Electronic Colloquium on Computational Complexity, Report No. 92 (2017)

A Duality Between Depth-Three Formulas and Approximation by Depth-Two

Shuichi Hirahara*

Abstract

We establish an explicit link between depth-3 formulas and one-sided-error approximation by depth-2 formulas, which were previously studied independently. Specifically, we show that the minimum size of depth-3 formulas is (up to a factor of n) equal to the inverse of the maximum, over all depth-2 formulas, of one-sided-error correlation bound divided by the size of the depth-2 formula, on a certain hard distribution. We apply this duality to obtain several consequences:

1. Any function f can be approximated by a CNF formula of size $O(\epsilon 2^n/n)$ with one-sided error and advantage ϵ for some ϵ, which is tight up to a constant factor.
2. There exists a monotone function f such that f can be approximated by some polynomial-size CNF formula, whereas any *monotone* CNF formula approximating f requires exponential size.
3. Any depth-3 formula computing the parity function requires $\Omega(2^{2\sqrt{n}})$ gates, which is tight up to a factor of $\sqrt{n}$. This establishes a quadratic separation between depth-3 *circuit* size and depth-3 *formula* size.
4. We give a characterization of the depth-3 monotone circuit complexity of the majority function, in terms of a natural extremal problem on hypergraphs. In particular, we show that a known extension of Turán's theorem gives a tight (up to a polynomial factor) circuit size for computing the majority function by a monotone depth-3 circuit with bottom fan-in 2.
5. $\mathsf{AC}^0[p]$ has exponentially small one-sided correlation with the parity function for odd prime p.

1 Introduction

The main theme of this paper is a new connection between approximation by depth-2 formulas and exact computation of depth-3 formulas. We first review these two (previously independent) lines of research.

1.1 Approximation by Depth-2 Formulas

A depth-2 formula is a CNF formula or a DNF formula. By De Morgan's law, we can often assume without loss of generality that a depth-2 formula is a CNF formula; hence, in this paper, we will focus on CNF formulas. The size $|\varphi|$ of a CNF formula φ is the number of clauses in the formula.

*Department of Computer Science, The University of Tokyo, Tokyo, Japan, hirahara@is.s.u-tokyo.ac.jp

ISSN 1433-8092

A CNF formula is one of the simplest computational models in complexity theory, and its computational power is well understood; for example, the parity function requires CNF formulas of size 2^{n-1}, which is exactly tight; any function can be computed by a CNF formula of size 2^{n-1}, and hence the parity function is the hardest problem against a CNF formula.

Nonetheless, the situation becomes quite different when, instead of exact computation, *approximation* by CNF formulas is concerned. Namely, we allow a CNF formula to err on some fraction of inputs. We say that a CNF formula φ ϵ-approximates a function f if $\Pr[f(x) \neq \varphi(x)] < \epsilon$. Initiated by O'Donnell and Wimmer [OW07], a line of research [OW07, BT15, BHST14] has highlighted surprising differences between exact computation of depth-2 formulas and approximation by depth-2 formulas. Two of them are reviewed below.

Quine's Theorem for Approximation. Blais, Håstad, Servedio, and Tan [BHST14] studied whether an analog of Quine's theorem holds. Quine's theorem states that the minimum CNF formula computing a monotone function is achieved by monotone CNF formulas (i.e., CNF formulas without any negated literal).

Theorem (Quine [Qui54])**.** *A smallest CNF formula computing a monotone function* exactly *is monotone as well.*

Given this fact, it is tempting to guess that a smallest CNF formula *approximating* a monotone function is monotone. However, in [BHST14], it was shown that there is at least a quadratic gap between the size of the smallest CNF formula and the smallest monotone CNF formula approximating a monotone function f.

Theorem ([BHST14])**.** *There are a parameter $\epsilon(n)$ and a monotone function on n variables that can be $\epsilon(n)$-approximated by some* nonmonotone *CNF formula of size $O(n)$, but cannot be $\epsilon(n)$-approximated by any* monotone *CNF formula of size less than n^2.*

In this paper, we will significantly improve this bound and exhibit an *exponential* separation.

Universal Bounds. Blais and Tan [BT15] studied the universal bound on the size of CNF formulas approximating a function: they showed that any function f can be ϵ-approximated by any CNF formula of size $O_\epsilon(2^n/\log n)$ for any constant ϵ (where a constant O_ϵ depends on ϵ). They also gave a lower bound of $\Omega_\epsilon(2^n/n)$ for a CNF formula ϵ-approximating a random function, and thus leaving a gap. Namely, the maximum complexity of CNF formulas approximating functions was not well understood.

1.2 Depth-3 Circuits

As with approximation by depth-2, another simple computational model whose power is still "mystery" in complexity theory is *depth-3 circuits* (or formulas). A depth-3 (AC^0) circuit here is a directed acyclic graph consisting of alternating 3 layers of AND and OR gates and an input layer whose gates are labelled by literals (i.e., an input x_i or its negation $\neg x_i$). A depth-3 *formula* is a depth-3 circuit whose gates have fan-out 1 (i.e., a computational model in which the intermediate computation cannot be reused). For simplicity, we assume that the top gate of depth-3 circuits and formulas is an OR gate (i.e., OR $\circ$ AND $\circ$ OR circuits).

A counting argument [RS42] shows that most functions f requires $\Omega(2^n/\log n)$ literals for a formula to compute f. And this is tight for depth-3 formulas, as a classical theorem by Lupanov

[Lup65] shows that any function can be computed by a depth-3 formula with $O(2^n/\log n)$ literals. Therefore, the complexity of depth-3 formulas computing *random functions* is fairly well understood.

In sharp contrast, it is wide open to prove a lower bound $2^{\omega(\sqrt{n})}$ of a depth-3 circuit (or formula) computing some *explicit* function (e.g., a function in P, or even E^{NP}). In 1980s, there has been significant progress in the understanding of the computational power of unbounded fan-in constant depth circuits (e.g., [Ajt83, FSS84, Yao85, Hås89, Raz87, Smo87]). These results give a depth-d circuit lower bound of the form $2^{\Omega(n^{1/(d-1)})}$ for explicit functions, such as the parity and majority functions. After 30 years, this remains asymptotically the current best circuit lower bounds against constant depth circuits.

There are several reasons why depth-3 circuits are interesting. One notable reason was given by Valiant [Val77], who showed that any linear-size log-depth circuit can be transformed into a depth-3 circuit of size $2^{O(n/\log\log n)}$. Hence, a strongly exponential depth-3 circuit lower bound will yield a super-linear log-depth circuit lower bound simultaneously, the former of which appears to be easier to deal with. Considerable efforts (e.g., [HJP95, PPZ99, PSZ00, PPSZ05, IPZ01, GW13, GT16]) have been thus made to obtain strongly exponential depth-3 circuit lower bounds.

Another reason is that depth-3 circuits are closely related to CNF-SAT algorithms. For example, Paturi, Pudlák and Zane [PPSZ05] proved that the minimum size of a depth-3 circuit computing the parity function is $\Theta(n^{1/4}2^{\sqrt{n}})$, and simultaneously developed a simple and fast CNF-SAT algorithm; Paturi, Pudlák, Saks and Zane [PPSZ05] improved the CNF-SAT algorithm and simultaneously gave a depth-3 circuit lower bound $2^{1.282\dots\sqrt{n}}$ for some explicit function, which is the current best depth-3 circuit lower bound. (A relationship between a fast CNF-SAT algorithm and a circuit lower bound was made formal by Williams [Wil13].)

1.3 Duality Theorem

The main theme of this paper is to weave these two lines of research together; thereby we advance the two lines simultaneously. To this end, we will exploit a general result below, which establishes an equivalence between exact computation of $\mathrm{OR} \circ \mathcal{C}$ and one-sided-error approximation by $\mathcal{C}$ for any circuit class $\mathcal{C}$. (Here, $\mathrm{OR} \circ \mathcal{C}$ denotes the class of circuits that consist of a top OR gate fed by disjoint $\mathcal{C}$ circuits. In most of our applications, we take $\mathcal{C} = \{\text{ CNF formulas }\}$ and hence $\mathrm{OR} \circ \mathcal{C}$ is the class of depth-3 formulas.)

First, we need to introduce one new notion. For any circuit class $\mathcal{C}$ and a Boolean function f and a distribution μ on $f^{-1}(1)$, define *maximum (one-sided-error) correlation*[1] *per size* of f as

$$D^{\mathcal{C}}_{\mu}(f) := \max_{\substack{\varphi\in\mathcal{C}\\ \varphi\le f}} \Pr_{x\sim\mu}[\varphi(x) = f(x)]/|\varphi| \quad \left(= \max_{\substack{\varphi\in\mathcal{C}\\ \varphi\le f}} \Pr_{x\sim\mu}[\varphi(x) = 1]/|\varphi| \right),$$

and define $D^{\mathcal{C}}(f) := \min_\mu D^{\mathcal{C}}_\mu(f)$. Intuitively, (one-sided-error) correlation per size measures a trade-off between the size of a circuit $\varphi \in \mathcal{C}$ and how well f can be approximated by a circuit $\varphi \in \mathcal{C}$ with one-sided error in the sense that $\varphi^{-1}(1) \subseteq f^{-1}(1)$. Our general result relates the complexity of exact computation of $\mathrm{OR} \circ \mathcal{C}$ and the complexity of one-sided-error approximation by $\mathcal{C}$:

[1]In a usual context of circuit lower bounds, a correlation between functions f and g is defined as $2\Pr[f(x) = g(x)] - 1$. While the definition here is slightly different, we borrow the terminology.

Theorem 1 (Duality (informal)). *For any circuit class $\mathcal{C}$ and a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$, let $\mathrm{L}^{\mathcal{C}}(f)$ denote the minimum size of an OR $\circ\, \mathcal{C}$ circuit computing f. Then,*

$$\mathrm{L}^{\mathcal{C}}(f) \approx (D^{\mathcal{C}}(f))^{-1},$$

up to a factor of n.

Here, the size of an OR $\circ\, \mathcal{C}$ circuit $C_1 \vee \cdots \vee C_m$ is defined as the sum $\sum_{i=1}^{m} |C_i|$.

In fact, its proof is quite simple: Regard a computation by an OR $\circ\, \mathcal{C}$ circuit as a set cover instance, consider the linear programming relaxation, and then take its *dual*, whose optimal value corresponds to $(D^{\mathcal{C}}(f))^{-1}$.

Related Work on Duality. It should be noted that a similar result was known before in the context of threshold circuits. Indeed, the discriminator lemma by Hajnal, Maass, Pudlák, Szegedy and Turán [HMP+93] states that a circuit with top threshold gate that computes a function f must have a subcircuit which has a high correlation with f. A converse direction of the discriminator lemma was proved by Goldmann, Håstad, and Razborov [GHR92] in the context of threshold circuits, and by Freund [Fre95] in the context of boosting. Our duality can be regarded as a version of the discriminator lemma and its converse specialized to the case of circuits with top AND or OR gate (instead of threshold gate), with a *tighter quantitative trade-off* between correlation and size. To the best of our knowledge, the notion of correlation per size was neither defined nor recognized as important quantity before; and the direction $\mathrm{L}^{\mathcal{C}}(f) \lesssim (D^{\mathcal{C}}(f))^{-1}$ was not known before.

In what follows, we explain how our duality theorem advances the lines of research on depth-3 formulas and approximation by depth-2 formulas.

1.4 Universal Bounds on Depth-2 Approximator

We first apply the duality in order to obtain a tight universal bound on a depth-2 approximator. As mentioned before, Blais and Tan [BT15] left a gap between an upper bound $O(2^n/\log n)$ of a CNF formula approximating any function and a lower bound $\Omega(2^n/n)$ for a random function.

In contrast, the depth-3 formula complexity of random functions is well understood [RS42, Lup65], at least when the size of depth-3 formulas is defined as the number of *literals* (and the bound is $\Theta(2^n/\log n)$). For our purpose, we need to redefine the size of a depth-3 formula as the number of AND gates at the middle layer, so that it is consistent with the fact that the size of a CNF formula is measured as the number of clauses. In fact, it turns out that Lupanov's construction [Lup65] can be adapted to this size measure, and we obtain an improved upper bound $O(2^n/n)$ of depth-3 formulas computing any function. Our duality theorem transfers this result into approximation by depth-2, and we obtain the following tight bound:

Theorem 2 (informal).

- *For any function f, there exists some CNF formula φ of size $O(\epsilon \cdot 2^n/n)$ approximating f with one-sided error and advantage ϵ for some ϵ.*
- *There exists a function f such that any CNF formula approximating f with one-sided error and advantage ϵ must have size $\Omega(\epsilon \cdot 2^n/n)$.*

Here, the constants hidden in O and Ω are universal (and, in particular, do not depend on ϵ nor f).

1.5 Approximating Monotone Functions

Next, we apply the duality theorem in order to obtain an exponential separation between non-monotone CNF formulas and monotone CNF formulas computing a monotone function, which significantly improves [BHST14]:

Theorem 3 (informal)**.** *There exists a monotone function f such that*

1. *there is some polynomial-size CNF formula that approximates f with one-sided error, whereas*
2. *any* monotone *CNF formulas approximating f with one-sided error requires exponential size.*

That is, Quine's theorem fails badly for approximation by depth-2 formulas.

Our proof is in fact an immediate consequence of a recent result by Chen, Oliveira and Servedio [COS15]: they showed that there is a monotone function f such that f can be computed by some depth-3 formula of polynomial size, whereas any *monotone* depth-3 formula computing f requires exponential size. Our duality theorem transfers their result to Theorem 3.

1.6 Depth-3 Formula Lower Bound for Parity

In the second half of this paper, we regard the duality theorem as a general approach for understanding the depth-3 formula complexity. That is, we aim at proving an upper bound on the correlation per size $D(f)$ of a CNF formula, and use it to obtain a depth-3 formula lower bound. We prove that any depth-3 formula computing the parity function requires $\Omega(2^{2\sqrt{n}})$ gates,[2] which is tight up to a factor of $\sqrt{n}$.

Our proof follows from an almost tight one-sided-error correlation bound, and our correlation bound improves another result of Blais and Tan [BT15]: They studied the minimum size of CNF formulas that compute $\textsc{Parity}_n$ all but an ϵ fraction of inputs. They showed an upper bound of $2^{(1-2^{-\lceil \log 1/2\epsilon \rceil} n)} \approx 2^{(1-2\epsilon)n}$ (moreover, with one-sided error in our sense) and an lower bound of $(\frac{1}{2}-\epsilon)2^{\frac{1-2\epsilon}{1+2\epsilon}n}$. In this paper, we focus on approximation with one-sided error and high error regimes (e.g., $\epsilon = \frac{1}{2} - 2^{-\sqrt{n}}$). By using the satisfiability coding lemma [PPZ99] and width reduction techniques [Sch05, CIP06], we obtain an lower bound of $2^{n/(k+1)-3}$ for $k := \log \frac{2}{1-2\epsilon}$ and an upper bound of $k' 2^{\lceil n/k' \rceil - 1}$ for $k' := \lfloor k \rfloor$, which significantly improves the results of [BT15] when the error fraction ϵ is close to $\frac{1}{2}$.

In terms of correlation bounds, our result shows that any CNF formula φ that computes $\textsc{Parity}_n$ and does not err on inputs in $\textsc{Parity}_n^{-1}(0)$ has the correlation with $\textsc{Parity}_n$ at most $2^{-n/(\log|\varphi|+3)+2}$. This improves the previous bounds $2^{-n/O(\log s)^{d-1}}$ on the correlation between the parity function and depth-d circuits of size s in the case of $d = 2$ (Beame, Impagliazzo and Srinivasan [BIS12], Håstad [Hås14] and Impagliazzo, Matthews, and Paturi [IMP12]).

Given the almost optimal one-sided-error correlation bound, our duality theorem immediately implies a depth-3 formula lower bound as follows: The one-sided-error correlation per size of a CNF formula φ with $\textsc{Parity}_n$ is at most $2^{-n/(\log|\varphi|+3)+2}/|\varphi| = 2^{5-n/(\log|\varphi|+3)-(\log|\varphi|+3)}$, which is bounded above by $2^{5-2\sqrt{n}}$ from the inequality of the arithmetic and geometric means. Hence, any depth-3 formula computing the parity function requires $2^{2\sqrt{n}-5}$ gates.

[2] Independently of our work, Rahul Santhanam and Srikanth Srinivasan (personal communication) obtained the same lower bound.

1.7 Depth-3 Circuits vs. Formulas

Our formula lower bound on the parity function is of interest from yet another literature: Questions whether circuits are strictly more powerful than formulas are one of central questions in complexity theory. It is straightforward to see that any unbounded fan-in depth-d circuit of size s has an equivalent representation as a depth-d formula of s^{d-1}, by simply replicating overlapping subcircuits. (Thus, depth-3 formula size and depth-3 circuit size are at most quadratically different.) The converse direction, namely, whether this naïve simulation is optimal, is closely related to the NC^1 vs. AC^1 problem. Indeed, if one could show that there exists a language that can be computed by log-depth unbounded fan-in polynomial-size circuits (i.e., AC^1 circuits) but requires a formula[3] of size $n^{\Omega(\log n)}$, then $\mathsf{NC}^1 \neq \mathsf{AC}^1$. Although we thus cannot hope that the circuit vs. formula question can be solved for $d = O(\log n)$ anytime soon (due to the natural proof barrier [RR97]), the question was solved affirmatively in some restricted cases.

The monotone circuit vs. monotone formula question was solved by Karchmer and Wigderson [KW90]. They showed that monotone formulas computing st-connectivity require $n^{\Omega(\log n)}$ gates, whereas there is a polynomial-size monotone circuit computing st-connectivity. Their communication complexity theoretic approach (Karchmer-Wigderson games) has been quite successful in monotone settings. However, little was known in nonmonotone settings until recent results by Rossman [Ros14, Ros15] (see [Ros14] for further background). He showed that the simulation of depth-d circuits of size s by depth-d formulas requires $s^{\Omega(d)}$ $(> s)$ gates for a sufficiently large d (say, $d \geq 108$) up to $d = o(\frac{\log n}{\log \log n})$. More specifically, he [Ros15] showed that any depth-d formula computing $\textsc{Parity}_n$ requires $2^{\Omega(d(n^{1/(d-1)}-1))}$ gates, whereas it is known that $\textsc{Parity}_n$ can be computed by a depth-d circuit of size $n2^{n^{1/(d-1)}}$.

Motivated by Rossman's results, we may ask whether depth-d circuits are more powerful than depth-d formulas for a small constant d. And we may ask whether depth-d circuits of size s *cannot* be simulated by depth-d formulas of size even slightly better than the naïve simulation, say, $s^{d-1.01}$. We answer these questions affirmatively for $d = 3$: Our formula lower bound immediately implies that simulating depth-3 circuits by depth-3 formulas requires a *quadratic* slowdown, thereby separating depth-3 circuit and formula size almost optimally.

Corollary. *Let $s (= \Theta(n^{1/4} 2^{\sqrt{n}}))$ be the minimum depth-3 circuit size for computing $\textsc{Parity}_n$. Any depth-3 formula computing $\textsc{Parity}_n$ requires $\Omega(s^2/\log s)$ gates.*

We note that this does not seem to follow from Rossman's techniques as he used the switching lemma [Hås89], which may lose some constant factor in the exponent.

1.8 Computing Majority by Depth-3 Circuits

We also make a step towards better understanding of depth-3 formulas computing the majority function. We characterize the monotone complexity of the majority function in terms of a natural extremal problem on hypergraphs: For a hypergraph $\mathcal{F}$, let $|\mathcal{F}|$ denote the number of edges in $\mathcal{F}$, $\tau(\mathcal{F})$ denote the minimum size of hitting sets of $\mathcal{F}$ (i.e., sets that intersect with every edge in $\mathcal{F}$), and $t(\mathcal{F})$ denote the number of minimum hitting sets of $\mathcal{F}$. Let $T(n, \tau) := \max_{\mathcal{F} \colon \tau(\mathcal{F}) = \tau} t(\mathcal{F})/|\mathcal{F}|$, where the maximum is taken over all hypergraphs of n vertices. In Section 3, we show that $2^n/T(n, n/2)$ is asymptotically equal to the minimum size of monotone depth-3 formulas computing the majority

[3] Note that, by Spira's theorem [Spi71], NC^1 can be characterized as languages computable by polynomial-size formulas.

function. The main idea of the proof is that one of the hardest distributions against monotone CNF formulas can be exactly determined.

It is known that any depth-3 formula (or circuit) computing the majority function requires $2^{\Omega(\sqrt{n})}$ gates (e.g., [Hås89, HJP95]), whereas the current best upper bound on the size of a depth-3 formula computing the majority function is $2^{O(\sqrt{n \log n})}$ (which can be constructed by partitioning variables into blocks of equal size; see [KPPY84]), and the depth-3 formula is monotone. Thus, it follows that $2^{n-O(\sqrt{n \log n})} \le T(n, n/2) \le 2^{n-O(\sqrt{n})}$. To the best of our knowledge, there has been essentially no improvement on the minimum depth-3 circuit computing the majority function over 20 years since Håstad *et al.* [HJP95] posed the question explicitly, even when restricted to the case of *monotone* circuits. We propose an open problem of determining the asymptotic behavior of $T(n, n/2)$ as the first step towards determining the minimum depth-3 nonmonotone circuits computing the majority function.

We note that, in the case of graphs (instead of hypergraphs), $T(n, \tau)$ and its extremal structure are well understood in the context of extensions of Turán's theorem [Tur41]. We show that a known extension of Turán's theorem implies the optimal circuit lower bound for computing the majority function by monotone depth-3 circuits with bottom fan-in 2.

1.9 Worst-case to Average-case Connection

Our duality theorem can be viewed as a *generic* equivalence between worst-case complexity and one-sided-error average-case complexity: It applies to any computational model that is capable of simulating an unbounded fan-in OR or AND gate.

As a concrete application, we apply it to the case of $\mathsf{AC}^0[p]$. Here, $\mathsf{AC}^0[p]$ denotes the class of constant-depth circuits consisting of NOT gates and unbounded fan-in AND, OR, and MOD_p gates. Razborov [Raz87] and Smolensky [Smo87] established celebrated exponential lower bounds of $\mathsf{AC}^0[p]$ circuits computing $\textsc{Parity}_n$ for odd prime p. Their techniques also give an average-case lower bound such that any $\mathsf{AC}^0[p]$ circuit of size at most $2^{-n^{o(1)}}$ has correlation with $\textsc{Parity}_n$ at most $n^{-1/2+o(1)}$ (Smolensky [Smo93]; see also [Fil10]). This correlation bound remains the strongest known, and it is a long-standing open problem whether the correlation bound can be improved. See [FSUV13] for more detailed backgrounds.

While we were not able to obtain (two-sided-error) correlation bounds, we do obtain a *one-sided-error* correlation bound. Our duality theorem transfers the worst-case lower bound of Razborov and Smolensky [Raz87, Smo87] against $\mathrm{OR} \circ \mathsf{AC}^0[p]$ circuits to the following one-sided-error correlation bound:

Theorem 4. *Let C be any $\mathsf{AC}^0[p]$ circuit of depth d for some odd prime p such that $C^{-1}(1) \subseteq \textsc{Parity}_n^{-1}(1)$. Then we have $\Pr_{x \sim \{0,1\}^n}[C(x) = \textsc{Parity}_n(x)] \le \frac{1}{2} + |C| \cdot 2^{-n^{\Omega(1/d)}}$, where $|C|$ denotes the number of gates of C.*

In particular, any $\mathsf{AC}^0[p]$ circuit of depth d and size at most $2^{n^{O(1/d)}}$ that does not err on $\textsc{Parity}_n^{-1}(0)$ has correlation with $\textsc{Parity}_n$ at most $2^{-n^{\Omega(1/d)}}$.

1.10 Preliminaries and Notations

Throughout this paper, we assume that the top gate of a depth-3 circuit or formula is an OR gate. Note that this assumption does not lose any generality for computing the parity or majority function, due to De Morgan's laws. A *depth-3 formula* is thus an OR of CNF formulas, and its size is the sum of the size of CNF formulas in the depth-3 formula. A *CNF formula* is an AND of clauses, and its size is the number of clauses in the CNF formula. A clause is an OR of literals. (Alternatively, the size of depth-3 formulas is defined to be the number of OR gates at the bottom layer when represented as a rooted tree. Note that our measure is the same with the number of gates up to a factor of 2 and is the same with the number of literals up to a factor of $2n$.) For a function $f\colon \{0,1\}^n \to \{0,1\}$, we write $\mathrm{L}_3(f)$ for the minimum size of depth-3 formulas computing f. We often identify a formula with the function computed by the formula.

$\textsc{Parity}_n\colon \{0,1\}^n \to \{0,1\}$ is the function such that $\textsc{Parity}_n(x) = 1$ if and only if the number of ones in $x \in \{0,1\}^n$ is odd. Similarly, $\textsc{Maj}_n\colon \{0,1\}^n \to \{0,1\}$ is the function such that $\textsc{Maj}_n(x) = 1$ if and only if the number of ones in x is at least $n/2$.

A *distribution* μ on a finite set X is a function $\mu\colon X \to [0,1]$ such that $\sum_{x\in X} \mu(x) = 1$. We write $x \sim \mu$ to indicate that x is a random variable sampled from a distribution μ. Abusing notation, we identify a finite set X with the uniform distribution on X. For example, $x \sim X$ means that x is a uniform sample from X.

1.11 Organization

The rest of this paper is organized as follows: In Section 2, we prove our duality theorem which links depth-3 formulas with one-sided approximation by depth-2 formulas. In Section 3, we compute one of hardest distributions in some cases, and prove Theorem 4. In Sections 4 and 5, Theorems 3 and 2 are proved, respectively. In Section 6, we study the one-sided-error correlation bound of the parity function. We conclude with some open problems in Section 7.

2 Proof of Duality

In this section, we prove that the complexity of depth-3 formulas is closely related to correlation bounds of CNF formulas with one-sided error. As mentioned, our results hold for general settings (e.g., worst-case complexity of depth-$(d+1)$ formulas is almost equivalent to one-sided-error average-case complexity of depth-d formulas); however, for simplicity, we focus on the case of CNF formulas. We first define correlation bounds of CNF formulas.

Definition 5. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a function.

- Let CNF_f be the set of all CNF formulas φ such that $\varphi^{-1}(1) \subseteq f^{-1}(1)$ (i.e., φ does not err on inputs x such that $f(x) = 0$).
- Let μ be a distribution on $f^{-1}(1)$. The *maximum one-sided correlation per size*[4] $D_\mu(f)$ with f with respect to μ is
$$D_\mu(f) := \max_{\varphi\in\mathrm{CNF}_f} \Pr_{x\sim\mu}[\varphi(x) = 1] \,/\, |\varphi|.$$
- The *maximum one-sided correlation per size* $D(f)$ with f is $D(f) := \min_\mu D_\mu(f)$, where the minimum is taken over all distribution μ on $f^{-1}(1)$.
- We denote by $D^+(f)$ the maximum one-sided correlation per size with f for *monotone* CNF formulas (i.e., formulas without negated literals).

[4] In order to make this definition well-defined, we regard $0/0$ as 0.

The following duality is the main principle that we will exploit.

Theorem 6. *Let* $f\colon \{0,1\}^n \to \{0,1\}$ *be a function such that* $\mathrm{L}_3(f) > 0$*. The following holds:*

$$(D(f))^{-1} \le \mathrm{L}_3(f) \le (1 + n \cdot \ln 2) \cdot (D(f))^{-1}.$$

Proof. The idea is to regard the depth-3 formulas as a set cover instance, apply a linear programming relaxation and take the dual of the LP relaxation.

Specifically, if f is computed by a depth-3 formula, f can be written as a disjunction of CNF formulas $\varphi_1, \ldots, \varphi_m$ (i.e., $f = \bigvee_{i=1}^m \varphi_i$, or equivalently, $f^{-1}(1) = \bigcup_{i=1}^m \varphi_i^{-1}(1)$) such that $\varphi_i^{-1}(1) \subseteq f^{-1}(1)$. Moreover, the size of the depth-3 formula is $\sum_{i=1}^m |\varphi_i|$. Therefore, the problem of finding the minimum depth-3 formula computing f is equivalent to the following set cover instance: We want to cover a universe $U := f^{-1}(1)$ by using a collection of sets $\varphi^{-1}(1)$ such that φ is a CNF formula and $\varphi^{-1}(1) \subseteq f^{-1}(1)$ (i.e., $\varphi \in \mathrm{CNF}_f$), where the cost of $\varphi^{-1}(1)$ is defined to be $|\varphi|$. The minimum cost of this set cover instance is exactly equal to the minimum size of depth-3 formulas computing f.

Now we consider a linear programming relaxation of the set cover instance on variables x_φ for each $\varphi \in \mathrm{CNF}_f$:

$$\begin{aligned}
&\text{minimize} && \sum_{\varphi \in \mathrm{CNF}_f} |\varphi| \cdot x_\varphi && \\
&\text{subject to} && \sum_{\varphi \in \mathrm{CNF}_f \colon \varphi(e)=1} x_\varphi \ge 1 && \text{for all } e \in f^{-1}(1), \\
& && x_\varphi \ge 0 && \text{for all } \varphi \in \mathrm{CNF}_f.
\end{aligned}$$

Let s^* denote the optimal value of this linear programming. Since it is a linear programming relaxation of the set cover problem, it holds that $s^* \le \mathrm{L}_3(f)$. Moreover, it is well known that the integrality gap of the set cover problem is at most $1 + \ln |U| \le 1 + n \cdot \ln 2$ (Lovász [Lov75] and Chvatal [Chv79]; see also Vazirani [Vaz01]). Thus, we have $s^* \le \mathrm{L}_3(f) \le (1 + n \cdot \ln 2) \cdot s^*$.

It remains to claim that $s^* = (D(f))^{-1}$. By the strong duality of linear programming, s^* is equal to the optimal value of the following dual problem on variables y_e for each $e \in f^{-1}(1)$:

$$\begin{aligned}
&\text{maximize} && \sum_{e \in f^{-1}(1)} y_e && \\
&\text{subject to} && \sum_{e \in \varphi^{-1}(1)} y_e \le |\varphi| && \text{for all } \varphi \in \mathrm{CNF}_f, \\
& && y_e \ge 0 && \text{for all } e \in f^{-1}(1).
\end{aligned}$$

Let $v := \sum_{e \in f^{-1}(1)} y_e$. Define[5] μ to be the distribution on $f^{-1}(1)$ such that $\mu(e) := y_e / v$ for each $e \in f^{-1}(1)$. Since $\sum_{e \in \varphi^{-1}(1)} y_e = v \cdot \Pr_{e \sim \mu}[\varphi(e) = 1]$, we can rewrite the dual problem as the following optimization problem over all distributions μ on $f^{-1}(1)$:

$$\begin{aligned}
&\text{maximize} && v && \\
&\text{subject to} && v \le \left(\Pr_{e \sim \mu}[\varphi(e) = 1] / |\varphi| \right)^{-1} && \text{for all } \varphi \in \mathrm{CNF}_f,
\end{aligned}$$

[5] Since $0 < \mathrm{L}_3(f) \le (1 + n \cdot \ln 2) \cdot s^*$, we have $s^* > 0$. Thus, we may assume, without loss of generality, that $v > 0$.

whose optimal value is clearly equal to $(D(f))^{-1}$. Hence, the optimal value s^* of the dual problem is equal to $(D(f))^{-1}$. $\square$

3 Computing Hard Distributions

In this section, we give some examples for which one of hard distributions can be determined. If a function f is "symmetric" in a certain sense, one can reduce the number of variables, as in Yao's minimax principle [Yao77]. The following lemma gives such a general method.

Lemma 7. *Let $f\colon \{0,1\}^n \to \{0,1\}$. Let Π be a subgroup of the symmetric group on $\{0,1\}^n$ such that, for each $\pi \in \Pi$ and each $\varphi \in \mathrm{CNF}_f$, there exists a formula $\varphi_\pi \in \mathrm{CNF}_f$ such that $\varphi = \varphi_\pi \circ \pi$, $f = f \circ \pi$, and $|\varphi_\pi| = |\varphi|$. Then, there exists a distribution μ on $f^{-1}(1)$ such that $\mu \circ \pi = \mu$ for any $\pi \in \Pi$ and $D(f) = D_\mu(f)$.*

Proof. Let μ^* be a "hard" distribution μ such that $D_{\mu^*}(f) = D(f)$. Define a distribution μ on $f^{-1}(1)$ as $\mu(x) = \mathbb{E}_{\pi\sim\Pi}[\mu^*(\pi(x))]$ for each $x \in f^{-1}(1)$, where the probability is taken over the uniform distribution over Π. Since Π is a subgroup of the symmetric group on $\{0,1\}^n$, for any $\pi \in \Pi$ and $x \in f^{-1}(1)$, we have $\mu(\pi(x)) = \mathbb{E}_{\pi'\sim\Pi}[\mu^*(\pi'(\pi(x)))] = \mathbb{E}_{\sigma\sim\Pi}[\mu^*(\sigma(x))] = \mu(x)$, where, in the second equality, we replaced $\pi' \circ \pi$ by σ and used the fact that Π is a group.

By the definition of μ^*, we have $D_{\mu^*}(f) \le D_\mu(f)$. Hence, it suffices to claim that $D_\mu(f) \le D_{\mu^*}(f)$. Indeed, for any $\varphi \in \mathrm{CNF}_f$,

$$
\begin{aligned}
\Pr_{x\sim\mu}[\varphi(x)=1] &= \sum_{x\in f^{-1}(1)} \mu(x)\varphi(x) \\
&= \sum_{x\in f^{-1}(1)} \mathop{\mathbb{E}}_{\pi\sim\Pi}[\mu^*(\pi(x))]\varphi(x) && \text{(by the definition of } \mu) \\
&= \sum_{y\colon \pi^{-1}(y)\in f^{-1}(1)} \mathop{\mathbb{E}}_{\pi\sim\Pi}[\mu^*(y)]\varphi(\pi^{-1}(y)) && \text{(by defining } y := \pi(x)) \\
&= \sum_{y\in f^{-1}(1)} \mathop{\mathbb{E}}_{\pi\sim\Pi}[\mu^*(y)]\varphi_\pi(y) && \text{(by } f\circ\pi = f \text{ and } \varphi = \varphi_\pi\circ\pi) \\
&= \mathop{\mathbb{E}}_{\pi\sim\Pi}\left[\sum_{y\in f^{-1}(1)} \mu^*(y)\varphi_\pi(y)\right] \\
&\le \mathop{\mathbb{E}}_{\pi\sim\Pi}[D_{\mu^*}(f)\cdot|\varphi_\pi|] = D_{\mu^*}(f)\cdot|\varphi| && \text{(by } |\varphi| = |\varphi_\pi|)
\end{aligned}
$$

It follows that $D_\mu(f) = \max_{\varphi\in\mathrm{CNF}_f} \Pr_{x\sim\mu}[\varphi(x)=1]/|\varphi| \le D_{\mu^*}(f)$. $\square$

In particular, if a function $f\colon \{0,1\}^n \to \{0,1\}$ is symmetric in the sense that the value of $f(x)$ depends only on the number of ones in x, then one of the hardest distributions μ is also symmetric.

Corollary 8. *If a function $f\colon \{0,1\}^n \to \{0,1\}$ is symmetric, then there exists a symmetric distribution μ such that $D(f) = D_\mu(f)$.*

Proof. Let Π be the set of all permutations that are induced by the $n!$ permutations on $(x_1, \cdots, x_n)$, and apply Lemma 7. $\square$

In the case of $\textsc{Parity}_n$, we can completely determine one of the hardest distributions.

Corollary 9. *Let μ be the uniform distribution on $\textsc{Parity}_n^{-1}(1)$. Then we have $D(\textsc{Parity}_n) = D_\mu(\textsc{Parity}_n)$.*

Proof. Let Π be the set of all permutations $\pi\colon \{0,1\}^n \to \{0,1\}^n$ that negates an even number of coordinates (e.g., $\pi(x) = \neg x_1 \cdot \neg x_2 \cdot x_3 \cdots x_n$). Note that any permutation $\pi \in \Pi$ does not change the parity. By Lemma 7, there exists a distribution on $f^{-1}(1)$ such that $\mu \circ \pi = \mu$ for any $\pi \in \Pi$ and $D(f) = D_\mu(f)$. We claim that $\mu(x) = \mu(10^{n-1})$ for any $x \in \textsc{Parity}_n^{-1}(1)$. Indeed, it is easy to see that, for any input x whose parity is 1, there exists a permutation $\pi_x \in \Pi$ that maps x to 10^{n-1}. Since $\mu \circ \pi_x = \mu$, we have $\mu(x) = \mu(\pi_x(x)) = \mu(10^{n-1})$. Hence, μ is the uniform distribution on $\textsc{Parity}_n^{-1}(1)$. □

More generally, Corollary 9 establishes an equivalence between depth-$(d+1)$ formula lower bounds and one-sided-error correlation bounds of depth-d formulas with $\textsc{Parity}_n$ on the uniform distribution. In particular, we obtain one-sided-error correlation bounds for $\mathsf{AC}^0[p]$ as stated in Theorem 4.

Proof of Theorem 4. We apply the duality theorem (Theorem 6) to $\mathrm{OR} \circ \mathsf{AC}^0[p]$ (namely, the class of circuits that consist of a top OR gate fed by disjoint $\mathsf{AC}^0[p]$ circuits). Recall that $D^{\mathsf{AC}^0[p]}(f)$ denotes the one-sided-error correlation per size of $\mathsf{AC}^0[p]$ circuits with the function f. Theorem 6 shows that $D^{\mathsf{AC}^0[p]}(f)$ is at most $O(n)$ times the inverse of the size of $\mathrm{OR} \circ \mathsf{AC}^0[p]$ circuits for computing f. Now we apply Smolensky's lower bound [Smo87] on the parity function to $\mathrm{OR}\circ\mathsf{AC}^0[p]$ circuits, and obtain $\Pr[C(x)=1]/|C| = D^{\mathsf{AC}^0[p]}(\textsc{Parity}_n) \le (1 + n \cdot \ln 2) \cdot 2^{-n^{\Omega(1/(d+1))}} = 2^{-n^{\Omega(1/d)}}$ for any $\mathsf{AC}^0[p]$ circuit C of depth d. Here, by Corollary 9, we may assume that the probability is taken over the uniform distribution μ on $\textsc{Parity}_n^{-1}(1)$. Therefore,

$$\begin{aligned}
&\Pr_{x\sim\{0,1\}^n}[C(x) = \textsc{Parity}_n(x)] \\
&= \frac{1}{2} + \frac{1}{2}\Pr_{x\sim\mu}[C(x)=1] && \text{(since } C(x)=0 \text{ if } \textsc{Parity}_n(x)=0) \\
&\le \frac{1}{2} + \frac{1}{2}\cdot|C|\cdot 2^{-n^{\Omega(1/d)}}.
\end{aligned}$$

□

We can also compute one of the hardest distributions for computing the majority function by *monotone* formulas. We write S^n_k for the set of all inputs $x \in \{0,1\}^n$ such that the number of ones in x is k.

Proposition 10. *Let μ be the uniform distribution on $S^n_{\lceil n/2\rceil}$. Then, $D^+(\textsc{Maj}_n) = D^+_\mu(\textsc{Maj}_n)$.*

Proof. As observed in [BHST14], for any monotone function f, it holds that $\Pr_{x\sim S^n_k}[f(x)=1] \le \Pr_{x\sim S^n_{k+1}}[f(x)=1]$ for any $k < n$. Indeed, by double counting,

$$\begin{aligned}
\#\{\, x \in S^n_k \mid f(x)=1 \,\}\cdot(n-k) &= \#\{\, (x,y) \in S^n_k \times S^n_{k+1} \mid f(x)=1, x \le y \,\} \\
&\le \#\{\, (x,y) \in S^n_k \times S^n_{k+1} \mid f(y)=1, x \le y \,\} \\
&= \#\{\, y \in S^n_{k+1} \mid f(y)=1 \,\}\cdot(k+1),
\end{aligned}$$

and hence $\Pr_{x\sim S_k^n}[f(x)=1] \le \Pr_{x\sim S_{k+1}^n}[f(x)=1]$ (by noting that $\frac{n-k}{k+1}\binom{n}{k} = \binom{n}{k+1}$). Therefore, the uniform distribution on S_k^n is "harder" than S_{k+1}^n for monotone formulas. Details follow.

Since $\textsc{Maj}_n$ is a symmetric function, by Corollary 8 there exists a symmetric distribution μ^* such that $D^+(\textsc{Maj}_n) = D^+_{\mu^*}(\textsc{Maj}_n)$. (Note that, while Corollary 8 is stated for nonmonotone formulas, the same proof can be applied to the case of monotone formulas.) Let φ be a monotone CNF formula that achieves the maximum one-sided correlation per size $D^+_{\mu^*}(\textsc{Maj}_n)$. Then,

$$\begin{aligned}
D^+_{\mu^*}(\textsc{Maj}_n)\cdot|\varphi| &= \Pr_{x\sim\mu^*}[\varphi(x)=1] \\
&= \sum_{k=\lceil n/2\rceil}^{n} \Pr_{x\sim\mu^*}[x\in S_k^n]\cdot \Pr_{x\sim S_k^n}[\varphi(x)=1] && \text{(since μ^* is symmetric)} \\
&\ge \sum_{k=\lceil n/2\rceil}^{n} \Pr_{x\sim\mu^*}[x\in S_k^n]\cdot \Pr_{x\sim S_{\lceil n/2\rceil}^n}[\varphi(x)=1] && \text{(since φ is monotone)} \\
&= \Pr_{x\sim\mu}[\varphi(x)=1].
\end{aligned}$$

Hence, we have $D^+(\textsc{Maj}_n) \le D^+_{\mu}(\textsc{Maj}_n) \le D^+_{\mu^*}(\textsc{Maj}_n) = D^+(\textsc{Maj}_n)$. $\square$

We can state the one-sided correlation per size $D^+(\textsc{Maj}_n)$ in terms of hypergraphs. Specifically, the following holds.

Corollary 11. *For a hypergraph $\mathcal{F}$, let $|\mathcal{F}|$ denote the number of edges in $\mathcal{F}$, $\tau(\mathcal{F})$ denote the minimum size of hitting sets of $\mathcal{F}$, and $t(\mathcal{F})$ denote the number of minimum hitting sets of $\mathcal{F}$. Let $T(n,\tau) := \max_{\mathcal{F}\colon \tau(\mathcal{F})=\tau} t(\mathcal{F})/|\mathcal{F}|$, where the maximum is taken over all hypergraphs of n vertices. Let $\mathrm{L}_3^+(\textsc{Maj}_n)$ be the minimum depth-3 monotone formula size for computing $\textsc{Maj}_n$. Then, the following holds:*

$$\frac{1}{T(n,\lceil n/2\rceil)} \le \frac{\mathrm{L}_3^+(\textsc{Maj}_n)}{\binom{n}{\lceil n/2\rceil}} \le \frac{1+\ln 2\cdot n}{T(n,\lceil n/2\rceil)}.$$

Proof. Given a monotone CNF formula φ, we define a hypergraph $\mathcal{F}$ so that each edge of $\mathcal{F}$ is the set of literals in a clause of φ. (Note that, since φ is a monotone CNF, every literal is a positive.) We can naturally identify $x\in\{0,1\}^n$ with $x\subseteq[n]$. It is easy to see that φ accepts $x\in\{0,1\}^n$ if and only if $x\subseteq[n]$ is a hitting set of $\mathcal{F}$. Therefore, the constraint that $\varphi^{-1}(1)\subseteq \textsc{Maj}_n^{-1}(1)$ corresponds to the constraint that $\tau(\mathcal{F})=\lceil n/2\rceil$. The number $|\mathcal{F}|$ of edges in $\mathcal{F}$ is equal to $|\varphi|$. Moreover, $t(\mathcal{F}) = \binom{n}{\lceil n/2\rceil}\cdot\Pr_{x\sim S^n_{\lceil n/2\rceil}}[\varphi(x)=1]$. Therefore, $T(n,\lceil n/2\rceil) = \binom{n}{\lceil n/2\rceil}D^+(\textsc{Maj}_n)$ and the result follows from Theorem 6. $\square$

In the case of graphs (instead of hypergraphs), the extremal structure that maximizes the number of hitting sets under the constraint that the minimum size of hitting sets is bounded from below is well understood. Thus, in the case of graphs, or equivalently, in the case of bottom fan-in 2, we are able to determine the minimum depth-3 circuit size for computing the majority function.

Proposition 12. *The minimum size of depth-3 monotone circuits with bottom fan-in 2 for computing $\textsc{Maj}_n$ is $\widetilde{\Theta}(2^{n/2})$.*

Proof. It is well known that the extremal structure that maximizes the number of hitting sets (of each cardinality) given the constraint that $\tau(\mathcal{F}) = \tau$ is the complement of Turán's graph [Tur41] (i.e., a disjoint union of cliques of almost equal size; hence, in our case, it is a disjoint union of $n/2$ edges). (This fact was independently proved by many people, such as Zykov [Zyk49], Erdős [Erd62], Sauer [Sau71], Hadžiivanov [Had76], and Roman [Rom76]. For a recent simple proof, see Cutler and Radcliffe [CR11].)

The number of minimum hitting sets is $O(2^{n/2})$. Since the number of edges in a graph of n vertices is at most $n^2/2$, the maximum correlation per size is $\widetilde{\Theta}(2^{-n/2})$. Moreover, depth-3 circuit size and depth-3 formula with bottom fan-in 2 are the same up to a factor of $O(n^2)$ (e.g., see [PSZ00, Proposition 2.2]). Thus, the minimum circuit size is equal to $2^{n/2}$ up to a polynomial factor. $\square$

We remark that the same quantitative bound can be obtained by using the techniques of [HJP95]. Proposition 12 merely gives its alternative proof from the techniques of the different literature.

4 Approximating Monotone Functions by Depth-2 Formulas

It is reasonable to conjecture that an analog of Proposition 10 holds for *nonmonotone* CNF formulas: that is, we conjecture that the uniform distribution on $S^n_{\lceil n/2 \rceil}$ is close to the hardest distribution of CNF formulas approximating $\textsc{Maj}_n$. And it is tempting to guess that, since $\textsc{Maj}_n$ is a monotone function, an optimal CNF formula approximating $\textsc{Maj}_n$ should be a monotone CNF formula. This is true when *exact* complexity of computing monotone function is concerned, by Quine's theorem [Qui54]. It is, however, not true when approximation is concerned: Blais, Håstad, Servedio and Tan [BHST14] showed that there is a monotone function on n variables that can be approximated by some CNF formula of size $O(n)$, but cannot be approximated by any monotone CNF formula of size less than n^2.

Here we significantly improve their bounds:

Theorem 13 (Theorem 3, restated formally). *There exists a function $\epsilon \colon \mathbb{N} \to (0,1)$ and a family of monotone Boolean functions $f = \{f_n \colon \{0,1\}^n \to \{0,1\}\}_{n\in\mathbb{N}}$ and a family of distributions $\{\mu_n\}_{n\in\mathbb{N}}$ on $f_n^{-1}(1)$ such that*

- *there is some CNF formula $\{\varphi_n\}_{n\in\mathbb{N}}$ of size at most $n^{O(1)}$ satisfying $\varphi_n^{-1}(1) \subseteq f_n^{-1}(1)$ and $\Pr_{x\sim\mu_n}[\varphi_n(x) = 1] \le \epsilon(n)$, whereas*
- *any monotone CNF formula $\{\varphi_n^+\}_{n\in\mathbb{N}}$ satisfying $(\varphi_n^+)^{-1}(1) \subseteq f_n^{-1}(1)$ and $\Pr_{x\sim\mu_n}[\varphi_n^+(x) = 1] \le \epsilon(n)$ requires size $2^{n^{\Omega(1)}}$.*

In fact, this theorem is an immediate consequence of our duality theorem together with recently improved Ajtai-Gurevich's theorem [AG87].

Theorem 14 (Chen, Oliveira and Servedio [COS15]). *There exists a family of monotone functions $f = \{f_n \colon \{0,1\}^n \to \{0,1\}\}_{n\in\mathbb{N}}$ such that*

1. $\mathrm{L}_3(f_n) = n^{O(1)}$*, and*

2. *any monotone depth-d circuit*[6] *computing f requires size $2^{n^{\Omega(1/d)}}$.*

Using our duality theorem, we now transfer Theorem 14 about depth-3 circuits to Theorem 13 about approximation by depth-2 formulas.

Proof of Theorem 13. By the second item of Theorem 14, $\mathrm{L}_3^+(f_n) \geq 2^{n^{\Omega(1)}}$. By Theorem 6, $D^+(f_n) \leq O(n)/\mathrm{L}_3^+(f_n) = 2^{-n^{\Omega(1)}}$. That is, there exists a distribution μ_n on $f_n^{-1}(1)$ such that

$$\Pr_{x\sim\mu_n}[\varphi^+(x) = 1] \leq |\varphi^+| \cdot 2^{-n^{\Omega(1)}}$$

for any monotone CNF formula φ^+. In particular, for any monotone CNF formula φ^+ of size less than $2^{\frac{1}{2}\cdot n^{\Omega(1)}}$, we have $\Pr_{x\sim\mu_n}[\varphi^+(x) = 1] \leq 2^{-\frac{1}{2}n^{\Omega(1)}} = 2^{-n^{\Omega(1)}}$.

On the other hand, by Theorem 6 and the first item of Theorem 14, we obtain $D(f) \geq 1/\mathrm{L}_3(f) \geq n^{-O(1)}$. Hence, for the distribution μ_n defined above, there exists a CNF formula φ_n such that

$$\Pr_{x\sim\mu_n}[\varphi_n(x) = 1] \geq |\varphi_n| \cdot n^{-O(1)} \geq n^{-O(1)} \tag{1}$$

since $|\varphi_n| \geq 1$. Define $\epsilon(n) = n^{-O(1)}$ to be the rightmost lower bound in (1). □

5 Universal Bounds on Approximation by CNFs

In this section, we prove a tight bound on the maximum size of CNF formulas approximating a function with one-sided error. We say that a CNF formula φ approximates a function f with (one-sided error and) advantage ϵ if $\Pr_{x\sim f^{-1}(1)}[\varphi(x) = 1] \geq \epsilon$ and $\varphi^{-1}(1) \subseteq f^{-1}(1)$.

Theorem 15 (Theorem 2, restated formally)**.** *For all large $n \in \mathbb{N}$, the following holds.*

- *For any function $f: \{0,1\}^n \to \{0,1\}$, there exists some $\epsilon \in (0,1]$ and some CNF formula of size at most $\epsilon 2^{n+3}/n$ approximating f with one-sided error and advantage ϵ.*
- *There exists a function $f: \{0,1\}^n \to \{0,1\}$ such that, for any $\epsilon \in (0,1]$, any CNF formula approximating f with one-sided error and advantage ϵ must have size at least $\epsilon 2^{n-7}/n$.*

In order to obtain the upper bound, we prove the following lemma.

Lemma 16. $\mathrm{L}_3(f) \leq 2^{n+3}/n$ *for any function f.*

A classical theorem by Lupanov [Lup65] states that, for any function f, the number of literals in the smallest depth-3 formula computing f is at most $O(2^n/\log n)$ (and this is tight). In contrast, Lemma 16 deals with the number of AND gates at the middle layer. To the best of our knowledge, for this size measure, the universal upper bound was not studied before; however, Lupanov's construction can be adapted to our size measure as well, by changing some parameters in the construction.

[6] In fact, their result gives a lower bound for monotone *majority* circuits. For our purpose, it is sufficient to use a lower bound for monotone OR ∘ AND ∘ OR circuits.

Proof of Lemma 16. The main idea is to cover the whole space by a sphere of diameter 1, and then to describe the function inside each sphere by a relatively small CNF formula. Take a parameter $D = 2^d \ (< n)$ (which is fixed later) for some $d \in \mathbb{N}$. A sphere of diameter 1 with center a is denoted by $S_a \subseteq \{0,1\}^D$. That is, S_a consists of all the strings $y \in \{0,1\}^D$ such that y and a differ in exactly one coordinate. We use several simple facts from [Lup65].

Fact 17. *There is some subset $A_D \subseteq \{0,1\}^D$ such that $\{ S_a \mid a \in A_D \}$ partitions $\{0,1\}^n$ (i.e., $\coprod_{a \in A_D} S_a = \{0,1\}^n$).*

Proof Sketch. Let $H \in \mathrm{GF}(2)^{d \times 2^d}$ be a $d \times 2^d$ matrix whose ith column is the d-bit binary representation of $i \in [2^d]$. Define A_D to be the kernel of a linear map H regarded as $H \colon \mathrm{GF}(2)^{2^d} \to \mathrm{GF}(2)^d$. The fact follows from the properties of the Hamming code [Ham50]. □

Fact 18. *For any $a \in \{0,1\}^D$, the characteristic function of S_a can be computed by a CNF formula φ_a of size at most D^2.*

Proof Sketch. $x \in S_a$ if and only if (1) there is some $i \in [D]$ such that $x_i \neq a_i$ and (2) for any pair of $i < j$, $x_i = a_i$ or $x_j = a_j$. □

The next fact states that, if restricted to a sphere S_a, any function g can be described by a single clause C_a^g.

Fact 19. *For any function $g \colon \{0,1\}^D \to \{0,1\}$ and $a \in \{0,1\}^D$, there exists a clause C_a^g such that $g(y) = C_a^g(y)$ for any $y \in S_a$.*

Proof Sketch. For $b \in \{0,1\}$, let y_i^b denote a positive literal y_i if $b = 1$ and a negative literal $\neg y_i$ otherwise. Define $C_a^g(y) := \bigvee_i y_i^{1-a_i}$ where the disjunction is taken over all $i \in [D]$ such that g evaluates to 1 if the ith coordinate of a is flipped (i.e., $(a_1, \dots, 1 - a_i, \dots a_D) \in g^{-1}(1)$). □

Using these facts, we can now describe any function $f \colon \{0,1\}^n \to \{0,1\}$ by a small depth-3 formula. Regard $\{0,1\}^n = \{0,1\}^D \times \{0,1\}^{n-D}$. For $y \in \{0,1\}^D$ and $z \in \{0,1\}^{n-D}$,

$$
\begin{aligned}
f(y,z) &= f(y,z) \wedge \left(\bigvee_{a \in A_D} \varphi_a(y) \right) && \text{(by Fact 17)} \\
&= \bigvee_{a \in A_D} \varphi_a(y) \wedge f(y,z) \\
&= \bigvee_{a \in A_D} \left\{ \varphi_a(y) \wedge \bigwedge_{w \in \{0,1\}^{n-D}} \left(f(y,w) \vee \bigvee_{j \in [n-D]} z_j^{1-w_j} \right) \right\} \\
&= \bigvee_{a \in A_D} \left\{ \varphi_a(y) \wedge \bigwedge_{w \in \{0,1\}^{n-D}} \left(C_a^{f(\text{-},w)}(y) \vee \bigvee_{j \in [n-D]} z_j^{1-w_j} \right) \right\}. && \text{(by Fact 19)}
\end{aligned}
$$

This is a depth-3 formula of size at most $|A_D| \cdot (|\varphi_a| + 2^{n-D}) \leq 2^D D + 2^n/D$. Define $d := \lfloor \log(n/2) \rfloor$. Then, $n/4 \leq D = 2^d \leq n/2$ and hence the size is at most $2^{n/2} \cdot n/2 + 4 \cdot 2^n/n \leq 2^{n+3}/n$. This completes the proof of Lemma 16. □

Remark. Lemma 16 is tight up to a constant factor. Indeed, a simple counting shows that the number of depth-3 formulas of size s is at most $2^{O(s \log s + sn)}$, which is much less than 2^{2^n} for $s \ll 2^n/n$.

Proof of Theorem 15.

1. By Lemma 16 and Theorem 6, we obtain $(D(f))^{-1} \le \mathrm{L}_3(f) \le 2^{n+3}/n$. Hence, for any distribution μ on $f^{-1}(1)$ (and in particular the uniform distribution on $f^{-1}(1)$), there exists some CNF formula φ such that

$$|\varphi| \le 2^{n+3}/n \cdot \Pr_{x\sim\mu}[\varphi(x) = 1].$$

Therefore, φ approximates f with advantage ϵ for $\epsilon := \Pr_{x\sim\mu}[\varphi(x) = 1]$, and $|\varphi| \le \epsilon \cdot 2^{n+3}/n$.

2. Let $f\colon \{0,1\}^n \to \{0,1\}$ be a random function. That is, for each input $x \in \{0,1\}^n$, pick $f(x) \sim \{0,1\}$ uniformly at random and independently. Fix any CNF formula φ and advantage ϵ. We will bound the probability that φ approximates f with one-sided error and advantage ϵ.

Claim 20. $\Pr_f\left[\ \varphi^{-1}(1) \subseteq f^{-1}(1)\ \textit{and}\ \Pr_{x\sim f^{-1}(1)}[\varphi(x) = 1] \ge \epsilon\ \right] \le 2^{-\epsilon 2^{n-4}}$

By definition, $\Pr_f[\varphi^{-1}(1) \subseteq f^{-1}(1)] = 2^{-|\varphi^{-1}(1)|}$. This probability is bounded above by $2^{-\epsilon 2^{n-2}}$ if $|\varphi^{-1}(1)| > \epsilon 2^{n-2}$, in which case the claim holds.

It remains to consider the case when $|\varphi^{-1}(1)| \le \epsilon 2^{n-2}$. First, note that $\Pr_{x\sim f^{-1}(1)}[\varphi(x) = 1] \ge \epsilon$ is equivalent to $|\varphi^{-1}(1)| \ge \epsilon |f^{-1}(1)|$ under the assumption that $\varphi^{-1}(1) \subseteq f^{-1}(1)$. Therefore, the probability in the claim can be bounded above by

$$\Pr\left[\epsilon |f^{-1}(1)| \le |\varphi^{-1}(1)|\right] \le \Pr\left[|f^{-1}(1)| \le 2^{n-2}\right] \le 2^{-2^{n-4}} \le 2^{-\epsilon 2^{n-4}},$$

where the second last inequality follows from the Chernoff bound. This completes the proof of the claim.

Fix any size $s \in \mathbb{N}$. Since there are at most 3^{ns} CNF formulas of size s,

$$\Pr_f\left[\ \exists\,\varphi \text{ of size } s \text{ such that } \varphi^{-1}(1) \subseteq f^{-1}(1) \text{ and } \Pr_{x\sim f^{-1}(1)}[\varphi(x) = 1] \ge \epsilon\ \right] \le 3^{ns} \cdot 2^{-\epsilon 2^{n-4}}.$$

by the union bound. Define $s_\epsilon := \epsilon 2^{n-6}/n$. Then, for any fixed ϵ, the probability that there exists a CNF formula φ of size s_ϵ approximating f with advantage ϵ is at most $3^{ns_\epsilon} \cdot 2^{-\epsilon 2^{n-4}} \le 2^{\epsilon 2^{n-5} - \epsilon 2^{n-4}} = 2^{-\epsilon 2^{n-5}}$.

Let $\mathcal{E} := \{\, 2^{-n+5}, 2^{-n+6}, \cdots, 2^0 \,\}$. By the union bound over all $\epsilon \in \mathcal{E}$,

$$\begin{aligned}&\Pr_f[\ \exists\,\epsilon \in \mathcal{E},\ \exists\,\varphi \text{ of size } s_\epsilon \text{ such that } \varphi \text{ approximates } f \text{ with advantage } \epsilon\ ]\\ &\le \sum_{i=0}^{n-5} 2^{-2^i} < \sum_{i=1}^{\infty} 2^{-i} = 1.\end{aligned}$$

Hence, there exists a function f such that f cannot be approximated by any CNF formula of size s_ϵ with any advantage $\epsilon \in \mathcal{E}$. While this gives us inapproximability for discrete values of $\epsilon \in \mathcal{E}$, we can extend it to an arbitrary advantage $\epsilon \in (0,1]$ as follows: Given an arbitrary

$\epsilon \geq 2^{-n+5}$, we take $\epsilon' \in \mathcal{E}$ such that $\epsilon/2 \leq \epsilon' \leq \epsilon$. Then, f cannot be approximated by any CNF formula of size $s_\epsilon/2$ $(\leq s_{\epsilon'})$ with any advantage ϵ $(\geq \epsilon')$. On the other hand, if $\epsilon < 2^{-n+5}$ then $s_\epsilon < 1$ and hence f cannot be approximated by any CNF formula of size at most s_ϵ (i.e., a constant formula) with positive advantage, as we may assume that f is not constant. Therefore, in any case, f cannot be approximated by any CNF formula of size $s_\epsilon/2 = \epsilon 2^{n-7}/n$ with any advantage $\epsilon \in (0,1]$.

□

6 Satisfiability Coding Lemma With Width Reduction

In this section, we modify the satisfiability coding lemma [PPZ99] and obtain an upper bound on the number of isolated solutions of a CNF formula. Here, we say that an assignment $x \in \{0,1\}^n$ is an *isolated solution* of a function φ if $\varphi(x) = 1$ and $\varphi(y) = 0$ for any adjacent assignment y of x (i.e., x and y differ on exactly one coordinate). We note that an upper bound on the number of isolated solutions immediately implies a one-sided correlation bound of PARITY_n.

Paturi, Pudlák, and Zane [PPZ99] developed the satisfiability coding lemma, which states that an isolated solution has a short description (and thus the number of isolated solutions is small). We say that a randomized algorithm E is a *randomized prefix-free encoding* if, for any fixed randomness r of E, the image of the algorithm E_r that uses r as randomness is prefix-free (i.e., no two strings in the image of E_r contain the other as a prefix).

Lemma 21 (Satisfiability Coding Lemma [PPZ99])**.** *Let φ be any k-CNF formula on n variables, and let $T \subseteq \{0,1\}^n$ be the set of isolated solutions of φ. Then, there exists a randomized prefix-free encoding $E(\text{-};\varphi)\colon T \to \{0,1\}^*$ such that $\mathbb{E}[E(x;\varphi)] \leq n - n/k$ for any $x \in T$, where the expectation is taken over the coin flips of E. In particular, $|T| \leq 2^{n-n/k}$.*

In order to derive a depth-3 formula lower bound of PARITY_n, we need an upper bound in terms of the formula size $|\varphi|$ instead of the width k. In the context of satisfiability algorithms (i.e., decoding algorithms of satisfying assignments), Schuler [Sch05] gave a variant of the PPZ algorithm [PPZ99] that runs in time $2^{n-n/(\log|\varphi|+1)}\mathsf{poly}(n)$ for a CNF formula φ (instead of the running time $2^{n-n/k}\mathsf{poly}(n)$ of the PPZ algorithm where k is the width of φ). We note that Calabro, Impagliazzo and Paturi [CIP06] gave another analysis of Schuler's width reduction technique. Their analysis gives a satisfiability algorithm that runs in time $2^{n-n/O(\log(|\varphi|/n))}\mathsf{poly}(n)$ (see [DH09]). However, it seems that their analysis does not improve our depth-3 formula lower bound on the parity function because the "O" notation hides some factor. We thus incorporate Schuler's width reduction technique into the satisfiability coding lemma, and obtain the following:

Theorem 22. *Let φ be a CNF formula of size at most 2^s on n variables. Then the number of isolated solutions of φ is at most $2^{n-n/(s+2)+1}$.*

Proof. We will construct a randomized prefix-free encoding of the isolated solution of φ. Let T be the set of isolated solutions of φ. Let $k := s+2$.

The idea is to cut off long clauses of length greater than k in φ and then apply the satisfiability coding lemma. We will define $\widetilde{\varphi}$ to be a formula produced by cutting off long clauses in φ so that the width of $\widetilde{\varphi}$ is at most k. The satisfiability coding lemma yields a short encoding whose average length is $n - n/k$. However, since we cut off long clauses in φ, it is possible that a satisfying assignment x of φ is not necessarily a satisfying assignment of $\widetilde{\varphi}$. In this case, we specify the index of the clause in $\widetilde{\varphi}$ that is not satisfied by x. While it costs us additional s bits on the code length, we can obtain the k bits of information about x from the fact that x does not satisfy the clause.

To be more precise, the definition of $\widetilde{\varphi}$ is as follows: For each clause $C = l_1 \vee l_2 \vee \cdots \vee l_m$, define $\widetilde{C}$ as $\widetilde{C} := l_1 \vee l_2 \vee \cdots \vee l_k$ if $m > k$, and $\widetilde{C} := C$ otherwise. (We fix an arbitrary ordering of literals so that $\widetilde{C}$ is uniquely determined.) Then, for a CNF formula $\varphi = \bigwedge_{i=1}^{|\varphi|} C_i$, let us define a k-CNF formula $\widetilde{\varphi}$ as $\widetilde{\varphi} := \bigwedge_{i=1}^{|\varphi|} \widetilde{C}_i$.

Now we describe the prefix-free encoding algorithm $\widetilde{E}(x;\varphi)$ for encoding an isolated solution x of φ. If $\widetilde{\varphi}(x) = 1$, then apply the satisfiability coding lemma, that is, output $0 \cdot E(x;\varphi)$ ("0" is a marker indicating the first case). Otherwise, there exists a clause $\widetilde{C}$ in $\widetilde{\varphi}$ such that $\widetilde{C}(x) = 0$. Let $\widetilde{C} = \bigvee_{i=1}^{k} l_i$ be the lexicographically first clause in $\widetilde{\varphi}$ such that $\widetilde{C}(x) = 0$. (Note that the width of $\widetilde{C}$ is exactly equal to k because $\widetilde{C} \neq C$.) Let $i \in \{0,1\}^s$ be the binary encoding of the index of $\widetilde{C}$ in $\widetilde{\varphi}$. Let $\varphi' := \varphi|_{l_1=0,\ldots,l_k=0}$ (i.e., assign the variables in $\widetilde{C}$ so that $\widetilde{C}$ is falsified), and let $x' \in \{0,1\}^{n-k}$ be the part of the assignment x other than variables in $l_1, \ldots, l_k$. Output $1 \cdot i \cdot \widetilde{E}(x';\varphi')$. (That is, recursively call this procedure on input x' and φ'.)

We claim that $\widetilde{E}(\text{-};\varphi)\colon T \to \{0,1\}^*$ is a (randomized) prefix-free encoding. To this end, we describe a decoding algorithm $\widetilde{D}(z;\varphi)$: Given an input $z = \widetilde{E}(x;\varphi)$, read the first bit $b \in \{0,1\}$ of z. If $b = 0$, then use the decoding algorithm for $E(\text{-},\varphi)$. Otherwise, read the next s bits $i \in \{0,1\}^s$ of z so that $z = 1 \cdot i \cdot z'$, let $\widetilde{C}$ denote the ith clause in $\widetilde{\varphi}$, define $\varphi' := \varphi|_{\widetilde{C}=0}$ as in the encoding algorithm of $\widetilde{E}$, and recursively call $\widetilde{D}(z';\varphi')$. It is easy to see that $\widetilde{D}(\widetilde{E}(x;\varphi);\varphi) = x$ for any $x \in T$, and thus $\widetilde{E}$ is a prefix-free encoding.

Finally, we analyze the average code length of the prefix-free encoding $\widetilde{E}$. We claim, by induction on the number of recursive calls of $\widetilde{E}$, that $\mathbb{E}\left[\widetilde{E}(x;\varphi)\right] \le n - n/k + 1$ for any isolated solution x of φ. Here, the expectation is taken over the randomness of $\widetilde{E}$ (i.e., the randomness of E). If $\widetilde{\varphi}(x) = 1$ (i.e., there is no recursive call), then $|\widetilde{E}(x;\varphi)| = 1 + |E(x;\varphi)|$, and thus $\mathbb{E}\left[|\widetilde{E}(x;\varphi)|\right] \le 1 + n - n/k$ by Lemma 21. Otherwise, we have $|\widetilde{E}(x;\varphi)| = 1 + s + |\widetilde{E}(x';\varphi')|$, where x' and φ' are defined as in the encoding algorithm. Since x' is an isolated solution of φ', we can apply the induction hypothesis for $x' \in \{0,1\}^{n-k}$ and φ'. Hence,

$$\begin{aligned}\mathbb{E}[|\widetilde{E}(x;\varphi)|] &\le 1 + s + (n-k) - (n-k)/k + 1 \\ &= 1 + n - n/k,\end{aligned}$$

which completes the induction.

We have proved that $\mathbb{E}_r\left[\widetilde{E}_r(x;\varphi)\right] \le n - n/k + 1$ for any $x \in T$, where r denotes the internal randomness of $\widetilde{E}$. In particular, the same holds even if the expectation is taken over $x \in T$ as well as internal randomness of E. By averaging, there exists some internal randomness r' of E such that $\mathbb{E}_{x\sim T}\left[\widetilde{E}_{r'}(x;\varphi)\right] \le n - n/k + 1$. Thus, $\widetilde{E}_{r'}(\text{-};\varphi)\colon T \to \{0,1\}^*$ is a deterministic prefix-free encoding whose average code length is at most $n - n/k + 1$. By Kraft's inequality and Jensen's inequality, we have

$$1 \ge \sum_{x\sim T} 2^{-|\widetilde{E}_{r'}(x;\varphi)|} = |T| \cdot \mathop{\mathbb{E}}_{x\sim T}\left[2^{-|\widetilde{E}_{r'}(x;\varphi)|}\right] \ge |T| \cdot 2^{\mathbb{E}_{x\sim T}\left[-|\widetilde{E}_{r'}(x;\varphi)|\right]} \ge |T| \cdot 2^{-(n-n/k+1)}$$

Hence, $|T| \le 2^{n-n/k+1}$. $\square$

Corollary 23. *Let $n \in \mathbb{N}$ and $0 \le \epsilon < \frac{1}{2}$. Let s be the minimum size of a CNF formula φ that computes $\textsc{Parity}_n$ all but an ϵ fraction of inputs with one-sided error in the sense that $\varphi^{-1}(1) \subseteq \textsc{Parity}_n^{-1}(1)$. Then, for $k := \log \frac{2}{1-2\epsilon}$ and $k' = \lfloor k \rfloor$,*

$$2^{\frac{n}{k+1}-3} \le s \le k' 2^{\lceil \frac{n}{k'} \rceil - 1}.$$

Proof. We first prove the lower bound. Since $\varphi^{-1}(1) \subseteq \textsc{Parity}_n^{-1}(1)$, any satisfying assignment of φ is an isolated solution. Thus, by Lemma 21, $\Pr_{x \sim \textsc{Parity}_n^{-1}(1)}[\varphi(x) = 1] \le 2^{n-n/(\lceil \log s \rceil + 2)+1}/2^{n-1} \le 2^{-n/(\log s + 3)+2}$. By the assumption,

$$\epsilon \ge \Pr_{x \sim \{0,1\}^n}[\varphi(x) \ne \textsc{Parity}_n(x)] = \frac{1}{2} \Pr_{x \sim \textsc{Parity}_n^{-1}(1)}[\varphi(x) = 0] \ge \frac{1}{2}(1 - 2^{-n/(\log s + 3)+2}),$$

and thus $s \ge 2^{\frac{n}{k+1}-3}$.

For the upper bound, we divide an n-bit input into k' blocks the ith block of which contains n_i bits so that $\sum_{i=1}^{k'} n_i = n$ and $n_i \le \lceil n/k' \rceil$. For the ith block, we add $2^{n_i - 1}$ clauses that check whether the parity of the block is 0 if $i \ne 1$ and 1 if $i = 1$. The constructed CNF formula φ is of size $\sum_{i=1}^{k'} 2^{n_i-1} \le k' 2^{\lceil \frac{n}{k'} \rceil - 1}$ and does not accept any input whose parity is even. The number of the satisfying assignments of φ is exactly equal to $2^{n-k'}$, and hence $\Pr_x[\varphi(x) \ne \textsc{Parity}_n(x)] = \frac{1}{2}(1 - 2^{-k'+1}) \le \epsilon$. □

Now we can determine the depth-3 formula size for computing $\textsc{Parity}_n$.

Theorem 24. $\mathrm{L}_3(\textsc{Parity}_n) = \widetilde{\Theta}(2^{2\sqrt{n}})$. *More precisely,* $2^{2\sqrt{n}-5} \le \mathrm{L}_3(\textsc{Parity}_n) \le O(2^{2\sqrt{n}}\sqrt{n})$.

Proof. Let φ denote any CNF formula such that $\varphi^{-1}(1) \subseteq \textsc{Parity}_n^{-1}(1)$. Let T denote the set of satisfying assignments of φ. Since T corresponds to the set of isolated solutions of φ, by Theorem 22, it follows that $|T| \le 2^{n-n/(\lceil \log |\varphi| \rceil + 2)+1} \le 2^{n-n/(\log|\varphi|+3)+1}$. Therefore, for the uniform distribution μ on $\textsc{Parity}_n^{-1}(1)$, the one-sided correlation per size is

$$\begin{aligned}\Pr_{x\sim\mu}\left[\varphi(x) = 1\right]/|\varphi| &= |T| \,/\, (2^{n-1} \cdot |\varphi|) \\ &\le 2^{5-n/(\log|\varphi|+3)-(\log|\varphi|+3)} \le 2^{5-2\sqrt{n}},\end{aligned}$$

where the last inequality follows from the arithmetic and geometric means. That is, $D(\textsc{Parity}_n) \le D_\mu(\textsc{Parity}_n) \le 2^{5-2\sqrt{n}}$ and thus $2^{2\sqrt{n}-5} \le (D(\textsc{Parity}_n))^{-1} \le \mathrm{L}_3(\textsc{Parity}_n)$ by Theorem 6.

A depth-3 formula for computing $\textsc{Parity}_n$ can be constructed as follows: Divide n variables into n/k blocks each of which contains k variables. For each string $z \in \{0,1\}^{n/k}$ such that $\textsc{Parity}_{n/k}(z) = 1$, we construct a CNF formula φ_z that checks whether, for each $i \in [n/k]$, the parity of the ith block is z_i. Then it is easy to see that $\textsc{Parity}_n = \bigvee_z \varphi_z$. Since $|\varphi_z| \le n/k \cdot 2^{k-1}$, we have $\mathrm{L}_3(\textsc{Parity}_n) \le n/k \cdot 2^{k+n/k-1}$. Choosing $k := \sqrt{n}$, we obtain $\mathrm{L}_3(\textsc{Parity}_n) \le 2^{2\sqrt{n}-1}\sqrt{n}$. □

7 Concluding Remarks

In this paper, we proved that depth-3 circuit size and depth-3 formula size are quadratically different. Rossman [Ros15] separated the depth-d circuits from the depth-d formulas for a sufficiently large d (say, $d \geq 108$). These results leave a mysterious regime $3 < d \ll 108$ as an interesting open problem. Can one prove that depth-d circuits are more powerful than depth-d formulas for such a regime, and in particular, for $d = 4$?

Can one prove a depth-3 formula lower bound of $\textsc{Parity}_n$ that is tight up to a constant factor? Can one extend our lower bound of approximating $\textsc{Parity}_n$ by CNF formulas with one-sided error to the case of two-sided error? What is the asymptotic behavior of $T(n, n/2)$? The techniques from the literature of Turán's theorem might be helpful here.

Similarly, the techniques of O'Donnell and Wimmer [OW07] who studied approximation of the majority function by DNF formulas (under the uniform distribution) might be helpful for studying $T(n, n/2)$ and the depth-3 circuit size of majority. We believe that the explicit link established between depth-3 formula size and the complexity of approximation by depth-2 is useful to advance both of these lines of research further. Can one find another application of the duality?

References

[AG87] Miklós Ajtai and Yuri Gurevich. Monotone versus positive. *J. ACM*, 34(4):1004–1015, 1987.

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

[BHST14] Eric Blais, Johan Håstad, Rocco A. Servedio, and Li-Yang Tan. On DNF approximators for monotone boolean functions. In *Proceedings of the 41st International Colloquium on Automata, Languages, and Programming (ICALP), Part I*, pages 235–246, 2014.

[BIS12] Paul Beame, Russell Impagliazzo, and Srikanth Srinivasan. Approximating AC^0 by small height decision trees and a deterministic algorithm for $\#\mathrm{AC}^0\mathrm{SAT}$. In *Proceedings of the 27th Conference on Computational Complexity (CCC)*, pages 117–125, 2012.

[BT15] Eric Blais and Li-Yang Tan. Approximating boolean functions with depth-2 circuits. *SIAM J. Comput.*, 44(6):1583–1600, 2015.

[Chv79] Vasek Chvatal. A greedy heuristic for the set-covering problem. *Math. Oper. Res.*, 4(3):233–235, 1979.

[CIP06] Chris Calabro, Russell Impagliazzo, and Ramamohan Paturi. A duality between clause width and clause density for SAT. In *Proceedings of the 21st Annual IEEE Conference on Computational Complexity (CCC)*, pages 252–260, 2006.

[COS15] Xi Chen, Igor Carboni Oliveira, and Rocco A. Servedio. Addition is exponentially harder than counting for shallow monotone circuits. *CoRR*, abs/1508.03061, 2015.

[CR11] Jonathan Cutler and A. J. Radcliffe. Extremal problems for independent set enumeration. *Electr. J. Comb.*, 18(1), 2011.

[DH09] Evgeny Dantsin and Edward A. Hirsch. Worst-case upper bounds. In Armin Biere, Marijn Heule, Hans van Maaren, and Toby Walsh, editors, *Handbook of Satisfiability*, chapter 12, pages 359–380. IOS Press, 2009.

[Erd62] Paul Erdős. On the number of complete subgraphs contained in certain graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl*, 7:459–464, 1962.

[Fil10] Yuval Filmus. Smolensky's lower bound. 2010.

[Fre95] Yoav Freund. Boosting a weak learning algorithm by majority. *Inf. Comput.*, 121(2):256–285, 1995.

[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.

[FSUV13] Bill Fefferman, Ronen Shaltiel, Christopher Umans, and Emanuele Viola. On beating the hybrid argument. *Theory of Computing*, 9:809–843, 2013.

[GHR92] Mikael Goldmann, Johan Håstad, and Alexander A. Razborov. Majority gates vs. general weighted threshold gates. *Computational Complexity*, 2:277–300, 1992.

[GT16] Oded Goldreich and Avishay Tal. Matrix rigidity of random toeplitz matrices. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC)*, pages 91–104, 2016.

[GW13] Oded Goldreich and Avi Wigderson. On the size of depth-three boolean circuits for computing multilinear functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:43, 2013.

[Had76] N. G. Hadžiivanov. A generalization of Turáns theorem on graphs. *CR Acad. Bulgare Sci*, 29(11):1567–1570, 1976.

[Ham50] R. W. Hamming. Error detecting and error correcting codes. *Bell System Technical Journal*, 29(2):147–160, 1950.

[Hås89] Johan Håstad. Almost optimal lower bounds for small depth circuits. In S. Micali, editor, *Randomness and Computation*, volume 5 of *Advances in computing research*, pages 143–170. JAI Press, 1989.

[Hås14] Johan Håstad. On the correlation of parity and small-depth circuits. *SIAM J. Comput.*, 43(5):1699–1708, 2014.

[HJP95] Johan Håstad, Stasys Jukna, and Pavel Pudlák. Top-down lower bounds for depth-three circuits. *Computational Complexity*, 5(2):99–112, 1995.

[HMP+93] András Hajnal, Wolfgang Maass, Pavel Pudlák, Mario Szegedy, and György Turán. Threshold circuits of bounded depth. *J. Comput. Syst. Sci.*, 46(2):129–154, 1993.

[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 (SODA)*, pages 961–972, 2012.

[IPZ01] Russell Impagliazzo, Ramamohan Paturi, and Francis Zane. Which problems have strongly exponential complexity? *J. Comput. Syst. Sci.*, 63(4):512–530, 2001.

[KPPY84] Maria M. Klawe, Wolfgang J. Paul, Nicholas Pippenger, and Mihalis Yannakakis. On monotone formulae with restricted depth. In *Proceedings of the 16th Annual ACM Symposium on Theory of Computing (STOC)*, pages 480–487, 1984.

[KW90] Mauricio Karchmer and Avi Wigderson. Monotone circuits for connectivity require super-logarithmic depth. *SIAM J. Discrete Math.*, 3(2):255–265, 1990.

[Lov75] László Lovász. On the ratio of optimal integral and fractional covers. *Discrete mathematics*, 13(4):383–390, 1975.

[Lup65] Oleg Lupanov. On the realization of functions of logical algebra by formulae of finite classes (formulae of limited depth) in the basis $\wedge, \vee, \neg$. *Probl. Cybernetics*, 6:1–14, 1965.

[OW07] Ryan O'Donnell and Karl Wimmer. Approximation by DNF: Examples and counterexamples. In Lars Arge, Christian Cachin, Tomasz Jurdzinski, and Andrzej Tarlecki, editors, *ICALP*, volume 4596 of *Lecture Notes in Computer Science*, pages 195–206. Springer, 2007.

[PPSZ05] Ramamohan Paturi, Pavel Pudlák, Michael E. Saks, and Francis Zane. An improved exponential-time algorithm for k-SAT. *J. ACM*, 52(3):337–364, 2005.

[PPZ99] Ramamohan Paturi, Pavel Pudlák, and Francis Zane. Satisfiability coding lemma. *Chicago J. Theor. Comput. Sci.*, 1999, 1999.

[PSZ00] Ramamohan Paturi, Michael E. Saks, and Francis Zane. Exponential lower bounds for depth three boolean circuits. *Computational Complexity*, 9(1):1–15, 2000.

[Qui54] W. V. Quine. Two theorems about truth-functions. *Journal of Symbolic Logic*, 19(2):142–143, 1954.

[Raz87] Alexander Razborov. Lower bounds on the size of bounded depth circuits over a complete basis with logical addition. *Mathematical notes of the Academy of Sciences of the USSR*, 41(4):333–338, 1987.

[Rom76] Steven Roman. The maximum number of q-cliques in a graph with no p-clique. *Discrete Mathematics*, 14(4):365–371, 1976.

[Ros14] Benjamin Rossman. Formulas vs. circuits for small distance connectivity. In *Proceedings of the 46th Symposium on Theory of Computing (STOC)*, pages 203–212, 2014.

[Ros15] Benjamin Rossman. The average sensitivity of bounded-depth formulas. In *Proceedings of the IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 424–430, 2015.

[RR97] Alexander A. Razborov and Steven Rudich. Natural proofs. *J. Comput. Syst. Sci.*, 55(1):24–35, 1997.

[RS42] John Riordan and C E Shannon. The number of two-terminal series-parallel networks. *Journal of Mathematics and Physics*, 21(1-4):83–93, 1942.

[Sau71] Norbert Sauer. A generalization of a theorem of Turán. *Journal of Combinatorial Theory, Series B*, 10(2):109–112, 1971.

[Sch05] Rainer Schuler. An algorithm for the satisfiability problem of formulas in conjunctive normal form. *J. Algorithms*, 54(1):40–44, 2005.

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In Alfred V. Aho, editor, *STOC*, pages 77–82. ACM, 1987.

[Smo93] Roman Smolensky. On representations by low-degree polynomials. In *34th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 130–138, 1993.

[Spi71] Philip M Spira. On time-hardware complexity tradeoffs for boolean functions. In *Proceedings of the 4th Hawaii Symposium on System Sciences*, pages 525–527, 1971.

[Tur41] Paul Turán. On an extremal problem in graph theory. *Mat. Fiz. Lapok*, 48(436-452):137, 1941.

[Val77] Leslie G. Valiant. Graph-theoretic arguments in low-level complexity. In *Proceedings of the 6th Symposium on Mathematical Foundations of Computer Science (MFCS)*, pages 162–176, 1977.

[Vaz01] Vijay V. Vazirani. *Approximation algorithms*. Springer, 2001.

[Wil13] Ryan Williams. Improving exhaustive search implies superpolynomial lower bounds. *SIAM J. Comput.*, 42(3):1218–1244, 2013.

[Yao77] Andrew Chi-Chih Yao. Probabilistic computations: Toward a unified measure of complexity. In *Proceedings of the 18th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 222–227, 1977.

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

[Zyk49] Alexander Aleksandrovich Zykov. On some properties of linear complexes. *Matematicheskii sbornik*, 66(2):163–188, 1949.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il