Hardness Amplification Proofs Require Majority

Ronen Shaltiel[*] Emanuele Viola[†]

March 3, 2008

Abstract

Hardness amplification is the fundamental task of converting a δ-hard function $f : \{0,1\}^n \to \{0,1\}$ into a $(1/2-\epsilon)$-hard function $Amp(f)$, where f is γ-hard if small circuits fail to compute f on at least a γ fraction of the inputs. Typically, ϵ, δ are small (and $\delta = 2^{-k}$ captures the case where f is worst-case hard). Achieving $\epsilon = 1/n^{\omega(1)}$ is a prerequisite for cryptography and most pseudorandom-generator constructions.

In this paper we study the complexity of black-box proofs of hardness amplification. A class of circuits $\mathcal{D}$ *proves* a hardness amplification result if for any function h that agrees with $Amp(f)$ on a $1/2+\epsilon$ fraction of the inputs there exists an oracle circuit $D \in \mathcal{D}$ such that D^h agrees with f on a $1-\delta$ fraction of the inputs. We focus on the case where every $D \in \mathcal{D}$ makes *non-adaptive* queries to h. This setting captures most hardness amplification techniques. We prove two main results:

1. The circuits in $\mathcal{D}$ "can be used" to compute the majority function on $1/\epsilon$ bits. In particular, these circuits have large depth when $\epsilon \leq 1/\text{poly}\log n$.
2. The circuits in $\mathcal{D}$ must make $\Omega\left(\log(1/\delta)/\epsilon^2\right)$ oracle queries.

Both our bounds on the depth and on the number of queries are tight up to constant factors.

Our results explain why hardness amplification techniques have failed to transform known lower bounds against constant-depth circuit classes into strong average-case lower bounds. When coupled with the celebrated "Natural Proofs" result by Razborov and Rudich (J. CSS '97) and the pseudorandom functions by Naor and Reingold (J. ACM '04), our results show that *standard techniques for hardness amplification can only be applied to those circuit classes for which standard techniques cannot prove circuit lower bounds.*

Our results reveal a contrast between *Yao's XOR Lemma* ($Amp(f) := f(x_1) \oplus \ldots \oplus f(x_t) \in \{0,1\}$) and the *Direct-Product Lemma* ($Amp(f) := f(x_1) \circ \ldots \circ f(x_t) \in \{0,1\}^t$; here $Amp(f)$ is non-Boolean). Our results (1) and (2) apply to Yao's XOR lemma, whereas known proofs of the direct-product lemma violate both (1) and (2).

One of our contributions is a new technique to handle "non-uniform" reductions, i.e. the case when $\mathcal{D}$ contains many circuits.

[*] ronen@cs.haifa.ac.il. This research was supported by BSF grant 2004329 and ISF grant 686/07.

[†] viola@cs.columbia.edu. Supported by grants NSF award CCF-0347282 and NSF award CCF-0523664. Research partially done while at Harvard University, supported by grants NSF CCR-0133096, US-Israel BSF 2002246, and ONR N-00014-04-1-0478, and at the Institute for Advanced Study, supported by NSF grant CCR-0324906.

1 Introduction

Proving circuit lower bounds is a major goal of Complexity Theory. However, the celebrated "Natural Proofs" result by Razborov and Rudich [RR], coupled with the pseudorandom functions by Naor and Reingold [NR], marks the class of polynomial-size constant-depth circuits *with majority gates* (TC^0) as a fundamental limit for most currently available lower bounding techniques. This limitation already applies to *worst-case* lower bounds, where one seeks a function that small circuits fail to compute on *at least one* input. In particular, it applies to *average-case* lower bounds, where one seeks a function that small circuits fail to compute on *many* inputs. Average-case hard functions are especially important as they are a prerequisite for most modern cryptography and can be used to construct pseudorandom generators [NW] which in turn have a striking variety of applications (see, e.g., the books by Goldreich [Gol2, Gol3]). We stress that both these applications require *strongly* average-case hard functions. That is functions that small circuits cannot compute with even a small advantage over random guessing, for a randomly chosen input. (For concreteness, the reader may think of a function $f : \{0,1\}^n \to \{0,1\}$ that any small circuit fails to compute with probability $1/2 - 1/n^{\omega(1)}$ over the choice of the input).

As we do not know how to prove unconditional lower bounds for general circuit classes, a long line of research has focused on *hardness amplification.* This is the task of transforming worst-case hard functions (or sometimes *mildly* average-case hard functions) into average-case hard functions [Yao1, Lip, BF, BFL, BFNW, Imp, GNW, FL, IW1, IW2, CPS, STV, TV, SU1, Tre1, O'D, Vio1, Tre3, HVV, SU2, GK, IJK, IJKW, GG]. This research was largely successful in its goal. In particular, it provided worst-case to average-case connections within many complexity classes. Many of these connections give *strongly* average-case hard functions. This research also spurred fruitful interaction with coding theory (see, e.g., the survey by Trevisan [Tre2]).

Complexity theory has produced many exciting and useful lower bounds for restricted computational models, most notably against classes of circuits with unbounded fan-in and constant depth with various gates [FSS, Yao2, Hås, Raz, Smo, HG, HMP$^+$, ABFR, HM]. In some of these classes we in fact can prove worst-case lower bounds, but cannot prove *strongly* average-case lower bounds (e.g. [Raz, Smo, ABFR]). Several such examples are surveyed in Section 7 and in [Vio3, Chapter 6]; for concreteness, an example is the lower bound against constant-depth circuits with And, Or and Parity gates [Raz, Smo]. One would expect that hardness amplification techniques could be used to produce strongly average-case lower bounds from the known lower bounds (which would in turn give pseudorandom generators for these classes [NW]). But in fact "standard hardness amplification techniques" fail.

In this paper we show that:

"standard hardness amplification techniques" only apply when starting with hardness against circuits that can compute the majority function.

This explains the following *"lose-lose" phenomenon:* For classes that are weaker than TC^0 (e.g. constant-depth circuits, or constant-depth circuits with parity gates) we can prove

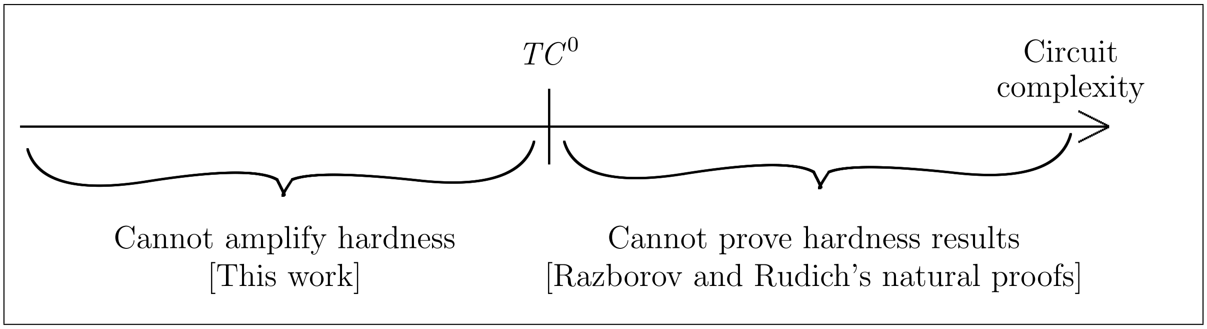
TC^0
Circuit complexity
Cannot amplify hardness
[This work]
Cannot prove hardness results
[Razborov and Rudich's natural proofs]

Figure 1: Reach of "standard techniques." Recall TC^0 is the class of polynomial-size constant-depth circuits with majority gates.

lower bounds, however we do not have hardness amplification theorems, while for classes at least as powerful as TC^0 we have hardness amplification theorems but cannot prove circuit lower bounds; see Figure 1.

A couple of remarks is in order. First, our results likely do not apply to "every conceivable" class of circuits, but rather they apply to the most well-studied ones. Second, we note that, just like Razborov and Rudich's result [RR] is not claiming that it is impossible to prove lower bounds for classes like TC^0, but rather that certain techniques will not do, this work is not claiming that it is impossible to prove strong average-case hardness results for circuit classes weaker than TC^0, but that we cannot obtain such results by "standard hardness amplification techniques." We elaborate on these techniques next.

1.1 Hardness amplification

In this section we review the notion of hardness amplification. Let us start by formalizing our notion of hardness.

Definition 1.1 (Average-case hardness)**.** *A function $f : \{0,1\}^k \to \{0,1\}$ is δ-hard for a class of circuits $\mathcal{C}$ (e.g., all circuits of size s) if for every circuit $C \in \mathcal{C}$ we have* $\Pr_{x\in\{0,1\}^k}[C(x) \neq f(x)] \geq \delta$.

Hardness amplification is the generic task of transforming a given function $f : \{0,1\}^k \to \{0,1\}$ that is δ-hard for a class of circuits $\mathcal{C}$ into another function $Amp(f) : \{0,1\}^n \to \{0,1\}$ that is $(1/2-\epsilon)$-hard for a related class of circuits $\mathcal{C}'$, where one wants ϵ as small as possible and n not much larger than k. The first and most important example of hardness amplification is *Yao's XOR lemma* (cf. [GNW]), which works as follows. We let $n := t \cdot k$ for a parameter t and on input $(x_1, \ldots, x_t) \in \left(\{0,1\}^k\right)^t = \{0,1\}^n$ we define

$$Amp(f)(x_1, \ldots, x_t) := f(x_1) \oplus \cdots \oplus f(x_t),$$

where $\oplus$ denotes exclusive OR. The lemma states that if f is δ-hard for (the class of) circuits of size s, then choosing $t := O(\log(1/\epsilon)/\delta)$ one has that $Amp(f)$ is $(1/2-\epsilon)$-hard for circuits of size $s \cdot \text{poly}(\epsilon \cdot \delta/k)$. In particular, if f is 1/3-hard for circuits of superpolynomial size $s = n^{\omega(1)}$, then by choosing a suitable $t := \omega(\log n)$ we obtain a $(1/2 - 1/n^{\omega(1)})$-hard function, (recall that such a function is a prerequisite of most cryptography and can be used to construct pseudorandom generators [NW]).

Yao's XOR lemma is not useful when starting from worst-case hard functions, i.e., when $\delta = 2^{-k}$. Hardness amplification from worst-case hardness is still possible (e.g., [Lip, BF, BFL, BFNW, FL, CPS, STV, TV]) but is more difficult. This distinction is not relevant to our work which, jumping ahead, proves limitations on hardness amplification that already apply when amplifying from constant hardness $\delta = \Omega(1)$ (and in particular apply when amplifying from worst-case hardness $\delta = 2^{-k}$).

1.2 Black-box hardness amplification

We now explain what we mean by "standard techniques" for proving hardness amplification theorems. To explain this, we use the classical notion of an *oracle circuit* $D^h(x)$, where $h : \{0,1\}^n \to \{0,1\}$. This is simply a circuit with special oracle gates that on input $y \in \{0,1\}^n$ return the value $h(y) \in \{0,1\}$. We note that this notion also makes sense when restricting the depth of the circuit D. It has been observed several times (see, e.g., [Tre1]) that most proofs of hardness amplification in the literature are *black-box* in the following sense.

Definition 1.2 (Black-box hardness amplification). *A $\delta \to (1/2 - \epsilon)$ black-box hardness* amplification *with input lengths k and n is a pair $(Amp, \mathcal{D})$ such that Amp is a map from functions $f : \{0,1\}^k \to \{0,1\}$ to functions $Amp(f) : \{0,1\}^n \to \{0,1\}$, $\mathcal{D}$ is a class of oracle circuits on k input bits (e.g., all oracle circuits of size s), and the following holds:*

For every function $f : \{0,1\}^k \to \{0,1\}$ and every function $h : \{0,1\}^n \to \{0,1\}$ such that

$$\Pr_{y \in \{0,1\}^n} [h(y) \neq Amp(f)(y)] < 1/2 - \epsilon$$

there is an oracle circuit $D \in \mathcal{D}$ such that

$$\Pr_{x \in \{0,1\}^k} \left[D^h(x) \neq f(x)\right] < \delta.$$

The black-box hardness amplification is non-adaptive q-query *if every circuit $D \in \mathcal{D}$ makes q non-adaptive queries to h. Finally, we say that a class of circuits $\mathcal{D}$* proves *a black-box hardness amplification (with certain parameters) if there is a map Amp such that $(Amp, \mathcal{D})$ is a black-box hardness amplification (with the same parameters).*

Why black-box hardness amplification lets us amplify hardness. It is instructive to verify that black-box hardness amplification indeed lets us amplify hardness. To see this, suppose that $(Amp, \mathcal{D})$ is a q-query $\delta \to (1/2 - \epsilon)$ black-box hardness amplification where $\mathcal{D}$ is the class of circuits of size s. Now let $f : \{0,1\}^k \to \{0,1\}$ be δ-hard for (the class of) circuits of size $t \geq 2 \cdot s$. Observe that indeed the function $Amp(f) : \{0,1\}^n \to \{0,1\}$ is $(1/2 - \epsilon)$-hard for circuits of size $t/(2 \cdot q)$ (where recall q is the number of oracle queries made by circuits in $\mathcal{D}$). This is proven by a standard counterpositive argument. Suppose for the sake of contradiction that there exists a circuit h of size $t/(2 \cdot q)$ that computes $Amp(f)$ on more than a $1/2 + \epsilon$ fraction of the inputs. Then by definition of black-box hardness amplification there is a circuit $D \in \mathcal{D}$ such that D^h computes f on more than a $1 - \delta$ fraction of the inputs. Since D has size s and makes q oracle queries, by replacing each query with a copy of the circuit for h we see that D^h can be computed by a circuit of size $q \cdot t/(2 \cdot q) + s \leq t/2 + s \leq t$, contradicting our assumption that f was δ-hard for circuits of size t.

It is also instructive to remark that, in the language of Definition 1.2, Yao's XOR lemma is a $\delta \to (1/2 - \epsilon)$ black-box hardness amplification $(Amp, \mathcal{D})$ with input lengths k and n, where $n = O(k \cdot \log(1/\epsilon)/\delta)$ and $\mathcal{D}$ is the class of circuits of size $\mathrm{poly}(k/(\epsilon \cdot \delta))$.

The complexity of $\mathcal{D}$. We want to stress that the complexity of the class $\mathcal{D}$ plays a crucial role when deriving average-case hardness results using a black-box hardness amplification. Specifically, to obtain hardness amplification the initial function $f : \{0,1\}^k \to \{0,1\}$ must be hard for a class of circuits that contains $\mathcal{D}$. This is a key point for our results which will essentially show that $\mathcal{D}$ has to be at least as powerful as TC^0, the class of constant-depth circuits with majority gates. Thus for hardness amplification we need to start from a lower bound against TC^0.

Non-uniformity. Another aspect we wish to stress is the *non-uniformity* of the notion of black-box hardness amplification. In Definition 1.2 the circuit $D \in \mathcal{D}$ is allowed to depend arbitrarily on both the δ-hard function f and the function h that approximates $Amp(f)$. It can be shown that some non-uniformity is *necessary* for black-box hardness amplification: $|\mathcal{D}| \geq (1/\epsilon)^{\Omega(1)}$ [TV]. Establishing hardness amplification results with small non-uniformity (e.g. $|\mathcal{D}| = \mathrm{poly}(1/\epsilon)$) is important for achieving "uniform hardness amplification within NP" and is the focus of a lot of recent attention (see Section 1.4 on related work). In this work we give impossibility results for black-box hardness amplification and therefore are interested in handling *any* black-box hardness amplification, including ones which use large non-uniformity (e.g. $|\mathcal{D}| = \exp(1/\epsilon)$).

1.3 Our results

The main result of this paper applies to *non-adaptive* black-box hardness amplification and can be stated informally as follows:

> If a set of circuits $\mathcal{D}$ proves non-adaptive $\delta \to (1/2 - \epsilon)$ black-box hardness amplification then $\mathcal{D}$ "can be used" to compute majority on $1/\epsilon$ bits. $(\star)$

The formal statement of the above result requires a bit of notation, and is deferred to Theorem 1.6 at the end of this section where, intuitively, we show how oracle access to the circuits $\mathcal{D}$ is sufficient to compute majority. For now we state a qualitatively weaker result which requires less notation. Specifically, the next theorem shows that if $\mathcal{D}$ proves non-adaptive $\delta \to (1/2 - \epsilon)$ black-box hardness amplification, then the depth of the circuits in $\mathcal{D}$ must be large whenever ϵ is small (cf. Definition 1.2 for the definition of "proves"). This weak form of the theorem intuitively follows from $(\star)$ by using the well-known fact that computing the majority function on $m := 1/\epsilon$ bits by circuits of depth d requires size $s \geq \exp\left(m^{\Omega(1/d)}\right) = \exp\left((1/\epsilon)^{\Omega(1/d)}\right)$, i.e. exponential in $1/\epsilon$ [Hås, Raz, Smo].

Theorem 1.3 (Decoding requires majority, stated in terms of circuit depth)**.** *Suppose that a class of non-adaptive oracle circuits $\mathcal{D}$ proves a $(\delta = 1/3) \to (1/2 - \epsilon)$ black-box hardness amplification $(Amp, \mathcal{D})$ with input lengths k and n.*

Suppose that every circuit $D \in \mathcal{D}$ has size s and depth d. Then

$$s \geq \min\left\{\exp\left((1/\epsilon)^{\Omega(1/d)}\right), 2^{\Omega(k)}\right\}.$$

In particular, Theorem 1.3 implies that poly(n)-size constant-depth circuits (i.e., d is fixed and $s = \text{poly}(n)$ grows) can only prove hardness amplification up to $1/2 - \epsilon \leq 1/2 - 1/\text{poly}\log n$. This should be contrasted with standard hardness amplifications (e.g., [GNW]) that show that if we do not put any restriction on the depth of the circuits in $\mathcal{D}$ then circuits of size poly(n) can prove hardness amplification up to $1/2 - 1/n$.

We remark that, for constant-depth circuits, the size bound in Theorem 1.3 is tight. This follows easily from Impagliazzo's beautiful hard-core set theorem [Imp] when amplifying from constant hardness $\delta = \Omega(1)$. Moreover, Impagliazzo's result [Imp] conceptually matches our result $(\star)$ by showing that computing majority on poly$(1/\epsilon)$ bits is "all that is needed" for proving hardness amplification. Precisely this feature was exploited a few times in complexity theory, for example in Klivans' elegant work [Kli]. When amplifying from worst-case hardness $\delta = 2^{-k}$, the construction by Goldwasser et al. [GGH$^+$][1] again matches the size bound in our Theorem 1.3.

Our second main result is a lower bound on the number of queries made by circuits $\mathcal{D}$ in any black-box hardness amplification $(Amp, \mathcal{D})$. One reason for studying the number of queries necessary for proving hardness amplification is the loss in circuit size, i.e. the difference between the circuit sizes that come up in the assumption and conclusion of the hardness amplification theorem. The question of how much loss is necessary has been raised a number of times (see, e.g., [GNW, KS]) but was never answered in generality until this paper. Additional motivation is discussed in Sections 7, 8.

Theorem 1.4 (Decoding requires many queries)**.** *There is a universal constant $C > 1$ such that the following holds. Let $(Amp, \mathcal{D})$ be a non-adaptive q-query $\delta \to (1/2 - \epsilon)$ black-box hardness amplification. Suppose that $\log|\mathcal{D}| \leq 2^{k/C}$, and $n, k \geq C^2$, and that both δ and ϵ are between $2^{-k/C}$ and $1/3$.*

Then

$$q \geq \frac{1}{C} \cdot \frac{\log(1/\delta)}{\epsilon^2}.$$

[1] See Theorem 5.20 in the full version of [GGH$^+$].

We also note that the lower bound of Theorem 1.4 is tight (up to constants) even when only considering XOR-lemmas. This is because Impagliazzo's proof of the XOR-lemma [Imp] can be made to work with $q = O\left(\log(1/\delta)/\epsilon^2\right)$ queries matching our lower bound.[2]

It has been observed (see e.g. [Tre1]) that black-box hardness amplification is closely related to list-decodable codes. Using this connection our results can be seen as lower bounds on the "complexity of decoding" locally-decodable codes. We explain this view in Section 8.

XOR lemma vs. direct product: A qualitative difference. So far we have discussed hardness amplification where the amplified function $Amp(f)$ is Boolean, i.e. its range is $\{0,1\}$, and our leading example was Yao's XOR lemma which recall is defined as $Amp(f) := f^{\oplus t}(x_1,\ldots,x_t) = f(x_1) \oplus \ldots \oplus f(x_t) \in \{0,1\}$, where $\oplus$ denotes exclusive-or.

Hardness amplification where the amplified function $Amp(f) : \{0,1\}^n \to \{0,1\}^t$ is *not* Boolean, i.e. $t \geq 1$, is also widely studied. The first and most important example of this is the *direct product* which is defined as follows $Amp(f) := f^{\circ t}(x_1,\ldots,x_t) = f(x_1) \circ \ldots \circ f(x_t) \in \{0,1\}^t$, where $\circ$ denotes concatenation. Recall that in XOR-lemmas we are interested in amplifying hardness from δ to $1/2 - \epsilon$, whereas in direct-product lemmas we are interested in amplifying from δ to $1 - \epsilon$.

The direct product and the XOR lemma, and more generally Boolean and non-Boolean hardness amplification, have often been regarded as essentially interchangeable. In fact, many proofs of Boolean hardness amplification proceed by proving the direct product first and then transforming the amplified function $f^{\circ t}$ into a Boolean function (see, e.g., [GNW, IW1, STV, O'D, Tre1, HVV]), often using the remarkable Goldreich-Levin Theorem [GL]. The converse, proving a direct product lemma from an XOR lemma, is much easier [VW].

By contrast, *our results show that Yao's XOR lemma and the direct product lemma are qualitatively different.*

The main difference is that the proof of Yao's XOR lemma requires majority, whereas the proof of the direct product lemma does not. Specifically, our results show that if a class $\mathcal{D}$ proves a $(\delta = 1/3) \to (1/2 - \epsilon)$ black-box hardness amplification, such as Yao's XOR lemma, then "$\mathcal{D}$ can compute majority," and in particular $\mathcal{D}$ requires either large depth or exponential size in $1/\epsilon$ (Theorem 1.3). On the other hand, there are black-box proofs of the $\delta \to (1-\epsilon)$ direct-product lemma that can be implemented by small constant-depth circuits for arbitrary $\epsilon > 0$. For example, this is achieved by the proof of Goldreich et al. [GNW].[3]

Another difference can be seen in the number of queries. The proof of the direct-product lemma in [GNW] uses $q = O\left(\log(1/\delta)/\epsilon\right)$ queries, and note that for small ϵ this beats our $\Omega\left(\log(1/\delta)/\epsilon^2\right)$ lower bound that applies to XOR lemmas (Theorem 1.4).

[2] The proof in Impagliazzo's paper gives $q = O\left(\log(1/\epsilon\delta)/\epsilon^2\right)$ (when using the min-max proof for the hard-core theorem). However, a more efficient version (in terms of queries) of the hard-core theorem is given in [KS], and using it one can push the number of queries to $q = O(\log(1/\delta)/\epsilon^2)$.

[3] We remark that the proof that appears in [GNW] does not directly achieve this. However, several researchers have independently observed that this is possible via a simple modification. We also mention that an unpublished manuscript [SVW] gives an alternative proof of the direct-product lemma that is also implementable by constant-depth circuits.

Finally, we point out that the techniques in this paper show that $q = \Omega(\log(1/\delta)/\epsilon)$ queries are necessary for black-box proofs of the direct-product lemma (details omitted), which again matches the upper bound in [GNW].

Our main result: The general form. We now state our main result that hardness amplification requires majority in its full generality. Previously, we had stated a corollary of it that was tailored to circuit depth (Theorem 1.3). The general form of our results shows that the circuits $\mathcal{D}$ in a black-box hardness amplification $(Amp, \mathcal{D})$ can be used to compute the majority function by a small constant-depth circuit. The way in which we are going to use a circuit $D \in \mathcal{D}$ is simple and explained next. First, let us remark that since the circuit makes non-adaptive oracle queries, for a fixed $x \in \{0,1\}^k$ the output of $D^h(x)$ is a function $D_x : \{0,1\}^q \to \{0,1\}$ of q evaluations of h at fixed points $y_1, y_2, \ldots, y_q \in \{0,1\}^n$ (again, the y_i's depend on x only): $D^h(x) = D_x(h(y_1), \ldots, h(y_q))$. Let us formally state this key definition.

Definition 1.5. *Let $D^h(x)$ be an oracle circuit that makes q non-adaptive queries to its oracle. For a fixed input x we denote by $D_x : \{0,1\}^q \to \{0,1\}$ the function that maps the q oracle answers to the output $D^h(x) \in \{0,1\}$.*

We are going to show that having access to the above functions $D_x : \{0,1\}^q \to \{0,1\}$ for a few distinct $D \in \mathcal{D}$ and $x \in \{0,1\}^k$ is sufficient to compute majority.

Theorem 1.6 (Decoding requires majority)**.** *There is a universal constant $C > 1$ such that the following holds. Let $(Amp, \mathcal{D})$ be a q-query non-adaptive $(1/2 - \gamma) \to (1/2 - \epsilon)$ black-box hardness amplification. Suppose that $q, \log|\mathcal{D}|, 1/\gamma \leq 2^{k/C}$, and $n, k \geq C^2$, and that $\gamma \geq 1/\log(1/\epsilon)$.*

Then there is a circuit of depth C and size $(q/\epsilon)^C$ with oracle access to (at most $(q/\epsilon)^C$ of) the functions $\{D_x : \{0,1\}^q \to \{0,1\}\}_{D \in \mathcal{D}, x \in \{0,1\}^k}$ that computes majority on inputs of length $1/\epsilon$.

To understand the above theorem, let us briefly see how to obtain Theorem 1.3 from it. Suppose that $\mathcal{D}$ consists of circuits of size s and depth d, that $1/2 - \gamma = 1/3$, and that $s \leq 2^{\gamma \cdot k}$ for a suitable universal constant γ. First, we verify that the hypothesis of Theorem 1.6 is satisfied. This is because the circuits in $\mathcal{D}$ make at most $q \leq s \leq 2^{\gamma \cdot k} \leq 2^{k/C}$ queries – where the last inequality holds for a small enough γ – and $|\mathcal{D}| \leq 2^{s^{O(1)}}$ which implies $\log|\mathcal{D}| \leq s^{O(1)} \leq 2^{k/C}$ – where again the last inequality holds for a small enough γ. At this point, observe that the functions $D_x : \{0,1\}^q \to \{0,1\}$ are also computable by circuits of size s and depth d. Substituting these circuits for the oracle gates in the circuit of depth C and size $(q/\epsilon)^C$ given by the above theorem, we obtain a circuit of depth $C \cdot d = O(d)$ and size $(q/\epsilon)^C \cdot s = \mathrm{poly}(s/\epsilon)$ that computes the majority function on inputs of length $1/\epsilon$. As we mentioned earlier, by known lower bounds for the majority function [Hås, Raz, Smo] we obtain Theorem 1.3: $s \geq \exp\left((1/\epsilon)^{\Omega(1/d)}\right)$.

1.4 Related work

The inapplicability of hardness amplification techniques against low-complexity classes seems to have been observed independently by several researchers, and is also pointed out in [Agr] and in [Vio1, Section 10]. The latter paper informally conjectures the main result of this work that proving hardness amplification requires computing majority (Theorem 1.6). A preliminary version of this work [Vio3, Chapter 6] proves the conjecture in the special case where the class $\mathcal{D}$ in Definition 1.2 is small. The main result in this paper addresses for the first time the general case when there is no bound on the size of $\mathcal{D}$. The same preliminary version [Vio3, Chapter 6] also proved a qualitatively weaker lower bound on the number of queries. We note that a recent work by Lu et al. [LTW4] addresses the necessity of both majority and many queries in proofs of Impagliazzo's hard-core set theorem [Imp]. Specifically, [LTW4] introduces two notions of black-box proof of the hard-core set theorem, and shows that one proof cannot be implemented by small constant-depth circuits, and that the other requires many oracle queries. Their arguments only apply to proofs of the hard-core set theorem, whereas our work addresses arbitrary black-box hardness amplification.

We remark that there is a variety of features that it is interesting to study and optimize of q-query $\delta \to (1/2 - \epsilon)$ hardness amplification $(Amp, \mathcal{D})$ with input lengths k, n. We discuss the most relevant ones next.

Optimizing the ratio between k and n: E.g. [BFNW, Imp, IW1, STV]. This is in particular relevant to obtain conclusions such as $P = BPP$ under the assumption that E requires exponential-size circuits [IW1].

Optimizing $|\mathcal{D}|$ = advice = list size: [IW2, STV, TV, Tre1, Tre3, IJK, IJKW]. This is in particular relevant when $\mathcal{D}$ is a class of uniform machines (as opposed to circuits).

Optimizing the number of queries q: [Imp, KS], as well as the literature on locally-decodable codes (see, e.g., [Tre2]). As discussed in Section 1.1, this is particularly relevant to the loss in circuit size incurred by hardness amplification.

The complexity of Amp: [O'D, Tre1, Vio1, Vio2, Tre3, HVV, LTW, LTW3, LTW1, LTW2] This line of research is orthogonal to this paper which studies the complexity of $\mathcal{D}$ and does not place any restriction on Amp. For context, we mention that the complexity of Amp is a key issue when we want to guarantee that the amplified function $Amp(f)$ lies in a specific class whenever the starting function f does. An example of this is the line of work on hardness amplification within NP [O'D, Tre1, HVV, Tre3, LTW3] which started with the remarkable result by O'Donnell [O'D].

Relaxed definitions of hardness amplification: There are other works that study different, less demanding models of hardness amplification which are tailored to important questions such as worst-case to average-case connections within NP [BT2, BT1, Vio2]. These works are incomparable with ours, one key difference being that they impose computational restrictions on the starting function f and the amplified function $Amp(f)$, whereas our results do not.

Finally, we would like to mention that there is a long line of research that is devoted to proving average-case hardness results for circuit classes below TC^0, e.g. [Hås, HMP$^+$, Kli, AB, Bou, GRS, VW]. With a few exceptions (discussed below) this research has been independent of hardness amplification, and our results may be interpreted as a partial explanation for this independence. The work by Klivans [Kli] stands out. Exploiting precisely the fact that computing majority is all that is needed for hardness amplification, Klivans uses a lower bound for constant-depth circuits with *one* majority gate [ABFR] to give an alternative proof of the strong average-case hardness of parity for constant-depth circuits *without* majority gates. We remark that [Kli] does not contradict the results in this paper, but rather matches them by showing that a lower bound for a class with majority gates is sufficient for hardness amplification; see Section 7 for more on the status of lower bounds for constant-depth circuits with few (e.g. one) majority gates.

2 Overview of the proof

In this section we give a high level overview of the ideas that come into the proofs of our main results (Theorems 1.6,1.4). Within this section we allow ourselves to oversimplify and ignore some technicalities; the reader is referred to the formal proofs for precise details.

The Zoom Theorem. Both the result about the necessity of majority (Theorem 1.6) and our lower bound on the number of queries (Theorem 1.4) rely on a theorem that we call "the Zoom Theorem." Let us first recall the setup. We are given a non-adaptive q-query $\delta \to (1/2-\epsilon)$ black-box hardness amplification $(Amp, \mathcal{D})$ where Amp maps functions $f : \{0,1\}^k \to \{0,1\}$ into functions $Amp(f) : \{0,1\}^n \to \{0,1\}$ (think $n = k^{O(1)}$). Recall that $\mathcal{D}$ is a class of oracle circuits and that for any circuit $D \in \mathcal{D}$ and input $x \in \{0,1\}^k$, Definition 1.5 defines a function $D_x : \{0,1\}^q \to \{0,1\}$ which captures the way D uses the answer to its q oracle queries to compute its output.

An informal statement of the Zoom Theorem follows (see Theorem 4.2 for a precise statement).

Informal Theorem 2.1 (Zoom Theorem)**.** *There exists a circuit $D \in \mathcal{D}$ and an input $x \in \{0,1\}^k$ such that there is a function $T : \{0,1\}^q \to \{0,1\}$ of roughly the same complexity as D_x that satisfies:*

1. $\Pr[T(N^1_{1/2}, \dots, N^q_{1/2}) = 1] \geq 0.49$*, where* $(N^1_{1/2}, \dots, N^q_{1/2})$ *is a vector of q independent bits with probability of being* 1 *equal to* $1/2$ *(i.e., the vector is uniform in* $\{0,1\}^q$*).*

2. $\Pr[T(N^1_{1/2-\epsilon}, \dots, N^q_{1/2-\epsilon}) = 1] \leq 2\delta$*, where* $(N^1_{1/2-\epsilon}, \dots, N^q_{1/2-\epsilon})$ *is a vector of q independent bits with probability of being* 1 *equal to* $1/2-\epsilon$*.*

We refer to the distributions $(N^1_{1/2}, \dots, N^q_{1/2})$ and $(N^1_{1/2-\epsilon}, \dots, N^q_{1/2-\epsilon})$ above as "uniform noise" and "bounded noise," respectively. Loosely speaking, the theorem says that T (which has the same complexity as circuits in $\mathcal{D}$) distinguishes between uniform noise and bounded noise.

Usefulness of the Zoom Theorem. Our two main results follow from the Zoom Theorem. On an intuitive level, it seems that the natural way to decide whether a string $w \in \{0,1\}^q$ was chosen according to uniform noise or according to bounded noise is to compute the Hamming weight of w (which we denote by $\mathit{weight}(w)$) and decide according to whether $\mathit{weight}(w) \leq (1/2 - \epsilon/2)q$. Note that if T implements this strategy then it can indeed be used to compute majority. Furthermore note that when implementing this strategy, a Chernoff-style bound shows that $q = O(\log(1/\delta)/\epsilon^2)$ independent variables are sufficient in order to distinguish uniform noise from bounded noise (at rate $1/2 - \epsilon$) with confidence $1 - \delta$. Our bound on the number of queries essentially follows from the fact that this bound on q is tight.

Let us be more precise in explaining how the "necessity of majority" Theorem 1.6 follows from the Zoom Theorem. We would like to argue that T can be used to compute majority on inputs z of length $\ell := 1/\epsilon$. For simplicity, we explain how to use T to accomplish a slightly easier task, namely distinguishing between inputs z with $\mathit{weight}(z) = \ell/2$ and inputs z with $\mathit{weight}(z) = \ell/2 - 1$ (in the formal proof we essentially show that computing majority can be reduced to this simpler task). Given an input $z \in \{0,1\}^\ell$ we generate a string $w \in \{0,1\}^q$ where w_i is obtained by picking a random index $j \in [\ell]$ and setting $w_i = z_j$. In words, each bit in w is filled with a bit from a random position in z. Note that if $\mathit{weight}(z) = \ell/2$ then w is distributed like uniform noise, whereas if $\mathit{weight}(z) = \ell/2 - 1$ then w is distributed like bounded noise, because $\mathit{weight}(z)/\ell = 1/2 - 1/\ell = 1/2 - \epsilon$. It follows that we can use T to distinguish between the two cases (and recall that T has roughly the same complexity as circuits in $\mathcal{D}$).

This key idea was communicated to us by Madhu Sudan.

Finally, we point out that although the above reduction is randomized, at the end we obtain a *deterministic* circuit that computes majority. For this we also exploit that the relevant probabilities in the above reduction are sufficiently bounded away that they can be amplified using circuits of constant-depth by the result [Ajt1] (see also [Ajt2, Vio5]).

2.1 Proving the Zoom Theorem when $\mathcal{D}$ contains a single circuit

The proof of the Zoom Theorem is the main technical contribution of this paper. What makes this problem challenging is that the class $\mathcal{D}$ can be very large (e.g. $|\mathcal{D}| = \exp(k)$). We explain how we handle such large $\mathcal{D}$ later on. As a warm-up, we outline of the argument in the case that $\mathcal{D}$ contains only one circuit D. We consider a probability space with four independent random variables:

- A uniformly chosen function $F : \{0,1\}^k \to \{0,1\}$. We think of F as the original hard function.
- An input $X \in \{0,1\}^k$ that is uniformly distributed. We think of X as a random input to F.
- A uniformly chosen function $UN : \{0,1\}^n \to \{0,1\}$. We refer to UN as "uniform noise function."

- A function $BN : \{0,1\}^n \to \{0,1\}$ where for every $y \in \{0,1\}^n$, $BN(y)$ is an independent bit with probability of being 1 equal to $1/2 - \epsilon$. We refer to BN as "bounded noise function."

We first consider the setting in which D is run with oracle $Amp(F) \oplus UN$. (This is an oracle that on input $y \in \{0,1\}^n$ returns $Amp(F)(y) \oplus UN(y)$). Note that the uniform noise function UN "masks out" the values of $Amp(F)$ and therefore the circuit D receives no information about F. Thus, D cannot possibly compute a function that is correlated with F:

$$\Pr\left[D^{Amp(F)\oplus UN}(X) \neq F(X)\right] = \Pr\left[D^{UN}(X) \neq F(X)\right] \geq 0.49. \tag{1}$$

We also consider the setting in which D is run with oracle $Amp(F) \oplus BN$. (This is an oracle that on input $y \in \{0,1\}^n$ returns $Amp(F)(y) \oplus BN(y)$). Since BN corresponds to bounded noise at rate $1/2 - \epsilon$, we have that this oracle agrees with $Amp(F)$ on a $(1/2 + \epsilon)$ fraction of inputs and therefore, by the definition of black-box hardness amplification:

$$\Pr\left[D^{Amp(F)\oplus BN}(X) \neq F(X)\right] \leq \delta. \tag{2}$$

Intuitively, the inequalities (1), (2) are going to translate into the two items of the Zoom Theorem. We now explain this part of the argument. Let us examine the computation of D on an input $x \in \{0,1\}^k$ with the two different oracles: In both cases D prepares the same q queries $y_1, \ldots, y_q \in \{0,1\}^n$ to the oracle and receives answers $a_1, \ldots, a_q$ from the oracle. It then outputs $D_x(a_1, \ldots, a_q)$. The high level idea is that when run on random $X \in \{0,1\}^k$, D_X distinguishes between the two oracles and therefore distinguishes between uniform noise and bounded noise. More precisely, by an averaging argument we can fix the random variables F and X and obtain a fixed function T that essentially equals D_X and distinguishes between bounded noise and uniform noise.

2.2 Extending the argument to the case when $\mathcal{D}$ is large

We would like to imitate the proof above when the class $\mathcal{D}$ contains many circuits. For concreteness let us assume that $\mathcal{D}$ contains 2^{k^2} circuits, i.e. $|\mathcal{D}| = \exp(k^2)$. In this general case the definition of black-box hardness amplification only says that for any choice of f, h where h agrees with $Amp(f)$ on a $(1/2+\epsilon)$ fraction of inputs *there exists* a circuit $D \in \mathcal{D}$ such that D^h agrees with f on a $1-\delta$ fraction of inputs. Note that the circuit D is a function of both f and h, and let us denote this function by *circuit*(f, h).

We would like to imitate the previous argument. However, when we use oracle $Amp(F) \oplus BN$, we do not know which circuit $D \in \mathcal{D}$ is the "correct circuit", i.e. *circuit*$(F, Amp(F) \oplus BN)$. More formally, we have that *circuit*$(F, Amp(F) \oplus BN)$ is a random variable that in particular depends on BN. In the previous argument we applied a *fixed* function D_x on the answers $a_1, \ldots, a_q$ that were returned by the oracle. However, the function D_x for $D =$ *circuit*$(F, Amp(F) \oplus BN)$ that we want to apply on $a_1, \ldots, a_q$ is now a random variable that *depends* on $a_1, \ldots, a_q$ and we cannot use the argument above.

Going back to the case of a single circuit. To avoid the aforementioned problem we start by fixing the random variable $\mathit{circuit}(F, \mathit{Amp}(F) \oplus BN)$ to its most likely value. That is, let D be the most likely value of $\mathit{circuit}(F, \mathit{Amp}(F) \oplus BN)$ and let $E = E(F, BN)$ be the event

$$E := \{\mathit{circuit}(F, \mathit{Amp}(F) \oplus BN) = D\}\,.$$

Note that the probability of E is at least $1/|\mathcal{D}| = 2^{-k^2}$ (which is small but not *too* small). We have that $\mathit{circuit}(F, \mathit{Amp}(F) \oplus BN)$ is *fixed* in E (which means that in E we only need to consider *one fixed circuit* D). From now on we restrict our attention to E. That is, let F', BN' denote the distribution of F, BN when conditioned on the event E. Note that this conditioning can skew the distribution of F', BN' and that these variables are no longer distributed like the original variables F, BN and in particular may become dependent. For the purpose of explaining the argument let us assume the unjustified assumption that F' and BN' are independent. (In the actual argument we bypass this problem by fixing F to some fixed function f before conditioning on the event E).

We would like to imitate the argument of the previous section in this new probability space. Indeed, we are back to dealing with one fixed circuit D. However, the previous argument critically relies on properties of BN: Most notably that for any $y_1, \ldots, y_q \in \{0,1\}^n$, the random variable $(BN(y_1), \ldots, BN(y_q))$ is distributed like bounded noise. This may not necessarily hold for BN'.

An information-theoretic lemma. In order to handle this problem, we use the following Lemma (stated informally; cf. Section 3 for a precise statement).

Informal Lemma 2.2. *Let $V_1, \ldots, V_t$ be independent and identically distributed random variables. Let E be an event whose probability is "not too small." Then for any integer q there exists a "large" set $G \subseteq [t]$ such that for every $i_1, \ldots, i_q \in G$, the distribution $(V_{i_1}, \ldots, V_{i_q})$ "does not change significantly" when conditioning on E.*

This lemma can be viewed as a generalization of a Lemma by Raz (in which $q = 1$) that is used in his parallel repetition theorem [Raz]. We have recently found out that this lemma follows easily from the results in [EIRS, Section 4].

We apply the lemma on the random variables $\{BN(y)\}_{y \in \{0,1\}^n}$. We conclude that there exists a large set $G \subseteq \{0,1\}^n$ such that for any $y_1, \ldots, y_q \in G$ the variable

$$(BN'(y_1), \ldots, BN'(y_q))$$

is statistically close to

$$(BN(y_1), \ldots, BN(y_q)).$$

This lemma intuitively helps us recover the previous argument in the new probability space: We consider the operation of $D^{\mathit{Amp}(F') \oplus BN'}$ on an input $x \in \{0,1\}^k$. If the queries $y_1, \ldots, y_q \in \{0,1\}^n$ that D makes are all in the "good set" G, then the rest of the proof essentially goes through. This is because on these q queries the distribution of the bounded noise function is statistically close to its initial distribution and we can continue with the previous argument.

However, even though the set G of "good queries" is large, it may be the case that on every input $x \in \{0,1\}^k$, D makes a "bad query" $y' \notin G$. We have no control on the distribution $BN'(y')$ when $y' \notin G$; for example, it may be correlated with the value of BN' on another query y, and so we cannot relate this distribution to that of bounded noise (in which different coordinates are independent and distributed in the same way).

Fixing bad queries. In order to address this issue we further refine the probability space by fixing the value of BN' at some bad queries. The high level idea is that by fixing the bounded noise function on these queries we "remove dependencies" between the answers that the circuit D sees when making its queries. This part of the argument is more technical and we will not describe it in detail. However, we point out that fixing bad queries is a tricky business as whenever we fix a bad query we change the probability space, which in turn skews the distribution of the bounded noise function and may result in introducing new bad queries (and it seems that we make no progress as we can never fully get rid of bad queries). In the actual argument we fix the bounded noise function only on those queries that are *heavy* in the sense that they are "asked frequently" by D. The rationale is that even if fixing the bounded noise function on these queries skews the distribution and introduces new bad queries we do make progress as the new bad queries are queries that are not asked frequently by D. Finally, we argue that bad queries that are not asked frequently by D do not hurt us too much when implementing the initial argument (because on an intuitive level, this means that D asks good queries "most of the time").

One technical point that we want to make is that for implementing the approach above we must make sure that the number of bad queries that are introduced after fixing the frequent queries does not depend on the number of frequent queries that we fix. This is because in the actual argument we do a union bound over all bad queries and argue that the probability that a random input queries *any* bad query (that is not already fixed) is low. This allows us to ignore bad queries as the weight of inputs which query bad queries is small.

2.3 Organization of the paper

In Section 3 we state and prove the information-theoretic lemma (Informal Lemma 2.2). In Section 4 we state and prove the Zoom Theorem which is the main technical theorem of this paper. In Section 5 we show how our result on necessity of majority follows from the Zoom Theorem. In Section 6 we show how our lower bound on the number of queries follows from the Zoom Theorem. In Section 7 we explain the significance of our results to various circuit classes. In Section 8 we explain that our results can be viewed as lower bounds on the complexity of decoding locally (list-)decodable codes. Finally, Section 9 discusses some open problems.

3 The information-theoretic lemma

In this section we prove the following lemma which, loosely speaking, says that if one conditions uniformly distributed random variables $V_1, \ldots, V_t$ on an event that happens with noticeable probability, then even following the conditioning most groups of q variables are close to being *jointly* uniformly distributed. We need the following definition.

Definition 3.1. *We say that two random variables V, W over the same set S are ϵ-close if for every event $E \subseteq S$, $|\Pr[V \in E] - \Pr[W \in E]| \leq \epsilon$.*

Given a random variable V over a set S and an event E we use $(V|E)$ to denote the probability distribution of V conditioned to E, that is for any event $A \subseteq E$, $\Pr_{(V|E)}[A] = \Pr[V \in A | V \in E]$.

We are now ready to state the Lemma.

Lemma 3.2. *Let $V = (V_1, \ldots, V_t)$ be a collection of independent random variables where each one of them is uniformly distributed over a set S. Let $A \subseteq S^t$ be an event such that $\Pr[V \in A] \geq 2^{-a}$. Then for any $\eta > 0$ and integer q there exists a set $G \subseteq [t]$ such that $|G| \geq t - 16 \cdot q \cdot a / \eta^2$ and for any $i_1, \ldots, i_q \in G$ the distribution $(V_{i_1}, \ldots, V_{i_q} | V \in A)$ is η-close to uniform.*

Lemma 3.2 can be viewed as a generalization of a Lemma by Raz [Raz, Section 3] that implies Lemma 3.2 for the special case of $q = 1$. We mention again that we have recently found out that Lemma 3.2 follows easily from results in [EIRS, Section 4].

We now discuss the proof of Lemma 3.2. The proof relies on the notion of *entropy* H of a random variable X, defined as $H(X) := \sum_x \Pr[X = x] \cdot \log(1/\Pr[X = x])$ (cf. [CT, Chapter 2]). We list next a few standard properties of entropy that we will use in the proof.

Fact 1. *Entropy satisfies the following.*

1. Chain rule: *For any random variables $X_1, \ldots, X_n$ we have*

$$H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i | X_{i-1}, \ldots, X_1)$$

[CT, Theorem 2.5.1].

2. Conditioning reduces entropy: *For any random variables X, Y, Z we have $H(X|Y) \geq H(X|Y, Z)$ (follows easily from the definition).*

3. High entropy implies uniform: *Let V be a random variable taking values in a set S and suppose that $H(V) \geq \log|S| - \alpha$; then V is $4\sqrt{\alpha}$-close to uniform [CK, Chapter 3; Exercise 17].*

We now prove Lemma 3.2

Proof of Lemma 3.2. Let $n := \log|S|$ and let $V' := (V'_1,\ldots,V'_t) := (V_1,\ldots,V_t|V_1,\ldots,V_t \in A)$. Note that V' is uniformly distributed over A and therefore $H(V') = \log|A| = \log(2^{n\cdot t}\cdot \Pr[V\in A]) \geq n\cdot t - a$. By the chain rule (Item (1) in Fact 1) for entropy we have that:

$$n\cdot t - a \leq H(V') = H(V'_1,\ldots,V'_t) = \sum_{1\leq i\leq t} H(V'_i|V'_1,\ldots,V'_{i-1}). \tag{3}$$

Let

$$\ell_i = H(V'_i|V'_1,\ldots,V'_{i-1}).$$

By (3) we have that $\frac{1}{t}\cdot\sum_{1\leq i\leq t}\ell_i \geq n - a/t$. Let $b := \eta^2\cdot t/(16q\cdot a)$. By a Markov argument at most t/b of the indices i are such that $\ell_i < n - b\cdot a/t$. Let

$$G := \{i : \ell_i \geq n - b\cdot a/t\}.$$

We have that $|G| \geq t - t/b = t - 16q\cdot a/\eta^2$. Let $i_1 < i_2 < \ldots < i_q$ be arbitrary indices in G. By the chain rule for entropy:

$$H(V'_{i_1},\ldots,V'_{i_q}) = \sum_{1\leq j\leq q} H(V'_{i_j}|V'_{i_1},\ldots,V'_{i_{j-1}}).$$

We now use the fact that "conditioning reduces entropy," i.e. Item (2) in Fact 1, and deduce that:

$$H(V'_{i_1},\ldots,V'_{i_q}) \geq \sum_{1\leq j\leq q} H(V'_{i_j}|V'_1,\ldots,V'_{i_j-1}) = \sum_{1\leq j\leq q}\ell_{i_j} \geq q\cdot n - \frac{q\cdot b\cdot a}{t}.$$

We have that $(V'_{i_1},\ldots,V'_{i_q})$ is a random variable taking values in a set of size $2^{q\cdot n}$ and that its entropy is at least $q\cdot n - \frac{q\cdot b\cdot a}{t}$. By Item (3) in Fact 1 it then follows that $(V'_{i_1},\ldots,V'_{i_q})$ is $4\sqrt{\frac{q\cdot b\cdot a}{t}}$-close to uniform. To conclude, note that $4\sqrt{\frac{q\cdot b\cdot a}{t}} = \eta$. □

4 Statement and proof of the Zoom Theorem 4.2

Both our result about the necessity of majority (Theorem 1.6) and our lower bound on the number of queries (Theorem 1.4) rely on the following Zoom Theorem which is our main technical contribution. This theorem shows that given a non-adaptive q-query black-box $\delta \rightarrow (1/2-\epsilon)$ hardness amplification $(Amp,\mathcal{D})$ we can "zoom in" on a particular function $D_x : \{0,1\}^q \rightarrow \{0,1\}$, where $D \in \mathcal{D}, x \in \{0,1\}^k$ (cf. Definition 1.5 for the definition of D_x) that is distinguishing noise rate $1/2$ from noise rate $1/2-\epsilon$. The distinguisher will not quite be a function D_x but rather (a distribution on) *projections* of such functions, which are simply functions that can be obtained from D_x by fixing some input variables to constants and complementing others. We give the formal definition of a projection and then we state the zoom theorem.

Definition 4.1. *Let $d = d(y_1, \ldots, y_q) : \{0,1\}^q \to \{0,1\}$ be a function. A* projection *of d is a function $d' : \{0,1\}^q \to \{0,1\}$ that can be obtained from d by fixing some input variables to constants and complementing others, and possibly complementing the output. Formally, there are $a_1, \ldots, a_q, b_1, \ldots, b_q, c \in \{0,1\}$ such that for any $y_1, \ldots, y_q \in \{0,1\}$, $d'(y_1, \ldots, y_q) = d((y_1 \cdot a_1) \oplus b_1, \ldots, (y_q \cdot a_q) \oplus b_q) \oplus c$.*

Theorem 4.2 (Zoom theorem)**.** *There is a universal constant $C > 1$ such that the following holds. Let $(Amp, \mathcal{D})$ be a non-adaptive q-query $\delta \to (1/2 - \epsilon)$ hardness amplification scheme. Suppose that $q, \log|\mathcal{D}| \le 2^{k/C}$, and $n, k \ge C^2$.*

Then there is a distribution T on functions $t : \{0,1\}^q \to \{0,1\}$ such that

1. *$\Pr_{T, N^1_{1/2}, \ldots, N^q_{1/2}}[T(N^1_{1/2}, \ldots, N^q_{1/2}) = 1] \ge 1/2 - 2^{-k/C}$, where $(N^1_{1/2}, \ldots, N^q_{1/2})$ is a vector of q independent bits with probability of being 1 equal to $1/2$ (i.e., the vector is uniform in $\{0,1\}^q$),*

2. *$\Pr_{T, N^1_{1/2-\epsilon}, \ldots, N^q_{1/2-\epsilon}}[T(N^1_{1/2-\epsilon}, \ldots, N^q_{1/2-\epsilon}) = 1] \le \delta + 2^{-k/C}$, where $(N^1_{1/2-\epsilon}, \ldots, N^q_{1/2-\epsilon})$ is a vector of q independent bits with probability of being 1 equal to $1/2 - \epsilon$, and*

3. *each $t \in T$ is a projection of a function D_x for some $D \in \mathcal{D}$ and $x \in \{0,1\}^k$. I.e., every $t \in T$ can be obtained from D_x for some $D \in \mathcal{D}, x \in \{0,1\}^k$ by fixing some input variables to constants and complementing others, and possibly complementing the output.*

4.1 Proof

Let $a := \log|\mathcal{D}|$ and $C > 1$ be a constant to be determined later. Recall that we are only interested in the case that $q, a \le 2^{k/C}$, and $n, k \ge C^2$. We will assume that his holds throughout the proof. In various places in the proof we will want certain inequalities to hold and will observe that each one of them holds for a sufficiently large constant $C > 1$. In the end, we will choose C to be sufficiently large so that all the conditions throughout the proof hold simultaneously.

Let F be the uniform distribution on functions $f : \{0,1\}^k \to \{0,1\}$. We now would like a distribution on "noise functions" that perturbs each bit with probability $1/2 - \epsilon$. For the later application of Lemma 3.2, which deals with random variables uniformly distributed, it is convenient to adopt the following approach to define our noise function. Let M be the uniform distribution on functions mapping $\{0,1\}^n$ to $[\epsilon^{-1}]$; let $\Theta : [\epsilon^{-1}] \to \{0,1\}$ be the function such that $\Theta(y) = 1$ if and only if $y \le \epsilon^{-1}/2 - 1$.[4] We think of our noise function as $\Theta \circ M : \{0,1\}^n \to \{0,1\}$, which indeed satisfies $\Pr_M[(\Theta \circ M)(y) = 1] = (\epsilon^{-1}/2 - 1)/\epsilon^{-1} = 1/2 - \epsilon$ for every $y \in \{0,1\}^n$. For readability, we simply write ΘM for $\Theta \circ M$ and also $\Theta M(y)$ for $\Theta(M(y))$.

We think of the circuits as trying to compute F from the oracle $Amp(F) \oplus \Theta M$, whose value at y is $Amp(F)(y) \oplus \Theta M(y)$. We also think of ΘM as corrupting at most a $1/2 - \epsilon$

[4] If $\epsilon^{-1}/2$ is not an integer we can replace $[\epsilon^{-1}]$ with $[A]$ for a sufficiently large integer A and carry through a similar argument: Our argument is not affected by the magnitude of A.

fraction of the values of $Amp(F)$, namely those for which $\Theta M(y) = 1$, but for this we have to deal with the technicality that our Definition 1.2 of black-box hardness amplification puts $1/2 - \epsilon$ as a sharp threshold for the noise, whereas with some probability ΘM will map more than a $1/2 - \epsilon$ fraction of the inputs to 1. However, a loose bound, which is sufficient for our purposes, shows that ΘM will indeed map at most a $1/2 - \epsilon$ fraction of the inputs to 1 with probability at least $1/2^n$.[5]

By assumption and the above, we have that

$$\Pr_{F,M}\left[\exists D \in \mathcal{D} : \Pr_{x \in \{0,1\}^k}[D^{Amp(F) \oplus \Theta M}(x) \neq F(x)] \leq \delta\right] \geq 2^{-n}. \tag{4}$$

It is convenient for the rest of the proof to introduce the following notation for the error made by a circuit $D \in \mathcal{D}$ on computing a function $f : \{0,1\}^k \to \{0,1\}$ from an oracle $g : \{0,1\}^n \to \{0,1\}$:

$$\mathit{\Delta D}\,(f, g) := \Pr_{x \in \{0,1\}^k}[D^g(x) \neq f(x)].$$

This lets us rewrite Equation (4) as

$$\Pr_{F,M}\left[\exists D \in \mathcal{D} : \mathit{\Delta D}\,(F, Amp(F) \oplus \Theta M) \leq \delta\right] \geq 2^{-n}. \tag{5}$$

From Equation (5) we see that there must exist a *fixed* circuit $D \in \mathcal{D}$ such that

$$\Pr_{F,M}\left[\mathit{\Delta D}\,(F, Amp(F) \oplus \Theta M) \leq \delta\right] \geq 2^{-a} \cdot 2^{-n} \geq 2^{-2 \cdot a}, \tag{6}$$

where the last inequality holds for $a := \log |D| \geq n$, which we can assume without loss of generality.

For the rest of the proof we fix $D \in \mathcal{D}$ to the particular circuit given by the above Equation (6). Since D is non-adaptive, the queries $y \in \{0,1\}^n$ made by D depend only on the input x. Again, note here we crucially use the non-adaptivity of D. Let us now call a query $y \in \{0,1\}^n$ *heavy* if a $\tau := \sqrt{2^{-k}}$ fraction of the x's query y. That is, we make the following definition.

Definition 4.3. *We say that $y \in \{0,1\}^n$ is* heavy *if*

$$\Pr_{x \in \{0,1\}^k}[D^g(x) \text{ queries } g(y)] \geq \tau := \sqrt{2^{-k}}.$$

Claim 4.3.1. *There are at most $h := q/\tau$ heavy queries $y \in \{0,1\}^n$.*

[5] Specifically, the probability that ΘM maps exactly a $1/2 - \epsilon$ fraction of its inputs to 1 is at least $1/\sqrt{8 \cdot 2^n \cdot (1/2 - \epsilon) \cdot (1/2 + \epsilon)}$ (see, e.g., [CT, Lemma 17.5.1]), which is at least $1/2^n$ for sufficiently large $n \geq C$. (Again, we assume without loss of generality that $(1/2 - \epsilon) \cdot 2^n$ is an integer.) This argument will let us derive a conclusion which involves noise rate $1/2 - \epsilon$, as opposed to $1/2 - O(\epsilon)$ which can be obtained by modifying the definition of Θ and using a Chernoff Bound instead of the above bound.

Proof.

$$q \geq E_{x\in\{0,1\}^k}[\text{number of queries made by } D(x)] \geq (\text{number of heavy } y \in \{0,1\}^n) \cdot \tau.$$

□

Let $r_1, \ldots, r_h \in \{0,1\}^n$ be the heavy queries, and let us call *light* the other queries. Conditioning on the values $m_1, \ldots, m_h \in [\epsilon^{-1}]$ of M on the heavy queries $r_1, \ldots, r_h \in \{0,1\}^n$, we can write Equation (6) as follows

$$E_{m_1,\ldots,m_h\in[\epsilon^{-1}]}\left[\Pr_{F,M}\left[\mathit{\Delta D}\left(F, \mathit{Amp}(F) \oplus \Theta M\right) \leq \delta \middle| M(r_1) = m_1, \ldots, M(r_h) = m_h\right]\right] \geq 2^{-2\cdot a}. \tag{7}$$

Therefore, by an averaging argument, we see that there exists a fixing

$$M(r_1) = m_1, \ldots, M(r_h) = m_h$$

of the values of M on the heavy queries such that, if we denote by $\bar{M}$ the distribution

$$\bar{M} := M \,\Big|\, M(r_1) = m_1, \ldots, M(r_h) = m_h,$$

it still holds that

$$\Pr_{F,\bar{M}}\left[\mathit{\Delta D}\left(F, \mathit{Amp}(F) \oplus \Theta\bar{M}\right) \leq \delta\right] \geq 2^{-2\cdot a}. \tag{8}$$

Now we would like to fix a particular choice for the function $F = f$ that simultaneously accomplishes two things. First, D should still sufficiently often compute f well over random $\bar{M}$ (as in Equation (8)), and second D should *not* compute f well when given access to only the values of $\mathit{Amp}(f)$ on the h heavy y's in $\{0,1\}^n$. The next claim gives such an f. To formalize our second point, namely the inability of D to compute f well when given access only to the values of $\mathit{Amp}(f)$ on the heavy y's, let us make the following definition.

Definition 4.4. *We denote by $\bar{N}_{1/2} : \{0,1\}^n \to \{0,1\}$ a uniformly distributed random function consistent with our fixing of the noise on the heavy queries $r_1, \ldots, r_h$, i.e. $\bar{N}_{1/2}(r_i) := \Theta\bar{M}(r_i) = \Theta(m_i) \in \{0,1\}$ for every $i \leq h$.*

Note that $\bar{N}_{1/2}(y)$ is defined to be a random bit if y is light, thereby hiding the value of $\mathit{Amp}(f) \oplus \bar{N}_{1/2}$ on y.

Claim 4.4.1. *There exists $f : \{0,1\}^k \to \{0,1\}$ such that both the following claims are true:*

1. $\Pr_{\bar{M}}\left[\mathit{\Delta D}\left(f, \mathit{Amp}(f) \oplus \Theta\bar{M}\right) \leq \delta\right] \geq 2^{-2\cdot a}/2$, *and*

2. $E_{\bar{N}_{1/2}}\left[\mathit{\Delta D}\left(f, \mathit{Amp}(f) \oplus \bar{N}_{1/2}\right)\right] \geq 1/2 - 2^{-k/C}$.

The proof of Claim 4.4.1 is a relatively standard counting argument that is deferred to the end of this section.

We now refine our probability space as follows. First, we fix $F = f$ as given by Claim 4.4.1. Second, we condition $\bar{M}$ on the event that D is successful, namely

$$M' := \bar{M} \mid \mathit{\Delta D}\left(f, \mathit{Amp}(f) \oplus \Theta\bar{M}\right) \leq \delta.$$

From now on we look more locally to the action of D, and specifically we consider the output of D on input x as the function $D_x : \{0,1\}^q \to \{0,1\}$ that maps the output of the q queries to the output bit (cf. Definition 1.5). It is convenient to introduce the following shorthands. For $x \in \{0,1\}^k$ we denote by $Q(x) \in (\{0,1\}^n)^q$ the vector $(y_1, \ldots, y_q)$ of the q queries $y_1, \ldots, y_q \in \{0,1\}^n$ made by D on input x. For a function $g : \{0,1\}^n \to \{0,1\}$ and $Q(x) = (y_1, \ldots, y_q) \in (\{0,1\}^n)^q$ we write $g(Q(x)) \in \{0,1\}^q$ for the vector of the q evaluations of g on the q vectors in $Q(x)$. For example, $\mathit{Amp}(f)(Q(x))$ is the vector of the q evaluations of $\mathit{Amp}(f)$ on the queries made by D on input x.

Let us now proceed with our proof. We can express the success in computing f from noise M' using the functions D_x and the above notation as follows (where we are just using the above definition of M' and the fixing of f given by Claim 4.4.1):

$$\Pr_{M', x \in \{0,1\}^k}\left[D_x\Big(\mathit{Amp}(f)(Q(x)) \oplus \Theta M'(Q(x))\Big) \neq f(x)\right] \leq \delta. \tag{9}$$

We now would like to assert that D also computes f well when $\mathit{Amp}(f)$ is perturbed with $\Theta\bar{M}$, as opposed to $\Theta M'$ in Equation (9), where recall $\Theta\bar{M}$ is independent noise at rate $1/2 - \epsilon$ on the light queries, and is fixed on the heavy queries. In other words, we want to replace M' with $\bar{M}$ in Equation (9). We argue this by showing that M' and $\bar{M}$ look locally the same. This is the step of the proof where we make use of the information-theoretic Lemma 3.2. Specifically, let $\ell := 2^n - h$ be the number of light y's in $\{0,1\}^n$, and let $(z_1, \ldots, z_\ell)$ be an enumeration of such y's. Now consider the vector of ℓ random variables

$$(\bar{M}(z_1), \ldots, \bar{M}(z_\ell)),$$

(which is just a vector of uniform and independent random variables in $[\epsilon^{-1}]$). Now recall that M' is defined as $\bar{M}$ conditioned on the event $A :=$ "$\mathit{\Delta D}\left(f, \mathit{Amp}(f) \oplus \Theta\bar{M}\right) \leq \delta$," and that $\Pr[A] \geq 2^{-2 \cdot a}/2$ by Item (1) in Claim 4.4.1. Therefore we can apply Lemma 3.2 with

$$\eta := \frac{2^{-k/C}}{2} \tag{10}$$

to conclude that there is a set $G \subseteq \{0,1\}^n$ of size

$$|G| \geq \ell - 16q \cdot (2a+1)/\eta^2 \tag{11}$$

such that for any $i_1, \ldots, i_q \in G$, the vector

$$(M'(i_1), \ldots, M'(i_q)) \text{ is } \eta\text{-close to } (\bar{M}(i_1), \ldots, \bar{M}(i_q)), \tag{12}$$

where the distance is in statistical difference.

Let us now call x *good* if D, on input x, only makes queries that are either heavy or in G.

Definition 4.5. *An input $x \in \{0,1\}^k$ is* good *if all the q queries $y_1, \ldots, y_q \in \{0,1\}^n$ made by $D(x)$ are either heavy or in the set G.*

Since M' and $\bar{M}$ agree on the heavy queries by definition, we have by Equation (12) that for every good $x \in \{0,1\}^k$ D performs as well with noise $\Theta M'$ as with noise $\Theta \bar{M}$. Specifically, we have the following inequality:

$$\Big| \Pr_{M'}\left[D_x(Amp(f)(Q(x)) \oplus \Theta M'(Q(x))) \neq f(x)\right] - \\ \Pr_{\bar{M}}\left[D_x(Amp(f)(Q(x)) \oplus \Theta \bar{M}(Q(x))) \neq f(x)\right] \Big| \leq \eta \qquad \left(\text{recall } \eta = \frac{2^{-k/C}}{2}\right). \quad (13)$$

We now observe that most x's in $\{0,1\}^k$ are good. We have

$$\begin{aligned}
\Pr_x\left[x \text{ is not good}\right] &= \Pr_x\left[\exists y \in \{0,1\}^n : D(x) \text{ queries } y \text{ and } y \text{ is both light and not in } G\right] \\
&\leq (\ell - |G|) \max_{y \text{ light}} \Pr_x\left[D(x) \text{ queries } y\right] \\
&\leq \frac{16 \cdot q \cdot (2 \cdot a + 1)}{\eta^2} \cdot \tau \qquad \text{(By Equation (11) and the Definition 4.3 of heavy.)} \\
&\leq \frac{16 \cdot 2^{k/C} \cdot 3 \cdot 2^{k/C}}{2^{-2 \cdot k/C}/4} \cdot 2^{-k/2} \\
&\qquad \text{(By our assumption that } a, q \leq 2^{k/C}\text{, and Definitions (10) and (4.3).)} \\
&\leq \frac{2^{-k/C}}{2}. \qquad \text{(For sufficiently large } C\text{.)} \qquad (14)
\end{aligned}$$

Note in the above calculation we make crucial use of the fact that the size of G is independent from our choice of the threshold τ in the definition of heavy queries.

Using the above bound on the probability that x is good we now have

$$\begin{aligned}
&\Pr_{\bar{M},x}\left[D_x(Amp(f)(Q(x)) \oplus \Theta \bar{M}(Q(x))) \neq f(x)\right] \\
&\leq \Pr_{\bar{M},x}\left[D_x(Amp(f)(Q(x)) \oplus \Theta \bar{M}(Q(x))) \neq f(x) \bigwedge x \text{ is good}\right] + \Pr_x[x \text{ is not good}] \\
&\leq \Pr_{M',x}\left[D_x(Amp(f)(Q(x)) \oplus \Theta M'(Q(x))) \neq f(x) \bigwedge x \text{ is good}\right] + \frac{2^{-k/C}}{2} + \frac{2^{-k/C}}{2} \\
&\qquad \text{(By Equations (13) and (14).)} \\
&\leq \Pr_{M',x}\left[D_x(Amp(f)(Q(x)) \oplus \Theta M'(Q(x))) \neq f(x)\right] + 2^{-k/C} \\
&\leq \delta + 2^{-k/C} \qquad \text{(By Equation (9).)} \qquad (15)
\end{aligned}$$

The desired distribution $T : \{0,1\}^q \to \{0,1\}$ on functions in the conclusion of the theorem is defined as follows. On input $z \in \{0,1\}^q$, pick a random $x \in \{0,1\}^k$, and let $Q(x) = (y_1, \ldots, y_q) \in (\{0,1\}^n)^q$ be the vector of the q queries made by $D(x)$. Output $T(z) := 1$ if and only if $D_x(Amp(f)(Q(x)) \oplus \bar{z}) \neq f(x)$ where $\bar{z}$ is defined as follows: the i-th bit $\bar{z}_i$ equals z_i if y_i is light, and otherwise is set to $\Theta\bar{M}(y_i)$. To conclude the proof, we verify the properties of T claimed in the statement of the theorem:

1. By Item (2) in Claim 4.4.1, $\Pr_{T,N^1_{1/2},\ldots,N^q_{1/2}}[T(N^1_{1/2}, \ldots, N^q_{1/2}) = 1] \geq 1/2 - 2^{-k/C}$;
2. by Equation (15), $\Pr_{T,N^1_{1/2-\epsilon},\ldots,N^q_{1/2-\epsilon}}[T(N^1_{1/2-\epsilon}, \ldots, N^q_{1/2-\epsilon}) = 1] \leq \delta + 2^{-k/C}$;
3. by definition of T, each $t \in T$ is a projection of a function D_x for some x, or its complement.

This concludes the proof of the theorem except for the proof of Claim 4.4.1, which is given next.

Proof of Claim 4.4.1. From Equation (8) and a Markov argument we obtain

$$\Pr_F \left[\Pr_{\bar{M}} \left[\mathit{\Delta D} \left(F, Amp(F) \oplus \Theta\bar{M} \right) \leq \delta \right] \geq 2^{-2\cdot a}/2 \right] \geq 2^{-2\cdot a}/2. \tag{16}$$

Thus, a random F satisfies Item (1) of the claim with probability at least $2^{-2\cdot a}/2$. To conclude the proof, we show by a counting argument that a random F satisfies Item (2) of the claim with probability bigger than $1 - 2^{-2\cdot a}/2$. This guarantees the existence of a function f that satisfies both items.

We now proceed with the counting argument. For this, let

$$\beta := \frac{2^{-k/C}}{2}$$

and define B to be the set of functions $g : \{0,1\}^k \to \{0,1\}$ such that the probability over $\bar{N}_{1/2}$ that D differs from g on less than a $1/2 - \beta$ fraction of the inputs is at least β. We intuitively think of B as the set of functions for which, given random noise, unreasonably often D computes g well on average. Formally:

$$B := \left\{ g : \Pr_{\bar{N}_{1/2}} \left[\mathit{\Delta D} \left(g, Amp(g) \oplus \bar{N}_{1/2} \right) \leq 1/2 - \beta \right] \geq \beta \right\}.$$

Note that if $f \notin B$ then f satisfies Item (2) by Markov inequality:

$$\begin{aligned} E_{\bar{N}_{1/2}} \left[\mathit{\Delta D} \left(f, Amp(f) \oplus \bar{N}_{1/2} \right) \right] &\geq (1/2 - \beta) \cdot \Pr \left[\mathit{\Delta D} \left(f, Amp(f) \oplus \bar{N}_{1/2} \right) \geq 1/2 - \beta \right] \\ &\geq (1/2 - \beta)(1 - \beta) = (1/2 - 2^{-k/C}/2)(1 - 2^{-k/C}/2) \geq 1/2 - 2^{-k/C}. \end{aligned}$$

Thus, for our goal is enough to show that a random $f \in F$ falls in B with low probability at most $2^{-2\cdot a}/2$. This bound is slightly complicated by the fact that D is getting some information about g, namely its values on the heavy queries. However, since there are only few heavy queries, we can tolerate this information as the following standard argument shows. Consider a uniform random function $V : \{0,1\}^n \to \{0,1\}$. Note that for every g we have that $(Amp(g) \oplus V)(y)$ is distributed like $\bar{N}_{1/2}(y)$ if y is light, whereas if y is heavy we have that, with probability $1/2$ over $V(y)$, $V(y) = (Amp(g) \oplus \bar{N}_{1/2})(y)$. Therefore:

$$B \subseteq B' := \left\{ g : \Pr_V \left[\Delta D\left(g, V\right) \le 1/2 - \beta \right] \ge \beta/2^h \right\}.$$

We now bound the size of B'. Averaging over all g's in B' we see that we can fix a value $v = V$ so that a $\beta/2^h$ fraction of the g's in B' will fall in the set

$$B'' := \left\{ g : \Delta D\left(g, v\right) \le 1/2 - \beta \right\}$$

(here we are using that V is independent of g). By a standard Chernoff Bound (see, e.g. [DP, Theorem 1.1]), the probability that a random function $F : \{0,1\}^k \to \{0,1\}$ falls in B'' is at most $2^{-\beta^2 \cdot 2^k}$. This is just the probability that, tossing 2^k coins, one gets "heads" more than $2^k(1/2 + \beta)$ times, where in our case "heads" means "$F(x) = D^v(x)$." Consequently, we can bound

$$\Pr_F[F \in B] \le \Pr_F[F \in B'] \le 2^h/\beta \cdot \Pr_F[F \in B''] \le 2^h/\beta \cdot 2^{-\beta^2 2^k}.$$

To conclude, we need to verify that

$$2^h/\beta \cdot 2^{-\beta^2 2^k} \le 2^{-2 \cdot a}/2 \;\Leftrightarrow\; 2 \cdot a + 1 + h + \log(1/\beta) \le \beta^2 2^k.$$

Indeed, recalling our assumptions on the parameters $q, a \le 2^{k/C}$, the definition of $\tau := 2^{k/2}$, the bound on $h \le q/\tau \le 2^{k(1/2+1/C)}$ from Claim 4.1, and the definition of $\beta := \frac{2^{-k/C}}{2}$, we can bound

$$2 \cdot a + 1 + h + \log(1/\beta) \le 2 \cdot 2^{k/C} + 1 + 2^{k(1/2+1/C)} + 1 + k/C \le 4 \cdot 2^{k(1-2/C)} = \beta^2 \cdot 2^k,$$

for sufficiently large C (recall $k \ge C^2$). This proves the claim. □

5 Proof of Theorem 1.6 from the Zoom Theorem 4.2

In this section we prove Theorem 1.6. We restate the theorem for the reader's convenience.

Theorem 1.6 (Decoding requires majority)**.** *There is a universal constant $C > 1$ such that the following holds. Let $(Amp, \mathcal{D})$ be a q-query non-adaptive $(1/2 - \gamma) \to (1/2 - \epsilon)$ black-box hardness amplification. Suppose that $q, \log|\mathcal{D}|, 1/\gamma \le 2^{k/C}$, and $n, k \ge C^2$, and that $\gamma \ge 1/\log(1/\epsilon)$.*

Then there is a circuit of depth C and size $(q/\epsilon)^C$ with oracle access to (at most $(q/\epsilon)^C$ of) the functions $\{D_x : \{0,1\}^q \to \{0,1\}\}_{D \in \mathcal{D}, x \in \{0,1\}^k}$ that computes majority on inputs of length $1/\epsilon$.

We present the proof as a series of claims that are based on the zoom theorem and lead to the conclusion of Theorem 1.6. The outline of the sequence of claims is as follows.

1. First, in Claim 5.0.1 we exhibit, for every Hamming weight w, a distribution on (small constant-depth) circuits that distinguishes balanced strings from strings of Hamming weight w.

2. Then, in Claim 5.0.2 using standard amplification techniques coupled with Ajtai's small constant-depth circuits for approximate majority [Ajt1], we obtain, for every Hamming weight w, a deterministic circuit that distinguishes balanced inputs from inputs of weight w.

3. Applying the previous item for every w, we construct in Claim 5.1.1 a deterministic circuit that distinguishes balanced inputs from inputs of any Hamming weight less than $1/2$.

4. Finally, in Claim 5.1.2 we use the circuits from the previous item to compute majority.

We now proceed with the formal proof. Throughout, we speak of circuits with "oracle access to $\{D_x\}_{x,D}$." What we mean by this is that such a circuit can have oracle gates for the functions $D_x : \{0,1\}^q \to \{0,1\}$ for any $x \in \{0,1\}^k, D \in \mathcal{D}$, where of course the number of such oracle gates is limited by the size of the circuit.

Claim 5.0.1. *There is a universal constant c such that for every integer Hamming weight $w < \epsilon^{-1}/2$, there is a distribution S_w on functions $s : \{0,1\}^{1/\epsilon} \to \{0,1\}$ such that*

1. *Each $s \in S_w, s : \{0,1\}^{1/\epsilon} \to \{0,1\}$ is computable by a circuit of depth c and size $(q/\epsilon)^c$ with oracle access to $\{D_x\}_{x,D}$, and*

2. *S_w tells balanced strings from strings of Hamming weight w: There are $\alpha, \beta \in [0,1]$ such that $|\alpha - \beta| \geq \gamma/2$ and for every $y \in \{0,1\}^{1/\epsilon}$ of Hamming weight $\epsilon^{-1}/2$, $\Pr_{s \in S_w}[s(y) = 1] = \alpha$, whereas for every $z \in \{0,1\}^{1/\epsilon}$ of Hamming weight w, $\Pr_{s \in S_w}[s(z) = 1] = \beta$.*

Proof. Note that a $\delta \to (1/2-\epsilon)$ black-box hardness amplification is trivially also a $\delta \to (w\cdot\epsilon)$ black-box hardness amplification whenever $w \cdot \epsilon \leq 1/2 - \epsilon$. Therefore, we are in the position to apply the Zoom Theorem 4.2. By the zoom theorem, we can find a distribution T_w on projections of the functions $\{D_x\}_{x,D}$ such that

$$\left|\Pr[T_w(N^1_{1/2}, \ldots, N^q_{1/2}) = 1] - \Pr[T_w(N^1_{w\cdot\epsilon}, \ldots, N^q_{w\cdot\epsilon}) = 1]\right| \geq \gamma - 2 \cdot 2^{-k/C_Z} \geq \gamma/2, \qquad (\star)$$

where $(N^1_{1/2}, \ldots, N^q_{1/2})$ is a vector of q independent bits with probability of being 1 equal to $1/2$ and $(N^1_{w\cdot\epsilon}, \ldots, N^q_{w\cdot\epsilon})$ is a vector of q independent bits with probability of being 1 equal to $w \cdot \epsilon$. Above, C_Z denotes the constant in the statement of the Zoom Theorem, and the last inequality holds by our assumption that $\gamma \geq 2^{-k/C}$ and for a choice of C that is sufficiently larger than C_Z.

Now consider the following distribution S_w on functions mapping $y \in \{0,1\}^{1/\epsilon}$ to $\{0,1\}$. Let $Y^1, \ldots, Y^q$ be independent random variables each of which is uniformly distributed over the set of names of input variables $\{y_1, \ldots, y_{1/\epsilon}\}$. Then, S_w is defined as

$$S_w := T_w(Y^1, \ldots, Y^q).$$

In other words, $S_w(y)$ selects q random bits from the input y and applies T_w to them. Item (2) in the claim follows by ($\star$). This is because for every input y of Hamming weight $\epsilon^{-1}/2$ the distribution of the evaluations of $(Y^1, \ldots, Y^q)$ equals the distribution $(N^1_{1/2}, \ldots, N^q_{1/2})$, and a similar argument applies to the case of an input z of weight $w \cdot \epsilon$.

Item (1) follows from the definition of S_w and the fact that T_w is a distribution on projections of D_x (possibly negated) where we use the straightforward fact that any projection of D_x is easily implementable by a small constant-depth circuit with oracle access to D_x. □

Claim 5.0.2. *There is a universal constant c such that for every Hamming weight $w < \epsilon^{-1}/2$, there is a deterministic circuit C_w of depth c and size $(q/\epsilon)^c$, with oracle access to $\{D_x\}_{x,D}$ such that C_w tells balanced strings from strings of Hamming weight w: For every $y \in \{0,1\}^{1/\epsilon}$ of Hamming weight $\epsilon^{-1}/2$, $C_w(y) = 1$, whereas for every $y \in \{0,1\}^{1/\epsilon}$ of Hamming weight w, $C_w(y) = 0$.*

The proof of the above claim follows by amplifying the success probability of the circuits S_w given by the previous claim. This amplification is accomplished by the standard method of taking several independent samples of S_w and then computing a majority of their outputs. At first glance, taking a majority may seem problematic, since the whole point of this proof is to turn the decoder into a circuit for majority. The key observation is that the success probabilities of S_w (on balanced and unbalanced inputs) are sufficiently bounded away that we can use the remarkable construction by Ajtai of small constant-depth circuits for approximate majority [Ajt1] (see [ABO, Ajt2, Vio5] for alternative proofs of Ajtai's original construction.) It is the approximation in Ajtai's circuits that forces upon us the assumption that γ is large compared to ϵ, i.e. that $\gamma \geq 1/\log(1/\epsilon)$.

Lemma 5.1 (Constant-depth circuits for approximate majority; [Ajt1])**.** *There is a constant c such that for every $m, a \leq m$ there is a circuit A of depth c and size m^c satisfying the following:*

For every input $x \in \{0,1\}^m$ of Hamming weight at least $a + a/(100 \cdot \log m)$, $A(x) = 1$, while

For every input $x \in \{0,1\}^m$ of Hamming weight at most $a - a/(100 \cdot \log m)$, $A(x) = 0$.

Proof of Claim 5.0.2. Let S_w, α, and β be as in the previous claim. Namely, $\alpha := \Pr_{s \in S_w}[s(y) = 1]$ and $\beta := \Pr_{s \in S_w}[s(z) = 1]$, where y and z are any two strings with Hamming weights $\epsilon^{-1/2}/2$ and w respectively. Also from the previous claim, we know that $|\alpha - \beta| \geq \gamma/2$, and let us suppose that $\alpha < \beta$ without loss of generality.

Consider taking m independent copies of S_w, denoted by $S^1_w, \ldots, S^m_w$. By a Chernoff bound (see, e.g. [DP, Theorem 1.1]) for $m = \text{poly}(\gamma^{-1} \cdot \epsilon^{-1})$ we have that for every y of weight $\epsilon^{-1}/2$ the probability that $\sum_i S^i_w(y) \geq m(\alpha + \gamma/8)$ is less than $2^{-1/\epsilon}$, and similarly for every z of weight w the probability that $\sum_i S_w^i(y) \leq m(\beta - \gamma/8)$ is less than $2^{-1/\epsilon}$. Therefore, by a union bound, we can fix a choice of $s_w^1 = S_w^1, \dots, s_w^m = S_w^m$ such that for every y of weight $\epsilon^{-1}/2$ we have $\sum_i s_w^i(y) \leq m(\alpha + \gamma/8)$, whereas for every z of weight w we have $\sum_i s_w^i(y) \geq m(\beta - \gamma/8)$.

Definition of the circuit C_w: Let A be the circuit from Lemma 5.1 with input length m, and $a := m(\alpha+\beta)/2$. The circuit C_w runs A on the outputs of the s_w's. Specifically,

$$C_w := A(s_w^1, \dots, s_w^m).$$

The bounds on the size and depth of the circuit C_w follow easily from those for the s_w's (in the previous claim) and for A (in Lemma 5.1).

Correctness of C_w: By Lemma 5.1, A distinguishes strings of weight at least $a + a/(100 \cdot \log m)$ from those of weight at most $a - a/(100 \cdot \log m)$. In particular, since $a \leq m$, A distinguishes strings of weight at least $a + m/(100 \cdot \log m)$ from those of weight at most $a - m/(100 \cdot \log m)$. Now recall that $\gamma \geq 1/\log(1/\epsilon)$ (cf. the statement of Theorem 1.6) and that $m \geq 1/\epsilon$. Thus, $\gamma \geq 1/\log(m)$. Consequently, A distinguishes strings of weight at least $a + m \cdot \gamma/100$ from those of weight at most $a - m \cdot \gamma/100$. To conclude, we only have to verify that the precision given by Ajtai's circuit is fine enough for the parameters given by the above Chernoff bound. Specifically, we need to verify that

$$a + m \cdot \gamma/100 \leq m \cdot (\beta - \gamma/8) \text{ and } a - a \cdot \gamma/100 \geq m \cdot (\alpha + \gamma/8).$$

We now verify the first inequality above; the verification of the second is similar. Recalling that $a := m(\alpha+\beta)/2$ and that $|\alpha - \beta| = \gamma/2$, we can verify the first inequality as follows

$$a + m \cdot \gamma/100 \leq m \cdot (\beta - \gamma/8) \Leftrightarrow \frac{\alpha+\beta}{2} + \gamma/100 \leq \beta - \gamma/8 \Leftrightarrow \gamma/100 \leq \gamma/4 - \gamma/8,$$

which is true. □

Claim 5.1.1. *There is a universal constant c and a circuit $C_* : \{0,1\}^{1/\epsilon} \to \{0,1\}$ of depth c and size $(q/\epsilon)^c$, with oracle access to $\{D_x\}_{x,D}$, that distinguishes balanced strings from strings of relative weight less than $1/2$: For every $y \in \{0,1\}^{1/\epsilon}$ of Hamming weight $\epsilon^{-1}/2$, $C_*(y) = 1$, whereas for every $y \in \{0,1\}^{1/\epsilon}$ of Hamming weight $w < \epsilon^{-1}/2$, $C_*(y) = 0$.*

Proof. For every weight w less than $\epsilon^{-1}/2$, run in parallel the circuit C_w from the previous claim. Take the AND of these circuits. In other words,

$$C_* := AND(C_1, C_2, \dots, C_{\epsilon^{-1}/2-1}).$$

The bounds on the depth and size of C_* are straightforward, so let us proceed with the analysis of correctness. If y is balanced then all the circuits $C_1, \dots, C_{\epsilon^{-1}/2-1}$ evaluate to 1 and hence their AND C_* also evaluates to 1. If y has weight $w < \epsilon^{-1}/2$ then $C_w(y) = 0$, and hence $C_*(y) = 0$. □

Claim 5.1.2. *There is a universal constant c and a circuit $C : \{0,1\}^{1/\epsilon} \to \{0,1\}$ of depth c and size $(q/\epsilon)^c$, with oracle access to $\{D_x\}_{x,D}$, that computes majority on $1/\epsilon$ bits.*

Proof. Given an input $y \in \{0,1\}^{1/\epsilon}$ we compute associated inputs $y^1, \ldots, y^{1/\epsilon}$, where y^i is obtained from y by setting the first i bits to 0. Thus, $y^0 = y$ and $y^{1/\epsilon} = 0^{1/\epsilon}$. We then run copies of the circuit C_* from Claim 5.1.1 in parallel on each of the y^i's and we output their OR. Namely,

$$C(y) := OR\left(C_*(y^0), C_*(y^1), \ldots, C_*(y^{1/\epsilon})\right).$$

Again, the bounds on the depth and size of C are straightforward, so let us proceed with the analysis of correctness. If y has weight less than $\epsilon^{-1}/2$ then all the inputs $y^0, \ldots, y^{1/\epsilon}$ also have weight less than $\epsilon^{-1}/2$; thus the circuits $C_*(y^0), C_*(y^1), \ldots, C_*(y^{1/\epsilon}$ all output 0 and so does $C(y)$.

Conversely, if y has weight at least $\epsilon^{-1}/2$, then for some i the input y^i has weight exactly $\epsilon^{-1}/2$, and thus $C(y) = C_*(y^i) = 1$. □

Why this proof cannot be easily simplified. One may wonder whether the proof in this section can be simplified. In particular, one may wonder whether it is really necessary to argue separately for each Hamming weight w, by invoking Claim 5.0.1 several times. A natural attempt to simplification would be to try to fix a particular Hamming weight w, then directly use circuits that distinguish balanced inputs from inputs of weight w to compute majority. In fact this cannot be done: The ability to distinguish two Hamming weights is *not* enough to compute majority. As a simple example of this phenomenon, observe that the parity function distinguishes inputs of Hamming weight w from inputs of Hamming weight $w+2\cdot a+1$, for every w, a, but it is known that small constant-depth circuits with parity gates cannot compute majority. In conclusion, it is only the availability *for every* w of circuits that distinguish balanced inputs from inputs of weight w that lets us compute majority.

Finally, we remark that in Theorem 1.6 the assumption $\gamma \geq \log(1/\epsilon)$ can be replaced by $\gamma \geq \log^i(1/\epsilon)$, where the depth of the circuit in the conclusion of the theorem depends on i.

6 Proof of queries lower bound (Theorem 1.4)

In this section we prove our queries lower bound (Theorem 1.4). We restate the theorem for the reader's convenience.

Theorem 1.4 (Decoding requires many queries)**.** *There is a universal constant $C > 1$ such that the following holds. Let $(Amp, \mathcal{D})$ be a non-adaptive q-query $\delta \to (1/2 - \epsilon)$ black-box hardness amplification. Suppose that $\log|\mathcal{D}| \leq 2^{k/C}$, and $n, k \geq C^2$, and that both δ and ϵ are between $2^{-k/C}$ and $1/3$.*

Then

$$q \geq \frac{1}{C} \cdot \frac{\log(1/\delta)}{\epsilon^2}.$$

The proof of the queries lower bound (Theorem 1.4) follows from the Zoom Theorem 4.2 and the next lemma.[6]

In this section **we write $N^q_{1/2-\epsilon}$ for a vector of q independent $0-1$ random variables whose probability of being 1 is $1/2-\epsilon$.**

Lemma 6.1. *Let T be a distribution on functions $t : \{0,1\}^q \to \{0,1\}$ such that*

1. $\Pr_{T,N^q_{1/2-\epsilon}}\left[T(N^q_{1/2-\epsilon}) = 1\right] \leq p \leq 0.4$, *and*

2. $\Pr_{T,N^q_{1/2}}\left[T(N^q_{1/2}) = 1\right] \geq 0.49$.

Then $q \geq \Omega(\log(1/p)/\epsilon^2)$.

In the next section we prove Lemma 6.1, thus completing the proof of Theorem 1.4. We remark that statements similar to Lemma 6.1 have appeared often in the literature, however we are unaware of a simple self-contained proof. Our proof is inspired by an argument of [CEG] that shows a lower bound on the query complexity of "sampling procedures" [Gol1]. While it does not seem that one can reduce our

setup to that of "sampling procedures" the approach of [CEG] can be applied in our setup and gives the lower bound.

6.1 Proof of Lemma 6.1

The family T that we are given distinguishes $N^q_{1/2}$ from $N^q_{1/2-\epsilon}$ in the sense that it answers "one" with probability at most $p \leq 0.4$ on $N^q_{1/2-\epsilon}$, whereas it answers "one" with probability at least 0.49 on $N^q_{1/2}$. We would like to replace 0.49 with a constant that is larger than $1/2$ (say 0.99). This can be achieved by "amplification." More specifically, we can increase q to $c \cdot q$ (for some universal constant $c > 1$), take c independent copies of T that are run on c independent copies of the input (these copies are identically distributed and could come from either $N^q_{1/2-\epsilon}$ or $N^q_{1/2}$), and combine the outputs appropriately. Performing this amplification gives the following claim.

Claim 6.1.1. *There is a universal constant c and a distribution T' on functions $t : \{0,1\}^{q'} \to \{0,1\}$, where $q' := c \cdot q$, such that*

1. $\Pr_{T',N^{q'}_{1/2-\epsilon}}\left[T'(N^{q'}_{1/2-\epsilon}) = 1\right] \leq p/4$, *and*

2. $\Pr_{T',N^{q'}_{1/2}}\left[T'(N^{q'}_{1/2}) = 1\right] \geq 0.99$.

[6] We note that the Zoom Theorem can only be applied when $q \leq 2^{k/C}$, whereas the hypothesis of Theorem 1.4 does not place any restriction on q. This is not a problem because by choosing the constant C in Theorem 1.4 sufficiently larger than the constant C in the Zoom Theorem one has that if q does not satisfy the conclusion of Theorem 1.4 then one is indeed in the position to apply the Zoom Theorem.

Proof. Let d be a large constant to be determined later, and set $b := \log_{1/0.4} d + 3$. The constant c in the statement of the claim is defined as $c := b \cdot d$. We think of the input as c blocks of q bits. The distribution on functions T' is defined as follows. We take c independent copies of T, and we evaluate each on the corresponding input block. This results in an output vector of length c. We divide up this vector into d blocks of b bits each. We take the AND in each block of b bits, then we complement the result and we take the AND over the d blocks. Finally, for consistency in the notation we complement the result again.

Analysis. We have

$$\Pr_{T', N^{q'}_{1/2-\epsilon}} \left[T'(N^{q'}_{1/2-\epsilon}) = 0\right] \geq \left(1 - p^b\right)^d \geq 1 - d \cdot p^b = 1 - d \cdot p^{b-3} \cdot p^3$$

$$\geq 1 - d \cdot (0.4)^{b-3} \cdot p \cdot (0.4)^2 = 1 - p \cdot (0.16) \geq 1 - p/4,$$

where we used that $p \leq 0.4$ and $b := \log_{1/0.4} d + 3$.

Also, for an integer K and $d := (1/0.4)^K$ (and hence $b = K + 3$), which we can set without spoiling the above derivation, we have:

$$\Pr_{T', N^{q'}_{1/2}} \left[T'(N^{q'}_{1/2}) = 0\right] \leq \left(1 - (0.49)^b\right)^d =$$

$$\left(1 - (0.49)^{K+3}\right)^{(1/0.4)^K} < \exp\left(-\left(\frac{0.49}{0.4}\right)^K \cdot (0.49)^3\right) < 0.01,$$

for sufficiently large K. □

Claim 6.1.2. *There is a fixed function $t : \{0,1\}^{q'} \to \{0,1\}$ such that*

1. $\Pr_{N^{q'}_{1/2-\epsilon}} \left[t\left(N^{q'}_{1/2-\epsilon}\right) = 1\right] \leq p$, *and*
2. $\Pr_{N^{q'}_{1/2}} \left[t\left(N^{q'}_{1/2}\right) = 1\right] \geq 1/2 + 1/4$.

Proof. We show by Markov arguments that a random $t \in T'$ satisfies both items in the conclusion of the lemma with positive probability.

First, by a Markov argument, the first item in the conclusion of the previous claim implies that

$$\Pr_{T'}\left[\Pr_{N^{q'}_{1/2-\epsilon}} \left[T'(N^{q'}_{1/2-\epsilon}) = 1\right] \geq p\right] \leq 1/4.$$

Second, another Markov argument shows that the second item in the conclusion of the previous claim implies that

$$\Pr_{T'}\left[\Pr_{N^{q'}_{1/2}} \left[T'(N^{q'}_{1/2}) = 0\right] \geq 1/4\right] \leq 4/100.$$

By a union bound, since $1/4 + 4/100 < 1$, there is a t that satisfies the conclusion of the claim. □

Claim 6.1.3. *There is a set $S \subseteq \{0,1\}^{q'}$ of relative size $1/4$ such that every string $s \in S$ satisfies: (1) $t(s) = 1$ and (2) the Hamming weight of s is at most $q'/2$.*

Proof. From the previous claim we know that t evaluates to 1 on a $1/2 + 1/4$ fraction of the inputs in $\{0,1\}^{q'}$. On the other hand, the strings with weight bigger than $q'/2$ have measure exactly $1/2$. Therefore such a set S exists. □

We will conclude the proof by observing that $N^{q'}_{1/2-\epsilon}$ falls in S with probability at least $\exp\left(-O\left(\epsilon^2 \cdot q'\right)\right)$, then recalling that on the other hand this probability must be at most p by the first item in the conclusion of the previous claim. Details follow. First, note that

$$\Pr_{N^{q'}_{1/2-\epsilon}}\left[N^{q'}_{1/2-\epsilon} \in S\right] = \sum_{s \in S} \Pr_{N^{q'}_{1/2-\epsilon}}\left[N^{q'}_{1/2-\epsilon} = s\right] \geq |S| \cdot (1/2-\epsilon)^{q'/2}(1/2+\epsilon)^{q'/2}$$
$$= |S| \cdot 2^{-q'}(1-2\epsilon)^{q'/2}(1+2\epsilon)^{q'/2} = (1/4)(1-4\epsilon^2)^{q'/2} \geq (1/4) \cdot \exp\left(-8 \cdot \epsilon^2 \cdot q'/2\right). \qquad (\star)$$

Above, we used that S has relative size $1/4$, and that $\Pr_{N^{q'}_{1/2-\epsilon}}\left[N^{q'}_{1/2-\epsilon} = s\right] \geq (1/2 - \epsilon)^{q'/2}(1/2+\epsilon)^{q'/2}$ for every $s \in S$ because strings in s have weight at most $q'/2$. Finally, the last inequality holds whenever $\epsilon < 1/3$.

The above $(\star)$ implies that $\Pr_{N^{q'}_{1/2-\epsilon}}\left[t\left(N^{q'}_{1/2-\epsilon}\right) = 1\right] \geq (1/4)\exp\left(-4 \cdot \epsilon^2 \cdot q'\right)$. On the other hand, this probability must be at most p by the previous claim. Consequently,

$$(1/4)\exp\left(-4 \cdot \epsilon^2 \cdot q'\right) \leq p.$$

Since $q' = c \cdot q$, this means that $q = \Omega\left(\log(1/p)/\epsilon^2\right)$, concluding the proof of the lemma.

7 Case studies and significance of our results

In this section we give background on a few widely-studied circuit classes and we discuss what our results have to say about them.

Constant-depth circuits AC^0: The class of constant-depth circuits with And, Not, and Or gates is one of the central classes studied in circuit complexity. Lower bounds are known for this class, and even strong average-case lower bounds. In particular, the celebrated result in [Hås] (cf. [FSS, Yao2]) shows that the parity function on n bits is $(1/2 - 1/s)$-hard for AC^0 circuits of depth d and size s, where

$$s := \exp\left(n^{\Theta(1/d)}\right).$$

This result has a number of consequences, and in particular it was used to construct pseudorandom generators that fool constant-depth circuits [Nis].

One can ask for stronger lower bounds, for example for lower bounds for size

$$s' := \exp\left(\Theta(n)\right).$$

Such lower bounds are not known even for depth-3 circuits. It is natural to wonder whether one can use hardness amplification techniques to relate average-case lower bounds for size s' to worst-case lower bounds for circuits of size s'. This kind of connection does hold for circuits of unrestricted depth (e.g., [IW1, STV]) but fails when we restrict the depth of the circuit. For more on this issue we refer the reader to the excellent discussion by Agrawal [Agr].

Our results explain the above failure of hardness amplification techniques: Proving such a connection using standard hardness amplification techniques would require the circuits to compute majority, and in particular their depth to be large (see Theorems 1.6 and 1.3).

Constant-depth circuits with parity gates $AC^0[2]$: If we augment AC circuits with parity gates we obtain the class of constant-depth circuits with And, Not, Or, and Parity gates, denoted $AC^0[2]$. A long-standing open problem about this class is whether there is an explicit function $f : \{0,1\}^n \to \{0,1\}$ that is $\left(1/2 - 1/n^{\omega(1)}\right)$-hard for polynomial-size constant-depth circuits with Parity gates,[7] i.e. such that for any depth-d polynomial-size $AC^0[2]$ circuit C:

$$\Pr_{x \in \{0,1\}^n}[C(x) \neq f(x)] \geq 1/2 - 1/n^{\omega(1)}.$$

This problem is perhaps most puzzling because we do know of explicit functions that are $\Omega(1)$-hard for depth-d $AC^0[2]$ circuits of size $2^{n^{\Omega(1/d)}}$. An example is the Mod 3 function, i.e. counting the number of 1's mod 3 in a given n-bit input (see [Raz, Smo] and the survey by Beigel [Bei1, Corollary 22]). Even for this class, standard hardness amplification techniques fail, and in particular we cannot amplify the hardness of the mod 3 function.

Our results again explain the above failure of hardness amplification techniques: Just like before, proving such a connection using standard hardness amplification techniques would require the circuits to compute majority, whereas $AC^0[2]$ circuits cannot [Raz, Smo].

Similar consideration hold for classes with mod p gates for p prime (cf. [Smo]); we focus on $p = 2$ for simplicity.

Constant-depth circuits with one Majority gate $MAJ \circ AC^0$: Another puzzling case is that of constant-depth circuits with *one* majority gate ($MAJ \circ AC^0$ circuits), a class which in particular contains the widely studied *perceptrons*. It is known that the n-bits Parity function is $\Omega(1)$-hard for $MAJ \circ AC^0$ circuits of depth d and size $2^{n^{\Omega(1/d)}}$ [ABFR] (see also [Kli, Theorem 6]). However, it is not known whether there is an explicit function $f : \{0,1\}^n \to \{0,1\}$ that is $\left(1/2 - 1/2^{n^{\Omega(1)}}\right)$-hard for $MAJ \circ AC^0$ circuits of size $2^{n^{\Omega(1)}}$, or even $\left(1/2 - 1/n^{\omega(\log n)}\right)$-hard for $MAJ \circ AC^0$ circuits of size $n^{\omega(\log n)}$ (think of a fixed large depth). To the best of our knowledge, the strongest result in this direction is the one in [Vio6]

[7] For context, this problem is also open if C varies over GF(2) polynomials of degree $2\log(n)$. This is a slightly different setting because these polynomials in general correspond to circuits of size $n^{\Omega(\log n)}$. On the other hand, if the degree is $\epsilon \log(n)$ then an explicit $\left(1/2 - 1/n^{\omega(1)}\right)$-hard function is known [BNS, HG, Vio4, VW].

that in particular gives a function that is $\left(1/2 - 1/n^{\epsilon\cdot\log n}\right)$-hard for $MAJ \circ AC^0$ circuits of size $n^{\epsilon\cdot\log n}$.

Even though this class can compute majority, our query lower bound (Theorem 1.4) explains why we cannot amplify hardness against $MAJ \circ AC^0$ circuits. Specifically, it is easy to see that if the circuits $\mathcal{D}$ in the black-box hardness amplification make q query, then to obtain a function that is $(1/2 - \epsilon)$-hard for circuits with 1 majority gate we must start from a function that is δ-hard for circuits with q majority gates. Since $q \geq 1/\epsilon$ by our Theorem 1.4, we obtain that when $\epsilon \leq 1/n$ we must start from a function that is hard for constant-depth circuits with n majority gates, i.e. TC^0. (Our current lower bounding techniques [Bei2] let us handle circuits with poly log majority gates, and this in turn lets us amplify hardness up to $1/2 - 1/\text{poly}\log$.)

A final remark about this class is in order. For the specific hardness amplification given by Yao's XOR lemma (cf. Section 1.1) one can prove the stronger result that this lemma simply is *false* for $MAJ \circ AC^0$ circuits (for context, we mention that whether this lemma holds for other classes such as $AC^0[2]$ and AC^0 is a major open problem). This failure was pointed out to us by Adam Klivans (personal communication, 2002) and holds because applying Yao's XOR lemma to the parity function results again in the parity function, which is provably *not* strongly average-case hard for $MAJ \circ AC^0$ circuits. Our results apply more generally to arbitrary hardness amplifications. Moreover, they also apply when the single majority gate is replaced by a more powerful gate that can compute an arbitrary symmetric function (including parity) cf. [HG, RW, HM, Vio6].

Finally, we mention that an important motivation for obtaining strong average-case hardness results for these classes is that such results can be used to construct pseudorandom generators that fool the same classes, cf. [Nis, LVW, NW, Vio6].

8 Relationship to error-correcting codes

It has been observed (e.g. [STV, Tre1] that black-box hardness amplification is closely related to list-decodable error correcting codes. In particular, our lower bounds on black-box hardness amplification can be seen as lower bounds on the "complexity of decoding" (list-decodable) locally-decodable codes. The purpose of this section is to explain these connections at a high-level and we will not state precise theorems.

We start with the following definition of list-decodable codes.

Definition 8.1 (List-decodable codes). *A map $Amp : \{0,1\}^K \to \{0,1\}^N$ is a (ρ, ℓ)-list decodable code if for every $h \in \{0,1\}^N$, $\left|\left\{f \in \{0,1\}^K : \Delta(Amp(f), h) \leq \rho\right\}\right| \leq \ell$. (Here Δ is the relative Hamming distance.) A code is uniquely decodable with radius ρ if it is $(\rho, 1)$-list decodable.*

Let $(Amp, \mathcal{D})$ be a $(\delta = 2^{-k}) \to (1/2 - \epsilon)$ black box hardness amplification, then the map Amp is a $(1/2 - \epsilon, |\mathcal{D}|)$-list decodable code. (Here we think of Amp as a mapping that receives and outputs truth tables. That is we set $K = 2^k$ and $N = 2^n$.)

Our lower bounds immediately translate into lower bounds on the complexity of decoders of list-decodable locally decodable codes. Let us start with *uniquely decodable* locally decodable codes.

Definition 8.2 (Locally decodable codes)**.** *A mapping $Amp : \{0,1\}^K \to \{0,1\}^N$ is a locally decodable code with radius ρ that makes q queries and has error probability δ if Amp is a uniquely decodable code with radius ρ and there exists a randomized oracle procedure D such that for every $f : \{0,1\}^k \to \{0,1\}$ and $h : \{0,1\}^n \to \{0,1\}$ such that $\Delta(Amp(f), h) \leq \rho$ and every $x \in \{0,1\}^k$,*

$$\Pr[D^h(x) \neq f(x)] < \delta \tag{17}$$

(where the probability is over the coin tosses of D).

Let us write $D = D(x, r)$ where x is the input of D and r is the coin tosses of D. We define:

$$\mathcal{D} = \{D'(x) : \exists r \ s.t. \ D'(x) = D(x, r)\}$$

Note that $|\mathcal{D}|$ is bounded by the number of possible coin tosses of D. We now claim that the pair $(Amp, \mathcal{D})$ is a $\delta \to \rho$ black box hardness amplification. This follows because by an averaging argument for every $h : \{0,1\}^n \to \{0,1\}$ such that $\Delta(Amp(f), h) \leq \rho$ there exists a fixed string r_0 for which

$$\Pr_{x \in \{0,1\}^k}[D^h(x, r_0) \neq f(x)] < \delta$$

Furthermore, the function $D'(x) = D(x, r_0)$ indeed appears in $\mathcal{D}$.

Thus, our lower bounds on the complexity of black-box hardness amplification immediately translate into lower bounds on the complexity of the decoding algorithm D. More specifically, we show that D can be used to compute majority and that it needs to make many queries.[8]

We now explain that the same lower bounds apply to a very general notion of *list-decodable* locally decodable codes. Note that the argument above would work if instead of having one randomized oracle procedure we allowed the decoding procedure to be chosen from a "list" of many randomized oracle procedures $\{D_\alpha\}$, and only required that for any f and h there exists procedure D in the list is guaranteed to fulfill equation (17). (We only need to take the size of the list into account when measuring the size of $\mathcal{D}$.) We remark that the argument applies even when different procedures in the list are allowed to use different query distributions. Finally, our lower bounds apply even if the notion of decoding is relaxed and instead of equation (17) the "correct procedure" D is only guaranteed to decode in a weak sense, namely:

$$\Pr_{x \in \{0,1\}^k}[\Pr[D^h(x) \neq f(x)] < \delta/2] > 1 - \delta/2$$

[8] It is important to note that black-box hardness amplifications become uninteresting when $\delta > 1/2 - \epsilon$ (and indeed our results only apply when $\delta < 1/2 - \epsilon$). In contrast, locally decodable codes are interesting even when $\delta > \rho = 1/2 - \epsilon$. However, in this case the transformation above is transforming the code into an uninteresting black-box hardness amplification.

This is because the argument above still gives that there exists a fixing r_0 of the random coins of D for which it decodes correctly on a $(1-\delta)$-fraction of inputs $x \in \{0,1\}^k$.

9 Open problems

One weakness of our result is that we can only handle black-box hardness amplification which use *nonadaptive* circuits. While to the best of our knowledge most known black-box hardness amplification results use nonadaptive circuits, it is an interesting open problem to extend the results in this work to the case of *adaptive* circuits. We remark that, for some specific functions *Amp*, the techniques in this work already give some results on adaptive circuits when the amount of non-uniformity $|\mathcal{D}|$ of the black-box hardness amplification is small (e.g., $|\mathcal{D}| = \text{poly}(1/\epsilon)$). In particular, one can show that achieving the parameters of the hardness amplification in [GL] (based on the Hadamard code) or the parameters of the hardness amplification in [STV] (based on Reed-Muller codes), requires computing majority. The details of these results appear in [Vio3, Chapter 6].

Another problem that deserves more investigation is whether something similar to our results can be said about pseudorandom generator constructions. For example, is computing majority necessary for a black-box construction of a pseudorandom generator with constant error from a (1/3)-hard function?

Acknowledgements. We would like to thank Madhu Sudan, Salil Vadhan, and Avi Wigderson for many helpful conversations and comments. Ronen would like to thank Avi Wigderson for introducing him to the subject of hardness amplification and for many discussions. Emanuele would like to thank Salil Vadhan for many helpful discussions during the preliminary stages of this work [Vio3, Chapter 6], and Madhu Sudan also for helpful discussions and especially for a key suggestion (cf. Section 2, Claim 5.0.1, and [Vio3, Section 6.2]). We thank Russell Impagliazzo for pointing out [EIRS] to us. We also thank Hoeteck Wee for pointing out [CEG] to us.

References

[Agr] M. Agrawal. Hard Sets and Pseudo-random Generators for Constant Depth Circuits. In *Twenty First Foundations of Software Technology and Theoretical Computer Science, December 13-15, Bangalore, India*, pages 58–69. Springer-Verlag, 2001. 8, 30

[Ajt1] M. Ajtai. Σ_1^1-formulae on finite structures. *Ann. Pure Appl. Logic*, 24(1):1–48, 1983. 10, 23, 24

[Ajt2] M. Ajtai. Approximate counting with uniform constant-depth circuits. In *Advances in computational complexity theory (New Brunswick, NJ, 1990)*, pages 1–20. Amer. Math. Soc., Providence, RI, 1993. 10, 24

[ABO] M. Ajtai and M. Ben-Or. A Theorem on Probabilistic Constant Depth Computation. In ACM, editor, *Proceedings of the sixteenth annual ACM Symposium on Theory of Computing, Washington, DC, April 30–May 2, 1984*, pages 471–474, 1984. 24

[AB] N. Alon and R. Beigel. Lower bounds for approximations by low degree polynomials over Z_m. In *Proceedings of the Sixteenth Annual Conference on Computational Complexity*, pages 184–187. IEEE, June 18–21 2001. 8

[ABFR] J. Aspnes, R. Beigel, M. Furst, and S. Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):135–148, 1994. 1, 9, 30

[BFL] L. Babai, L. Fortnow, and C. Lund. Nondeterministic exponential time has two-prover interactive protocols. *Comput. Complexity*, 1(1):3–40, 1991. 1, 3

[BFNW] L. Babai, L. Fortnow, N. Nisan, and A. Wigderson. BPP Has Subexponential Time Simulations Unless EXPTIME has Publishable Proofs. *Computational Complexity*, 3(4):307–318, 1993. 1, 3, 8

[BNS] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Comput. System Sci.*, 45(2):204–232, 1992. Twenty-first Symposium on the Theory of Computing (Seattle, WA, 1989). 30

[BF] D. Beaver and J. Feigenbaum. Hiding Instances in Multioracle Queries. In *7th Annual Symposium on Theoretical Aspects of Computer Science*, volume 415 of *Lecture Notes in Computer Science*, pages 37–48, Rouen, France, 22–24 Feb. 1990. Springer. 1, 3

[Bei1] R. Beigel. The polynomial method in circuit complexity. In *Proceedings of the Eighth Annual Structure in Complexity Theory Conference (San Diego, CA, 1993)*, pages 82–95, Los Alamitos, CA, 1993. IEEE Comput. Soc. Press. 30

[Bei2] R. Beigel. When do extra majority gates help? polylog(N) majority gates are equivalent to one. *Comput. Complexity*, 4(4):314–324, 1994. Special issue devoted to the 4th Annual McGill Workshop on Complexity Theory. 31

[BT1] A. Bogdanov and L. Trevisan. Average-Case Complexity. *Foundations and Trends in Theoretical Computer Science*, 2(1), 2006. 8

[BT2] A. Bogdanov and L. Trevisan. On Worst-Case to Average-Case Reductions for NP Problems. *SIAM J. Comput.*, 36(4):1119–1159, 2006. 8

[Bou] J. Bourgain. Estimation of certain exponential sums arising in complexity theory. *C. R. Math. Acad. Sci. Paris*, 340(9):627–631, 2005. 8

[CPS] J.-Y. Cai, A. Pavan, and D. Sivakumar. On the Hardness of the Permanent. In *16th International Symposium on Theoretical Aspects of Computer Science*, Lecture Notes in Computer Science, Volume 1563, pages 90–99, Trier, Germany, 1999. Springer-Verlag. 1, 3

[CEG] R. Canetti, G. Even, and O. Goldreich. Lower Bounds for Sampling Algorithms for Estimating the Average. *Information Processing Letters*, 53(1):17–25, 1995. 27, 33

[CT] T. M. Cover and J. A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006. 14, 17

[CK] I. Csiszar and J. G. Korner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, Inc., Orlando, FL, USA, 1982. 14

[DP] D. Dubhashi and A. Panconesi. Concentration of Measure for the Analysis of Randomised Algorithms, 2005. Manuscript. Available from `http://http://www.dsi.uniroma1.it/~ale/papers.html`. 22, 24

[EIRS] J. Edmonds, R. Impagliazzo, S. Rudich, and J. Sgall. Communication complexity towards lower bounds on circuit depth. *Computational Complexity*, 10(3):210–246, 2001. 12, 14, 33

[FL] U. Feige and C. Lund. On the Hardness of Computing the Permanent of Random Matrices. *Computational Complexity*, 6(2):101–132, 1996. 1, 3

[FSS] M. L. Furst, J. B. Saxe, and M. Sipser. Parity, Circuits, and the Polynomial-Time Hierarchy. *Mathematical Systems Theory*, 17(1):13–27, April 1984. 1, 29

[Gol1] O. Goldreich. A Sample of Samplers - A Computational Perspective on Sampling (survey). *Electronic Colloquium on Computational Complexity (ECCC)*, 4(020), 1997. 27

[Gol2] O. Goldreich. *Modern cryptography, probabilistic proofs and pseudorandomness*, volume 17 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 1999. 1

[Gol3] O. Goldreich. *Foundations of Cryptography: Volume 1, Basic Tools*. Cambridge University Press, Cambridge, 2001. 1

[GL] O. Goldreich and L. A. Levin. A Hard-Core Predicate for all One-Way Functions. In *Proceedings of the Twenty First Annual ACM Symposium on Theory of Computing*, pages 25–32, Seattle, Washington, 15–17 May 1989. 6, 33

[GNW] O. Goldreich, N. Nisan, and A. Wigderson. On Yao's XOR lemma. Technical Report TR95–050, Electronic Colloquium on Computational Complexity, March 1995. `http://www.eccc.uni-trier.de/eccc`. 1, 2, 5, 6, 7

[GGH$^+$] S. Goldwasser, D. Gutfreund, A. Healy, T. Kaufman, and G. N. Rothblum. Verifying and decoding in constant depth. In *STOC*, pages 440–449, 2007. 5

[GG] P. Gopalan and V. Guruswami. Hardness amplification within NP against deterministic algorithms. In *Proceedings of the 23nd Annual Conference on Computational Complexity*. IEEE, June 23–26 2008. 1

[GRS] F. Green, A. Roy, and H. Straubing. Bounds on an exponential sum arising in Boolean circuit complexity. *C. R. Math. Acad. Sci. Paris*, 341(5):279–282, 2005. 8

[GK] V. Guruswami and V. Kabanets. Hardness Amplification Via Space-Efficient Direct Products. In J. R. Correa, A. Hevia, and M. A. Kiwi, editors, *LATIN*, volume 3887 of *Lecture Notes in Computer Science*, pages 556–568. Springer, 2006. 1

[HMP$^+$] A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. *J. Comput. System Sci.*, 46(2):129–154, 1993. 1, 8

[HM] K. A. Hansen and P. B. Miltersen. Some Meet-in-the-Middle Circuit Lower Bounds. In *Proceedings of the 29th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, Lecture Notes in Computer Science, Volume 3153, pages 334 – 345, August 22–27 2004. 1, 31

[Hås] J. Håstad. *Computational limitations of small-depth circuits*. MIT Press, 1987. 1, 5, 7, 8, 29

[HG] J. Håstad and M. Goldmann. On the power of small-depth threshold circuits. *Comput. Complexity*, 1(2):113–129, 1991. 1, 30, 31

[HVV] A. Healy, S. P. Vadhan, and E. Viola. Using Nondeterminism to Amplify Hardness. *SIAM J. Comput.*, 35(4):903–931, 2006. 1, 6, 8

[Imp] R. Impagliazzo. Hard-core distributions for somewhat hard problems. In *36th Annual Symposium on Foundations of Computer Science*, pages 538–545, Milwaukee, Wisconsin, 23–25 Oct. 1995. IEEE. 1, 5, 6, 8

[IJK] R. Impagliazzo, R. Jaiswal, and V. Kabanets. Approximately List-Decoding Direct Product Codes and Uniform Hardness Amplification. In *FOCS*, pages 187–196. IEEE Computer Society, 2006. 1, 8

[IJKW] R. Impagliazzo, R. Jaiswal, V. Kabanets, and A. Wigderson. Uniform Direct-Product Theorems: Simplified, Optimized, and Derandomized. In *Proceedings of the 40th Annual ACM Symposium on the Theory of Computing (STOC)*, Victoria, Canada, 17–20 May 2008. 1, 8

[IW1] R. Impagliazzo and A. Wigderson. $P = BPP$ if E Requires Exponential Circuits: Derandomizing the XOR Lemma. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*, pages 220–229, El Paso, Texas, 4–6 May 1997. 1, 6, 8, 30

[IW2] R. Impagliazzo and A. Wigderson. Randomness vs time: derandomization under a uniform assumption. *J. Comput. System Sci.*, 63(4):672–688, 2001. Special issue on FOCS 98. 1, 8

[KS] A. Klivans and R. A. Servedio. Boosting and Hard-Core Sets. *Machine Learning*, 53(3):217–238, 2003. 5, 6, 8

[Kli] A. R. Klivans. On the Derandomization of Constant Depth Circuits. In *Proceedings of the Fifth International Workshop on Randomization and Approximation Techniques in Computer Science*, August 18–20 2001. 5, 8, 9, 30

[LTW] H. Lin, L. Trevisan, and H. Wee. On Hardness Amplification of One-Way Functions. In J. Kilian, editor, *TCC*, volume 3378 of *Lecture Notes in Computer Science*, pages 34–49. Springer, 2005. 8

[Lip] R. Lipton. New Directions in Testing. In *Proceedings of DIMACS Workshop on Distributed Computing and Cryptography*, volume 2, pages 191–202. ACM/AMS, 1991. 1, 3

[LTW1] C.-J. Lu, S.-C. Tsai, and H.-L. Wu. On the Complexity of Hardness Amplification. In *Proceedings of the 20th Annual Conference on Computational Complexity*, pages 170–182. IEEE, June 12–15 2005. 8

[LTW2] C.-J. Lu, S.-C. Tsai, and H.-L. Wu. Impossibility Results on Weakly Black-Box Hardness Amplification. In E. Csuhaj-Varjú and Z. Ésik, editors, *FCT*, volume 4639 of *Lecture Notes in Computer Science*, pages 400–411. Springer, 2007. 8

[LTW3] C.-J. Lu, S.-C. Tsai, and H.-L. Wu. Improved hardness amplification in NP. *Theor. Comput. Sci.*, 370(1-3):293–298, 2007. 8

[LTW4] C.-J. Lu, S.-C. Tsai, and H.-L. Wu. On the Complexity of Hard-Core Set Constructions. In L. Arge, C. Cachin, T. Jurdzinski, and A. Tarlecki, editors, *ICALP*, volume 4596 of *Lecture Notes in Computer Science*, pages 183–194. Springer, 2007. 8

[LVW] M. Luby, B. Velickovic, and A. Wigderson. Deterministic Approximate Counting of Depth-2 Circuits. In *Proceedings of the 2nd Israeli Symposium on Theoretical Computer Science (ISTCS)*, pages 18–24, 1993. 31

[NR] M. Naor and O. Reingold. Number-theoretic constructions of efficient pseudo-random functions. *J. ACM*, 51(2):231–262, 2004. 1

[Nis] N. Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991. 29, 31

[NW] N. Nisan and A. Wigderson. Hardness vs Randomness. *J. Computer & Systems Sciences*, 49(2):149–167, Oct. 1994. 1, 3, 31

[O'D] R. O'Donnell. Hardness Amplification Within *NP*. *J. Comput. Syst. Sci.*, 69(1):68–94, Aug. 2004. 1, 6, 8

[Raz] R. Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803 (electronic), 1998. 12, 14

[RW] A. Razborov and A. Wigderson. $n^{\Omega(\log n)}$ lower bounds on the size of depth-3 threshold circuits with AND gates at the bottom. *Inform. Process. Lett.*, 45(6):303–307, 1993. 31

[Raz] A. A. Razborov. Lower bounds on the dimension of schemes of bounded depth in a complete basis containing the logical addition function. *Mat. Zametki*, 41(4):598–607, 623, 1987. 1, 5, 7, 30

[RR] A. A. Razborov and S. Rudich. Natural Proofs. *J. Comput. Syst. Sci.*, 55(1):24–35, Aug. 1997. 1, 2

[SU1] R. Shaltiel and C. Umans. Simple extractors for all min-entropies and a new pseudorandom generator. *J. ACM*, 52(2):172–216, 2005. 1

[SU2] R. Shaltiel and C. Umans. Pseudorandomness for Approximate Counting and Sampling. *Computational Complexity*, 15(4):298–341, 2006. 1

[SVW] R. Shaltiel, E. Viola, and A. Wigderson. Unpublished manuscript, 2005. 6

[Smo] R. Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, pages 77–82, New York City, 25–27 May 1987. 1, 5, 7, 30

[STV] M. Sudan, L. Trevisan, and S. Vadhan. Pseudorandom generators without the XOR lemma. *J. Comput. System Sci.*, 62(2):236–266, 2001. Special issue on the Fourteenth Annual IEEE Conference on Computational Complexity (Atlanta, GA, 1999). 1, 3, 6, 8, 30, 31, 33

[Tre1] L. Trevisan. List Decoding Using the XOR Lemma. In *44th Annual Symposium on Foundations of Computer Science*, pages 126–135, Cambridge, Massachusetts, 11–14 Oct. 2003. IEEE. 1, 3, 6, 8, 31

[Tre2] L. Trevisan. Some applications of coding theory in computational complexity. In *Complexity of computations and proofs*, volume 13 of *Quad. Mat.*, pages 347–424. Dept. Math., Seconda Univ. Napoli, Caserta, 2004. 1, 8

[Tre3] L. Trevisan. On uniform amplification of hardness in NP. In H. N. Gabow and R. Fagin, editors, *STOC*, pages 31–38. ACM, 2005. 1, 8

[TV] L. Trevisan and S. Vadhan. Pseudorandomness and Average-Case Complexity Via Uniform Reductions. *Comput. Complex.*, 16(4):331–364, 2007. 1, 3, 4, 8

[Vio1] E. Viola. The Complexity of Constructing Pseudorandom Generators from Hard Functions. *Comput. Complexity*, 13(3-4):147–188, 2004. 1, 8

[Vio2] E. Viola. On Constructing Parallel Pseudorandom Generators from One-Way Functions. In *Proceedings of the 20th Annual Conference on Computational Complexity*, pages 183–197. IEEE, June 12–15 2005. 8

[Vio3] E. Viola. *The Complexity of Hardness Amplification and Derandomization*. PhD thesis, Harvard University, 2006. `http://www.eccc.uni-trier.de/eccc`. 1, 8, 33

[Vio4] E. Viola. New correlation bounds for GF(2) polynomials using Gowers uniformity. *Electronic Colloquium on Computational Complexity*, Technical Report TR06-097, 2006. `http://www.eccc.uni-trier.de/eccc`. 30

[Vio5] E. Viola. On approximate majority and probabilistic time. In *Proceedings of the 22nd Annual Conference on Computational Complexity*, pages 155–168. IEEE, June 13–16 2007. 10, 24

[Vio6] E. Viola. Pseudorandom Bits for Constant-Depth Circuits with Few Arbitrary Symmetric Gates. *SIAM Journal on Computing*, 36(5):1387–1403, 2007. 30, 31

[VW] E. Viola and A. Wigderson. Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. In *Proceedings of the 22nd Annual Conference on Computational Complexity*. IEEE, June 13–16 2007. 6, 8, 30

[Yao1] A. C. Yao. Theory and Applications of Trapdoor Functions (Extended Abstract). In *23rd Annual Symposium on Foundations of Computer Science*, pages 80–91, Chicago, Illinois, 3–5 Nov. 1982. IEEE. 1

[Yao2] A. C.-C. Yao. Separating the polynomial-time hierarchy by oracles. In *Proc. 26th annual symposium on Foundations of computer science*, pages 1–10. IEEE Press, 1985. 1, 29