," *Electronic Colloquium on Computational Complexity*, Tech. Rep. TR95–050, March 1995, www.eccc.uni-trier.de/.

[13] M. Sudan, L. Trevisan, and S. Vadhan, "Pseudorandom generators without the XOR lemma," *J. of Computer and System Sciences*, vol. 62, no. 2, pp. 236–266, 2001.

[14] E. Viola, "The complexity of constructing pseudorandom generators from hard functions," *Computational Complexity*, vol. 13, no. 3-4, pp. 147–188, 2004.

[15] ——, "The complexity of hardness amplification and derandomization," Ph.D. dissertation, Harvard University, 2006.

[16] A. Razborov and S. Rudich, "Natural proofs," *J. of Computer and System Sciences*, vol. 55, no. 1, pp. 24–35, Aug. 1997.

[17] M. Naor and O. Reingold, "Number-theoretic constructions of efficient pseudo-random functions," *J. of the ACM*, vol. 51, no. 2, pp. 231–262, 2004.

[18] E. Miles and E. Viola, "Substitution-permutation networks, pseudorandom functions, and natural proofs," *J. of the ACM*, vol. 62, no. 6, 2015.

[19] N. Nisan, "Pseudorandom bits for constant depth circuits," *Combinatorica. An Journal on Combinatorics and the Theory of Computing*, vol. 11, no. 1, pp. 63–70, 1991.

[20] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán, "Threshold circuits of bounded depth," *J. of Computer and System Sciences*, vol. 46, no. 2, pp. 129–154, 1993.

[21] B. Fefferman, R. Shaltiel, C. Umans, and E. Viola, "On beating the hybrid argument," *Theory of Computing*, vol. 9, pp. 809–843, 2013.

[22] C. Lu, S. Tsai, and H. Wu, "Complexity of hard-core set proofs," *Computational Complexity*, vol. 20, no. 1, pp. 145–171, 2011. [Online]. Available: https://doi.org/10.1007/s00037-011-0003-7

[23] D. Gutfreund and G. Rothblum, "The complexity of local list decoding," in *12th Intl. Workshop on Randomization and Computation (RANDOM)*, 2008.

[24] S. Artemenko and R. Shaltiel, "Lower bounds on the query complexity of non-uniform and adaptive reductions showing hardness amplification," *Computational Complexity*, vol. 23, no. 1, pp. 43–83, 2014. [Online]. Available: https://doi.org/10.1007/s00037-012-0056-2

[25] A. Klivans and R. A. Servedio, "Boosting and hard-core sets," *Machine Learning*, vol. 53, no. 3, pp. 217–238, 2003.

[26] R. Impagliazzo, "Hard-core distributions for somewhat hard problems," in *IEEE Symp. on Foundations of Computer Science (FOCS)*, 1995, pp. 538–545.

[27] A. Razborov, "Lower bounds on the dimension of schemes of bounded depth in a complete basis containing the logical addition function," *Akademiya Nauk SSSR. Matematicheskie Zametki*, vol. 41, no. 4, pp. 598–607, 1987, english translation in Mathematical Notes of the Academy of Sci. of the USSR, 41(4):333-338, 1987.

[28] A. R. Klivans, "On the derandomization of constant depth circuits," in *Workshop on Randomization and Computation (RANDOM)*. Springer, 2001.

[29] E. Viola, "Gems of theoretical computer science," 2009, lecture notes of the class taught at Northeastern University. Available at http://www.ccs.neu.edu/home/viola/classes/gems-08/index.html.

[30] R. Impagliazzo, R. Jaiswal, V. Kabanets, and A. Wigderson, "Uniform direct product theorems: Simplified, optimized, and derandomized," *SIAM J. on Computing*, vol. 39, no. 4, pp. 1637–1665, 2010. [Online]. Available: https://doi.org/10.1137/080734030

[31] A. Bogdanov and M. Safra, "Hardness amplification for errorless heuristics," in *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2007, pp. 418–426. [Online]. Available: https://doi.org/10.1109/FOCS.2007.25

[32] T. Watson, "Query complexity in errorless hardness amplification," *Computational Complexity*, vol. 24, no. 4, pp. 823–850, 2015. [Online]. Available: https://doi.org/10.1007/s00037-015-0117-4

[33] R. Shaltiel and C. Umans, "Simple extractors for all min-entropies and a new pseudorandom generator," *J. of the ACM*, vol. 52, no. 2, pp. 172–216, 2005.

[34] C. Umans, "Pseudo-random generators for all hardnesses," *J. of Computer and System Sciences*, vol. 67, no. 2, pp. 419–440, 2003, special issue on STOC2002 (Montreal, QC).

[35] S. Goldwasser, D. Gutfreund, A. Healy, T. Kaufman, and G. N. Rothblum, "Verifying and decoding in constant depth," in *ACM Symp. on the Theory of Computing (STOC)*, 2007, pp. 440–449.

[36] B. Applebaum, S. Artemenko, R. Shaltiel, and G. Yang, "Incompressible functions, relative-error extractors, and the power of nondeterministic reductions (extended abstract)," in *Conf. on Computational Complexity (CCC)*, 2015, pp. 582–600. [Online]. Available: https://doi.org/10.4230/LIPIcs.CCC.2015.582

[37] A. Grinberg, R. Shaltiel, and E. Viola, "Indistinguishability by adaptive procedures with advice, and lower bounds on hardness amplification proofs," *Electronic Colloquium on Computational Complexity (ECCC)*, vol. 25, p. 61, 2018. [Online]. Available: https://eccc.weizmann.ac.il/report/2018/061

[38] R. Raz, "A parallel repetition theorem," *SIAM J. on Computing*, vol. 27, no. 3, pp. 763–803, 1998.

[39] S. Goldwasser and S. Micali, "Probabilistic encryption," *J. of Computer and System Sciences*, vol. 28, no. 2, pp. 270–299, Apr. 1984.

[40] M. Blum and S. Micali, "How to generate cryptographically strong sequences of pseudo-random bits," *SIAM J. on Computing*, vol. 13, no. 4, pp. 850–864, Nov. 1984.

[41] A. Yao, "Theory and applications of trapdoor functions," in *23rd IEEE Symp. on Foundations of Computer Science (FOCS)*. IEEE, 1982, pp. 80–91.

[42] Y. Freund, "Boosting a weak learning algorithm by majority," *Information and Computation*, vol. 121, no. 2, pp. 256–285, 1995.

[43] M. Goldmann, J. Håstad, and A. A. Razborov, "Majority gates vs. general weighted threshold gates," *Computational Complexity*, vol. 2, pp. 277–300, 1992.

[44] J. Aspnes, R. Beigel, M. Furst, and S. Rudich, "The expressive power of voting polynomials," *Combinatorica. An Journal on Combinatorics and the Theory of Computing*, vol. 14, no. 2, pp. 135–148, 1994.

[45] D. Gutfreund, R. Shaltiel, and A. Ta-Shma, "If NP languages are hard on the worst-case, then it is easy to find their hard instances," *Computational Complexity*, vol. 16, no. 4, pp. 412–441, 2007. [Online]. Available: https://doi.org/10.1007/s00037-007-0235-8

[46] A. Atserias, "Distinguishing SAT from polynomial-size circuits, through black-box queries," in *IEEE Conf. on Computational Complexity (CCC)*, 2006, pp. 88–95. [Online]. Available: https://doi.org/10.1109/CCC.2006.17

[47] D. Gutfreund, "Worst-case vs. algorithmic average-case complexity in the polynomial-time hierarchy," in *Workshop on Randomization and Computation (RANDOM)*, 2006, pp. 386–397. [Online]. Available: https://doi.org/10.1007/11830924_36

[48] D. Gutfreund and A. Ta-Shma, "Worst-case to average-case reductions revisited," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 10th International Workshop, APPROX 2007, and 11th International Workshop, RANDOM 2007, Princeton, NJ, USA, August 20-22, 2007, Proceedings*, 2007, pp. 569–583. [Online]. Available: https://doi.org/10.1007/978-3-540-74208-1_41

[49] E. Viola, "Constant-error pseudorandomness proofs from hardness require majority," 2018, available at http://www.ccs.neu.edu/home/viola/.