Electronic Colloquium on Computational Complexity, Report No. 51 (2013)

Approximating Boolean functions with depth-2 circuits

Eric Blais *
CSAIL
MIT
Cambridge, MA
eblais@csail.mit.edu

Li-Yang Tan †
Department of Computer Science
Columbia University
New York, NY
liyang@cs.columbia.edu

April 2, 2013

Abstract

We study the complexity of approximating Boolean functions with DNFs and other depth-2 circuits, exploring two main directions: universal bounds on the approximability of all Boolean functions, and the approximability of the parity function. In the first direction, our main positive results are the first non-trivial universal upper bounds on approximability by DNFs:

- Every Boolean function can be ε-approximated by a DNF of size $O_\varepsilon(2^n/\log n)$.
- Every Boolean function can be ε-approximated by a DNF of width $c_\varepsilon\, n$, where $c_\varepsilon < 1$.

Our techniques extend broadly to give strong universal upper bounds on approximability by various depth-2 circuits that generalize DNFs, including the intersection of halfspaces, low-degree PTFs, and unate functions. We show that the parameters of our constructions come close to matching the information-theoretic inapproximability of a random function.

In the second direction our main positive result is the construction of an explicit DNF that approximates the parity function:

- PAR_n can be ε-approximated by a DNF of size $2^{(1-2\varepsilon)n}$ and width $(1-2\varepsilon)n$.

Using Fourier analytic tools we show that our construction is essentially optimal not just within the class of DNFs, but also within the far more expressive classes of the intersection of halfspaces and intersection of unate functions.

1 Introduction

The study of the DNF complexity of Boolean functions is one of the great success stories in complexity theory. Among the many remarkably precise results in this area, let us highlight three:

Lupanov's Theorem ([Lup61]). *Any DNF computing the parity function PAR_n has size 2^{n-1} and width n. Furthermore, every Boolean function can be computed by a DNF of size 2^{n-1} and width n so the parity function has the largest DNF size and width complexity of all Boolean functions.*

Quine's Theorem ([Qui54]). *Any DNF computing the majority function MAJ_n has size at least $\binom{n}{n/2}$ and width at least $n/2$. Furthermore,* every *monotone Boolean function can be computed by a DNF of size $\binom{n}{n/2}$ so MAJ_n is the hardest monotone function to compute with respect to DNF size.*

---

*Research supported by a Simons Postdoctoral Fellowship. Part of this research was completed while the author was at CMU.

†Part of this research was completed while the author was visiting CMU.

ISSN 1433-8092

Korshunov–Kuznetsov Theorem ([Kor83, Kuz83][1]). *The optimal DNF size for a random Boolean function is* $(K + o(1))(2^n/\log n \log\log n)$, *where* $1 \le K \le 1.54169$.

Our understanding of the DNF complexity of Boolean functions extends beyond the minimum size of the DNFs computing specific functions or classes of functions. Notably, we have a good understanding of the maximum possible correlation of small-size DNFs with the parity function. Building on a long line of work originally motivated by the goal of showing that PAR_n is not in AC^0 [FSS84, Ajt83, Yao85, Hås86, Cai89, LMN89], and improving on a recent result of Beame *et al.* [BIS12], Impagliazzo *et al.* and Håstad recently showed that any DNF of size s has correlation at most $2^{-\Omega(n/\log s)}$ with PAR_n [IMP12, Hås12].

In this work we are interested in the DNF complexity of *approximating* Boolean functions. Specifically, we say that a DNF ε-*approximates* the function $f : \{0,1\}^n \to \{0,1\}$ if the function $g : \{0,1\}^n \to \{0,1\}$ computed by the DNF satisfies $f(x) = g(x)$ for all but an ε fraction of the inputs $x \in \{0,1\}^n$. The explicit study of the DNF complexity of approximating Boolean functions was initiated by O'Donnell and Wimmer [OW07], who showed that for any constant $\varepsilon > 0$ there is a DNF of size $2^{O(\sqrt{n})}$ that ε-approximates MAJ_n. They also showed that there exists a monotone function for which every DNF that $\frac{1}{100}$-approximates it must have size $2^{\Omega(n/\log n)}$. A comparison of these results with Quine's theorem shows that the DNF complexity of Boolean functions is strikingly different in the exact and approximate computation models: quantitatively, we see that it is possible to approximate the majority function with much smaller DNFs than those that compute majority exactly, and qualitatively we see that, unlike in the exact computation model, the majority function is far from the hardest monotone function to approximate in terms of DNF size.

We continue the study of the DNF complexity of approximating Boolean functions, focusing especially on the analogues of Lupanov's theorem and the Korshunov–Kuznetsov theorem in the approximation model. As we discuss next, our results further illustrate how different DNF complexity can be in exact and approximate computation models.

Universal bounds on approximability We begin with a simple question: is there a non-trivial universal upper bound on the size or width of DNFs for approximating *any* Boolean function? To the best of our knowledge, this question has not been considered explicitly before and a large gap exists between the trivial upper bounds and the known limitations on any potential universal upper bound. In fact, we believe that it is not even known whether there is a Boolean function for which every ε-approximating DNF has size[2] $\Omega_\varepsilon(2^n)$, whether every function can be ε-approximated by a DNF of size $2^{c_\varepsilon n}$ for some $c_\varepsilon < 1$ (*i.e.* whether every function can be approximated by a DNF that is *exponentially smaller* than the trivial DNF), or whether the best universal upper bound lies somewhere in the middle. Likewise, it is unknown whether there is a Boolean function for which every ε-approximating DNF has width $n - O_\varepsilon(1)$, whether every function can be ε-approximated by a DNF of width $c_\varepsilon n$ for some $c_\varepsilon < 1$ (*i.e.* whether every function can be approximated by a union of subcubes that each have dimension *linear* in n) or, once again, whether the best universal upper bound on the width of approximating DNFs is somewhere in between.

We answer these questions in the first part of the paper. Our main positive results are the first non-trivial universal upper bounds on the approximability of all Boolean functions with respect to

[1]The Korshunov–Kuznetsov theorem is the culmination of a long line of research [Gla67, Kor69, Sap72, Kor81, Kor83, Kuz83, Pip03]. For more discussion of the history of this theorem and elegant proofs of its components, we highly recommend Pippenger's article [Pip03].

[2]For clarity of presentation in informal discussions, we use the notation $O_\varepsilon(\cdot)$ and $\Omega_\varepsilon(\cdot)$ to represent asymptotic behaviors when ε is a fixed constant. For the dependencies on ε in the bounds, see the corresponding theorem statements in the main body of the paper.

Complexity measure	Upper bound for all functions	Lower bound for almost all functions
DNF width	$c_\varepsilon n \quad (c_\varepsilon < 1)$	$\Omega(n)$
DNF size	$O\big(2^n/\log n\big)$	$\Omega\left(2^n/n\right)$
AND of halfspaces	$(1+o(1)) \cdot 2^n/n$	$\Omega\left(2^n/n^2\right)$
AND of unate functions	$2^{c_\varepsilon n} \quad (c_\varepsilon < 1)$	$2^{\Omega(n)}$ for $\varepsilon < \frac{1}{16}$
AND of degree-d PTFs	$O\left(2^n/n^d\right)$	$\Omega\left(2^n/n^{d+1}\right)$

Table 1: Our universal bounds on approximating any Boolean function to constant accuracy.

both DNF size and width. Our universal bound on DNF width is asymptotically optimal, and we accomplish this via a connection between approximating DNFs and low-density coverings of the Boolean hypercube by Hamming balls. We show that this technique extends rather broadly to give universal bounds on approximability by various generalizations of DNFs, including the intersection of halfspaces, low-degree PTFs, and unate functions. We complement these upper bounds with near-matching information-theoretic lower bounds against a random function.

Approximating the parity function In the second part of the paper we turn our attention to the parity function. Despite decades of intensive study of the circuit complexity of this function, large gaps remain between the minimal size and width of DNFs that ε-approximate PAR_n for constant values of $\varepsilon > 0$. For instance, while the random restriction method shows that any DNF that ε-approximates PAR_n must have size $2^{\Omega_\varepsilon(n)}$, the precise dependence on ε in this bound is unclear, leaving open a wide range of possibilities. On one extreme, it is possible that the true lower bound is $\Omega_\varepsilon(2^n)$ (so that we only gain a linear savings on the size of DNFs by requiring them to approximate parity rather than compute it directly) and on the other it is possible that the true bound is $2^{c_\varepsilon n}$ for some $c_\varepsilon < 1$ (so that DNFs that approximate PAR_n are *exponentially* smaller than those that compute the same function exactly).

We resolve this question by showing, perhaps somewhat surprisingly, that the right answer falls in the latter extreme. Our construction of an explicit DNF approximator for PAR_n, when combined with information-theoretic lower bounds against a random function, also shows that the landscape of DNF complexity changes dramatically when we move from exact to approximate computation: while PAR_n is the hardest function to compute exactly by DNFs, it can be approximated by DNFs of size exponentially smaller than that required for almost every other function. The remainder of the paper is then devoted to proving the optimality of our DNF approximator for PAR_n. Using Fourier analytic tools, we show that the dependence on ε in our DNF construction is essentially the best possible, and furthermore, our construction is near-optimal even within the far more expressive classes of the intersection of halfspaces and the intersection of unate functions.

1.1 Our results

Universal bounds Our first result is the first non-trivial universal upper bound on the size of DNF approximators for all Boolean functions.

Theorem 1. *Every Boolean function can be ε-approximated by a DNF of size $O_\varepsilon(2^n/\log n)$.*

The proof of Theorem 1, presented in Section 3, is obtained with a randomized algorithm that constructs an explicit approximating DNF. In Section 4 we complement Theorem 1 with an asymptotically optimal universal upper bound on the width of DNFs required to approximate any function.

Theorem 2. *Every Boolean function can be ε-approximated by a DNF of width $c_\varepsilon n$, where $c_\varepsilon < 1$ depends only on ε.*

Theorem 2 highlights an interesting (and stark) contrast between exact and approximate computation with respect to DNF width: not only is the DNF width of a random Boolean function at least $n - \log(3n)$, *every* term in any DNF computing it has to have width at least $n - \log(3n)$ (see *e.g.* Theorem 3.21 of [CH11]). Therefore, while every 1-monochromatic subcube in a random function has dimension at most $\log(3n)$, Theorem 2 shows that every Boolean function can be ε-approximated by the union of 1-monochromatic subcubes *all* of which have dimension $\Omega_\varepsilon(n)$.

Theorem 2 is obtained by exploiting a connection between approximating DNFs and low-density coverings of the Boolean cube by Hamming balls. This technique extends rather broadly to give strong universal bounds on approximability by the intersection of unate functions and low-degree PTFs.

Theorem 3. *Every Boolean function can be ε-approximated by the intersection of $2^{c_\varepsilon n}$ unate functions, where $c_\varepsilon < 1$ depends only on ε.*

Theorem 4. *For every positive integer d, every Boolean function can be $O(1/n)$-approximated by the intersection of $O(2^n/n^d)$ degree-d PTFs.*

Using a theorem of Kabatyanski and Pachenko on the existence of asymptotically perfect covering codes of radius 1 [KP88], we obtain improvements on both the accuracy and size of our approximators in Theorem 4 when $d = 1$ (*i.e.* the intersection of LTFs). See Section 4 for the details.

In Section 5 we turn our attention to lower bounds, giving a lower bound on the size complexity of approximating DNFs for almost all functions that nearly matches the universal upper bound of Theorem 1.

Theorem 5. *For almost every Boolean function f, any DNF that ε-approximates f has size $\Omega_\varepsilon(2^n/n)$.*

The proof of Theorem 5 is obtained by extending Pippenger's elegant information-theoretic proof [Pip03] of Kuznetsov's theorem showing that the DNF size of a random Boolean function is at least $(1+o(1))(2^n/\log n \log\log n)$ [Kuz83]. Notably, our extension is rather general, and also implies strong bounds on the inapproximability of a random function by other types of depth-2 circuits; due to space considerations we do not list the associated corollaries here. We remark that our construction of small-width DNF approximators in Theorem 2 is asymptotically optimal. We prove in the second part of the paper that any ε-approximator for PAR_n must have width at least $(1-2\varepsilon)n$, and also show that the same proof extends easily to show that almost every function has ε-approximating DNF width $\Omega_\varepsilon(n)$.

Approximating the parity function In Section 6 we turn our focus to the complexity of approximating the parity function. We begin with a deterministic construction of a DNF that approximates PAR_n with one-sided error.

Theorem 6. *The parity function can be ε-approximated by a DNF f of width $(1-2\varepsilon)n$ and size $2^{(1-2\varepsilon)n}$. Furthermore, f has one-sided error: if $\mathsf{PAR}_n(x) = 1$ then $f(x) = 1$.*

Complexity measure	Upper bound	Lower bound
DNF width	$(1-2\varepsilon)n$	$(1-2\varepsilon)n$
DNF size	$2^{(1-2\varepsilon)n}$	$\max_{\delta>0}\{\delta\, 2^{(1-2\varepsilon-2\delta)n},\ (\frac{1}{2}-\varepsilon)2^{\frac{1-2\varepsilon}{1+2\varepsilon}n}\}$
AND of halfspaces	$2^{(1-2\varepsilon)n}$	$2^{\Omega_\varepsilon(n)}$ assuming Klivans *et al.*'s conjecture
AND of unate functions	$2^{(1-2\varepsilon)n}$	$2^{\Omega(n)}$ for $\varepsilon < \frac{1}{16}$

Table 2: Our bounds for approximating PAR_n to accuracy ε.

We point out the interesting contrast between this upper bound on size and the lower bound of Theorem 5: although PAR_n is the hardest function to compute exactly with respect to DNF size, Theorems 6 and 5 together show that it is in fact *exponentially* easier to approximate than *almost every* other function. We prove the optimality of our construction by giving matching lower bounds on the size and width of DNF approximators for PAR_n.

Theorem 7. *Any DNF that ε-approximates PAR_n has width at least $(1-2\varepsilon)n$.*

Theorem 8. *Any DNF that ε-approximates PAR_n has size at least*

$$s \geq \max\left\{\max_{\delta>0} \delta\, 2^{(1-2\varepsilon-2\delta)n}, \left(\tfrac{1}{2}-\varepsilon\right) 2^{\frac{1-2\varepsilon}{1+2\varepsilon}n}\right\}.$$

The width lower bound is obtained by applying Amano's bound on the total influence of small-width DNFs [Ama11], and the first size lower bound is obtained by combining Amano's theorem with an elementary truncation argument. The second lower bound on size uses a sharpening of Boppana's bound on the total influence of small-size DNFs [Bop97], obtained in concurrent work by the present authors using the entropy method [BTW13].

In Section 7 we provide further evidence of the optimality of our DNF approximators for PAR_n. Assuming a noise sensitivity conjecture of Klivans *et al.* [KOS04], we prove that ε-approximating PAR_n even with the intersection of halfspaces requires size $2^{\Omega_\varepsilon(n)}$, matching the size of our DNF approximators in Theorem 6.[3]

Theorem 9. *Assume the KOS conjecture holds. Let f be computed by the intersection of k halfspaces, and suppose f ε-approximates PAR_n. Then $k = 2^{\Omega_\varepsilon(n)}$.*

Naturally we would like an unconditional proof of Theorem 9. We are able to accomplish this for all $\varepsilon < \frac{1}{16}$, and in fact, our proof holds against the more expressive class of the intersection of unate functions.

Theorem 10. *Fix $\varepsilon < \frac{1}{16}$. Let f be computed by the intersection of k unate functions and suppose f ε-approximates PAR_n. Then $k = 2^{\Omega(n)}$.*

2 Preliminaries

All probabilities and expectations are with respect to the uniform distribution and logarithms are base 2 unless otherwise stated. For strings $x, y \in \{0,1\}^n$, we write $\mathrm{dist}(x,y)$ to denote the

[3] Since the class of halfspaces is closed under negation, universal bounds on approximability by the intersection of halfspaces immediately imply identical universal bounds for the disjunction of halfspaces, a strict superclass of DNFs. Likewise for the intersection of unate functions, a further generalization of the intersection of halfspaces.

Hamming distance between x and y. We write $\mathrm{Vol}(d)$ to denote the quantity $\sum_{i=0}^{d} \binom{n}{d}$, the volume of a Hamming ball of radius d.

We say that a Boolean function f is *sensitive* at coordinate $i \in [n]$ on input x if $f(x^{i=0}) \neq f(x^{i=1})$, where $x^{i=b}$ is x with its i-th coordinate set to b. We write $s(f, x, i)$ to denote the indicator for this event, and $s(f, x)$ to denote $\sum_{i=1}^{n} s(f, x, i)$. We say that f is *monotone* in direction i if $f(x^{i=0}) \leq f(x^{i=1})$ for all x, and *anti-monotone* in direction i if $f(x^{i=0}) \geq f(x^{i=1})$ for all x. A Boolean function f is *unate* if for all $i \in [n]$, f is either monotone or anti-monotone in direction i.

The *subcube* $C \subseteq \{0,1\}^n$ corresponding to the pair of disjoint sets $S_0, S_1 \subseteq [n]$ is the set of elements $x \in \{0,1\}^n$ for which $x_i = 0$ for every $i \in S_0$ and $x_i = 1$ for every $i \in S_1$. The *free coordinates* of C are the coordinates in $[n] \setminus (S_0 \cup S_1)$. The subcube C is 1-*monochromatic* with respect to f if $f(x) = 1$ for every x in C. Each term in a DNF corresponds to a 1-monochromatic subcube, and so a DNF may be viewed geometrically as a union of 1-monochromatic subcubes.

A degree-d *polynomial threshold function* (PTF) is a Boolean function $f(x) = \mathrm{sgn}(p(x))$, where $p : \{0,1\}^n \to \mathbb{R}$ is a degree-d polynomial. If $d = 1$ we refer to f as a *linear threshold function* (LTF), or a *halfspace*. It is straightforward to verify that halfspaces are unate.

2.1 Fourier analysis over the Boolean hypercube

Every Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ has a Fourier expansion $f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \chi_S(x)$, where the numbers $\widehat{f}(S) \in [-1,1]$ are the Fourier coefficients of f. We write $\mathbf{W}_k[f] = \sum_{|S|=k} \widehat{f}(S)^2$ to denote the total Fourier weight of f at level k.

Definition 11. The influence of coordinate $i \in [n]$ on a Boolean function f, denoted $\mathbf{Inf}_i[f]$, is the probability $\mathbf{Pr}[f(\boldsymbol{x}) \neq f(\boldsymbol{x}^{\oplus i})]$, where $x^{\oplus i}$ denotes x with its i-th bit flipped. The total influence of f is $\mathbf{Inf}[f] = \sum_{i=1}^{n} \mathbf{Inf}_i[f]$.

Definition 12. The noise sensitivity of a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ at noise rate δ, denoted $\mathbf{NS}_\delta[f]$, is the probability $\mathbf{Pr}[f(\boldsymbol{x}) \neq f(\boldsymbol{y})]$, where $\boldsymbol{x} \sim \{-1,1\}^n$ is uniformly random and $\boldsymbol{y}$ is obtained from $\boldsymbol{x}$ by flipping each coordinate independently with probability δ.

Both total influence and noise sensitivity have well-known Fourier formulas:

$$\begin{aligned}
\mathbf{Inf}[f] &= \sum_{S \subseteq [n]} |S| \cdot \widehat{f}(S)^2. \\
\mathbf{NS}_\delta[f] &= \frac{1}{2} \sum_{k=0}^{n} (1 - (1-2\delta)^k) \cdot \mathbf{W}_k[f].
\end{aligned}$$

3 Universal upper bound on DNF size

In this section we prove that every Boolean function can be ε-approximated by a DNF of size $O_\varepsilon(2^n / \log n)$. The proof of this result is obtained via a randomized construction that, given a function f, constructs an explicit approximating DNF for f of the required size.

Before presenting the construction, let us first informally describe the intuition behind it. In order to build a good approximator for f, the construction must identify a small family of subcubes that (i) cover almost all of the inputs $x \in \{0,1\}^n$ for which $f(x) = 1$, and (ii) cover almost none of the inputs x for which $f(x) = 0$. For most functions f, these constraints are roughly equivalent to the requirement that the subcubes in the family should have relatively small overlap with each other over $f^{-1}(1)$ while having large overlap with each other over $f^{-1}(0)$.

Our construction meets these apparently conflicting requirements with a two-stage process. In the first stage, the algorithm selects a (small) random subset S of $f^{-1}(0)$ and defines the random function $\boldsymbol{g}$ to take the value 1 on every input in $f^{-1}(1) \cup S$. The second stage selects a random subset of the large subcubes that are 1-monochromatic in $\boldsymbol{g}$. The union of those subcubes corresponds to a small DNF that computes a function $\boldsymbol{h}$ that is close to f provided that S is small enough (in which case constraint (ii) is satisfied) and that most elements in $f^{-1}(1)$ are covered by many large subcubes that are 1-monochromatic in $\boldsymbol{g}$ (in which case constraint (i) is also satisfied). As we see below, with the right parameter settings, we can guarantee that both those events happen with large probability.

Theorem 1. *Let $\varepsilon \geq 10/n$.*[4] *Every Boolean function $f : \{0,1\}^n \to \{0,1\}$ can be ε-approximated by a DNF of size $4\ln(4/\varepsilon) \cdot 2^{n-d}$ and width $n-d$, where*

$$d = \log\log_{2/\varepsilon}\left(\frac{n}{\ln(4/\varepsilon)\log\log_{2/\varepsilon} n}\right).$$

That is, every f can be ε-approximated by a DNF of size $O_\varepsilon(2^n/\log n)$ and width $n - \Omega_\varepsilon(\log\log n)$.

Proof. We may assume that $\min\{\mathbf{Pr}[f(\boldsymbol{x}) = 0], \mathbf{Pr}[f(\boldsymbol{x}) = 1]\} \geq \varepsilon$, since otherwise f is ε-close to constant and the claim is trivially true. Let $\boldsymbol{g} : \{0,1\}^n \to \{0,1\}$ be the random function obtained by setting $\boldsymbol{g}(x) = 1$ for all $x \in f^{-1}(1)$, and for each $x \in f^{-1}(0)$, we independently set $\boldsymbol{g}(x) = 1$ with probability $\varepsilon/2$ and $\boldsymbol{g}(x) = 0$ otherwise. Let $\mathcal{G}$ denote the induced distribution over all Boolean functions. Since $\mathbf{E}_{\mathcal{G}}[\mathbf{Pr}_{x\in f^{-1}(0)}[\boldsymbol{g}(x) = 1]] = \varepsilon/2$, we apply the Chernoff bound to deduce that

$$\mathop{\mathbf{Pr}}_{\mathcal{G}}\left[\mathop{\mathbf{Pr}}_{f^{-1}(0)}[f(x) \neq \boldsymbol{g}(x)] \geq \varepsilon\right] \leq e^{-\varepsilon^2 \cdot 2^n/3}. \tag{1}$$

Let us call a subcube C *special* if C has dimension exactly d and the d free coordinates of C are $\{dk+1, \ldots, dk+d\}$ for some $k = 0, \ldots, \lfloor n/d \rfloor - 1$. There are $\lfloor n/d \rfloor \cdot 2^{n-d}$ special subcubes in total, and every $x \in \{0,1\}^n$ is contained in exactly $\lfloor n/d \rfloor$ special subcubes. Let $\boldsymbol{h} : \{0,1\}^n \to \{0,1\}$ be the union of a random subset of the 1-monochromatic special subcubes in $\boldsymbol{g}$ where each 1-monochromatic special subcube C in $\boldsymbol{g}$ is included in $\boldsymbol{h}$ with probability $(\varepsilon/2)^{\#\{x\in C\,:\,f(x)=1\}}$. The probability, over the randomness of $\boldsymbol{g}$ and $\boldsymbol{h}$, that a fixed special subcube C is included in $\boldsymbol{h}$ is therefore exactly $(\varepsilon/2)^{2^d}$, since the probability that C is 1-monochromatic in $\boldsymbol{g}$ is $(\varepsilon/2)^{\#\{x\in C\,:\,f(x)=0\}}$, and the probability that C is then included in $\boldsymbol{h}$ is $(\varepsilon/2)^{\#\{x\in C\,:\,f(x)=1\}}$. Note that $\boldsymbol{h}$ is a DNF of width $n-d$, and $\boldsymbol{h}^{-1}(1) \subseteq \boldsymbol{g}^{-1}(1)$; in particular, the error of $\boldsymbol{h}$ on the 0-inputs of f is at most that of $\boldsymbol{g}$, and (1) remains true with $\boldsymbol{h}$ in place of $\boldsymbol{g}$. Next we argue that

$$\mathop{\mathbf{Pr}}_{\mathcal{G}}\left[\mathop{\mathbf{Pr}}_{f^{-1}(1)}[f(x) \neq \boldsymbol{h}(x)] \geq \varepsilon\right] \leq 1/4. \tag{2}$$

Fix $x \in f^{-1}(1)$. The probability that $\boldsymbol{h}(x) = 0$ (*i.e.* $\boldsymbol{h}(x) \neq f(x)$) is the probability that none of the $\lfloor n/d \rfloor$ special subcubes containing x are included in $\boldsymbol{h}$. Since any two of these $\lfloor n/d \rfloor$ special subcubes intersect only at x, their inclusion probability are independent and so we have

$$\begin{aligned} \mathop{\mathbf{Pr}}_{\mathcal{G}}[\boldsymbol{h}(x) = 0] &= \left(1 - \left(\frac{\varepsilon}{2}\right)^{2^d}\right)^{\lfloor n/d \rfloor} \\ &\leq e^{-(\varepsilon/2)^{2^d} n/d} < \varepsilon/4, \end{aligned}$$

[4] Our lower bounds in Section 6 imply that any DNF that ε-approximates PAR_n has size $\Omega(2^n)$ and width $n - O(1)$ when $\varepsilon = O(1/n)$, and so the universal upper bounds are only of interest for $\varepsilon = \omega(1/n)$.

where we have used our choice of d in the final inequality. This gives us

$$\mathop{\mathbf{E}}_{\mathcal{G}}\left[\mathop{\mathbf{Pr}}_{f^{-1}(1)}[f(x) \neq \boldsymbol{h}(x)]\right] < \varepsilon/4,$$

and so

$$\mathop{\mathbf{Pr}}_{\mathcal{G}}\left[\mathop{\mathbf{Pr}}_{f^{-1}(1)}[f(x) \neq \boldsymbol{h}(x)] \geq \varepsilon\right] \leq 1/4,$$

matching the claimed bound in (2) above. It remains to bound the DNF size of $\boldsymbol{h}$. Since there are $\lfloor n/d \rfloor \cdot 2^{n-d}$ special subcubes, and each is included with probability $(\varepsilon/2)^{2^d}$, we have

$$\begin{aligned}\mathop{\mathbf{E}}_{\mathcal{G}}[\text{DNF-size}[\boldsymbol{h}]] &= \left(\frac{\varepsilon}{2}\right)^{2^d}\left\lfloor\frac{n}{d}\right\rfloor 2^{n-d} \\ &\leq \frac{\ln(4/\varepsilon)\log\log_{2/\varepsilon} n}{d} \cdot 2^{n-d} \\ &\leq 2\ln(4/\varepsilon)\cdot 2^{n-d}.\end{aligned}$$

Here in the final inequality we use the fact that $d \geq (\log\log_{2/\varepsilon} n) - 1$ for n sufficiently large (which in turn holds since $\ln(4/\varepsilon)\log\log_{2/\varepsilon} n < \sqrt{n}$). Again, we apply Markov's inequality to say that

$$\mathop{\mathbf{Pr}}_{\mathcal{G}}\left[\text{DNF-size}[\boldsymbol{h}] \geq 4\ln(4/\varepsilon)\cdot 2^{n-d}\right] \leq 1/2. \tag{3}$$

Taking a union bound over (1), (2), and (3), we conclude that there must exist some h such that DNF-size$[h] \leq 4\ln(4/\varepsilon)\cdot 2^{n-d}$, DNF-width$[h] = n-d$, and $\mathbf{Pr}[f(x) \neq h(x)] \leq \varepsilon$ and this completes the proof. □

4 Approximation via Hamming ball covers of the hypercube

In this section we introduce a general method for constructing small depth-2 circuits with top gate OR that approximate an arbitrary Boolean function f,[5] to which there are three main components. We first show that for any radius d, all but an ε fraction of $\{0,1\}^n$ can be covered with $O_\varepsilon(2^n/\mathrm{Vol}(d))$ Hamming balls of radius d. Next, we approximate f restricted to each Hamming ball with the desired second layer gate (*e.g.* an LTF, degree-d PTF, or unate function[6]); these sub-approximators approximate f to high accuracy within the ball, and label all points outside the ball 0. Finally, our overall approximator for f is simply the $O_\varepsilon(2^n/\mathrm{Vol}(d))$-wise disjunction of these sub-approximators. We begin with a short proof of the first claim:

Lemma 4.1. *For every $\varepsilon > 0$ and $d > 0$ there is a collection of $\ln(1/\varepsilon)(2^n/\mathrm{Vol}(d))$ Hamming balls of radius d that covers all but an ε fraction of $\{0,1\}^n$.*

Proof. Let $\mathcal{C}$ be a random collection of $\ln(1/\varepsilon)(2^n/\mathrm{Vol}(d))$ Hamming balls of radius d with centers picked uniformly at random with replacement from $\{0,1\}^n$. For any $x \in \{0,1\}^n$, the probability that x is not covered by $\mathcal{C}$ is $(1-\mathrm{Vol}(d)\cdot 2^{-n})^{\ln(1/\varepsilon)(2^n/\mathrm{Vol}(d))} \leq \exp(-\ln(1/\varepsilon)) = \varepsilon$. Therefore $\mathcal{C}$ covers all but an ε fraction of points in $\{0,1\}^n$ in expectation, and so there is a collection of $\ln(1/\varepsilon)(2^n/\mathrm{Vol}(d))$ many Hamming balls of radius d that covers all but an ε fraction of $\{0,1\}^n$. □

[5] As alluded to in the introduction, since the classes of LTFs, degree-d PTFs, and unate functions are each closed under negation, if f can be approximated by a disjunction of k of them then $\neg f$ can be approximated by an intersection (*i.e.* conjunction) of k of them. We focus our exposition on depth-2 approximators with top gate OR, noting that they immediately imply the existence of universal approximators with top gate AND of the same size.

[6] For our construction of DNF approximators the sub-approximators are themselves DNFs instead of conjunctions; we may do this since the disjunction of DNFs is itself a DNF.

4.1 Approximating Boolean functions restricted to Hamming balls

Next we construct sub-approximators for Boolean functions restricted to Hamming balls. To be precise, when we write "f restricted to a Hamming ball B of radius d", we mean the function f_B that agrees with f on all points within B and takes value 0 on all points outside B. Since the bulk of the points in B lie on its surface, our sub-approximators may err on all the points in the interior of B (*i.e.* the points at distance $< d$ from the center of B). We begin with sub-approximators that are small-width DNFs:

Proposition 4.2. *Let $z \in \{0,1\}^n$ and $d \in [n]$. Let $f : \{0,1\}^n \to \{0,1\}$ be the characteristic function of a subset of the Hamming ball of radius d centered at z. There is a DNF g of width $n-d$ satisfying $g(y) = f(y)$ for all y such that $\mathrm{dist}(y,z) = d$, and $g(y) = f(y) = 0$ for all y such that $\mathrm{dist}(y,z) > d$.*

Proof. For each y such that $\mathrm{dist}(y,z) = d$ and $f(y) = 1$ we include a term T_y in the DNF g defined as follows: T_y is the conjunction of literals ℓ_i for each $i \in [n]$ such that $y_i \neq z_i$, where $\ell_i = x_i$ if $z_i = 1$, and $\neg x_i$ otherwise. The key property of T_y is that it accepts only y among all $\binom{n}{d}$ inputs at distance exactly d from the center z, and it rejects any input at distance greater than d from z. Since each term T_y has width exactly $n-d$, it follows that g is a DNF of width $n-d$ that satisfies the claimed properties. $\square$

Note that the DNFs g constructed in Proposition 4.2 are unate since no variable occurs both positively and negated in them; if the literal ℓ_i occurs in g then $\neg \ell_i$ does not occur in g. This observation yields the following corollary of Proposition 4.2:

Corollary 4.3. *Let $z \in \{0,1\}^n$ and $d \in [n]$. Let $f : \{0,1\}^n \to \{0,1\}$ be the characteristic function of a subset of the Hamming ball of radius d centered at z. There is a unate function g satisfying $g(y) = f(y)$ for all y such that $\mathrm{dist}(y,z) = d$, and $g(y) = f(y) = 0$ for all y such that $\mathrm{dist}(y,z) > d$.*

Finally we construct sub-approximators that are low-degree PTFs. These sub-approximators have the nice feature that they have one-sided error.

Proposition 4.4. *Let $z \in \{0,1\}^n$ and $d \in [n]$. Let $f : \{0,1\}^n \to \{0,1\}$ be the characteristic function of a subset of the Hamming ball of radius d centered at z. There is a degree-d PTF $g(x) = \mathrm{sgn}(p(x))$ satisfying $g(y) = f(y)$ for all y such that $\mathrm{dist}(y,z) = d$, and $g(y) = f(y) = 0$ for all y such that $\mathrm{dist}(y,z) > d$. Furthermore, $g(y) = 1$ for all y such that $\mathrm{dist}(y,z) < d$.*

Proof. The polynomial $p(x)$ will be $L(x) + D(x) - \theta$, where $\theta = n - d + \frac{1}{2\binom{n}{d}}$,

$$L(x) = \sum_{i \in [n]\,:\,z_i = 1} x_i + \sum_{i \in [n]\,:\,z_i = 0} (1 - x_i),$$

and

$$D(x) = \frac{1}{2\binom{n}{d}} \sum_{\substack{y\,:\,\mathrm{dist}(y,z) = d \\ f(y) = 1}} \mathbf{1}_{\mathrm{differ}(y,z)}(x),$$

where $\mathbf{1}_{\mathrm{differ}(y,z)}(x) = 1$ iff x agrees with y on the coordinates that y and z differ on (and 0 otherwise). For any y such that $\mathrm{dist}(y,z) = d$, the indicator $\mathbf{1}_{\mathrm{differ}(y,z)}$ is a function of d variables and hence is computed by a degree-d polynomial. Note that $L(y) = n - \mathrm{dist}(y,z)$ for all $y \in \{0,1\}^n$. Since $D(y) \in [0, 1/2]$ for all $y \in \{0,1\}^n$ and the threshold θ is set to be $n - d + \frac{1}{2\binom{n}{d}}$, it follows that

$p(y) > 0$ for all y such that $\mathrm{dist}(y, z) > d$, and $p(y) < 0$ for all y such that $\mathrm{dist}(y, z) < d$. Finally for all y such that $\mathrm{dist}(y, z) = d$, we have $D(y) = 0$ if $f(y) = 0$ and $D(y) = \frac{1}{2\binom{n}{d}}$ otherwise (since $\mathbf{1}_{\mathrm{differ}(y,z)}(y) = 1$), and so p takes the correct sign on these inputs. □

4.2 Combining the sub-approximators

To combine the sub-approximators from the last section, we use the following bound on the surface-to-volume ratio of Hamming balls.

Lemma 4.5. *Fix $1 \leq d \leq n$. Then $\mathrm{Vol}(d-1) \leq \frac{d}{n-d+1}\mathrm{Vol}(d)$. In particular, if $d = \rho n$ for some $\rho \leq 1/2$, then $\mathrm{Vol}(d-1) \leq 2\rho \cdot \mathrm{Vol}(d)$.*

Proof. Let $\boldsymbol{x} \in \{0,1\}^n$ be drawn uniformly at random from the Hamming ball of radius $d-1$ around 0, let $\mathbf{i} \in [n]$ be drawn uniformly at random from the coordinates j for which $x_j = 0$, and let $\boldsymbol{y} \in \{0,1\}^n$ be the input obtained by flipping the $\mathbf{i}$-th coordinate of $\boldsymbol{x}$.

By the chain rule, the joint entropy of $\boldsymbol{x}$, $\boldsymbol{y}$, and $\mathbf{i}$ is

$$\begin{aligned} H(\boldsymbol{x}, \boldsymbol{y}, \mathbf{i}) &= H(\boldsymbol{x}) + H(\mathbf{i} \mid \boldsymbol{x}) + H(\boldsymbol{y} \mid \boldsymbol{x}, \mathbf{i}) \\ &\geq \log \mathrm{Vol}(d-1) + \log(n-d+1) \end{aligned} \tag{4}$$

since $\boldsymbol{x}$ is drawn uniformly at random from a set of size $\mathrm{Vol}(d-1)$, $\mathbf{i}$ is drawn uniformly from a set of size at least $n-d+1$, and $\boldsymbol{y}$ is completely determined by $\boldsymbol{x}$ and $\mathbf{i}$. A different application of the chain rule also yields

$$\begin{aligned} H(\boldsymbol{x}, \boldsymbol{y}, \mathbf{i}) &= H(\boldsymbol{y}) + H(\mathbf{i} \mid \boldsymbol{y}) + H(\boldsymbol{x} \mid \boldsymbol{y}, \mathbf{i}) \\ &\leq \log \mathrm{Vol}(d) + \log d \end{aligned} \tag{5}$$

since the support of $\boldsymbol{y}$ has size $\mathrm{Vol}(d)$, the support of $\mathbf{i}$ has size at most d, and $\boldsymbol{x}$ is completely determined by $\boldsymbol{y}$ and $\mathbf{i}$. Combining (4) and (5) and re-arranging the terms completes the proof. □

Theorem 2. *Every Boolean function f can be ε-approximated by a DNF of width $(1-\rho)n$, where $\rho = \varepsilon/(4\ln(2/\varepsilon))$. In particular, every Boolean function can be 0.01-approximated by a DNF of width $c \cdot n$, where $c < 1$ is an absolute constant.*

Proof. Let $d = \rho n$ where $\rho \leq 1/2$ will be chosen later. By Lemma 4.1, there is a collection of $\ln(2/\varepsilon)(2^n/\mathrm{Vol}(d))$ Hamming balls of radius d that cover all but an $\varepsilon/2$ fraction of $\{0,1\}^n$. We approximate f restricted to each Hamming ball with the DNF of width $n-d$ given by Proposition 4.2. Note that the disjunction g of these DNFs is itself a DNF of width $n-d$, and

$$\begin{aligned} \mathbf{Pr}[f(x) \neq g(x)] &\leq \frac{\varepsilon}{2} + \ln(2/\varepsilon)\frac{\mathrm{Vol}(d-1)}{\mathrm{Vol}(d)} \\ &\leq \frac{\varepsilon}{2} + 2\rho\ln(2/\varepsilon), \end{aligned}$$

where the final inequality is by Lemma 4.5. Here the first inequality is a union bound over the $\varepsilon/2$ fraction of uncovered points and the error of the $\ln(2/\varepsilon)(2^n/\mathrm{Vol}(d))$ sub-approximators, each of which errs on at most $\mathrm{Vol}(d-1)$ points. It suffices to ensure that $2\rho\ln(2/\varepsilon) \leq \varepsilon/2$, and so we may take $\rho = \varepsilon/(4\ln(2/\varepsilon))$. □

As noted in Corollary 4.3 the DNF sub-approximators in Theorem 2 are unate. Viewing our overall approximator as a disjunction of unate functions instead of a disjunction of DNFs gives us the following:

Theorem 3. *Every Boolean function f can be ε-approximated by the disjunction (equivalently, the intersection) of $\ln(2/\varepsilon)\, 2^{(1-H(\rho))n}$ unate functions, where $\rho = \varepsilon/(4\ln(2/\varepsilon))$. In particular, every Boolean function can be 0.01-approximated by the intersection of $O(2^{c\,n})$ unate functions, where $c < 1$ is an absolute constant.*

Using the PTF sub-approximators of Proposition 4.4 in place of the DNF sub-approximators of Proposition 4.2, an identical proof to the one for Theorem 2 yields approximators that are the intersection of degree-d PTFs:

Theorem 4. *Let $d = \rho n$ where $\rho \leq 1/2$. Every Boolean function f can be $(\varepsilon + O(\ln(1/\varepsilon)\rho))$-approximated by the disjunction (equivalently, the intersection) of $\ln(1/\varepsilon)(2^n/\mathrm{Vol}(d))$ degree-d PTFs. In particular, for any constant d every Boolean function can be $O(1/n)$-approximated by the intersection of $O(2^n/n^d)$ degree-d PTFs.*

Improvements via covering codes. It is natural to ask if Lemma 4.1 can be improved. The strongest possible improvement is a covering of all of $\{0,1\}^n$ with $(1+o(1))(2^n/\mathrm{Vol}(d))$ Hamming balls of radius d, a covering code with efficiency asymptotically approaching that of a perfect code. This is a longstanding open problem in the field of covering codes [CHLL05], and even the $d = 2$ case remains open. The $d = 1$ case was resolved in the affirmative by Kabatyanski and Pachenko [KP88]:

Theorem 13 (Kabatyanski and Pachenko)**.** *All of $\{0,1\}^n$ can be covered by $(1 + o(1))(2^n/n)$ Hamming balls of radius 1.*

Using Theorem 13 in place of the approximate cover given by Lemma 4.1 allows us to sharpen the parameters of Theorem 4 in the case of $d = 1$ (*i.e.* intersection of LTFs). As an added bonus, since all of $\{0,1\}^n$ is covered and our LTF sub-approximators in Proposition 4.4 have one-sided error within each Hamming ball, our overall approximator has one-sided error as well.

Theorem 14. *For every Boolean function $f : \{0,1\}^n \to \{0,1\}$ there is a Boolean function g computed by the disjunction (equivalently, the intersection) of $(1 + o(1))(2^n/n)$ halfspaces that $((1 + o(1))/n)$-approximates f. Furthermore, these approximators have one sided error: $g(x) = 1$ whenever $f(x) = 1$.*

A generalization of Theorem 13 for Hamming balls of larger radii $d > 1$ would imply analogous improvements of Theorem 4 for the intersection of degree-d PTFs. The current best bound for general d is due to Krivelevich *et al.* who show that all of $\{0,1\}^n$ can be covered with $O(d \log d)(2^n/\mathrm{Vol}(d))$ many Hamming balls of radius d [KSV03]; however, using this in place of Lemma 4.1 does not yield any improvements on the parameters of our constructions.

5 Inapproximability of a random function

We write $\boldsymbol{f}_n$ to denote a uniformly random Boolean function with arity n.

Theorem 15. *Let $F_1, F_2, F_3, \ldots$ be an infinite sequence of Boolean functions indexed by $\mathbb{N}$, where F_k has arity k. Let $\mathcal{C}$ be a class of Boolean functions. For any $\varepsilon > 0$ and $f : \{0,1\}^n \to \{0,1\}$, let $\mathrm{opt}_\varepsilon[f]$ denote the smallest k such that there exists $g_1, \ldots, g_k \in \mathcal{C}$ and $F_k(g_1, \ldots, g_k)$ is an ε-approximator for f. Then*

$$\mathbf{E}[\mathrm{opt}_\varepsilon[\boldsymbol{f}_n]] \geq (1 - H(\varepsilon)) \cdot \frac{2^n}{\log(e \cdot |\mathcal{C}|)}.$$

Proof. Let $t = |\mathcal{C}|$, and fix a numbering $g_1, \ldots, g_t$ of the t Boolean functions in $\mathcal{C}$. Let $\mathbf{X} = \langle \mathbf{X}_1, \ldots, \mathbf{X}_t \rangle$ be a vector of indicator random variables where $\mathbf{X}_i = 1$ iff g_i occurs in the optimal ε-approximator $F_k(g_1, \ldots, g_k)$ for $\boldsymbol{f}_n$ (if there are multiple ε-approximators achieving minimal size we fix an arbitrary one to be the optimal).

Since $\boldsymbol{f}_n$ determines $\mathbf{X}$ and $H(\boldsymbol{f}_n) = 2^n$, we have $H(\boldsymbol{f}_n, \mathbf{X}) = 2^n$ as well. Furthermore, since $\mathbf{X}$ uniquely determines a Boolean function and there are $2^{H(\varepsilon)\cdot 2^n}$ Boolean functions that are ε-close to it, we have $H(\boldsymbol{f}_n \,|\, \mathbf{X}) \leq H(\varepsilon) \cdot 2^n$. Applying the chain rule, we see that

$$H(\mathbf{X}) = H(\boldsymbol{f}_n, \mathbf{X}) - H(\boldsymbol{f}_n \,|\, \mathbf{X}) \geq (1 - H(\varepsilon)) \cdot 2^n. \tag{6}$$

For each $i \in [t]$, we write $p_i = \mathbf{E}[\mathbf{X}_i]$ to denote that probability that g_i occurs in the optimal ε-approximator for $\boldsymbol{f}_n$, and so $\mathbf{E}_t[p_t] = \mathbf{E}[\mathrm{opt}_\varepsilon[\boldsymbol{f}_n]]/t$. and note that

$$\begin{aligned} H(\mathbf{X}) \;\leq\; \sum_{i=1}^{t} H(\mathbf{X}_i) &= \sum_{i=1}^{t} H(p_i) = t \cdot \mathop{\mathbf{E}}_t[H(p_t)] \\ &\leq t \cdot H(\mathop{\mathbf{E}}_t[p_t]) \\ &= t \cdot H\left(\frac{\mathbf{E}[\mathrm{opt}_\varepsilon[\boldsymbol{f}_n]]}{t}\right) \end{aligned}$$

where we have used the concavity of the binary entropy function. Finally, using the inequality $H(p) \leq p \log(e/p)$, we get

$$\begin{aligned} H(\mathbf{X}) &\leq t \cdot H\left(\frac{\mathbf{E}[\mathrm{opt}_\varepsilon[\boldsymbol{f}_n]]}{t}\right) \\ &\leq \mathbf{E}[\mathrm{opt}_\varepsilon[\boldsymbol{f}_n]] \cdot \log\left(\frac{e \cdot t}{\mathbf{E}[\mathrm{opt}_\varepsilon[\boldsymbol{f}_n]]}\right) \\ &\leq \mathbf{E}[\mathrm{opt}_\varepsilon[\boldsymbol{f}_n]] \cdot \log(e \cdot t). \end{aligned}$$

Combining this upper bound with the lower bound of (6) completes the proof. ☐

Since there are 3^n conjunctions over $\{0,1\}^n$ and $2^{n^{d+1}+O(n)}$ degree-d PTFs over $\{0,1\}^n$ [Cho61], applying Theorem 15 immediately implies strong bounds on the inapproximability of a random function by DNFs and the intersection of low-degree PTFs, respectively.

Theorem 5. *Suppose $\boldsymbol{f}_n$ is ε-approximated by a size-s DNF. Then $s = \Omega(1 - H(\varepsilon)) \cdot 2^n/n$.*

Theorem 16. *Suppose $\boldsymbol{f}_n$ is ε-approximated by the intersection of k degree-d PTFs. Then $k = \Omega(1 - H(\varepsilon)) \cdot 2^n/n^{d+1}$.*

The two remaining lower bounds in Table 1 are both witnessed explicitly by the parity function. We show in Section 6 that any DNF that ε-approximates PAR_n must have width $(1-2\varepsilon)n$ (Theorem 7), and in Section 7 that the intersection of $2^{\Omega_\varepsilon(n)}$ unate functions is required to ε-approximate PAR_n for any $\varepsilon < \frac{1}{16}$ (Theorem 10). In fact, we will see that our proof of the former extends easily to show that DNFs that ε-approximate a random function have width $\Omega_\varepsilon(n)$.

6 Approximating the parity function

We begin this section, and the second part of our paper, with a deterministic construction of a small-size small-width DNF that ε-approximates PAR_n with one-sided error.

Lemma 6.1. *Let $B_1, B_2, \ldots, B_\ell \subseteq [n]$ be linearly independent* [7] *sets of coordinates. Define $k = |B_1 \cup B_2 \cup \cdots \cup B_\ell|$. There is a DNF of size $2^{k-\ell}$ and width k that accepts exactly all $x \in \{0,1\}^n$ such that $\bigoplus_{i\in B_1} x_i = \bigoplus_{i\in B_2} x_i = \cdots = \bigoplus_{i\in B_\ell} x_i = 0$ (i.e. the set of all strings with even Hamming weight within B_1, B_2, ..., and B_ℓ). We write* $\mathrm{even}(B_1, \ldots, B_\ell)$ *to denote this DNF.*

Proof. By linear independence, there are exactly $2^{-\ell} \cdot 2^{|B_1\cup\cdots\cup B_\ell|} = 2^{k-\ell}$ possible settings of the k coordinates in $B_1 \cup \cdots \cup B_\ell$ that satisfy $\bigoplus_{i\in B_1} x_i = \cdots = \bigoplus_{i\in B_\ell} x_i = 0$. The DNF will be a disjunction of all $2^{k-\ell}$ such settings, each of which is computed by a conjunction of k literals. □

Theorem 6. *For every $\varepsilon > 0$ there is a DNF f of width $(1-2\varepsilon)n$ and size $2^{(1-2\varepsilon)n}$ that ε-approximates PAR_n. Furthermore f has one-sided error: if $\mathsf{PAR}_n(x) = 1$ then $f(x) = 1$.*

Proof. Let $k = \log(\frac{1}{2\varepsilon})$. For $i = 1, \ldots, k$, let $B_{i,0} = \{j \in [n] : j_i = 0\}$ and $B_{i,1} = \{j \in [n] : j_i = 1\} = [n] \setminus B_{i,0}$.[8] Consider the function $f = \bigvee_{z\in\{0,1\}^k} \mathrm{even}(B_{1,z_1}, \ldots, B_{k,z_k})$. For any $z \in \{0,1\}^n$, the union of the blocks $B_{1,z_1}, \ldots, B_{k,z_k}$ has size $(1-2^{-k})n = (1-2\varepsilon)n$. So by Lemma 6.1, f is a DNF of width $(1-2\varepsilon)n$ and size $2^k \cdot 2^{(1-2\varepsilon)n-k} = 2^{(1-2\varepsilon)n}$.

For any $x \in \{0,1\}^n$ such that $\mathsf{PAR}_n(x) = 1$ and every $i \in [k]$, either $\mathrm{even}(B_{i,0})$ or $\mathrm{even}(B_{i,1})$ is true. Thus, for each such x there is some $z \in \{0,1\}^k$ for which $\mathrm{even}(B_{1,z_1}, \ldots, B_{k,z_k}) = 1$ and we have $f(x) = 1$. Finally, when $\mathsf{PAR}_n(x) = 0$, then the probability that it has even Hamming weight in all of the blocks B_{i,z_i} is $2^{-k} = 2\varepsilon$. Therefore, the probability that $f(x) \neq \mathsf{PAR}_n(x)$ is $\frac{1}{2} \cdot 2\varepsilon = \varepsilon$, as we wanted to show. □

We remark that Theorem 6 actually establishes a stronger result: every s-sparse $\mathbb{F}_2$-polynomial (*i.e.* the parity of s conjunctions) can be ε-approximated by a DNF of size $2^{(1-2\varepsilon)s}$. For example, for any $\varepsilon > 0$ the inner-product-mod-2 function $(x_1 \wedge y_1) \oplus \ldots \oplus (x_{n/2} \wedge y_{n/2})$ can be ε-approximated by a DNF of size $2^{(\frac{1}{2}-\varepsilon)n}$.

6.1 Lower bounds on DNF width and size

We begin with a simple lemma relating the total influence of close Boolean functions.

Lemma 6.2. *For any functions $f, g : \{0,1\}^n \to \{0,1\}$, the total influence of f and g satisfies $|\mathbf{Inf}[f] - \mathbf{Inf}[g]| \leq 2\,\mathbf{Pr}[f(x) \neq g(x)] \cdot n$. In particular, if f is ε-close to PAR_n then $\mathbf{Inf}[f] \geq (1-2\varepsilon)n$.*

Proof. We think of g as being obtained from f by flipping the values of an $\varepsilon = \mathbf{Pr}[f(x) \neq g(x)]$ fraction of inputs. Let f^* denote f with the value of a single input x^* flipped, and note that $|s(f, x^*) - s(f^*, x^*)| \leq n$, that $|s(f, y) - s(f^*, y)| \leq 1$ for all y such that $\mathrm{dist}(x^*, y) = 1$, and that $s(f, z) = s(f^*, z)$ for all other z such that $\mathrm{dist}(x^*, z) \geq 2$. It follows that $|\mathbf{Inf}[f] - \mathbf{Inf}[f^*]| \leq n + n = 2n$, and taking a union bound over all x^* such that $f(x^*) \neq g(x^*)$ completes the proof. □

Building on the bounds on Boppana and Traxler [Bop97, Tra09] and resolving an open problem of O'Donnell, Amano proved that the total influence of a Boolean function is at most its DNF width [Ama11].

Theorem 17 (Amano)**.** *Let f be a width-w DNF. Then $\mathbf{Inf}[f] \leq w$.*

Combining Lemma 6.2 and Theorem 17 yields a lower bound on the width of any DNF that ε-approximates PAR_n, matching the width of our construction in Theorem 6 exactly.

[7] More formally, let these sets correspond to linearly independent vectors under the usual correspondence between subsets of $[n]$ and vectors in $\mathbb{F}_2^n$.

[8] Here, the notation j_i represents the ith bit of the binary representation of j.

Theorem 7. *Let f be a width-w DNF that ε-approximates PAR_n. Then $w \geq (1-2\varepsilon)n$.*

Straightforward Fourier-analytic computations show that the total influence of a random Boolean function is $n/2$ in expectation (see *e.g.* Theorem 6 of [BCS97]), and so by the same reasoning used to establish Theorem 7, we see that DNFs that ε-approximate a random function have width at least $(\frac{1}{2}-2\varepsilon)n = \Omega_\varepsilon(n)$, as had been claimed at the end of Section 5. Next we turn to lower bounds on size of DNFs that ε-approximate PAR_n, giving two incomparable bounds. The first simply combines Amano's theorem with an elementary truncation argument.

Theorem 8. *(first lower bound) Let f be a size-s DNF that ε-approximates PAR_n. Then $s \geq \delta\, 2^{(1-2\varepsilon-2\delta)n}$ for all $\delta > 0$.*

Proof. We use the folklore observation that dropping all terms of width greater than $\log(s/\delta)$ in f yields a DNF g that is δ-close to f (to see this, note that each dropped term is satisfied with probability less than $2^{-\log(s/\delta)} = \delta/s$, and taking a union bound yields an approximation error that is less than δ). Since g is an s-term DNF of width $\log(s/\delta)$ that is $(\varepsilon+\delta)$-close to PAR_n, and we may apply Theorem 7 to get $\log(s/\delta) \geq (1-2(\varepsilon+\delta))n$; rearranging completes the proof. $\square$

The second lower bound on size uses a sharpening of the Boppana's $O(\log s)$ bound on the total influence of size-s DNFs [Bop97], obtained in concurrent work by the present authors via the entropy method [BTW13]. For the sake of completeness, we include its short proof here.

Theorem 18. *Let $f : \{0,1\}^n \to \{0,1\}$ be a size-s DNF with $\mathbf{E}[f] = \mu$. Then $\mathbf{Inf}[f] \leq 2\mu\log(s/\mu)$.*

Proof. Consider a fixed ordering of the terms in the DNF for f, and define three random variables: $\mathbf{X}$ is a uniform random $x \in f^{-1}(1)$, $\mathbf{Y}$ is the indicator of uniform random subset of sensitive coordinates of $\mathbf{X}$, and $\mathbf{T}$ is the first term in the DNF satisfied by $\mathbf{X}$. We consider the joint entropy of these three random variables. Applying the chain rule, we see that

$$\begin{aligned} H(\mathbf{X},\mathbf{Y},\mathbf{T}) &= H(\mathbf{X}) + H(\mathbf{Y}\,|\,\mathbf{X}) + H(\mathbf{T}\,|\,\mathbf{X},\mathbf{Y}) \\ &= H(\mathbf{X}) + H(\mathbf{Y}\,|\,\mathbf{X}) + 0 \\ &= n - \log(1/\mu) + \mathbf{Inf}[f]/(2\mu). \end{aligned}$$

We claim that $H(\mathbf{X},\mathbf{Y},\mathbf{T}) \leq n + \log s$, noting that this implies the claimed upper bound by rearranging the terms. Again applying the chain rule, we have

$$\begin{aligned} H(\mathbf{X},\mathbf{Y},\mathbf{T}) &= H(\mathbf{T}) + H(\mathbf{X}\,|\,\mathbf{T}) + H(\mathbf{Y}\,|\,\mathbf{X},\mathbf{T}) \\ &\leq \log s + \mathop{\mathbf{E}}_{\mathbf{T}}[n - |\mathbf{T}|] + \mathop{\mathbf{E}}_{\mathbf{T}}[|\mathbf{T}|] \\ &= n + \log s. \end{aligned}$$

Here the expectations are with respect to the distribution where the weight of a term T is the probability that a uniformly random $x \in f^{-1}(1)$ satisfies T. The inequality uses that fact that the number of inputs satisfying a term T is $2^{n-|T|}$, and if x satisfies T it can only be sensitive on the coordinates fixed by T. $\square$

Theorem 8. *(second lower bound) Let f be a size-s DNF that ε-approximates PAR_n. Then $s \geq (\frac{1}{2}-\varepsilon)\cdot 2^{\frac{(1-2\varepsilon)n}{(1+2\varepsilon)}}$.*

Proof. Let $\mathbf{E}[f] = \mu$. Since PAR_n is balanced and f is ε-close to PAR_n, we have $\mu \in [\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon]$. Combining Lemma 6.2 and Theorem 18, we get $2\mu\log(s/\mu) \geq \mathbf{Inf}[f] \geq (1-2\varepsilon)n$, and rearranging yields the claimed lower bound: $s \geq \mu \cdot 2^{(1-2\varepsilon)n/2\mu} \geq (\frac{1}{2}-\varepsilon)\cdot 2^{(1-2\varepsilon)n/(1+2\varepsilon)}$. $\square$

7 Lower bounds for intersection of LTFs and unate functions

In this section we provide further evidence of the optimality of our DNF approximators for PAR_n. We begin by showing that a noise sensitivity conjecture of Klivans *et al.* [KOS04] implies that ε-approximating PAR_n even with the intersection of halfspaces requires size $2^{\Omega_\varepsilon(n)}$, matching the size of our DNF approximators in Theorem 6.

Conjecture 1 (Klivans-O'Donnell-Servedio)**.** *Let f be a Boolean function computed by the intersection of k halfspaces. Then $\mathbf{NS}_\delta[f] \leq O(\sqrt{\log k}\sqrt{\delta})$.*

Theorem 9. *Assume the KOS conjecture and let $\varepsilon < 1/2$. Let f be a Boolean function computed by the intersection of k halfspaces and suppose f ε-approximates PAR_n. Then $k = 2^{\Omega_\varepsilon(n)}$.*

Proof. Since f is ε-close to PAR, we have $\mathbf{W}_n[f] = \widehat{f}([n])^2 \geq (1-2\varepsilon)^2$, and so by the Fourier expression for noise sensitivity at noise rate δ we get

$$\mathbf{NS}_\delta[f] \geq \tfrac{1}{2} - \tfrac{1}{2}\left((1-2\varepsilon)^2(1-2\delta)^n + (1-(1-2\varepsilon)^2)\right).$$

Assuming upper bound on $\mathbf{NS}_\delta[f]$ given by Conjecture 1 and taking $\delta = 1/n$, we have $O(\sqrt{\log(k)/n}) \geq \mathbf{NS}_{1/n}[f]$ and

$$\mathbf{NS}_{1/n}[f] \geq \frac{1}{2} - \frac{1}{2}\left(\frac{(1-2\varepsilon)^2}{e} + (1-(1-2\varepsilon)^2)\right).$$

The quantity on the RHS is positive for any $\varepsilon < 1/2$, and so rearranging yields the claimed lower bound on k. $\square$

Naturally, we would like an unconditional proof of Theorem 9. We are able to accomplish this for any fixed $\varepsilon < \frac{1}{16}$, and in fact, our lower bound holds against the more expressive class of intersection of unate functions.

Lemma 7.1. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function computed by the intersection of k unate functions. Suppose $S^* \subseteq \{0,1\}^n$ is a set of 0-inputs of f such that for all pairs $x, y \in S^*$, there exists a coordinate $i \in [n]$ such that $x_i \neq y_i$ and $s(f,x,i) = s(f,y,i) = 1$. Then $k \geq |S^*|$.*

Proof. Let f be computed by the intersection of k unate functions $g_1, \ldots, g_k$. We will show that for any $x \in S^*$ there must be some g_j such that $g_j(x) = 0$ and $g_j(y) = 1$ for all other $y \in S^*$, noting that the claimed lower bound $k \geq |S^*|$ follows immediately.

Fix $x \in S^*$. Since $f(x) = 0$ and f is computed by the intersection of $g_1, \ldots, g_k$, certainly there must be some g_j such that $g_j(x) = 0$. We claim that $g_j(y) = 0$ for all other $y \in S^*$. Seeking a contradiction, suppose there is some $y \in S^*$, $y \neq x$ such that $g_j(y) = 0$. Since $x, y \in S^*$, there is some coordinate $i \in [n]$ such that $x_i \neq y_i$ and $s(f,x,i) = s(f,y,i) = 1$. Without loss of generality, suppose g_j is monotone in the i-th direction, and that $x_i = 0$ whereas $y_i = 1$. It follows that $g_j(y^{\oplus i}) = 0$ and hence $f(y^{\oplus i}) = 0$, contradicting the assumption that $s(f,y,i) = 1$. $\square$

Note that Lemma 7.1 can be used to lower bound the number of unate functions whose intersection computes PAR_n exactly: taking S^* to be the set of all 0-inputs of PAR_n, we get an optimal lower bound of $k \geq |S^*| = 2^{n-1}$. Next we use Lemma 7.1 to show that even functions that approximate PAR_n require exponentially many unate functions.

Theorem 10. *Fix $\varepsilon < \frac{1}{16}$. Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function that is computed by the intersection of k unate functions and suppose f ε-approximates PAR_n. Then $k = 2^{\Omega_\varepsilon(n)}$.*

Proof. Let $\varepsilon = \frac{1}{16} - \gamma$, where $\gamma > 0$. We first note that since f is ε-close to PAR_n, the expected 0-sensitivity of f must be large:

$$\begin{aligned}\mathbf{E}[s(f,x) \mid f(x) = 0] &= \frac{\mathbf{Inf}[f]}{\mathbf{Pr}[f(x) = 0]} \\ &\geq \frac{(1-2\varepsilon)n}{(1+2\varepsilon)} \\ &\geq (1-4\varepsilon)n = \left(\frac{3}{4} + 4\gamma\right)n.\end{aligned}$$

Therefore a constant fraction of the 0-inputs of f have sensitivity at least $(\frac{3}{4}+\gamma)n$. Since $\mathbf{Pr}[f(x) = 0] \geq \frac{1}{2} - \varepsilon$, it follows that there is a set $S \subseteq \{0,1\}^n$ of $\Omega(2^n)$ inputs such that $f(x) = 0$ and $s(f,x) \geq (\frac{3}{4} + \gamma)n$ for all $x \in S$. Note that for every pair $x, y \in S$ share at least $(\frac{1}{2} + 2\gamma)n$ sensitive coordinates. Next we note that for every $x \in S$, there are at most $2^{H(\frac{1}{2}+\gamma)n}$ inputs $y \in S$ that are at distance $(\frac{1}{2} + \gamma)n$ from x. It follows that there is a subset $S^* \subseteq S$ of size at least $|S|/2^{H(\frac{1}{2}+\gamma)n} = 2^{\Omega(n)}$ such that for all $x, y \in S^*$, there exists some $i \in [n]$ such that $x_i \neq y_i$ and $s(f,x,i) = s(f,y,i) = 1$. Applying Lemma 7.1 yields the claimed lower bound. □

8 Conclusion

Having obtained asymptotically matching universal bounds on the approximability of all Boolean functions with respect to DNF width in this work, the natural next step would be to do likewise for DNF size, closing the current gap between $\Omega_\varepsilon(2^n/n)$ and $O_\varepsilon(2^n/\log n)$.

Open Problem 1. *Obtain matching universal bounds on the approximability of all Boolean functions with respect to DNF size. That is, determine the function $\varphi(n)$ such that every Boolean function f can be ε-approximated by a DNF of size $O_\varepsilon(2^n/\varphi(n))$, and there exists an f such that any ε-approximator for f has size $\Omega_\varepsilon(2^n/\varphi(n))$.*

Another open problem is to prove that the size of our DNF approximators for PAR_n in Theorem 6 is optimal even up to the exact dependence on ε in the exponent, closing the current gap between $(1-2\varepsilon)n$ and $\frac{(1-2\varepsilon)}{(1+2\varepsilon)}n$. One way to accomplish this is to further improve on the sharpening of Boppana's influence bound on small-size DNFs we obtained in [BTW13]; we believe that understanding this basic complexity measure of DNFs is a fundamental question in its own right. We restate here the conjectured bound from [BTW13], which would be tight by considering the parity function on $\log(s) + 1$ variables.

Conjecture 2. *Let f be computed by a size-s DNF. Then* $\mathbf{Inf}[f] \leq \log(s) + 1$.

Acknowledgment

We thank Johan Håstad, Ryan O'Donnell, and Rocco Servedio for numerous helpful discussions, and the anonymous referees for useful feedback. We also thank Mahdi Cheraghchi and Benny Sudakov for pointers to the literature on covering codes.

References

[Ajt83] Miklós Ajtai. Σ_1^1-formulae on finite structures. *Annals of Pure and Applied Logic*, 24(1):1–48, 1983. 1

[Ama11] Kazuyuki Amano. Tight bounds on the average sensitivity of k-CNF. *Theory of Computing*, 7(1):45–48, 2011. 1.1, 6.1

[BCS97] A. Bernasconi, B. Codenotti, and J. Simon. On the Fourier analysis of Boolean functions. Technical report, Istituto di Matematica Computazionale, 1997. 6.1

[BIS12] Paul Beame, Russell Impagliazzo, and Srikanth Srinivasan. Approximating AC^0 by small height decision trees and a deterministic algorithm for $\#\mathsf{AC}^0\mathsf{SAT}$. In *IEEE Conference on Computational Complexity*, pages 117–125, 2012. 1

[Bop97] Ravi Boppana. The average sensitivity of bounded-depth circuits. *Information Processing Letters*, 63(5):257–261, 1997. 1.1, 6.1, 6.1

[BTW13] Eric Blais, Li-Yang Tan, and Andrew Wan. Edge-isoperimetric inequalities via the entropy method. Manuscript, 2013. 1.1, 6.1, 8

[Cai89] Jin-Yi Cai. With probability one, a random oracle separates PSPACE from the polynomial-time hierarchy. *J. Comput. Syst. Sci.*, 38(1):68–85, 1989. 1

[CH11] Yves Crama and Peter L. Hammer. *Boolean Functions - Theory, Algorithms, and Applications*, volume 142 of *Encyclopedia of mathematics and its applications*. Cambridge University Press, 2011. 1.1

[CHLL05] G. Cohen, I. Honkala, S. Litsyn, and A. Lobstein. *Covering Codes*. North-Holland Mathematical Library. Elsevier Science, 2005. 4.2

[Cho61] Chao-Kong Chow. On the characterization of threshold functions. In *Proceedings of the 2nd Annual Symposium on Switching Circuit Theory and Logical Design (FOCS)*, pages 34–38, 1961. 5

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

[Gla67] V. V. Glagolev. Nekotorye otsenki dizyunktivnykh normalnykh form funktsiĭ algebry logiki. *Problemy Kibernetiki*, 19:75–95, 1967. 1

[Hås86] Johan Håstad. Almost optimal lower bounds for small depth circuits. In *STOC*, pages 6–20, 1986. 1

[Hås12] Johan Håstad. On the correlation of parity and small-depth circuits. Technical Report TR12-137, Electronic Colloquium on Computational Complexity, 2012. 1

[IMP12] Russell Impagliazzo, William Matthews, and Ramamohan Paturi. A satisfiability algorithm for AC^0. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 961–972, 2012. 1

[Kor69] Aleksej Dmitrievich Korshunov. Verkhnyaya otsenka slozhnosti kratchaĭshikh dnf pochti vsekh bulevykh funktsiĭ. *Kibernetika*, 6:1–8, 1969. 1

[Kor81] Aleksej Dmitrievich Korshunov. O slozhnosti kratchaĭshikh dizyunktivnykh normalnykh form bulevykh funktsiĭ. *Metody Diskretnogo Anal*, 37:9–41, 1981. 1

[Kor83] Aleksej Dmitrievich Korshunov. O slozhnosti kratchaĭshikh dizyunktivnykh normalnykh form sluchanykh bulevykh funktsiĭ. *Metody Diskretnogo Anal*, 40:25–53, 1983. 1, 1

[KOS04] Adam Klivans, Ryan O'Donnell, and Rocco Servedio. Learning intersections and thresholds of halfspaces. *Journal of Computer and System Sciences*, 68(4):808–840, 2004. 1.1, 7

[KP88] G. A. Kabatyanski and V.I. Panchenko. Packings and coverings of the hamming space by unit balls. *Dokl. Akad. Nauk SSSR*, 303(3):550–552, 1988. 1.1, 4.2

[KSV03] Michael Krivelevich, Benny Sudakov, and Van H. Vu. Covering codes with improved density. *IEEE Transactions on Information Theory*, 49(7):1812–1815, 2003. 4.2

[Kuz83] S. E. Kuznetsov. O nizhneĭ otsenke dliny kratchaĭsheĭ dnf pochti vsekh bulevykh funktsiĭ. *Veroyatnoste Metody Kibernetiki*, 19:44–47, 1983. 1, 1, 1.1

[LMN89] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform and Learnability. In *Proceedings of the 30th Annual IEEE Symposium on Foundations of Computer Science*, pages 574–579, 1989. 1

[Lup61] Oleg Lupanov. Implementing the algebra of logic functions in terms of constant depth formulas in the basis &, $\vee$, $\neg$. *Dokl. Ak. Nauk. SSSR*, 136:1041–1042, 1961. 1

[OW07] Ryan O'Donnell and Karl Wimmer. Approximation by DNF: examples and counterexamples. In *Proceedings of the 34th Annual International Colloquium on Automata, Languages and Programming*, pages 195–206, 2007. 1

[Pip03] Nicholas Pippenger. The shortest disjunctive normal form of a random boolean function. *Random Struct. Algorithms*, 22(2):161–186, 2003. 1, 1.1

[Qui54] Willard Van Orman Quine. Two theorems about truth functions. *Bol. Soc. Math. Mexicana*, 10:64–70, 1954. 1

[Sap72] A. A. Sapozhenko. O slozhnosti dizyunktivnykh normalnykh form, poluchaemykh s pomoshchyu gradientnogo algoritma. *Diskretnyĭ Anal*, 21:62–71, 1972. 1

[Tra09] Patrick Traxler. Variable influences in conjunctive normal forms. In *SAT*, pages 101–113, 2009. 6.1

[Yao85] Andrew Yao. Separating the polynomial time hierarchy by oracles. In *Proceedings of the 26th Annual IEEE Symposium on Foundations of Computer Science*, pages 1–10, 1985. 1

ECCC ISSN 1433-8092
http://eccc.hpi-web.de