Exponential Lower Bounds for Threshold Circuits of Sub-Linear Depth and Energy

Kei Uchizawa ✉
Graduate School of Science and Engineering, Yamagata University, Yonezawa-shi Yamagata, Japan

Haruki Abe
Graduate School of Science and Engineering, Yamagata University, Yonezawa-shi Yamagata, Japan

Abstract

In this paper, we investigate computational power of threshold circuits and other theoretical models of neural networks in terms of the following four complexity measures: size (the number of gates), depth, weight and energy. Here, the energy of a circuit measures sparsity of their computation, and is defined as the maximum number of gates outputting non-zero values taken over all the input assignments.

As our main result, we prove that any threshold circuit C of size s, depth d, energy e and weight w satisfies $\log(rk(M_C)) \le ed(\log s + \log w + \log n)$, where $rk(M_C)$ is the rank of the communication matrix M_C of a $2n$-variable Boolean function that C computes. Thus, such a threshold circuit C is able to compute only a Boolean function of which communication matrix has rank bounded by a product of logarithmic factors of s, w and linear factors of d, e. This implies an exponential lower bound on the size of even sublinear-depth and sublinear-energy threshold circuit. For example, we can obtain an exponential lower bound $s = 2^{\Omega(n^{1/3})}$ for threshold circuits of depth $n^{1/3}$, energy $n^{1/3}$ and weight $2^{o(n^{1/3})}$. We also show that the inequality is tight up to a constant factor when the depth d and energy e satisfies $ed = o(n/\log n)$.

For other models of neural networks such as a discretized ReLU circuits and descretized sigmoid circuits, we define energy as the maximum number of gates outputting non-zero values. We then prove that a similar inequality also holds for a discretized circuit C: $rk(M_C) = O(ed(\log s + \log w + \log n)^3)$. Thus, if we consider the number gates outputting non-zero values as a measure for sparse activity of a neural network, our results suggest that larger depth linearly helps neural networks to acquire sparse activity.

2012 ACM Subject Classification Theory of computation → Models of computation

Keywords and phrases Circuit complexity, disjointness function, equality function, neural networks, threshold circuits, ReLU cicuits, sigmoid circuits, sparse activity

Digital Object Identifier 10.4230/LIPIcs.MFCS.2023.85

Related Version *Full Version*: `https://arxiv.org/abs/2107.00223`

Funding *Kei Uchizawa*: This work was supported by JSPS KAKENHI Grant Number JP22K11897.

Acknowledgements We thank the anonymous reviewers for their careful reading and helpful comments.

1 Introduction

Background. DiCarlo and Cox argued that constructing good internal representations is crucial to perform visual information processing, such as object recognition, for neural networks in the brain [5]. Here, an internal representation is described by a vector in a very high dimensional space, where each axis is one neuron's activity and the dimensionality equals to the number (e.g., ∼1 million) of neurons in a feedforward neural network. They call a representation good if, for a given pair of two images that are hard to distinguish at the input space, there exist representations for them that are easy to separate by simple classifiers such as a linear classifier. While such internal representations are likely to play fundamental role in information processing in the brain, it is also known that a neuron needs relatively high energy to be active [18, 27], and hence neural networks are forced to acquire representations supported by only a small number of active neurons [7]. These observations pose a question: for what information processing can neural networks construct good internal representations?

In the paper [42], Uchizawa *et al.* address the question from the viewpoint of circuit complexity. More formally, they employed threshold circuits as a model of neural networks [23, 24, 28, 31, 34, 35, 36], and introduced a complexity measure, called energy complexity, for sparsity of their internal representations. A threshold circuit is a feedforward logic circuit whose basic computational element computes a linear threshold function, and energy of a circuit is defined as the maximum number of internal gates outputting ones over all the input assignments. (See also [6, 16, 33, 37, 45] for studies on energy complexity of other types of logic circuits). Uchizawa *et al.* then show that the energy complexity is closely related to the rank of linear decision trees. In particular, they prove that any linear decision tree of l leaves can be simulated by a threshold circuit of size $O(l)$ and energy $O(\log l)$. Thus, even logarithmic-energy threshold circuits have certain computational power: any linear decision tree of polynomial number of leaves can be simulated by a polynomial-size and logarithmic-energy threshold circuit.

Following the paper [42], a sequence of papers show relations among other major complexity measures such as size (the number of gates), depth, weight and fan-in [22, 38, 39, 43, 41, 40, 44]. In particular, Uchizawa and Takimoto [43] showed that any threshold circuit C of depth d and energy e requires size $s = 2^{\Omega(n/e^d)}$ if C computes a high bounded-error communication complexity function such as Inner-Product function. Even for low communication complexity functions, an exponential lower bound on the size is known for constant-depth threshold circuits: any threshold circuit C of depth d and energy e requires size $s = 2^{\Omega(n/e2^{e+d}\log^e n)}$ if C computes the parity function [41]. These results provide exponential lower bounds if the depth is constant and energy is sub-linear [43] or sub-logarithmic [41], while both Inner-Product function and Parity function are computable by linear-size, constant-depth, and linear-energy threshold circuits. Thus these results imply that the energy complexity strongly related to representational power of threshold circuits. However these lower bounds break down when we consider threshold circuits of larger depth and energy, say, non-constant depth and sub-linear energy.

Our Results for Threshold Circuits. In this paper, we prove that simple Boolean functions are hard even for sub-linear depth and sub-linear energy threshold circuits. Let C be a threshold circuit with Boolean input variables $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$. A communication matrix M_C of C is a $2^n \times 2^n$ matrix where each row (resp., each column) is indexed by an assignment $\mathbf{a} \in \{0,1\}^n$ to $\mathbf{x}$ (resp., $\mathbf{b} \in \{0,1\}^n$ to $\mathbf{y}$), and the value $M_C[\mathbf{a}, \mathbf{b}]$ is defined to be the output of C given $\mathbf{a}$ and $\mathbf{b}$. We denote by $rk(M_C)$ the rank of M_C over $\mathbb{F}_2$. Our main result is the following relation among size, depth energy and weight.

▶ **Theorem 1.** *Let s, d, e and w be integers satisfying $2 \le s, d$, $10 \le e$, $1 \le w$. If a threshold circuit C computes a Boolean function of $2n$ variables, and has size s, depth d, energy e and weight w, then it holds that*

$$\log(rk(M_C)) \le ed(\log s + \log w + \log n). \quad (1)$$

The theorem implies exponential lower bounds for sub-linear depth and sub-linear energy threshold circuits. As an example, let us consider a Boolean function CD_n defined as follows: For a $2n$ input variables $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$,

$$\mathrm{CD}_n(\mathbf{x}, \mathbf{y}) = \bigvee_{i=1}^{n} x_i \wedge y_i.$$

We note that CD_n is a biologically motivated Boolean function: Maass [21] defined CD_n to model coincidence detection or a pattern matching, and Lynch and Musco [19] introduced a related problem, called Filter problem, for studying theoretical aspect of spiking neural networks. Since CD_n is the complement of the disjointness function and has rank 2^n, the theorem implies that

$$n \leq ed(\log s + \log w + \log n) \tag{2}$$

holds if a threshold circuit C computes CD_n. Arranging Eq. (2), we can obtain a lower bound $2^{n/(ed)}/(wn) \leq s$ which is exponential in n if both d and e are sub-linear and w is sub-exponential. For example, we can obtain an exponential lower bound $s = 2^{\Omega(n^{1/3})}$ even for threshold circuits of depth $n^{1/3}$, energy $n^{1/3}$ and weight $2^{o(n^{1/3})}$. We can obtain similar lower bounds for the Inner-Product function and the equality function, since they have linear rank.

Comparing the lower bound $s = 2^{\Omega(n/e^d)}$ given in [43] to ours, our lower bound is meaningful only for sub-exponential weight, but improves on it in two-fold: the lower bound is exponential even if d is sub-linear, and provide a nontrivial lower bound for Boolean functions with much weaker condition: Threshold circuits need exponential size even for Boolean functions of the standard rank $\Omega(n)$.

Threshold circuits have received considerable attention in circuit complexity, and a number of lower bound arguments have developed for threshold circuits under some restrictions on computational resources including size, depth, energy and weight [1, 2, 3, 9, 10, 13, 15, 22, 26, 30, 32, 41, 43, 44]. However, the arguments for lower bounds are designated for constant-depth threshold circuits, and hence cannot provide meaningful ones when the depth is not constant. In particular, CD_n is computable by a depth-2 and linear-size threshold circuit. Thus, directly applying known techniques are unlikely to yield an exponential lower bound for CD_n.

To complement Theorem 1, we also show that the lower bound is tight up to a constant factor if the product of e and d are small:

▶ **Theorem 2.** *For any integers e and d such that $2 \leq e$ and $2 \leq d$, CD_n is computable by a threshold circuit of size*

$$s \leq (e-1)(d-1) \cdot 2^{\frac{n}{(e-1)(d-1)}}.$$

depth d, energy e and weight

$$w \leq \left(\frac{n}{(e-1)(d-1)}\right)^2.$$

Substituting s, d, e and w of a threshold circuit given in Theorem 2 to the right hand side of Eq. (2), we have

$$\begin{aligned} & ed(\log s + \log w + \log n) \\ \leq\ & ed\left(\frac{n}{(e-1)(d-1)} + \log(e-1)(d-1) + \log\left(\frac{n}{(e-1)(d-1)}\right)^2 + \log n\right) \\ \leq\ & 4n + O(ed \log n), \end{aligned}$$

which almost matches the left hand side of Eq. (2) if $ed = o(n/\log n)$. Thus, Theorem 1 neatly captures the computational aspect of threshold circuits computing CD_n. Recall that any linear decision tree of polynomial number of leaves can be simulated by a polynomial-size and logarithmic-energy threshold circuit [42]. Also, it is known that any Boolean function is computable by a threshold circuit of depth two and energy one if an exponential size is allowed [22]. Thus, we believe that the situation $ed = o(n/\log n)$ is not too restrictive. We also show that the lower bound is also tight for the equality function.

Our Result for Discretized Circuits. Besides threshold circuits, we consider other well-studied models of neural network, where an activation function and weights of an computational element are discretized (such as, discretized sigmoid or ReLU circuits). The size, depth, energy and weight are important parameters also for artificial neural networks. The size and depth are major topics on success of deep learning. The energy is related to important techniques for deep learning method such as regularization, sparse coding, or sparse autoencoder [11, 17, 25]. The weight resolution is closely related to chip resources in neuromorphic hardware systems [29], and quantization schemes received attention [4, 12].

We define energy for a discretized circuit as the maximum number of gates outputting non-zero values, and show that any discretized circuit can be simulated by a threshold circuit with a moderate increase in size, depth, energy, and weight. Consequently, combining with Theorem 1, we can show that its rank is bounded by a product of the polylogarithmic factors of s, w and linear factors of d, e for discretized circuits. For example, we can obtain the following proposition for discretized sigmoid circuits:

▶ **Theorem 3.** *If a discretized sigmoid circuit C of size s, depth d, energy e, and weight w computes a Boolean function f, then it holds that*

$$\log(rk(M_C)) = O(ed(\log s + \log w + \log n)^3).$$

Maass, Schnitger and Sontag [20] showed that a sigmoid circuit could be simulated by a threshold circuit, but their simulation was optimized to be depth-efficient and did not consider energy. Thus, their result does not fit into our purpose.

Theorems 1 and 3 imply that a threshold circuit or discretized circuit are able to compute a Boolean function of bounded rank. Thus, we can consider these theorems as bounds on corresponding concept classes. According to the bound, c times larger depth is comparable to 2^c times larger size. Thus, large depth could enormously help neural networks to increase its expressive power. Also, the bound suggests that increasing depth could also help a neural network to acquire sparse activity when we have hardware constraints on both the number of neurons and the weight resolution. These observations may shed some light on the reason for the success of deep learning.

Organization. The rest of the paper is organized as follows. In Section 2, we define terms needed for analysis. In Section 3, we present our main lower bound result. In Section 4, we show the tightness of the lower bound. In Section 5, we show a bound for discretized circuits. In Section 6, we conclude with some remarks.

2 Preliminaries

For an integer n, we denote by $[n]$ a set $\{1, 2, \ldots n\}$. The base of the logarithm is two unless stated otherwise. In Section 2.1, we define terms on threshold circuits and discretized circuits. In Section 2.2, we define communication matrix, and give some known facts.

2.1 Circuit Model

In Sections 2.1.1 and 2.1.2, we give definitions of threshold and discritized circuits, respectively.

2.1.1 Threshold Circuits

Let k be a positive integer. A *threshold gate* g with k input variables $\xi_1, \xi_2, \ldots, \xi_k$ has weights $w_1, w_2, \ldots, w_k$, and a threshold t. We define the output $g(\xi_1, \xi_2, \ldots, \xi_k)$ of g as

$$g(\xi_1, \xi_2, \ldots, \xi_k) = \mathrm{sign}\left(\sum_{i=1}^{k} w_i \xi_i - t\right) = \begin{cases} 1 & \text{if } t \leq \sum_{i=1}^{k} w_i \xi_i; \\ 0 & \text{otherwise} \end{cases}$$

To evaluate the weight resolution, we assume single synaptic weight to be discrete, and that $w_1, w_2, \ldots, w_n$ are integers. The *weight* w_g *of* g is defined as the maximum of the absolute values of $w_1, w_2, \ldots, w_k$. In other words, we assume that $w_1, w_2, \ldots, w_k$ are $O(\log w_g)$-bit coded discrete values. Throughout the paper, we allow a gate to have both positive and negative weights, although biological neurons are either excitatory (all the weights are positive) or inhibitory (all the weights are negative). As mentioned in [21], this relaxation has basically no impact on circuit complexity investigations, unless one cares about constant blowup in computational resources.

A *threshold circuit* C is a combinatorial circuit consisting of threshold gates, and is expressed by a directed acyclic graph. The nodes of in-degree 0 correspond to input variables, and the other nodes correspond to gates. Let G be a set of the gates in C. For each gate $g \in G$, the *level* of g, denoted by $\mathrm{lev}(g)$, is defined as the length of a longest path from an input variable to g on the underlying graph of C. For each $l \in [d]$, we define G_l as a set of gates in the lth level: $G_l = \{g \in G \mid \mathrm{lev}(g) = l\}$.

In this paper, we consider a threshold circuit C for a Boolean function $f : \{0,1\}^{2n} \to \{0,1\}$. Thus, C has $2n$ Boolean input variables $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, and a unique output gate, denoted by g^{clf}, which is a linear classifier separating internal representations given by the gates in the lower levels (possibly together with input variables). Consider a gate g in C. Let $w_1^x, w_2^x, \ldots, w_n^x$ (resp., $w_1^y, w_2^y, \ldots, w_n^y$) be the weights for $x_1, x_2, \ldots, x_n$ (resp., $y_1, y_2, \ldots, y_n$), and t_g be threshold of g. For each gate h directed to g, let $w_{h,g}$ be a weight of g for the output of h. Then the output $g(\mathbf{x}, \mathbf{y})$ of g is defined as

$$g(\mathbf{x}, \mathbf{y}) = \mathrm{sign}\left(p_g(\mathbf{x}, \mathbf{y}) - t_g\right)$$

where $p_g(\mathbf{x}, \mathbf{y})$ denotes a potentials of g invoked by the input variables and gates:

$$p(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} w_i^x x_i + \sum_{i=1}^{n} w_i^y y_i + \sum_{l=1}^{\mathrm{lev}(g)-1} \sum_{h \in G_l} w_{h,g} h(\mathbf{x}, \mathbf{y}).$$

We sometimes write $p_g^x(\mathbf{x})$ (resp., $p_g^y(\mathbf{y})$) for the potential invoked by $\mathbf{x}$ (resp., $\mathbf{y}$):

$$p_g^x(\mathbf{x}) = \sum_{i=1}^{n} w_i^x x_i \quad \text{and} \quad p_g^y(\mathbf{y}) = \sum_{i=1}^{n} w_i^y y_i.$$

Although the inputs to g are not only $\mathbf{x}$ and $\mathbf{y}$ but the outputs of gates in the lower levels, we write $g(\mathbf{x}, \mathbf{y})$ for the output of g, because $\mathbf{x}$ and $\mathbf{y}$ inductively decide the output of g. We say that C *computes* a Boolean function $f : \{0,1\}^{2n} \to \{0,1\}$ if $g^{\mathrm{clf}}(\mathbf{a}, \mathbf{b}) = f(\mathbf{a}, \mathbf{b})$ for every $(\mathbf{a}, \mathbf{b}) \in \{0,1\}^{2n}$.

Let C be a threshold circuit. We define *size* s of C as the number of the gates in C, and *depth* d of C as the level of g^{clf}. We define the *energy* e of C as

$$e = \max_{(\mathbf{a},\mathbf{b})\in\{0,1\}^{2n}} \sum_{g\in G} g(\mathbf{a},\mathbf{b}).$$

We define *weight* w of C as the maximum of the weights of the gates in C: $w = \max_{g\in G} w_g$.

2.1.2 Discretized Circuits

Let φ be an activation function. Let δ be a discretizer that maps a real number to a number representable by a bitwidth b. We define a discretized activation function $\delta \circ \varphi$ as a composition of φ and δ, that is, $\delta \circ \varphi(x) = \delta(\varphi(x))$ for any number x. We say that $\delta \circ \varphi$ has *silent range for an interval* I if $\delta \circ \varphi(x) = 0$ if $x \in I$, and $\delta \circ \varphi(x) \neq 0$, otherwise. For example, if we use the ReLU function as the activation function φ, then $\delta \circ \varphi$ has silent range for $I = (-\infty, 0]$ for any discretizer δ. If we use the sigmoid function as the activation function φ and linear partition as discretizer δ, then $\delta \circ \varphi$ has silent range for $I = (-\infty, t_{\max}]$ where $t_{\max} = \ln(1/(2^b - 1))$ where ln is the natural logarithm.

Let $\delta \circ \varphi$ be a discretized activation function with silent range. A $(\delta \circ \varphi)$-*gate* g with k input variables $\xi_1, \xi_2, \ldots, \xi_k$ has weights $w_1, w_2, \ldots, w_k$ and a threshold t, where each of the weights and threshold are discretized by δ. The output $g(\xi_1, \xi_2, \ldots, \xi_k)$ of g is then defined as

$$g(\xi_1, \xi_2, \ldots, \xi_k) = \delta \circ \varphi \left(\sum_{i=1}^{k} w_i \xi_i - t \right).$$

A $(\delta \circ \varphi)$-*circuit* is a combinatorial circuit consisting of $(\delta \circ \varphi)$-gates except that the top gate g^{clf} is a threshold gate, that is, a linear classifier. We define size and depth of a $(\delta \circ \varphi)$-circuit same as the ones for a threshold circuit. We define energy e of a $(\delta \circ \varphi)$-circuit as the maximum number of gates outputting non-zero values in the circuit:

$$e = \max_{(\mathbf{a},\mathbf{b})\in\{0,1\}^{2n}} \sum_{g\in G} [\![g(\mathbf{a},\mathbf{b}) \neq 0]\!]$$

where $[\![\mathrm{P}]\!]$ for a statement P denote a notation of the function which outputs one if P is true, and zero otherwise. We define *weight* w of C as $w = 2^{2b}$, where $2b$ is the bitwidth possibly needed to represent a potential value invoked by a single input of a gate in C.

2.2 Communication Matrix and its Rank

Let $Z \subseteq \{0,1\}^n$. For a Boolean function $f : Z \times Z \to \{0,1\}$, we define a communication matrix M_f over Z as a $|Z| \times |Z|$ matrix where each row and column are indexed by $\mathbf{a} \in Z$ and $\mathbf{b} \in Z$, respectively, and each entry is defined as $M_f(\mathbf{a},\mathbf{b}) = f(\mathbf{a},\mathbf{b})$. We denote by $rk(M_f)$ the rank of M_f over $\mathbb{F}_2$. If a circuit C computes f, we may write M_C instead of M_f. If a Boolean function f does not have an obvious separation of the input variables to $\mathbf{x}$ and $\mathbf{y}$, we may assume a separation so that $rk(M_f)$ is maximized.

Let k and n be natural numbers such that $k \leq n$. Let

$$Z_k = \{\mathbf{a} \in \{0,1\}^n \mid \text{The number of ones in } \mathbf{a} \text{ is at most } k\}.$$

A k-disjointness function $\mathrm{DISJ}_{n,k}$ over Z_k is defines as follows:

$$\mathrm{DISJ}_{n,k}(\mathbf{x},\mathbf{y}) = \bigwedge_{i=1}^{n} \overline{x_i} \vee \overline{y_i}$$

where the input assignments are chosen from Z_k. The book [14] contains a simple proof showing that $\mathrm{DISJ}_{n,k}$ has full rank [14].

▶ **Theorem 4.** *$rk(M_{\mathrm{DISJ}_{n,k}}) = \sum_{i=0}^{k} \binom{n}{i}$. In particular, $rk(M_{\mathrm{DISJ}_{n,n}}) = 2^n$.*

CD_n is the complement of $\mathrm{DISJ}_{n,n}$. We can obtain the same bound for CD_n, as follows:

▶ **Corollary 5.** *$rk(M_{\mathrm{CD}_n}) = 2^n$.*

We also use well-known facts on the rank. Let A and B be two matrices of same dimensions. We denote by $A + B$ the summation of A and B, and by $A \circ B$ the Hadamard product of A and B.

▶ **Fact 1.** *For two matrices A and B of same dimensions, we have*
(i) $rk(A + B) \leq rk(A) + rk(B)$;
(ii) $rk(A \circ B) \leq rk(A) \cdot rk(B)$.

3 Lower Bound for Threshold Circuits

In this section, we give the inequality relating the rank of the communication matrix to the size, depth, energy and weight.

▶ **Theorem 6** (Theorem 1 restated)**.** *Let s, d, e and w be integers satisfying $2 \leq s, d$, $10 \leq e$, $1 \leq w$. Suppose a threshold circuit C computes a Boolean function of $2n$ variables, and has size s, depth d, energy e, and weight w. Then it holds that*

$$\log(rk(M_C)) \leq ed(\log s + \log w + \log n).$$

We prove the theorem by showing that M_C is a sum of matrices each of which corresponds to an internal representation that arises in C. Since C has bounded energy, the number of internal representations is also bounded. We then show by the inclusion-exclusion principle that each matrix corresponding to an internal representation has bounded rank. Thus, Fact 1 implies the theorem.

Proof. Let C be a threshold circuit that computes a Boolean function of $2n$ variables, and has size s, depth d, energy e and weight w. Let G be a set of the gates in C. For $l \in [d]$, let G_l be a set of the gates in l-th level of C. Without loss of generality, we assume that $G_d = \{g^{\mathrm{clf}}\}$. We evaluate the rank of M_C, and prove that

$$rk(M_C) \leq \left(\frac{c \cdot s}{e-1}\right)^{e-1} \cdot \left(\left(\frac{c \cdot s}{e-1}\right)^{e-1} \cdot (2nw+1)^{e-1}\right)^{d-1} \cdot (2nw+1) \tag{3}$$

where $c < 3$. Equation (3) implies that

$$rk(M_C) \leq \left(\frac{c \cdot s}{e-1} \cdot (2nw+1)\right)^{(e-1)d} \leq (snw)^{ed},$$

where the last inequality holds if $e \geq 10$. Taking the logarithm of the inequality, we obtain the theorem.

Below we verify that Eq. (3) holds. Let $\mathbf{P} = (P_1, P_2, \ldots, P_d)$, where P_l is a subset of G_l for each $l \in [d]$. Given an input $(\mathbf{a}, \mathbf{b}) \in \{0,1\}^{2n}$, we say that an *internal representation* $\mathbf{P}$ *arises for* $(\mathbf{a}, \mathbf{b})$ if, for every $l \in [d]$, $g(\mathbf{a}, \mathbf{b}) = 1$ for every $g \in P_l$, and $g(\mathbf{a}, \mathbf{b}) = 0$ for every $g \notin P_l$. We denote by $\mathbf{P}^*(\mathbf{a}, \mathbf{b})$ the internal representation that arises for $(\mathbf{a}, \mathbf{b}) \in \{0,1\}^{2n}$. We then define $\mathcal{P}_1$ as a set of the internal representations that arise for $(\mathbf{a}, \mathbf{b})$ such that $g^{\mathrm{clf}}(\mathbf{a}, \mathbf{b}) = 1$:

$$\mathcal{P}_1 = \{\mathbf{P}^*(\mathbf{a}, \mathbf{b}) \mid g^{\mathrm{clf}}(\mathbf{a}, \mathbf{b}) = 1\}.$$

Note that, for any $\mathbf{P} = (P_1, P_2, \ldots, P_d) \in \mathcal{P}_1$, we have $|P_1| + |P_2| + \cdots + |P_{d-1}| \leq e - 1$ and $|P_d| = 1$. Thus a standard upper bound on a sum of binomial coefficients implies that

$$|\mathcal{P}_1| \leq \sum_{k=0}^{e-1} \binom{s}{k} \leq \left(\frac{c \cdot s}{e-1} \right)^{e-1}. \tag{4}$$

For each $\mathbf{P} \in \mathcal{P}_1$, let $M_{\mathbf{P}}$ be a $2^n \times 2^n$ matrix such that, for every $(\mathbf{a}, \mathbf{b}) \in \{0,1\}^{2n}$,

$$M_{\mathbf{P}}(\mathbf{a}, \mathbf{b}) = \begin{cases} 1 & \text{if } \mathbf{P} = \mathbf{P}^*(\mathbf{a}, \mathbf{b}); \\ 0 & \text{if } \mathbf{P} \neq \mathbf{P}^*(\mathbf{a}, \mathbf{b}). \end{cases}$$

By the definitions of $\mathcal{P}_1$ and $M_{\mathbf{P}}$, we have

$$M_C = \sum_{\mathbf{P} \in \mathcal{P}_1} M_{\mathbf{P}},$$

and hence Fact 1(i) implies that

$$rk(M_C) \leq \sum_{\mathbf{P} \in \mathcal{P}_1} rk(M_{\mathbf{P}}).$$

Thus Eq. (4) implies that

$$rk(M_C) \leq \left(\frac{c \cdot s}{e-1} \right)^{e-1} \cdot \max_{\mathbf{P} \in \mathcal{P}_1} rk(M_{\mathbf{P}}).$$

We complete the proof by showing that, for any $\mathbf{P} \in \mathcal{P}_1(C)$, it holds that

$$rk(M_{\mathbf{P}}) \leq \left(\left(\frac{c \cdot s}{e-1} \right)^{e-1} \cdot (2nw+1)^{e-1} \right)^{d-1} \cdot (2nw+1).$$

In the following argument, we consider an arbitrary fixed internal representation $\mathbf{P} = (P_1, P_2, \ldots, P_d)$ in $\mathcal{P}_1$. We call a gate a *threshold function* if the inputs of the gate consists of only $\mathbf{x}$ and $\mathbf{y}$. For each $g \in G$, we denote by $\tau[g, \mathbf{P}]$ a threshold function defined as

$$\tau[g, \mathbf{P}](\mathbf{x}, \mathbf{y}) = \operatorname{sign}\left(p_g^x(\mathbf{x}) + p_g^y(\mathbf{y}) + t_g[\mathbf{P}] \right).$$

where $t_g[\mathbf{P}]$ is a threshold of g, being assumed that the internal representation $\mathbf{P}$ arises:

$$t_g[\mathbf{P}] = \sum_{l=1}^{\mathrm{lev}(g)-1} \sum_{h \in P_l} w_{h,g} - t_g.$$

For each $l \in [d]$, we define a set T_l of threshold functions as $T_l = \{\tau[g, \mathbf{P}] \mid g \in G_l\}$. Since every gate in G_1 is a threshold function, T_1 is identical to G_1.

For any set T of threshold functions, we denote by $M[T]$ a $2^n \times 2^n$ matrix such that, for every $(\mathbf{a}, \mathbf{b}) \in \{0,1\}^{2n}$,

$$M[T](\mathbf{a}, \mathbf{b}) = \begin{cases} 1 & \text{if } \forall \tau \in T, \tau(\mathbf{a}, \mathbf{b}) = 1; \\ 0 & \text{if } \exists \tau \in T, \tau(\mathbf{a}, \mathbf{b}) = 0. \end{cases}$$

It is well-known that the rank of $M[T]$ is bounded [8, 9], as follows. We give a proof for completeness.

▷ Claim 7. $rk(M[T]) \leq (2nw+1)^{|T|}$.

Proof. Let $z = |T|$, and $\tau_1, \tau_2, \ldots, \tau_z$ be an arbitrary order of threshold functions in T. For each $k \in [z]$, we define

$$R_k = \{p^x_{\tau_k}(\mathbf{a}) \mid \mathbf{a} \in \{0,1\}^n\}.$$

Since a threshold function receives a value between $-w$ and w from a single input, we have $|R_k| \le 2nw + 1$. For $\mathbf{r} = (r_1, r_2, \ldots, r_z) \in R_1 \times R_2 \times \cdots \times R_z$, we define $R(\mathbf{r}) = X(\mathbf{r}) \times Y(\mathbf{r})$ as a combinatorial rectangle where

$$X(\mathbf{r}) = \{\mathbf{x} \mid \forall k \in [z], p_{\tau_k}(\mathbf{x}) = r_k\}$$

and

$$Y(\mathbf{r}) = \{\mathbf{y} \mid \forall k \in [z], t_{\tau_k} \le r_k + p^y_{\tau_k}(\mathbf{y})\}.$$

Clearly, all the rectangles are disjoint, and hence $M[T]$ can be expressed as a sum of rank-1 matrices given by $R(\mathbf{r})$'s taken over all the $\mathbf{r}$'s. Thus Fact 1(i) implies that its rank is at most $|R_1 \times R_2 \times \cdots \times R_z| \le (2nw + 1)^z$. ◁

For each $l \in [d]$, based on P_l in $\mathbf{P}$, we define a set Q_l of threshold functions as

$$Q_l = \{\tau[g, \mathbf{P}] \mid g \in P_l\} \subseteq T_l$$

and a family $\mathcal{T}(Q_l)$ of sets T of threshold functions as

$$\mathcal{T}(Q_l) = \{T \subseteq T_l \mid Q_l \subseteq T \text{ and } |T| \le e - 1\}.$$

Following the inclusion-exclusion principle, we define a $2^n \times 2^n$ matrix

$$H[Q_l] = \sum_{T \in \mathcal{T}(Q_l)} (-1)^{|T| - |Q_l|} M[T].$$

We can show that $M_{\mathbf{P}}$ is expressed as the Hadamard product of $H[Q_1], H[Q_2], \ldots, H[Q_d]$:

▷ Claim 8. $M_{\mathbf{P}} = H[Q_1] \circ H[Q_2] \circ \cdots \circ H[Q_d]$.

Proof. Consider an arbitrary fixed assignment $(\mathbf{a}, \mathbf{b}) \in \{0,1\}^{2n}$. We show that

$$H[Q_1](\mathbf{a}, \mathbf{b}) \circ H[Q_2](\mathbf{a}, \mathbf{b}) \circ \cdots \circ H[Q_d](\mathbf{a}, \mathbf{b}) = 0,$$

if $M_{\mathbf{P}}(\mathbf{a}, \mathbf{b}) = 0$, and

$$H[Q_1](\mathbf{a}, \mathbf{b}) \circ H[Q_2](\mathbf{a}, \mathbf{b}) \circ \cdots \circ H[Q_d](\mathbf{a}, \mathbf{b}) = 1,$$

if $M_{\mathbf{P}}(\mathbf{a}, \mathbf{b}) = 1$. We write $\mathbf{P}^* = (P^*_1, P^*_2, \ldots, P^*_d)$ to denote $\mathbf{P}^*(\mathbf{a}, \mathbf{b})$ for a simpler notation.

Suppose $M_{\mathbf{P}}(\mathbf{a}, \mathbf{b}) = 0$. In this case, we have $\mathbf{P} \ne \mathbf{P}^*$, and hence there exists a level $l \in [d]$ such that $P_l \ne P^*_l$ while $P_{l'} = P^*_{l'}$ for every $l' \in [l-1]$. For such l, it holds that

$$\tau[g, \mathbf{P}^*](\mathbf{a}, \mathbf{b}) = \tau[g, \mathbf{P}](\mathbf{a}, \mathbf{b}) \tag{5}$$

for every $g \in G_l$. We show that $H[Q_l](\mathbf{a}, \mathbf{b}) = 0$ by considering two cases: $P_l \backslash P^*_l \ne \emptyset$ and $P_l \subset P^*_l$.

Consider the case where $P_l \backslash P^*_l \ne \emptyset$, then there exists $g \in P_l \backslash P^*_l$. Since $g \notin P^*_l$, we have $\tau[g, \mathbf{P}^*](\mathbf{a}, \mathbf{b}) = 0$. Thus, Eq. (5) implies that $\tau[g, \mathbf{P}](\mathbf{a}, \mathbf{b}) = 0$, and hence $M[T](\mathbf{a}, \mathbf{b}) = 0$ for every T such that $Q_l \subseteq T$. Therefore, for every $T \in \mathcal{T}(Q_l)$, we have $M[T](\mathbf{a}, \mathbf{b}) = 0$, and hence

$$H[Q_l](\mathbf{a}, \mathbf{b}) = \sum_{T \in \mathcal{T}(Q_l)} M[T](\mathbf{a}, \mathbf{b}) = 0.$$

Consider the other case where $P_l \subset P_l^*$. Let $Q_l^* = \{\tau[g, \mathbf{P}^*] \mid g \in P_l^*\}$. Equation (5) implies that $M[T](\mathbf{a}, \mathbf{b}) = 1$ if T satisfies $Q_l \subseteq T \subseteq Q_l^*$, and $M[T](\mathbf{a}, \mathbf{b}) = 0$, otherwise. Thus,

$$H[Q_l](\mathbf{a}, \mathbf{b}) = \sum_{T \in \mathcal{T}(Q_l)} (-1)^{|T|-|Q_l|} M[T] = \sum_{Q_l \subseteq T \subseteq Q_l^*} (-1)^{|T|-|Q_l|}$$

Therefore, by the binomial theorem,

$$H[Q_l](\mathbf{a}, \mathbf{b}) = \sum_{k=0}^{|Q_l^*|-|Q_l|} \binom{|Q_l^*| - |Q_l|}{k} (-1)^k = (1-1)^{|Q_l^*|-|Q_l|} = 0.$$

Suppose $M_{\mathbf{P}}(\mathbf{a}, \mathbf{b}) = 1$. In this case, we have $\mathbf{P} = \mathbf{P}^*$. Thus, for every $l \in [d]$, Eq. (5) implies that $M[T](\mathbf{a}, \mathbf{b}) = 1$ if $T = Q_l$, and $M[T](\mathbf{a}, \mathbf{b}) = 0$, otherwise. Therefore,

$$H[Q_l](\mathbf{a}, \mathbf{b}) = \sum_{T \in \mathcal{T}(Q_l)} (-1)^{|T|-|Q_l|} M[T](\mathbf{a}, \mathbf{b}) = (-1)^{|Q_l|-|Q_l|} = 1.$$

Consequently, $H[Q_1](\mathbf{a}, \mathbf{b}) \circ H[Q_2](\mathbf{a}, \mathbf{b}) \circ \cdots \circ H[Q_d](\mathbf{a}, \mathbf{b}) = 1$, as desired. ◁

We finally evaluate $rk(M_{\mathbf{P}})$. Claim 8 and Fact 1(ii) imply that

$$rk(M_{\mathbf{P}}) = rk\left(H[Q_1] \circ H[Q_2] \circ \cdots \circ H[Q_d]\right) \leq \prod_{l=1}^{d} rk(H[Q_l]). \tag{6}$$

Since

$$|\mathcal{T}(Q_l)| \leq \left(\frac{c \cdot s}{e-1}\right)^{e-1}$$

Fact 1(i) and Claim 7 imply that

$$\begin{aligned} rk(H[Q_l]) &\leq \sum_{T \in \mathcal{T}(Q_l)} rk(M[T]) \\ &\leq \left(\frac{c \cdot s}{e-1}\right)^{e-1} \cdot (2nw+1)^{e-1} \end{aligned} \tag{7}$$

for every $l \in [d-1]$, and

$$rk(H[Q_d]) \leq 2nw + 1. \tag{8}$$

Equations (6)-(8) imply that

$$rk(M_{\mathbf{P}}) \leq \left(\left(\frac{c \cdot s}{e-1}\right)^{e-1} \cdot (2nw+1)^{e-1} \right)^{d-1} \cdot (2nw+1)$$

as desired. We thus have verified Eq. (3). ◀

Combining Corollary 5 and Theorem 6, we obtain the following corollary:

▶ **Corollary 9.** *Let s, d, e and w be integers satisfying $2 \leq s, d$, $10 \leq e$, $1 \leq w$. Suppose a threshold circuit C of size s, depth d, energy e, and weight w computes CD_n. Then it holds that*

$$n \leq ed(\log s + \log w + \log n).$$

Equivalently, we have $2^{n/(ed)}/(nw) \leq s$.

Theorem 6 implies lower bounds for other Boolean functions with linear rank. For example, consider another Boolean function EQ_n asking if $\mathbf{x} = \mathbf{y}$:

$$\mathrm{EQ}_n(\mathbf{x}, \mathbf{y}) = \bigwedge_{i=1}^{n} \overline{x_i \oplus y_i}$$

Since M_{EQ_n} is the identity matrix with full rank, we have the same lower bound.

▶ **Corollary 10.** *Let s, d, e and w be integers satisfying $2 \le s, d$, $10 \le e$, $1 \le w$. Suppose a threshold circuit C of size s, depth d, energy e, and weight w computes EQ_n. Then it holds that*

$$n \le ed(\log s + \log w + \log n).$$

Equivalently, we have $2^{n/(ed)}/(nw) \le s$.

4 Tightness of the Lower Bound

In this section, we show that the lower bound given in Theorem 6 is tight if the depth and energy are small.

4.1 Definitions

Let z be a positive integer, and f be a Boolean function of $2n$ variables. We say that f is z-piecewise with $f_1, f_2, \dots, f_z$ if the following conditions are satisfied: Let

$$B_j = \{i \in [n] \mid x_i \text{ or } y_i \text{ are fed into } f_j\},$$

then

(i) $B_1, B_2, \dots, B_z$ compose a partition of $[n]$;
(ii) $|B_j| \le \lceil n/z \rceil$ for every $j \in [z]$;
(iii)

$$f(\mathbf{x}, \mathbf{y}) = \bigvee_{j=1}^{z} f_j(\mathbf{x}, \mathbf{y}) \quad \text{or} \quad f(\mathbf{x}, \mathbf{y}) = \overline{\bigvee_{j=1}^{z} f_j(\mathbf{x}, \mathbf{y})}.$$

We say that a set of threshold gates sharing input variables is a neural set, and a neural set is selective if at most one of the gates in the set outputs one for any input assignment. A selective neural set S computes a Boolean function f if for every assignment in $f^{-1}(0)$, no gates in S outputs one, while for every assignment in $f^{-1}(1)$, exactly one gate in S outputs one. We define the size and weight of S as $|S|$ and $\max_{g \in S} w_g$, respectively.

By a DNF-like construction, we can obtain a selective neural set of exponential size that computes f for any Boolean function f.

▶ **Theorem 11.** *For any Boolean function f of n variables, there exists a selective neural set of size 2^n and weight one that computes f.*

4.2 Upper Bounds

The following proposition shows that we can construct threshold circuits of small energy for piecewise functions.

▶ **Lemma 12.** *Let e and d be integers satisfying $2 \leq e$ and $2 \leq d$, and z be an integer. Suppose $f : \{0,1\}^{2n} \to \{0,1\}$ is a z-piecewise function with $f_1, f_2, \ldots, f_z$. If f_j is computable by a selective neural set of size at most s' and weight w' for every $j \in [z]$, f is computable by a threshold circuit of size*

$$s \leq z \cdot s' + 1,$$

depth d, energy e and weight

$$w \leq \frac{2n}{z} \cdot w'.$$

Clearly, CD_n is a piecewise function, and so the lemma gives our upper bound for CD_n.

▶ **Theorem 13** (Theorem 2 restated)**.** *For any integers e and d such that $2 \leq e$ and $2 \leq d$, CD_n is computable by a threshold circuit of size*

$$s \leq (e-1)(d-1) \cdot 2^{\frac{n}{(e-1)(d-1)}}.$$

depth d, energy e and weight

$$w \leq \left(\frac{n}{(e-1)(d-1)}\right)^2.$$

We can also obtain a similar proposition for EQ_n.

▶ **Theorem 14.** *For any integers e and d such that $2 \leq e$ and $2 \leq d$, EQ_n is computable by a threshold circuit of size*

$$s \leq (e-1)(d-1) \cdot 2^{\frac{2n}{(e-1)(d-1)}}.$$

depth d, energy e and weight

$$w \leq \frac{n}{(e-1)(d-1)}.$$

5 Simulating Discretized Circuits

In this section, we show that any discretized circuit can be simulated using a threshold circuit with a moderate increase in size, depth, energy, and weight. Thus, a similar inequality holds for discretized circuits, as follows.

▶ **Theorem 15.** *Let δ be a discretizer and φ be an activation function such that $\delta \circ \varphi$ has a silent range. If a $(\delta \circ \varphi)$-circuit C of size s, depth d, energy e, and weight w computes a Boolean function f, then it holds that*

$$\log(rk(M_C)) = O(ed(\log s + \log w + \log n)^3).$$

We prove the theorem by showing that, given a $(\delta \circ \varphi)$-circuit C, we can safely replace any $(\delta \circ \varphi)$-gate g in C by a set of threshold gates that simulate g. Our simulation is based on a binary search of the potentials of a discretized gate, and employ a conversion technique from a linear decision tree to a threshold circuit given in [42]. We omit our proof of the theorem due to the page limitation.

6 Conclusions

In this paper, we prove that a threshold circuit is able to compute only a Boolean function of which communication matrix has rank bounded by a product of logarithmic factors of size and weight, and linear factors of depth and energy. This bound implies that any threshold circuit of sub-linear depth, sub-linear energy and sub-exponential weight needs exponential size to compute CD_n, EQ_n, and the Inner-Product function. We show that the bounds are tight up to a constant factor. We also prove that a similar bound holds for discretized circuits. Thus, increasing depth could help a neural network to acquire sparse activity. This observation may shed some light on the reason for the success of deep learning.

References

1 Kazuyuki Amano. On the size of depth-two threshold circuits for the Inner Product mod 2 function. In Alberto Leporati, Carlos Martín-Vide, Dana Shapira, and Claudio Zandron, editors, *Language and Automata Theory and Applications*, pages 235–247, Cham, 2020. Springer International Publishing.

2 Kazuyuki Amano and Akira 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.

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

4 Matthieu Courbariaux, Yoshua Bengio, and Jean-Pierre David. Binaryconnect: Training deep neural networks with binary weights during propagations. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'15, pages 3123–3131, Cambridge, MA, USA, 2015. MIT Press.

5 James J. DiCarlo and David D. Cox. Untangling invariant object recognition. *Trends in Cognitive Sciences*, 11(8):333–341, 2007. doi:10.1016/j.tics.2007.06.010.

6 Krishnamoorthy Dinesh, Samir Otiv, and Jayalal Sarma. New bounds for energy complexity of boolean functions. *Theoretical Computer Science*, 845:59–75, 2020.

7 Peter 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, 2003.

8 Jürgen Forster, Matthias Krause, Satyanarayana V. Lokam, Rustam Mubarakzjanov, Niels Schmitt, and Hans Ulrich 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.

9 András Hajnal, Wolfgang Maass, Pavel Pudlák, Márió Szegedy, and György Turán. Threshold circuits of bounded depth. *Journal of Computer and System Sciences*, 46:129–154, 1993.

10 Johan Håstad and Mikael Goldmann. On the power of small-depth threshold circuits. *Computational Complexity*, 1(2):113–129, 1991.

11 Yunlong He, Koray Kavukcuoglu, Yun Wang, Arthur Szlam, and Yanjun Qi. Unsupervised feature learning by deep sparse coding. In *Proc. of SIAM International Conference on Data Mining*, pages 902–910, 2014.

12 Itay Hubara, Matthieu Courbariaux, Daniel Soudry, Ran El-Yaniv, and Yoshua Bengio. Quantized neural networks: Training neural networks with low precision weights and activations. *Journal of Machine Learning Research*, 18(187):1–30, 2018. URL: http://jmlr.org/papers/v18/16-456.html.

13 Russell Impagliazzo, Ramamohan Paturi, and Michael E. Saks. Size-depth tradeoffs for threshold circuits. *SIAM Journal on Computing*, 26(3):693–707, 1997.

14 Stasys Jukna. *Extremal Combinatorics with Applications in Computer Science*. Springer-Verlag Berlin Heidelberg, 2011.

15 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.

16 O. M. Kasim-zade. On a measure of active circuits of functional elements (russian). *Mathematical problems in cybernetics "Nauka"*, 4:218?–228, 1992.

17 Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y. Ng. Efficient sparse coding algorithms. In *Proc. of the 19th Advances in Neural Information Processing Systems*, pages 801–808, 2006.

18 Peter Lennie. The cost of cortical computation. *Current Biology*, 13:493–497, 2003.

19 Nancy Lynch and Cameron Musco. *A Basic Compositional Model for Spiking Neural Networks*, pages 403–449. Springer Nature Switzerland, 2022.

20 Wolffanf Maass, Georg. Schnitger, and Eduardo D. Sontag. On the computational power of sigmoid versus boolean threshold circuits. In *Proc. of 32nd Annual Symposium of Foundations of Computer Science*, pages 767–776, 1991. doi:10.1109/SFCS.1991.185447.

21 Wolfgang Maass. Networks of spiking neurons: The third generation of neural network models. *Neural Networks*, 10(9):1659–1671, 1997.

22 Hiroki Maniwa, Takayuki Oki, Akira Suzuki, Kei, and Xiao Zhou. Computational power of threshold circuits of energy at most two. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E101.A(9):1431–1439, 2018.

23 Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133, 1943.

24 Marvin Minsky and Seymour Papert. *Perceptrons: An Introduction to Computational Geometry.* MIT Press, 1988.

25 Andrew Ng. Sparse autoencoder. CS294A Lecture notes, 2011.

26 Noam Nisan. The communication complexity of threshold gates. *Proceeding of "Combinatorics, Paul Erdös is Eighty"*, pages 301–315, 1993.

27 Bruno A Olshausen and David J Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4):481–487, 2004.

28 Ian Parberry. *Circuit Complexity and Neural Networks.* MIT Press, 1994.

29 Thomas Pfeil, Tobias Potjans, Sven Schrader, Wiebke Potjans, Johannes Schemmel, Markus Diesmann, and Karlheinz Meier. Is a 4-bit synaptic weight resolution enough? - constraints on enabling spike-timing dependent plasticity in neuromorphic hardware. *Frontiers in Neuroscience*, 6:90, 2012. doi:10.3389/fnins.2012.00090.

30 Alexander A. Razborov and Alexander A. Sherstov. The sign-rank of AC^0. *SIAM Journal on Computing*, 39(5):1833–1855, 2010.

31 Frank Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

32 Alexander A. Sherstov. Separating AC^0 from depth-2 majority circuits. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, STOC '07, page 294?301, New York, NY, USA, 2007. Association for Computing Machinery. doi:10.1145/1250790.1250834.

33 Janio Carlos Nascimento Silva and Uéverton S. Souza. Computing the best-case energy complexity of satisfying assignments in monotone circuits. *Theoretical Computer Science*, 932:41–55, 2022.

34 K. Y. Siu and J. Bruck. On the power of threshold circuits with small weights. *SIAM Journal on Discrete Mathematics*, 4(3):423–435, 1991.

35 Kai-Yeung. Siu, Vwani Roychowdhury, and Thomas Kailath. *Discrete Neural Computation: A Theoretical Foundation.* Prentice Hall, 1995.

36 Kai-Yeung Siu and Vwani P. Roychowdhury. On optimal depth threshold circuits for multiplication and related problems. *SIAM Journal on Discrete Mathematics*, 7(2):284–292, 1994.

37 Xiaoming Sun, Yuan Sun, Kewen Wu, and Zhiyu 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.

38 Akira Suzuki, Kei Uchizawa, and Xiao Zhou. Energy-efficient threshold circuits computing MOD functions. In *Proc. of the 17th Computing: the Australasian Theory Symposium*, pages 105–110, 2011.

39 Akira Suzuki, Kei Uchizawa, and Xiao Zhou. Energy-efficient threshold circuits computing MOD functions. *International Journal of Foundations of Computer Science*, 24(1):15–29, 2013.

40 Kei Uchizawa. Lower bounds for threshold circuits of bounded energy. *Interdisciplinary Information Sciences*, 20(1):27–50, 2014.

41 Kei Uchizawa. Size, Depth and Energy of Threshold Circuits Computing Parity Function. In Yixin Cao, Siu-Wing Cheng, and Minming Li, editors, *31st International Symposium on Algorithms and Computation (ISAAC 2020)*, volume 181 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 54:1–54:13, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ISAAC.2020.54.

42 Kei Uchizawa, Rodney Douglas, and Wolfgang Maass. On the computational power of threshold circuits with sparse activity. *Neural Computation*, 18(12):2994–3008, 2008.

43 Kei Uchizawa and Eiji 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.

44 Kei Uchizawa, Eiji Takimoto, and Takao Nishizeki. Size-energy tradeoffs of unate circuits computing symmetric Boolean functions. *Theoretical Computer Science*, 412:773–782, 2011.

45 M. N. Vaintsvaig. On the power of networks of functional elements (Russian). *Doklady Akademii Nauk*, 139(2):320?323, 1961.