Electronic Colloquium on Computational Complexity, Report No. 60 (2012)

DNF Sparsification and a Faster Deterministic Counting Algorithm

Parikshit Gopalan
Microsoft Research
Silicon Valley
parik@microsoft.com

Raghu Meka*
Institute for Advanced Study
Princeton
raghu@ias.edu

Omer Reingold
Microsoft Research
Silicon Valley
Omer.Reingold@microsoft.com

Abstract

Given a DNF formula f on n variables, the two natural size measures are the number of terms or size $s(f)$, and the maximum width of a term $w(f)$. It is folklore that short DNF formulas can be made narrow. We prove a converse, showing that narrow formulas can be sparsified. More precisely, any width w DNF irrespective of its size can be ε-approximated by a width w DNF with at most $(w \log(1/\varepsilon))^{O(w)}$ terms.

We combine our sparsification result with the work of Luby and Velikovic [LV91, LV96] to give a faster deterministic algorithm for approximately counting the number of satisfying solutions to a DNF. Given a formula on n variables with $\mathrm{poly}(n)$ terms, we give a deterministic $n^{\tilde{O}(\log\log(n))}$ time algorithm that computes an additive ε approximation to the fraction of satisfying assignments of f for $\varepsilon = 1/\mathrm{poly}(\log n)$. The previous best result due to Luby and Velickovic from nearly two decades ago had a run-time of $n^{\exp(O(\sqrt{\log\log n}))}$ [LV91, LV96].

*Work done while an intern at Microsft Research, Silicon Valley.

ISSN 1433-8092

1 Introduction

A natural way to represent a Boolean function $f : \{0,1\}^n \to \{0,1\}$ is to write it as a CNF or DNF formula. The class of functions that admit compact representations of this form (aka polynomial size CNF and DNF formulae) are central to Boolean function analysis, computational complexity and machine learning.

Given a DNF formula f on n variables, the two natural size measures are the number of terms or size $s(f)$, and the maximum width of a term $w(f)$. The analogous measures for a CNF, are the number of clauses and clause width. It is folklore that every DNF formula f with m terms can be ε-approximated by another DNF g where $s(g) \leq m$ and $w(g) \leq \log(m/\varepsilon)$, regardless of $w(f)$. The formula g is a sparsification of f obtained by simply discarding all terms of width larger than $\log(m/\varepsilon)$. In other words, short DNF formulas can be made narrow. An analogous statement can be derived for CNFs.

In this work, we show the reverse connection: narrow formulae can be made short. Indeed, we prove the existence of a strong form of approximation known as sandwiching approximations which are important in pseudorandomness. In this work we only consider approximators which are also Boolean functions.

Definition 1.1. *Let $f : \{0,1\}^n \to \{0,1\}$. We say that functions $f_u, f_\ell : \{0,1\}^n \to \{0,1\}$ are ε-sandwiching approximators for f if $f_\ell(x) \leq f(x) \leq f_u(x)$ for every $x \in \{0,1\}^n$, and*

$$\Pr_{x\in\{0,1\}^n}[f_\ell(x) \neq f(x)] = \Pr_{x\in\{0,1\}^n}[(f_\ell(x) = 0) \wedge (f(x) = 1)] \leq \varepsilon,$$
$$\Pr_{x\in\{0,1\}^n}[f_u(x) \neq f(x)] = \Pr_{x\in\{0,1\}^n}[(f_u(x) = 1) \wedge (f(x) = 0)] \leq \varepsilon.$$

Our main result is the existance of ε-sandwiching approximators for arbitrary width w DNFs using short width w DNFs where the number of clauses depends only on w and ε.

Theorem 1.1. *For every width-w DNF formula f and every $\varepsilon > 0$, there exist DNF formulae f_ℓ, f_u each of width w and size at most $(w \log(1/\varepsilon))^{O(w)}$ which are ε-sandwiching approxmiators for f.*

Our result is proved by a sparsification procedure for DNF formulae which uses the notion of quasi-sunflowers due to Rossman [Ros10]. The best previously known result along these lines was due to Trevisan [Tre04], who built on previous work by Ajtai and Wigderson [AW85]. Trevisan shows that every width w DNF has ε-sandwiching approximators that are decision trees of depth $d = O(w2^w \log(1/\varepsilon))$.

A k-junta is a function which depends only on k variables. We say that $f, g : \{0,1\}^n \to \{0,1\}$, we say that g ε-approximates f if

$$\Pr_{x\in\{0,1\}^n}[f(x) \neq g(x)] \leq \varepsilon.$$

A corollary of our result is the following *junta theorem* for DNFs.

Corollary 1.2. *Every width-w DNF formula is ε-approximated by a $(w \log(1/\varepsilon))^{O(w)}$-junta.*

A similar but incomparable statement can be derived from Friedgut's junta theorem [Fri98]. It is easy to see that width w DNFs have average sensitivity at most $2w$ [1], so by Friedgut's theorem any width w DNF is ε-close to a $2^{\tilde{O}(w/\varepsilon)}$-junta. Friedgut's result gives better dependence on w, whereas we achieve much better dependence on ε. Friedgut's approximator is not *a priori* a small-width DNF, and one does not get sandwiching approximations. Trevisan's result implies that any width w DNF is ε-approximated by a k-junta for $k = \exp(O(w2^w \log(1/\varepsilon)))$ [Tre04].

Theorem 1.1 has interesting consequences for other parameter settings. One example is the following:

Corollary 1.3. *Every width-$O(\log n)$ DNF formula on n variables is $n^{-O(1)}$ close to a DNF of width $O(\log n)$ and size $n^{O(\log\log(n))}$.*

In Section 6, we conjecture that a better bound should be possible in Theorem 1.1, which is singly exponential in w. If true, this conjecture will give better bounds for both Corollaries 1.2 and 1.3.

1.1 DNF Counting and Pseudorandom Generators

The problem of estimating the number of satisfying solutions to CNF and DNF formulae is closely tied to the problem of designing pseudorandom generators for such formulae with short seed-length. These problems have been studied extensively [KL83, AW85, NW94, Nis91, LV91, LV96, LVW93, Tre04, Baz09, Raz09, DETT10].

For a formula f, let

$$\mathsf{Bias}(f) = \Pr_{x\in\{0,1\}^n}[f(x) = 1].$$

Given a formula f from a class $\mathcal{F}$ of functions, the goal of a counting algorithm for the class $\mathcal{F}$ is to compute $\mathsf{Bias}(f)$. We refer to the counting problems for CNFs and DNFs as #CNF and #DNF respectively. The problem of computing $\mathsf{Bias}(f)$ exactly is #P-hard [Val79], hence we look to approximate $\mathsf{Bias}(f)$.

An algorithm gives an ε-additive approximation for $\mathsf{Bias}(f)$ if its output is in the range $[\mathsf{Bias}(f)-\varepsilon, \mathsf{Bias}(f)+\varepsilon]$. It is easy to see that additive approximations for CNFs and DNFs are equivalent. There is a trivial solution based on random sampling, but finding a deterministic polynomial time algorithm has proved challenging.

Computing multiplicative approximations to $\mathsf{Bias}(f)$ is harder, and here the complexities of #CNF and #DNF are very different. An algorithm is said to be a c-approximation algorithm if its output lies in the range $[\mathsf{Bias}(f), c\mathsf{Bias}(f)]$. It is easy to see that obtaining a multliplicative approximation for #CNF is NP-hard. Karp and Luby gave the first multiplicative approximation for #DNF, their algorithm is randomized [KL83]. There is a reduction between additive and multiplicative approximations for #DNF: for DNF formulae with m terms, the problem of computing a $(1+\varepsilon)$-multiplicative approximation can be

[1] [Ama11] shows a sharp bound of w

reduced deterministically to the problem of computing an (ε/m)-additive approximation to #DNF. This reduction is stated explicitly in [LV96], where is attributed to [KL83, KLM89]

Derandomizing thes Karp-Luby algorithm is an important problem in derandomization that has received a lot of attention starting form the work of Ajtai and Wigderson [AW85, LN90, LV91, LVW93, LV96, Tre04]. The best previous result is due to Luby and Velickovic [LV91, LV96] from nearly two decades ago: they gave a deterministic $n^{\exp(O(\sqrt{\log\log n}))}$ time algorithm that can compute an ε-additive approximation for any fixed constant ε.

A natural approach to this problem is to design pseudorandom generators (PRGs) with small seeds that can ε fool depth two circuits. This problem and its generalization to constant depth circuits are central problems in pseudorandomness [AW85, NW94, Nis91, LV96, LVW93, Tre04, Baz09, Raz09, Bra10, DETT10].

Definition 1.4. *A generator $G : \{0,1\}^r \to \{0,1\}^n$ δ-fools a class $\mathcal{F}$ of functions if*

$$\left| \Pr_{y \in \{0,1\}^r} [f(G(y))] - \mathsf{Bias}(f) \right| \leq \delta$$

for all $f \in \mathcal{F}$. The genrator is said to be explicit if G is computable in time polynomial in r and n.

A generator with seed-length r that ε-fools DNFs with m clauses gives an ε-additive approximation for $\mathsf{Bias}(f)$ in $\mathrm{poly}(m, n, 2^r)$ time by enumerating over all seeds. Such an algorithm only requires black-box access to f. The reduction form [KL83, KLM89] implies that an optimal pseduorandom generator for DNFs with seedlength $O(\log(mn/\varepsilon))$ will give a deterministic multiplicative approximation algorithm for #DNF. However, the best known generator currently due to De, Etesami, Trevisan and Tulsiani [DETT10] requires seed length $O((\log(mn/\varepsilon)^2)$. The Luby-Velikovic algorithm is a not a black-box algorithm, but PRGs for small-width DNFs are an important ingredient.

Our Results

We use our sparsification lemma to give a better PRG for the class of width w DNF formulae on n variables, which we denote by $\mathsf{DNF}(w, n)$. [2]

Theorem 1.5. *For all δ, there exists an explicit generator $G : \{0,1\}^r \to \{0,1\}^n$ that δ-fools $\mathsf{DNF}(w, n)$ and has seed-length*

$$r = \tilde{O}\left(w^2 + w \log\left(\frac{1}{\delta}\right) + \log\log(n)\right).$$

In comparison, Luby and Velickovic [LV96] give a PRG with seed-length $O(2^w + \log\log n)$ for fooling width w DNFs. Note that for $w = O(\log\log n)$ and δ constant, the seed-length of the our generator is $\tilde{O}((\log\log n)^2)$, whereas Luby and Velickovic need seed-length $O(\log^{O(1)} n)$. For $w = \log\log(n)$ and $\delta \geq 1/\mathrm{poly}(n)$, our seed-length is still $\tilde{O}(\log n)$.

[2] The $\tilde{O}()$ notation is used to hide terms that are logarithmic in the arguments.

The improved generator for small-width DNFs is obtained by using our sparsification result to reduce fooling width w DNFs with an arbitrary number of terms to fooling width w DNFs with $2^{\tilde{O}(w)}$ terms. We then apply recent results by De *et al.* on fooling DNF formulas using small-bias spaces. The fact that our sparsification gives sandwiching approximators is critical for this result.

The Luby-Velickovic counting algorithm can be viewed as a (non black-box) reduction from fooling DNFs of size $\mathrm{poly}(n)$ to fooling DNFs of smaller width. Given Theorem 1.5, we can improve and simplify their analysis to get a faster deterministic counting algorithm. This is the first progress on this well-studied problem in nearly two decades. In addition, we can allow for smaller values of ε.

Theorem 1.6. *There is a deterministic algorithm which when given a* DNF *formula on n variables of size m as input, returns an $O(\varepsilon)$-additive approximation to* $\mathsf{Bias}(f)$ *in time*

$$\left(\frac{mn}{\varepsilon}\right)^{\tilde{O}(\log\log(n)+\log\log(m)+\log(1/\varepsilon))}$$

For $m \leq \mathrm{poly}(n)$ and $\varepsilon \geq 1/\mathrm{poly}(\log n)$, the running time is $O(n^{\tilde{O}(\log\log(n))})$.

Håstad's celebrated Switching Lemma [Hås86] is a powerful tool in proving lower bounds for small-depth circuits. It also has applications in computational learning [LMN93, Man95] and PRG constructions [AW85, GMR+12]. As an additional application of our sparsification result, we give a partial derandomization of the switching lemma. The parameters we obtain are close to that of the previous best results due to Ajtai and Wigderson [AW85] and perhaps more importantly, our argument is conceptually simpler, involving iterative applications of our sparsification result and a naive union bound. We defer the details to Section 5.

2 DNF Sparsification

We will consider DNF formulas that are specified as $f = \vee_{i=1}^{m} T_i$ where the representation is minimal in the following sense:

- Each T_i is non-constant. Hence each term is non-empty (else we replace it by 1), and does not contain a variable and its negation (else we replace it by 0). This guarantess that $\mathsf{Pr}_x[T_i = 1] \leq 1/2$.
- Each that T_i is not implied by some other T_j; if this is so, we can simply drop T_i from the definition of f. This means that when viewed as a set of literals, $T_j \not\subset T_i$. A consequence is that $T_i \cap T_j \subsetneq T_j$.

If some stage of our sparsification produces a representation which is not minimal, we can convert it to a minimal represntation without increasing the number of terms.

We call a DNF f *unate* if it does not contain a variable and its negation.

2.1 Sparsification using Sunflowers

We will first show the following weaker version of Theorem 1.1 with a bound of $(w2^w \ln(m/\varepsilon))^w$, and assumes that f is unate. The proof will illustrate the key ideas behind our sprsification procedure.

Theorem 2.1. *For every unate* DNF *formula f with width w and size m every $\varepsilon > 0$, there exist* DNF *formulae f_ℓ, f_u each with width w and at most $(w \log(m/\varepsilon))^{O(w)}$ which are ε-sandwiching approxmiators for f.*

The starting point of our sparsification result is the Erdős-Rado Sunflower Lemma [ER60].

Definition 2.1. *Let $k \geq 3$. A collection of subsets $S_1, \ldots, S_k \subseteq [n]$ is a sunflower with core Y if $Y \subsetneq S_i$ for all i and $S_i \cap S_j = Y$ for all $i \neq j$. The sets $S_i \setminus Y$ are called the petals.*

The set systems that we consider will arise from the terms in some minimal representation of a monotone DNF. This will ensure that the petals are always non-empty, although the core might be empty.

The celebrated Erdős-Rado Sunflower Lemma guarantees that every sufficiently large set system of bounded size sets contains large sunflowers.

Theorem 2.2. *(Sunflower Lemma, [ER60]) Let $\mathcal{F} = \{S_1, \ldots, S_m\}$ be a collection of subsets of $[n]$, each of cardinality at most w. If $m > w!(k-1)^w$, then $\mathcal{F}$ has a sunflower of size k.*

The lemma and its variants have found several applications in complexity theory, we refer the reader to [Juk01, Chapter 7] for more details. We will use it to prove Theorem 2.1.

Proof. (Proof of Theorem 2.1.) Fix a unate, width w DNF $f = T_1 \vee T_2 \vee \cdots \vee T_m$ and for simplicity suppose that f is monotone. Since f is monotone, we can think of each term T_i as a set of variables of size at most w. Set $k = 2^w \ln(m/\varepsilon)$. Provided

$$m \geq \left(w2^w \ln\left(\frac{m}{\varepsilon}\right)\right)^w \geq w!(k-1)^w \tag{2.1}$$

the Sunflower Lemma guarantees the existance of a collection of terms $T_{i_1}, \ldots, T_{i_k}$ with a core $Y = \cap_{j=1}^k T_{i_j}$ and disjoint petals $T_{i_j} \setminus Y$. Hence we can write

$$\vee_{j=1}^k T_{i_j} = Y \wedge \left(\vee_{j=1}^k (T_{i_j} \setminus Y)\right) = Y \wedge g \text{ where } g = \vee_{i=1}^k (T_{i_j} \setminus Y).$$

Note that g is a read-once DNF of width w and size $k = 2^w \ln(m/\varepsilon)$, so it is almost surely satisfied by a random assignment:

$$\Pr_x[g(x) = 0] = \prod_{i=1}^k \Pr_x[T_{i_j} \setminus Y = 0] \leq \left(1 - \frac{1}{2^w}\right)^k \leq \frac{\varepsilon}{m}.$$

The first inequality holds because each $T_{i_j} \setminus Y$ is a term with width at most w, and the second by our choice of k.

Thus a natural way to get an upper sandwiching approximation is to replace $g(x)$ by the constant 1, which is equivalent to replacing $\vee_{j=1}^{k} T_{i_j}$ with Y. Let $f' : \{0,1\}^n \to \{0,1\}$ be the DNF formula obtained by this replacement. It is clear that $f(x) \leq f'(x)$. Further,

$$\Pr_x[f(x) = 0 f'(x) = 1] \leq \Pr_x[g(x) = 0] \leq \frac{\varepsilon}{m}.$$

Finally, we have $s(f') \leq s(f) - (k-1)$.

We can now iteratively apply the above argument as long as the number of terms is larger than the bound in Equation (2.1). In each iteration we reduce $s(f)$ by $k-1$. Thus, we repeat the process at most $m/(k-1)$ times, obtaining an upper approximating formula f_u where

$$\begin{aligned} f(x) &\leq f_u(x) \ \forall x \in \{0,1\}^n, \\ \Pr_x[f(x) \neq f_u(x)] &\leq \frac{m}{k-1} \cdot \frac{\varepsilon}{m} = \varepsilon, \\ s(f_u) &\leq \left(w 2^w \ln\left(\frac{m}{\varepsilon}\right)\right)^w. \end{aligned}$$

We next describe the construction of the lower approximating formula f_ℓ. We start with the sunflower $T_{i_1}, \cdots, T_{i_k}$ with core Y. Now consider the formula f'' obtained from f by dropping one of the terms, say T_{i_1}. Then, $f''(x) \leq f(x)$. Further, the two of them differ only if $f''(x) = 0$ and $f(x) = 1$, which happens if $T_{i_1} = 1$ whereas $T_{i_j} = 0$ for $j \in \{2, \ldots, k\}$. Hence we can bound this probability by

$$\begin{aligned} \Pr_x[f''(x) \neq f(x)] &= \Pr_x[T_{i_j} = 1] \cdot \Pr_x[(\vee_{j=2}^{k} T_{i_j}) = 0 | T_{i_j} = 1] \\ &= \frac{1}{2} \Pr_x[(\vee_{j=2}^{k} T_{i_j} \setminus Y) = 0] = \frac{1}{2}\left(1 - \frac{1}{2^w}\right)^{k-1} \leq \frac{\varepsilon}{m} \end{aligned}$$

where the second inequality holds since by the sunflower property, conditioning on $T_{i_1} = 1$ fixes the core $Y = 1$, but does not affect the other petals. Note that $s(f'') \leq s(f) - 1$. We now iterate this step no more than m times to obtain a formula f_ℓ where

$$\begin{aligned} f_\ell(x) &\leq f(x) \ \forall x \in \{0,1\}^n, \\ \Pr_x[f_\ell(x) \neq f(x)] &\leq m \cdot \frac{\varepsilon}{m} = \varepsilon, \\ s(f_u) &\leq \left(w 2^w \ln\left(\frac{m}{\varepsilon}\right)\right)^w. \end{aligned}$$

□

Theorem 2.1 is weaker than Theorem 1.1 in the assumption of unateness, the dependence on m and the dependence on w. We briefly sketch how one can handle the first two issues.

1. **Unateness.** One can remove this assumption by using Lemma 2.7 which guarantees that any DNF formula contains a large sub-formula which is unate. The resulting statement already suffices for Corollary 1.3, since any width $\log(n)$ DNF can have at most $n^{O(\log(n))}$ many clauses.

2. **Dependence on m.** The size of the approximators depends logarithmically on m. One can avoid this by observing that when the formula size is large, the error resulting from each step of the sparsification is tiny. One can use this argument to get a size bound of $(2^w \ln(1/\varepsilon))^{O(w)}$ which is independent of m.

3. **Dependence on w.** The final bound is exponential in w^2 rather than w. This comes from the $(k-1)^w$ term in the Sunflower Lemma, which we apply for $k = 2^w$. The question of whether the $w!$ term in the Sunflower Lemma is necessary is a well-known open problem in combinatorics. But there is a lower bound of $(k-1)^w$ [Juk01]. So even if the lower bound were to be right answer, it does not (directly) imply a better bound for Theorem 2.1.

2.2 Sparsification using Quasi-Sunflowers.

The main property of the sunflower system we used in Theorem 2.1 is that the formula g on the petals is highly biased towards 1. As shown by Rossman [Ros10], one can guarantee the existence of such "quasi-sunflower" systems satisfying this weaker property, even when the number of terms is much smaller than in the usual sunflower lemma. We adapt our argument to use quasi-sunflowers instead of sunflowers, to obtain Theorem 1.1.

We shall use the notion of quasi-sunflower due to Rossman [Ros10].

Definition 2.2. *(Quasi-Sunflowers, [Ros10]) A unate* DNF *formula $h = \vee_{i=1}^{k} T_i$ where $k \geq 2$ is a γ-quasi-sunflower with core $Y = \cap_{j=1}^{k} T_i$, and petals $\{T_i \setminus Y\}_{i=1}^{k}$ if*

$$\Pr_x[\vee_{i=}^{k}(T_i \setminus Y) = 1] \geq 1 - e^{-\gamma}.$$

Quasi-sunflowers extend the notion of a sunflower in the sense that even though the "petals" $(T_{i_j} \setminus Y)$ are not necessarily disjoint, the probability that none of them is satisfied is small. We disallow $k = 1$, since otherwise every term is trivially a quasi-sunflower. Since we insist that no term of a DNF is contained in another, the petals are non-empty. Hence each petal is satified with probability at most $1/2$, so every γ-sunflower has $k = \Omega(\gamma)$ petals.

Lemma 2.3. *(Quasi-Sunflower Lemma, [Ros10]) Any unate width w* DNF *formula with m terms contains a $\gamma(m)$-quasi-sunflower where*

$$\gamma(m) := \frac{1}{5}\left(\frac{m}{w!}\right)^{1/w}. \tag{2.2}$$

Rossman states the result in the language of set systems, which we have rephrased in the language of DNFs. We show the equivalence of the two in the appendix.

The following lemma will be used to analyze a single step of our sparsification.

Lemma 2.4. *Let $g = \vee_{i=1}^{m} T_i$ be a unate* DNF*. Then*

$$\Pr_x[(T_1 = 1) \wedge ((\vee_{i=2}^{k} T_i) = 0)] \leq \Pr_x[(\vee_{i=1}^{k} T_i) = 0].$$

Proof. Without loss of generality suppose that g is monotone. Since every term in g is also monotone, Kleitman's lemma [AS11, Chapter 6] implies that

$$\begin{aligned}\Pr_x[(T_1 = 0) \wedge ((\vee_{i=2}^k T_i) = 0)] &\geq \Pr_x[T_1 = 0] \cdot \Pr_x[(\vee_{i=2}^k T_i) = 0] \\ \Pr_x[(T_1 = 1) \wedge ((\vee_{i=2}^k T_i) = 0)] &\leq \Pr_x[T_1 = 1] \cdot \Pr_x[(\vee_{i=2}^k T_i) = 0]\end{aligned}$$

Hence we have

$$\frac{\Pr_x[(T_1 = 0) \wedge ((\vee_{i=2}^k T_i) = 0)]}{\Pr_x[T_1 = 0]} \geq \Pr_x[(\vee_{i=2}^k T_i) = 0] \geq \frac{\Pr_x[(T_1 = 1) \wedge ((\vee_{i=2}^k T_i) = 0)]}{\Pr_x[T_1 = 1]}.$$

But this implies that

$$\Pr_x[(T_1 = 1) \wedge ((\vee_{i=2}^k T_i) = 0)] \leq \Pr_x[(\vee_{i=1}^k T_i) = 0] \cdot \frac{\Pr_x[T_1 = 1]}{\Pr_f[T_1 = 0]} \leq \Pr_x[(\vee_{i=1}^k T_i) = 0]$$

where the last inequality follows because for any (non-empty) term T,

$$\Pr_x[T = 1] \leq \frac{1}{2} \leq \Pr_x[T = 0]. \tag{2.3}$$

□

The only property of T_1 that we use is that $\Pr_x[T_1 = 1] \leq \Pr_x[T_1 = 0]$. Indeed, we can drop any set of terms $\{T_i\}_{i \in S}$ which satisfies $\Pr_x[\vee_{i \in S} T_i = 1] \leq \Pr_x[\vee_{i \in S} T_i = 0]$.

The following is our key technical lemma. It applies to unate formulae and allows us to reduce the size of formula by (at least) 1.

Lemma 2.5. *For every unate width-w DNF formula g of size m, there exist width-w DNF formulae g_ℓ, g_u each of size at most $m - 1$ that are $e^{-\gamma(m)}$ sandwiching approximators for g.*

Proof. Let $g = \vee_{i=1}^m T_i$. Lemma 2.3 guarantees the existance of a $\gamma(m)$-quasi-sunflower $h = \vee_{i=1}^k T_{i_j}$ where $\gamma(m)$ is given by Equation (2.2). Letting $p(x) = \vee_{i=1}^k (T_{i_j} \setminus Y)$ be the formula on the petals, we have $\Pr_x[p(x) = 0] \leq e^{-\gamma(m)}$. We can write

$$h(x) \;=\; \vee_{j=1}^k T_{i_j} \;=\; Y \wedge \left(\vee_{j=1}^k (T_{i_j} \setminus Y)\right) \;=\; Y \wedge p(x)$$

We get an upper sandwiching DNF formula $g_u : \{0,1\}^n \to \{0,1\}$ from $g(x)$ by replacing $p(x)$ by the constant 1, which is equivalent to replacing $h(x)$ with the core Y. It is clear that

$$g(x) \leq g_u(x), \quad s(g_u) \leq s(g) - (k-1) \leq s(g) - 1.$$

Further,

$$\begin{aligned}\Pr_x[g(x) \neq g_u(x)] &= \Pr_x[(g(x) = 0) \wedge (g_u(x) = 1)] \\ &\leq \Pr_x[p(x) = 0] \\ &\leq e^{-\gamma(m)}.\end{aligned}$$

We now construct the lower sandwiching approximation. Let g_ℓ be the formula obtained from g by dropping the term T_{i_1}. Then, it is clear that

$$g_\ell(x) \leq g(x), \quad s(g_\ell) \leq s(g) - 1.$$

Further,

$$\begin{aligned}
\Pr_x[g(x) \neq g_\ell(x)] &= \Pr_x[g(x) = 1 \wedge g_\ell(x) = 0] \\
&\leq \Pr_x[((T_{i_1} \setminus Y) = 1) \wedge (\vee_{j=2}^{k}(T_{i_j} \setminus Y)) = 0] \\
&\leq \Pr_x[p(x) = 0] \qquad \text{(By Lemma 2.4)} \\
&\leq e^{-\gamma(m)}.
\end{aligned}$$

□

One can prove Theorem 1.1 for unate DNFs by repeated applications of this Lemma. To handle the general case, we use the following simple lemmas to reduce the problem of constructing sandwiching approximations to the unate case.

Lemma 2.6. *Let $f, g, h : \{0,1\}^n \to \{0,1\}$ be such that $f = g \vee h$. Let g_ℓ, g_u be ε-sandwiching approximators for g. Then $g_\ell \vee h$ and $g_u \vee h$ are ε-sandwiching approximators for f.*

Proof. It is easy to see that for every $x \in \{0,1\}^n$,

$$g_\ell(x) \vee h(x) \leq g(x) \vee h(x) \leq g_u(x) \vee h(x).$$

We bound the approximation error for $g_\ell \vee h$, the proof for $g_u \vee h$ is similar.

$$\begin{aligned}
\Pr_x[(g_\ell(x) \vee h(x)) \neq (g(x) \vee h(x))] &= \Pr_x[(g_\ell(x) \vee h(x) = 0) \wedge (g(x) \vee h(x) = 1)] \\
&= \Pr_x[(g_\ell(x) = 0) \wedge (g(x) = 1) \wedge (h(x) = 0)] \\
&\leq \Pr_x[(g_\ell(x) = 0) \wedge (g(x) = 1)] \\
&\leq \varepsilon.
\end{aligned}$$

□

Lemma 2.7. *For every width w DNF $f = \vee_{i=1}^{m} T_i$ of size m, there exists $S \subseteq [m]$ where $|S| \geq m/2^w$ such that the formula $g = \vee_{j \in S} T_{i_j}$ is unate.*

Proof. Pick a random set of literals S as follows: for each of the variables x_i add one of x_i or $\bar{x}_i$ to S uniformly at random. Let g_S be the sub-formula of f formed of terms containing only literals from S. Then, g_S is always unate.

Each term has at least a 2^{-w} chance of being in g_S. By linearity of expectation

$$\mathop{\mathsf{E}}_S[s(g_S)] \geq \frac{m}{2^w}.$$

□

We will use the following asymptotic bound whose proof is a calculation and is deferred to the appendix.

Fact 2.8. *For $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ defined by Equation 2.2, $W = (2w)^{3w}(50\log(1/\varepsilon))^w$, and $\varepsilon \leq 1/4$,*

$$\sum_{j=W+1}^{m} e^{-\gamma(j/2^w)} \leq \varepsilon.$$

We can now prove Theorem 1.1:

Proof. Let $f = \vee_{i=1}^{m} T_i$. By applying Lemma 2.7, we can write $f = g \vee h$ where g is unate and has $m' \geq m/2^w$ terms. By Lemma 2.5, there exist sandwiching approximators g_ℓ, g_u each of width w and size at most $m' - 1$, whose error is bounded by

$$e^{-\gamma(m')} \leq e^{-\gamma(m/2^w)}.$$

By Lemma 2.6, $f_\ell^1 = g_\ell \vee h$ and $f_u^1 = g_u \vee h$ are $e^{-\gamma(m')}$ sandwiching approximations for f. Further

$$s(f_\ell^1) = s(g_\ell) + s(h) \leq s(g) - 1 + s(h) \leq s(f) - 1$$

and similarly $s(f_u^1) \leq s(f) - 1$.

We iterate this construction separately for the upper and lower approximator till the size of the formulae drops below W. This gives the sequence

$$f(x) \leq f_u^1(x) \cdots \leq f_u^{k_u}(x) := f^u(x)$$
$$f(x) \geq f_\ell^1(x) \cdots \geq f_\ell^{k_\ell}(x) := f_\ell(x)$$

where $s(f_\ell), s(f_u) \leq W$. We can bound the error of these approximators by

$$\sum_{j=W+1}^{m} e^{-\gamma(j/2^w)} \leq \varepsilon. \tag{2.4}$$

where the inequality is from Fact 2.8. This completes the proof of Theorem 1.1. □

3 Fooling Small-Width DNFs

We next use our sparsification result to construct a pseudorandom generator for small-width DNFs, obtaining an exponential improvement in terms of the width over the generator of Luby and Velickovic [LV96]. We restate Theorem 1.5 with the exact asymptotics for r.

Theorem 3.1. *For all δ, there exists an explicit generator $G : \{0,1\}^r \to \{0,1\}^n$ that δ-fools all width w DNFs and has seed-length*

$$r = O\left(w^2 \log^2(w) + w\log(w)\log\left(\frac{1}{\delta}\right) + \log\log(n)\right)$$

We prove the theorem as follows: we first use our sparsification result to reduce the case of fooling width w DNFs with an arbitrary number of terms to that of fooling width w DNFs with $2^{\tilde{O}(w)}$ terms and then apply the recent results due to De *et al.* [DETT10] showing that small-bias spaces fool DNFs with few terms.

Definition 3.2 (k-wise ε-biased spaces)**.** *A distribution $\mathcal{D}$ over $\{0,1\}^n$ is said to be (k,ε)-biased space if for every non-empty subset $I \subseteq [n]$ of size at most k,*

$$\left| \Pr_{x \leftarrow \mathcal{D}}[\oplus_{i \in I} x_i = 1] - \frac{1}{2} \right| \leq \varepsilon.$$

Naor and Naor [NN93] constructed explicit (k,ε)-biased spaces that require only $O(k + \log(1/\varepsilon) + \log\log n)$ bits to sample from.

Next, we need the following result of De *et al.* [DETT10] showing that (k,ε)-biased spaces fool DNFs for suitable choices of k and ε.

Theorem 3.3. *[DETT10, Theorem 4.1] For every $\delta > 0$, every DNF with width w and size m is δ-fooled by (k,ε)-biased distributions for*

$$k = O\left(w \log\left(\frac{m}{\delta}\right)\right),$$
$$\log\left(\frac{1}{\varepsilon}\right) = O\left(w \log(w) \log\left(\frac{m}{\delta}\right)\right).$$

De *et al.* prove the above statement only for the case of $k = n$, and they use the bound $w \leq \log(m/\delta)$. Their proof proceeds by constructing small ℓ_1-norm sandwiching approximators. The above statement is obtained by repeating their proof keeping w and m separate, and bounding both the degree and the ℓ_1 norm of the resulting approximators. It is easy to see from their proof that the approximators have degree $k \leq O(w \log(m/\delta))$ and ℓ_1-norm bounded $(m/\delta)^{O(w \log(w))}$.

We use the fact that to fool a class of functions, it suffices to fool sandwiching approximators [BGGP07, Baz09].

Fact 3.4. *Let $\mathcal{F}, \mathcal{G}$ be classes of functions such that every $f \in \mathcal{F}$ has ε-sandwiching approximators in $\mathcal{G}$. Let $G : \{0,1\}^r \to \{0,1\}^n$ be a pseudorandom generator that ε-fools $\mathcal{G}$. Then G $(\varepsilon + \delta)$-fools $\mathcal{F}$.*

We are now ready to prove the main result of this section.

Proof of Theorem 3.1. Recall that $\mathsf{DNF}(w,n)$ denotes the class of all width w DNF s on n variables. Let $\mathcal{G} \subset \mathsf{DNF}(w,n)$ denote the subset of all formulae with size at most $m = (w \log(1/\delta))^{cw}$ for some sufficiently large constant c. By Theorem 1.1, every $f \in \mathsf{DNF}(w,n)$ can has δ-sandwiching approximators in $\mathcal{G}$.

Next, we apply Theorem 3.3 with $m = (w \log(1/\delta))^{cw}$. Note that

$$\log\left(\frac{m}{\delta}\right) = O\left(w \log(w) + \log\left(\frac{1}{\delta}\right)\right).$$

So we conclude that (k, ε)-biased distributions δ-fool $\mathcal{G}$ where

$$k = O\left(w^2 \log(w) + w \log\left(\frac{1}{\delta}\right)\right)$$
$$\log\left(\frac{1}{\varepsilon}\right) = O\left(w^2 \log^2 w + w \log(w) \log\left(\frac{1}{\delta}\right)\right).$$

Note that we can sample from such a distribution using a seed of length

$$\begin{aligned} r &= O\left(k + \log\left(\frac{1}{\varepsilon}\right) + \log\log(n)\right) \\ &= O\left(w^2 \log^2(w) + w \log(w) \log\left(\frac{1}{\delta}\right) + \log\log(n)\right) \end{aligned}$$

Finally, by Fact 3.4, such distributions 2δ fool the class $\mathsf{DNF}(w, n)$. □

4 Deterministic Counting for DNFs

We now use the PRG for small-width DNFs from the previous section in the Luby-Velickovic counting algorithm [LV96]. The better seed-length means that we do not need to balance various parameters as carefully, and can redo their arguments with simpler and better settings of parameters.

The input to our algorithm is a DNF formula $f = \vee_{j=1}^{m} T_j$ on n variables with size m and width w, and the output is an ε-additive approximation to $\mathsf{Bias}(f)$. We set the following parameters

$$k := \log\left(\frac{w}{\varepsilon}\right), \; t := \frac{w}{k}, \; w' = 6k, \delta = \frac{\varepsilon}{t}$$

Let $\mathcal{H} = \{h : [n] \to [t]\}$ be a family of k-wise independent hash functions. Fix a hash function $h \in \mathcal{H}$ and let $B_j = \{i : h(i) = j\}$. We say the term T_i bad for h if

$$\max_{j \in [t]} |B_j \cap T_i| > w'$$

where we view T_i as a set of variables. Let f_h be the formula obtained from f by dropping all terms that are bad for h.

Let $G : \{0,1\}^r \to \{0,1\}^n$ be the generator from Theorem 1.5 that fools $\mathsf{DNF}(w', n)$ with error at most δ. Define a new generator $G_h : (\{0,1\}^r)^t \to \{0,1\}^n$ as follows:

$$G_h(z_1, \ldots, z_t) = x, \text{ where for } j \in [t], \; x_{|B_j} = G(z_j). \tag{4.1}$$

Thus G_h applies an independent copy of G to each bucket defined by the hash function h.

We now state the counting algorithm:

Algorithm DNFCount

For each $h \in \mathcal{H}$,
  Drop all bad terms for h from f to obtain f_h.
  By enumeration over all $z \in \{0,1\}^{rt}$, compute

$$p_h = \Pr_{z \in \{0,1\}^{rt}}[f_h(G_h(z)) = 1]. \tag{4.2}$$

Return $p_{\mathcal{H}} = \max_{h \in \mathcal{H}} p_h$.

We need the following lemma about k-wise independent hash functions.

Lemma 4.1. *Let $\mathcal{H} : [n] \to [t]$ be a k-wise independent family of hash functions. Then, for every set $S \subseteq [n]$ of size $|S| \leq kt$, and every $j \in [t]$,*

$$\Pr_{h \in_u \mathcal{H}} \left[|h^{-1}(j) \cap S| \geq 6k \right] \leq 2^{-k}.$$

Proof. Fix $j \in [t]$. Let $S = \{1, \ldots, kt\}$ without loss of generality. Let $\{X_i\}_{i=1}^{kt}$ be indicator random variables that are 1 if $h(i) = j$ and 0 otherwise. Then

$$\mathop{\mathsf{E}}_{h \in \mathcal{H}} \left[\sum_{I \subseteq S, |I| = k} \prod_{i \in I} X_i \right] \leq \binom{kt}{k} \cdot \frac{1}{t^k} \leq e^k.$$

Applying Markov's inequality,

$$\Pr_{h \in_u \mathcal{H}} \left[|h^{-1}(j) \cap S| \geq 6k \right] \leq \frac{e^k}{\binom{6k}{k}} \leq 2^{-k}.$$

□

Our analysis requires two Lemmas from [LV96]. Since their terminology and notation differs from ours, we provide proofs of both these Lemmas in Appendix B.

The first Lemma relates the bias of f_h with that of f.

Lemma 4.2. *[LV96, Lemma 11] We have*

$$\forall h \in \mathcal{H}, \quad \mathsf{Bias}(f_h) \leq \mathsf{Bias}(f),$$
$$\mathop{\mathsf{E}}_{h \in \mathcal{H}}[\mathsf{Bias}(f_h)] \geq \mathsf{Bias}(f_h) - \varepsilon.$$

The next lemma showing that G_h fools the formula f_h is essentially [LV96, Lemma 7]. Recall that by Equation (4.2), p_h is the bias of f_h under distribution generated by G_h.

Lemma 4.3. *[LV96, Lemma 7] We have $|p_h - \mathsf{Bias}(f_h)| \leq \varepsilon$.*

With these Lemmas in hand, we now analyze the algorithm.

Theorem 4.4. *Algorithm* DNFCount *when given a* DNF *on n variables with width w and size m as input, returns an $O(\varepsilon)$-additive approximation to* $\mathsf{Bias}(f)$ *in time*

$$O(n^{O(\log(w/\varepsilon))}(\log n)^{O(w)}2^{\tilde{O}(w\log(1/\varepsilon))}m).$$

Proof. The correctness of the algorithm is easy to argue. For every $h \in \mathcal{H}$,

$$\begin{aligned} p_h &\leq \mathsf{Bias}(f_h) + \varepsilon \quad \text{(By Lemma 4.3)} \\ &\leq \mathsf{Bias}(f) + \varepsilon \quad \text{(By Lemma 4.2)} \end{aligned}$$

Further by Lemma 4.2, there exists $h \in \mathcal{H}$ such that

$$\mathsf{Bias}(f_h) \geq \mathsf{Bias}(f) - \varepsilon,$$

hence by Lemma 4.3,

$$p_h \geq \mathsf{Bias}(f_h) - \varepsilon \geq \mathsf{Bias}(f) - 2\varepsilon.$$

Thus $p_{\mathcal{H}}$ is a 2ε-additive approximation $\mathsf{Bias}(f)$.

We now bound the running time. Computing f_h for any $h \in \mathcal{H}$ and evaluating it on $G_h(z)$ for $z \in \{0,1\}^{rt}$ can be done in time $O(mn)$. Thus the running time is dominated by $|\mathcal{H}|2^{rt}$. By standard constructions of k-wise independent hash functions,

$$|\mathcal{H}| \leq n^{O(k)}.$$

Next we bound the seed-length r. Recall that

$$\begin{aligned} k = \log\left(\frac{w}{\varepsilon}\right), \delta = \frac{\varepsilon}{t} = \frac{k\varepsilon}{w} \\ \text{Hence} \quad \log\left(\frac{1}{\delta}\right) = \log\left(\frac{w}{\varepsilon k}\right) = k - \log(k). \end{aligned}$$

Further, $w' = 6k$. Hence Theorem 3.1,

$$\begin{aligned} r &= O\left(w'^2\log^2(w') + w'\log(w')\log\left(\frac{1}{\delta}\right) + \log\log(n)\right) \\ &= O(k^2\log^2(k) + \log\log(n)) \\ rt &= O\left(\frac{w}{k}(k^2\log^2(k) + \log\log(n))\right) \\ &= O(wk\log^2 k + w\log\log(n)). \end{aligned}$$

So we get

$$|\mathcal{H}|2^{rt} \leq \exp(O(k\log(n) + wk\log^2 k + w\log\log(n))).$$

Overall the runtime is bounded by

$$O(mn)|\mathcal{H}|2^{rt} = \exp(O(\log(w/\varepsilon)\log(n) + w\log(w/\varepsilon)(\log\log(w/\varepsilon))^2 + w\log\log(n) + \log(m)))$$
$$= n^{O(\log(w/\varepsilon))}(\log n)^{O(w)}2^{\tilde{O}(w\log(1/\varepsilon))}m.$$

□

Theorem 1.6 is obtained from Theorem 4.4 by setting parameters appropriately.

Proof. (Proof of Theorem 1.6.) Given a DNF formula with size m, we can ignore all terms of width larger than $\log(m/\varepsilon)$ while only changing the bias by ε. Plugging in $w = \log(m/\varepsilon)$, we can bound the running time by

$$\left(\frac{mn}{\varepsilon}\right)^{\tilde{O}(\log\log(n)+\log\log(m)+\log(1/\varepsilon))}$$

For $m = \text{poly}(n), \varepsilon = 1/\text{poly}(\log n)$, this gives $n^{\tilde{O}(\log\log(n))}$. □

5 A Derandomized Switching Lemma

Håstad's celebrated Switching Lemma [Hås86] is a powerful tool in proving lower bounds for small-depth circuits. It also has applications in computational learning [LMN93, Man95] and PRG constructions [AW85, GMR+12]. This lemma builds on earlier work due to Ajtai [Ajt83], Furst, Saxe and Sipser [FSS84] and Yao [Yao85].

To state the Switching lemma, we need to set up some notation. Given $L \subseteq [n]$ and $x \in \{0,1\}^{[n]\setminus L}$ define a restriction $\rho := \rho_{L,x} \in \{*,0,1\}^n$ by $\rho_i = *$ if $i \in L$ and $\rho_i = x_i$ otherwise. We call the set $L \equiv L(\rho)$ as the set of "live" variables. For $f : \{0,1\}^n \to \{0,1\}$, and $\rho \in \{*,0,1\}^n$, define $f_\rho : \{0,1\}^{L(\rho)} \to \{0,1\}$ by $f_\rho(y) = f(x)$, where $x_i = y_i$ for $i \in L(\rho)$ and $x_i = \rho_i$ otherwise.

Given a distribution $\mathcal{D}$ on $2^{[n]}$, let $\mathcal{D}$ (abusing notation, the meaning will be clear from context) denote the distribution on $\rho \in \{*,0,1\}^n$ by setting $\rho = \rho_{L,x}$ where $L \leftarrow \mathcal{D}$ and $x \in_u \{0,1\}^{[n]\setminus L}$. Call a distribution $\mathcal{D}$ as above p-regular if for each $i \in [n]$, $\mathsf{Pr}_{L\leftarrow\mathcal{D}}[i \in L] = p$. Let $\mathcal{D}_p(n)$ (we omit n if clear from context) denote the p-regular distribution on subsets L of $[n]$ where each element $i \in [n]$ is present in L independently with probability p. For $f : \{0,1\}^n \to \{0,1\}$, let $\mathsf{DT}(f)$ denote the minimum depth of a decision tree computing f.

Theorem 5.1 (Switching Lemma, [Hås86])**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a* DNF *of width w and let $\rho \leftarrow \mathcal{D}_p(n)$. Then,*

$$\mathsf{Pr}[\mathsf{DT}(f_\rho) \geq s] < (5pw)^s.$$

There has been work on finding a *derandomized* version of the switching lemma, motivated by better PRG constructions . Such a lemma would choose the set of live variables in a pseudorandom way, as in [AW85]. One could even ask for a stronger derandomization where the assignments to the non-live variables are also chosen pseudorandomly, this is done in [GMR+12]. We limit ourselves to the former case here.

Derandomized switching lemmas were first studied in the seminal work of Ajtai and Wigderson [AW85], with the aim of constructing better PRGs for constant depth circuits.

Theorem 5.2 ([AW85]). *For all $\gamma \in (0,1]$, $p < 1/n^{\gamma}$, there is a p-regular distribution $\mathcal{D}$ on $2^{[n]}$ with $L \leftarrow \mathcal{D}$ samplable using $O_{\gamma}(\log n)$ random bits, and $k = O_{\gamma}(1)$ such that for $\rho \leftarrow \mathcal{D}$, and any polynomial size* DNF *f,*

$$\mathsf{Pr}[f_{\rho} \textit{is not an } k\textit{-junta}] \leq 1/\mathrm{poly}(n).$$

A very recent result along these lines is due to the authors together with Trevisan and Vadhan, which gives a near-optimal derandomization in the special case of read-once DNFs [GMR+12]. They use this to give near PRGs for read-once DNFs with seed-length $\tilde{O}(\log n)$.

We remark that if instead of finding a small set of restrictions that work for all formulas f, we are given the formula f as input, Agrawal et al. [AAI+01] give a polynomial-time algorithm to find a restriction that simplifies the formula as well as the bounds given by the switching lemma Theorem 5.1.

5.1 Our Result

We give a different argument that essentially recovers the result of Ajtai and Wigderson and further gives a trade-off between the survival probability p, the complexity of the restricted function and the failure probability of the restriction. Our argument is through repeated applications of Theorem 1.1 and it seems to us to be simpler than those of Hastad [Hås86] and Ajtai-Wigderson [AW85].

Theorem 5.3. *There exists a constant C such that for any $w, s, \delta > 0$ and all p such that*

$$p \leq \frac{\delta}{(w \log(1/\varepsilon))^{C \log w}},$$

there is a distribution $\mathcal{D}$ on $2^{[n]}$ such that $L \leftarrow \mathcal{D}$ can be sampled using r random bits where

$$r = r(n, s, \varepsilon, \delta) = O\left((\log w) \cdot (\log n + s \log(1/\delta)) + w \log(w \log(1/\varepsilon))\right),$$

the indicator events $\mathbf{1}\{i \in L\}$ are p-biased and the following holds: for any width w DNF *$f : \{0,1\}^n \to \{0,1\}$, and $\rho \leftarrow \mathcal{D}$,*

$$\mathsf{Pr}[f_{\rho} \textit{ does not have } \varepsilon\textit{-sandwiching approximations in } \mathsf{DNF}(s,n)] < \varepsilon + \delta^{s/4}$$

In particular, by setting $\delta = 1/n^{\gamma}$, $s = \Theta(1/\gamma)$, $\varepsilon = 1/\mathrm{poly}(n)$, $w = O(\log n)$, we almost recover the derandomized switching lemma of Ajtai and Wigderson, with the main difference being that we need $O((\log n)(\log\log n))$ bits to sample from $\mathcal{D}$ and we only get f_{ρ} has sandwiching approximations by width $O_{\gamma}(1)$ DNFs.

Our derandomization is based on the intuition that the switching lemma is easy to show when the number of terms in the original DNF f is small in terms of the width w of f. Let

$f = \vee_{j=1}^{2^w} T_j$ be a width w DNF. Note that for $0 < p < 1$, and $\rho \leftarrow \mathcal{D}_p$, the probability that a single term T_i survives the restriction f_ρ (is not set to be a constant) is at most

$$\sum_{i=1}^{w} \binom{w}{i} p^i \left(\frac{(1-p)}{2}\right)^{w-i} \leq \left(\frac{1+p}{2}\right)^w.$$

In particular if $p \leq 1/w$, the above probability is at most $e/2^w$. Thus, by linearity of expectation, the expected number of terms that survive the restriction is at most $O(1)$. Hence, by Markov's inequality, the restricted DNF f_ρ has very few surviving terms with high probability. Further, as we are only using Markov's inequality, the above argument would work even if the restriction ρ is sampled from a distribution where the choices for different variables are only k-wise independent for $k = O(w)$.

We use Theorem 1.1 to reduce the case of arbitrary DNFs of small-width to that of DNFs with a small number of terms and then use an argument similar to the above. Unfortunately, the bound in Theorem 1.1 is not sufficiently strong, so we need to use somewhat stronger restrictions where the sruvival probability is $p = w^{-r}$ for $r \geq 1$. Such a restriction can be viewed as a sequence of r rounds of random restrictions, leaving with a $1/w$ fraction of live variables. We argue that in each round, the width of the formula decreases by $1/2$ with high probability and then iteratively apply the argument to the new width $w/2$ formulas. After $O(\log w)$ rounds, the width reduces to a constant. This corresponds to a random restriction where the probability of being alive is $\exp(-\Omega(\log^2 w))$. Moreover, this argument works even when the random restrictions only have limited independence, yielding Theorem 5.3.

For $k \leq n$, let $\mathcal{D}_p(k)$ denote the class of p-regular distributions on $2^{[n]}$ such that for $L \leftarrow \mathcal{D} \in \mathcal{D}_p(k)$, $\mathsf{Pr}[I \subseteq L] \leq 2p^{|I|}$ for all $I \subseteq [n], |I| \leq k$. There exist explicit distributions $\mathcal{D} \in \mathcal{D}_p(k)$ that can be sampled using $O(k \log(1/p) + \log n)$-random bits. For instance, one can use p^k-almost k-wise independent p-biased variables from [NN93].

Claim 5.4. *There exists a constant $c < 1$ such that the following holds for all $\delta, \varepsilon > 0$, $0 < s \leq w$ and*

$$p \leq p(w,s) := \frac{c\delta^{s/2w}}{(w^3 \log(1/\varepsilon))^2}$$

For any width w DNF $f : \{0,1\}^n \to \{0,1\}$ and $\rho \leftarrow \mathcal{D} \in \mathcal{D}_p(w)$, with probability at least $1 - \delta^{s/4} - \varepsilon$ there exist width $w/2$ DNFs $f_\rho^\ell, f_\rho^u : \{0,1\}^{L_\rho} \to \{0,1\}$ that are ε-sandwiching approximators for f_ρ.

Proof of Claim 5.4. Let f^ℓ, f^u be width w DNFs with at most $h(w) = w^{3w}(C \log(1/\varepsilon))^w$ terms that are $\varepsilon^2/2$-sandwiching approximators for f as guaranteed by Theorem 1.1 for C a large constant. Consider a random restriction ρ sampled from a distribution in $\mathcal{D}_p(w/2)$. Then, the probability that a fixed term of f^ℓ has more than $w/2$ live variables under ρ is at most $2^w \cdot p^{w/2}$. Therefore, by a union bound, the probability that f_ρ^ℓ has width more than $w/2$ is at most $h(w)2^w p^{w/2} < \delta^{s/4}/2$ for a sufficiently small constant c. Similarly, the probability that f_ρ^u has width more than $w/2$ is at most $\delta^{s/4}/2$.

Note that as $f^\ell \le f \le f^u$, $f^\ell_\rho \le f_\rho \le f^u_\rho$. We now need to show that f^ℓ_ρ, f^u_ρ are close to f_ρ with high probability. Let $\rho \equiv \rho_{L,x}$ and consider a fixing of the set of live variables L. Then as f^ℓ, f^u are $\varepsilon^2/2$-sadwiching approximators for f,

$$\begin{aligned}\mathop{\mathsf{E}}_{x\in\{0,1\}^{[n]\setminus L}}[\mathsf{Bias}(f_\rho)] &= \mathsf{Bias}(f)\\ &\le \mathsf{Bias}(f^\ell) + \frac{\varepsilon^2}{2}\\ &= \mathop{\mathsf{E}}_{x\in\{0,1\}^{[n]\setminus L}}[\mathsf{Bias}(f^\ell_\rho)] + \frac{\varepsilon^2}{2}.\end{aligned}$$

Therefore,

$$\mathop{\mathsf{E}}_{x\in\{0,1\}^{[n]\setminus L}}[\mathsf{Bias}(f_\rho) - \mathsf{Bias}(f^\ell_\rho)] \le \frac{\varepsilon^2}{2}.$$

Thus, by Markov's inequality,

$$\mathop{\mathsf{Pr}}_{x\in\{0,1\}^{[n]\setminus L}}[\mathsf{Bias}(f_\rho) - \mathsf{Bias}(f^\ell_\rho) \ge \varepsilon] \le \frac{\varepsilon}{2}.$$

Using a similar argument to f^u, and a union bound, we get that f_ρ is ε-sandwiched by (f^ℓ_ρ, f^u_ρ) with probability at leat $1 - \delta^{s/4} - \varepsilon$. □

We now prove Theorem 5.3.

Proof of Theorem 5.3. Let t be such that $w/2^t = s$ (we ignore the minor technicality of t being non-integral) and for $r = 1, \dots, t$, let $p_r = p(w/2^r, s)$ as defined in the above claim. For $i \in [t]$, let L_i be chosen independently from a distribution in $\mathcal{D}_{p_i}(w/2^i)$. Let $L = \cap_{i=1}^t L_i$ and for $x \in_u \{0,1\}^n$, let $\rho = \rho_{L,x}$. Then, ρ is a random restriction with indicator variables $1\{i \in L\}$ having bias

$$q = \prod_{i=1}^t p_i \ge \frac{c^{\log w} \cdot \delta^{\sum_{i=1}^t s2^i/2w}}{(w^3 \log(1/\varepsilon))^{2\log w}} > \frac{\delta}{(w \log(1/\varepsilon))^{C \log w}},$$

for C a sufficiently large constant.

Define the composition of two restrictions $\rho' \in \{*,0,1\}^L$ and $\rho'' \in \{*,0,1\}^{L(\rho')}$ in the natural way by $(\rho' \circ \rho'')_i = \rho''_i$ if $i \in L(\rho')$ and $(\rho' \circ \rho'')_i = \rho'_i$ otherwise. Then, by definition, we can view ρ as a composition of independently chosen random restrictions $\rho_t \circ \rho_{t-1} \circ \cdots \circ \rho_1$, where $\rho_j \equiv \rho_{L_j, x^j}$ (with $x^j \in_u \{0,1\}^n$). Further, for any function g, $g_\rho \equiv (((g_{\rho_1})_{\rho_2})_{\dots})_{\rho_t}$.

Therefore, by iteratively applying the Claim 5.4 t times with the random restrictions $\rho_1, \dots, \rho_t$ and a union bound, we get that with probability at least $1 - t(\delta^{s/4} + \varepsilon)$, there exists a lower approximating DNF $f^\ell : \{0,1\}^L \to \{0,1\}$ of width at most $w/2^{t+1}$ such that $f^\ell \le f_\rho$ and $\mathsf{Bias}(f_\rho) - \mathsf{Bias}(f^\ell) < t\varepsilon$. Similarly, by iteratively applying the claim to the upper approximators given by the claim, we get that with probability at least $1 - 2t(\delta^{s/4} + \varepsilon)$, f_ρ has $(t\varepsilon)$-sandwiching approximators that are width-s DNFs.

Finally, the number of bits needed to sample L is

$$\begin{aligned} r(n,s,\varepsilon,\delta) &= \sum_{v=1}^{t} O\left(\frac{w}{2^v}\cdot\log(1/p(w/2^v,s))+\log n\right) \\ &= O\left((\log n)(\log w)\right)+\sum_{v=1}^{t}\frac{w}{2^v}\left(\frac{s2^r}{2w}O\left(\log(1/\delta)\right)+O\left(\log(w\log(1/\varepsilon))\right)\right) \\ &= O\left((\log w)\cdot(\log n+s\log(1/\delta))+w\log\left(w\log(1/\varepsilon)\right)\right). \end{aligned}$$

The theorem now follows from applying the above argument to $\delta' = \delta/2t$, $\varepsilon' = \varepsilon/t$ and noting that this only changes the constant terms in the final bounds. □

6 Open Problems

Better DNF Sparsification. A natural open question is to show optimal bounds for DNF sparsification. We believe this question is interesting of its own right, even without the the sandwiching requirement. Formally, let $m(w,\varepsilon)$ be the smallest integer such that every width-w DNF formula can be ε-approximated by a width-w DNF with m terms. Theorem 1.1 shows that $m(w,\varepsilon) \le (w\log(1/\varepsilon)^{O(w)}$. Rocco Servedio [Ser11] observed that the Majority function on $2w$ variables (which is a width-w DNF) shows that $m(w,\varepsilon) \ge 4^{w-o(w)}$ for any constant ε. We are unaware of a better lower bound, and it is conceivable that the right bound is exponential in w. We pose this as a conjecture:

Conjecture 6.1. (Weaker Version) *There exists a function $c(\varepsilon)$ such that*

$$m(w,\varepsilon) \le O(c(\varepsilon)^w).$$

(Stronger Version) *There exists a constant c such that*

$$m(w,\varepsilon) \le O(\log(1/\varepsilon)^{cw}).$$

The weaker version, if true, will imply that $\log(n)$ width DNFs can be ε-approxmiated by $n^{O_\varepsilon(1)}$ size DNFs for any constant ε. Currently Theorem 1.1 gives the weaker bound of

$$m(\log(n),\varepsilon) \le n^{O(\log\log(n)+\log\log(1/\varepsilon))}.$$

The stronger version, if true, will strengthen Freidgut's theorem in the context of DNFs.

Conjecture 6.1 is similar in spirit to Mansour's conjecture which also asserts that DNF formulas admit concise representations, but in the Fourier domain. It also implies reductions between the formulations of his conjecture for small width DNFs and small-size DNFs. We discuss Mansour's conjecture and its relation to ours further in Section 6.1

Sparsification using the Greedy Algorithm. A natural approach to sparsifying a DNF formula f is to view it as a set-covering problem, where we wish to cover $f^{-1}(1) \subseteq \{0,1\}^n$ by width w terms. One could use the greedy algorithm in the hope that it constructs a sparse cover. It woule be interesting to analyze its performance. In this direction, Jan Vondrak has pointed out that one can use the analysis of greedy set cover to argue that if there is a lower sandwiching DNF formula of size $m_\ell(w,\varepsilon)$ which is ε-close to f, then greedy returns a 2ε approximation of size $m_\ell(w,\varepsilon)\ln(1/\varepsilon)$ [Von12].

Deterministic DNF counting. The question of finding a deterministic polynomial time algorithm for approximate DNF counting remains open. One approach towards this goal would be to construct pseudorandom generators for DNFs formulas with seed-length $O(\log(n) + \log(m) + \log(1/\varepsilon))$. Such constructions are currently not known even for read-once DNFs. A recent result by the Trevisan, Vadhan and the authors gets a seed-length of $\tilde{O}(\log(n) + \log(1/\varepsilon))$ in the read-once case [GMR+12].

6.1 Mansour's Conjecture.

We say that $f : \{0,1\}^n \to \{0,1\}$ has a t-sparse ε-approximation if there exists $p : \{0,1\}^n \to \mathbb{R}$ with at most t non-zero Fourier coefficients such that

$$\Pr_{x \in \{0,1\}^n}\left[(f(x) - p(x))^2\right] \leq \varepsilon.$$

Conjecture 6.2. *(Mansour's Conjecture for size) [Man94]*
(Weaker version) *There exists a function $c(\varepsilon)$ such that every DNF of size m has an $m^{c(\varepsilon)}$-sparse ε-approximation.*
(Stronger version) *Every DNF of size m has an $m^{O(\log(1/\varepsilon))}$-sparse ε-approximation.*

Mansour originally stated the stronger version of the conjecture, the weaker version appears in [O'D12]. The following analogue of Mansour's conjecture for small width suggests itself. To our knowledge, this conjecture has not appeared explicitly in the literature.

Conjecture 6.3. *(Mansour's conjecture for width)*
(Weaker version) *There exists a function $c(\varepsilon)$ such that every DNF of width w has an $2^{c(\varepsilon)w}$-sparse ε-approximation.*
(Stronger version) *Every DNF of width w has an $2^{O(w\log(1/\varepsilon))}$-sparse ε-approximation.*

The best known bounds for both size and width are due to Mansour, who shows that every DNF of width w has an $w^{O(w\log(1/\varepsilon))}$-sparse ε-approximation and then derives a bound for size using $w = O(\log(m/\varepsilon))$ [Man95].

We feel that this width analogue of Mansour's Conjecture is natural; indeed most results on DNFs proceed by first tackling the width-w case, and then translating it to DNFs of size m using $w \leq \log(m/\varepsilon)$ [Hås86, LMN93, Man95]. This substitution also shows that

- The weaker version of Mansour's Conjecture for width implies the weaker version of Mansour's Conjecture for size.

- The stronger version of Mansour's Conjecture for width implies the stronger version of Conjecture for size, as long as $\varepsilon \geq 1/\mathrm{poly}(m)$.

Conjecture 6.1 implies the reverse equivalence.

Lemma 6.4.
- *Aussme the stronger version of Conjecture 6.1. Then the stronger version of Mansour's Conjecture for size implies that every width w DNF formula has a $2^{O(w \log(1/\varepsilon) \log\log(1/\varepsilon))}$-sparse ε-approximation.*
- *Assume the weaker version of Conjecture 6.1. Then the weaker version of Mansour's Conjecture for size implies the weaker version of Mansour's Conjecture for width.*

Note that if we replace Conjecture 6.1 with Theorem 1.1, this does not improve on the bound from [Man95]. So in this context, the improved dependence on w in Conjecture 6.1 is crucial.

Acknowledgements

We thank Adam Klivans, Ryan O'Donnell, Rocco Servedio, Avi Wigderson and David Zuckerman for valuable discussions. We thank Rocco for drawing our attention to Friedgut's theorem in this context.

References

[AAI+01] Manindra Agrawal, Eric Allender, Russell Impagliazzo, Toniann Pitassi, and Steven Rudich. Reducing the complexity of reductions. *Computational Complexity*, 10(2):117–138, 2001.

[Ajt83] Miklos Ajtai. Σ^2_1-formula on finite structures. *Ann. Pure. Appl. Logic*, 24:1–48, 1983.

[Ama11] Kazuyuki Amano. Tight bounds on the average sensitivity of k-CNF. *Theory of Computing*, 7(1):45–48, 2011.

[AS11] N. Alon and J.H. Spencer. *The Probabilistic Method*. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, 2011.

[AW85] Miklós Ajtai and Avi Wigderson. Deterministic simulation of probabilistic constant depth circuits (preliminary version). In *FOCS*, pages 11–19, 1985.

[Baz09] Louay M. J. Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM J. Comput.*, 38(6):2220–2272, 2009.

[BGGP07] I. Benjamini, O. Gurel-Gurevich, and R. Peled. On k-wise independent distributions and Boolean functions. available at http://www.wisdom.weizmann.ac.il/ origurel/, 2007.

[Bra10] Mark Braverman. Polylogarithmic independence fools AC^0 circuits. *J. ACM*, 57(5), 2010.

[DETT10] Anindya De, Omid Etesami, Luca Trevisan, and Madhur Tulsiani. Improved pseudorandom generators for depth 2 circuits. In *APPROX-RANDOM*, pages 504–517, 2010.

[ER60] P. Erdős and R. Rado. Intersection theorems for systems of sets. *Journal of the London Mathematical Society*, s1-35(1):85–90, 1960.

[Fri98] Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[FSS84] Merrick L. Furst, James B. Saxe, and Michael Sipser. Parity, circuits, and the polynomial-time hierarchy. *Mathematical Systems Theory*, 17(1):13–27, 1984.

[GMR+12] Parikshit Gopalan, Raghu Meka, Omer Reingold, Luca Trevisan, and Salil Vadhan. Better pseudoranom generators from milder pseudorandom restrictions, 2012. Under submission.

[Hås86] Johan Håstad. Almost optimal lower bounds for small depth circuits. In *STOC*, pages 6–20, 1986.

[Juk01] S. Jukna. *Extremal Combinatorics: With Applications in Computer Science*. Texts in Theoretical Computer Science. Springer, 2001.

[KL83] Richard M. Karp and Michael Luby. Monte-carlo algorithms for enumeration and reliability problems. In *FOCS*, pages 56–64, 1983.

[KLM89] Richard M. Karp, Michael Luby, and Neal Madras. Monte-carlo approximation algorithms for enumeration problems. *J. Algorithms*, 10(3):429–448, 1989.

[LMN93] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform and learnability. *Journal of the ACM*, 40(3):607–620, 1993.

[LN90] N. Linial and N. Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10:349–365, 1990.

[LV91] Michael Luby and Boban Velickovic. On deterministic approximation of DNF. In *STOC*, pages 430–438, 1991.

[LV96] Michael Luby and Boban Velickovic. On deterministic approximation of DNF. *Algorithmica*, 16(4/5):415–433, 1996.

[LVW93] Michael Luby, Boban Velickovic, and Avi Wigderson. Deterministic approximate counting of depth-2 circuits. In *ISTCS*, pages 18–24, 1993.

[Man94] Y. Mansour. *Learning Boolean functions via the Fourier transform*, pages 391–424. Kluwer Academic Publishers, 1994.

[Man95] Y. Mansour. An $o(n^{\log\log n})$ learning algorithm for DNF under the uniform distribution. *Journal of Computer and System Sciences*, 50:543–550, 1995.

[Nis91] Noam Nisan. Pseudorandom bits for constant depth circuits. *Combinatorica*, 11(1):63–70, 1991.

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, 1993.

[NW94] Noam Nisan and Avi Wigderson. Hardness vs randomness. *J. Comput. Syst. Sci.*, 49(2):149–167, 1994.

[O'D12] Ryan O'Donnell. Open problems in analysis of Boolean functions. *CoRR*, abs/1204.6447, 2012.

[Raz09] Alexander A. Razborov. A simple proof of Bazzi's theorem. *TOCT*, 1(1), 2009.

[Ros10] Benjamin Rossman. The monotone complexity of k-clique on random graphs. In *FOCS*, pages 193–201, 2010.

[Ser11] Rocco Servedio, 2011. Personal communication.

[Tre04] Luca Trevisan. A note on approximate counting for k-DNF. In *APPROX-RANDOM*, pages 417–426, 2004.

[Val79] L.G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8(2):189 – 201, 1979.

[Von12] Jan Vondrak, 2012. Personal communication.

[Yao85] Andrew C. Yao. Separating the polynomial-time hierarchy by oracles (preliminary version). In *FOCS*, pages 1–10, 1985.

A Proofs from Section 2

We first show that Lemma 2.3 is equivalent to Lemma A.2 below from [Ros10].

Definition A.1 ([Ros10]). *Let $\mathcal{F}$ be a family of sets over a universe U and let $Y = \cap_{T\in\mathcal{F}} S$. Call $\mathcal{F}$ a γ-sunflower if for a random set $W \subseteq U$, with each element of U present in W independently with probability $1/2$,*

$$\mathsf{Pr}[\exists T \in \mathcal{F}, (T \setminus Y) \cap W = \emptyset] \geq 1 - \gamma.$$

Lemma A.2 ([Ros10]). *Let $\mathcal{F}$ be a family of sets over a universe U each of size at most w. If $|\mathcal{F}| > w! \cdot (2.47 \log(1/\gamma))^w$, then $\mathcal{F}$ contains a γ-sunflower.*

Proof of Lemma 2.3. As f is unate, without loss of generality suppose that f is monotone. Let $U = [n]$ and $\mathcal{F} = \{T_i : 1 \le i \le m\}$. By the above lemma, there exists a γ-sunflower $\mathcal{F}' = \{T_{i_1}, \ldots, T_{i_s}\}$ for

$$\gamma = \mu^{(m/w!)^{1/w}} \text{ where } \mu = \frac{1}{2^{1/2.47}}.$$

We claim that the lemma holds for the terms in $\mathcal{F}'$ and $Y = \cap_{j=1}^{s} T_{i_j}$. Let $x \in_u \{0,1\}^n$ and let $W = \{i : x_i = 0\}$. Then, each element of U is present in W independently with probability $1/2$. Therefore, as $\mathcal{F}'$ is a γ-sunflower

$$\Pr_x[\vee_{j=1}^{c}(T_{i_j} \setminus Y) = 1] = \Pr_W[\exists T \in \mathcal{F}',\ (T \setminus Y) \cap W = \emptyset\,] \ge 1 - \gamma.$$

□

We next show Fact 2.8.

Proof of Fact 2.8. From the definition of $\gamma(\ )$ from Equation 2.2, it is easy to check that $\gamma(j/2^w) \ge j^{1/w}/10w$. We shall also use the following inequality that follows from partial integration: for any $\theta \ge k \ge 0$,

$$\int_{\theta}^{\infty} x^k e^{-x} dx = \sum_{i=0}^{k} \binom{k}{i} \cdot (i!) \cdot \left(\theta^{k-i} e^{-\theta}\right) \le (k+1)\theta^k \cdot e^{-\theta}. \tag{A.1}$$

Therefore, for $\theta = W^{1/w}/10w$,

$$\begin{aligned}
\sum_{j=W+1}^{\infty} e^{-\gamma(j/2^w)} &\le \sum_{j=W+1}^{\infty} e^{-\left(j^{1/w}/10w\right)} \\
&\le \int_{W}^{\infty} e^{-\left(x^{1/w}/10w\right)}\, dx \\
&= 10w^2 \cdot (10w)^{w-1} \cdot \int_{\theta}^{\infty} y^{w-1} \cdot e^{-y} dy \qquad \text{(substituting } y \equiv x^{1/w}/10w) \\
&\le 10w^2 \cdot (10w)^{w-1} \cdot w \cdot \theta^{w-1} e^{-\theta} \qquad \text{(by Equation A.1)} \\
&\le 10w^3 \cdot W \cdot \exp(-10w^2 \log(1/\varepsilon)) \\
&= \exp\left(\log(10w^3) + w\log 2 + 3w \log w + w \log(50 \log(1/\varepsilon)) - 10w^2 \log(1/\varepsilon)\right) \\
&< \exp(-\log(1/\varepsilon)) = \varepsilon
\end{aligned}$$

where the last inequality can be checked numerically for $w \ge 1$ and $\varepsilon \le 1/4$. □

B Proofs from Section 4

In this section, we prove the two Lemmas from [LV96] that are used in our analysis. We restate them here for the reader's convenience.

Lemma B.1. *(Lemma 4.2 Restated) We have*

$$\forall h \in \mathcal{H}, \quad \mathsf{Bias}(f_h) \leq \mathsf{Bias}(f),$$
$$\mathop{\mathsf{E}}_{h\in\mathcal{H}}[\mathsf{Bias}(f_h)] \geq \mathsf{Bias}(f_h) - \varepsilon.$$

Proof. As f_h is obtained by dropping terms in f, we have $f_h(x) \leq f(x)\ \forall x \in \{0,1\}^n$, so $\mathsf{Bias}(f_h) \leq \mathsf{Bias}(f)$. This also implies that

$$\mathsf{Bias}(f_h) = \frac{1}{2^n}\left(\sum_{x\in f^{-1}(1)} f_h(x)\right). \tag{B.1}$$

Taking expectation over h, we have

$$\mathop{\mathsf{Pr}}_{h\in\mathcal{H}}[\mathsf{Bias}(f_h)] = \frac{1}{2^n}\left(\sum_{x\in f^{-1}(1)} \mathop{\mathsf{Pr}}_{h\in\mathcal{H}}[f_h(x)]\right). \tag{B.2}$$

Fix an $x \in f^{-1}(1)$ and a term T_i of f that it satisfies. If T_i is included in f_h, which happens unless T_i is bad for h, then $f_h(x) = 1$. By Lemma 4.1 and a union bound,

$$\mathop{\mathsf{Pr}}_{h\in\mathcal{H}}[T_i \text{ is bad for } h] \leq t \cdot 2^{-k} \leq \frac{\varepsilon}{w}\cdot\frac{w}{k} \leq \varepsilon.$$

Hence we have

$$\mathop{\mathsf{Pr}}_{h\in\mathcal{H}}[f_h(x)] \geq 1 - \varepsilon.$$

Plugging this into Equation (B.2) gives

$$\mathop{\mathsf{Pr}}_{h\in\mathcal{H}}[\mathsf{Bias}(f_h)] \geq \frac{1}{2^n}\left(\sum_{x\in f^{-1}(1)} (1-\varepsilon)\right) = (1-\varepsilon)\frac{|f^{-1}(1)|}{2^n} = (1-\varepsilon)\mathsf{Bias}(f).$$

□

Lemma B.2. *(Lemma 4.3 restated) We have*

$$|p_h - \mathsf{Bias}(f_h)| \leq \varepsilon.$$

Proof. Let $\mathcal{D}_0$ be the uniform distribution over $\{0,1\}^n$. For $j \in [t]$, let $\mathcal{D}_j$ be the distribution obtained from $\mathcal{D}_{j-1}$ by replacing the uniform distirbution on variables in bucket B_j with an independent copy of output of the generator G. Thus $\mathcal{D}_t$ is the output distribution of G_h.

We claim that for $j \in [t]$,

$$\left| \Pr_{x \in \mathcal{D}_{j-1}} [f_h(x) = 1] - \Pr_{x \in \mathcal{D}_j} [f_h(x) = 1] \right| \leq \delta. \tag{B.3}$$

Since $\mathcal{D}_{j-1}$ and $\mathcal{D}_j$ differ only on the distribution over bucket B_j, we first sample assignments for the other buckets. The resulting formula on the variables in B_j is a DNF with width at most w'. Hence it is δ-fooled by G, which gives Equation (B.3).

We now have

$$\begin{aligned}
|\mathsf{Bias}(f_h) - p_h| &= \left| \Pr_{x \in \mathcal{D}_0} [f_h(x)] - \Pr_{x \in \mathcal{D}_t} [f_h(x)] \right| \\
&\leq \sum_{j=1}^{t} \left| \Pr_{x \in \mathcal{D}_{j-1}} [f_h(x)] - \Pr_{x \in \mathcal{D}_j} [f_h(x)] \right| \\
&\leq t\delta \\
&\leq \varepsilon.
\end{aligned}$$

□

ECCC ISSN 1433-8092

http://eccc.hpi-web.de