Comparing Computational Entropies Below Majority (Or: When Is the Dense Model Theorem False?)

Russell Impagliazzo
CSE Department, University of California San Diego, La Jolla, CA, USA
russell@cs.ucsd.edu

Sam McGuire
CSE Department, University of California San Diego, La Jolla, CA, USA
shmcguir@eng.ucsd.edu

Abstract

Computational pseudorandomness studies the extent to which a random variable $\mathbf{Z}$ looks like the uniform distribution according to a class of tests $\mathcal{F}$. Computational entropy generalizes computational pseudorandomness by studying the extent which a random variable looks like a *high entropy* distribution. There are different formal definitions of computational entropy with different advantages for different applications. Because of this, it is of interest to understand when these definitions are equivalent.

We consider three notions of computational entropy which are known to be equivalent when the test class $\mathcal{F}$ is closed under taking majorities. This equivalence constitutes (essentially) the so-called *dense model theorem* of Green and Tao (and later made explicit by Tao-Zeigler, Reingold et al., and Gowers). The dense model theorem plays a key role in Green and Tao's proof that the primes contain arbitrarily long arithmetic progressions and has since been connected to a surprisingly wide range of topics in mathematics and computer science, including cryptography, computational complexity, combinatorics and machine learning. We show that, in different situations where $\mathcal{F}$ is *not* closed under majority, this equivalence fails. This in turn provides examples where the dense model theorem is *false*.

2012 ACM Subject Classification Theory of computation → Pseudorandomness and derandomization

Keywords and phrases Computational entropy, dense model theorem, coin problem

Digital Object Identifier 10.4230/LIPIcs.ITCS.2021.2

Related Version Full Version: `https://arxiv.org/abs/2011.06166`

Funding Research supported by NSF award 1909634 and a Simons Investigator award.

1 Introduction

Computational pseudorandomness is a central topic in theoretical computer science. In this scenario, one has a class $\mathcal{F}$ of boolean functions $f : \{0,1\}^n \to \{0,1\}$ (which we'll refer to as *tests*) and random variable $\mathbf{Z}$ over $\{0,1\}^n$. We say that $\mathbf{Z}$ is ϵ-pseudorandom with respect to $\mathcal{F}$) if $\max_{f \in \mathcal{F}} |\mathbb{E}[f(\mathbf{Z})] - \mathbb{E}[f(\mathbf{U})]| \leq \varepsilon$ where $\mathbf{U}$ is the uniform distribution over $\{0,1\}^n$ and $\varepsilon > 0$ is small. In this case, we think of $\mathbf{Z}$ as "behaving like the uniform distribution" according to tests in $\mathcal{F}$. In general, say that two random variables $\mathbf{X}$, $\mathbf{Y}$ ε-indistinguishable by $\mathcal{F}$ if $\max_{f \in \mathcal{F}} |\mathbb{E}[f(\mathbf{X})] - \mathbb{E}[f(\mathbf{Y})]|$ (and so ε-pseudorandom distributions are exactly those which are ε-indistinguishable from $\mathbf{U}$). Constructing explicit $\mathbf{Z}$'s which behave like the uniform distribution according to different test classes is among the central goals of complexity theory, with sufficiently strong constructions leading to, for example, derandomization of BPP. One way in which the theory of pseudo-randomness is rich is that there are multiple equivalent formulations of pseudo-randomness, such as Yao's next bit test ([51]).

The various notions of pseudo-entropy and pseudo-density generalize pseudo-randomness to formalize how much randomness a distribution looks like it has as far as this class of tests can perceive. Many of these notions were first introduced as stepping stones towards pseudo-randomness, giving properties of sub-routines within constructions of pseudo-random generators. However, measuring seeming randomness quantitatively is important in many other contexts, so these notions have found wider application. For example, in mathematical subjects such as combinatorics and number theory, there is a general phenomenon of "structure vs. randomness", where a deterministically defined object such as a graph or set of integers can be decomposed into a structured part and a random part. Pseudo-entropy quantifies how much randomness the "random part" has. Notions of pseudo-density were used in this context by Green, Tao, and Ziegler [18, 48] to show that the primes contain arbitrarily long arithmetic progressions. We can also use pseudo-entropy notions to characterize the amount of seeming randomness remains n a cryptographic key after it has been compromised with a side-channel attack. A data set used in a machine learning algorithm might not have much randomness in itself, and might not be completely random looking, but is hopefully representative of the much larger set of inputs that the results of the algorithm will be applied to, so we can use notions of pseudo-entropy to say when such algorithms will generalize. There are many possible definitions of this intuitive idea, and as with pseudo-randomness, the power of pseudo-entropy is that many of these notions have been related or proven equivalent.

In particular, the dense model theorem provides such a basic equivalence. Here, the intuitive concept we are trying to capture is the density (or relative min-entropy) of the target distribution within a larger distribution, what fraction of the larger distribution is within the target. We say that $\mathbf{Z}$ is δ-dense if $\mathbb{E}[\mu(x)] = 2^{-n}\sum_x \mu(x) \geq \delta$ where $\mu : \{0,1\}^n \to [0,1]$ is density function defining $\mathbf{Z}$ (in the sense that $\Pr[\mathbf{Z} = z] = \mu(z)/(2^n\mathbb{E}[\mu(x)])$). One application of indistinguishability from a dense distribution is as a stepping stone to pseudorandomness: if $\mathbf{Z}$ is indistinguishable from a distribution $\mathbf{M}$ with density δ within the uniform distribution, then applying a randomness extractor with min-entropy rate $n - \log(1/\delta)$ to $\mathbf{Z}$ is a pseudorandom distribution. A more sophisticated application comes from additive number theory. It is not hard to show that a *random* subset of $[N] = \{1, 2, ..., N\}$ (including each element with probability 1/2, say) contains many arithmetic progressions (which are sets of the form $\{a, a+b, a+2b, a+3b, ...\}$). Szemerédi [45] showed that, in fact, sufficiently *dense* subsets of the integers also contain such arithmetic progressions: specifically, that for any k, the size of the largest subsets of $[N]$ which *doesn't* contain an arithmetic progression grows like $o(N)$.

So we would like some technology to reason about random variables $\mathbf{Z}$ which "behave like dense distributions". It turns out, however, that formalizing what it means for $\mathbf{Z}$ to "behave like a dense distribution" is subtle. Here are three perfectly legitimate candidates:

Candidate 1: $\mathbf{Z}$ *behaves like a δ-dense distribution if it behaves like something that's δ-dense.* Formally, this means that $\mathbf{Z}$ is ε-indistinguishable from some δ-dense distribution. In this case, we say that $\mathbf{Z}$ has a *δ-dense ε-model.*

Candidate 2: $\mathbf{Z}$ *behaves δ-dense if it's δ-dense inside of something that behaves like the uniform distribution.* Formally this means there's an ε-pseudorandom distribution $\mathbf{X}$ in which $\mathbf{Z}$ is δ-dense. In this case, we say that $\mathbf{Z}$ is *δ-dense in an ε-pseudorandom set.*

Candidate 3: $\mathbf{Z}$ *behaves δ-dense if it appears to be the case that conditioning on $\mathbf{Z}$ increases the size of any set by at most (roughly) a $1/\delta$-factor.* This is an operational definition: conditioning on a (truly) dense set increases the set by at most a $1/\delta$-fraction, so we should expect the same behavior from things that behave like a dense set. Formally, this means that $\delta\mathbb{E}[f(\mathbf{Z})] \leq \mathbb{E}[f(\mathbf{U})] + \varepsilon$ for any f in our test class $\mathcal{F}$. In this case, we say that $\mathbf{Z}$ has *(ε, δ)-pseudodensity.*

Precisely which definition you pick will depend on what you know about $\mathbf{Z}$ and in what sense you would like it to behave like a δ-dense distribution. Indeed, each of these definitions have appeared in different applications ([25], [18], [13], respectively), so there are scenarios where each of these types of behavior is desired. In general, the first candidate is the strongest (and, arguably, the most natural), but it is sometimes hard to establish that a distribution has the property. The following claim gives some simple relationships between the definitions:

▷ Claim 1. For any $\mathcal{F}$, the following hold:

1. If $\mathbf{Z}$ has a δ-dense ε-model, then $\mathbf{Z}$ is δ-dense in a ε-pseudorandom set.
2. If $\mathbf{Z}$ is δ-dense in an ε-pseudorandom set, then $\mathbf{Z}$ has (ϵ, δ)-pseudodensity.

Proof sketch.

1. Let $\mathbf{M}$ be the δ-dense ε-model for $\mathbf{Z}$. Note that $\mathbf{U} = \delta\mathbf{M}+(1-\delta)\overline{\mathbf{M}}$. So $\mathbf{U}' = \delta\mathbf{Z}+(1-\delta)\overline{\mathbf{M}}$ is ε-pseudorandom and $\mathbf{Z}$ is δ-dense within it.
2. Suppose $\mathbf{Z}$ is δ-dense in $\mathbf{Z}'$ which ε-pseudorandom for $\mathcal{F}$. Then for any $f \in \mathcal{F}$, $\delta\mathbb{E}[f(\mathbf{Z})] \leq \mathbb{E}[f(\mathbf{Z}')] \leq \mathbb{E}[f(\mathbf{U}] + \varepsilon$. ◁

The marvelous quality of these three candidates in particular is that, for many natural $\mathcal{F}$, all of them are *equivalent*, and so establishing even (ϵ', δ)-pseudodensity is enough to guarantee the existence of a δ-dense ε-model.

This equivalence holds for $\mathcal{F}$ which are *closed under majority*, meaning for any k (which we can think of as $k = O(1)$ for now), if $f_1, ..., f_k \in \mathcal{F}$ then $\mathsf{MAJ}_k(f_1, ..., f_k) \in \mathcal{F}$, where $\mathsf{MAJ} : \{0,1\}^n \to \{0,1\}$ is 1 if at least half of its input bits are 1. In fact, it holds for more general $\mathcal{F}$ if we allow the distinguishing parameter (ε' in (ε', δ)-pseudodensity) to be exponentially small (as in the original formulation, which we'll dicuss later on). In this case, the subtelty in defining what it means to behave like a dense set vanishes. These equivalences constitute (essentially) what is known as the *dense model theorem*, originating in the work of Green-Tao [18] and Tao-Zeigler [48], and independently in Barak et al. [8] (though in different guises). This result has been fruitfully applied in many seemingly unrelated areas of mathematics and computer science: additive number theory [18, 48] where $\mathcal{F}$ encodes additive information about subsets of $\{1, ..., N\}$ (or possibly a more general group), graph theory [49, 38] where $\mathcal{F}$ encodes cuts in a fixed graph, circuit complexity [49], Fourier analysis [29], machine learning [29] and leakage-resilient cryptography [14]. The ubiquity of the dense model theorem motivates a simple question: are there natural scenarios in which the dense model theorem is *false*?

We show that the answer to this question is *yes*. In particular, we show that for either implication from Claim 1 there is a class $\mathcal{F}$ and a random variable $\mathbf{Z}$ so that converse fails to hold. From the computational entropy perspective, we show that the three computational entropies we've discussed are inequivalent for certain test classes $\mathcal{F}$. Necessarily (with ε' not exponentially small) these classes are *not* closed under majority and so we will need to look "below" majority in order to find our counterexamples.

1.1 The dense model theorem

We turn to discuss the dense model theorem in some more detail to better contextualize our work. Restricting our attention to random variable over $\{0,1\}^n$, the dense model theorem states the following:

▶ **Theorem 1.1** (Dense model theorem). *Let $\mathcal{F}$ be a class of tests $f : \{0,1\}^n \to \{0,1\}$ and $\mathbf{Z}$ a random variable over $\{0,1\}^n$ with $(\varepsilon\delta, \delta)$-pseudodensity with respect to $\mathsf{MAJ}_k \circ \mathcal{F}$ for $k = O(\log(1/\delta)/\varepsilon^2)$. Then $\mathbf{Z}$ has a δ-dense ε-model with respect to $\mathcal{F}$.*

We will generally also consider a parameter ε', which in this case is $\varepsilon\delta$, the additive error in pseudodensity. To get an intuition for what this is saying, let's conisder a setting where it's false but for trivial reasons. As a simple example given in [52], pick a set $\mathbf{Z}$ some set as a $(1-\varepsilon)$ fraction of another set $\mathbf{S}$ of size $\delta 2^n$. Then $\mathbf{Z}$ doesn't have a δ-dense ε-model (i.e. $\mathbf{S}$) with respect to $\mathbf{Z}$'s indicator function, which we'll call f. On the other hand, the distribution $\mathbf{W}$ obtained by sampling $\mathbf{Z}$ with probability δ and sampling from $\mathbf{S}$'s complement with probability $1-\delta$ is at most $\varepsilon\delta$-distinguishable from $\mathbf{S}$ for any function, since $\varepsilon\delta$ is simply the measure of the difference between $\mathbf{S}$ and $\mathbf{Z}$. In particular $\mathbf{Z}$ is δ-dense in the $\varepsilon\delta$-pseudorandom $\mathbf{W}$ (which implies, via Claim 1, that it is $(\varepsilon\delta, \delta)$-pseudodense). This means that the Theorem 1.1 is tight for the dependence on $\varepsilon' = \varepsilon\delta$, in that it becomes false for $\Omega(\varepsilon\delta)$. In many instances, we think of $\varepsilon = 1/\mathsf{poly}(n)$, δ constant (or perhaps with mild dependences on n) and $\varepsilon' = \delta\varepsilon$.

Originally, the dense model theorem was proved with a different (and stronger) assumption; namely, that $\mathbf{Z}$ is dense in a pseudorandom set. Green and Tao, in proving that the primes contain arbitrarily long arithmetic progressions, used it to the following effect: if $\mathbf{Z}$ are the prime numbers up to n, then its density is known to behave like $\Theta(1/\log n)$. On the other hand, Szemerédi [45] showed that sufficiently dense subsets of $\mathbb{Z}$ contain arbitrarily long arithmetic progressions. The best bounds for Szemerédi 's theorem require density $\omega(1/\log\log n))$, which is much larger than the primes (see [16] and the recent [9] for more on the rich history on this and related problems). Not all is lost, however: the only property of dense sets that we're interested in is that they contain arithemtic progressions. So Green and Tao construct a class $\mathcal{F}$ of tests which can "detect" arithmetic progressions and under which the primes are dense inside of a $\mathcal{F}'$-pseudorandom set (more on $\mathcal{F}'$ later). By applying the dense model theorem, we conclude that the primes "look like" a dense set (themselves having long arithemtic progressions) with respect to the class $\mathcal{F}$. As $\mathcal{F}$ detects arithmetic progressions, it must be the case that the primes possess them. Of course, many details need to be filled in, but we hope this example shows the reader the "spirit" of the dense model theorem.

A primary source of interest in the dense model theorem is in the connections it shares with seemingly unrelated branches of mathematics and computer science. The original application was in additive number theory, but it was independently discovered and proved in the context of cryptography ([8, 14]). RTTV [38] and Gowers [17] observed proofs of the dense model theorem which use linear programming duality, which is in turn related to Nisan's proof of the hardcore lemma from circuit complexity [28]. In fact, Impagliazzo [29] shows in unpublished work that optimal-density versions of the hardcore lemma due to Holenstein [26] actually *imply* the dense model theorem. Klivans and Servedio [32] famously observed the relationship betweeen the hardcore lemma and *boosting*, a fundamental technique for aggregating weak learners in machine learning [15]. Together with the result of Impagliazzo, this connection means that dense model theorems can be proved by a particular type of boosting algorithm. A boosting argument for the existence of dense models also gives us *constructive* versions of the dense model theorem, which are needed for algorithmic applications. Zhang [52] (without using Impagliazzo's reduction from the dense model theorem to the hardcore lemma) used the boosting algorithm of [7] directly to prove the dense model theorem with optimal query complexity (k).

In addition to its connections to complexity, machine learning, additive number theory and cryptography, the dense model theorem (and ideas which developed from the dense model theorem, chiefly the approximation theorem of [49]), have been used to understand the weak graph regularity lemma of Frieze and Kannan [29], notions of computational differential privacy [36] and even generalization in generative adversarial networks (GANs) [5]. We now turn to discussing the complexity-theoretic aspects of the dense model theorem, specifically regarding our question of whether the MAJ_k from the statement is optimal.

As alluded to earlier, Green and Tao actually worked in a setting where $\mathcal{F}'$ doesn't need to compute majorities but where $\varepsilon\delta$ (that is, the distinguishing parameter in the pseudodensity assumption in the statement of Theorem 1.1) needs to be replaced by some $\varepsilon' = \exp(-\mathsf{poly}(1/\varepsilon, 1/\delta))$ (with $k = \mathsf{poly}(1/\delta, 1/\varepsilon)$ experiencing a small increase). We state this result, as proved in Tao and Zeigler [48] and stated this way in RTTV [38], for comparison. For a test class $\mathcal{F}$, let $\prod_k \mathcal{F}$ be the set of tests of the form $\prod_{i\in[k]} f_i$ for $f_i \in \mathcal{F}$.

▶ **Theorem 1.2** (Computationally simple dense-model theorem, strong assumption). *Let $\mathcal{F}$ be a class of tests $f : \{0,1\}^n \to [0,1]$ and $\mathbf{Z}$ a random variable over $\{0,1\}^n$ which is δ-dense in a set ε'-pseudorandom for $\prod_k \mathcal{F}$ with $k = \mathsf{poly}(1/\delta, 1/\varepsilon)$ and $\varepsilon' = \exp(-1/\delta, 1/\varepsilon)$. Then $\mathbf{Z}$ has a δ-dense ε-model with respect to $\mathcal{F}$.*

RTTV [38] observe that this proof can be adapted to work for ε' have polynomial dependence on ε, δ by restricting to the case of boolean-valued tests. Doing so, however, makes $\mathcal{F}'$ much more complicated (essentially requiring circuits of size exponential in k). In Theorem 1.1, we can obtain the best of both worlds: ε' has polynomial dependence on ε, δ and the complexity blow-up is rather small. However, in this more picturesque circumstance, we need to be able to compute majorities. Is such a tradeoff necessary? Our results suggest that the answer is yes. Theorem 1.6 (stated in the following section) tells us that if the dense model theorem is true for $\mathcal{F}$, then there's a small, constant-depth circuit with $\mathcal{F}$-oracle gates approximating majority on $O(1/\varepsilon^2)$ bits.

Another important aspect of the dense model theorem is how the different assumptions are related. As mentioned, the original assumption was that $\mathbf{Z}$ is δ-dense in an ε-pseudorandom set, but the proof can be extended to the case where $\mathbf{Z}$ is (ε, δ)-pseudodense. Claim 1 showed that the former assumption implies that latter assumption. When the dense model theorem is true, the latter also implies the former: simply apply the dense model theorem to $\mathbf{Z}$ which is (ε, δ)-dense to obtain a δ-dense ε-model. Then, by the first part of Claim 1, we're done.

First, we give examples of situations where these two notions are distinct. For example, we show in Theorem 1.4 and Theorem 1.5 that they are inequivalent when $\mathcal{F}$ is constant-depth polynomial size circuits or when $\mathcal{F}$ is a low-degree polynomial over a finite field. Note that a separation between pseudodensity and being dense in a pseudorandom set also implies a separation between pseudodensity and having a dense model, as being dense in a pseudorandom set is a necessary condition for having a dense model.

Second, we show that the dense model theorem is false even when we make the stronger assumption that the starting distribution $\mathbf{Z}$ is dense in a pseudorandom set. Specifically, in Theorem 1.3 we can show that some distributions $\mathbf{Z}$ are dense in a pseudorandom set but fail to have a dense model when $\mathcal{F}$ consists of constant-depth, polynomial size circuits.

Having contextualized our work some, we now turn to describe our contributions in more detail.

1.2 Contributions

We separate the previously described notions of computational entropy, giving examples where the dense model theorem is false. We are able to prove different separations when $\mathcal{F}$ is constant-depth unbounded fan-in circuits, low-degree polynomials over a finite field, and, in one case, any test class $\mathcal{F}$ which cannot efficiently approximate majority (in some sense made explicit later on). The only known separation prior was between pseudodensity and having a dense model for bounded-width read-once branching programs, due to Barak et al. [8].

Let $\mathcal{C}(S,d)$ denote the class of unbounded fan-in, size S, depth d circuits. We are generally thinking of $S = \mathsf{poly}(n)$ and $d = O(1)$, which corresponds to the complexity class AC^0. Theorem 1.3 shows that $\mathbf{Z}$ being δ-dense in an ε-pseudorandom set need not imply that $\mathbf{Z}$ has a δ-dense ε-model when the test class is $\mathcal{C}(S,d)$:

▶ **Theorem 1.3.** *Let $\varepsilon, \varepsilon' > 0$ be arbitrary, $\delta \geq \varepsilon'/8$ and*

$$S \leq \exp\left(O\Big(\frac{\sqrt{\varepsilon'}}{\varepsilon} \cdot \frac{\sqrt{\log(1/\delta)}}{\log(1/\varepsilon')} \Big)^{1/(d-1)} \right).$$

Then for $\mathcal{F} = \mathcal{C}(S,d)$, there is a random variable $\mathbf{D}$ over $\{0,1\}^n$ with $n = O(\log(1/\delta)/\varepsilon^2)$ so that $\mathbf{D}$ is δ-dense in an ε'-pseudorandom set but does not have a δ-dense ε-model. In particular, the dense model theorem is false in this setting.

Recall that the dense model theorem is false when $\varepsilon' = \Omega(\varepsilon\delta)$, which makes the restriction $\delta \geq \varepsilon'/8$ extremely mild. A common regime is $\varepsilon = 1/\mathsf{poly}(n)$, $\delta = O(1)$ and $\varepsilon' = \delta\varepsilon = \Theta(\varepsilon)$, in which case this gives us (essentially) a lower bound of weakly exponential in $1/\sqrt{\varepsilon} \approx 1/\sqrt{\varepsilon'}$.

Let $\mathbf{N}_\alpha$ denote the product distribution of n Bernoulli random variables with success probability $1/2 - \alpha$. Recall that density in a pseudorandom set readily implies pseudodensity, and one can use the dense model theorem to show the converse. We show that (ε,δ)-pseudodensity need not imply δ-density in an ε-pseudorandom set when the test class is $\mathcal{C}(S,d)$:

▶ **Theorem 1.4.** *Fix $\varepsilon, \varepsilon', \delta > 0$, $d \in \mathbb{N}$, and*

$$S \leq \exp\left(O\Big(\frac{\sqrt{\delta}}{\sqrt{\varepsilon}} \cdot \frac{\log(1/\delta)}{\log(1/\varepsilon')} \Big)^{1/(d-1)} \right).$$

Then $\mathbf{N}_{\sqrt{\varepsilon/\delta}}$ over $\{0,1\}^n$ with $n = O(1/\varepsilon)$ is (ε',δ)-pseudodense and yet $\mathbf{N}_{\sqrt{\varepsilon/\delta}}$ is not δ-dense inside of any ε-pseudorandom set.

The dependence ε' means that we can take ε' exponentially smaller than ε and still obtain a separation. This case corresponds to $\mathcal{F}$ being "very" fooled by $\mathbf{N}_\alpha$ but still not being δ-dense in a "mildly" pseudorandom set. This result draws on a recent line of work in the pseudorandomness literature – often referred to as "the coin problem" and studied in, e.g., [42, 12, 1, 46] – which concerns the ability of a test class $\mathcal{F}$ unable to compute majority has in distinguising $\mathbf{N}_\alpha$ and $\mathbf{U}$. We will discuss this connection in more detail during the proof overviews.

We prove a similar separation for degree-d $\mathbb{F}_p$-polynomials (on n variables), which generalizes (and uses techniques from) a recent result of Srinivasan [44] in the case where $\delta = 1$. In this case, we think of a distribution $\mathbf{Z}$ as being (ε',δ)-pseudodense for degree-d $\mathbb{F}_p$-polynomials when $\delta \Pr[P(\mathbf{Z}) \neq 0] - \varepsilon' \geq \Pr[P(\mathbf{U}) \neq 0]$ for any degree-d polynomial $P \in \mathbb{F}_p[X_1, ..., X_n]$ (noting that we are only evaluating P over $\{0,1\}^n$).

▶ **Theorem 1.5.** *Fix a finite field $\mathbb{F}$ with characteristic $p = O(1)$, $\varepsilon, \varepsilon' > 0$ and let $c > \delta > 0$ where $c \approx 1/200$ is an absolute constant. Suppose that*

$$d \leq O(\sqrt{\delta/\varepsilon}).$$

Then when $\mathcal{F}$ is the n-variate degree-d polynomials over $\mathbb{F}$ with $n = 1/\varepsilon$, and $\alpha = O(\sqrt{\varepsilon/\delta})$, $\mathbf{N}_\alpha$ is (ε',δ)-pseudodense but is not δ-dense inside of an ε-pseudorandom set.

This implies lower bounds for constant-depth circuits with MOD_p gates by the classical lower bounds of Razborov [37] and Smolensky [43]. Perhaps more interestingly, this holds even over non-prime fields. Also notably, there is no dependence on $\varepsilon' \leq \varepsilon\delta$, so we can take it to be arbitrarily small.

We also prove a more general separation between pseudodensity and density in a pseudorandom set. This result, drawing from the work of [42], provides a more specific characterization of the sense in which dense model theorems are "required" to compute majority.

▶ **Theorem 1.6.** *Let $\varepsilon, \delta > 0$. Suppose $\mathcal{F}$ is a test class of boolean functions $f : \{0,1\}^n \to \{0,1\}$ with the following property: there is no AC^0 $\mathcal{F}$-oracle circuit of size $\mathsf{poly}(n \cdot \frac{\sqrt{\delta}}{\varepsilon^{3/2}})$ computing majority on $O(\sqrt{\delta/\varepsilon})$ bits.*

Then $\mathbf{N}_{\sqrt{\varepsilon/\delta}}$ is $(\epsilon\delta, \delta)$-pseudodense and yet does not have a δ-dense ε-model. In particular, when the hypotheses are met, the dense model theorem is false.

Informally, this says that any $\mathcal{F}$ which can refute the pseudodensity of $\mathbf{N}_\alpha$ is only "a constant-depth circuit away" from computing majority.

1.3 Related work

Computational entropy

Computational entropy was studied systematically in [8] and is relevant to various problems in complexity and cryptography such as leakage-resilience [14], constructions of PRGs from one-way functions [25, 21, 20]. and derandomization [13].

There are a number of definitions of computational entropy which we *don't* consider in this work. For example, Yao pseudoentropy [51] (see also [8]), corresponding to random variables which are "compressible" by a class of tests $\mathcal{F}$, in the sense that $\mathcal{F}$ can encode and decode the random variable by encoding into a small number of bits. Yao pseudoentropy was recently used in time-efficient hardness-to-randomness tradeoffs [13], where (randomness-efficient) samplers for pseudodense distributions were used with an appropriate extractor to construct a pseudorandom distribution. Another example is *inaccessible entropy* of Haitner et al. [21], corresponding to the entropy of a message at some round in a two-player protocol conditioned on the prior messages and the randomness of the players, which is used in efficient constructions of statistically hiding commitment schemes from one-way functions [20].

Separating notions of computational entropy has been studied before in [8], who prove a separation of pseudodensity and having a dense model for bounded-width read-once branching programs. Separating notions of *conditional* computational entropy was studied in [27], showing separations between conditional variants of Yao pseudoentropy and having a dense model.

As mentioned in [27], citing [49] and personal communication with Impagliazzo, another question of interest is whether Yao pseudoentropy (corresponding to efficient encoding/-decoding algorithms) implies having dense model. It is not hard to see that small Yao pseudoentropy implies small pseudodensity, with some mild restrictions on $\mathcal{F}$. It would be interesting to see if the techniques from this paper can be used to understand Yao pseudoentropy in more detail. We leave this to future work.

Complexity of dense model theorems and hardness amplification

Prior work on the complexity of dense model theorems has included a tight lower bound on the query complexity [52] and a lower bound on the advice complexity [50]. As far as we are aware, this is the first work to consider the computational complexity of dense model theorems.

There has also been prior work on the computational complexity of hardness amplification, establishing that various known strategies for hardness amplification require the computation of majority [34, 42, 19, 41]. It is known that a particular type of hardness amplification given by the *hardcore lemma* implies the dense model theorem [29].

Our results are stronger in the following sense: previous work [34, 42, 19] shows that *black-box hardness amplification proofs* require majority. This means that if you amplify the hardness of f in some black-box way, then this can be used to compute majority. In our case, we simply show (in different settings) that the dense model theorem is *false*, regardless of how we tried to prove it. By the connection between the hardcore lemma and the dense model theorem, our results also provide scenarios where the hardcore lemma is false. As far as we are aware, these are the first such scenarios recorded in the literature.

1.4 Technical overview

We discuss two general themes that appear consistently in the proofs and then discuss each of the main theorems in some more detail.

1.4.1 Dense distributions have mostly unbiased bits

A commonly-used observation in theoretical computer science is that most bit positions of a δ-dense random variable over $\{0,1\}^n$ have bias $O(\sqrt{\log(1/\delta)/n})$ (see, for example, the introduction of [35]). Relevant to our purposes, it provides a *necessary* condition for having a δ-dense ε-model with respect to any class $\mathcal{F}$ containing the projections $z \mapsto z_i$. $\mathbf{Z}$ has a δ-dense ε-model, then most bits of $\mathbf{Z}$ have bias $\varepsilon + O(\sqrt{\log(1/\delta)/n})$. In particular, if all of the bits of $\mathbf{Z}$ have *large* bias, then it can't have a dense model.

This is used directly in the proof of Theorem 1.3. In this case, we construct a distribution $\mathbf{Z}$ which is δ-dense in a set which is ε-pseudorandom for AC^0 but where the each bit is noticeably biased away from 1/2.

In order to prove separations between pseudodensity and being dense in a pseudorandom set – as in Theorem 1.4, Theorem 1.5 and Theorem 1.6 – we need to consider the bias of larger subsets of variables. Considering just two bits is sufficient to prove mild concentration bounds on the weight of pseudorandom strings. This implies that the tails of dense subsets of pseudorandom sets should not be too heavy.

1.4.2 Biased coin distribution

The *biased coin distribution*, $\mathbf{N}_\alpha$ over $\{0,1\}^n$ is the product of n Bernoulli random variables with success probability $1/2 - \alpha$. $\mathbf{N}_\alpha$ has recently garnered significant interest in the pseudorandomness literature (see [2, 12, 46, 10, 1]). Shaltiel and Viola [42] showed that if f is a test which ε-distinguishes $\mathbf{N}_\alpha$ from $\mathbf{U}$, then there is a small, constant-depth circuit C with f-oracle gates which computes majority on $O(1/\varepsilon)$ bits. A similar, but qualitatively different, connection due to Limaye et al [33] – extended to any choice of $\varepsilon > 0$ by Srinivasan [44] – shows that any $\mathbb{F}_p$-polynomial with advantage $1-2\varepsilon$ in distinguishing $\mathbf{N}_\alpha$ from $\mathbf{U}$ must have degree $\Omega(\log(1/\varepsilon)/\alpha)$. We extend some of these pseudorandomness results regarding $\mathbf{N}_\alpha$ to *pseudodensity* results.

First, we extend the observation of Shaltiel and Viola to apply to tests f for which $\mathbb{E}[f(\mathbf{Z})] \geq \delta\mathbb{E}[f(\mathbf{U})] + \varepsilon$ (which corresponds to pseudorandomness when $\delta = 1$). This gives us unconditional pseudodensity for test classes $\mathcal{F}$ which can't be used in small, constant-depth oracle circuits approximating majority. We also extend the observation of [33] to show lower bounds on the $\mathbb{F}_p$-degree for any function f which refutes the pseudodensity of $\mathbf{N}_\alpha$.

In Lemma 13, we show that $\mathbf{N}_\alpha$ exhibits (ε, δ)-pseudodensity for $\varepsilon = (p \cdot O(\log S)^{d-1})^k$ and $\delta = e^{-\alpha k/p}$. This can be seen as a generalization of Tal's result, building on [12, 1, 42] that $\mathbf{N}_\alpha$ is $3\alpha \cdot O(\log S)^{d-1}$-pseudorandom for $\mathcal{C}(S, d)$.

Tal uses a Fourier analytic proof which becomes very simple given tail bounds on the Fourier spectrum of AC^0 (the latter being the main contribution of [46]). More generally, any $\mathcal{F}$ enjoying sufficiently strong tail bounds on the Fourier spectrum (in the ℓ_1 norm) cannot distinguish between $\mathbf{N}_\alpha$ and uniform. It turns out, as proved by Tal and recorded in Agarwal [2], that if $\mathcal{F}$ is closed under restrictions than even bounding the first level of the Fourier spectrum works. The proof of Lemma 13 based specifically on the switching lemma for constant-depth circuits. While switching lemmas can be used to show Fourier concentration, it would be intersting to find a proof which only uses the assumption of Fourier concentration (or some Fourier-analytic assumption).

1.4.3 Theorem 1.3

Our goal is to construct a random variable $\mathbf{D}$ which is dense inside of an AC^0-pseudorandom set but where each bit is biased away from 0. In this case, $\mathbf{D}$ would be distinguishable from any dense set, since the average bit of a dense set is roughly unbiased. Doing so requires two steps.

The first step is constructing an appropriate distribution $\mathbf{Z}$ that fools AC^0 circuits. For this we adopt a general strategy of Ajtai and Wigderson [3] (and applied in many contexts in pseudorandomness since; see, e.g., [40]): to fool a circuit C, we start by producing a random restriction to simpify C to a short decision tree (via the switching lemma), and then we fool the decision tree on the remaining bits using a k-wise independent distribution $\mathbf{S}$. If we wanted $\mathbf{Z}$ to have small support size, we would need some way of producing random restrictions with a small amount randomness (which is precisely the approach of Ajtai-Wigderson and later work). Fortunately, we only care about the existence of $\mathbf{Z}$ and are therefore content to use the "non-derandomized" switching lemma.

The second step is finding a dense subset $\mathbf{D}$ of $\mathbf{S}$ with biased bits. We do this by constructing $\mathbf{S}$ so that each bit has bias roughly $\sqrt{\log(1/\delta)/K}$, where $k \ll K \ll n$ is a parameter. This is achieved by randomly bucketing the indices into K buckets and assigning each bucket a random bit, which reduces the dimension of the problem from n to K. This means we can pick a δ-dense event in $\{0, 1\}^K$ with extremal bias – met (up to constants) by the function accepting all strings with weight less than $K/2 - K\sqrt{\log(1/\delta)}$ – in order to find a dense subset of $\mathbf{S}$ with large bias. The bucketing construction introduces some error when a small set $I \subseteq [n]$ hits to distinct elements in some buckets.

1.4.4 Theorem 1.4

We will show $\mathbf{N}_\alpha$ has (δ, ε')-pseudodensity for AC^0 for $\delta = \varepsilon' = O(1)$, $\alpha = 1/\mathsf{poly}\log(n)$. The idea is that $\mathbf{N}_\alpha$ can be sampled by first sampling a random restriction which leaves a p fraction of the bits unset (and is unbiased on the restricted bits) and then setting the remaining bits with bias α/p. Applying the switching lemma, we conclude that $\mathbb{E}[f(\mathbf{N}_\alpha)] \approx \mathbb{E}[f'(\mathbf{N}_{\alpha/p})]$ where f' is a short decision tree (which doesn't not depend on all of its inputs). A simple calculation reveals that acceptance probability of f' can increase by at a most a factor $(1 + \alpha/p)^d \leq e^{\alpha d/p}$ when passing from the uniform distribution to $\mathbf{N}_{\alpha/p}$. By incorporating the error from the switching lemma (i.e. the advantage lost by conditioning on the switching lemma succeeding), we get (δ, ϵ)-pseudodensity.

To prove the separation, we use the fact that the Hamming weight of a random variable fooling $\mathcal{C}(S,d)$ is concentrated around its expectation. This means in particular that if $\mathbf{N}_\alpha$ *were* δ-dense in a pseudorandom distribution, then the tails of $\mathbf{N}_\alpha$ couldn't be too heavy and therefore α couldn't be too large.

1.4.5 Theorem 1.5 and Theorem 1.6

Theorem 1.5 and Theorem 1.6 draw from related work of Srinivasan [44] and Shaltiel-Viola [42] respectively.

With $\epsilon > 0$ and $\mathcal{F}$ an arbitrary class of tests $f : \{0,1\}^n \to \{\pm 1\}$, suppose that $f \in \mathcal{F}$ witnesses that $\mathbf{N}_\varepsilon$ *fails* to have (ε', δ)-pseudo-density in the sense that

$$\mathbb{E}[f(\mathbf{U})] \leq \delta \mathbb{E}[f(\mathbf{N}_\beta)] - \gamma.$$

[44] and [42] both make use of the following simple observation. Given two strings $u, v \in \{0,1\}^m$ with $\mathrm{wt}(u) = (1/2 - \varepsilon)m$ and $\mathrm{wt}(v) = m/2$, a uniformly random index $i \in [m]$ has u_i distributed as a $(1/2-\varepsilon)$-biased coin and v_i as an unbiased coin. In our case, applying f to sufficiently many random samples from u or v "distinguishes" the two of them, but in a weaker sense.

In the case of Theorem 1.6, we can amplify acceptance probabilities by increasing the size of the circuit by a factor $1/\varepsilon\delta$, after which we can apply [42] saying that constant-error distinguishers between $\mathbf{N}_\alpha$ and $\mathbf{U}$ can be used to compute majority.

For Theorem 1.5, we apply a beautiful recent result of Srinivasan [44] showing that any m-variate polynomial (over a finite field) which vanishes on most points on the slice $1/2 - \alpha$ and doesn't vanish on most points on the slice $1/2$ must have high degree $\Omega(\alpha m)$. One way of interpreting this result is that low-degree polynomials can't approximately solve certain "promise" versions of majority.

In this latter case, we need to open up the error reduction procedure we use for Theorem 1.6 and show how to approximate it using low-degree polynomials. This will ultimately be achieved by approximating OR with a probabilistic polynomial, as in [37, 43]. The detailed proofs of these results are deferred to the full version of the paper.

2 Technical tools

We write $[n] = \{1, ..., n\}$ and use boldface to denote random variables. Let $\mathcal{C}(S,d)$ be the set of size S, depth-d unbounded fan-in circuits. For a boolean function $f : \{0,1\}^n \to \{0,1\}$, let $DT(f)$ denote the depth of the shortest decision tree computing f.

2.1 Biased coins

As before, let $\mathbf{N}_\alpha$ denote the random variable corresponding to the product of n independent coins with bias $(1/2 - \alpha)$. That is,

$$\Pr[\mathbf{N}_\alpha = z] = (1/2 - \alpha)^{\mathrm{wt}(z)}(1/2 + \alpha)^{n - \mathrm{wt}(z)}$$

where $\mathrm{wt}(z)$ denotes the Hamming weight of z.

For a random variable $\mathbf{Z}$ over $\{0,1\}^n$ and $i \in [n]$, let $\mathrm{bias}_i(\mathbf{Z}) = |\Pr[\mathbf{Z}_i = 1] - \Pr[\mathbf{Z}_i = 0]|/2$. Let $\mathcal{B} = \{z \mapsto z_i : i \in [n]\}$ be the set of monotone projections. A random variable $\mathbf{Z} = (\mathbf{Z}_1, ..., \mathbf{Z}_n)$ is ϵ-pseudorandom with respect to $\mathcal{B}$ precisely when each marginal $\mathbf{Z}_i$ has the property that $\mathrm{bias}_i(\mathbf{Z}) = |\Pr[\mathbf{Z}_i = 1] - 1/2| \leq \varepsilon$ for each $i \in [n]$. In particular,

▷ Claim 2. For any $\varepsilon > 0$, $\mathbf{N}_\varepsilon$ is ε-pseudorandom with respect to $\mathcal{B}$.

2.2 Information theory

The *(Shannon) entropy* of a random variable is defined as

$$H(\mathbf{Z}) = -\sum_{x\in\{0,1\}^n} p_{\mathbf{Z}}(x)\log p_{\mathbf{Z}}(x),$$

where $p_{\mathbf{Z}}$ is the probability density function corresponding to $\mathbf{Z}$. The Shannon entropy of random vector is sub-additive, in that $H(\mathbf{Z}) \leq \sum_{i\in[n]} \mathbf{Z_i}$. When $\mathbf{Z} \in \{0,1\}$ and $\Pr[\mathbf{Z}=1] = p$, we use $h(p) = H(\mathbf{Z}) = -(p\log p + (1-p)\log(1-p))$ to denote the binary entropy function.

The *min-entropy* is defined as

$$H_\infty(\mathbf{Z}) = -\min_{x\in\{0,1\}^n} \log p_{\mathbf{Z}}(x).$$

If $\mathbf{Z}$ is δ-dense inside of $\mathbf{U}$, then its min-entropy is $n - \log(1/\delta)$ and for any random variable $\mathbf{Z}$, $H_\infty(\mathbf{Z}) \leq H(\mathbf{Z})$.

By this latter inequality and subadditivity, the average entropy of $\mathbf{Z}$'s bits is at least $1 - \log(1/\delta)/n$. Appealing to a quadratic approximation of binary entropy, we learn that the bias must be at most $\sqrt{\log(1/\delta)/n}$. This result has been referred to as *Chang's inequality* and the *Level-1 inequality*, having been observed in different forms and with different proofs in, for example, [47, 11, 22, 31]. Because it is so simple, we provide a proof here:

▷ Claim 3. If $\mathbf{Z}$ is δ-dense in $\mathbf{U}$, then $\mathbb{E}_i[\text{bias}_i(\mathbf{Z})] \leq \sqrt{\log(1/\delta)/n}$.

Proof. As δ-density is equivalent to $n - \log(1/\delta)$ min-entropy,

$$n - \log(1/\delta) = H_\infty(\mathbf{Z}) \leq H(\mathbf{Z}) \leq \sum_{i\in[n]} H(\mathbf{Z}_i),$$

by subadditivity of entropy. The entropy of $\mathbf{Z}_i$'s bits, therefore, is at least $1 - \log(1/\delta)/n$ on average. Taking the Taylor series, we can approximate the binary entropy function $h(p)$ around $1/2$ by a quadratic function as $h(1/2+\varepsilon) \leq 1 - (2/\ln 2)\varepsilon^2$. Comparing this bound with the average, we get

$$1 - \log(1/\delta) \leq 1 - (2/\ln 2)\varepsilon^2,$$

meaning $\varepsilon \leq \sqrt{(\ln 2/2)\cdot(\log(1/\delta)/n)} \leq \sqrt{\log(1/\delta)/n}$. ◁

2.3 Random variables lacking computational entropy

It follows directly from Claim 3 that if $\text{bias}_i(\mathbf{Z}_i)$ exceeds $\varepsilon + \sqrt{\log(1/\delta)}/n$ for every i, then $\mathbf{Z}$ does not have a δ-dense ε-model with respect to the projections $\mathcal{B}$.

▶ **Lemma 4.** *Let $\mathbf{Z}$ be a random variable with $\text{bias}_i(\mathbf{Z}) \leq \gamma$ for every $i \in [n]$. Then for any $\delta > 0$ and $\gamma \geq \epsilon + \sqrt{\frac{\log(1/\delta)}{n}}$, $\mathbf{Z}$ does not have a δ-dense ϵ-model with respect to $\mathcal{B}$.*

This is used for the separation in Theorem 1.3. We would also like a necessary condition for being dense in a pseudorandom set. Towards this end, we note that pseudorandom distributions for even very simple test classes have mild concentration properties.

▷ Claim 5. Suppose $\mathcal{F}$ can compute $x_i \oplus x_j$ for every $i, j \in [n]$ and let $\mathbf{Z}$ over $\{0,1\}^n$ be ε-pseudorandom for $\mathcal{F}$. Then

$$\Pr\left[\sum_i \mathbf{Z_i} \leq n/2 - \alpha n\right] \leq \frac{1}{4\alpha^2 n} + \frac{\varepsilon}{4\alpha^2}.$$

Proof. We work over $\{\pm 1\}$ instead of $\{0,1\}$ to make calculations easier. We can compute the second moment as

$$\mathbb{E}[(\sum_i \mathbf{Z}_i)^2] = \sum_i \mathbb{E}[\mathbf{Z}_i^2] + \sum_{i \neq j} \mathbb{E}[\mathbf{Z}_i \mathbf{Z}_j] \leq n + \varepsilon n^2.$$

Applying Markov's inequality to $(\sum_i \mathbf{Z}_i)^2$, we see that

$$\Pr\left[|\sum_i \mathbf{Z}_i| \geq 2\alpha n\right] = \Pr\left[(\sum_i \mathbf{Z}_i)^2 \geq (2\alpha n)^2\right] \leq \mathbb{E}[(\sum_i \mathbf{Z}_i)^2]/(2\alpha n)^2.$$

We use $2\alpha n$ because it maps back to $n/2 - \alpha n$ in $\{0,1\}$. Then the conclusion follows from our second moment calculation and converting back to $\{0,1\}$. ◁

The tails of a dense subset can't be too much larger than the original distribution, by definition of density. This gives us a test for being dense in a pseudorandom set, which we specialize to $\mathbf{N}_\alpha$.

▶ **Lemma 6.** *Let $\varepsilon, \delta > 0$ be arbitrary. Suppose $\mathcal{F}$ can compute $x_i \oplus x_j$ for any $i, j \in [n]$ and $\alpha \geq \sqrt{1/(8\delta) \cdot (1/n + \varepsilon)}$. Then $\mathbf{N}_\alpha$ is not δ-dense in any set which is ε-pseudorandom for $\mathcal{F}$.*

Proof. Under $\mathbf{N}_\alpha$, the volume of the threshold $\mathbf{1}[\sum_i \mathbf{Z}_i \leq n/2 - \alpha n]$ is $1/2$. Taking Claim 5 in the contrapositive, we reach the desired conclusion when

$$\begin{aligned} 1/2 &> \frac{1}{4\delta\alpha^2 n} + \frac{\varepsilon}{4\delta\alpha^2} \\ \alpha^2 &> \frac{1}{8\delta}(1/n + \varepsilon). \end{aligned}$$ ◀

2.4 Random restrictions and the switching lemma

A restriction over $[n]$ is a function $\rho : [n] \to \{0, 1, *\}$. Indices in $\rho^{-1}(*)$ can be thought of as *unset* and each other index as *set*. For another restriction z so that $\rho^{-1}(*) \subseteq z^{-1}(\{0,1\})$, let $\rho \circ z \in \{0,1\}^n$ be defined by

$$(\rho \circ z)_i = \begin{cases} z_i \text{ if } i \in \rho^{-1}(*), \\ \rho_i \text{ otherwise.} \end{cases}$$

Define the restricted function $f|_\rho : \{0,1\}^{\rho^{-1}(*)} \to \{0,1\}$ over ρ's unset indices by

$$f|_\rho(z) = f(\rho \circ z).$$

Let R_p be the distribution on restrictions over $[n]$ obtained by setting $\rho(i) = *$ independently with probability p, and then setting each bit not assigned to $*$ a random bit. The switching lemma we use is due to Rossman [39], building on a long line of work [3, 23, 24, 30]:

▶ **Theorem 2.1** (Rossman [39]). *Suppose $f \in \mathcal{C}(S, d)$. Then*

$$\Pr_{\boldsymbol{\rho} \sim R_p}[DT(f|_{\boldsymbol{\rho}}) \geq k] \leq (p \cdot O(\log S)^{d-1})^k.$$

By considering a random restriction $\boldsymbol{\rho} \sim R_p$ over $[n]$ and a random variable $\mathbf{Z}$ over $\{0,1\}^n$, the definition of a restricted function implies that

$$\mathbb{E}[f(\boldsymbol{\rho} \circ \mathbf{Z})] = \mathbb{E}[f|_{\boldsymbol{\rho}}(\mathbf{Z})].$$

We make crucial use of two simple corollaries of the switching lemma, which allow us to reason about distinguishability for AC^0 circuits in terms of distinguishability for short decision trees.

▶ **Lemma 7.** *Suppose $f \in \mathcal{C}(S,d)$. Then there is a distribution over depth k decision trees so that*

$$|\mathbb{E}[f(\boldsymbol{\rho} \circ \mathbf{Z})] - \mathbb{E}[h_{\boldsymbol{\rho}}(\mathbf{Z})]| \leq (p \cdot O(\log S)^{d-1})^k.$$

Proof. Let $g_{\boldsymbol{\rho}}$ denote the optimal decision tree for $f|_{\boldsymbol{\rho}}$. Let E denote the event that $g_{\boldsymbol{\rho}}$ has depth at most k and $\Pr[E] = 1-q$. Let $h_{\boldsymbol{\rho}}$ be the distribution over depth at most k decision trees obtained by sampling $g_{\boldsymbol{\rho}}$ conditioned on E. Then

$$\begin{aligned}\mathbb{E}[f(\boldsymbol{\rho} \circ \mathbf{Z})] &= \mathbb{E}[f|_{\boldsymbol{\rho}}(\mathbf{Z})]\\ &= (1-q)\mathbb{E}[g_{\boldsymbol{\rho}}(\mathbf{Z})|E] + q\mathbb{E}[g_{\boldsymbol{\rho}}(\mathbf{Z})|\neg E]\\ &= (1-q)\mathbb{E}[h_{\boldsymbol{\rho}}(\mathbf{Z})] + q\mathbb{E}[g_{\boldsymbol{\rho}}(\mathbf{Z})|\neg E]\\ &= \mathbb{E}[h_{\boldsymbol{\rho}}(\mathbf{Z})] - q(\mathbb{E}[h_{\boldsymbol{\rho}}(\mathbf{Z})] - \mathbb{E}[g_{\boldsymbol{\rho}}(\mathbf{Z})|\neg E]).\end{aligned}$$

The right-hand term is bounded in absolute value by q because f is Boolean. By Theorem 2.1, $q \leq (p \cdot O(\log S)^{d-1})^k$. ◀

▶ **Lemma 8.** *Suppose $f \in \mathcal{C}(S,d)$. Then there's a depth k decision tree h so that*

$$|\mathbb{E}[f(\mathbf{U})] - \mathbb{E}[f(\boldsymbol{\rho} \circ \mathbf{Z})]| \leq |\mathbb{E}[f'(\mathbf{U})] - \mathbb{E}[f'(\mathbf{Z})]| + (p \cdot O(\log S)^{d-1})^k.$$

Proof. Lemma 7 gives us the following upper bound.

$$\begin{aligned}|\mathbb{E}[f(\mathbf{U})] - \mathbb{E}[f(\boldsymbol{\rho} \circ \mathbf{Z})]| &\leq |(\mathbb{E}[h_{\boldsymbol{\rho}}(\mathbf{U})] \pm q) - (\mathbb{E}[h_{\boldsymbol{\rho}}(\mathbf{Z})] \pm q)| && \text{(Lemma 7)}\\ &\leq |\mathbb{E}[h_{\boldsymbol{\rho}}(\mathbf{U})] - \mathbb{E}[h_{\boldsymbol{\rho}}(\mathbf{Z})]| + 2q. && \text{(triangle inequality)}\end{aligned}$$

We can continue to upper bound the right-hand term by

$$\begin{aligned}|\mathbb{E}[h_{\boldsymbol{\rho}}(\mathbf{U})] - \mathbb{E}[h_{\boldsymbol{\rho}}(\mathbf{Z})]| &= |\mathbb{E}_{\boldsymbol{\rho}}[\mathbb{E}[h_\rho(\mathbf{U})] - \mathbb{E}[h_\rho(\mathbf{Z})]]|\\ &\leq \mathbb{E}_{\boldsymbol{\rho}}[|\mathbb{E}[h_\rho(\mathbf{U})] - \mathbb{E}[h_\rho(\mathbf{Z})]|] && \text{(triangle inequality)}\\ &\leq |\mathbb{E}[h(\mathbf{U})] - \mathbb{E}[h(\mathbf{Z})]|\end{aligned}$$

where the last line holds for some h in the support of $h_{\boldsymbol{\rho}}$ by averaging. ◀

3 Proof of Theorem 1.3

We start by reducing the problem of constructing a pseudorandom $\mathbf{Z}$ for AC^0 to constructing a pseudorandom $\mathbf{Z}$ for small-depth decision trees. This can be immediately achieved by applying Lemma 8.

▷ Claim 9. Let $p \in [0,1]$ be arbitrary and suppose $\mathbf{Z}$ is a random variable over $\{0,1\}^n$ which is ϵ-pseudorandom for depth-k decision trees. Then for $\boldsymbol{\rho} \sim R_p$, $\boldsymbol{\rho} \circ \mathbf{Z}$ is ϵ'-pseudorandom for $\mathcal{C}(S,d)$ for

$$\epsilon' = \epsilon + (p \cdot O(\log S)^{d-1})^k.$$

The next lemma constructs a pseudorandom distribution for depth-k decision trees with each bit having significant bias.

▶ **Lemma 10.** *For any $k \in \mathbb{N}, \delta > 0$ and $K \geq 1/2\delta$, there is a k-wise independent random variable $\mathbf{S}$ over $\{0,1\}^n$ and a δ-dense subset $\mathbf{D}$ of $\mathbf{S}$ with the property that*

1. *$\mathbf{D}$ is δ-dense in $\mathbf{S}$.*
2. *For all $i \in [n]$, $\mathrm{bias}_i(\mathbf{D}) = \Omega(\sqrt{\log(1/\delta)/8K})$.*
3. *$\mathbf{S}$ is k^2/K-pseudorandom for depth-k decision trees.*

We will use the following standard lower bound on the lower tail of a binomial distribution:

▷ Claim 11 ([6]). For $0 < \alpha < 1$ and let $\mathbf{Z}_1, ..., \mathbf{Z}_K$ be independent unbiased coins ($\{0,1\}$-valued). Then any γ with $1/2 - \gamma = r/K$ for some positive integer r satisfies

$$\frac{2^{-K(1-h(1/2-\gamma))}}{\sqrt{2K}} \leq \Pr\left[\sum_{i\in[K]} \mathbf{Z}_i \leq K/2 - K\gamma\right].$$

Proof of Lemma 10. We sample $\mathbf{S}$ in two stages. First, randomly partition $[n]$ into K parts $\mathbf{A}_1, ..., \mathbf{A}_K$ for $K > k^2$. Second, assign to each A_i a uniformly random bit $\mathbf{b}_i$.

Let $\mathbf{D}$ be $\mathbf{S}$ conditioned on $\mathbf{b} = (\mathbf{b}_1, ..., \mathbf{b}_K)$ having weight less than $K/2 - \sqrt{K\log(1/\delta)/8}$. Since the $\mathbf{b}_i$'s are unbiased random bits, we can apply Claim 11 to lower bound $\mathbf{D}$'s density: for any γ,

$$\Pr\left[\sum_{i\in[k]} \mathbf{b}_i \leq \gamma K\right] \geq \frac{2^{-Kh(1/2-\gamma)}}{\sqrt{2K}}.$$

This is at least δ when

$$\begin{aligned}\frac{2^{-K(1-h(1/2-\gamma))}}{\sqrt{2K}} &\geq \delta \\ 1 - h(1/2-\gamma) &\geq \log(1/\delta)/K - \log(2K)/2K \\ 4\gamma^2 &\geq \log(1/\delta)/K - \log(2K)/2K\end{aligned}$$

with the upper bound in the last line following from $h(1/2-\gamma) \geq 1 - 4\gamma^2$. Hence, if the set of strings with weight at most $K/2 - \gamma K$ is δ-dense, we have $\gamma \geq \frac{1}{2}\sqrt{\log(1/\delta)/K - \log(2K)/2K}$. $\log(2K)/2K$ is at most $\log(1/\delta)/2K$ when $2K \leq 1/\delta$, in which case $\gamma \geq \sqrt{\log(1/\delta)/8K}$. In particular, this lower bounds the bias of $\mathbf{D}$'s bits.

To see why it's k^2/K-pseudorandom for depth-k decision trees, consider a depth-k decision tree T. Over $\mathbf{U}$, we can imagine evaluting T "on-line" as follows: whenever T queries the ith bit, determine the value of z_i by flipping an unbiased coin. Over $\mathbf{S}$, we can imagine evaluating T similarly, where we determine the bucket A_j that i lives in and the value b_j of that bucket.

By conditioning $\mathbf{S}$ on *not* placing two distinct indices i, j in the same bucket – call this conditioned random variable $\mathbf{S}'$ – then T doesn't have *any* distinguish advantage over $\mathbf{S}'$, as all of the bits it queries are independent and uniform. By a union bound, $\mathbf{S}$ places two distinct indices in the same bucket with probability at most k^2/K. T's distinguishing advantage is therefore at most k^2/K. ◀

In principle, we could have used other pseudorandom distributions for decision trees such as the ε-almost k-wise independent distributions from [4]. The construction here is used to obtain better dependence on the parameters of interest. We will also need a claim to expresses the bias of the bits in $\boldsymbol{\rho} \circ \mathbf{Z}$. The proof can be found in the full version of the paper.

▷ Claim 12. Fix $p \in [0,1]$ and a random variable $\mathbf{Z}$. Let E be an event which is independent from ρ (in that the conditional distribution of ρ is identical to the unconditioned distribution). Then

$$\Pr[(\boldsymbol{\rho} \circ \mathbf{Z})_i = 1|E] = p\Pr[\mathbf{Z}_i = 1|E] + (1-p)/2.$$

Theorem 1.3, which we restate here, is obtained by an appropriate setting of parameters.

▶ **Theorem 1.3.** *Let $\varepsilon, \varepsilon' > 0$ be arbitrary, $\delta \geq \varepsilon'/8$ and*

$$S \leq \exp\left(O\left(\frac{\sqrt{\varepsilon'}}{\varepsilon} \cdot \frac{\sqrt{\log(1/\delta)}}{\log(1/\varepsilon')}\right)^{1/(d-1)}\right).$$

Then for $\mathcal{F} = \mathcal{C}(S,d)$, there is a random variable $\mathbf{D}$ over $\{0,1\}^n$ with $n = O(\log(1/\delta)/\varepsilon^2)$ so that $\mathbf{D}$ is δ-dense in an ε'-pseudorandom set but does not have a δ-dense ε-model. In particular, the dense model theorem is false in this setting.

Proof. Let $n = \log(1/\delta)/\varepsilon^2$, $k = \log(2/\varepsilon')$ and $K = (2k^2)/\varepsilon'$. We also need $K \geq 1/2\delta$ by the restriction in Lemma 10, which explains the restriction $8\delta k^2 \geq \varepsilon'$, simplified by using $8\delta \geq \varepsilon'$ (a stronger restriction) instead. Let $\mathbf{S}$ and $\mathbf{D}$ be the random variables from Lemma 10. By Claim 12, the bias of $\boldsymbol{\rho} \circ \mathbf{S}$ (where $\boldsymbol{\rho} \sim R_p$) is $p\sqrt{\log(1/\delta)/8K}$. By Claim 9 and Lemma 10, $\boldsymbol{\rho} \circ \mathbf{S}$ is $\varepsilon' = k^2/K + (pO(\log S)^{d-1})^k$ pseudorandom. We can also ensure that $\boldsymbol{\rho} \circ \mathbf{S}$ does *not* have a δ-dense ε-model when $p\sqrt{\log(1/\delta)/8K} \geq \varepsilon + \sqrt{\log(1/\delta)/n}$, by Lemma 4.

By substituting, $p \geq 2\sqrt{K/n} = 2\sqrt{(k\varepsilon)^2/\varepsilon' \cdot \log(1/\delta)}$. In comparison, $\varepsilon' \geq k^2/K + (pO(\log S)^{d-1})^k$. Recalling that $k^2/K = \varepsilon'/2$, we get that

$$\begin{aligned}
\varepsilon'/2 &\geq (2\sqrt{K/n}O(\log S)^{d-1})^k \\
\frac{\sqrt{n}}{2\sqrt{K}}(\varepsilon'/2)^{1/k} &\geq O(\log S)^{d-1} \\
\frac{\sqrt{\log 1/\delta}}{\varepsilon} \cdot \frac{\sqrt{\varepsilon'}}{2\sqrt{2}k} \cdot (\varepsilon'/2)^{1/k} &\geq O(\log S)^{d-1} \\
\frac{\sqrt{\log 1/\delta}}{\varepsilon} \cdot \frac{\sqrt{\varepsilon'}}{\sqrt{32}\log(1/\varepsilon')} &\geq O(\log S)^{d-1}.
\end{aligned}$$

The claim follows by solving for S. ◀

4 Proof of Theorem 1.4

Theorem 1.4 follows by combining Lemma 6 and the following lemma:

▶ **Lemma 13.** *$\mathbf{N}_\alpha$ has (ε, δ)-pseudodensity for $\mathcal{C}(S,d)$ for $\varepsilon = (p \cdot O(\log S)^{d-1})^k$ and $\delta = e^{-\alpha k/p}$.*

Of note, the only additive error depends on the error from the switching lemma. Compare this with the claim that $\mathbf{N}_\alpha$ is $(3\alpha \cdot O(\log S)^{d-1})$-pseudorandom (and therefore has the same pseudodensity for $\delta = 1$) for $\mathcal{C}(S,d)$, due to Tal [46].

To prove the lemma, we need a few claims.

▷ Claim 14. Suppose $f \in \mathcal{C}(S,d)$. Then there is a depth-k decision tree h with the property that:

$$\mathbb{E}[f(\mathbf{N}_\alpha)] \le \mathbb{E}[h(\mathbf{N}_{\alpha/p})] + (p \cdot O(\log S)^{d-1})^k.$$

Proof. Take $\mathbf{Z} = \mathbf{N}_{\alpha/p}$ in Lemma 7, so we have $\boldsymbol{\rho} \circ \mathbf{N}_{\alpha/p} = \mathbf{N}_\alpha$ and

$$\mathbb{E}[f(\mathbf{N}_\alpha)] \le \mathbb{E}[h_{\boldsymbol{\rho}}(\mathbf{N}_{\alpha/p})] + (p \cdot O(\log S)^{d-1})^k.$$

Averaging over ρ yields the fixed decision tree. ◁

Second, we can upper bound the extent to which the acceptance probability of a short decision tree increases when passing from the uniform distribution $\mathbf{U}$ to the biased distribution $\mathbf{N}_\gamma$.

▷ Claim 15. Suppose $f : \{0,1\}^n \to \{-1,1\}$ is a depth-k decision tree. Then

$$\mathbb{E}[f(\mathbf{N}_\gamma)] \le (1+\gamma)^k \cdot \mathbb{E}[f(\mathbf{U})] \le e^{\gamma k} \cdot \mathbb{E}[f(\mathbf{U})].$$

The proof is simple and can be found in the full version. We're now in a position to prove the lemma.

Proof of Lemma 13. Directly applying Claim 14, we get

$$\mathbb{E}[f(\mathbf{N}_\alpha)] \le \mathbb{E}[f'(\mathbf{N}_{\alpha/p})] + (p \cdot O(\log S)^{d-1})^k.$$

Applying Claim 15 to $\mathbb{E}[f'(\mathbf{N}_{\alpha/p})]$, we get

$$\begin{aligned}\mathbb{E}[f(\mathbf{N}_\alpha)] &\le (1+\alpha/p)^k \mathbb{E}[f'(\mathbf{U})] \\ &\le e^{\alpha k/p} \mathbb{E}[f'(\mathbf{U})].\end{aligned}$$

Putting these together finishes the proof. ◀

We can now prove Theorem 1.4, restated here:

▶ **Theorem 1.4.** *Fix $\varepsilon, \varepsilon', \delta > 0$, $d \in \mathbb{N}$, and*

$$S \le \exp\left(O\Big(\frac{\sqrt{\delta}}{\sqrt{\varepsilon}} \cdot \frac{\log(1/\delta)}{\log(1/\varepsilon')}\Big)^{1/(d-1)}\right).$$

Then $\mathbf{N}_{\sqrt{\varepsilon/\delta}}$ over $\{0,1\}^n$ with $n = O(1/\varepsilon)$ is (ε', δ)-pseudodense and yet $\mathbf{N}_{\sqrt{\varepsilon/\delta}}$ is not δ-dense inside of any ε-pseudorandom set.

Proof of Theorem 1.4. Let $n = 1/(7\varepsilon)$, $k = \log(1/\varepsilon')$ and $\alpha = \sqrt{\varepsilon/\delta}$. These choices satisfy $\alpha \ge \sqrt{\frac{1}{8\delta}}(1/n + \varepsilon)$, meaning $\mathbf{N}_\alpha$ is not δ-dense in any ε-pseudorandom set for $\mathcal{C}(S,d)$, by Lemma 6.

By Lemma 13, $\mathbf{N}_\alpha$ has (ε', δ)-pseudodensity for $\delta = e^{-\alpha k/p}$ and $\varepsilon' = (p \cdot O(\log S)^{d-1})^k$. The constraint on the density implies

$$\begin{aligned}\delta &= e^{-\alpha k/p} \\ \log(1/\delta) &= \alpha k/p \\ \log(1/\delta) &= \sqrt{\varepsilon/\delta}\log(1/\varepsilon')/p \\ p &= \frac{\sqrt{\varepsilon/\delta}\log(1/\varepsilon')}{\log(1/\delta)}.\end{aligned}$$

Plugging this value of p into the expression for ε', we get

$$\begin{aligned} \varepsilon' &= (p \cdot O(\log S)^{d-1})^k \\ (\varepsilon')^{1/k}/p &= O(\log S)^{d-1} \\ (\varepsilon')^{1/\log(1/\varepsilon')} \cdot \frac{\sqrt{\delta}\log(1/\delta)}{\sqrt{\varepsilon}\log(1/\varepsilon')} &= O(\log S)^{d-1}. \end{aligned}$$

Note that $(\varepsilon')^{1/\log(1/\varepsilon')} = 2^{-\log(1/\varepsilon')/\log(1/\varepsilon')} = 1/2$. Solving for S gives the claimed bound. ◀

5 Discussion of Theorem 1.5 and Theorem 1.6

This section briefly discusses Theorem 1.5 and Theorem 1.6, deferring a more detailed discussion to the full version of the paper. The basic idea underlying both proofs is to use tests which solve the coin problem to construct a test which "computes majority" in some problem-dependent sense.

Theorem 1.5 shows that the dense model theorem can fail for low-degree polynomials over finite fields.

▶ **Theorem 1.5.** *Fix a finite field $\mathbb{F}$ with characteristic $p = O(1)$, $\varepsilon, \varepsilon' > 0$ and let $c > \delta > 0$ where $c \approx 1/200$ is an absolute constant. Suppose that*

$$d \leq O(\sqrt{\delta/\varepsilon}).$$

Then when $\mathcal{F}$ is the n-variate degree-d polynomials over $\mathbb{F}$ with $n = 1/\varepsilon$, and $\alpha = O(\sqrt{\varepsilon/\delta})$, $\mathbf{N}_\alpha$ is (ε', δ)-pseudodense but is not δ-dense inside of an ε-pseudorandom set.

The main tool used in the proof is a special case of the robust Hegëdus lemma, discovered recently by Srinivasan [44].

▶ **Lemma 16** (Robust Hegëdus lemma (special case), [44])**.** *Let $\mathbb{F}$ be a finite field. Let $2^{-m/100} \leq \lambda \leq c$ where $c < 1$ is a (small) absolute constant. Let $\alpha^2 m$ be an integer so that $2^{-2\alpha^2 m} \geq \lambda$. Then if $P : \mathbb{F}^n \to \mathbb{F}$ is a degree d polynomial for which:*

1. $\Pr\left[P(\mathbf{Sp}_{m,\alpha m}) \neq 0\right] \leq \lambda$
2. $\Pr\left[P(\mathbf{Sp}_{m,0}) = 0\right] \leq 1 - e^{-\alpha^2 m/2}$.

Then $d = \Omega(\alpha m)$.

The idea is to use a low-degree polynomial distinguishing the biased coin distribution from uniform to construct another low-degree polynomial satisfying the conditions in the above lemma. Our particular approach uses random sampling and the approximation of OR by low-degree probabilistic polynomials [37]. We defer the details to the full version.

Theorem 1.6 gives a generic condition under which the dense model theorem is false, being witnessed by biased coins.

▶ **Theorem 1.6.** *Let $\varepsilon, \delta > 0$. Suppose $\mathcal{F}$ is a test class of boolean functions $f : \{0,1\}^n \to \{0,1\}$ with the following property: there is no AC^0 $\mathcal{F}$-oracle circuit of size $\mathsf{poly}(n \cdot \frac{\sqrt{\delta}}{\varepsilon^{3/2}})$ computing majority on $O(\sqrt{\delta/\varepsilon})$ bits.*

Then $\mathbf{N}_{\sqrt{\varepsilon/\delta}}$ is $(\epsilon\delta, \delta)$-pseudodense and yet does not have a δ-dense ε-model. In particular, when the hypotheses are met, the dense model theorem is false.

The approach is the same as Theorem 1.5, this time using a small circuit distinguishing biased from uniform to build an only-slightly-larger circuit computing majority. In this case, we use the following result of Shaltiel & Viola:

▶ **Theorem 5.1** ([42]). *Let $f : \{0,1\}^n \to \{0,1\}$ be a function that distinguishes between $\mathbf{U}$ and $\mathbf{N}_\alpha$ with constant distinguishing probability. Then there is an AC^0-circuit of size $\mathsf{poly}(n/\alpha)$ using f-oracle gates which computes majority on $O(1/\alpha)$ bits.*

Once again, we defer the details to the full version.

References

1 Scott Aaronson. Bqp and the polynomial hierarchy. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 141–150, 2010.

2 Rohit Agrawal. Coin theorems and the fourier expansion. *arXiv preprint*, 2019. `arXiv:1906.03743`.

3 Miklos Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constant depth circuits. In *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*, pages 11–19. IEEE, 1985.

4 Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.

5 Sanjeev Arora, Rong Ge, Yingyu Liang, Tengyu Ma, and Yi Zhang. Generalization and equilibrium in generative adversarial nets (gans). In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 224–232. JMLR. org, 2017.

6 R.B. Ash. *Information Theory*. Dover books on advanced mathematics. Dover Publications, 1990. URL: `https://books.google.com/books?id=nJ3UmGvdUCoC`.

7 Boaz Barak, Moritz Hardt, and Satyen Kale. The uniform hardcore lemma via approximate bregman projections. In *Proceedings of the twentieth annual ACM-SIAM symposium on Discrete algorithms*, pages 1193–1200. SIAM, 2009.

8 Boaz Barak, Ronen Shaltiel, and Avi Wigderson. Computational analogues of entropy. In *Approximation, Randomization, and Combinatorial Optimization.. Algorithms and Techniques*, pages 200–215. Springer, 2003.

9 Thomas F Bloom and Olof Sisask. Breaking the logarithmic barrier in roth's theorem on arithmetic progressions. *arXiv preprint*, 2020. `arXiv:2007.03528`.

10 Joshua Brody and Elad Verbin. The coin problem and pseudorandomness for branching programs. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 30–39. IEEE, 2010.

11 Mei-Chu Chang et al. A polynomial bound in freiman's theorem. *Duke mathematical journal*, 113(3):399–419, 2002.

12 Gil Cohen, Anat Ganor, and Ran Raz. Two sides of the coin problem. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2014.

13 Dean Doron, Dana Moshkovitz, Justin Oh, and David Zuckerman. Nearly optimal pseudorandomness from hardness. Technical report, ECCC preprint TR19-099, 2019.

14 Stefan Dziembowski and Krzysztof Pietrzak. Leakage-resilient cryptography. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 293–302. IEEE, 2008.

15 Yoav Freund and Robert Schapire. A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780):1612, 1999.

16 W. T. Gowers. A new proof of szemerédi's theorem. *Geometric & Functional Analysis GAFA*, 11(3):465–588, 2001.

17 W. T. Gowers. Decompositions, approximate structure, transference, and the hahn–banach theorem. *Bulletin of the London Mathematical Society*, 42(4):573–606, 2010.

18 Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, pages 481–547, 2008.

19 Aryeh Grinberg, Ronen Shaltiel, and Emanuele Viola. Indistinguishability by adaptive procedures with advice, and lower bounds on hardness amplification proofs. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 956–966. IEEE, 2018.

20 Iftach Haitner, Omer Reingold, and Salil Vadhan. Efficiency improvements in constructing pseudorandom generators from one-way functions. *SIAM Journal on Computing*, 42(3):1405–1430, 2013.

21 Iftach Haitner, Omer Reingold, Salil Vadhan, and Hoeteck Wee. Inaccessible entropy. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 611–620, 2009.

22 Lianna Hambardzumyan and Yaqiao Li. Chang's lemma via pinsker's inequality. *Discrete Mathematics*, 343(1):111496, 2020.

23 Johan Håstad. *Computational limitations of small-depth circuits*. MIT Press, 1987.

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

25 Johan Håstad, Russell Impagliazzo, Leonid A Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.

26 Thomas Holenstein. Key agreement from weak bit agreement. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 664–673, 2005.

27 Chun-Yuan Hsiao, Chi-Jen Lu, and Leonid Reyzin. Conditional computational entropy, or toward separating pseudoentropy from compressibility. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 169–186. Springer, 2007.

28 Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 538–545. IEEE, 1995.

29 Russell Impagliazzo. Connections between pseudo-randomness and machine learning: boosting, dense models, and regularity, 2020.

30 Russell Impagliazzo, William Matthews, and Ramamohan Paturi. A satisfiability algorithm for ac0. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 961–972. SIAM, 2012.

31 Russell Impagliazzo, Cristopher Moore, and Alexander Russell. An entropic proof of chang's inequality. *SIAM Journal on Discrete Mathematics*, 28(1):173–176, 2014.

32 Adam R Klivans and Rocco A Servedio. Boosting and hard-core set construction. *Machine Learning*, 51(3):217–238, 2003.

33 Nutan Limaye, Karteek Sreenivasaiah, Srikanth Srinivasan, Utkarsh Tripathi, and S Venkitesh. A fixed-depth size-hierarchy theorem for $\mathsf{AC}^0[\oplus]$ via the coin problem. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 442–453, 2019.

34 Chi-Jen Lu, Shi-Chun Tsai, and Hsin-Lung Wu. Complexity of hard-core set proofs. *computational complexity*, 20(1):145–171, 2011.

35 Or Meir and Avi Wigderson. Prediction from partial information and hindsight, with application to circuit lower bounds. *computational complexity*, 28(2):145–183, 2019.

36 Ilya Mironov, Omkant Pandey, Omer Reingod, and Salil Vadhan. Computational differential privacy. In *Annual International Cryptology Conference*, pages 126–142. Springer, 2009.

37 Alexander A Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical Notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987.

38 Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Dense subsets of pseudorandom sets. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 76–85. IEEE, 2008.

39 Benjamin Rossman. An entropy proof of the switching lemma and tight bounds on the decision-tree size of ac0, 2017.

40 Rocco A Servedio and Li-Yang Tan. Improved pseudorandom generators from pseudorandom multi-switching lemmas. *arXiv preprint*, 2018. arXiv:1801.03590.

41 Ronen Shaltiel. Is it possible to improve yao's xor lemma using reductions that exploit the efficiency of their oracle? In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.

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

43 Roman Smolensky. On representations by low-degree polynomials. In *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science*, pages 130–138. IEEE, 1993.

44 Srikanth Srinivasan. A robust version of hegedus's lemma, with applications. In Konstantin Makarychev, Yury Makarychev, Madhur Tulsiani, Gautam Kamath, and Julia Chuzhoy, editors, *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, pages 1349–1362. ACM, 2020. doi:10.1145/3357713.3384328.

45 Endre Szemerédi. On sets of integers containing no four elements in arithmetic progression. *Acta Mathematica Academiae Scientiarum Hungarica*, 20(1-2):89–104, 1969.

46 Avishay Tal. Tight bounds on the fourier spectrum of ac0. In *32nd Computational Complexity Conference (CCC 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

47 Michel Talagrand. How much are increasing sets positively correlated? *Combinatorica*, 16(2):243–258, 1996.

48 Terence Tao, Tamar Ziegler, et al. The primes contain arbitrarily long polynomial progressions. *Acta Mathematica*, 201(2):213–305, 2008.

49 Luca Trevisan, Madhur Tulsiani, and Salil Vadhan. Regularity, boosting, and efficiently simulating every high-entropy distribution. In *2009 24th Annual IEEE Conference on Computational Complexity*, pages 126–136. IEEE, 2009.

50 Thomas Watson. Advice lower bounds for the dense model theorem. *ACM Transactions on Computation Theory (TOCT)*, 7(1):1–18, 2015.

51 Andrew C Yao. Theory and application of trapdoor functions. In *23rd Annual Symposium on Foundations of Computer Science (SFCS 1982)*, pages 80–91. IEEE, 1982.

52 Jiapeng Zhang. On the query complexity for showing dense model. *Electron. Colloquium Comput. Complex.*, 2011.