Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 58 (2013)

Improved Average-Case Lower Bounds for De Morgan Formula Size*

Matching Worst-Case Lower Bound

Ilan Komargodski† Ran Raz† Avishay Tal†

Abstract

We give an explicit function $h\colon \{0,1\}^n \to \{0,1\}$ such that every De Morgan formula of size $n^{3-o(1)}/r^2$ agrees with h on at most a fraction of $\frac{1}{2} + 2^{-\Omega(r)}$ of the inputs.

Our technical contributions include a theorem that shows that the "expected shrinkage" result of Håstad (SIAM J. Comput., 1998) actually holds with very high probability (where the restrictions are chosen from a certain distribution that takes into account the structure of the formula), using ideas of Impagliazzo, Meka and Zuckerman (FOCS, 2012).

1 Introduction

Proving lower bounds on the complexity of classical computational models for Boolean functions is the holy grail in theoretical computer science. One of the simplest and most natural non-uniform computational models that is of great interest is the model of Boolean De Morgan formulas. It is well known that the De Morgan formula size of almost all functions on n variables is at least $\Omega\left(2^n/\log n\right)$. Nevertheless, no explicit function (constructible deterministically in polynomial time) with super-polynomial lower bounds on the De Morgan formula size has been found yet. Providing such a function would separate P from NC^1.

A De Morgan formula is represented by a binary tree such that every leaf is labeled by an input variable or its negation, and every internal node is labeled by an operation from $\{\vee, \wedge\}$. A formula is said to compute a function $f\colon \{0,1\}^n \to \{0,1\}$ if on all inputs $x \in \{0,1\}^n$ it outputs $f(x)$. The computation is done in the natural way from the leaves to the root. The size of a formula F, denoted by $L(F)$, is defined as the number of leaves it contains. The De Morgan formula size of a function $f\colon \{0,1\}^n \to \{0,1\}$ is the size of the minimal De Morgan formula that computes f.

Previous works considered two types of lower bounds on De Morgan formula size: *worst-case* lower bounds and *average-case* lower bounds.

Worst-case lower bounds are lower bounds on the size of the minimal De Morgan formula that computes an explicit function $f\colon \{0,1\}^n \to \{0,1\}$. The first non-trivial lower bound in this model was achieved by Subbotovskaya [Sub61] that proved an $\Omega(n^{1.5})$ lower bound on the size of any De Morgan formula that computes the parity function on n variables. Subbotovskaya also introduced the concept of random restrictions that has had many applications since. In fact,

*This is a final draft of [KRT17]. A preliminary version of this work appeared in the 45th ACM Symposium on Theory of Computing (STOC 2013), pp. 171–180 [KR13] and in the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2013), pp. 588–597 [KRT13].

†Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel. Email: {ilan.komargodski,ran.raz,avishay.tal}@weizmann.ac.il. Research supported by an Israel Science Foundation grant and by the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation.

ISSN 1433-8092

Subbotovskaya showed a lower bound of $\Omega(n^{\Gamma})$ where Γ is referred to as the shrinkage exponent of De Morgan formulas under random restrictions and showed that $\Gamma \geq 1.5$. In [Khr71] Khrapchenko was able to improve the lower bound of [Sub61] and to prove, using a completely different method, that the parity function on n variables requires a De Morgan formula of size at least n^2 (which is tight, up to constant factors). In [And87], Andreev was able to cleverly combine some previous techniques (including the method of [Sub61]) in order to prove an $\Omega(n^{1+\Gamma-o(1)})$ lower bound on the size of the minimal De Morgan formula that computes an explicit function, later referred to as Andreev's function. Subsequent improvements on the constant Γ led to improved lower bounds on the De Morgan formula size of Andreev's function. Impagliazzo and Nisan [IN93] proved that $\Gamma \geq 1.55$, Paterson and Zwick [PZ93] proved that $\Gamma \geq 1.63$, Håstad [Hås98] proved that $\Gamma \geq 2-o(1)$ and Tal [Tal14] removed logarithmic factors from Håstad's work to achieve the tight result $\Gamma = 2$. This, in turn, yields a lower bound of $\Omega(n^{3-o(1)})$ on the size of De Morgan formulas that compute Andreev's function.

Average-case lower bounds (a.k.a., correlation bounds) are lower bounds on the size of the minimal De Morgan formula that only approximates an explicit function $f\colon \{0,1\}^n \to \{0,1\}$. An approximation of f is a computation that agrees with f on some fraction larger than $1/2$ of the inputs (rather than on all inputs). The first explicit average-case lower bound for De Morgan formulas appears in the work of Santhanam [San10]. There, it is shown that any family of linear-size De Morgan formulas has correlation of at most $\frac{1}{2} + 2^{-\Omega(n)}$ with the parity function, and moreover, his technique could be extended to show a correlation of at most $\frac{1}{2} + 2^{-n^{\Omega(1)}}$ between any De Morgan formula of size $O(n^{1.5})$ and the parity function. In addition works regarding the degree of approximating polynomials also imply correlation bounds for De Morgan formulas. Specifically, from the works of [BBC+01, Rei11] it follows that any formula of size $o\left((n/\log(1/\varepsilon))^2\right)$ has correlation of at most $\frac{1}{2} + \varepsilon$ with the parity function on n variables.

In this work, we construct an explicit function $h\colon \{0,1\}^n \to \{0,1\}$ such that any De Morgan formula of size at most $O(n^{2.99})$ computes h correctly on a fraction of at most $\frac{1}{2} + 2^{-n^{\Omega(1)}}$ of the inputs. More generally, our main theorem gives the following trade-off between the size of the formula and the quality of approximation:

Theorem 1.1. *There exists a constant $c \geq 8$ such that for any large enough integer n and any $c \cdot \log(n) \leq r \leq n^{1/3}$ the following holds. There is an explicit (computable in polynomial time) Boolean function $h\colon \{0,1\}^{6n} \to \{0,1\}$ such that any formula of size $\frac{n^3}{r^2 \cdot \mathsf{polylog}(n)}$ computes h correctly on a fraction of at most $1/2 + 2^{-\Omega(r)}$ of the inputs.*

1.1 Techniques

We start by informally defining restrictions and shrinkage (more formal definitions can be found in Section 2). Given a function $f\colon \{0,1\}^n \to \{0,1\}$, a vector $\rho \in \{0,1,*\}^n$ defines a restriction of f, denoted by $f|_\rho$, in the following way: if $\rho_i \in \{0,1\}$ then the i-th input variable of f is fixed (or assigned) to 0 or 1, respectively, and otherwise it is still a variable. Let $\mathcal{D}$ be a distribution of restrictions that leave k variables unassigned; we say that De Morgan formulas have s-shrinkage with probability γ over $\mathcal{D}$ if any De Morgan formula shrinks by a factor of at least $c \cdot (k/n)^s$ with probability γ over $\mathcal{D}$, for some universal constant c.

Next, we present a framework for proving average-case lower bounds. In order to explain it, we first begin by describing the worst-case lower bound of Andreev [And87].

1.1.1 Andreev's Worst-Case Lower Bound

Andreev's function $\mathcal{A}\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is defined as follows. $\mathcal{A}$ views the second input as a $\log n$ by $n/\log n$ matrix and computes the XOR of the input bits in every row. $\mathcal{A}$ uses the resulting $\log n$ bits to address an index in the first input ($\log n$ bits are enough to represent a cell in a vector of length n) and returns that bit. The analysis of[And87, IN93, PZ93, Hås98, Tal14] relies on the following 4 facts:

1. There exists an n bit vector h that represents a Boolean function on $\log n$ bits which requires formulas of size $\Omega(n/\log\log n)$.
2. It holds that $L(\mathcal{A}) \geq L(\mathcal{A}_h)$, where $\mathcal{A}_h$ is the function $\mathcal{A}$ such that the first input is fixed to the hard function h from Item 1.
3. Γ-shrinkage occurs with probability at least $3/4$ (over completely random restrictions). That is, for a function $f\colon \{0,1\}^n \to \{0,1\}$ and for a random restriction ρ that leaves k variables unassigned it holds that $L(f|_\rho) \leq c\cdot\left(\frac{k}{n}\right)^\Gamma \cdot L(f)$ with probability at least $3/4$ for some universal constant $c > 0$.

   This fact, that was first proved by [Sub61] (for $\Gamma = 1.5$), was gradually improved throughout [IN93, PZ93, Hås98, Tal14].
4. After applying a completely random restriction that leaves $k = \Theta\left(\log n \cdot \log\log n\right)$ variables unrestricted, with probability at least $3/4$ every row in the matrix (represented by the second input to $\mathcal{A}$) has at least one variable that is not restricted.

Andreev derived the lower bound as follows. Since Item 3 and Item 4 occur with probability at least $3/4$, there exists a restriction ρ such that both items hold simultaneously. Hence,

$$L(\mathcal{A}) \overset{\text{Item 2}}{\geq} L(\mathcal{A}_h) \overset{\text{Item 3}}{\geq} \frac{1}{c}\cdot\left(\frac{n}{k}\right)^\Gamma \cdot L(\mathcal{A}_h|_\rho) \overset{\text{Items 1 and 4}}{\geq} n^{\Gamma+1-o(1)}\,,$$

where $\mathcal{A}_h|_\rho$ denotes the function $\mathcal{A}_h$ after applying the restriction ρ.

1.1.2 A Scheme for Average-Case Lower Bounds

Define a function $\mathcal{B}\colon \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ that views its second input as an n^ε by $n^{1-\varepsilon}$ matrix and computes the XOR of the input bits in every row. $\mathcal{B}$ encodes the first n input bits, using an error correcting code, into a string of 2^{n^ε} bits (or alternatively, to a function from n^ε bits into 1). Then, $\mathcal{B}$ uses the resulting n^ε bits of the XORs to address an index in the encoded first input and returns that bit.

In order to prove an average-case lower bound one has to prove the following four facts (stated informally):

1. Most strings of length n are encoded, using an error correcting code with large relative distance, into strings of length 2^{n^ε} that represent functions (from n^ε bits into 1) that cannot be approximated by formulas of size at most $n/\log n$.
2. If a formula F approximates $\mathcal{B}$ well, then there exists a string $h \in \{0,1\}^n$ such that its encoding is hard to approximate (see Item 1) and F_h approximates $\mathcal{B}_h$ well, where $\mathcal{B}_h$ (resp. F_h) is the function $\mathcal{B}$ (resp. F) such that the first input is fixed to h.

3. Γ-shrinkage occurs with probability exponentially close to 1, over a distribution of random restrictions that takes into account the structure of the formula.

4. After applying this restriction (from the same distribution as in Item 3) most rows in the matrix represented by the second input to $\mathcal{B}$ have at least one variable that is not restricted, with probability exponentially close to 1.

Deriving a lower bound given the above facts is conceptually similar to Andreev's lower bound, but may sometimes be more difficult. Items 1 and 2 are not very hard to prove. Specifically, Item 1 follows by an application of the Johnson bound and Item 2 is a standard averaging argument. Items 3 and 4 are usually more challenging. For example, the distribution over restrictions in Item 3 may depend on the structure of the formula (i.e., it is not a uniformly random restriction), making Item 4 in the scheme above non-trivial to prove.

In this work, our technical contributions are twofold. First, we prove that 1.99-shrinkage occurs with probability exponentially close to 1 over a certain distribution (that satisfies some additional properties; see the discussion in Section 1.1.3 and Remark 1.2). Second, we decouple the dependence between Items 3 and 4 by giving a slightly different construction for the function $\mathcal{B}$ for which we prove the lower bound. Namely, instead of XORing the rows in the aforementioned matrix to derive the n^{ε} index bits, we derive the index bits by applying a bit-fixing extractor to the second input (see more details below).[1] We believe that the insights in the latter contribution might be of independent interest.

1.1.3 Our Techniques

We first describe how we prove that 1.99-shrinkage occurs with high probability and then explain our modification to the function $\mathcal{B}$ which helps us simplify the analysis of Item 4.

Shrinkage with high probability. We borrow ideas from the work of Impagliazzo, Meka and Zuckerman [IMZ12]. Impagliazzo et al. prove a theorem (among other interesting results) showing that $(2 - o(1))$-shrinkage occurs with probability that is polynomially close to 1. In [IMZ12], the theorem is proved for certain pseudorandom distributions and is used to construct pseudorandom generators with seed of length $O(s)$ that fool De Morgan formulas of size $s^{3-o(1)}$.

We use the proof technique of [IMZ12] applied to a distribution of random restrictions that takes the structure of the formula into account. In order to show that shrinkage occurs with high probability, [IMZ12] show that formulas with no "heavy" variables (variables that appear a lot more than an average variable) can be split into many "medium size" sub-formulas (subtrees). Thus, it suffices to show that the total size of all sub-formulas after applying a random restriction is small with high probability. The crucial point is that (assuming that there are no "heavy" variables) these subtrees can be gathered into large sets, such that in each set the sub-formulas are defined over disjoint variables. Thus, in each such set of sub-formulas, the sizes of the sub-formulas after restriction are independent random variables, and one can apply Chernoff-Hoeffding inequalities.

The main problem with this argument is the assumption that no "heavy" variables exist. [IMZ12] treat the "heavy" variables separately, showing overall that shrinkage occurs with probability at least $1 - 1/\mathsf{poly}(n)$. Their analysis can even be pushed to show that shrinkage over uniformly random restrictions holds with probability at least $1 - 1/2^{o(\log^2 n)}$ but not further.[2]

[1] We have learnt that the same idea in this context was suggested independently in [CKK+14] (private communication with the authors).

[2] The limitation comes from the following example. Take a random function on a junta of $O(\log n)$ variables; such

In this work, we use the formula structure to derive our restriction. We first restrict all heavy variables one by one, in each step ensuring that shrinkage occurs with probability 1. When no heavy variables are left, we apply a random restriction and analyze similarly to [IMZ12]. This technique ensures shrinkage with very high probability ($\geq 1 - 2^{-n^{\Omega(1)}}$).

Remark 1.2. *As we have stated, we prove that* 1.99*-shrinkage occurs with probability exponentially close to* 1 *over a certain distribution of random restrictions that takes into account the structure of the formula (see Theorem 4.2). We note that without additional requirements on the distribution, achieving this goal is pretty easy since it follows directly from [Hås98].*

However, in order to use "shrinkage with high probability" to prove an average-case lower bound in the scheme from Section 1.1.2, one needs to prove that shrinkage occurs with very high probability over a distribution with additional properties. More specifically, the following property is sufficient: the distribution is defined by some process that at each step chooses a variable (possibly depending on the structure of the formula) and then fixes the value of the chosen variable uniformly at random to either 0 *or* 1*. We refer to such a distribution as a distribution of* random valued restrictions*. In the proof, we use the fact that the restrictions are* random valued *to derive correlation bounds for the original function from correlation bounds for the restricted function, using an averaging argument (see Section 7).*

In Section 4, we define a distribution of random valued restrictions, and prove that 1.99*-shrinkage occurs with probability exponentially close to* 1 *over this distribution.*

Simplification of Item 4. As our restriction depends on the structure of the formula, it is not a uniformly random restriction, and one needs to work hard in order to show Item 4 in the proof scheme of Section 1.1.2. In this work, we generalize Andreev's scheme (see Section 1.1.1) to work with any distribution over restrictions, as long as it keeps $k = 100n^{\varepsilon}$ variables unrestricted (where n^{ε} is the input length to the hard function). Instead of generating the index to the hard function by simply XORing bits of the n bits of the second input, we apply a more complicated function on those variables, i.e., we apply a bit-fixing extractor.

Bit-fixing extractors were introduced in [CGH$^+$85] and then later constructed in [KZ07, GRS06, Rao09] with better and better parameters. Intuitively, a bit-fixing extractor is a function which takes n bits of input, outputs n^{ε} bits and ensures that if k of the input variables are truly random and the rest are fixed to some constants, then the output is very close to the uniform distribution over n^{ε} bits. This allows us to argue that a hard function defined on n^{ε} bits, is hard on the output of the bit-fixing extractor as well.

We use the fact that we can also use an advice (seed) for the bit-fixing extractor as part of the input and give a construction of a bit-fixing extractor with better parameters (smaller error and lower min-entropy) than bit-fixing extractors that do not assume access to an advice [KZ07, GRS06, Rao09].[3]

We hope that the idea to use a bit-fixing extractor can be helpful in other works. In general, instead of arguing that formulas (or other models) shrink under random restrictions to derive lower

a function has, with high probability, formula size $\mathsf{poly}(n)$. Applying a random restriction leaving each variable alive with probability $1/n$, independently, leaves the formula with the same size with probability $1/n^{O(\log n)} = 1/2^{O(\log^2 n)}$. This example shows the difference between taking totally random restriction and random restrictions that take into account the formula structure.

[3]This allows us to get that our function cannot be computed on $1/2+2^{-r}$ fraction of the inputs by formulas of size at most $n^3/(r^2\mathsf{polylog}n)$. Using the explicit bit-fixing extractors instead would result in a function that cannot be computed on $1/2+2^{-r^d}$ fraction of the inputs by formulas of size at most $n^3/(r^2\mathsf{poly}\log n)$, for a small (unspecified) constant d.

bounds, using a bit-fixing extractor one only needs to argue that there *exists* some restriction leaving k variables unrestricted under which the formula shrinks well. In other words, when proving worst-case lower bound, one can consider *best-case restrictions* instead of random restrictions. When proving average-case lower bounds, one can consider any distribution of *random valued restrictions* (as in Remark 1.2) for which the formula shrinks well with high probability.

The techniques of [KR13]. The construction and proof of [KR13] follow the blueprint given in Section 1.1.2. There, for Item 3 it is shown that 1.5-shrinkage occurs with very high probability over some distribution that takes into account the structure of the formula. Technically, the proof uses the Azuma inequality and it is very different from the proof in this version. Furthermore, unlike in the current version, the proof of Item 4 depends heavily on the structure of the distribution and requires a series of reductions to bins and balls adversary games. Even though this work simplifies both steps, we believe that these techniques may still be interesting in their own right. See [KR13] for more information.

1.2 Related Work

Independent of our work, Chen et al. [CKK$^+$14] simplified the average-case lower bound of [KR13] using a bit-fixing extractor. This is quite similar to our simplification of Item 4 from Section 1.1.2. We emphasize that [CKK$^+$14] obtain an average-case lower bound of $n^{2.49}$ (as [KR13]), whereas we obtain an average-case lower bound of $n^{2.99}$ using our improved shrinkage result. The quantitive difference between our result and theirs is due to the fact that they prove that 1.49-shrinkage occurs with high probability, whereas we prove that 1.99-shrinkage occurs with high probability.

1.3 Paper Organization

The rest of the paper is organized as follows. In Section 2 we give some general notations that are used throughout the paper and some preliminary material and definitions. In Section 3 we give the construction of the hard function. In Section 4 we prove our "shrinkage with very high probability" theorem. In Section 5 we provide a simple construction of a bit-fixing extractor that uses an advice. In Section 6 we prove that composing an error correction code with a bit-fixing extractor almost always yields a function that is hard to approximate under any restriction. Finally, in Section 7 we prove the main theorem of this paper (Theorem 1.1).

2 Preliminaries

We start with some general notations. Throughout the paper we only consider De Morgan formulas and not always explicitly mention it.

We denote by $[n]$ the set of numbers $\{1, 2, \ldots, n\}$. For $i \in [n]$ and for $x \in \{0,1\}^n$, denote by x_i the i-th bit of x. We denote by $e_i \in \{0,1\}^n$ the vector with one on the i-th coordinate and zero elsewhere. We use logarithms to base two by default. We denote by $\mathbf{U}_k$ the uniform distribution over $\{0,1\}^k$. For a distribution $\mathcal{D}$ we denote by $x \sim \mathcal{D}$ a random element sampled according to $\mathcal{D}$. For a set X we denote by $x \sim X$ a random element sampled according to the uniform distribution from X. For two functions $f\colon \{0,1\}^s \to \{0,1\}^m$ and $g\colon \{0,1\}^n \to \{0,1\}^s$, we denote by $f \circ g\colon \{0,1\}^n \to \{0,1\}^m$ the composition of f and g, i.e., $f \circ g(x) = f(g(x))$.

Boolean Formulas

Definition 2.1. *A* De Morgan formula *is a binary tree with OR and AND gates with fan-in 2 on the internal nodes, and variables or their negations on the leaves.*

Definition 2.2. *The* size *of a De Morgan formula F is the number of leaves in it and is denoted by $L(F)$. For a function $f\colon \{0,1\}^n \to \{0,1\}$, we denote by $L(f)$ the size of the smallest De Morgan formula computing the function f.*

Definition 2.3 (Restriction)**.** *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a Boolean function. A* restriction *ρ is a vector of length n of elements from $\{0,1,*\}$. We denote by $f|_\rho$ the function f restricted according to ρ in the following sense: if $\rho_i = *$, then the i-th input bit of f is unassigned, and otherwise the i-th input bit of f is assigned to be ρ_i.*

We denote by $\mathfrak{R}_k$ the set of restrictions that leave k variables unassigned.

Definition 2.4 (p-Random Restriction)**.** *A p-*random restriction *is a restriction as in Definition 2.3 that is sampled in the following way. For every $i \in [n]$, independently with probability p set $\rho_i = *$ and with probability $\frac{1-p}{2}$ set ρ_i to be 0 and 1, respectively. We denote this distribution of restrictions by $\mathcal{R}_p$.*

We extend the definition of a restriction (Definition 2.3) to De Morgan formulas as well. For a given formula F over n variables computing a Boolean function f, and a restriction $\rho \in \{0,1,*\}^n$, we denote by $F|_\rho$ the formula of minimal size that computes $f|_\rho$. In other words, we assign values to the variables fixed by the restriction and then simplify the formula to the minimal size formula that is equivalent to it.[4]

Definition 2.5 (Average-Case Hardness)**.** *A function $f\colon \{0,1\}^n \to \{0,1\}$ is said to be (s,ε)-*hard *if for any De Morgan formula F of size at most s*

$$\Pr_{x\sim\{0,1\}^n}[F(x) = f(x)] \leq \frac{1}{2} + \varepsilon\ .$$

Probability

We state three variants of the well-known Chernoff/Hoeffding inequality.

Proposition 2.6 (Chernoff/Hoeffding Inequalities, [Che52, Hoe63, MR95])**.** *Let $X = \sum_{i=1}^n X_i$ be a sum of independent random variables $X_1, \ldots, X_n$. If for every $i \in [n]$ there exists $a_i, b_i \in \mathbb{R}$ such that $a_i \leq X_i \leq b_i$, then for $t > 0$,*

$$\Pr[X - \mathbf{E}[X] \geq t] \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right). \tag{2.1}$$

If we further assume that for every $i \in [n]$ we have $0 \leq X_i \leq b$, and $\mathbf{E}[X_i] \leq E$, then

$$\Pr[X \geq (1+\delta)\cdot n \cdot E] \leq \begin{cases} \exp(-\delta^2 \cdot n \cdot E/3b) & 0 \leq \delta < 1 \\ \exp(-\delta^2 \cdot n \cdot E/((2+\delta)b)) & 1 \leq \delta\ . \end{cases} \tag{2.2}$$

If for all $i \in [n]$ we have $0 \leq X_i \leq b$, then for $\delta \in (0,1)$,

$$\Pr[X \leq (1-\delta)\cdot \mathbf{E}[X]] \leq \exp(-\delta^2 \cdot \mathbf{E}[X]/2b)\ . \tag{2.3}$$

[4] We remark in the literature it is common to define the restricted formula by plugging constants to the tree and then applying a set of simplification rules on the resulting formula to reduce its size (see e.g., the simplification rules of Håstad [Hås98]). This approach is more algorithmic in nature (and has its advantages – see the comment before Theorem 4.3), but is not required for our results, as we are using restrictions only in the analysis.

We will use the following simple lemma.

Lemma 2.7. *Let X be a random variable taking values in the range $[0,1]$ and let B be an event such that $\mathbf{Pr}[B] > 0$. Then, $\mathbf{E}[X|B] \geq \mathbf{E}[X] - \mathbf{Pr}[\neg B]$. In particular, if X is an indicator of an event A, then $\mathbf{Pr}[A|B] \geq \mathbf{Pr}[A] - \mathbf{Pr}[\neg B]$.*

Proof. If $\mathbf{Pr}[B] = 1$ this is obvious so we can assume $\mathbf{Pr}[B] \in (0,1)$ and get

$$\mathbf{E}[X] = \mathbf{E}[X|B] \cdot \mathbf{Pr}[B] + \mathbf{E}[X|\neg B] \cdot \mathbf{Pr}[\neg B] \leq \mathbf{E}[X|B] + \mathbf{Pr}[\neg B] \ ,$$

as needed. □

We will use the notion of statistical distance.

Definition 2.8 (Statistical Distance)**.** *Let Ω be some finite set. Let P and Q be two distributions on Ω. The* statistical distance *between P and Q is defined as*

$$|P - Q| = \max_{A \subseteq \Omega} \left| \mathop{\mathbf{Pr}}_{P}(A) - \mathop{\mathbf{Pr}}_{Q}(A) \right| \ .$$

If $|P - Q| \leq \varepsilon$ we say that P is ε-close to Q.

We define k-wise independent distributions.

Definition 2.9 (k-wise Independent Distribution)**.** *A distribution $\mathcal{D}$ over $\{0,1\}^n$ is* k-wise independent *if and only if for all distinct $i_1, \ldots, i_k \in [n]$ and all $a_1, \ldots, a_k \in \{0,1\}$*

$$\mathop{\mathbf{Pr}}_{x \sim \mathcal{D}}[x_{i_1} = a_1 \wedge \cdots \wedge x_{i_k} = a_k] = \frac{1}{2^k} \ .$$

Bit-Fixing and Affine Extractors

Two ingredients of our construction are a bit-fixing extractor and an affine extractor, which we define next.

Definition 2.10 (Bit-Fixing Source)**.** *A distribution X over $\mathbb{F}_2^n$ is an (n,k)-*bit-fixing source *if there exist k distinct indices $i_1, \ldots, i_k$ such that the k-tuple $(X_{i_1}, \ldots, X_{i_k})$ is uniformly distributed over $\{0,1\}^k$, and for $i \notin \{i_1, \ldots, i_k\}$, X_i is a fixed constant. We refer to k as the entropy of the source.*

An affine source, that we define next, is a generalization of a bit-fixing source.

Definition 2.11 (Affine Source)**.** *A distribution X over $\mathbb{F}_2^n$ is an (n,k)-*affine source *if there exist k linearly independent vectors $v_1, \ldots, v_k \in \mathbb{F}_2^n$ and another vector $v_0 \in \mathbb{F}_2^n$ such that X is uniformly distributed over $v_0 + \mathsf{span}\{v_1, \ldots, v_k\}$. We refer to k as the dimension or entropy of the source.*

Definition 2.12 (Bit-Fixing Extractor, Affine Extractor)**.** *An (n,k)-*bit-fixing extractor *(*affine extractor*, resp.) with error ε and output length r is a function $\mathsf{Ext}\colon \{0,1\}^n \to \{0,1\}^r$ such that for every (n,k)-*bit-fixing source *((n,k)-*affine source*, resp.) X the distribution of $\mathsf{Ext}(X)$ is ε-close to the uniform distribution in statistical distance, i.e.,*

$$|\mathsf{Ext}(X) - \mathbf{U}_r| \leq \varepsilon \ .$$

We will later denote bit-fixing extractors by BFExt and affine extractors by AExt.

Coding Theory

We recall basic terminology and results from coding theory.

Definition 2.13. *Let Σ be an* alphabet *(i.e., a finite set) of size q. An $(n,k,d)_q$* code *C is a subset of Σ^n, where $k = \log(|C|)/\log(q)$ is called the* dimension *of the code, and d is the* minimal (Hamming) distance *between any two different strings in C. A code C can be thought of as a mapping from Σ^k to Σ^n such that every two outputs of the mapping differ in at least d locations. The mapping procedure sometimes referred to as the* encoding function *of C. The* relative distance *of C is $\delta = d/n$. We say that a code is an $[n,k,d]_q$* linear code *if $\Sigma = \mathbb{F}_q$ is a finite field and the mapping is linear over $\mathbb{F}_q$.*

We state some well known codes and recall the definition of code concatenation. A Reed-Solomon code is a $[n, k, n-k+1]_q$, where $q \geq n$ (the size of the alphabet symbols) is a power of a prime. Given a finite field $\mathbb{F}$ on size q, and $n \leq q$ distinct elements $a_1, \ldots, a_n \in \mathbb{F}$, the Reed-Solomon code encodes $(x_0, \ldots, x_{k-1}) \in \mathbb{F}^k$ into the vector $(p(a_1), p(a_2), \ldots, p(a_n))$ where $p(a) = \sum_{j=0}^{k-1} x_j \cdot a^j$. In other words, the message is interpreted as the set of coefficients of a polynomial $p(\cdot)$ with $\deg(p) \leq k-1$, and encoded by taking the values of this polynomial on $a_1, \ldots, a_n$.

A Hadamard code is a $[2^k, k, 2^{k-1}]_2$ code that is defined as follows. A message $(x_1, \ldots, x_k) \in \mathbb{F}_2^k$ is encoded by a vector of length 2^k whose entries are indexed by vectors $y \in \mathbb{F}_2^k$; The y's entry of the encoded message equals $\langle x, y \rangle := \sum_i x_i y_i$.

Let C_{in} be an $[n,k,d]_q$-code and let C_{out} be an $[N,K,D]_{q^k}$-code. Then, the concatenation of C_{out} with C_{in} is an $[nN, kK, \geq dD]_q$-code which is defined as follows. A message $x \in \left(\mathbb{F}_{q^k}\right)^K \cong (\mathbb{F}_q)^{kK}$ is encoded using the outer code into $C_{\mathsf{out}}(x) \in \left(\mathbb{F}_{q^k}\right)^N$. Then, each of the N symbols in $C_{\mathsf{out}}(x)$ is encoded using the inner code, resulting in a vector in $(\mathbb{F}_{q^n})^N \cong (\mathbb{F}_q)^{nN}$. It is well known that the distance of the concatenation of two codes is at least dD.

Definition 2.14. *Let $0 \leq \rho \leq 1$ and $L \geq 1$. A code $C \subset \{0,1\}^n$ is (ρ, L)-*list decodable *if for every $y \in \{0,1\}^n$,*

$$|\{c \in C \ : \ \Delta(y,c) \leq \rho n\}| \leq L \ ,$$

where Δ denotes the Hamming distance.

We state the well known Johnson bound for codes with binary alphabet. This version of the bound was taken from [Rud07] for the case of binary alphabet.

Proposition 2.15 (Johnson Bound)**.** *Let $C \subseteq \{0,1\}^n$ be an $(n,k,d)_2$ code with relative distance $\delta = d/n \leq 1/2$. It holds that C is $(\rho, 2dn)$-list decodable for any*

$$\rho < \frac{1}{2} \cdot \left(1 - \sqrt{1-2\delta}\right) \ .$$

3 Construction of the Function

Our construction is parameterized by n and r, and can be thought of as a family of functions as it is defined for infinitely many possibilities for these parameters. Let $c \geq 8$ be a large enough constant. We assume that $c \cdot \log(n) \leq r \leq n^{1/3}$ and that n is large enough.

We define a function $h \colon \{0,1\}^{4n} \times \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ that takes three inputs: $x \in \{0,1\}^{4n}$, $y \in \{0,1\}^n$ and $s \in \{0,1\}^n$. We use two ingredients in our construction: an error correcting code and a bit-fixing extractor.

Let $\mathcal{C}$ be a $[2^r, 4n, d]_2$ error correcting code with $d \geq 2^r \left(\frac{1}{2} - \frac{1}{2^{r/4}}\right)$. $\mathcal{C}$ encodes $x \in \{0,1\}^{4n}$ to $\mathrm{Enc}_x^{\mathcal{C}} \in \{0,1\}^{2^r}$ and has relative distance $\delta = d/2^r \geq 1/2 - 2^{-r/4}$. One may view each codeword also as a Boolean function $\mathrm{Enc}_x^{\mathcal{C}} : \{0,1\}^r \to \{0,1\}$. The construction of the code $\mathcal{C}$ is described in Section 3.1.

Our bit-fixing extractor is a function $\mathsf{BFExt}_s \colon \{0,1\}^n \to \{0,1\}^r$ parameterized by the input s, which is of length n. The exact definition of this function and its properties are described in Section 5.

The function h is defined as

$$h(x, y, s) = \mathrm{Enc}_x^{\mathcal{C}}(\mathsf{BFExt}_s(y)) \ .$$

We remark that both computations $z = \mathsf{BFExt}_s(y)$ and $\mathrm{Enc}_x^{\mathcal{C}}(z)$ can be done in polynomial time given the inputs (s, y) and (x, z), respectively (see Sections 3.1 and 5).

3.1 The Error Correcting Code

As part of the construction we need a code $\mathcal{C}$ with the following parameters. $\mathcal{C}$ should be a $[2^r, 4n, d]_2$ code with $d \geq 2^r \left(\frac{1}{2} - \frac{1}{2^{r/4}}\right)$ where $8 \log n \leq r \leq n^{1/3}$.

The code $\mathcal{C}$ can be obtained in the following way. Consider a concatenation of a Reed Solomon $\left[2^{r/2}, 8n/r, 2^{r/2} - 8n/r + 1\right]_{2^{r/2}}$ code R and a Hadamard $\left[2^{r/2}, r/2, 2^{r/2-1}\right]_2$ code H. Concatenating R and H we get a $[2^r, 4n, d]_2$ code $\mathcal{C}$ where

$$d \geq 2^{r/2-1}\left(2^{r/2} - 8n/r\right) = 2^r\left(\frac{1}{2} - \frac{4n}{r2^{r/2}}\right) \geq 2^r\left(\frac{1}{2} - \frac{1}{2^{r/4}}\right)$$

for large enough r and n, where $8 \log n \leq r \leq n^{1/3}$.

We note that for any given $x \in (\mathbb{F}_2)^{4n}$ and $i \in [2^r]$, computing the encoding of x at coordinate i (according to $\mathcal{C}$) can be done in $\mathsf{poly}(n, r)$ time by computing a symbol of a Reed-Solomon code and then computing a symbol of the Hadamard encoding of that symbol.

4 Shrinkage with Very High Probability

In this section we prove that the shrinkage property of De Morgan formulas holds with very high probability. We begin by stating the main theorem of this section.

Definition 4.1 (Random Valued Restrictions)**.** *A distribution of random restrictions (as in Definition 2.3) is called* random valued *if it can be defined by an iterative process such that at each step a variable is chosen (possibly depending on the structure of the formula and previous choices) and then the value of the chosen variable is randomly fixed to* 0 *or* 1*.*

Usually, we will associate a parameter k with a random valued restriction which states that the restriction leaves k variables unassigned (i.e., it is in $\mathfrak{R}_k$).

Theorem 4.2. *Let $c > 0$ be any constant and let F be a formula over n variables of size at most n^c. Then, for large enough n, there exists a constant $c'' > 0$ (where c'' depends only on c) such that for any k in the range $c'' \cdot \log(n) \leq k \leq n$ there is a distribution $\mathcal{T}_k$ of* random valued restrictions*, each restriction leaving exactly k variables alive, such that*

$$\Pr_{\rho \sim \mathcal{T}_k}\left[L(F|_\rho) \leq \mathsf{poly}\log(n) \cdot \left(\frac{k}{n}\right)^2 \cdot L(F)\right] \geq 1 - \varepsilon_{\mathrm{shr}} \ ,$$

where $\varepsilon_{\mathrm{shr}} = 2^{-\Omega(k)}$.

Our proof is based on the results of Håstad [Hås98] and Tal [Tal14] that state that shrinkage of De Morgan formulas occurs in expectation. In Theorem 4.3, we use the theorem of [Tal14] which gives a quantitatively slightly better result than [Hås98]. Eventually, this yields a better lower bound in Theorem 4.2 and consequently in our main theorem (improving a $2^{O(\log^2 \log n)}$ term into a $\mathsf{polylog}(n)$ term). However, the result of [Hås98] is algorithmic (i.e., the simplification rules can be efficiently implemented) and can be used to get #SAT algorithms for small formulas (see [CKK+14]).

Theorem 4.3 ([Hås98], [Tal14])**.** *Let f be a Boolean function. For every $p > 0$,*

$$\mathop{\mathbf{E}}_{\rho \sim \mathcal{R}_p} [L(f|_\rho)] \leq O\left(p^2 \cdot L(f) + p\sqrt{L(f)}\right) .$$

We define a restriction process for a formula F as follows. If F contains a heavy variable (i.e., a variable that appears in the formula many times), then we just restrict it (assign to it 0 or 1 at random). Otherwise, we treat all variables as equal and use a truly random restriction on the remaining variables. In the analysis, a removal of a "heavy" variable is pretty easy to handle since we are guaranteed that the formula shrinks well, and the second step is harder. In the analysis of the second step, we split the formula (which is just a binary tree) into parts (sub-formulas) that are almost independent, in the sense that every variable does not appear in too many parts. We show that this small dependence does not affect much, and thus we can apply Hoeffding's inequality to get the result.

Formally, for a given formula F on n variables and a parameter r, we present a randomized algorithm that samples a restriction keeping exactly r variables alive. The algorithm takes the structure of F into account, and defines a distribution of *random valued restrictions* that we denote by $\mathcal{T}'_r$. The final distribution $\mathcal{T}_k$ (as in Theorem 4.2) is obtained by iteratively sampling restrictions from $\mathcal{T}'_r$ with varying values of r (more details below and in Remark 4.8).

$F_0 \leftarrow F$.
$i \leftarrow 0$.
while $n - i > r$ **AND** there is a variable x_j in F_i that appears more than $\frac{2L(F_i)}{n-i}$ times **do**
    Assign x_j at random and let F_{i+1} be the formula F_i restricted by x_j
    $i \leftarrow i + 1$,
end while
Sample a random $\rho' \sim \mathfrak{R}_r$ on the remaining variables.

Algorithm 1: $\mathcal{T}'_r$ distribution on restrictions.

First, we argue that shrinkage occurs with very high probability for formulas that do not contain any "heavy" variable. Actually, we claim that such shrinkage holds under random restrictions sampled from $\mathcal{R}_p$, keeping each variable alive independently with probability p (where $p \approx r/n$). In Lemma 4.6, we show that a similar bound holds for restrictions sampled from $\mathfrak{R}_r$ (i.e, when the number of alive variables is exactly r), as required by the last step in $\mathcal{T}'_r$.

Lemma 4.4. *There exists a universal constant $c > 0$ such that the following holds. Let F be a De Morgan formula over n variables such that any variable appears in F in at most $2L(F)/n$ leaves. Then, for any $0 < p \leq 1$,*

$$\Pr_{\rho \sim \mathcal{R}_p} \left[L\left(F|_\rho\right) \geq c \cdot p^2 \cdot L(F)\right] \leq L(F) \cdot e^{-n \cdot p^6} .$$

The proof of Lemma 4.4 is based on ideas from [IMZ12] and [Tal14]. We use the following claim from [Tal14], that is based on ideas from [IMZ12].

Claim 4.5 ([Tal14])**.** *Let F be a De Morgan formula over the set of variables $X = \{x_1, \ldots, x_n\}$, and let $\ell \in \mathbb{N}$ be a parameter. Then, there exist $m \leq O(L(F)/\ell)$ formulas over X, denoted by $G_1, \ldots, G_m$, each of size at most ℓ, and there exists a read-once formula F' of size m such that $F'(G_1(x), \ldots, G_m(x)) = F(x)$ for all $x \in \{0,1\}^n$.*

We proceed with the proof of Lemma 4.4.

Proof of Lemma 4.4. The claim is vacuous for $p < n^{-1/6}$ since it says that something occurs with probability less than something that is greater than 1. So we may assume without loss of generality that $p \geq n^{-1/6}$. We set $\ell := 1/p^2$.

Recall that the formula F does not contain "heavy" variables. Decompose the formula F into $F'(G_1, \ldots, G_m)$ as in Claim 4.5. Form a graph whose vertices are the G_i's, and put an edge between G_i and G_j, for $i \neq j$, iff the formulas G_i and G_j share a variable. This graph has $m = O(L(F)/\ell)$ vertices (see Claim 4.5). Since F does not contain "heavy" variables, the degree of this graph is at most $d \triangleq \ell \cdot \left(\frac{2L(F)}{n} - 1\right)$. It is well-known that one can greedily color a graph of degree d with $d+1$ colors, inducing a partition of the vertices into $d+1$ independent sets $\{I_1, \ldots, I_{d+1}\}$.

Denote by $f, g_1, \ldots, g_m$ the functions computed by $F, G_1, \ldots, G_m$, respectively. Consider a specific independent set I_i. Let $\rho \sim \mathcal{R}_p$ be a random restriction. By our design, each of the functions $\{g_j\}_{j \in I_i}$ is defined on a distinct set of variables, thus we get that shrinkage in each one of them occurs independently and we may apply Chernoff-Hoeffding's inequality. Let $X_j = L(g_j|_\rho)$ for $j \in [m]$. Then, $\{X_j\}_{j \in I_i}$ are independent random variables bounded in $[0, \ell]$ whose expectancy is $O(p^2 \cdot \ell + 1) = O(1)$, by Theorem 4.3. Denote the constant hidden in the $O(1)$-notation by c_2, and assume without loss of generality that $c_2 \geq 1$. Denote by $X^{(i)} = \sum_{j \in I_i} X_j$. If $|I_i| < n/\ell^2$, then we add dummy random variables $Y_{|I_i|+1}, \ldots, Y_{n/\ell^2}$, that are always equal to 0, to the sum that defines $X^{(i)}$. By doing so, we guarantee that $X^{(i)}$ is the sum of at least n/ℓ^2 independent random variables. By the Chernoff-Hoeffding's bound (see Equation (2.2) with $\delta = 2$ and $E = c_2$), we get that

$$\Pr_{\rho \sim \mathcal{R}_p}\left[X^{(i)} \geq 3 \cdot \max\{n/\ell^2, |I_i|\} \cdot c_2\right] \leq e^{-\max\{n/\ell^2, |I_i|\} \cdot c_2/\ell} \leq e^{-n/\ell^3} .$$

Taking a union bound over all $d+1$ independent sets, we get that with probability at least $1 - (d+1) \cdot e^{-n/\ell^3}$ it holds that

$$X^{(i)} < 3 \cdot \max\{n/\ell^2, |I_i|\} \cdot c_2$$

for all $i \in [d+1]$. In such a case,

$$\begin{aligned} L(f|_\rho) \leq \sum_{i=1}^{d+1} X^{(i)} &\leq 3 \cdot c_2 \cdot \left(\sum_i |I_i| + (d+1) \cdot \frac{n}{\ell^2}\right) \\ &\leq 3 \cdot c_2 \cdot \left(m + \frac{2\ell \cdot L(F)}{n} \cdot \frac{n}{\ell^2}\right) = O\left(m + \frac{L(F)}{\ell}\right) = O(L(F) \cdot p^2) . \end{aligned}$$

Overall, there exists a constant c such that

$$\Pr_{\rho \sim \mathcal{R}_p}\left[L\left(f|_\rho\right) \geq c \cdot p^2 \cdot L(F)\right] \leq (d+1) \cdot e^{-n/\ell^3} \leq L(F) \cdot e^{-n \cdot p^6} .$$

Since $L(f|_\rho) = L(F|_\rho)$, this completes the proof. □

In the proof of Lemma 4.4, it was convenient to treat variables in the restriction as independent random variables, which advocated the usage of $\mathcal{R}_p$. For the rest of this section, and in particular for Theorem 4.2, it is important to consider restrictions leaving exactly k variables alive (i.e., in $\mathfrak{R}_k$), as we later argue that under such restrictions the bit-fixing extractor outputs a close to uniform string. The next lemma derives a shrinkage result for $\mathfrak{R}_r$ from the shrinkage result for $\mathcal{R}_{2r/n}$.

Lemma 4.6. *There exists a universal constant $c' > 0$ such that the following holds. Let F be a De Morgan formula over n variables such that any variable appears in F in at most $2L(F)/n$ leaves. Then, for any integer $1 \leq r \leq n$,*

$$\mathop{\mathbf{Pr}}_{\rho' \sim \mathfrak{R}_r} \left[L\left(F|_{\rho'}\right) \geq c' \cdot (\tfrac{r}{n})^2 \cdot L(F) \right] \leq 2 \cdot L(F) \cdot e^{-r^6/n^5} \ .$$

Proof. Let c be the constant guaranteed by Lemma 4.4, and assume without loss of generality that $c \geq 1$. Take $c' = 4c \geq 4$. If $r > n/2$, then the claim surely holds as $L(F|_{\rho'}) \leq L(F) < L(F) \cdot c' \cdot (r/n)^2$. Hence, for the rest of the proof we assume without loss of generality that $r \leq n/2$.

We let $p = 2r/n$, and couple the distributions $\mathcal{R}_p$ and $\mathfrak{R}_r$. To do so, we draw n real numbers $a_1, \ldots, a_n$ uniformly from $[0, 1]$. In addition, we draw n bits $b_1, \ldots, b_n \in \{0, 1\}$ uniformly at random. To sample a restriction ρ from $\mathcal{R}_p$ simply assign $\rho(i) = *$ if $a_i < p$, and $\rho(i) = b_i$ otherwise. Let p' be the minimum value such that exactly r out of $\{a_1, \ldots, a_n\}$ are smaller than p'. To sample a restriction ρ' from $\mathfrak{R}_r$ assign $\rho'(i) = *$ if $a_i < p'$, and $\rho(i) = b_i$ otherwise.

Whenever $p' \leq p$ we have that ρ' is a refinement of ρ, i.e., any bit that is fixed to a constant in ρ is fixed to the same constant ρ', and possibly some variables that were unrestricted in ρ are fixed in ρ'. In such a case $L(F|_{\rho'}) \leq L(F|_{\rho})$, since we may first fix the variables that ρ fixes, and then potentially fix additional variables that ρ' fixes, which can only reduce the formula size. Thus, we have

$$\mathop{\mathbf{Pr}}_{\rho' \sim \mathfrak{R}_r} [L(F|_{\rho'}) \geq c' \cdot (\tfrac{r}{n})^2 \cdot L(F)] \leq \mathop{\mathbf{Pr}}_{\rho \sim \mathcal{R}_p} [L(F|_{\rho}) \geq c' \cdot (\tfrac{r}{n})^2 \cdot L(F)] + \mathbf{Pr}[p' > p]. \tag{4.1}$$

To bound the first term in the RHS of Eq. (4.1) we note that $c' \cdot (\frac{r}{n})^2 = c \cdot p^2$ and apply Lemma 4.4, which gives

$$\mathop{\mathbf{Pr}}_{\rho \sim \mathcal{R}_p} [L(F|_{\rho}) \geq c' \cdot (\tfrac{r}{n})^2 \cdot L(F)] = \mathop{\mathbf{Pr}}_{\rho \sim \mathcal{R}_p} [L(F|_{\rho}) \geq c \cdot p^2 \cdot L(F)] \leq L(F) \cdot e^{-n \cdot p^6} \leq L(F) \cdot e^{-64r^6/n^5}.$$

To bound the second term in the RHS of Eq. (4.1), we analyze the probability that less than r of the a_i's got value at most p. We expect $p \cdot n = 2r$ of the a_i's to get value smaller than p. Since the a_i's are independent random variables, we may apply Chernoff's bound (Eq. (2.3)) to get that $\mathbf{Pr}[p' > p] \leq \exp(-r/4)$.

Plugging the bounds on both terms in Eq. (4.1) and using the assumption that $r \leq n/2$, gives

$$\mathop{\mathbf{Pr}}_{\rho' \sim \mathfrak{R}_r} [L(F|_{\rho'}) \geq c' \cdot (\tfrac{r}{n})^2 \cdot L(F)] \leq L(F) \cdot e^{-64r^6/n^5} + e^{-r/4} \leq 2 \cdot L(F) \cdot e^{-r^6/n^5}. \qquad \square$$

Corollary 4.7. *Let c' be the constant from Lemma 4.6. Let F be a De Morgan formula over n variables. For any $1 \leq r \leq n$ it holds that*

$$\mathop{\mathbf{Pr}}_{\rho \sim \mathcal{T}'_r} \left[L(F|_{\rho}) \geq c' \cdot \left(\tfrac{r}{n}\right)^2 \cdot L(F) \right] \leq 2 \cdot L(F) \cdot e^{-r^6/n^5} \ .$$

Proof. Assume that we have $0 \leq h \leq n - r$ heavy variables that cause $\rho \sim \mathcal{T}'_r$ enter the while loop in Algorithm 1. Let $z_1, z_2, \ldots, z_h$ be the variables assigned in the while loop and denote by $F' = F_h$ the formula F after restricting $z_1, z_2, \ldots, z_h$. Each z_i, conditioned on the previous choices of values,

must reduce the size of the formula by a factor of at least $\left(1-\frac{2}{n-(i-1)}\right) \leq \left(1-\frac{1}{n-(i-1)}\right)^2$, hence the size of F' is

$$L\left(F'\right) \leq L(F) \cdot \left(1-\frac{1}{n}\right)^2 \cdot \left(1-\frac{1}{n-1}\right)^2 \cdots \left(1-\frac{1}{n-h+1}\right)^2 = L(F) \cdot \left(\frac{n-h}{n}\right)^2 . \qquad (4.2)$$

Apply Lemma 4.4 on the formula F' that contains $n-h$ variables. Since

$$2 \cdot L(F') \cdot e^{-r^6/(n-h)^5} \leq 2 \cdot L(F') \cdot e^{-r^6/n^5} \leq 2 \cdot L(F) \cdot e^{-r^6/n^5}$$

it follows that

$$\mathop{\mathbf{Pr}}_{\rho' \sim \mathcal{R}_r}\left[L(F'|_{\rho'}) \geq c' \cdot \left(\tfrac{r}{n-h}\right)^2 \cdot L(F')\right] \leq 2 \cdot L(F) \cdot e^{-r^6/n^5} .$$

Following our notations and Algorithm 1, every restriction $\rho \sim \mathcal{T}'_r$ and a formula F corresponds to a restriction $\rho' \sim \mathcal{R}_r$ and a formula F'. So,

$$\begin{aligned}
2 \cdot L(F) \cdot e^{-r^6/n^5} &\geq \mathop{\mathbf{Pr}}_{\rho' \sim \mathcal{R}_r}\left[L(F'|_{\rho'}) \geq c' \cdot \left(\tfrac{r}{n-h}\right)^2 \cdot L(F')\right] \\
&\geq \mathop{\mathbf{Pr}}_{\rho' \sim \mathcal{R}_r}\left[L(F'|_{\rho'}) \geq c' \cdot \left(\tfrac{r}{n-h}\right)^2 \cdot L(F) \cdot \left(\tfrac{n-h}{n}\right)^2\right] && \text{(Equation (4.2))} \\
&\geq \mathop{\mathbf{Pr}}_{\rho \sim \mathcal{T}'_r}\left[L(F|_{\rho}) \geq c' \cdot \left(\tfrac{r}{n}\right)^2 \cdot L(F)\right], && (L(F'|_{\rho'}) = L(F|_{\rho}))
\end{aligned}$$

which concludes the proof of the corollary. □

Remark 4.8. *Note that Corollary 4.7 is useful only for $r > n^{5/6}$. This range of r's is not enough to derive the lower bound of Theorem 1.1, so we need to be able to argue a similar statement for much smaller values of r. This is what we achieve in Theorem 4.2.*

Next, we prove the main theorem of this section (Theorem 4.2).

Proof of Theorem 4.2. We apply Corollary 4.7 $t \geq 1$ times where t will be determined later. Let $F_0 = F$ and $n_0 = n$. We define a sequence of formulas $F_0, F_1, \ldots, F_t$, where for $i = 1, \ldots, t$, the formula F_i is a restriction of F_{i-1} and is defined on n_i variables. We state the following claim that suggests the existence of good parameters t and $n_0, n_1, \ldots, n_t$, and defer its proof for later.

Claim 4.9. *There exist $t \in \mathbb{N}$ and a sequence of integers $n = n_0 \geq n_1 \geq \ldots \geq n_t = k$ such that*

1. *for $1 \leq i \leq t$, $\frac{(n_i)^6}{(n_{i-1})^5} = \Omega(k)$*
2. *$t = O(\log \log n)$*

For $i = 1, \ldots, t$ we apply a random restriction $\rho_i \sim \mathcal{T}'_{n_i}$ on the formula F_{i-1} to get $F_i := F_{i-1}|_{\rho_i}$. Our final formula F' equals to F_t. Let $\mathcal{E}_i$ be the event that $L(F_i) \leq c' \cdot \left(\frac{n_i}{n_{i-1}}\right)^2 \cdot L(F_{i-1})$. Corollary 4.7 and Claim 4.9 gives

$$\mathbf{Pr}\left[\mathcal{E}_i | \mathcal{E}_1, \ldots, \mathcal{E}_{i-1}\right] > 1 - 2 \cdot L(F) \cdot \exp\left(\frac{-(n_i)^6}{(n_{i-1})^5}\right) = 1 - 2 \cdot L(F) \cdot \exp\left(-\Omega(k)\right) .$$

Standard calculation shows that all events $\mathcal{E}_1, \ldots, \mathcal{E}_t$ hold simultaneously with probability at least $1 - 2t \cdot L(F) \cdot \exp\left(-\Omega(k)\right)$. By the assumption that $L(F) \leq n^c$, there exists a constant c'' such that if $k \geq c'' \cdot \log n$, then $1 - 2t \cdot L(F) \cdot \exp\left(-\Omega(k)\right) \geq 1 - 2^{-\Omega(k)}$.

In the case that all events $\mathcal{E}_i$ hold, we have

$$L(F') \leq L(F_t) \leq (c')^t \cdot \left(\prod_{i=1}^{t} \left(\frac{n_i}{n_{i-1}} \right)^2 \right) \cdot L(F_0) = (c')^t \cdot \left(\frac{n_t}{n_0} \right)^2 \cdot L(F_0) = (c')^t \cdot \frac{k^2}{n^2} \cdot L(F_0) \ .$$

Using Item 2 of Claim 4.9, $t = O(\log \log n)$, it follows that

$$L(F') \leq 2^{O(\log \log n)} \cdot \frac{k^2}{n^2} \cdot L(F) = \mathsf{polylog}(n) \cdot \frac{k^2}{n^2} \cdot L(F) \ .$$

We get that with probability at least $1 - 2^{\Omega(k)}$ the formula size of F' is at most $\mathsf{polylog}(n) \cdot \frac{k^2}{n^2} \cdot L(F)$. Since we applied a sequence of t random valued restrictions, we got a random valued restriction overall. In addition, we left exactly k variables alive, which completes the proof. □

Proof of Claim 4.9. Let $\{x_i\}_{i \in \mathbb{N}}$ be the infinite sequence of real numbers defined by $x_i = \frac{k}{2} \cdot 2^{(6/5)^i}$. We have that

$$\frac{x_i^6}{x_{i+1}^5} = \tfrac{k}{2} \cdot 2^{(6/5)^i \cdot 6 - (6/5)^{i+1} \cdot 5} = \tfrac{k}{2} \ . \tag{4.3}$$

Let t be the least such that $x_t \geq n$. Using the definition of x_i we get that $t = O(\log \log n)$.

Next, we define the sequence $n_0, \ldots, n_t$. Set $n_0 := n$ and for $i = 1, \ldots, t$, set $n_i := \lceil x_{t-i} \rceil$. Notice that $n = n_0 \geq n_1 \geq n_2 \ldots \geq n_t = k$, since the sequence $\{x_i\}_{i=0}^t$ is monotone increasing and $x_0 = k$. For $i = 1, \ldots, t$, we have

$$\frac{(n_i)^6}{(n_{i-1})^5} \geq \frac{(x_{t-i})^6}{(1 + x_{t-i+1})^5} \geq \frac{(x_{t-i})^6}{(2 \cdot x_{t-i+1})^5} = \Omega(k) \ ,$$

which completes the proof. □

5 Extractors for Bit-Fixing Sources

One of the ingredients in the construction of our hard function is an extractor for bit-fixing sources (recall Definitions 2.10 and 2.12). We wish to construct a bit-fixing extractor $\mathsf{BFExt} \colon \{0,1\}^n \to \{0,1\}^r$ such that for every (n,k)-bit-fixing source X the output $\mathsf{BFExt}(X)$ is very close to the uniform distribution in statistical distance. Such an extractor was constructed by Rao.

Theorem 5.1 ([Rao09])**.** *There exist constants c and d such that for every $k(n) > \log^c n$, there exists a polynomial time computable function $\mathsf{BFExt} \colon \{0,1\}^n \to \{0,1\}^r$ that is an (n,k)-bit-fixing extractor with output length $r = k - o(k)$ and error 2^{-k^d}.*

We give a construction with better parameters which uses $O(k^2 \cdot \log n)$ bits of advice. Note that this is not an explicit bit-fixing extractor. Nevertheless, since we can have advice of size $O(n)$ without increasing the input size by more than a constant factor, we can use this advantage.

One ingredient of our construction is the following.

Definition 5.2 (Linear Condenser)**.** *An $(n, m, k_{\text{in}}, k_{\text{out}})$ $\mathsf{linear\ condenser}$ is a linear mapping $T \colon \{0,1\}^n \to \{0,1\}^m$ such that for any $S \subseteq [n]$ of size $\geq k_{\text{in}}$ we have*

$$\dim \left(T \left(\mathsf{span}\{e_i : i \in S\} \right) \right) \geq k_{\text{out}} \ ,$$

where $e_i \in \{0,1\}^n$ is the i-th unit vector in which all entries are 0 except the i-th entry which is 1.

The output of an $(n, m, k_{\text{in}}, k_{\text{out}})$ linear condenser on an (n, k_{in})-bit-fixing source is distributed uniformly over an affine subspace of $\mathbb{F}_2^m$ of dimension at least k_{out}, i.e., it is an (m, k_{out})-affine source. Thus, we can compose this linear condenser with an (m, k_{out}) affine extractor and get altogether an (n, k_{in})-bit-fixing extractor. The affine extractor that we use was given by Bourgain.[5]

Theorem 5.3 ([Bou07])**.** *Let $\delta \in (0,1)$ be any constant. There exists a constant $\lambda_\delta \in (0,1)$ such that for any m large enough there is an explicit polynomial time computable $(m, \delta m)$ affine extractor $\mathsf{AExt}\colon \{0,1\}^m \to \{0,1\}^r$, that extracts $r = \lambda_\delta \cdot m$ bits with error 2^{-r}.*

Next, we show that a random matrix is actually a good linear condenser.

Lemma 5.4. *For $k \geq 2\log n$, a random Boolean $k \times n$ matrix is an $(n, k, k, k - \sqrt{k \cdot 2\log n})$ linear condenser with probability $\geq 1 - 2^{-k \cdot \log n}$.*

Proof. We will first count the number of $k \times k$ matrices of rank $\leq d$ over $\mathbb{F}_2$. Any $k \times k$ matrix of rank $\leq d$ can be described unambiguously by specifying a subset of d rows, choosing the vectors for these rows, and then choosing the remaining $k - d$ rows as linear combinations of those d rows. This shows that there are at most

$$\binom{k}{d} \cdot 2^{dk} \cdot 2^{(k-d)d} = \binom{k}{d} \cdot 2^{2dk - d^2}$$

such matrices. Thus, the probability that a random $k \times k$ matrix over $\mathbb{F}_2$ has rank $\leq d$ is at most $\binom{k}{d} \cdot 2^{2dk-d^2}/2^{k^2} = \binom{k}{d} \cdot 2^{-(k-d)^2}$. By a union bound, the probability that any subset of k columns in a random $k \times n$ matrix induces a matrix of rank $\leq d$ is at most

$$\binom{n}{k} \cdot \binom{k}{d} \cdot 2^{-(k-d)^2} \leq n^k \cdot 2^{-(k-d)^2} = 2^{k \cdot \log(n) - (k-d)^2} .$$

For $d \leq k - \sqrt{2k \cdot \log n}$, this probability is at most $2^{-k \cdot \log n}$ which finishes the proof. □

The analysis above only relied on the fact that every $k \times k$ submatrix is uniformly random. Hence, we can replace the requirement that the $k \times n$ matrix is completely random with the requirement that the entries of the $k \times n$ matrix are sampled from a k^2-wise independent distribution (see Definition 2.9).

Corollary 5.5. *For $k \geq 2\log n$, a $k \times n$ matrix whose values are bits sampled from a k^2-wise distribution, is an $(n, k, k, k - \sqrt{k \cdot 2\log n})$ linear condenser with probability $\geq 1 - 2^{-k \cdot \log n}$.*

To summarize things in this section we state the following theorem.

Theorem 5.6. *Let n be a large enough integer and r, k be integers such that $8 \cdot \log n \leq r \leq n^{1/3}$ and $k = r/\lambda_{1/2}$ (where $\lambda_{1/2}$ was given by Theorem 5.3).*

Then, there exists a family of efficiently computable functions $\{\mathsf{BFExt}_s\colon \{0,1\}^n \to \{0,1\}^r\}_{s \in \{0,1\}^n}$ such that all but $2^{-r \cdot \log n}$ fraction of the seeds $s \in \{0,1\}^n$ are good where s is a good seed *if and only if s is in the set*

$$\mathcal{S} \triangleq \{s \in \{0,1\}^n\colon \mathsf{BFExt}_s \text{ is an } (n,k) \text{ bit-fixing extractor with error } 2^{-r}\} .$$

[5] We could have also used the inner product construction over large fields (see [Rao07]) of affine extractors. This requires k_{out} to be bigger than $m/2$, or even say $3m/4$, which can be arranged by adjusting the constants we choose later.

Proof. We use the most standard k^2-wise independent sample space that outputs $n \cdot k$ bits. The sample space is generated by polynomials of degree $k^2 - 1$ over $\mathbb{F}_{2^m}$, where m is the least integer such that $2^m \geq n \cdot k$.[6] This construction requires seeds of length $k^2 \cdot m = O(n^{2/3} \cdot \log n)$ and this is smaller than n, for a large enough n. By Corollary 5.5, at least $1 - 2^{-k \cdot \log n}$ (and this is at least $1 - 2^{-r \cdot \log n}$ since $k \geq r$) fraction of the choices of s yields an $(n, k, k, k - \sqrt{k \cdot 2 \cdot \log n})$ linear condenser. Next, we show that these seeds are *good*.

For a specific choice of s which defines an $(n, k, k, k - \sqrt{k \cdot 2 \cdot \log n})$ linear condenser, the output of the linear condenser is an affine source of dimension at least

$$k - \sqrt{k \cdot 2 \cdot \log n} \geq k - \sqrt{k \cdot k/4} = k/2 \ ,$$

using the assumption $k \geq r \geq 8 \cdot \log n$. By this guarantee on the dimension of the condenser output, the composition of the condenser with the affine extractor AExt stated in Theorem 5.3 yields an (n, k)-bit-fixing extractor that outputs $r = \lambda_{1/2} \cdot k$ bits with error 2^{-r}. □

6 Most Functions are Hard to Approximate on any Restriction

Throughout this section we state and prove theorems on the function h defined in Section 3. The first lemma states that if we restrict our attention only to *good* seeds s, then for almost all inputs $x \in \{0,1\}^{4n}$, it holds that $\mathrm{Enc}_x^{\mathcal{C}} \circ (\mathsf{BFExt}_s|_\rho)$ is hard to approximate for any restriction ρ leaving k inputs unassigned.

Lemma 6.1. *Let $\{\mathsf{BFExt}_s \colon \{0,1\}^n \to \{0,1\}^r\}_{s \in \{0,1\}^n}$ be the family of functions as in Theorem 5.6. Recall that $\mathcal{S} = \{s \in \{0,1\}^n \colon \mathsf{BFExt}_s$ is an (n,k) bit-fixing extractor with error $2^{-r}\}$. Let $\mathcal{C}$ be the $\left[2^r, 4n, 2^r\left(\frac{1}{2} - \frac{1}{2^{r/4}}\right)\right]_2$ code from Section 3.1. Let $n' = n/\log n$, $\varepsilon_{\mathrm{app}} = 2^{-r/10}$. For $s \in \mathcal{S}$, let*

$$\mathcal{H}_s = \{x \in \{0,1\}^{4n} \ : \ \mathrm{Enc}_x^{\mathcal{C}} \circ (\mathsf{BFExt}_s|_\rho) \text{ is } (n', \varepsilon_{\mathrm{app}})\text{-hard, for all } \rho \in \mathfrak{R}_k\} \ .$$

Then, $|\mathcal{H}_s| \geq 2^{4n} - 2^{3n}$.

(In other words, for any $s \in \mathcal{S}$, there exists a set $\mathcal{H}_s \subseteq \{0,1\}^{4n}$ of size at least $2^{4n} - 2^{3n}$ such that for all $x \in \mathcal{H}_s$ and all $\rho \in \mathfrak{R}_k$, the function $\mathrm{Enc}_x^{\mathcal{C}} \circ (\mathsf{BFExt}_s|_\rho)$ cannot be computed correctly by any formula of size at most n' on a fraction larger than $1/2 + \varepsilon_{\mathrm{app}}$ of the inputs.)

The second lemma is a simple averaging argument.

Lemma 6.2. *Let $\varepsilon > 0$ and let $F(x, y, s)$ be a formula such that*

$$\Pr_{x,y,s}[F(x,y,s) = h(x,y,s)] \geq 1/2 + \varepsilon \ .$$

Then, there exist $s_0 \in \mathcal{S}$ and $x_0 \in \mathcal{H}_{s_0}$ such that

$$\Pr_{y \sim \{0,1\}^n}[F(x_0,y,s_0) = h(x_0,y,s_0)] \geq 1/2 + \varepsilon - \varepsilon_{\mathrm{avg}} \ ,$$

where $\varepsilon_{\mathrm{avg}} = 2^{-r \cdot \log n} + 2^{-n}$.

We begin with the simple proof of Lemma 6.2.

[6] We can ignore extra bits if needed.

Proof of Lemma 6.2. Notice that

$$\begin{aligned}
&\mathop{\mathbf{Pr}}_{x,y,s}[F(x,y,s) = h(x,y,s) \mid s \in \mathcal{S}, x \in \mathcal{H}_s] \\
&\geq \mathop{\mathbf{Pr}}_{x,y,s}[F(x,y,s) = h(x,y,s)] - \mathbf{Pr}[s \notin \mathcal{S} \vee x \notin \mathcal{H}_s] && \text{(Using Lemma 2.7)} \\
&\geq 1/2 + \varepsilon - (2^{-r \cdot \log n} + 2^{-n}) \,. && \text{(Using Theorem 5.6 and Lemma 6.1)}
\end{aligned}$$

Using an averaging argument, there exist $s_0 \in \mathcal{S}$ and $x_0 \in \mathcal{H}_{s_0}$ such that

$$\mathop{\mathbf{Pr}}_{y}[F(x_0,y,s_0) = h(x_0,y,s_0)] \geq 1/2 + \varepsilon - (2^{-r \cdot \log n} + 2^{-n}) \,,$$

as needed. □

The rest of the section is devoted for the proof of Lemma 6.1. We begin with a definition.

Definition 6.3. *For a set of functions $\mathcal{F} \subseteq \{f \colon \{0,1\}^t \to \{0,1\}\}$ the code defined by this set $C_{\mathcal{F}} \subseteq \{0,1\}^{2^t}$ is the set of truth-tables of these functions. Alternatively, any code $\mathcal{C} \subseteq \{0,1\}^{2^t}$ defines a set of functions $\subseteq \{f \colon \{0,1\}^t \to \{0,1\}\}$.*

Next, we prove a useful lemma that states that the composition of a code with large relative distance with a function whose output is close to being uniformly distributed, results in a code with a large relative distance.

Lemma 6.4. *Let $g \colon \{0,1\}^k \to \{0,1\}^r$ be a function such that $|\mathbf{U}_r - g(\mathbf{U}_k)| < \varepsilon$. Let $\mathcal{F} \subseteq \{f \colon \{0,1\}^r \to \{0,1\}\}$ such that $C_{\mathcal{F}}$ has relative distance δ. Let*

$$\mathcal{G} = \{f \circ g \mid f \in \mathcal{F}\} \,.$$

Then, $C_{\mathcal{G}} \subseteq \{0,1\}^{2^k}$ is a code with relative distance $\geq \delta - \varepsilon$.

Proof. Let c_1, c_2 be two codewords in $C_{\mathcal{G}}$. Then, there exist $f_1, f_2 \colon \{0,1\}^r \to \{0,1\}$ such that $c_i = \mathrm{tt}(f_i \circ g)$ for $i = 1,2$ where $\mathrm{tt}(f_i \circ g) \in \{0,1\}^{2^k}$ is the truth table[7] of the function $f_i \circ g$. Let $A = \{y \in \{0,1\}^r : f_1(y) \neq f_2(y)\}$. Then, $|A| \geq 2^r \cdot \delta$ by the assumption on the relative distance of $C_{\mathcal{F}}$. By the definition of statistical distance (see Definition 2.8)

$$\mathop{\mathbf{Pr}}_{x \sim \{0,1\}^k}[g(x) \in A] \geq \mathop{\mathbf{Pr}}_{y \sim \{0,1\}^r}[y \in A] - \varepsilon \geq \delta - \varepsilon \,.$$

Thus, the number of inputs for which $f_1 \circ g$ and $f_2 \circ g$ disagree is at least $(\delta - \varepsilon) \cdot 2^k$, which completes the proof. □

We are now ready to prove Lemma 6.1.

Proof of Lemma 6.1. Let $s \in \mathcal{S}$ be some fixed seed. For any fixed $\rho \in \mathfrak{R}_k$ we upper bound the size of the set

$$\mathrm{EASY}_\rho \triangleq \{x \in \{0,1\}^{4n} \;:\; \mathrm{Enc}^{\mathcal{C}}_x \circ (\mathsf{BFExt}_s|_\rho) \text{ is not } (n', 2^{-r/10})\text{-hard}\} \,.$$

[7] More formally, for a function $f \colon \{0,1\}^n \to \{0,1\}$ we denote by $\mathrm{tt}(f) \in \{0,1\}^{2^n}$ the string which represent the truth-table of f, i.e., $\mathrm{tt}(f) = f(x_1) \ldots f(x_{2^n})$ where $x_i \in \{0,1\}^n$ is the i-th string in lexicographic order of length n.

Then, we apply a union bound over all $\rho \in \mathfrak{R}_k$ and the identity

$$\mathcal{H}_s = \{0,1\}^{4n} \setminus \bigcup_{\rho \in \mathfrak{R}_k} \mathrm{EASY}_\rho \tag{6.1}$$

to conclude that $|\mathcal{H}_s| \geq 2^{4n} - 2^{3n}$.

By definition of $\mathcal{S}$, $\mathsf{BFExt}_s|_\rho$ is a function $\{0,1\}^k \to \{0,1\}^r$ such that the statistical distance between $\mathbf{U}_r$ and $\mathsf{BFExt}_s|_\rho(\mathbf{U}_k)$ is at most $\varepsilon = 2^{-r}$. In addition, by the definition of our construction, for any two different $x_1, x_2 \in \{0,1\}^{4n}$, the encodings $\mathrm{Enc}^{\mathcal{C}}_{x_1}$ and $\mathrm{Enc}^{\mathcal{C}}_{x_2}$ have relative distance $\delta \geq 1/2 - 2^{-r/4}$. Using Lemma 6.4, the relative distance between $\mathrm{Enc}^{\mathcal{C}}_{x_1} \circ (\mathsf{BFExt}_s|_\rho)$ and $\mathrm{Enc}^{\mathcal{C}}_{x_2} \circ (\mathsf{BFExt}_s|_\rho)$ is at least $\delta - \varepsilon \geq 1/2 - 2^{-r/4} - 2^{-r}$.

Thus, for s and ρ as above, the set $\{\mathrm{Enc}^{\mathcal{C}}_x \circ \mathsf{BFExt}_s|_\rho\}_{x \in \{0,1\}^{4n}}$ defines a code with parameters $[N, K, D]_2$, where $N = 2^k$, $K = 4n$ and $D \geq N \cdot (1/2 - 2^{-r/4} - 2^{-r})$. Using Johnson bound (see Proposition 2.15), any ball of relative radius $1/2 - 2^{-r/10}$ has at most $2ND = \mathsf{poly}(2^k)$ codewords.

We will now upper bound the size of EASY_ρ. Any $x \in \mathrm{EASY}_\rho$ corresponds to a function $\mathrm{Enc}^{\mathcal{C}}_x \circ \mathsf{BFExt}_s|_\rho \colon \{0,1\}^k \to \{0,1\}$ whose relative distance is $\leq 1/2 - 2^{-r/10}$ from some function that can be computed using a formula of size n'. Let $N_{n',k}$ be the number of formulas of size n' on k variables. Then, $N_{n',k}$ is at most $(9k)^{n'}$ (see [Juk12, Theorem 1.23]). Overall

$$|\mathrm{EASY}_\rho| \leq N_{n',k} \cdot \mathsf{poly}(2^k) \leq (9k)^{n'} \cdot 2^{O(k)} \leq 9^{n'} \cdot 2^{n' \log k} \cdot 2^{O(k)} \leq 2^{n + o(n)} \ .$$

Applying a union bound over all $\rho \in \mathfrak{R}_k$ and using the fact that $|\mathfrak{R}_k| \leq 3^n$ gives

$$\left| \bigcup_{\rho \in \mathfrak{R}_k} \mathrm{EASY}_\rho \right| \leq 3^n \cdot 2^{n + o(n)} \leq 2^{3n} \ .$$

Plugging this into Equation (6.1) gives $|\mathcal{H}_s| \geq 2^{4n} - 2^{3n}$, as needed. □

7 Proof of Main Theorem

In this section we prove the main theorem of this paper (Theorem 1.1).

Theorem 7.1 (Restating Theorem 1.1)**.** *There exists a constant $c \geq 8$ such that for any large enough integer n and any $c \cdot \log(n) \leq r \leq n^{1/3}$ the following holds. There is an explicit (computable in polynomial time) Boolean function $h \colon \{0,1\}^{6n} \to \{0,1\}$ such that any formula of size $\frac{n^3}{r^2 \cdot \mathsf{polylog}(n)}$ computes h correctly on a fraction of at most $1/2 + 2^{-\Omega(r)}$ of the inputs.*

Proof. Consider the function h constructed in Section 3. Recall that k, the entropy of our bit-fixing extractor, is equal to $r/\lambda_{1/2}$, where $\lambda_{1/2}$ is some universal constant (see Theorem 5.6). Let

$$\varepsilon := \max\{\varepsilon_{\mathrm{avg}}, \varepsilon_{\mathrm{shr}}, \varepsilon_{\mathrm{app}}\} = \max\{2^{-n} + 2^{-r \cdot \log n}, 2^{-\Omega(k)}, 2^{-r/10}\} = 2^{-\Omega(r)} \ .$$

Assume that F is a formula that approximates h with probability $\geq 1/2 + 3\varepsilon$. According to Lemma 6.2, there exists $s_0 \in \mathcal{S}$ and $x_0 \in \mathcal{H}_{s_0}$ such that

$$\Pr_{y \sim \{0,1\}^n}[F(x_0, y, s_0) = h(x_0, y, s_0)] \geq 1/2 + 3\varepsilon - \varepsilon_{\mathrm{avg}} \geq 1/2 + 2\varepsilon \ .$$

Denote by $F_{x_0,s_0}(y) = F(x_0, y, s_0)$ and by $h_{x_0,s_0}(y) = h(x_0, y, s_0)$. Let ρ be a random restriction to F_{x_0,s_0} that is distributed according to $\mathcal{T}_k$ from Theorem 4.2, and denote by S_ρ the set of variables

unassigned by ρ. Since once a variable is chosen to be restricted, its value is determined randomly, we get that

$$\mathop{\mathbf{E}}_{\rho\sim\mathcal{T}_k}\mathop{\mathbf{Pr}}_{z\sim\{0,1\}^{S_\rho}}[F_{x_0,s_0}|_\rho(z)=h_{x_0,s_0}|_\rho(z)]\geq 1/2+2\varepsilon\ .$$

Let A be the set of restrictions in $\mathcal{T}_k$ that leave exactly k variables unrestricted and that shrinks F_{x_0,s_0} by a factor of $\mathsf{polylog}(n)\cdot\left(\frac{k}{n}\right)^2$. Theorem 4.2 gives $\mathbf{Pr}[\rho\in A]\geq 1-\varepsilon_{\text{shr}}$. Since $X_\rho\triangleq\mathbf{Pr}_{z\sim\{0,1\}^{S_\rho}}[F_{x_0,s_0}|_\rho(z)=h_{x_0,s_0}|_\rho(z)]$ is a random variable whose range is $[0,1]$, we can apply Lemma 2.7 and get that

$$\mathop{\mathbf{E}}_{\rho\sim\mathcal{T}_k}\left[\mathop{\mathbf{Pr}}_{z\sim\{0,1\}^{S_\rho}}[F_{x_0,s_0}|_\rho(z)=h_{x_0,s_0}|_\rho(z)]\ \middle|\ \rho\in A\right]\geq 1/2+2\varepsilon-\varepsilon_{\text{shr}}\geq 1/2+\varepsilon\ .$$

By averaging there must exist $\rho\in A$ such that

$$\mathop{\mathbf{Pr}}_{z\sim\{0,1\}^{S_\rho}}[F_{x_0,s_0}|_\rho(z)=h_{x_0,s_0}|_\rho(z)]\geq 1/2+\varepsilon\ .$$

Recall the definition of $\mathcal{S}$, $\mathcal{H}_{s_0}, n'$ and ε_{app} in Lemma 6.1. The fact that $s_0\in\mathcal{S}$, $x_0\in\mathcal{H}_{s_0}$, $\rho\in\mathfrak{R}_k$ and $\varepsilon\geq\varepsilon_{\text{app}}$ gives

$$L(F_{x_0,s_0}|_\rho)\geq n'=n/\log(n)\ . \tag{7.1}$$

By the definition of the set A,

$$L(F_{x_0,s_0}|_\rho)\leq\left(\frac{k}{n}\right)^2\cdot\mathsf{polylog}(n)\cdot L(F_{x_0,s_0})\ . \tag{7.2}$$

Thus,

$$L(F)\geq L(F_{x_0,s_0})\overset{(7.2)}{\geq}\left(\frac{n}{k}\right)^2\cdot\frac{1}{\mathsf{polylog}(n)}\cdot L(F_{x_0,s_0}|_\rho)\overset{(7.1)}{\geq}\frac{n^3}{k^2\cdot\mathsf{polylog}(n)}=\frac{n^3}{r^2\cdot\mathsf{polylog}(n)}\ ,$$

which completes the proof. □

8 Summary and Open Questions

In this paper we presented a tailor made construction that gives average-case hardness in the spirit of Andreev's function. Specifically, we presented an explicit function $f\colon\{0,1\}^n\to\{0,1\}$ such that any De Morgan formula of size at most $n^{3-o(1)}/r^2$ agrees with f on at most $\frac{1}{2}+2^{-r}$ fraction of the inputs. In particular, for a suitable choice of r, any formula of size $O(n^{2.99})$ agrees with f on at most $\frac{1}{2}+2^{-n^{\Omega(1)}}$ fraction of the inputs.

A natural question is whether this trade-off between the size of the formula and the approximation quality is necessary. More specifically, is there an explicit function $f'\colon\{0,1\}^n\to\{0,1\}$ such that any De Morgan formula of size $n^{3-o(1)}$ agrees with f' on at most $\frac{1}{2}+2^{-n^{\Omega(1)}}$ (or even $\frac{1}{2}+2^{-\Omega(n)}$) fraction of the inputs?

In addition, it is interesting whether there is a black-box reduction from worst-case hardness or even mild hardness (a function which is hard to calculate on 0.9 fraction of the inputs) to average-case hardness with similar guarantees as given here. We note that the standard analysis of the XOR lemma suffers a great loss in the parameters, and only allows to show hardness of computing a function on $\frac{1}{2}+\frac{1}{\mathsf{poly}(n)}$ fraction of the inputs.

Finally, improving Håstad's worst-case lower bound [Hås98] is a long standing open problem and any step towards it would be extremely interesting.

Acknowledgements

We thank Gil Cohen, Raghu Meka, Moni Naor and Eylon Yogev for helpful discussions. We thank the anonymous referees of SICOMP for helpful comments. In particular, we appreciate the referees' suggestion to use $\mathfrak{R}_k$ instead of $\mathcal{R}_p$ in Section 4, which simplifies the presentation.

References

[And87] A. E. Andreev. On a method for obtaining more than quadratic effective lower bounds for the complexity of π-schemes. *Moscow Univ. Math. Bull.*, 42:63–66, 1987. In Russian.

[BBC+01] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.

[Bou07] J. Bourgain. On the construction of affine extractors. *GAFA Geometric And Functional Analysis*, 17(1):33–57, 2007.

[CGH+85] B. Chor, O. Goldreich, J. Håstad, J. Friedman, S. Rudich, and R. Smolensky. The bit extraction problem of t-resilient functions (preliminary version). In *FOCS*, pages 396–407, 1985.

[Che52] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, pages 493–507, 1952.

[CKK+14] R. Chen, V. Kabanets, A. Kolokolova, R. Shaltiel, and D. Zuckerman. Mining circuit lower bound proofs for meta-algorithms. In *CCC*, pages 262–273, 2014.

[GRS06] A. Gabizon, R. Raz, and R. Shaltiel. Deterministic extractors for bit-fixing sources by obtaining an independent seed. *SIAM J. Comput.*, 36(4):1072–1094, 2006.

[Hås98] J. Håstad. The shrinkage exponent of De Morgan formulas is 2. *SIAM J. Comput.*, 27(1):48–64, 1998.

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[IMZ12] R. Impagliazzo, R. Meka, and D. Zuckerman. Pseudorandomness from shrinkage. In *FOCS*, pages 111–119, 2012.

[IN93] R. Impagliazzo and N. Nisan. The effect of random restrictions on formula size. *Random Struct. Algorithms*, 4(2):121–134, 1993.

[Juk12] S. Jukna. *Boolean function complexity: advances and frontiers*, volume 27. Springerverlag Berlin Heidelberg, 2012.

[Khr71] V. M. Khrapchenko. A method of determining lower bounds for the complexity of π schemes. *Matematischi Zametki*, 10:83–92, 1971. In Russian.

[KR13] I. Komargodski and R. Raz. Average-case lower bounds for formula size. In *STOC*, pages 171–180, 2013.

[KRT13] I. Komargodski, R. Raz, and A. Tal. Improved average-case lower bounds for De Morgan formula size. In *FOCS*, pages 588–597. IEEE Computer Society, 2013.

[KRT17] Ilan Komargodski, Ran Raz, and Avishay Tal. Improved average-case lower bounds for de morgan formula size: Matching worst-case lower bound. *SIAM J. Comput.*, 46(1):37–57, 2017.

[KZ07] J. Kamp and D. Zuckerman. Deterministic extractors for bit-fixing sources and exposure-resilient cryptography. *SIAM J. Comput.*, 36(5):1231–1247, 2007.

[MR95] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[PZ93] M. Paterson and U. Zwick. Shrinkage of De Morgan formulae under restriction. *Random Struct. Algorithms*, 4(2):135–150, 1993.

[Rao07] A. Rao. An exposition of bourgain's 2-source extractor. *Electronic Colloquium on Computational Complexity (ECCC)*, 14(034), 2007.

[Rao09] A. Rao. Extractors for low-weight affine sources. In *IEEE Conference on Computational Complexity*, pages 95–101, 2009.

[Rei11] B. Reichardt. Reflections for quantum query algorithms. In *SODA*, pages 560–569, 2011.

[Rud07] A. Rudra. Lecture notes on coding theory, 2007. `http://www.cse.buffalo.edu/~atri/courses/coding-theory/lectures/lect19.pdf`.

[San10] R. Santhanam. Fighting perebor: New and improved algorithms for formula and QBF satisfiability. In *FOCS*, pages 183–192, 2010.

[Sub61] B. A. Subbotovskaya. Realizations of linear function by formulas using $+, \cdot, -$. *Doklady Akademii Nauk SSSR*, 136:3:553–555, 1961. In Russian.

[Tal14] A. Tal. Shrinkage of de Morgan formulae from quantum query complexity. In *FOCS*, pages 551–560, 2014.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il