Advice Lower Bounds for the Dense Model Theorem*

Thomas Watson[1]

1 **University of California, Berkeley**
tom@cs.berkeley.edu

Abstract

We prove a lower bound on the amount of nonuniform advice needed by black-box reductions for the Dense Model Theorem of Green, Tao, and Ziegler, and of Reingold, Trevisan, Tulsiani, and Vadhan. The latter theorem roughly says that for every distribution D that is δ-dense in a distribution that is ϵ'-indistinguishable from uniform, there exists a "dense model" for D, that is, a distribution that is δ-dense in the uniform distribution and is ϵ-indistinguishable from D. This ϵ-indistinguishability is with respect to an arbitrary small class of functions F. For the natural case where $\epsilon' \geq \Omega(\epsilon\delta)$ and $\epsilon \geq \delta^{O(1)}$, our lower bound implies that $\Omega\big(\sqrt{(1/\epsilon)\log(1/\delta)} \cdot \log|F|\big)$ advice bits are necessary. There is only a polynomial gap between our lower bound and the best upper bound for this case (due to Zhang), which is $O\big((1/\epsilon^2)\log(1/\delta) \cdot \log|F|\big)$. Our lower bound can be viewed as an analog of list size lower bounds for list-decoding of error-correcting codes, but for "dense model decoding" instead. Our proof introduces some new techniques which may be of independent interest, including an analysis of a majority of majorities of p-biased bits. The latter analysis uses an extremely tight lower bound on the tail of the binomial distribution, which we could not find in the literature.

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes

Keywords and phrases Pseudorandomness, advice lower bounds, dense model theorem

Digital Object Identifier 10.4230/LIPIcs.STACS.2013.634

1 Introduction

The question of whether the prime numbers contain arbitrarily long arithmetic progressions was a long-standing and famous open problem until Green and Tao [9] answered the question in the affirmative in a breakthrough paper in 2004. A key ingredient in their proof is a certain transference principle which, very roughly, states the following. Let U denote the set of positive integers. Then for every $D \subseteq U$, if there exists an $R \subseteq U$ such that D is dense in R and R is "indistinguishable" from U, then there exists an $M \subseteq U$ such that M is dense in U and D is "indistinguishable" from M. Tao and Ziegler [13] proved a much more general version of the transference principle, which has come to be known as the Dense Model Theorem (since M is a dense "model" for D).

Reingold, Trevisan, Tulsiani, and Vadhan [12] demonstrated the relevance of the Dense Model Theorem to computer science, and they gave a new proof which is much simpler and achieves better parameters than the proof of Green, Tao, and Ziegler. Gowers [6] independently came up with a similar proof. In addition to the original application of

* This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-0946797 and by the National Science Foundation under Grant No. CCF-1017403.

30th Symposium on Theoretical Aspects of Computer Science (STACS'13).
Editors: Natacha Portier and Thomas Wilke; pp. 634–645
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany
SYMPOSIUM ON THEORETICAL ASPECTS OF COMPUTER SCIENCE

$$\begin{tikzcd} D \arrow[r, "\delta\text{-dense}"] \arrow[d, leftrightarrow, "{(\epsilon, F)\text{-indistinguishable}}"'] & R \arrow[d, leftrightarrow, "{(\epsilon', F')\text{-indistinguishable}}"] \\ M \arrow[r, "\delta\text{-dense}"'] & U \end{tikzcd}$$

Figure 1 Relations among distributions in the Dense Model Theorem

showing that the primes contain arbitrarily long arithmetic progressions, the Dense Model Theorem has found applications in differential privacy [11], pseudoentropy and leakage-resilient cryptography [2, 12, 3], and graph decompositions [12], as well as further applications in additive combinatorics [7, 8]. Subsequent variants of the Dense Model Theorem have found applications in cryptography [5] and pseudorandomness [14].

To formally state the Dense Model Theorem, we first need some definitions. We identify $\{0,1\}^{2^n}$ with the set of functions from $\{0,1\}^n$ to $\{0,1\}$. We use $\mathcal{D}_n$ to denote the set of all distributions on $\{0,1\}^n$. The domain $\{0,1\}^n$ could be replaced by any finite set of size 2^n; we use the domain $\{0,1\}^n$ for concreteness.

▶ **Definition 1.** We say $D_1 \in \mathcal{D}_n$ is δ-*dense* in $D_2 \in \mathcal{D}_n$ if for all $x \in \{0,1\}^n$, $\Pr_{D_1}[x] \leq \frac{1}{\delta}\Pr_{D_2}[x]$.

▶ **Definition 2.** We say $f \in \{0,1\}^{2^n}$ ϵ-*distinguishes* $D_1, D_2 \in \mathcal{D}_n$ if $\left|\mathrm{E}_{D_1}[f] - \mathrm{E}_{D_2}[f]\right| > \epsilon$.

▶ **Definition 3.** For $F \subseteq \{0,1\}^{2^n}$, we say $D_1, D_2 \in \mathcal{D}_n$ are (ϵ, F)-*indistinguishable* if there is no $f \in F$ that ϵ-distinguishes D_1 and D_2.

The following is quantitatively the best known version of the theorem, due to Zhang [15] (building on [12, 1]).

▶ **Theorem 4** (Dense Model Theorem). *For every $F \subseteq \{0,1\}^{2^n}$ and every $D \in \mathcal{D}_n$, if there exists an $R \in \mathcal{D}_n$ such that D is δ-dense in R and (R,U) are (ϵ', F')-indistinguishable where $U \in \mathcal{D}_n$ is the uniform distribution, then there exists an $M \in \mathcal{D}_n$ such that M is δ-dense in U and (D,M) are (ϵ, F)-indistinguishable, where $\epsilon' \geq \Omega(\epsilon\delta)$ and F' consists of all linear threshold functions with ± 1 coefficients applied to $O\big((1/\epsilon^2)\log(1/\delta)\big)$ functions from F.*

The relations among the four distributions in Theorem 4 are illustrated in Figure 1. We remark that the theorem also holds when we allow $[0,1]$-valued functions f rather than just $\{0,1\}$-valued functions f. The proof of [12] gives the same result but where $O\big((1/\epsilon^2)\log(1/\epsilon\delta)\big)$ functions from F are combined to get a function from F'. The original proof of [13] achieves an F' which is qualitatively simpler, namely all products of $\mathrm{poly}(1/\epsilon, 1/\delta)$ functions from F, but it only achieves $\epsilon' \geq \exp(-\mathrm{poly}(1/\epsilon, 1/\delta))$.[1] We note that the dependence $\epsilon' \geq \Omega(\epsilon\delta)$ is tight in two senses.

- The Dense Model Theorem is actually false when $\epsilon' > \epsilon\delta$, even if $F' = \{0,1\}^{2^n}$. See [15] for the simple argument.
- The following converse to the Dense Model Theorem holds: If there exists an $M \in \mathcal{D}_n$ such that M is δ-dense in U and (D,M) are (ϵ, F)-indistinguishable, then there exists

[1] Another proof that also achieves this is given in [12].

an $R \in \mathcal{D}_n$ such that D is δ-dense in R and (R, U) are (ϵ', F')-indistinguishable, where $\epsilon' = \epsilon\delta$ and $F' = F$. To see this, note that $U = \delta M + (1-\delta)\widehat{M}$ for some $\widehat{M} \in \mathcal{D}_n$, so we can let $R = \delta D + (1-\delta)\widehat{M}$; then D is δ-dense in R, and for every $f \in \{0,1\}^{2^n}$ we have $\mathrm{E}_R[f] - \mathrm{E}_U[f] = \delta\big(\mathrm{E}_D[f] - \mathrm{E}_M[f]\big)$ and thus if $\big|\mathrm{E}_R[f] - \mathrm{E}_U[f]\big| > \epsilon'$ then $\big|\mathrm{E}_D[f] - \mathrm{E}_M[f]\big| > \epsilon$.

The Dense Model Theorem has an undesirable feature: The class F' is more complex than the class F. Thus, if we wish to conclude that D and M are indistinguishable for a class F, we need to assume that R and U are indistinguishable for a more complex class F'. The less complex F' is, the stronger the theorem is. The reason for this loss in complexity is because the theorem is proved using a *black-box reduction.* In other words, the contrapositive is proved: We assume that for every M δ-dense in U there exists a function from F that ϵ-distinguishes D and M, and we show that some of these functions can be plugged into the reduction to get a function that ϵ'-distinguishes R and U. Thus the resulting function is necessarily more complex than the functions that get plugged into the reduction. There are three notions of complexity that are interesting to address here.

1. *Computational complexity.* If F consists of functions computed by small constant-depth circuits (AC^0), then can we let F' consist of functions computed by (slightly larger) constant-depth circuits? This is not known to be true when $\epsilon' \geq \Omega(\epsilon\delta)$, because the reductions of [12, 15] involve a linear threshold function, which cannot be computed by small constant-depth circuits. Is it necessary that the reduction computes a linear threshold function? The original result of [13] shows that this is *not* necessary if ϵ' is extremely small.
2. *Query complexity.* If F consists of functions computed by circuits of size s, then F' will need to consist of functions computed by circuits of a larger size s' — but how much larger? If the reduction makes q queries to functions from F, then plugging in size-s circuits for these functions yields a circuit of size $\geq q \cdot s$, and thus we must have $s' \geq q \cdot s$. Hence it is desirable to minimize q. Can we do better than $q \leq O\big((1/\epsilon^2)\log(1/\delta)\big)$ as in Theorem 4?
3. *Advice complexity.* Suppose F consists of functions computed by uniform algorithms running in time t (that is, a single algorithm computes a sequence of functions, one for each $n = 1, 2, 3, \ldots,$). Then can we let F' consist of functions computed by uniform algorithms running in some (slightly larger) time t'? (Here, the distributions D, M, R, U would need to be sequences of distributions, and a distinguisher would only be required to succeed for infinitely many n.) The proofs of [12, 15] do not yield this, because the reductions need a nonuniform advice string to provide some extra information about the nth distribution D. How many bits of advice are needed?

Before proceeding we draw attention to the fact that, as we just alluded to, the advice strings used by the reductions of [12, 15] depend on D *but do not depend on* R. Hence something a little stronger than Theorem 4 actually holds: Although the statement of Theorem 4 says we need to assume that for some R in which D is δ-dense, there is no function in F' that ϵ'-distinguishes R and U, we actually only need to assume that there is no function in F' that simultaneously ϵ'-distinguishes U from every R in which D is δ-dense (the quantifiers are swapped). We are interested in proving lower bounds on the complexity of this type of black-box reduction for the Dense Model Theorem, where the advice does not depend on R.

The *query complexity* was considered by Zhang [15], who showed that for a certain type of nonadaptive black-box reduction, $\Omega\big((1/\epsilon^2)\log(1/\delta)\big)$ queries are necessary when

$\epsilon' \geq \Omega(\epsilon\delta)$ and $\epsilon \geq \delta^{O(1)}$, matching the upper bound of $O\big((1/\epsilon^2)\log(1/\delta)\big)$ for this case. In this paper we consider the *advice complexity*. We show that for arbitrary black-box reductions, $\Omega\big(\sqrt{(1/\epsilon)\log(1/\delta)}\cdot\log|F|\big)$ advice bits are necessary when $\epsilon' \geq \Omega(\epsilon\delta)$ and $\epsilon \geq \delta^{O(1)}$, which comes close to matching the upper bound of $O\big((1/\epsilon^2)\log(1/\delta)\cdot\log|F|\big)$ for this case [15]. Our result also holds for much more general settings of the parameters (with some degradation in the lower bound). Proving lower bounds on the *computational complexity* remains open.

Let us formally state what we mean by a black-box reduction. Recall the standard notation $[k] = \{1,\ldots,k\}$.

▶ **Definition 5.** An $(n,\epsilon,\delta,\epsilon',k,\alpha)$*-reduction* (for the Dense Model Theorem) is a function

$$\mathrm{Dec} : \big(\{0,1\}^{2^n}\big)^k \times \{0,1\}^\alpha \to \{0,1\}^{2^n}$$

such that for all $f_1,\ldots,f_k \in \{0,1\}^{2^n}$ and all $D \in \mathcal{D}_n$, if for every $M \in \mathcal{D}_n$ that is δ-dense in the uniform distribution $U \in \mathcal{D}_n$ there exists an $i \in [k]$ such that f_i ϵ-distinguishes D and M, then there exists an advice string $a \in \{0,1\}^\alpha$ such that for every $R \in \mathcal{D}_n$ in which D is δ-dense, $\mathrm{Dec}(f_1,\ldots,f_k,a)$ ϵ'-distinguishes R and U.

The proofs of [12, 15] work by exhibiting such reductions. The functions $\{f_1,\ldots,f_k\}$ correspond to the class F (which, if we were considering uniform algorithms, would be the restrictions of all the algorithms in the class to a particular input length n). We now state our theorem.

▶ **Theorem 6** (Main). *If there exists an $(n,\epsilon,\delta,\epsilon',k,\alpha)$-reduction for the Dense Model Theorem, and if $w > 1$ is an integer such that $2^{w+2}\cdot\delta^{w/160} \leq \epsilon'$, then*

$$\alpha \;\geq\; \Big\lfloor \tfrac{1}{160w}\sqrt{(1/\epsilon)\log_2(1/\delta)}\Big\rfloor \cdot \log_2 k - \log_2 w - 1$$

provided $2^n \geq \frac{w\log_2 k}{\epsilon\delta^2(\epsilon')^2}$, $\epsilon \leq 1/64\log_2(1/\delta)$, and $k \geq 1/16\epsilon^4$.

For the case where $\epsilon' \geq \Omega(\epsilon\delta)$ and $\epsilon \geq \delta^{O(1)}$ (which is reasonable), the condition $2^{w+2}\cdot\delta^{w/160} \leq \epsilon'$ is met provided w is a sufficiently large constant and δ is less than a sufficiently small constant,[2] and thus we get a lower bound $\alpha \geq \Omega\big(\sqrt{(1/\epsilon)\log(1/\delta)}\cdot\log k\big)$. Note that the three conditions at the end of the statement of Theorem 6 are very generous.

Our proof of Theorem 6 is somewhat reminiscent of the proof of a lower bound due to Lu, Tsai, and Wu [10] on the advice complexity of black-box reductions for the Hardcore Lemma, but our proof diverges significantly. We now give a quick preview of some of our ingredients. We use the probabilistic method to find a class of functions $f_1,\ldots,f_k$ for which many advice strings are needed to "cover" all the distributions D that do not have dense models. The key technical ingredients in the analysis (which differ from the ingredients in [10] and which may be of independent interest) include (1) a combinatorial argument identifying when several distributions D cannot share the same advice string, and (2) an analysis of a majority of majorities applied to overlapping sets of p-biased bits, where the sets form an almost-disjoint family (see Figure 2). The latter analysis makes use of extremely tight lower bounds on the tail probabilities of the binomial distribution, which we also prove (but could not find in the literature).

In the full version we point out an analogy between our lower bound and list size lower bounds for list-decoding of error-correcting codes, and we summarize analogous previous work on lower bounds for hardness amplification and list-decoding. The rest of this paper is devoted to proving Theorem 6. In Section 2 we give some intuition for the proof, and then in Section 3 we give the formal proof.

[2] The statement of Theorem 6 requires $\delta < 2^{-160}$, but this constant can be drastically improved.

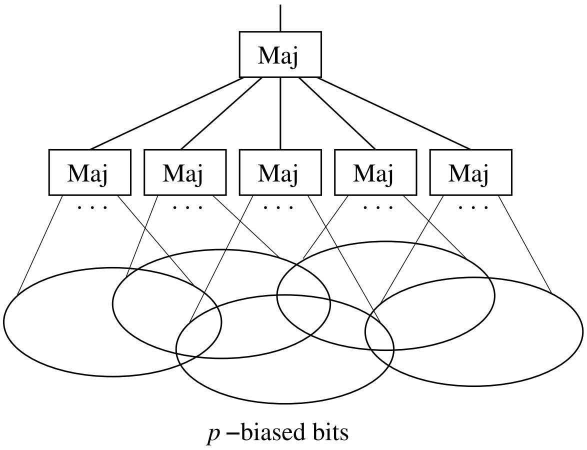

Figure 2 The majority of majorities

2 Intuition

According to Definition 5, for Dec to succeed as a reduction, it must be the case that for all $f_1, \ldots, f_k \in \{0,1\}^{2^n}$ and all $D \in \mathcal{D}_n$, if D has no "dense model" then there is some advice string a such that $\mathrm{Dec}(f_1, \ldots, f_k, a)$ "covers" D in a certain sense. To show that Dec needs many advice strings in order to succeed, we find functions $f_1, \ldots, f_k \in \{0,1\}^{2^n}$ and a large family of distributions in $\mathcal{D}_n$ such that

(i) each distribution in the family has no dense model (with respect to $f_1, \ldots, f_k$), and

(ii) each function $f \in \{0,1\}^{2^n}$ covers few of the distributions in the family.

So (i) implies that each distribution in the family needs to get covered, while (ii) implies that for each advice string a, $\mathrm{Dec}(f_1, \ldots, f_k, a)$ does not cover very many of them. Since the family is large, many advice strings are needed.

First we describe a technique for achieving (i), then we describe a technique for achieving (ii), and then we show how to consolidate the techniques to achieve both properties simultaneously. When we say D has no "dense model" we mean that for every $M \in \mathcal{D}_n$ that is δ-dense in U there exists an $i \in [k]$ such that f_i ϵ-distinguishes D and M. When we say a function "covers" D we mean that it ϵ'-distinguishes R and U for every $R \in \mathcal{D}_n$ in which D is δ-dense. The only distributions D we need to consider are uniform distributions over subsets of $\{0,1\}^n$.

Given $f_1, \ldots, f_k \in \{0,1\}^{2^n}$, what is an example of a distribution with no dense model? Suppose we pick any $I \subseteq [k]$ of size $1/4\epsilon$ and we let X_I be the set of all $x \in \{0,1\}^n$ such that $f_i(x) = 1$ for the majority of $i \in I$. Suppose we take D_I to be the uniform distribution over X_I. Then we have $\Pr_{x \sim D_I,\, i \sim I}[f_i(x) = 1] \geq 1/2 + 2\epsilon$ where $i \sim I$ means picking $i \in I$ uniformly at random. If X_I is roughly a $\delta/2$ fraction of $\{0,1\}^n$, then every distribution M that is δ-dense in U has at least half its mass outside of X_I, on strings x where $\Pr_{i \sim I}[f_i(x) = 1] \leq 1/2 - 2\epsilon$. It is possible to show that $\Pr_{x \sim M,\, i \sim I}[f_i(x) = 1] < \Pr_{x \sim D_I,\, i \sim I}[f_i(x) = 1] - \epsilon$ and thus there exists an $i \in I$ (depending on M) such that f_i ϵ-distinguishes D_I and M. So if $|X_I| \approx (\delta/2)2^n$ then D_I has no dense model. This is the technique we use for finding distributions without dense models.

Now, what is an example of a pair of distributions such that no function can cover both simultaneously? If we can show that every pair of distributions in the family is like this, then we will have achieved (ii). Because of an issue described below, we actually need to consider small collections of distributions rather than just pairs, but for now we consider pairs. Suppose D is uniform over some $X \subseteq \{0,1\}^n$ of size roughly $(\delta/2)2^n$, and similarly D' is uniform over some $X' \subseteq \{0,1\}^n$ of size roughly $(\delta/2)2^n$. If $X \cap X' = \emptyset$, then it can be shown that no function covers both D and D'.[3] Furthermore, if $|X \cap X'|$ is at most roughly $\epsilon' 2^n$ then this property still holds.

To consolidate the two techniques, we find a large family of sets $I \subseteq [k]$ each of size $1/4\epsilon$, where

(A) $|X_I| \approx (\delta/2)2^n$ for each I in the family, and
(B) the pairwise intersections of the X_I's (for I in the family) all have size at most roughly $\epsilon' 2^n$.

This would imply that the corresponding distributions D_I (for I in the family) have no dense models, and no function would cover more than one of them, so (i) and (ii) would be achieved.

We choose the functions $f_1, \ldots, f_k \in \{0,1\}^{2^n}$ randomly in some way, and we argue that for an appropriate family of sets I, properties (A) and (B) both hold with high probability. Property (A) suggests that we should choose p so that the probability a majority of $1/4\epsilon$ independent coins each with expectation p come up 1 is exactly $\delta/2$. Then we can set $f_i(x) = 1$ with probability p independently for each $i \in [k]$ and each $x \in \{0,1\}^n$, so for each I of size $1/4\epsilon$, $\Pr[x \in X_I] = \delta/2$. Then by concentration, $|X_I| \approx (\delta/2)2^n$ with high probability over $f_1, \ldots, f_k$.

If we choose $f_1, \ldots, f_k$ randomly in this way, how big will $|X_I \cap X_{I'}|$ be, for I and I' in the family? By concentration, we would have that with high probability over $f_1, \ldots, f_k$, $|X_I \cap X_{I'}|$ is roughly 2^n times $\Pr[x \in X_I \cap X_{I'}]$ (which is the same for all $x \in \{0,1\}^n$), so we would like the latter probability to be $\leq \epsilon'$. So what is the probability that the conjunction of two majorities of p-biased bits is 1? The best case is if $I \cap I' = \emptyset$, in which case the probability is exactly $(\delta/2)^2$. There are two problems with this.

(1) We cannot get a very large family of sets I if we require them to be pairwise disjoint.
(2) This requires $\epsilon' \geq (\delta/2)^2$. In a typical setting where $\epsilon' \geq \Omega(\epsilon\delta)$, this would require $\epsilon > \delta$, which is an odd and somewhat severe restriction.

To solve problem (1), we use the natural idea to allow the sets I to be pairwise almost-disjoint, rather than disjoint (which allows us to get a much larger family). So if $|I \cap I'|$ is at most some value b, how small does b have to be to ensure that the probability both majorities are 1 is not much more than $(\delta/2)^2$? We analyze this using the following trick: If both majorities are 1, then the fraction of coins that are 1 among $I \cup I'$ is at least q, where $q = 1/2 - 2\epsilon b = \frac{1/4\epsilon - b}{1/2\epsilon} \leq \frac{|I|/2 + |I'|/2 - b}{|I \cup I'|}$. Using an extremely tight characterization of the tail probabilities of the binomial distribution (which we prove using known techniques but which we could not find in the literature), we can show that $p \approx 1/2 - \sqrt{\epsilon \log(1/\delta)}$ and the probability of getting $\geq q$ fraction of 1's among the $|I \cup I'|$ coins is not much more than $(\delta/2)^2$ provided q is at least a constant factor closer to 1/2 than p is, say $q \approx 1/2 - \sqrt{\epsilon \log(1/\delta)}/4$. Thus it suffices to have $b \approx \sqrt{\epsilon \log(1/\delta)}/8\epsilon \geq \Omega\big(\sqrt{(1/\epsilon)\log(1/\delta)}\big)$. Since the family of sets I needs to be in the universe $[k]$, there exists such a family of roughly k^b many sets with pairwise intersections bounded in size by b. Since each function can cover D_I for only one I

[3] Actually, there is an issue having to do with the absolute value signs in the definition of distinguishing; this is dealt with in the formal proof.

in the family, roughly k^b advice strings are needed, which gives an advice lower bound of roughly $\log(k^b) \geq \Omega\big(\sqrt{(1/\epsilon)\log(1/\delta)} \cdot \log k\big)$.

Problem (2) is solved in the formal proof by considering small collections of sets from the family, rather than pairs. The parameter w in Theorem 6 is used to determine how big these collections should be. Then instead of considering the conjunction of two majorities, we need to consider the majority of several majorities, which explains where Figure 2 comes from.

3 Formal Proof

In Section 3.1, Section 3.2, and Section 3.3 we give preliminary lemmas, definitions, and notation. Then in Section 3.4 we give the proof of Theorem 6.

3.1 Binomial Distribution Tail

We let $\mathrm{Tail}(m, p, q)$ denote the probability that when m independent coins are flipped each with probability p of heads, at least a q fraction of the coins are heads (in other words, the probability the (m, p) binomial distribution is at least qm). For our proof of Theorem 6 we need extremely tight upper and lower bounds on the value of $\mathrm{Tail}(m, p, q)$. Such bounds can be given in terms of the fundamental quantity $\mathrm{RE}(q\|p) = q\log_2(\frac{q}{p}) + (1-q)\log_2(\frac{1-q}{1-p})$ which is known by a variety of names such as relative entropy, information divergence, and Kullback-Leibler distance.

We need the following fact, which can be seen using derivatives.

▶ Fact 7. For all $1/4 \leq p \leq q \leq 3/4$, we have $2(q-p)^2 \leq \mathrm{RE}(q\|p) \leq 4(q-p)^2$.

We also need the following standard and well-known form of the Chernoff-Hoeffding bound.

▶ **Lemma 8.** *For all $m \geq 1$ and all $0 \leq p \leq q \leq 1$, we have* $\mathrm{Tail}(m, p, q) \leq 2^{-\mathrm{RE}(q\|p)m}$.

The following lemma (see the full version for the proof) shows that Lemma 8 is very tight.

▶ **Lemma 9.** *For all $m \geq 1$ and all $1/4 \leq p \leq q \leq 1$, we have* $\mathrm{Tail}(m, p, q) \geq \frac{1}{48\sqrt{m}} \cdot 2^{-\mathrm{RE}(q\|p)m}$.

Although Lemma 9 is very simple and general, for our purpose we can only use it for a limited range of parameters, namely when $\epsilon \gg \delta$. This is because $\mathrm{RE}(q\|p)$ could be so close to 0 that $\frac{1}{48\sqrt{m}}$ completely swamps $2^{-\mathrm{RE}(q\|p)m}$, in which case Lemma 9 is not very tight. To handle the full range of ϵ and δ, we use the following stronger lower bound for the case $q = 1/2$. We prove this lemma in the full version.

▶ **Lemma 10.** *For all $m \geq 9$ and all $1/4 \leq p < 1/2$, we have*

$$\mathrm{Tail}(m, p, 1/2) \;\geq\; \min\left(\tfrac{1}{256}, \tfrac{1}{128\sqrt{m}(1/2-p)}\right) \cdot 2^{-\mathrm{RE}(1/2\|p)m}.$$

3.2 Combinatorial Designs

For our proof of Theorem 6 we need the existence of large families of almost-disjoint subsets of a finite set. Such combinatorial designs have numerous applications in theoretical computer science.

▶ **Definition 11.** An (ℓ, k, s, b)-*design* is a family of sets $I_1, \ldots, I_\ell \subseteq [k]$ all of size s such that $|I_j \cap I_{j'}| \leq b$ for every $j \neq j'$.

▶ **Lemma 12.** *For every k, s, b there exists an (ℓ, k, s, b)-design with $\ell \geq k^{b/8}$, provided $k \geq 16s^4$.*

There is nothing very novel about this lemma, and this precise version follows from a result in [4], but we provide a simple, self-contained proof in the full version. The proof uses the probabilistic method with a simple concentration bound for the hypergeometric distribution.

3.3 Notational Preliminaries

The parameters $n, \epsilon, \delta, \epsilon', k$, and w are fixed as in the statement of Theorem 6, and we always use D, M, R, U (possibly subscripted) to denote distributions in $\mathcal{D}_n$, in their roles as in Definition 5.

We let Maj denote the majority function on bit strings, and for even length strings we break ties by returning 1. We let And denote the and function on bit strings. We let Maj^t denote the function that takes t bit strings and returns their majorities as a length-t bit string. We use $\circ$ for function composition.

We also adhere to the following notational conventions. We use x for elements of $\{0,1\}^n$ and X for subsets of $\{0,1\}^n$. We use f for elements of $\{0,1\}^{2^n}$ (identified with functions from $\{0,1\}^n$ to $\{0,1\}$) and F for subsets of $\{0,1\}^{2^n}$. We use $[k]$ to index functions f, and we use i for elements of $[k]$ and I for subsets of $[k]$. We use $[\ell]$ to index subsets I (as in Definition 11), and we use j for elements of $[\ell]$ and J for subsets of $[\ell]$. We generally use s for the size of I, and t for the size of J.

The following notation is with respect to fixed $f_1, \ldots, f_k \in \{0,1\}^{2^n}$. Given $I \subseteq [k]$ we define

- f_I is the function that takes $x \in \{0,1\}^n$ and returns the length-$|I|$ bit string $(f_i(x))_{i \in I}$;
- X_I is the set of $x \in \{0,1\}^n$ on which $\text{Maj} \circ f_I$ returns 1;
- D_I is the uniform distribution over X_I (and if $X_I = \emptyset$ then D_I is undefined).

The following notation is with respect to fixed $f_1, \ldots, f_k \in \{0,1\}^{2^n}$ and fixed $I_1, \ldots, I_\ell \subseteq [k]$. Given $J \subseteq [\ell]$ we define

- f_{I_J} is the function that takes $x \in \{0,1\}^n$ and returns the $|J|$-tuple $(f_{I_j}(x))_{j \in J}$;
- X_{I_J} is the set of $x \in \{0,1\}^n$ on which $\text{Maj} \circ \text{Maj}^{|J|} \circ f_{I_J}$ returns 1.

We use $\sim$ to denote sampling from a distribution (for example $x \sim D$), and we use the convention that sampling from a set (for example $i \sim I$) means sampling from the uniform distribution over that set.

3.4 Proof of Theorem 6

Consider an arbitrary function $\text{Dec} : \left(\{0,1\}^{2^n}\right)^k \times \{0,1\}^\alpha \to \{0,1\}^{2^n}$. Supposing that $\alpha < \left\lfloor \frac{1}{160w} \sqrt{(1/\epsilon) \log_2(1/\delta)} \right\rfloor \cdot \log_2 k - \log_2 w - 1$, we show that Dec is not an $(n, \epsilon, \delta, \epsilon', k, \alpha)$-reduction. We first introduce some terminology to make things concise. Given $f_1, \ldots, f_k \in \{0,1\}^{2^n}$, a *dense model* for $D \in \mathcal{D}_n$ is an $M \in \mathcal{D}_n$ that is δ-dense in the uniform distribution $U \in \mathcal{D}_n$ and is such that for all $i \in [k]$, f_i does not ϵ-distinguish D and M. We say a function $f \in \{0,1\}^{2^n}$ *covers* $D \in \mathcal{D}_n$ if for every $R \in \mathcal{D}_n$ in which D is δ-dense, f ϵ'-distinguishes R and U.

Thus to show that Dec is not an $(n, \epsilon, \delta, \epsilon', k, \alpha)$-reduction, we need to find $f_1, \ldots, f_k \in \{0,1\}^{2^n}$ such that some D has no dense model but is not covered by $\text{Dec}(f_1, \ldots, f_k, a)$ for any advice string $a \in \{0,1\}^\alpha$.

3.4.1 Distributions Without Dense Models

The following claim is our tool for finding distributions that have no dense models. We prove this claim in the full version.

▶ Claim 13. For every $f_1, \ldots, f_k \in \{0,1\}^{2^n}$ and every $I \subseteq [k]$ of size $0 < s \leq 1/4\epsilon$ (for some s), if $0 < |X_I| \leq (2\delta/3)2^n$ then D_I has no dense model.

3.4.2 Distributions That Cannot Be Covered

We say a function $f \in \{0,1\}^{2^n}$ *positively covers* $D \in \mathcal{D}_n$ if for every $R \in \mathcal{D}_n$ in which D is δ-dense, $\mathrm{E}_R[f] - \mathrm{E}_U[f] > \epsilon'$ (note the absence of absolute value signs). Observe that if $f \in \{0,1\}^{2^n}$ covers D then either f or its complement positively covers D. This is because if there existed $R_1, R_2 \in \mathcal{D}_n$ in which D is δ-dense and such that $\mathrm{E}_{R_1}[f] < \mathrm{E}_U[f] < \mathrm{E}_{R_2}[f]$, then some convex combination R_3 of R_1 and R_2 would have $\mathrm{E}_{R_3}[f] = \mathrm{E}_U[f]$. However, D would be δ-dense in R_3 since the set of R in which D is δ-dense is convex, so f would not cover D.

▶ Claim 14. For every $f_1, \ldots, f_k \in \{0,1\}^{2^n}$, every $I_1, \ldots, I_\ell \subseteq [k]$ (for some ℓ), and every $J \subseteq [\ell]$ of size $t > 1$ (for some t), if $|X_{I_J}| \leq (\epsilon'/2)2^n$ and $|X_{I_j}| \geq (\delta/2 - \epsilon'/4)2^n$ for all $j \in J$ then there is no function that simultaneously positively covers D_{I_j} for all $j \in J$.

Proof. Assume that $|X_{I_J}| \leq (\epsilon'/2)2^n$ and $|X_{I_j}| \geq (\delta/2 - \epsilon'/4)2^n$ for all $j \in J$. Consider an arbitrary $f \in \{0,1\}^{2^n}$ and let X be the set of $x \in \{0,1\}^n$ such that $f(x) = 1$. For $\tau \in \{0, 1, \ldots, t\}$ let $X^{(\tau)}$ be the set of $x \in \{0,1\}^n$ such that there are exactly τ values of $j \in J$ for which $x \in X_{I_j}$ (in other words, $(\mathrm{Maj}^t \circ f_{I_J})(x)$ has Hamming weight τ). Note that $X_{I_J} = \bigcup_{\tau=t'}^{t} X^{(\tau)}$ where $t' = \lceil t/2 \rceil$. Let $\pi = \min_{j\in J}\left[\mathrm{E}_{D_{I_j}}[f]\right]$. Then for every $j \in J$ we have $|X \cap X_{I_j}| \geq \pi \cdot |X_{I_j}| \geq \pi \cdot (\delta/2 - \epsilon'/4)2^n$. We have

$$\begin{aligned}(t/2)\cdot\big(|X| + |X_{I_J}|\big) &\geq (t/2)\cdot|X \cap \overline{X_{I_J}}| + t\cdot|X \cap X_{I_J}| \\ &\geq \textstyle\sum_{\tau=0}^{t} \tau\cdot|X \cap X^{(\tau)}| \\ &= \textstyle\sum_{j\in J} |X \cap X_{I_j}| \\ &\geq t\cdot\pi\cdot(\delta/2 - \epsilon'/4)2^n\end{aligned}$$

which implies that

$$|X| \;\geq\; \pi\cdot(\delta - \epsilon'/2)2^n - |X_{I_J}| \;\geq\; \pi\delta 2^n - \epsilon' 2^n \;=\; (\pi - \epsilon'/\delta)\cdot\delta 2^n$$

since $\pi \leq 1$ and $|X_{I_J}| \leq (\epsilon'/2)2^n$. We might have $\pi - \epsilon'/\delta < 0$, but this is not problematic. Let M be a distribution δ-dense in U that maximizes $\mathrm{E}_M[f]$, and observe that

$$\mathrm{E}_M[f] \;=\; \min\big(|X|/\delta 2^n, 1\big) \;\geq\; \pi - \epsilon'/\delta.$$

We have $U = \delta M + (1-\delta)\widehat{M}$ for some $\widehat{M} \in \mathcal{D}_n$. Let $j \in J$ be such that $\mathrm{E}_{D_{I_j}}[f] = \pi$, and define the distribution $R = \delta D_{I_j} + (1-\delta)\widehat{M}$ so that D_{I_j} is δ-dense in R. Then we have

$$\mathrm{E}_R[f] \;=\; \delta\pi + (1-\delta)\,\mathrm{E}_{\widehat{M}}[f]$$

and

$$\mathrm{E}_U[f] \;=\; \delta\,\mathrm{E}_M[f] + (1-\delta)\,\mathrm{E}_{\widehat{M}}[f] \;\geq\; \delta\pi - \epsilon' + (1-\delta)\,\mathrm{E}_{\widehat{M}}[f] \;=\; \mathrm{E}_R[f] - \epsilon'$$

so f does not positively cover D_{I_j}. This finishes the proof of Claim 14. ◀

3.4.3 Setting the Parameters

Define $s = \lfloor 1/4\epsilon \rfloor$ and $t = w$ and $b = \big\lfloor \frac{1}{20t}\sqrt{(1/\epsilon)\log_2(1/\delta)}\big\rfloor$. By Lemma 12 there exists an (ℓ, k, s, b)-design $I_1, \ldots, I_\ell$ with $\ell = \lceil k^{b/8} \rceil$ (note that we do have $k \geq 16s^4$). Define p to be such that $\mathrm{Tail}(s, p, 1/2) = \delta/2$. We prove the following claim in the full version, using Lemma 10.

▶ Claim 15. $\frac{1}{2}\sqrt{\epsilon \log_2(1/\delta)} \;\leq\; 1/2 - p \;\leq\; 2\sqrt{\epsilon \log_2(1/\delta)} \;\leq\; 1/4$.

3.4.4 The Majority of Majorities

We choose $f_1, \ldots, f_k$ randomly by setting $f_i(x) = 1$ with probability p independently for each $i \in [k]$ and each $x \in \{0,1\}^n$.

▶ Claim 16. For every $J \subseteq [\ell]$ of size t and every $x \in \{0,1\}^n$, we have $\Pr_{f_1,\ldots,f_k}[x \in X_{I_J}] \leq \epsilon'/4$.

Proof. Define $t' = \lceil t/2 \rceil$. Note that if $(\mathrm{Maj} \circ \mathrm{Maj}^t \circ f_{I_J})(x) = 1$ then there exists a subset $J' \subseteq J$ of size t' such that $(\mathrm{And} \circ \mathrm{Maj}^{t'} \circ f_{I_{J'}})(x) = 1$. Thus we have

$$\begin{aligned} &\Pr_{f_1,\ldots,f_k}\big[(\mathrm{Maj} \circ \mathrm{Maj}^t \circ f_{I_J})(x) = 1\big] \\ \leq\ & 2^t \cdot \max_{J' \subseteq J\,:\,|J'|=t'} \Pr_{f_1,\ldots,f_k}\big[(\mathrm{And} \circ \mathrm{Maj}^{t'} \circ f_{I_{J'}})(x) = 1\big]. \end{aligned}$$

Consider an arbitrary $J' \subseteq J$ of size t'. Define $m = \big|\bigcup_{j \in J'} I_j\big|$ and notice that since $I_1, \ldots, I_\ell$ is an (ℓ, k, s, b)-design, by inclusion-exclusion we have

$$t's - \tbinom{t'}{2}b \;\leq\; m \;\leq\; t's. \tag{1}$$

Define $s' = \lceil s/2 \rceil$ and $q = 1/2 - t'b/2s$. If $(\mathrm{And} \circ \mathrm{Maj}^{t'} \circ f_{I_{J'}})(x) = 1$ then for each $j \in J'$ we have $\sum_{i \in I_j} f_i(x) \geq s'$ and so by inclusion-exclusion we have

$$\textstyle\sum_{i \in \bigcup_{j \in J'} I_j} f_i(x) \;\geq\; \Big(\sum_{j \in J'} \sum_{i \in I_j} f_i(x)\Big) - \binom{t'}{2}b \;\geq\; t's' - \binom{t'}{2}b \;\geq\; qt's \;\geq\; qm.$$

It follows that

$$\begin{aligned} \Pr_{f_1,\ldots,f_k}\big[(\mathrm{And} \circ \mathrm{Maj}^{t'} \circ f_{I_{J'}})(x) = 1\big] \;&\leq\; \Pr_{f_1,\ldots,f_k}\Big[\textstyle\sum_{i \in \bigcup_{j \in J'} I_j} f_i(x) \geq qm\Big] \\ &=\; \mathrm{Tail}(m, p, q) \\ &\leq\; 2^{-\mathrm{RE}(q\|p)m} \\ &=\; \big(2^{-\mathrm{RE}(1/2\|p)s}\big)^{(m/s)\cdot(\mathrm{RE}(q\|p)/\,\mathrm{RE}(1/2\|p))} \\ &\leq\; \big(\delta^{1/10}\big)^{(m/s)\cdot(\mathrm{RE}(q\|p)/\,\mathrm{RE}(1/2\|p))} \end{aligned}$$

where the third line follows by Lemma 8 and the fifth line follows by nonnegativity of RE and

$$2^{-\mathrm{RE}(1/2\|p)s} \;\leq\; 2^{-2(1/2-p)^2 s} \;\leq\; \delta^{\epsilon s/2} \;\leq\; \delta^{1/10}$$

which holds by Fact 7, Claim 15, and $\epsilon \leq 1/20$. We have

$$m/s \;\geq\; t' - (t')^2 b/2s \;\geq\; t'/2 \;\geq\; t/4 \tag{2}$$

by (1) and $b \leq s/t'$ (which can be shown using the final inequality in Claim 15). We also have $t'b/2s \leq \frac{1}{8}\sqrt{\epsilon \log_2(1/\delta)}$ and thus $q - p \geq \frac{3}{4}(1/2 - p)$ by Claim 15. Hence by Fact 7 we have

$$\mathrm{RE}(q\|p)/\,\mathrm{RE}(1/2\|p) \;\geq\; \tfrac{(q-p)^2}{2(1/2-p)^2} \;\geq\; \tfrac{(\frac{3}{4}(1/2-p))^2}{2(1/2-p)^2} \;\geq\; 1/4. \tag{3}$$

Using (2) and (3) we get

$$\Pr_{f_1,\ldots,f_k}\left[(\mathrm{And} \circ \mathrm{Maj}^{t'} \circ f_{I_{J'}})(x) = 1\right] \;\leq\; \left(\delta^{1/10}\right)^{(t/4)\cdot(1/4)} \;=\; \delta^{t/160}.$$

We conclude that $\Pr_{f_1,\ldots,f_k}[x \in X_{I_J}] \;\leq\; 2^t \cdot \delta^{t/160} \;\leq\; \epsilon'/4$. This finishes the proof of Claim 16. ◀

3.4.5 Putting It All Together

For every $j \in [\ell]$ and every $x \in \{0,1\}^n$, we have $\Pr_{f_1,\ldots,f_k}[x \in X_{I_j}] = \mathrm{Tail}(s, p, 1/2) = \delta/2$. Standard relative-error forms of the Chernoff bound give

$$\begin{aligned}
&\Pr_{f_1,\ldots,f_k}\left[|X_{I_j}| < (\delta/2 - \epsilon'/4)2^n\right] &&\leq\; e^{-2^n(\epsilon')^2/16\delta}\\
&\Pr_{f_1,\ldots,f_k}\left[|X_{I_j}| > (2\delta/3)2^n\right] &&\leq\; e^{-2^n\delta/54}\\
&\Pr_{f_1,\ldots,f_k}\left[|X_{I_J}| > (\epsilon'/2)2^n\right] &&\leq\; e^{-2^n\epsilon'/12}
\end{aligned}$$

where the latter holds for each $J \subseteq [\ell]$ of size t, using Claim 16. Thus by a union bound we have

$$\begin{aligned}
&\Pr_{f_1,\ldots,f_k}\begin{bmatrix}(\delta/2 - \epsilon'/4)2^n \leq |X_{I_j}| \leq (2\delta/3)2^n \text{ for all } j \in [\ell] \text{ and}\\ |X_{I_J}| \leq (\epsilon'/2)2^n \text{ for all } J \subseteq [\ell] \text{ of size } t\end{bmatrix}\\
&\geq 1 - \ell \cdot e^{-2^n(\epsilon')^2/16\delta} - \ell \cdot e^{-2^n\delta/54} - \binom{\ell}{t} \cdot e^{-2^n\epsilon'/12}\\
&> 0
\end{aligned}$$

since $2^n \geq \frac{t \log_2 k}{\epsilon\delta^2(\epsilon')^2}$. Fix a choice of $f_1, \ldots, f_k$ such that the above event occurs.

For every $J^* \subseteq [\ell]$ of size $2t - 1$, there is no $a \in \{0,1\}^\alpha$ such that $\mathrm{Dec}(f_1, \ldots, f_k, a)$ simultaneously covers D_{I_j} for all $j \in J^*$, because otherwise for some $J \subseteq J^*$ of size t, either $\mathrm{Dec}(f_1, \ldots, f_k, a)$ or its complement would simultaneously positively cover D_{I_j} for all $j \in J$, which would contradict Claim 14.

Therefore for each $a \in \{0,1\}^\alpha$, the number of $j \in [\ell]$ such that D_{I_j} is covered by $\mathrm{Dec}(f_1, \ldots, f_k, a)$ is at most $2t - 2$. This implies that the number of $j \in [\ell]$ for which there exists an $a \in \{0,1\}^\alpha$ such that $\mathrm{Dec}(f_1, \ldots, f_k, a)$ covers D_{I_j} is at most $2^\alpha \cdot (2t-2) < k^{b/8} \leq \ell$ since $\alpha \leq (b/8)\log_2 k - \log_2 t - 1$. Thus there exists a $j \in [\ell]$ such that D_{I_j} is not covered by $\mathrm{Dec}(f_1, \ldots, f_k, a)$ for any $a \in \{0,1\}^\alpha$. By Claim 13, D_{I_j} has no dense model, so Dec is not an $(n, \epsilon, \delta, \epsilon', k, \alpha)$-reduction. This finishes the proof of Theorem 6.

Acknowledgments I thank Anand Bhaskar, Siu Man Chan, and Siu On Chan for helpful discussions, and anonymous reviewers for useful comments.

References

1 Boaz Barak, Moritz Hardt, and Satyen Kale. The uniform hardcore lemma via approximate Bregman projections. In *Proceedings of the 20th ACM-SIAM Symposium on Discrete Algorithms*, pages 1193–1200, 2009.

2 Boaz Barak, Ronen Shaltiel, and Avi Wigderson. Computational analogues of entropy. In *Proceedings of the 7th International Workshop on Randomization and Computation*, pages 200–215, 2003.

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

4 Paul Erdős, Péter Frankl, and Zoltán Füredi. Families of finite sets in which no set is covered by the union of r others. *Israel Journal of Mathematics*, 51(1-2):79–89, 1985.

5 Craig Gentry and Daniel Wichs. Separating succinct non-interactive arguments from all falsifiable assumptions. In *Proceedings of the 43rd ACM Symposium on Theory of Computing*, pages 99–108, 2011.

6 Timothy Gowers. Decompositions, approximate structure, transference, and the Hahn–Banach Theorem. *Bulletin of the London Mathematical Society*, 42(4):573–606, 2010.

7 Timothy Gowers and Julia Wolf. Linear forms and higher-degree uniformity for functions on $\mathbb{F}_p^n$. *Geometric and Functional Analysis*, 21(1):36–69, 2011.

8 Timothy Gowers and Julia Wolf. Linear forms and quadratic uniformity for functions on $\mathbb{F}_p^n$. *Mathematika*, 57(2):215–237, 2012.

9 Ben Green and Terence Tao. The primes contain arbitrarily long arithmetic progressions. *Annals of Mathematics*, 167(2):481–547, 2008.

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

11 Ilya Mironov, Omkant Pandey, Omer Reingold, and Salil Vadhan. Computational differential privacy. In *Proceedings of the 29th International Cryptology Conference*, pages 126–142, 2009.

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

13 Terence Tao and Tamar Ziegler. The primes contain arbitrarily long polynomial progressions. *Acta Mathematica*, 201:213–305, 2008.

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

15 Jiapeng Zhang. On the query complexity for showing dense model. Technical Report TR11-038, Electronic Colloquium on Computational Complexity, 2011.