Size, Depth and Energy of Threshold Circuits Computing Parity Function

Kei Uchizawa
Graduate School of Science and Engineering, Yamagata University, Japan
uchizawa@yz.yamagata-u.ac.jp

Abstract

We investigate relations among the size, depth and energy of threshold circuits computing the n-variable parity function PAR_n, where the energy is a complexity measure for sparsity on computation of threshold circuits, and is defined to be the maximum number of gates outputting "1" over all the input assignments. We show that PAR_n is hard for threshold circuits of small size, depth and energy:

- If a depth-2 threshold circuit C of size s and energy e computes PAR_n, it holds that $2^{n/(e \log^e n)} \leq s$; and
- if a threshold circuit C of size s, depth d and energy e computes PAR_n, it holds that $2^{n/(e 2^{e+d} \log^e n)} \leq s$.

We then provide several upper bounds:

- PAR_n is computable by a depth-2 threshold circuit of size $O(2^{n-2e})$ and energy e;
- PAR_n is computable by a depth-3 threshold circuit of size $O(2^{n/(e-1)} + 2^{e-2})$ and energy e; and
- PAR_n is computable by a threshold circuit of size $O((e+d)2^{n-m})$, depth $d + O(1)$ and energy $e + O(1)$, where $m = \max(((e-1)/(d-1))^{d-1}, ((d-1)/(e-1))^{e-1})$.

Our lower and upper bounds imply that threshold circuits need exponential size if both depth and energy are constant, which contrasts with the fact that PAR_n is computable by a threshold circuit of size $O(n)$ and depth 2 if there is no restriction on the energy. Our results also suggest that any threshold circuit computing the parity function needs depth to be sparse if its size is bounded.

2012 ACM Subject Classification Theory of computation → Models of computation

Keywords and phrases Circuit complexity, neural networks, threshold circuits, sprase activity, tradeoffs

Digital Object Identifier 10.4230/LIPIcs.ISAAC.2020.54

Funding *Kei Uchizawa*: This work was supported by JSPS KAKENHI Grant Number JP19K11817.

Acknowledgements We thank the reviewers for their constructive criticisms and suggestions.

1 Introduction

Logic circuit is a traditional computational model consisting of a number of basic computational elements, called gates. In the standard type of logic circuits, a gate computes AND, OR or NOT function. A threshold circuit is a logic circuit consisting of gates computing linear threshold functions: a gate g with n input variables has weights $w_1, w_2, \ldots, w_n$ with a threshold t, and computes $g(\mathbf{x}) = \mathrm{sig}(\sum_{i=1}^{n} w_i x_i - t)$ for $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \{0,1\}^n$.

Threshold circuits are well-studied in the literature, and are known to have large expressive power: a single threshold gate is able to compute the majority function which the standard logic circuit of constant depth needs an exponential size to compute [9, 30], and threshold circuits of polynomial size (i.e., polynomial number of gates) and constant depth can compute basic arithmetic functions such as addiction, multiplication and division [22]. In fact, proving a limit of polynomial-size and constant-depth threshold circuits is one of the cutting-edge open problems in circuit complexity (See, for example, the paper [3]). Although several super-polynomial lower bounds are known for restricted settings [1, 7, 8, 10, 21], we can not even rule out the possibility that every decision problem in $\mathrm{EXP}^{\mathrm{NP}}$ is solvable by a threshold circuit of polynomial size and just depth 2.

Besides the considerable interest in circuit complexity, biology also motivates research of threshold circuits, since a threshold gate is traditionally considered to be a simple theoretical model of a biological neuron in the brain [17, 20]. The goal of this line of the research is to understand how the brain smoothly carry out tough computational tasks. We focus on the following interesting property that a biological neuron possesses: a neuron consumes more energy to emit a spike than not to emit a spike, and consequently their computation results from a relatively small number of simultaneously active neurons out of a large population in the nervous system [6, 16, 19], which implies that neural networks acquire the principles of computation with sparse activity. This may not be a burden on the brain, since machine learning techniques inspired by this property, such as sparse coding or sparse autoencoder, pushes recent progress of deep learning method [11, 15, 18].

The above observations address a natural question: What computational tasks are carried out by threshold circuits with sparse activity? In this paper, we employ a complexity measure for sparsity of computation, called energy complexity, and investigate the computational power of threshold circuits of small energy complexity. For a threshold circuit C, the energy of C is defined to be the maximum number of gates outputting "1" (i.e., emitting a spike) in C, where the maximum is taken over all the input assignments to C [27] (Energy complexity of the standard logic circuits is also studied in the literature. See [5, 23] and their references).

In the previous research, it turns out that the energy complexity relates to other major complexity measures such as size and depth. In particular, it is known that, if a n-variable Boolean function f has very high communication complexity (more formally, if f has bounded-error communication complexity $\Omega(n + \log \delta)$ to compute correctly with probability $1/2 + \delta$ for any δ, $0 < \delta < 1/2$), it holds that $s = 2^{\Omega(n/e^d)}$ for size s, depth d and energy e of any threshold circuit computing f [28]. Thus, any threshold circuit of constant depth and energy $n^{o(1)}$ requires an exponential size to compute f.

However, we believe that such a relation among the size, depth and energy is more universal, and exists independently of communication complexity. In this paper, as a first step towards finding such a relation, we focus on the n-variable parity function PAR_n which is a representative of low communication complexity Boolean functions.

Threshold circuits computing PAR_n have been studied in terms of either bounded depth or bounded energy (See Table 1). For bounded-depth circuits, it is shown that any threshold circuit of depth d has size $(n/2)^{1/2(d-1)}$ to compute PAR_n [12]. An upper bound $O(dn^{1/(d-1)})$ shows that these lower bounds are almost tight [22]. Moreover, any depth-2 threshold circuit needs size $\Omega(\sqrt{n})$ [14] and constant-depth circuits needs sublinear-size even to approximate the function [4]. For bounded-energy circuits, it is known that, in the case of energy $e = 1$, any threshold circuit needs size 2^{n-1} to compute PAR_n no matter how large depth of the circuit is [25], and in the case where $e \geq 2$ there is a lower bound $\Omega(n^{1/e})$ on the size of threshold circuit of energy e [29]. Upper bounds 2^{n-1} for $e = 1$ [26] and $O(en^{1/(e-1)})$ for $e \geq 2$ given in [25] show that these lower bounds are also almost tight.

However, the lower bounds given in [4, 12, 14] for bounded-depth circuits are based on random restriction, and symmetrically handle the outputs 0 and 1 of a gate, and seems independent of energy complexity. Also, the proofs given in [24, 29] for energy-bounded circuits are independent of depth. Thus we cannot directly apply the techniques to such a situation that both depth and energy are bounded. In fact, these results do not rule out the possibility that PAR_n is computable by a threshold circuit of size $O(\sqrt{n})$, depth 2 and energy 2, while, by looking into the construction of circuits given in [22, 24], we find the circuits need large energy to have small depth, and large depth to have small energy.

Table 1 Lower and upper bounds on the size of threshold circuits C computing the n-variable parity function, where d and e are depth and energy of C, respectively. In the bottom upper bound, $m = \max(((e-1)/(d-1))^{d-1}, ((d-1)/(e-1))^{e-1})$.

		Lower Bounds		Upper Bounds	
Bounded Depth	$d = 1$	not computable			[20, 22]
	$2 \le d$	$(n/2)^{1/2(d-1)}$	[12]	$O(dn^{1/(d-1)})$	[22]
Bounded Energy	$e = 1$	2^{n-1}	[24]	$2^{n-1} + 1$	[26]
	$2 \le e$	$\Omega(n^{1/e})$	[29]	$O(en^{1/(e-1)})$	[25]
Bounded Depth and Bounded Energy	$d = 2$	$2^{n/(e \log^e n)}$	[Ours]	$O(e2^{n-2e})$	[Ours]
	$d = 3$	$2^{n/(e2^{e+d} \log^e n)}$	[Ours]	$O(2^{e-2} + 2^{n/(e-1)})$	[Ours]
	$4 \le d$			$O((e+d)2^{n-m})$	[Ours]

In this paper, we show that there actually exists a relation among all the size, depth and energy of threshold circuits computing the parity function. More formally, we show that, for any depth-2 threshold circuit of size s and energy e, it holds that $2^{n/(e \log^e n)} \le s$. For $3 \le d$, we prove that, for any threshold circuit of size s, depth d and energy e, it holds that $2^{n/(e2^{e+d} \log^e n)} \le s$. These imply that a threshold circuit of constant depth and energy e requires an exponential size to compute PAR_n. Our proofs are based on a combination of depth and energy reduction. We first show that, if a threshold gate g outputs one for many input assignments, we can deterministically fix part of input variables so that the output of g become constant one. This statement plays key role in showing that there exists a partial input assignment which reduce either depth by one or energy by one. Since any threshold gate cannot compute the parity function of two variables, we can deduce that threshold circuits of small depth and energy requires large size.

In addition, we provide several upper bounds on the size of threshold circuits computing the parity function. We prove that the n-variable parity function is computable by a circuit of size $O(e2^{n-2e})$, depth 2 and energy e; a circuit of size $O(2^{n/(e-1)} + 2^{e-2})$, depth 3 and energy e; and a circuit of size $O((e+d)2^{n-m})$, depth $d + O(1)$, and energy $e + O(1)$, where $m = \max(((e-1)/(d-1))^{d-1}, ((d-1)/(e-1))^{e-1})$.

Although there is much room between our lower and upper bounds, these results imply that the size, depth and energy of threshold circuits computing the parity function are definitely involved. In particular, PAR_n is hard for threshold circuits of small size, depth and energy. Our results also suggest that any threshold circuit needs depth to be sparse even for computing the parity function if its size is bounded, which may shed light on computation of neural networks with sparse activity.

The rest of the paper is organized as follows. In Section 2, we define threshold circuits, and give several propositions used in the following sections. In Section 3, we prove lower bounds for depth-2 and depth-d circuits. In Section 4, we construct depth-2, depth-3 and depth-d circuits of small energy. In Section 5, we conclude with some remarks.

2 Preliminaries

In this section, we define some terms, and introduce several propositions. Throughout the paper, we denote by $[n] = \{1, 2, \ldots, n\}$ for any positive integer n. For a set I of Boolean input variables, we may denote by $\{0,1\}^I$ a set of the $2^{|I|}$ input assignments for I. For a set S, we denote by $|S|$ its cardinality. For a Boolean vector $\mathbf{a}$, we denote by $|\mathbf{a}|$ the hamming weight of $\mathbf{a}$ (i.e., the number of ones in $\mathbf{a}$).

In Section 2.1, we define threshold circuits, and show propositions on bounded-depth or bounded-energy circuits. In Section 2.2, we give Sauer's lemma that we will use to prove lower bounds.

2.1 Threshold Circuits

A *threshold gate* g is a logic gate computing a linear threshold function of an arbitrary integer n of inputs, which is identified by weights $w_1, w_2, \ldots, w_n$ for the n input variables and a threshold t. We define the output $g(\mathbf{x})$ of g as

$$g(\mathbf{x}) \quad = \quad \mathrm{sig}\left(\sum_{i=1}^{n} w_i x_i - t\right) = \begin{cases} 1 & \text{if } \sum_{i=1}^{z} w_i x_i \geq t; \\ 0 & \text{otherwise.} \end{cases}$$

A *threshold circuit* C is a feedforward circuit consisting of threshold gates, and is expressed by a directed acyclic graph. Let n be the number of inputs to C, then C has n input nodes of in-degree 0, each of which corresponds to one of the n input variables $x_1, x_2, \ldots x_n$, while the other nodes correspond to threshold gates. The inputs to a gate g in C consists of the inputs $x_1, x_2, \ldots x_n$ and the outputs of the gates directed to g.

We define the *size* s of a circuit C as the number of gates in C. Let $g_1, g_2, \ldots, g_s$ be the gates in C where the gates are numbered in the topological order on the underlying directed acyclic graph of C. We call g_s the *top gate* of C, and regard the output $g_s(\mathbf{x})$ of g_s as the *output* $C(\mathbf{x})$ *of* C, that is, $C(\mathbf{x}) = g_s(\mathbf{x})$ for every input $\mathbf{x} \in \{0,1\}^n$. A threshold circuit C *computes* a Boolean function $f : \{0,1\}^n \to \{0,1\}$ if $C(\mathbf{x}) = f(\mathbf{x})$ for every $\mathbf{x} \in \{0,1\}^n$. We say that a gate g_i, $1 \leq i \leq s$, is *in the l-th layer* of a circuit C if there are l gates (including g_i) on the longest path from an input node to g_i in the underlying graph. The *depth* d of C is the number of gates on the longest path to the top gate g_s. We define the *energy* e of C as

$$e = \max_{\mathbf{x} \in \{0,1\}^n} \sum_{i=1}^{s} g_i(\mathbf{x}).$$

We may assume throughout the paper that $1 \leq s, d, e$.

For every $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \{0,1\}^n$, the n-variable parity function PAR_n is defined to be

$$\mathrm{PAR}_n(\mathbf{x}) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} x_i \text{ is odd;} \\ 0 & \text{if } \sum_{i=1}^{n} x_i \text{ is even.} \end{cases}$$

For a Boolean function f of n variables, we define $S_0(f) = \{\mathbf{x} \in \{0,1\}^n \mid f(\mathbf{x}) = 0\}$ and $S_1(f) = \{\mathbf{x} \in \{0,1\}^n \mid f(\mathbf{x}) = 1\}$. Clearly, $|S_0(\mathrm{PAR}_n)| = |S_1(\mathrm{PAR}_n)| = 2^{n-1}$.

The following exponential lower bound is known for threshold circuits of arbitrary depth and energy one.

▶ **Theorem 1.** *If a threshold circuit of size s and energy one computes PAR_n (or its complement), then it holds that $2^{n-1} \leq s$.*

The lower bound in Theorem 1 is tight even for depth-2 circuits, as follows.

▶ **Theorem 2.** *PAR_n is computable by a threshold circuit of size $2^{n-1} + 1$, depth 2 and energy one.*

We also use the following two constructions for PAR_n.

▶ **Theorem 3.** *For every positive integer $d < \log n$, PAR_n is computable by a threshold circuit C of size at most $(d-1)\lceil n^{1/(d-1)}\rceil + 1$ and depth d. (In addition, the energy of C is same as its size.)*

▶ **Theorem 4.** *For every positive integer $e < \log n$, PAR_n is computable by a threshold circuit of size $(e-1)\lceil n^{1/(e-1)}\rceil + 1$ and energy e. (In addition, the depth of C is same as its size.)*

In these statements, we removed asymptotic notations from the original ones to refine the bounds.

2.2 Sauer's Lemma

For $S \subseteq \{0,1\}^n$ and $I \subseteq [n]$, we define a projected set S_I of S on I as

$$S_I = \{\mathbf{a} \cap I \mid \mathbf{a} \in S\}$$

where we consider $\mathbf{a}$ as the characteristic vector of a subset of $[n]$, and denote by $\mathbf{a} \cap I$ the intersection of the subset and I. If $|S_I| = 2^{|I|}$, we say that S *shatters* I. The VC dimension of S is the largest cardinality of I that C shatters. The following lemma is known as Sauer's lemma, and state that $|S|$ is bounded by its VC dimension (See, for example, Theorem 3.6 in [2] and its proof).

▶ **Lemma 5.** *Let $S \subseteq \{0,1\}^n$ and d be the VC dimension of S. Then*

$$|S| \le \sum_{i=0}^{d} \binom{n}{i} \le n^d.$$

The lemma immediately implies a lower bound on the largest cardinality of such I.

▶ **Corollary 6.** *Let $S \subseteq \{0,1\}^n$. Then, there exists I such that S shatters I, and*

$$\frac{\log |S|}{\log n} \le |I|.$$

3 Lower Bounds

In this section, we provide lower bounds for threshold circuits computing the parity function. In Section 3.1, we introduce a lemma which is simple, but useful for obtaining lower bounds of small-energy threshold circuits. We then give a lower bound for depth-2 circuits. In Section 3.2, using the bound for depth-2 circuits, we obtain a lower bound for depth-d circuits for $d \ge 3$.

3.1 Depth-2 Circuits

The following lemma implies that if a threshold gate outputs one for a large number of input assignments, we can deterministically fix a proper subset of variables so that g outputs one for all the remaining input assignments. Similarly to $S_1(f)$ for a Boolean function f, we define $S_1(g) = \{\mathbf{x} \in \{0,1\}^n \mid g(\mathbf{x}) = 1\}$ for a threshold gate g of n input variables.

▶ **Lemma 7.** *Let g be a threshold gate with n input variables. Then there exists $I \subseteq [n]$ and an input $\mathbf{b} \in \{0,1\}^{[n]\setminus I}$ such that, by fixing the variables for $[n]\setminus I$ to $\mathbf{b}$, g outputs one for every $\mathbf{a} \in \{0,1\}^I$ and*

$$\frac{\log |S_1(g)|}{\log n} \le |I|.$$

Proof. Corollary 6 implies that there exists a set $I \subseteq [n]$ such that $S_1(g)$ shatters I and

$$\frac{\log |S_1(g)|}{\log n} \leq |I|. \tag{1}$$

Thus, it suffices to show that there exists an input assignment in $\{0,1\}^{[n]\setminus I}$ that force the output of g to be constant one.

Let $w_1, w_2, \dots, w_n$ be the weights and t be the threshold of g. Since $S_1(g)$ shatters I, for each $\mathbf{a} \in \{0,1\}^I$, there exists $\mathbf{b} \in \{0,1\}^{[n]\setminus I}$ such that $(\mathbf{a}, \mathbf{b}) \in S_1(g)$. Let

$$\mathbf{a}^* = \arg \min_{\mathbf{a} \in \{0,1\}^I} \sum_{i \in I} w_i a_i,$$

and $\mathbf{b}^*$ be its counterpart: $(\mathbf{a}^*, \mathbf{b}^*) \in S_1(g)$. Clearly,

$$\sum_{i \in I} w_i a_i^* \leq \sum_{i \in I} w_i a_i$$

for every $\mathbf{a} \in \{0,1\}^I$, and hence we have not only $(\mathbf{a}^*, \mathbf{b}^*) \in S_1(g)$ but also

$$(\mathbf{a}, \mathbf{b}^*) \in S_1(g)$$

for every $\mathbf{a} \in \{0,1\}^I$. Thus, by fixing $[n]\setminus I$ to $\mathbf{b}^*$, g outputs one for every $\mathbf{a} \in \{0,1\}^I$, as desired. ◀

Using Lemma 7, we provide a lower bound on the size of depth-2 threshold circuits computing PAR_n and its complement. In the proof, we show that any circuit C computing the parity function contains a gate outputting ones for many input assignments, which implies that we can reduce energy of C by fixing some of input variables. Thus, Theorem 1 guarantees that C has exponential size if e is small.

▶ **Theorem 8.** *If a depth-2 threshold circuit of size s and energy e computes* PAR_n *(or its complement), then it holds that*

$$\frac{n}{e \log^e n} \leq \log s.$$

Proof. We prove the theorem for PAR_n, since the other case is similar. Our proof is by induction on the energy. Theorem 1 implies the base case where circuits have energy one. Suppose $e \geq 2$. We assume for the induction hypothesis that every depth-2 threshold circuit of size s and energy $e-1$ satisfies

$$\frac{n}{(e-1)\log^{e-1} n} \leq \log s.$$

Let C be a depth-2 circuit of size s and energy e, and G be a set of the gates in the first layer of C. We prove that there exists a gate in the first layer that outputs one for many input assignments:

▷ Claim 9. There exists $g \in G$ such that $2^n/(4n^5 s) \leq S_1(g)$.

Proof. Consider an error-correcting code $S \subseteq \{0,1\}^n$ of minimal distance five: each pair of vectors in S has mutual hamming distance at least or equal to five. It is known that there exists such S satisfying $2^n/n^5 \leq |S|$ (See Theorem 17.2 in [13]). Then we make a new set T from S by applying the following procedure to every vector $\mathbf{a} \in S$: If $|\mathbf{a}|$ is odd, flip an element in $\mathbf{a}$. Note that if two vectors have distance five, exactly one of them has an odd number of its weight. Thus, T is an error-correcting code of minimal distance four. Furthermore, we have $T \subseteq S_0(\mathrm{PAR}_n)$, and

$$2^n/n^5 \leq |S| = |T|. \tag{2}$$

Let

$$U = \{\mathbf{a} \in T \mid \text{There exists } g \in G \text{ such that } g(\mathbf{a}) = 1\ \}.$$

If $|T|/2 \leq |U|$, the pigeonhole principle implies the claim, since $|G| < s$. Thus, it suffices to verify the other case.

Suppose $|T|/2 > |U|$. Let $U' = T \backslash U$, then no gate in G outputs one for every $\mathbf{a} \in U'$, and Eq. (2) implies that $2^n/(2n^5) \leq |U'|$. Consider an arbitrary pair $(\mathbf{a}, \mathbf{b}) \in U' \times U'$ such that $\mathbf{a} \neq \mathbf{b}$. Let $i^* \in [n]$ be an index such that $a_{i^*} \neq b_{i^*}$. From $\mathbf{a}$ (resp., $\mathbf{b}$), we obtain $\mathbf{a}'$ (resp., $\mathbf{b}$) such that all the elements in $\mathbf{a}'$ (respectively, $\mathbf{b}'$) are same as $\mathbf{a}$ (resp., $\mathbf{b}$) except that the i^*-th element is flipped. Without loss of generality, we assume that $a_{i^*} = 1$ and $b_{i^*} = 0$. Note that $\mathbf{a}', \mathbf{b}' \in S_1(\mathrm{PAR}_n)$. We below show that a gate $g \in G$ outputs one for at least one of $\mathbf{a}'$ and $\mathbf{b}'$.

Suppose for the sake of contradiction that no gate in G outputs one for both of $\mathbf{a}'$ and $\mathbf{b}'$. Let $w_1, w_2, \ldots, w_n$ and t be the weights for the input variables and the threshold of the top gate of C. By the assumption, no gate in G outputs one for $\mathbf{a}, \mathbf{b}, \mathbf{a}'$ and $\mathbf{b}'$. Thus, on the one hand, we have $\mathbf{a}, \mathbf{b} \in T \subseteq S_0(\mathrm{PAR}_n)$, and hence

$$\sum_{i=1}^{n} w_i a_i < t \quad \text{and} \quad \sum_{i=1}^{n} w_i b_i < t. \tag{3}$$

On the other hand, since we flipped the elements, we have $\mathbf{a}', \mathbf{b}' \in S_1(\mathrm{PAR}_n)$, and hence

$$\sum_{i=1}^{n} w_i a'_i = \sum_{i=1}^{n} w_i a_i - w_{i^*} \geq t \quad \text{and} \quad \sum_{i=1}^{n} w_i b'_i = \sum_{i=1}^{n} w_i b_i + w_{i^*} \geq t. \tag{4}$$

The sum of the two inequalities in (3) contradicts the counterpart in (4).

We then make $|U'|/2$ disjoint pairs of $(\mathbf{a}, \mathbf{b})$. Note that different $(\mathbf{a}, \mathbf{b})$s yield different $(\mathbf{a}', \mathbf{b}')$s, since T is an error -correcting code of minimal distance four. Thus, there are at least

$$\frac{|U'|}{2} \geq 2^n/(4n^5)$$

input assignments for which a gate in G outputs one. Thus, the pigeonhole principle implies the claim. ◁

Lemma 7 and the claim imply that, by fixing an appropriate set of input variables, we can obtain a threshold circuit C' of size $s-1$ and $e-1$ that computes $\mathrm{PAR}_{n'}$ (or its complement) where

$$n' \geq \frac{n - (\log s + 5\log n + 2)}{\log n}.$$

Thus, by the induction hypothesis, we have

$$\frac{n - (\log s + 5\log n + 2)}{\log n} \cdot \frac{1}{(e-1)\log^{e-1} n} \leq \frac{n'}{(e-1)\log^{e-1} n'} \leq \log(s-1) < \log s.$$

Since it holds that $\log s + 5\log n + 2 \leq \log^e n \cdot \log s$ for sufficiently large n, we have

$$\frac{n}{e\log^e n} \leq \log s,$$

as desired. ◀

3.2 Depth-d Circuits

Based on the idea used for proving Theorem 1, we provide a lower bound for threshold circuits of depth d and energy e. Our proof is by induction on the depth and energy: we show that we can decrease either depth or energy of a given circuit by fixing part of input variables.

▶ **Theorem 10.** *If a threshold circuit of size s, depth d and energy e computes* PAR_n *(or its complement), then it holds that*

$$\frac{n}{e2^{e+d}\log^e n} \leq \log s.$$

Proof. We prove the theorem for PAR_n, since the other case is similar. Our proof is by induction on the energy and depth. Theorem 1 implies the base case where circuits have energy one, and Theorem 8 implies the base case where circuit have depth 2. We here assume that for every threshold circuit of size s, depth $d-1$ and energy e, it holds that

$$\frac{n}{e2^{e+d-1}\log^e n} \leq \log s, \tag{5}$$

and for every threshold circuit of size s, depth d and energy $e-1$, it holds that

$$\frac{n}{(e-1)2^{e+d-1}\log^{e-1} n} \leq \log s. \tag{6}$$

Let C be a circuit of size s, depth d, energy e. We denote by G a set of the gates in the first layer of C. We show that we can reduce either depth or energy by fixing some of the input variables.

Consider the set $\{0,1\}^n$ of input assignments as $A \times B$ such that $A = B = \{0,1\}^{n/2}$. Note that any input assignment can be considered as an element $(\mathbf{a}, \mathbf{b}) \in A \times B$. For every $\mathbf{a} \in A$, we define $B_\mathbf{a}$ as

$$B_\mathbf{a} = \{(\mathbf{a}, \mathbf{b}) \mid \mathbf{b} \in B\}.$$

We say that $B_\mathbf{a}$ is *clean* if no gate in G outputs one for every input assignment $(\mathbf{a}, \mathbf{b}) \in B_\mathbf{a}$. Consider the following two cases:

(i) *There exists* $\mathbf{a}^*$ *such that* $B_{\mathbf{a}^*}$ *is clean.*

In this case, by fixing A to $\mathbf{a}^*$, we can safely remove all the gates in G, and obtain a circuit C' that computes $\mathrm{PAR}_{n/2}$. Thus, by Eq. (5), we have

$$\frac{n}{e2^{e+d}\log^e n} = \frac{n/2}{e2^{e+d-1}\log^e n} \leq \log s$$

as desired.

(ii) *For every* $\mathbf{a} \in A$, $B_\mathbf{a}$ *is not clean.*

In this case, for every $\mathbf{a} \in A$, there exists $\mathbf{b} \in B$ such that some gate $g \in G_1$ outputs one. Thus, since $|G| < s$, the pigeonhole principle implies that there exists $g \in G$ such that

$$\frac{2^{n/2}}{s} \leq S_1(g).$$

Similarly to the case of depth-2 circuits, Lemma 7 implies that by fixing an appropriate set of input variables, we can obtain a threshold circuit C' of size $s-1$ and $e-1$ that computes $\mathrm{PAR}_{n'}$ where

$$n' \geq \frac{n/2 - \log s}{\log n}.$$

Thus, by the induction hypothesis, it holds that

$$\frac{n/2 - \log s}{\log n} \cdot \frac{1}{(e-1)2^{e+d-1}\log^{e-1} n} \le \frac{n'}{(e-1)2^{e+d-1}\log^{e-1} n} \le \log(s-1) < \log s.$$

Since we have $2\log s \le 2^{e+d}\log^e n \log s$, it holds that

$$\frac{n}{e2^{e+d}\log^e n} \le \log s,$$

as desired. ◀

4 Upper Bounds

In this section, we show upper bounds on the size of threshold circuits computing the parity function. In Section 4.1, we give depth-2 circuits. In Section 4.2, we show a better upper bound for depth-3 circuits. In Section 4.3, using the same idea given in Section 4.1, we provide circuits of depth d for $d \ge 4$.

4.1 Depth-2 Circuits

Before we give a construction of depth-2 circuits, we prove the following lemma showing that we can suppress any threshold gate for a subset of input assignments.

▶ **Lemma 11.** *Let l and r be positive integers, and g be a threshold gate with $l+r$ input variables. Then, for any $\mathbf{a} \in \{0,1\}^l$, there exists a threshold gate $g^{\mathbf{a}}$ such that for every $\mathbf{x} \in \{0,1\}^l$ and $\mathbf{y} \in \{0,1\}^r$*

$$g^{\mathbf{a}}(\mathbf{x},\mathbf{y}) = \begin{cases} 0 & \text{if } \mathbf{x} \ne \mathbf{a}; \\ g(\mathbf{a},\mathbf{y}) & \text{if } \mathbf{x} = \mathbf{a}. \end{cases} \tag{7}$$

We call $g^{\mathbf{a}}$ given in Lemma 11 *a suppressed gate of g for* $\mathbf{a}$.

Proof. Consider $\mathbf{a} = (a_1, a_2, \ldots, a_l) \in \{0,1\}^l$. Let $p_1, p_2, \ldots, p_l$ be weights for the l input variables, $q_1, q_2, \ldots, q_r$ be weights for the r input variables, and t be threshold of g: For every $\mathbf{x} = (x_1, x_2, \ldots, x_l)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_r)$,

$$g(\mathbf{x},\mathbf{y}) = \mathrm{sig}\left(p(\mathbf{x},\mathbf{y})\right).$$

where we denote by

$$p(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{l} p_i x_i + \sum_{i=1}^{r} q_i y_i - t.$$

Let W be a positive integer satisfying

$$1 + \max_{\mathbf{x},\mathbf{y}} p(\mathbf{x},\mathbf{y}) \le W. \tag{8}$$

We can obtain the desired gate $g^{\mathbf{a}}$ by modifying the weights for $x_1, x_2, \ldots, x_n$ as follows: For each $1 \le i \le l$, we define new weight p'_i for x_i as

$$p'_i = \begin{cases} p_i + W & \text{if } a_i = 1; \\ -W & \text{otherwise.} \end{cases}$$

and new threshold t' as

$$t' = t + W \sum_{i=1}^{l} a_i.$$

Consequently, we have $g^{\mathbf{a}}(\mathbf{x},\mathbf{y}) = \mathrm{sig}\,(p'(\mathbf{x},\mathbf{y}))$, where

$$\begin{aligned} p'(\mathbf{x},\mathbf{y}) &= \sum_{i:a_i=1} (p_i + W)x_i - \sum_{i:a_i=0} W x_i + \sum_{i=1}^{r} q_i y_i - \left(t + W \sum_{i=1}^{l} a_i\right) \\ &= \sum_{i:a_i=1} W(x_i - 1) - \sum_{i:a_i=0} W x_i + \sum_{i:a_i=1} p_i x_i + \sum_{i=1}^{r} q_i y_i - t \end{aligned} \tag{9}$$

We below verify Eq. (7).

Consider the case where $\mathbf{x} \neq \mathbf{a}$. Then there exists i such that either $0 = x_i \neq a_i = 1$ or $1 = x_i \neq a_i = 0$. Thus Eq. (9) implies that $p'(\mathbf{x},\mathbf{y})$ contains at least a term $-W$, and hence Eq. (8) implies that $p(\mathbf{x},\mathbf{y}) < 0$ for every $\mathbf{y} \in \{0,1\}^r$. Thus $g^{\mathbf{a}}(\mathbf{x},\mathbf{y}) = 0$. Consider the other case where $\mathbf{x} = \mathbf{a}$. Eq. (9) implies that all the $-W$s disappear, and hence we have $p(\mathbf{a},\mathbf{y}) = p'(\mathbf{a},\mathbf{y})$, as desired. ◀

Using Lemma 11, we construct depth-2 circuits computing the parity function.

▶ **Theorem 12.** *For any positive integer e, $3 \leq e \leq n$, PAR_n is computable by a depth-2 threshold circuit of size $O(e2^{n-2e})$ and energy e.*

Proof. We construct the desired circuit C. Let $l = n - 2(e-2)$ and $r = 2(e-2)$, and consider $\{0,1\}^n$ as $\{0,1\}^l \times \{0,1\}^r$.

We denote by g_j and h_j threshold gates such that g_j outputs one if and only if $j \leq \sum_{i=1}^{r} y_i$, and h_j outputs one if and only if $\sum_{i=1}^{r} y_i \leq j$: For every $\mathbf{x} \in \{0,1\}^l$ and $\mathbf{y} \in \{0,1\}^r$,

$$g_j(\mathbf{x},\mathbf{y}) = \mathrm{sig}\left(\sum_{i=1}^{r} y_i - j\right) \quad \text{and} \quad h_j(\mathbf{x},\mathbf{y}) = \mathrm{sig}\left(-\sum_{i=1}^{r} y_i + j\right).$$

For each $\mathbf{a} = (a_1, a_2, \ldots, a_n) \in \{0,1\}^l$, we make $2(e-2)$ or $2(e-2)+1$ suppressed gates depending on whether $|\mathbf{a}|$ is even or odd. If $|\mathbf{a}|$ is even, for every odd j satisfying $1 \leq j \leq r$, we make $g_j^{\mathbf{a}}$ and $h_j^{\mathbf{a}}$ (that is, we make $g_1^{\mathbf{a}}, h_1^{\mathbf{a}}, g_3^{\mathbf{a}}, h_3^{\mathbf{a}}$, and so on. Note that the weights for $x_1, x_2, \ldots, x_l$ are considered to be zeros when we apply the lemma.) Suppose $\mathbf{x} = \mathbf{a}$. Then, if $|\mathbf{y}|$ is even, exactly $e-2$ gates of $g_j^{\mathbf{a}}$s and $h_j^{\mathbf{a}}$s outputs one; otherwise, exactly $e-1$ gates of $g_j^{\mathbf{a}}$s and $h_j^{\mathbf{a}}$s outputs one. Similarly, if $|\mathbf{a}|$ is odd, for every even j satisfying $1 \leq j \leq r$, we make $g_j^{\mathbf{a}}$ and $g_j'^{\mathbf{a}}$ (that is, we make $g_2^{\mathbf{a}}, h_2^{\mathbf{a}}, g_4^{\mathbf{a}}, h_4^{\mathbf{a}}$, and so on). In addition, we add a suppressed gate $h_0^{\mathbf{a}}$. Suppose $\mathbf{x} = \mathbf{a}$. Then, if $|\mathbf{y}|$ is odd, exactly $e-2$ gates of $g_j^{\mathbf{a}}$s and $h_j^{\mathbf{a}}$s outputs one; otherwise, exactly $e-1$ gates of $g_j^{\mathbf{a}}$s and $g_j'^{\mathbf{a}}$s outputs one.

We then connect the outputs of $g_j^{\mathbf{a}}$ and $h_j^{\mathbf{a}}$ to the top gate of C with weight 1 for every $\mathbf{a} \in \{0,1\}^l$ and j. We finally set the threshold of the top gate to $r/2 + 1 = e - 1$. This completes the construction.

Consider an arbitrary fixed input $\mathbf{a}^*$ and $\mathbf{b}^*$. We verify that C computes PAR_n for the case where $|\mathbf{a}^*|$ is even, since the other case is similar. Lemma 11 implies that all the gates $g_j^{\mathbf{a}}$ and $h_j^{\mathbf{a}}$ such that $\mathbf{a}^* \neq \mathbf{a}$ output zeros, and hence the output of C is determined by $g_j^{\mathbf{a}^*}$s and $h_j^{\mathbf{a}^*}$s. Since the threshold of the top gate is $e-1$, C outputs one if and only if $|\mathbf{b}^*|$ is odd, as desired. Clearly, C has size at most $(r+1)2^l + 1 \leq 2(e-1)2^{n-2(e-2)} + 1$, depth 2 and energy e. ◀

4.2 Depth-3 Circuits

We here provide a construction of depth-3 circuits. The following construction shows that the exponent of the lower bound given in Theorem 10 is tight up to polylogarithmic factor for depth-3 circuits if energy e is constant.

▶ **Theorem 13.** *For any positive integer e, $2 \le e \le n$, PAR_n is computable by a depth-3 threshold circuit of size $O(e2^{n/(e-1)} + 2^{e-2})$ and energy e.*

Proof. Partition $[n]$ into $e-1$ disjoint sets $I_1, I_2, \ldots, I_{e-1}$, each of which has at most $\lceil n/(e-1) \rceil$ elements. Consider an input assignment $\mathbf{a}$ as $(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_{e-1}) \in \{0,1\}^{I_1} \times \{0,1\}^{I_2} \times \cdots \times \{0,1\}^{I_{e-1}}$.

Then, for each $j \in [e-1]$, we define

$$S_j = \{\mathbf{a}_j \in \{0,1\}^{I_j} \mid |\mathbf{a}_j| \text{ is odd}\}.$$

For every $j \in [e-1]$ and $\mathbf{a}_j \in S_j$, we make a threshold gate that has input variables x_i for all $i \in I_j$, and outputs one if and only if the gate receives $\mathbf{a}_j$.

Consequently, we have $2^{n/(e-1)}$ gates for each $j \in [e-1]$, and exactly one of the gates outputs one if $|\mathbf{a}_j|$ is odd, and no gate outputs one, otherwise. Thus, we can regard the outputs of the $2^{n/(e-1)}$ gates as a single input variable, and hence we can complete the construction by feeding them into a depth-2 circuit given by Theorem 2 which is of size 2^{e-2}, depth 2 and energy one, and computes PAR_{e-1}.

Since, the resulting circuit C has at most $(e-1)2^{n/(e-1)}$ gates in the first layer, 2^{e-2} gates in the second layer, and the top gate in the third layer, C has size $(e-1)2^{n/(e-1)} + 2^{e-2} + 1$ and depth 3. Moreover, at most $e-1$ gates outputs one in the first layer, and at most one gate outputs one in the second and third layers, and hence C has energy e. ◀

4.3 Depth-d Circuits

Using a similar idea to the one given in Section 4.1, we provide circuits of depth d and energy e. Let C be a threshold circuit. We define a *suppressed circuit $C^{\mathbf{a}}$ of C for $\mathbf{a}$* as a threshold circuit in which every gate g is replaced by the suppressed gate $g^{\mathbf{a}}$ of g for $\mathbf{a}$. Clearly, we have

$$C^{\mathbf{a}}(\mathbf{x}, \mathbf{y}) = \begin{cases} 0 & \text{if } \mathbf{x} \neq \mathbf{a}; \\ C(\mathbf{a}, \mathbf{y}) & \text{if } \mathbf{x} = \mathbf{a}, \end{cases} \tag{10}$$

and, moreover, if $\mathbf{x} \neq \mathbf{a}$ then no gate in $C^{\mathbf{a}}$ outputs one, and the the energy of $C^{\mathbf{a}}$ is at most the one of C.

Using the known circuit constructions given in Theorems 3 and 4, we show that we can decrease size if large depth or energy are allowed.

▶ **Theorem 14.** *For any positive integers d and e, $3 \le d, e \le n$, PAR_n is computable by a threshold circuit of size $O((e+d)2^{n-m})$, depth $d + O(1)$ and energy $e + O(1)$, where*

$$m = \max\left(\left(\frac{e-1}{d-1}\right)^{d-1}, \left(\frac{d-1}{e-1}\right)^{e-1}\right).$$

Proof. Consider first the case where it holds that

$$m = \left(\frac{e-1}{d-1}\right)^{d-1}.$$

For simplicity, we assume that m is an integer. In this case, Theorem 3 implies that a threshold circuit C of size

$$s = (d-1)\lceil m^{1/(d-1)} \rceil + 1 = e,$$

and depth d, and energy $s(= e)$ computes PAR_m. Moreover, it is known that a threshold gate is closed under complement [20], and hence there exists a circuit D of size e, depth d, and energy at most e that computes the complement of PAR_m. For each $\mathbf{a} \in \{0,1\}^{n-m}$, we construct a suppressed circuit as follows: If $|\mathbf{a}|$ is even, we make $C^{\mathbf{a}}$, and otherwise, we make $D^{\mathbf{a}}$. Equation (10) implies that by adding, as the top gate, a single threshold gate computing OR of all the $C^{\mathbf{a}}$s and $D^{\mathbf{a}}$s, we obtain the desired circuit C^* that computes PAR_n. Clearly, C^* has size $e2^{n-m} + 1$, depth $d + O(1)$, and energy $e + O(1)$.

For the other case where

$$m = \left(\frac{d-1}{e-1}\right)^{e-1},$$

we can similarly construct the desired circuit of size $d2^{n-m} + 1$, depth $d + O(1)$, and energy $e + O(1)$ based on Theorem 4. We omit the details. ◀

5 Conclusion

In this paper, we prove lower and upper bounds on the size of threshold circuits computing the parity function. Although the parity function has constant communication complexity, our lower bounds are exponential if depth and energy are constant. Since there is a large gap between our bounds, improving on these bounds is an interesting future work. We are also interested in whether similar relations among the size, depth and energy exists for other computational tasks.

References

1 K. Amano and A. Maruoka. On the complexity of depth-2 circuits with threshold gates. In *Proc. of the 30th international conference on Mathematical Foundations of Computer Science*, pages 107–118, 2005.

2 M. Anthony and P. L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999.

3 L. Chen and R. Tell. Bootstrapping results for threshold circuits "just beyond" known lower bounds. In *Proc. of the 51st Annual ACM Symposium on Theory of Computing*, pages 34–41, 2019.

4 R. Chen, R. Santhanam, and S. Srinivasan. Average-case lower bounds and satisfiability algorithms for small threshold circuits. *Theory of Computing*, 14(9):1–55, 2018.

5 K. Dinesh, S. Otiv, and J. Sarma. New bounds for energy complexity of boolean functions. In *Proc. of the 24th International Computing and Combinatorics Conference*, pages 738–750, 2018.

6 P. Földiák. Sparse coding in the primate cortex. In M. A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 1064–1068. MIT Press, 1995.

7 J. Forster, M. Krause, S. V. Lokam, R. Mubarakzjanov, N. Schmitt, and H. U. Simon. Relations between communication complexity, linear arrangements, and computational complexity. In *Proc. of the 21st International Conference on Foundations of Software Technology and Theoretical Computer Science*, pages 171–182, 2001.

8 A. Hajnal, W. Maass, P. Pudlák, M. Szegedy, and G. Turán. Threshold circuits of bounded depth. *Journal of Computer and System Sciences*, 46:129–154, 1993.

9 J. Håstad. *Computational Limitations of Small-depth Circuits*. MIT Press, 1987.
10 J. Håstad and M. Goldmann. On the power of small-depth threshold circuits. *Computational Complexity*, 1(2):113–129, 1991.
11 Y. He, K. Kavukcuoglu, Y. Wang, A. Szlam, and Y. Qi. Unsupervised feature learning by deep sparse coding. In *Proc. of SIAM International Conference on Data Mining*, pages 902–910, 2014.
12 R. Impagliazzo, R. Paturi, and M. E. Saks. Size-depth tradeoffs for threshold circuits. *SIAM Journal on Computing*, 26(3):693–707, 1997.
13 S. Jukna. *Extremal Combinatorics with Applications in Computer Science*. Springer-Verlag Berlin Heidelberg, 2011.
14 D. M. Kane and R. Williams. Super-linear gate and super-quadratic wire lower bounds for depth-two and depth-three threshold circuits. In *Proc. of the 48th Annual ACM Symposium on Theory of Computing*, pages 633–643, 2016.
15 H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. In *Proc. of the 19th Advances in Neural Information Processing Systems*, pages 801–808, 2006.
16 P. Lennie. The cost of cortical computation. *Current Biology*, 13:493–497, 2003.
17 M. Minsky and S. Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, 1988.
18 A. Y. Ng. Sparse autoencoder. CS294A Lecture notes, 2011.
19 B. Olshausen and D. J. Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4):481–487, 2004.
20 I. Parberry. *Circuit Complexity and Neural Networks*. MIT Press, 1994.
21 A. A. Razborov and A. A. Sherstov. The sign-rank of ac^0. *SIAM Journal on Computing*, 39(5):1833–1855, 2010.
22 K. Y. Siu, V. Roychowdhury, and T. Kailath. *Discrete Neural Computation: A Theoretical Foundation*. Prentice Hall, 1995.
23 X. Sun, Y. Sun, K. Wu, and Z. Xia. On the relationship between energy complexity and other boolean function measures. In *Proc. of the 25th International Computing and Combinatorics Conference*, pages 516–528, 2019.
24 A. Suzuki, K. Uchizawa, and X. Zhou. Energy-efficient threshold circuits computing MOD functions. In *Proc. of the 17th Computing: the Australasian Theory Symposium*, pages 105–110, 2011.
25 A. Suzuki, K. Uchizawa, and X. Zhou. Energy-efficient threshold circuits computing MOD functions. *International Journal of Foundations of Computer Science*, 24(1):15–29, 2013.
26 K. Uchizawa. Lower bounds for threshold circuits of bounded energy. *Interdisciplinary Information Sciences*, 20(1):27–50, 2014.
27 K. Uchizawa, R. Douglas, and W. Maass. On the computational power of threshold circuits with sparse activity. *Neural Computation*, 18(12):2994–3008, 2008.
28 K. Uchizawa and E. Takimoto. Exponential lower bounds on the size of constant-depth threshold circuits with small energy complexity. *Theoretical Computer Science*, 407(1–3):474–487, 2008.
29 K. Uchizawa, E. Takimoto, and T. Nishizeki. Size-energy tradeoffs of unate circuits computing symmetric Boolean functions. *Theoretical Computer Science*, 412:773–782, 2011.
30 A. C-C. Yao. Separating the polynomial-time hierarchy by oracles. In *Proc. of the 26th Annual Symposium on Foundations of Computer Science*, pages 1–10, 1985.