Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 58 (2013)

Improved Average-Case Lower Bounds for DeMorgan Formula Size

Matching Worst-Case Lower Bound

Ilan Komargodski* Ran Raz* Avishay Tal*

Abstract

We give a function $h : \{0,1\}^n \to \{0,1\}$ such that every deMorgan 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. This improves the previous average-case lower bound of Komargodski and Raz (STOC, 2013).

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), combining ideas of both Impagliazzo, Meka and Zuckerman (FOCS, 2012) and Komargodski and Raz. In addition, using a bit-fixing extractor in the construction of h allows us to simplify a major part of the analysis of Komargodski and Raz.[1]

1 Introduction

Proving lower bounds on the complexity of classical computational models for Boolean functions is a 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 deMorgan formulas. It is well known that the deMorgan 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 deMorgan formula size has been found yet. Providing such a function would separate P from NC^1.

A deMorgan formula is a Boolean formula over the basis $B_2 = \{\vee, \wedge, \neg\}$ with fan in at most 2. A deMorgan formula is represented by a tree such that every leaf is labeled by an input variable and every internal node is labeled by an operation from B_2. A formula is said to compute a function $f : \{0,1\}^n \to \{0,1\}$ 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 deMorgan formula size of a function $f : \{0,1\}^n \to \{0,1\}$ is the size of the minimal deMorgan formula that computes f.

Previous works considered the following two types of lower bounds on deMorgan formula size: *worst-case* lower bounds and *average-case* lower bound.

*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.

[1] We have learnt that the idea to use a bit-fixing extractor in this context was suggested independently in [CKK$^+$13] (private communication with the authors).

ISSN 1433-8092

Worst-case lower bounds are lower bounds on the size of the minimal deMorgan formula that computes an explicit function $f : \{0,1\}^n \to \{0,1\}$. The first lower bound was achieved by Subbotovskaya [Sub61] that proved an $\Omega(n^{1.5})$ lower bound on the size of deMorgan formulas that compute the parity function on n variables. Subbotovskaya also introduced the concept of random restrictions that has had many applications since. In fact, Subbotovskaya showed a lower bound of $\Omega(n^{\Gamma})$ where Γ is referred to as the *shrinkage exponent* of deMorgan 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 deMorgan formula of size $\Omega(n^2)$ (which is tight, up to constant factors). In [And87] Andreev was able to cleverly combine some previous techniques (including the method of [Sub61]) and to prove an $\Omega(n^{1+\Gamma-o(1)})$ lower bound on the size of the minimal deMorgan formula that computes an explicit function, later referred to as the Andreev function. Subsequent improvements on the constant Γ led to improved lower bounds on the deMorgan formula size of the Andreev function. Impagliazzo and Nisan [IN93] proved that $\Gamma \geq 1.55$, Peterson and Zwick [PZ93] proved that $\Gamma \geq 1.63$ and finally Håstad [Hås98] proved that $\Gamma \geq 2 - o(1)$ giving the lower bound of $\Omega(n^{3-o(1)})$ on the size of deMorgan formulas that compute the Andreev function. Since $\Gamma \leq 2$, Håstad's result is tight up to the $o(1)$ term.

Average-case lower bounds (a.k.a., correlation bounds) are lower bounds on the size of the minimal deMorgan formula that only approximates an explicit function $f : \{0,1\}^n \to \{0,1\}$. 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 deMorgan formulas appears in the work of Santhanam [San10]. There, it is shown that any family of linear-size deMorgan 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 deMorgan formula of size $O(n^{1.5})$ and the parity function. In addition, as pointed out in [KR12], works regarding the degree of approximating polynomials also imply correlation bounds for deMorgan 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. Recently, Komargodski and Raz [KR12] constructed an explicit function $f : \{0,1\}^n \to \{0,1\}$ such that any deMorgan formula of size at most $O(n^{2.499})$ computes f correctly on a fraction of at most $\frac{1}{2} + 2^{-n^{\Omega(1)}}$ of the inputs.

In this work, combining techniques from [IMZ12] and [KR12] together with some new ideas, we improve the result of Komargodski and Raz [KR12] and construct an explicit function $h : \{0,1\}^n \to \{0,1\}$ such that any deMorgan formula of size at most $O(n^{2.999})$ 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 is an explicit (computable in polynomial time) Boolean function $h : \{0,1\}^{6n} \to \{0,1\}$ and a constant $c \geq 8$ such that for any parameter r such that $c\log(n) \leq r \leq n^{1/3}$, any formula of size $\frac{n^{3-o(1)}}{r^2}$ 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 $f : \{0,1\}^n \to \{0,1\}$, a vector $\rho \in \{0,1,\star\}^n$ 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. We say that deMorgan formulas have s-shrinkage with probability γ over a distribution $\mathcal{D}$ of restrictions that leave k variables unassigned if any deMorgan formula shrinks by a factor of at least $c \cdot (k/n)^s$ with probability γ over $\mathcal{D}$ for some universal constant c.

Our technical contributions are twofold. First, we prove that 1.999-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), improving a theorem from [KR12]. Second, we simplify a major part of the proof of [KR12] by giving a different construction for the function for which we prove the lower bound. We believe that the insights in this simplification might be of independent interest.

In order to explain our techniques, we first begin by describing the worst-case lower bound of Andreev [And87] and then the average-case lower bound of Komargodski and Raz [KR12].

1.1.1 Andreev's Worst-Case Lower Bound

Andreev's function $\mathrm{A} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is defined as follows. 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. 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 return that bit. The analysis of [And87, IN93, PZ93, Hås98] relies on the following 4 facts:

1. There exists an n bit vector h that represents a Boolean function which is hard to compute by formulas of size $O(n)/\log\log n$.
2. It holds that $L(\mathrm{A}) \geq L(\mathrm{A}_h)$ where A_h is the function A when 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 $f : \{0,1\}^n \to \{0,1\}$ and for a random restriction ρ that leaves k variables unassigned it holds that $L(f|_\rho) \leq c \cdot \left(\frac{k}{n}\right)^\Gamma L(f)$ 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].
4. After applying a completely random restriction that leaves $k = \Theta(\log n \cdot \log\log n)$ variables unrestricted, with probability at least 3/4 every row in the matrix (represented by the second input to 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,

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

where $A_h|_\rho$ denotes the function A_h after applying the restriction ρ.

1.1.2 Komargodski and Raz's Average-Case Lower Bound

Komargodski and Raz's [KR12] function KR : $\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is similar to Andreev's function. KR views the second input as an n^ε by $n^{1-\varepsilon}$ matrix and computes the XOR of the input bits in every row. KR encodes the first n input bits, using an error correcting code, into 2^{n^ε} bits. Finally, KR uses the resulting n^ε bits of the XORs to address an index in the encoded first input and returns that bit. The analysis of [KR12] relies on the following 4 facts (stated informally):

1. Most strings of length n after being encoded, using an error correcting code with large relative distance, into strings of length 2^{n^ε} represent functions (from n^ε bits into 1) that are hard to approximate.

2. If a formula F approximates well KR, 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 KR_h where KR_h (resp. F_h) is the function KR (resp. F) where the first input is fixed to h.

3. 1.499-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 KR have at least one variable that is not restricted with probability exponentially close to 1.

Deriving the lower bound of [KR12] is conceptually similar to Andreev's lower bound, but technically it is a bit more complicated so we refer to [KR12] for additional details.

1.1.3 Our Techniques

In this work we improve the result of [KR12] by improving Item 3 in the proof scheme above. We prove that 1.999-shrinkage occurs with probability exponentially close to 1.

Our second contribution is mainly conceptual. We provide a more intuitive construction of the hard function and greatly simplify Item 4 in the proof of [KR12].

Improvement of Item 3 Komargodski and Raz [KR12] prove a theorem that shows that for deMorgan formulas 1.499-shrinkage occurs with probability exponentially close to 1 over a certain distribution of random restrictions that takes into account the structure of the formula.

Impagliazzo, Meka and Zuckerman [IMZ12] prove a theorem (among other interesting results) that shows 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 $O(s)$ for deMorgan formulas of size $s^{3-o(1)}$ as well as for several other models.

We combine the techniques of [KR12] and of [IMZ12] and prove that 1.999-shrinkage occurs with probability exponentially close to 1. That is, we use the proof technique of [IMZ12] applied to a distribution of random restrictions similar to that of [KR12] 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 (variable that appear a lot more than an average variable) can be split into many "medium size" sub-formulas (subtrees). Thus, if one shows that the total size of all sub-formulas after applying a random restriction is small with high probability we are done. The crucial point is that (conditioning on 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, the sizes of the subformulas after restriction is independent of one another and one can apply Chernoff-Hoeffding inequalities.

The main problem with this argument is the assumption that there are no "heavy" variables. [IMZ12] treat the "heavy" variables separately, showing overall that shrinkage occurs with probability $1 - 1/\text{poly}(n)$. 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.

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* 1.999*-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 4.1). 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 framework of [KR12], 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 such that at each step a variable is chosen (possibly depending on the structure of the formula) and then the value of the chosen variable is randomly fixed to 0 *or* 1*. We refer to such a distribution as a distribution of* random valued restrictions*.*

In Section 4 we define a distribution that has this property, and prove that 1.999*-shrinkage occurs with probability exponentially close to* 1 *over this distribution. Moreover, for possible future applications, we note that the process that defines our distribution can be efficiently implemented.*

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 harder in order to show Item 4 in the proof scheme above. [KR12] overcame this problem by a series of reductions to balls and bins games, heavily relying on the specific distribution of restrictions defined in Item 3. In this work, we view the restrictions distribution as a black-box, only ensuring that the number of variables left unrestricted is $k = 100n^{\varepsilon}$ with high probability (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, which is 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 than bit-fixing extractors that do not assume access to an advice [KZ07, GRS06, Rao09].

We think 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 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.

1.2 Related Work

Recently, Chen et al. [CKK+13][2] addressed the following problem (a.k.a, compression for "easy" Boolean functions): given the truth table of a Boolean function $f : \{0,1\}^n \to \{0,1\}$ that can be computed by a small unknown circuit from a given class $\mathcal{C}$, construct an explicit Boolean circuit (not necessarily in $\mathcal{C}$) that computes f and is of size $o(2^n/n)$.

Their results,[3] that rely on "shrinkage with high probability", are the following: (1) any Boolean n-variate function computable by a deMorgan formula of size at most $n^{2.49}$ is compressible in time $\text{poly}(2^n)$ to a circuit of size at most $2^{n-n^{\varepsilon}}$, for some $\varepsilon > 0$ and (2) there exists a deterministic #SAT-algorithm for n-variate deMorgan formulas of size at most $n^{2.49}$ that runs in time $2^{n-n^{\varepsilon}}$, for some $\varepsilon > 0$. Our shrinkage result (see Section 4) improves both of these results to hold for deMorgan formulas of size at most $n^{2.99}$. However, the resulting #SAT algorithm is zero-error randomized rather than deterministic.

Moreover, independent of our work, Chen et al. [CKK+13] simplify the average-case lower bound of [KR12] using a bit-fixing extractor. This is quite similar to some of our techniques.

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 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 represents 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 will only consider deMorgan 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 will 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 two functions $f : \{0,1\}^s \to \{0,1\}^m$

[2] Private communication with the authors. A preliminary version can be found in [KK13].

[3] [CKK+13] give results for several computational models such as branching programs, formulas over any complete basis and more. We only focus on their results regarding deMorgan formulas.

and $g : \{0,1\}^n \to \{0,1\}^s$, we denote by $f \circ g : \{0,1\}^n \to \{0,1\}^m$ the composition of f and g, i.e., $f \circ g(x) = f(g(x))$.

Boolean Formulas

Definition 2.1. *A deMorgan formula is a Boolean formula with AND, OR and NOT gates with fan in at most 2.*

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

Definition 2.3 (Restriction)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. A restriction ρ is a vector of length n of elements from $\{0,1,\star\}$. We denote by $f|_\rho$ the function f restricted according to ρ in the following sense: if $\rho_i = \star$ 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 $\rho_i = \star$ and with probability $\frac{1-p}{2}$ set ρ_i to be 0 and 1, respectively. We denote this distribution of restrictions by $\mathcal{R}_p$.*

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

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

Probability

We state a well known variant of Chernoff/Hoeffding inequality.

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

$$\Pr[X - \mathbb{E}[X] \ge t] \le \exp\left(\frac{-2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

If we assume further that $X_1, \dots, X_n \in \{0,1\}$ are identically distributed then for $\delta \in (0,1)$,

$$\Pr[X < (1-\delta) \cdot \mathbb{E}[X]] \le \exp\left(-\delta^2 \cdot \mathbb{E}[X]/2\right)$$

and

$$\Pr[X > (1+\delta) \cdot \mathbb{E}[X]] \le \exp\left(-\delta^2 \cdot \mathbb{E}[X]/3\right).$$

We will use this 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 $\Pr[B] > 0$, then $\mathbb{E}[X|B] \geq \mathbb{E}[X] - \Pr[\neg B]$. In particular if X is an indicator of an event A, then $\Pr[A|B] \geq \Pr[A] - \Pr[\neg B]$*

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

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

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*

$$|P - Q| = \max_{A \subseteq \Omega} \left| \Pr_P(A) - \Pr_Q(A) \right|$$

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 $a_1, \ldots, a_k \in \{0,1\}$*

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

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 distribution $(X_{i_1}, \ldots, X_{i_k})$ 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 a (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 distributed uniformly over $v_0 + \operatorname{span}\{v_1, \ldots, v_k\}$. 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) with error ε and output length r is a function $\mathrm{Ext} : \{0,1\}^n \to \{0,1\}^r$ such that for every (n,k)-bit-fixing source ((n,k)-affine source) the distribution of $\mathrm{Ext}(X)$ is ε-close to the uniform distribution in statistical distance, i.e.,*

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

Coding Theory

Definition 2.13. *A code C over an alphabet Σ of size q that has block length n, dimension k and minimal distance d is denoted as an $(n,k,d)_q$ code. 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 is sometimes referred to as the encoding function of C. The relative distance of C is $\delta = d/n$.*

Furthermore, 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 $\mathbb{F}_q$.

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

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

where Δ denotes the Hamming distance.

Next, 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 $\delta = d/n \le 1/2$. It holds that C is $(\rho, 2dn)$-list decodable for any*

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

3 Construction of the Function

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

We define a function $h : \{0,1\}^{4n} \times \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ 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 similar to the one in [KR12]. $\mathcal{C}$ encodes $x \in \{0,1\}^{4n}$ to $\mathrm{Enc}_x^{\mathcal{C}} \in \{0,1\}^{2^r}$ and has relative distance $\delta = \frac{d}{2^r} \ge 1/2 - 2^{-r/4}$. One may view each codeword also as a Boolean function $\mathrm{Enc}_x^{\mathcal{C}} : \{0,1\}^r \to \{0,1\}$. The exact definition and construction of the code $\mathcal{C}$ are described in Appendix B.

Our bit-fixing extractor is a function $\mathrm{BFExt}_s : \{0,1\}^n \to \{0,1\}^r$ 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

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

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

4 Shrinkage with Very High Probability

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

Theorem 4.1. *Let $c > 0$ be any constant and let F be a formula over n variables of size $\leq n^c$, for any n large enough. Then 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 *(see Remark 1.2) such that*

$$\Pr_{\rho \in \mathcal{T}_k}\left[L(F|_\rho) \leq 2^{O\left(\log^2 \log n\right)} \cdot \left(\frac{k}{n}\right)^2 \cdot L(F) \ \ \textbf{and} \ \ \rho \in \mathfrak{R}_k\right] \geq 1 - \varepsilon_{shr}$$

where $\varepsilon_{shr} = 2^{-\Omega(k)}$ (recall that $\mathfrak{R}_k$ is the set of restrictions leaving k variables unassigned).

Our proof is based on the result of Håstad [Hås98] that showed that shrinkage of deMorgan formulas occurs in expectation.

Theorem 4.2 ([Hås98])**.** *Let F be a deMorgan formula. For every $p > 0$ it holds that*

$$\mathop{\mathbb{E}}_{\rho \in_R \mathcal{R}_p}[L(F|_\rho)] \leq O\left(p^2\left(1 + \log^{3/2}\min\left\{\frac{1}{p}, L(F)\right\}\right)L(F) + p\sqrt{L(F)}\right).$$

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 (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, we define a "random" restriction algorithm with parameter p that takes the structure of F into account.[4] This algorithm defines a distribution of *random valued restrictions* that we denote by $\mathcal{T}'_p$.

```
$F_0 \leftarrow F$.
$i \leftarrow 0$.
while $n - i > pn$ AND there is a variable $x_j$ in $F_i$ that appears more than $t_i = \frac{2L(F_i)}{n-i}$ 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 $\rho' \in \mathcal{R}_{\frac{p \cdot n}{n-i}}$ and restrict the formula $F_i$ according to $\rho'$.
```

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

[4] For simplicity, we assume throughout this section that $p \cdot n$ is an integer.

First, we argue that shrinkage occurs with very high probability for formulas that do not contain any "heavy" variable. The proof of the following lemma can be found in Appendix A.

Lemma 4.3. *There exists a universal constant $c > 0$ such that for any formula F over n variables that does not contain any variable that appears more than $2L(F)/n$ times and for any $0 < p \leq 1$*

$$\Pr_{\rho \in \mathcal{R}_p} \left[L\left(F|_{\rho}\right) \geq c \cdot p^2 \log^{3/2}(n) \cdot L(F) \right] \leq L(F) \cdot e^{-n \cdot p^8}.$$

A corollary of Lemma 4.3 is that shrinkage also occurs under $\mathcal{T}'_p$ as follows.

Corollary 4.4. *Let c be the constant from Lemma 4.3. Let F be formula over n variables. For any $0 < p \leq 1$ it holds that*

$$\Pr_{\rho \in \mathcal{T}'_p} \left[L(F|_{\rho}) \geq c \cdot p^2 \log^{3/2}(n) \cdot L(F) \right] \leq L(F) \cdot e^{-n \cdot p^8}.$$

Proof. Assume that we have $h \geq 0$ heavy variables that cause $\rho \in \mathcal{T}'_p$ 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

$$\begin{aligned} L\left(F'\right) &\leq L(F) \left(1 - \frac{1}{n}\right)^2 \left(1 - \frac{1}{n-1}\right)^2 \cdots \left(1 - \frac{1}{n-h+1}\right)^2 \\ &= L(F) \left(\frac{n-h}{n}\right)^2 . \end{aligned} \tag{4.1}$$

Apply Lemma 4.3 on the formula F' that contains $n_l = n - h$ variables with $p' = \frac{pn}{n_l}$. Since

$$L(F') \cdot e^{-n_l \cdot (p')^8} = L(F') \cdot e^{-(np)^8/{n_l}^7} \leq L(F) \cdot e^{-n \cdot p^8} ,$$

it follows that

$$\Pr_{\rho' \in \mathcal{R}_{p'}} \left[L(F'|_{\rho'}) \geq c \cdot (p')^2 \log^{3/2}(n_l)\, L(F') \right] \leq L(F) \cdot e^{-n \cdot p^8}.$$

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

$$\begin{aligned} L(F) \cdot e^{-n \cdot p^8} &\geq \Pr_{\rho' \in \mathcal{R}_{p'}} \left[L(F'|_{\rho'}) \geq c \cdot (p')^2 \cdot \log^{3/2}(n_l) \cdot L(F') \right] \\ &\geq \Pr_{\rho' \in \mathcal{R}_{p'}} \left[L(F'|_{\rho'}) \geq c \cdot (p')^2 \cdot \log^{3/2}(n_l) \cdot L(F) \left(\frac{n_l}{n}\right)^2 \right] && \text{(eq. (4.1))} \\ &\geq \Pr_{\rho \in \mathcal{T}'_p} \left[L(F|_{\rho}) \geq c \cdot (p')^2 \cdot \log^{3/2}(n_l) \cdot L(F) \left(\frac{n_l}{n}\right)^2 \right] && (L(F'|_{\rho'}) = L(F|_{\rho})) \\ &\geq \Pr_{\rho \in \mathcal{T}'_p} \left[L(F|_{\rho}) \geq c \cdot p^2 \cdot \log^{3/2}(n) \cdot L(F) \right] && (p' = pn/n_l, n_l \leq n) \end{aligned}$$

which concludes the proof of the corollary. □

Remark 4.5. *Note that Corollary 4.4 is useful only for $p > n^{-1/8}$. This range of p'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 p. This is what we achieve in Theorem 4.1.*

Next, we prove the main theorem of this section (Theorem 4.1).

Proof of Theorem 4.1. We apply Corollary 4.4 $t \geq 1$ times where t will be determined later. Let $F_0 = F$ and for $1 \leq i \leq t$ let F_i be the formula after the i-th application of Corollary 4.4. If after t iterations we are left with more than k variables unrestricted,[5] we further randomly restrict variables until we are left with exactly k variables. We denote the resulting formula by F'. Set $n_0 = n$ and for every $1 \leq i \leq t$ think of n_i as a lower bound on the number of variables in F_i with high probability. For $0 \leq i \leq t-1$ denote by p_i the value of p used in the $(i+1)$-th application of Corollary 4.4. We state the following claim that suggests the existence of good parameters n_i, p_i, and defer its proof for later.

Claim 4.6. *There are parameters $t \in \mathbb{N}$, $p_0, \ldots, p_{t-1} \in \mathbb{R}$, $n_0, \ldots, n_t \in \mathbb{N}$ such that*

1. $n_0 = n$
2. $n_t = k$
3. *for* $0 \leq i \leq t-1$, $p_i \cdot n_i = 2 \cdot n_{i+1}$
4. *for* $0 \leq i \leq t-1$, $p_i^8 \cdot n_i = \Omega(k)$
5. *for* $0 \leq i \leq t-1$, $0 < p_i \leq 1$
6. $t = O(\log \log n)$

For $i = 1, \ldots, t$ let $\mathcal{D}_i$ be the event that after iteration i, the number of variables left unrestricted is at least n_i, and let $\mathcal{E}_i$ be the event that $L(F_i) \leq c \cdot \log^{3/2}(n) \cdot p_{i-1}^2 \cdot L(F_{i-1})$. Using Item 3 of Claim 4.6 and Chernoff's bound (Proposition 2.6) gives that for $i \in [t]$

$$\Pr\left[\mathcal{D}_i | \mathcal{D}_1, \ldots, \mathcal{D}_{i-1}\right] > 1 - \exp\left(-\Omega(n_{i-1} \cdot p_{i-1})\right) = 1 - \exp\left(-\Omega(n_i)\right) .$$

Using Item 4 of Claim 4.6 and Corollary 4.4 gives

$$\Pr\left[\mathcal{E}_i | \mathcal{D}_1, \ldots, \mathcal{D}_{i-1}\right] > 1 - L(F) \cdot \exp\left(-\Omega(n_{i-1} \cdot p_{i-1}^8)\right) = 1 - L(F) \cdot \exp\left(-\Omega(k)\right)$$

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

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

$$L(F') \leq L(F_t) \leq c^t \cdot \left(\prod_{i=0}^{t-1} p_i^2\right) \cdot \left(\log^{3/2}(n)\right)^t \cdot L(F_0) .$$

[5] We count variables which are unrestricted by our algorithm even if they do not appear in the restricted formula.

Using Item 3 of Claim 4.6, $\prod_{i=0}^{t-1} p_i^2 = \prod_{i=0}^{t-1} \frac{4n_{i+1}^2}{n_i^2} = 4^t \cdot \frac{k^2}{n^2}$. Using Item 6 of Claim 4.6 ($t = O(\log\log n)$) it follows that

$$L(F') \leq 2^{O\left(\log^2 \log n\right)} \cdot \frac{k^2}{n^2} \cdot L(F) \ .$$

Under the assumption that $\mathcal{D}_t$ holds, the number of unrestricted variables by the process is exactly k, which completes the proof. □

Proof of Claim 4.6. Let $\{x_i\}_{i\in\mathbb{N}}$ and $\{q_i\}_{i\in\mathbb{N}}$ be infinite sequences of real numbers defined as $x_0 = k$, $q_0 = 1/2$, and for $i \geq 1$, $q_i = (q_{i-1})^{8/7}$ and $x_i = 2 \cdot x_{i-1}/q_i$. We have

$$x_i \cdot q_i^8 = 2 \cdot x_{i-1} \cdot q_i^7 = 2 \cdot x_{i-1} \cdot q_{i-1}^8 = \ldots = 2^i \cdot x_0 \cdot q_0^8 = \Omega(k) \ . \tag{4.2}$$

Notice that the sequence $\{x_i\}_{i\in\mathbb{N}}$ is monotone increasing, and always greater or equal to k. The sequence $\{q_i\}_{i\in\mathbb{N}}$ is monotone decreasing and always in $(0, 1)$. For $i \geq 1$ we have

$$x_i = \frac{2x_{i-1}}{q_i} \geq \frac{2k}{q_i} \geq \frac{1}{q_i} = 2^{(8/7)^i} \ . \tag{4.3}$$

Let t be the least such that $x_t \geq n/2 - 1$. Using eq. (4.3) gives $t = O(\log\log n)$.

We are ready to define our n_is and p_is. Set $n_0 := n$ and for $i = 1, \ldots, t$, set $n_i := \lceil x_{t-i} \rceil$. For $i = 0, \ldots, t-1$, set $p_i := 2 \cdot n_{i+1}/n_i$.

Requirements 1-3 hold by our choice of p_i and n_i. $p_0 < 1$ since

$$p_0 = \frac{2n_1}{n_0} \leq \frac{2 \cdot (x_{t-1} + 1)}{n} < \frac{2 \cdot (n/2 - 1 + 1)}{n} = 1 \ ,$$

using the assumption that $x_{t-1} < n/2 - 1$. For $i = 1, \ldots, t-1$,

$$p_i = 2 \cdot n_{i+1}/n_i = 2 \cdot \lceil x_{t-i-1} \rceil / \lceil x_{t-i} \rceil = \Theta(q_{t-i}) \tag{4.4}$$

and one can verify that p_i is smaller than 1, hence Requirement 5 holds. Requirement 4 holds since $p_i^8 \cdot n_i \overset{(4.4)}{=} \Omega(q_{t-i}^8 \cdot x_{t-i}) \overset{(4.2)}{=} \Omega(k)$. Requirement 6 holds by the upper bound given on t, 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 $\mathrm{BFExt} : \{0,1\}^n \to \{0,1\}^r$ such that for every (n, k)-bit-fixing source, X, the output $\mathrm{BFExt}(X)$ 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* $\mathrm{BFExt} : \{0,1\}^n \to \{0,1\}^r$ *that is an (n, k)-bit-fixing extractor with output length $r = k - o(k)$ and error 2^{-k^d}*

We will show 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. None the less, 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_{in}, k_{out}) linear condenser is a linear mapping $T : \{0,1\}^n \to \{0,1\}^m$ such that for any $S \subseteq [n]$ of size $\geq k_{in}$ we have*

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

The output of an (n, m, k_{in}, k_{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., 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.

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 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 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

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

such matrices. Thus, the probability that a random $k \times k$ matrix has rank $\leq d$ is at most $2^{k+2dk-d^2-k^2} = 2^{k-(k-d)^2}$. 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

$$\binom{n}{k} \cdot 2^{k-(k-d)^2} \leq \left(\frac{en}{k}\right)^k \cdot 2^{k-(k-d)^2} = 2^{k \cdot \log(2en/k)-(k-d)^2} .$$

For $k \geq 2e$ and $d \leq k - \sqrt{2k \cdot \log n}$, this probability is at most $2^{k \cdot \log n - (k-d)^2} \leq 2^{-k \cdot \log n}$ 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 values of the $k \times n$ matrix are sampled from a k^2-wise independent distribution (Definition 2.9). Formally, we get the following immediate corollary.

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).*

There exists a family of efficiently computable functions $\{\text{BFExt}_s : \{0,1\}^n \to \{0,1\}^r\}_{s \in \{0,1\}^n}$ 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*

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

Proof. We will 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 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 n large enough. 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 gives an $(n, k, k, k - \sqrt{k \cdot 2 \cdot \log n})$ linear condenser. We will 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 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

In this section we prove two lemmas that are analogous to Theorems 5.1 and 5.2 in [KR12]. Since those theorems in [KR12] were tailored to their construction of the hard function, we can not apply them.

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 $\mathrm{Enc}_x^{\mathcal{C}} \circ (\mathrm{BFExt}_s|_\rho)$ is hard to approximate for any restriction ρ leaving k inputs unassigned.

Lemma 6.1. *Let $\{\mathrm{BFExt}_s : \{0,1\}^n \to \{0,1\}^r\}_{s \in \{0,1\}^n}$ be the family of functions as in Theorem 5.6. Recall that $\mathcal{S} = \{s \in \{0,1\}^n : \mathrm{BFExt}_s$ is an (n,k) bit-fixing extractor with error $2^{-r}\}$.*

Let $n' = n/\log n$, $\varepsilon_{app} = 2^{-r/10}$. For any seed $s \in \mathcal{S}$ denote by

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

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

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

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

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

[6] We can ignore extra bits, if needed.

Proof. Notice that

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

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

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

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

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 $g : \{0,1\}^k \to \{0,1\}^r$ be a function such that $|\mathbf{U}_r - g(\mathbf{U}_k)| < \varepsilon$. Let $\mathcal{F} \subseteq \{f : \{0,1\}^r \to \{0,1\}\}$ 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 $f_1, f_2 : \{0,1\}^r \to \{0,1\}$ 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 (Definition 2.8)

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

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 will upper bound the size of the following set

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

[7] More fomrally, for a function $f : \{0,1\}^n \to \{0,1\}$ we denote by $\mathrm{tt}(f) \in \{0,1\}^{2^n}$ the string which represent the truth-table of f, i.e., $\mathrm{tt}(f) = f(x_0)f(x_1)\ldots f(x_{2^n})$ where $x_i \in \{0,1\}^n$ is the i-th string in lexicographical order of length n.

We will then use a union bound over all $\rho \in \mathfrak{R}_k$

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

By definition of $\mathcal{S}$, $\mathrm{BFExt}_s|_\rho$ is a function $\{0,1\}^k \to \{0,1\}^r$ such that the statistical distance between $\mathbf{U}_r$ and $\mathrm{BFExt}_s|_\rho(\mathbf{U}_k)$ 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 $\mathrm{Enc}^{\mathcal{C}}_{x_1} \circ (\mathrm{BFExt}_s|_\rho)$ and $\mathrm{Enc}^{\mathcal{C}}_{x_2} \circ (\mathrm{BFExt}_s|_\rho)$ is at least $\delta - \varepsilon \geq 1/2 - 2^{-r/4} - 2^{-r}$.

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

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

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

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 eq. (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 is an explicit (computable in polynomial time) Boolean function $h : \{0,1\}^{6n} \to \{0,1\}$ and a constant $c \geq 8$ such that for any parameter r such that $c \log(n) \leq r \leq n^{1/3}$, any formula of size $\frac{n^{3-o(1)}}{r^2}$ 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 $1/\lambda_{1/2} \cdot r$ where $\lambda_{1/2}$ is some universal constant (see Theorem 5.6). Let

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

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

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

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} which is distributed according to $\mathcal{T}_k$ from Theorem 4.1, 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:

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

Let A be the set of restrictions in $\mathcal{T}_k$ that leave exactly k variables unrestricted and that make F_{x_0,s_0} shrink by a factor of $2^{O(\log^2 \log n)} \cdot \left(\frac{k}{n}\right)^2$. Theorem 4.1 gives $\Pr[\rho \in A] \geq 1 - \varepsilon_{shr}$. Since $X_\rho \triangleq \Pr_{z \in \{0,1\}^{S_\rho}} [F_{x_0,s_0}|_\rho(z) = h_{x_0,s_0}|_\rho(z)]$ is a random variable whose range is $[0,1]$, we can apply Lemma 2.7

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

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

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

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

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

By the definition of A

$$L(F_{x_0,s_0}|_\rho) \leq \left(\frac{k}{n}\right)^2 \cdot n^{o(1)} \cdot L(F_{x_0,s_0}) \tag{7.2}$$

Thus,

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

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 $f : \{0,1\}^n \to \{0,1\}$ such that any deMorgan 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 $O(n^{2.999})$ 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 $f' : \{0,1\}^n \to \{0,1\}$ such that any deMorgan 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 using 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 $\frac{1}{2} + \frac{1}{\mathrm{poly}(n)}$ 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.

9 Acknowledgments

We thank Gil Cohen, Moni Naor and Eylon Yogev for helpful discussions. We thank Raghu Meka for pointing out [HS70].

References

[And87] Alexander 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] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.

[Bou07] Jean Bourgain. On the construction of affine extractors. *Geometric and functional analysis*, 17(1):33–57, 2007.

[CGH$^+$85] Benny Chor, Oded Goldreich, Johan Håstad, Joel Friedman, Steven Rudich, and Roman Smolensky. The bit extraction problem of t-resilient functions (preliminary version). In *FOCS*, pages 396–407. IEEE Computer Society, 1985.

[CKK$^+$13] Ruiwen Chen, Valentine Kabanets, Antonina Kolokolova, Ronen Shaltiel, and David Zuckerman. Mining circuit lower bound proofs for meta-algorithms. *Manuscript*, 2013.

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

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

[HS70] András Hajnal and Endre Szemerédi. Proof of a conjecture of Erdős. *Combinatorial theory and its applications*, 2:601–623, 1970.

[IMZ12] Russell Impagliazzo, Raghu Meka, and David Zuckerman. Pseudorandomness from shrinkage. In *FOCS*, pages 111–119. IEEE Computer Society, 2012.

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

[Juk12] Stasys Jukna. *Boolean Function Complexity: Advances and Frontiers.* Springer 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.

[KK13] Valentine Kabanets and Antonina Kolokolova. Compression of boolean functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:24, 2013.

[KR12] Ilan Komargodski and Ran Raz. Average-case lower bounds for formula size. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:62, 2012. To appear in *STOC* 2013.

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

[PZ93] Mike Paterson and Uri Zwick. Shrinkage of de morgan formulae under restriction. *Random Struct. Algorithms*, 4(2):135–150, 1993.

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

[Rei11] Ben Reichardt. Reflections for quantum query algorithms. In Dana Randall, editor, *SODA*, pages 560–569. SIAM, 2011.

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

[San10] Rahul Santhanam. Fighting perebor: New and improved algorithms for formula and qbf satisfiability. In *FOCS*, pages 183–192. IEEE Computer Society, 2010.

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

A Proof of Lemma 4.3

In this section we prove Lemma 4.3. We note that the proof is highly based on ideas from [IMZ12]. We begin by restating the lemma.

Lemma A.1 (Restating Lemma 4.3)**.** *There exists a universal constant $c > 0$ such that for any formula F over n variables that does not contain any variable that appears more than $2L(F)/n$ times and for any $0 < p \leq 1$*

$$\Pr_{\rho \in \mathcal{R}_p}\left[L\left(F|_\rho\right) \geq c \cdot p^2 \log^{3/2}(n) \cdot L(F)\right] \leq L(F) \cdot e^{-n \cdot p^8}.$$

Recall Theorem 4.2. For ease of notation, we let

$$\mu(p, X) = p^2\left(1 + \log^{3/2} \min\left\{\frac{1}{p}, X\right\}\right) X + p\sqrt{X}.$$

We begin the proof with some technical claims that are needed to prove Lemma 4.3. The first theorem states that every graph with bounded degree can be partitioned into large independents sets. It was conjectured by Erdős, and proved by Hajnal and Szemerédi in [HS70].

Theorem A.2 (Equitable Coloring)**.** *The vertices of every graph $G = (V, E)$ with maximum degree d can be partitioned into $d + 1$ independent sets, such that the size of each independent set is at least $\left\lfloor \frac{|V|}{d+1} \right\rfloor$.*

The next lemma states that a small set of variables cannot affect (increase or decrease) the size of the formula too much.

Lemma A.3 ([IMZ12])**.** *Let F be a deMorgan formula on a set of n variables. Denote by $R \subseteq [n]$ a subset of coordinates of input variables. For $r \in \{0, 1\}^R$, denote by ρ_r the restriction formed by setting variables in R to r, leaving all other variables unassigned. Then,*

$$L(F) \leq \sum_{r \in \{0,1\}^R} \left(L\left(F|_{\rho_r}\right) + |R| \right) \leq 2^{|R|} \left(\max_{r \in \{0,1\}^R} L\left(F|_{\rho_r}\right) + |R| \right).$$

The next lemma states that every formula can be decomposed into smaller formulas with some overhead.

Lemma A.4 ([IMZ12])**.** *Let F be a deMorgan formula on a set of variables X such that $L(F) \geq \ell > 10$. There exist m deMorgan formulas $G_1, \ldots, G_m$ where $\frac{L(F)}{\ell} \leq m \leq \frac{6L(F)}{\ell}$ such that*

- *For $1 \leq i \leq m$ it holds that $L(G_i) \leq \ell$.*
- *For $1 \leq i \leq m$ it holds that G_i may depend on at most 2 "special" variables outside of X (that is, variables that F does not depend on).*
- *For any restriction $\rho \in \{0, 1, \star\}^X$, $L\left(F|_\rho\right) \leq \sum_{i=1}^m L\left(G_i|_{\rho'}\right)$ where $\rho'(x) = \rho(x)$ for $x \in X$ and $\rho'(x) = \star$ otherwise.*

The proofs of Lemmas A.3 and A.4 can be found in [IMZ12].

The last lemma that we give before the proof of Lemma 4.3 states that for any formula F that has a few special variables that are not allowed to be restricted, shrinkage in expectation still holds.

Lemma A.5. *Let F be a deMorgan formula that depends on n variables. Let S be a constant size subset of variables that are "special" in the sense that they are not allowed to be restricted. Denote by $\mathcal{R}_p^S$ a distribution of restrictions such that sampling $\rho' \in \mathcal{R}_p^S$ is done as follows: for every $x \in S$ set $\rho'(x) = \star$, while for every $x \notin S$, independently with probability p set $\rho'(x) = \star$ and with probability $\frac{1-p}{2}$ set $\rho'(x)$ to be 0 and 1, respectively. It holds that for $0 < p < 1$*

$$\mathop{\mathbb{E}}_{\rho' \in_R \mathcal{R}_p^S} \left[L(F|_{\rho'}) \right] \leq O\left(\mu(p, L(F)) + 1\right).$$

Proof. Recall that $|S|$ is constant. Let $\rho' \in \mathcal{R}_p^S$. For $r \in \{0, 1\}^S$ let ρ_r be the restriction induced by the vector r (as in Lemma A.3). Notice that $L((F|_{\rho'})|_{\rho_r}) = L((F|_{\rho_r})|_{\rho'})$ since ρ_r and ρ' restrict disjoint sets of input variables . Applying Lemma A.3 on $L(F|_{\rho'})$ we get

$$L(F|_{\rho'}) \leq \sum_{r \in \{0,1\}^S} L((F|_{\rho'})|_{\rho_r}) + O(1) = \sum_{r \in \{0,1\}^S} L((F|_{\rho_r})|_{\rho'}) + O(1)$$

now taking expectation over both sides of the inequality we get that

$$\begin{aligned}\mathop{\mathbb{E}}_{\rho' \in_R \mathcal{R}_p^S}[L(F|_{\rho'})] &\leq \mathop{\mathbb{E}}_{\rho' \in_R \mathcal{R}_p^S}\left[\sum_{r \in \{0,1\}^S} L((F|_{\rho_r})|_{\rho'}) + O(1)\right] \\ &= \mathop{\mathbb{E}}_{\rho \in_R \mathcal{R}_p}\left[\sum_{r \in \{0,1\}^S} L((F|_{\rho_r})|_{\rho}) + O(1)\right]\end{aligned}$$

where the last equality holds since $F|_{\rho_r}$ has no special variables. Using linearity of expectation and applying Theorem 4.2 with the formula $F|_{\rho_r}$ the lemma follows. □

At this point, we are ready to prove Lemma 4.3.

Proof of Lemma 4.3. The claim is vacuous for $p < n^{-1/8}$ 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 $p \geq n^{-1/8}$. Set $\ell := 1/p^2$.

Recall that the formula F does not contain "heavy" variables. Decompose the formula F into $G_1, \ldots, G_m$ as in Lemma A.4. Let $\rho \in \mathcal{R}_p$ be a random restriction. Form a graph whose vertices are the G_i's and there is an edge between G_i and G_j for $i \neq j$ if the formulas share a variable that is not special (that is, there is some x_k that appears in both). This graph has m vertices where $\frac{L(F)}{\ell} \leq m \leq \frac{6L(F)}{\ell}$ (see Lemma A.4) and since F does not contain "heavy" variables, the degree of this graph is at most $d \triangleq \ell \cdot \left(\frac{2L(F)}{n} - 1\right)$. Using Theorem A.2, it follows that this graph can be divided into z independent sets $\{I_1, \ldots, I_z\}$, each of size at least $\left\lfloor \frac{m}{d+1} \right\rfloor \geq \frac{n}{3\ell^2}$.

Denote by S the set of all "special" variables in $G_1, \ldots, G_m$. Denote by $\mathcal{R}_p^S$ the distribution $\mathcal{R}_p$ with the requirement that variables from S are always not restricted (as in Lemma A.5). Let ρ' be a restriction such that $\rho'(x) = \rho(x)$ for $x \notin S$ and $\rho'(x) = \star$ for $x \in S$. Hence, ρ' is distributed according to $\mathcal{R}_p^S$. Let $\{X_i \in \mathbb{N}\}_{i \in [m]}$ be a set of random variables such that $X_i = L(G_i|_{\rho'})$. Intuitively, partitioning the formula into independent components, enables us to apply Hoeffding's inequalities (Proposition 2.6) to show that shrinkage occurs with high probability for the formula induced by the vertices inside every such independent set.

Recall that every formula G_i contains at most 2 special variables that are not allowed to be restricted. From Lemma A.5, we get that shrinkage in expectation also holds for every such G_i, suffering from an additional constant factor inside the O. Formally, following the notation of Lemma A.5, for every $i \in [m]$ it holds that there exists a constant c'' such that

$$\begin{aligned}\mathop{\mathbb{E}}_{\rho' \in \mathcal{R}_p^S}[L\left(G_i|_{\rho'}\right)] &\leq c'' \cdot \left(\mu\left(p, L(G_i)\right) + 1\right) \\ &\leq c'' \cdot \left(\mu\left(p, \ell\right) + 1\right) \qquad \text{(A.1)}\end{aligned}$$

where the last inequality follows since μ is monotone increasing with respect to the second argument. Since $\ell = 1/p^2$ we have that $1 \leq \mu(p, \ell)$, so $\mu(p, \ell) + 1 \leq 2\mu(p, \ell)$. Plugging this into eq. (A.1) we get that there exists a constant $c' \geq 2$ such that

$$\mathop{\mathbb{E}}_{\rho' \in \mathcal{R}_p^S}[L\left(G_i|_{\rho'}\right)] \leq c' \cdot \mu\left(p, \ell\right).$$

Using Hoeffding's inequality (Proposition 2.6), we get that for every independent set I_i

$$\begin{aligned}
&\Pr_{\rho'\in\mathcal{R}_p^S}\left[\sum_{j\in I_i} X_j \geq 2\cdot|I_i|\cdot c'\cdot\mu(p,\ell)\right] \\
&\leq \Pr_{\rho'\in\mathcal{R}_p^S}\left[\sum_{j\in I_i} X_j - \mathbb{E}\left[\sum_{j\in I_i} X_j\right] \geq |I_i|\cdot c'\cdot\mu(p,\ell)\right] && (E[X_j]\leq c'\cdot\mu(p,\ell)) \\
&\leq \exp\left(\frac{-2\left(|I_i|\cdot c'\cdot\mu(p,\ell)\right)^2}{|I_i|\cdot\ell^2}\right) && (\text{Prop. 2.6, } 0\leq X_j\leq\ell) \\
&\leq \exp\left(\frac{-2c'^2\cdot\mu(p,\ell)^2\cdot n}{3\ell^4}\right) && (|I_i|\geq n/3\ell^2) \\
&\leq \exp\left(-n/\ell^4\right) && (\mu(p,\ell)\geq 1,\ 2c'^2\geq 3) \\
&\leq \exp\left(-n\cdot p^8\right) && (\ell = 1/p^2)
\end{aligned}$$

Next we apply a union bound on the different independent sets. Using the crude upper bound that the number of independent sets is $\leq L(F)$, we get that

$$\Pr_{\rho'\in\mathcal{R}_p^S}\left[\sum_{i=1}^{z}\sum_{j\in I_i} X_j \geq 2\cdot c'\cdot\mu(p,\ell)\cdot\sum_{i=1}^{z}|I_i|\right] \leq L(F)\cdot e^{-n\cdot p^8}\ . \tag{A.2}$$

The choice of ℓ and the assumption $p\geq n^{-1/8}$ gives $\mu(p,\ell)=O(\log^{3/2}(n))$. Hence,

$$\begin{aligned}
2\cdot c'\cdot\mu(p,\ell)\cdot\sum_{i=1}^{z}|I_i| &= 2\cdot c'\cdot\mu(p,\ell)\cdot m \\
&= O\left(\log^{3/2}(n)\cdot L(F)/\ell\right) \\
&= O\left(p^2\cdot\log^{3/2}(n)\cdot L(F)\right)\ .
\end{aligned} \tag{A.3}$$

From Lemma A.4 it follows that

$$\sum_{i=1}^{z}\sum_{j\in I_i} X_j \geq L\left(F|_\rho\right)\ . \tag{A.4}$$

Finally, plugging eq. (A.3) and eq. (A.4) into eq. (A.2), there exists a constant c such that

$$\Pr_{\rho\in\mathcal{R}_p}\left[L\left(F|_\rho\right)\geq c\cdot p^2\cdot\log^{3/2}(n)\cdot L(F)\right]\leq L(F)\cdot e^{-n\cdot p^8}$$

as needed. □

B The Error Correcting Code

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

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

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

We note that for any given $x \in \{0,1\}^{4n}$ and $i \in [2^r]$, computing the encoding of x at coordinate i (according to $\mathcal{C}$) can be done in poly(n) time.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de