The BNS-Chung Criterion for Multi-Party Communication Complexity

Ran Raz
ranraz@wisdom.weizmann.ac.il
Department of Applied Mathematics
Weizmann Institute
Rehovot 76100, ISRAEL

Abstract

The "Number on the Forehead" model of multi-party communication complexity was first suggested by Chandra, Furst and Lipton. The best known lower bound, for an explicit function (in this model), is a lower bound of $\Omega(n/2^k)$, where n is the size of the input of each player, and k is the number of players (first proved by Babai, Nisan and Szegedy). This lower bound has many applications in complexity theory. Proving a better lower bound, for an explicit function, is a major open problem. Based on the result of BNS, Chung gave a sufficient criterion for a function to have large multi-party-communication-complexity (up to $\Omega(n/2^k)$). In this paper, we use some of the ideas of BNS, and Chung, together with some new ideas, resulting in a new (easier and more modular) proof for the results of BNS and Chung. This gives a simpler way to prove lower bounds for the multi-party-communication-complexity of a function.

1 Multi-Party Communication Complexity

Multi-party communication complexity was first introduced by Chandra, Furst and Lipton [CFL], as a generalization of Yao's standard 2-parties communication model [Yao1]. In the k-parties model, we have k finite sets $X_1, \ldots, X_k$, and a function $f : X_1 \times \cdots \times X_k \to \{-1, 1\}$. We assume w.l.o.g. that $X_1 = \cdots = X_k = \{0,1\}^n$. We have k players of unlimited computational power, who wish to compute the value of f on the input $(x_1, \ldots, x_k) \in X_1 \times \cdots \times X_k$. The problem is that each one of the k players can see only $k-1$ of the input variables $x_1, \ldots, x_k$. Player i can see all the input variables except x_i (considered to be "on his forehead"). Moreover, initially Player i has no information about x_i. The players share a blackboard were they can exchange messages according to a protocol. The blackboard is viewed by all players. In each step, one player writes one bit on the blackboard (i.e. sends one bit of information about the input variables that he can see, to all the other players). In the end, they all have to know the value of $f(x_1, \ldots, x_k)$.

A *strategy* for the k players describes (in each step): 1) A way for the players to agree on a player that will speak in this step (this decision is based only on the messages already transmitted - so it is a function from the set of all possible messages to the set $\{1, \ldots, k\}$). 2) A way for this player to send one bit of information (based on the input variables that he can see, and on all the messages already transmitted).

A *protocol* for f is a strategy for the k players, such that after the last step they all know the value of $f(x_1, \ldots, x_k)$ (w.l.o.g. assume that the last symbol written on the board is the value $f(x_1, \ldots, x_k)$). The maximum number of steps in the protocol is called the *communication complexity of the protocol* (where the maximum is taken over all the possible inputs). The *deterministic k-parties communication complexity* of f, $C(f)$, is the minimal communication complexity of a deterministic protocol for f. The probabilistic case differs from the deterministic case in allowing the protocol to depend on flips of a coin (so the functions of the strategy may be probabilistic), and in allowing the protocol to make mistakes (in computing the value of $f(x_1, \ldots, x_k)$). We say that the probability of error of a probabilistic protocol is δ if for every input $f(x_1, \ldots, x_k)$, the protocol makes a mistake with probability at most δ. Denote by $C_\delta(f)$ the probabilistic k-parties communication complexity of f, with probability of error δ. We use here the public coin model of probabilistic communication complexity, i.e., the players share a common random string.

Multi-party communication complexity was studied in several papers, and was found to be relevant for many other complexity issues (such as, circuit complexity and derandomization, see for example [BNS, NW]). A major open problem is to prove a super-poly-logarithmic lower bound, for the k-parties communication complexity of an explicit function f, where the number of parties, k, is super-poly-logarithmic. Due to the results of [HG, Yao2], this will prove that the function f is not in the circuit-complexity-class ACC, which is a major open problem in circuit complexity. These results and others (e.g. [BNS, NW]) motivate the research for new lower bounds in the area, as well as understanding better the existing ones. For an excellent survey of communication complexity, and multi-party communication complexity see [KN].

The best known lower bound for the k-parties communication complexity of an explicit function f, due to [BNS], gives a non-trivial lower bound for k up to $\log n$. In this paper, we try to supply a better understanding of this lower bound.

2 Cylinder Intersections

Denote the space $X_1 \times \cdots \times X_k$ by $\bar{X}$. Let μ be the uniform probability distribution over $\bar{X}$. A function g, defined on $\bar{X}$, is called **cylindrical** in the i-th dimension iff it doesn't depend on x_i. A subset $T \subset \bar{X}$ is called a **cylinder** in the i-th dimension iff its characteristic function is cylindrical in the i-th dimension. That is, a subset T is a cylinder iff for every $(x_1, ..., x_i, ..., x_k) \in T$ and every $x'_i \in X_i$, we have $(x_1, ..., x'_i, ..., x_k) \in T$. A subset $T \subset \bar{X}$ is called a **cylinder intersection** iff it can be represented as $T = \bigcap_{i=1}^{k} T_i$, where T_i is a cylinder in the i-th dimension. The **discrepancy** of f on the cylinder intersection T, $\mathrm{Disc}_T(f)$, is defined by

$$\mathrm{Disc}_T(f) = \left|\mu\{\vec{x} : \vec{x} \in T, f(\vec{x}) = 1\} - \mu\{\vec{x} : \vec{x} \in T, f(\vec{x}) = -1\}\right|.$$

The discrepancy of f is defined by

$$\mathrm{Disc}(f) = \max_T \mathrm{Disc}_T(f),$$

where the maximum is taken over all cylinder intersections T.

Many known lower bounds for the probabilistic communication complexity of functions are proved by proving an upper bound for Disc(f). The connection between Disc(f), and the communication complexity of f is given by the following lemma:

Lemma 2.1 $\forall f : \bar{X} \to \{-1, 1\}$

$$C_{\frac{1}{2}-\delta}(f) = \Omega\left(\log_2(\delta/Disc(f))\right).$$

Proof (sketch):
The proof for the lemma is by a standard argument. Since it was used and proved many times before, and since it is not hard, we will give here just a short sketch that gives the main idea. For details see [BNS, KN].

Let P be the best probabilistic communication protocol for f. Let c be the maximal number of steps in P. Let r be the random string of P. Fix the random string r. The important fact is that given the first d bits communicated by the players, the $(d+1)^{th}$ bit of communication, (transmitted by one of the players), must be defined (as a function of the input) by a cylindrical function (because this bit doesn't depend on (at least) one of the coordinates of the input, $(x_1, \ldots, x_k)$, as the player doesn't see this coordinate). Therefore, it is not too hard to observe that the protocol P defines (in a natural way) a partition of the set of all possible inputs, into at most 2^c cylinder intersections. On every two inputs contained in the same cylinder intersection (of this partition), the observed behavior of the protocol is the same (i.e. an observer who cannot see the inputs, but can see all the messages transmitted, and can see which player sends which message will not be able to distinguish between different inputs in the same cylinder intersection of the partition). Also, on every two inputs contained in the same cylinder intersection, the protocol gives the same output. So, for each one of these cylinder intersections the output is constant. On average (on r), this output should be equal to the value of f, with probability of at least $1/2 + \delta$.

Take an average r, and an average cylinder intersection T (in the corresponding partition). $\mu(T)$ is around 2^{-c}, and since the output of the protocol on T is constant, and agrees with f with probability of around $1/2+\delta$, the discrepancy, $\mathrm{Disc}_T(f)$, is around $2\delta 2^{-c}$. Since $\mathrm{Disc}_T(f)$ is (by definition) smaller than Disc(f), the claims follows. □

The previous lemma is very useful. However, it doesn't give immediate lower bounds for the k-parties communication complexity, because for $k > 2$, Disc(f) is very hard to compute, or to estimate, or even to bound from above. Bounding Disc(f) is the important part of any lower bound proof using the previous lemma.

3 The BNS-Chung Criterion

A **cube** D is defined to be a multi-set $D = \{a_1, b_1\} \times \{a_2, b_2\} \times \cdots \times \{a_k, b_k\}$, where for all i : $a_i, b_i \in X_i$ (not necessarily distinct). Denote by $\mathcal{D}$ the family of all cubes D. We choose $D \in \mathcal{D}$ at random according to the uniform distribution. This can be done by choosing $\forall i : a_i, b_i \in X_i$ according to the uniform distribution. For the cube $D = \{a_1, b_1\} \times \{a_2, b_2\} \times \cdots \times \{a_k, b_k\}$, define its **sign**, $S_f(D)$, to be

$$S_f(D) = \prod_{d_1 \in \{a_1,b_1\}} \prod_{d_2 \in \{a_2,b_2\}} \cdots \prod_{d_k \in \{a_k,b_k\}} f(d_1, \ldots, d_k).$$

Clearly, the function $S_f : \mathcal{D} \to \{-1, 1\}$ gives 1 iff the number of elements in D on which f evaluates to 1 is even. Define

$$\mathcal{E}(f) = \mathbf{E}_D \left[S_f(D) \right],$$

where $\mathbf{E}_D$ denotes the expectation over the random variable D.

The following two theorems were first proved in [Chu]. The theorems are a generalization of the lower bound proved in [BNS], and the proofs are closely related.

Theorem 3.1 $\forall f : \bar{X} \to \{-1, 1\}$

$$C_{\frac{1}{2}-\delta}(f) = \Omega\left(2^{-k} \log_2(1/\mathcal{E}(f)) + \log_2 \delta\right).$$

Proof:
The proof follows as a conclusion of Lemma 2.1, and Theorem 3.2 stated below. □

The theorem is very useful because $\mathcal{E}$ is a much simpler object than the probabilistic (or deterministic) communication complexity. For many functions f, it is very easy to compute $\mathcal{E}(f)$ exactly. In [CT], $\mathcal{E}(f)$ was computed for some explicit functions, resulting in lower bounds of the type $\Omega(n/c^k)$ for the multi-party communication complexity of these functions. Note that these bounds are the best possible by Theorem 3.1, and that always for $k > 1$, $\mathcal{E}(f) \neq 0$. It is very interesting that for $k \leq \log n$, $\mathcal{E}(f)$ is *natural* in the sense of [RR]. That is, $\mathcal{E}(f)$ can be computed in a polynomial time in the number of possible inputs for f (i.e., exponential time in the length of the input), and Theorem 3.1 gives a lower bound for a random function.

Theorem 3.2 $\forall f : \bar{X} \to \{-1, 1\}$

$$\mathcal{E}(f) \geq Disc(f)^{2^k}.$$

This is the main theorem. In this paper we give a new proof for this theorem.

4 $\mathcal{E}(h)$, and $\Delta(h)$

We will first prove a lemma that gives a lower bound for $\mathcal{E}(h)$, for any function $h : \bar{X} \to \{-1, 1\}$. This lemma doesn't talk about discrepancy at all. The connection to discrepancy (and therefore to communication complexity) is given only in the next section.

Define

$$\Delta(h) = \mathbf{E}_{\vec{x} \in \bar{X}}[h(\vec{x})],$$

where $\mathbf{E}$ denotes the expectation over the uniform distribution μ.

Lemma 4.1 $\forall h : \bar{X} \to \{-1, 1\}$

$$\mathcal{E}(h) \geq |\Delta(h)|^{2^k}.$$

Proof:

$$\mathcal{E}(h) = \mathbf{E}_{\mathcal{D}}\left[S_h(D)\right] = \mathbf{E}_{a_1,b_1 \in X_1} \cdots \mathbf{E}_{a_k,b_k \in X_k} \prod_{d_1 \in \{a_1,b_1\}} \cdots \prod_{d_k \in \{a_k,b_k\}} h(d_1, \ldots, d_k).$$

For any function g, defined on X_k, we have

$$\mathbf{E}_{a_k,b_k \in X_k}[g(a_k)g(b_k)] = (\mathbf{E}_{a_k \in X_k}[g(a_k)]) \cdot (\mathbf{E}_{b_k \in X_k}[g(b_k)]) = (\mathbf{E}_{x_k \in X_k}[g(x_k)])^2.$$

For fixed $a_1, b_1, \ldots, a_{k-1}, b_{k-1}$, take

$$g(x_k) = \prod_{d_1 \in \{a_1,b_1\}} \cdots \prod_{d_{k-1} \in \{a_{k-1},b_{k-1}\}} h(d_1, \ldots, d_{k-1}, x_k)$$

to get

$$\mathcal{E}(h) = \mathbf{E}_{a_1,b_1 \in X_1} \cdots \mathbf{E}_{a_{k-1},b_{k-1} \in X_{k-1}} \left(\mathbf{E}_{x_k \in X_k} \prod_{d_1 \in \{a_1,b_1\}} \cdots \prod_{d_{k-1} \in \{a_{k-1},b_{k-1}\}} h(d_1, \ldots, d_{k-1}, x_k) \right)^2.$$

By the Cauchy-Schwartz inequality, for any random variable z, $\mathbf{E}[z^2] \geq (\mathbf{E}z)^2$. Therefore

$$\mathcal{E}(h) \geq \left(\mathbf{E}_{a_1,b_1 \in X_1} \cdots \mathbf{E}_{a_{k-1},b_{k-1} \in X_{k-1}} \mathbf{E}_{x_k \in X_k} \prod_{d_1 \in \{a_1,b_1\}} \cdots \prod_{d_{k-1} \in \{a_{k-1},b_{k-1}\}} h(d_1, \ldots, d_{k-1}, x_k) \right)^2$$

$$= \left(\mathbf{E}_{x_k \in X_k} \mathbf{E}_{a_1,b_1 \in X_1} \cdots \mathbf{E}_{a_{k-1},b_{k-1} \in X_{k-1}} \prod_{d_1 \in \{a_1,b_1\}} \cdots \prod_{d_{k-1} \in \{a_{k-1},b_{k-1}\}} h(d_1, \ldots, d_{k-1}, x_k) \right)^2.$$

Repeat the same argument k times to get

$$\mathcal{E}(h) \geq (\mathbf{E}_{x_k \in X_k} \cdots \mathbf{E}_{x_1 \in X_1} h(x_1, \ldots, x_k))^{2^k} = |\Delta(h)|^{2^k}.$$

□

5 The Connection to Discrepancy and Communication Complexity

In this section, we will prove that for every $f : \bar{X} \to \{-1,1\}$ there exists $h : \bar{X} \to \{-1,1\}$, with $\mathcal{E}(h) = \mathcal{E}(f)$, and $|\Delta(h)| \geq \mathrm{Disc}(f)$. This will connect $\mathcal{E}(f)$ to the discrepancy of f (and therefore also to the communication complexity of f). Theorem 3.2 will then follow. First we need the following two claims:

Claim 5.1 *$\forall f, g : \bar{X} \to \{-1,1\}$, if g is cylindrical then*

$$\mathcal{E}(fg) = \mathcal{E}(f),$$

(where fg is the product of f and g, i.e $fg(x) = f(x)g(x)$).

Proof:
By the definition of the sign function, for every cube D: $S_{fg}(D) = S_f(D)S_g(D)$. W.l.o.g. assume that g doesn't depend on x_k. Then $\forall d_1, \ldots, d_{k-1}, a_k, b_k$ we have $g(d_1, \ldots, d_{k-1}, a_k) = g(d_1, \ldots, d_{k-1}, b_k)$. Therefore, g gets the value 1 on an even number of elements of D. Therefore $S_g(D) = 1$, and we get $S_{fg}(D) = S_f(D)$. Since this is true for every cube D, we get $\mathcal{E}(fg) = \mathcal{E}(f)$. □

Claim 5.2 *$\forall f : \bar{X} \to \{-1,1\}$, and $\forall g_1, \ldots, g_k : \bar{X} \to \{-1,1\}$, if $g_1, \ldots, g_k$ are cylindrical then*

$$\mathcal{E}(fg_1g_2 \cdots g_k) = \mathcal{E}(f).$$

Proof:
Immediate from the previous claim. □

Lemma 5.1 *$\forall f : \bar{X} \to \{-1,1\}$ there exists $h : \bar{X} \to \{-1,1\}$, with $\mathcal{E}(h) = \mathcal{E}(f)$, and $|\Delta(h)| \geq \mathit{Disc}(f)$.*

Proof:
Take a cylinder intersection $T = \bigcap_{i=1}^{k} T_i$, with $\mathrm{Disc}_T(f) = \mathrm{Disc}(f)$. For every $1 \leq i \leq k$, define $g_i : \bar{X} \to \{-1,1\}$, as a random variable, in the following way: with probability 1/2, g_i is the constant function 1, and with probability 1/2, $g_i(\vec{x})$ gives the value 1 on all the elements $\vec{x} \in T_i$, and -1 otherwise. Then for $\vec{x} \in T_i$, $g_i(\vec{x}) = 1$ with probability 1, and for $\vec{x} \notin T_i$, $g_i(\vec{x}) = 1$ with probability 1/2, and $g_i(\vec{x}) = -1$ with probability 1/2.

Define $g = g_1g_2 \cdots g_k$. Then for $\vec{x} \in T$, $g(\vec{x}) = 1$ with probability 1, and for $\vec{x} \notin T$, $g(\vec{x}) = 1$ with probability 1/2, and $g(\vec{x}) = -1$ with probability 1/2 (this is true because $\vec{x} \notin T$ iff for some i, $\vec{x} \notin T_i$, and because the functions g_i are independent of each other). Thus for $\vec{x} \in T$, $\mathbf{E}_g[g(\vec{x})] = 1$, and for $\vec{x} \notin T$, $\mathbf{E}_g[g(\vec{x})] = 0$. Thus $\mathbf{E}_g[g(\vec{x})]$ is the characteristic function of the set T.

The lemma will follow by taking $h = fg$. First, since $\mathbf{E}_g[g(\vec{x})]$ is the characteristic function of T, we have

$$|\mathbf{E}_g\Delta(fg)| = |\mathbf{E}_g\mathbf{E}_{\vec{x}}[f(\vec{x})g(\vec{x})]| = |\mathbf{E}_{\vec{x}}\mathbf{E}_g[f(\vec{x})g(\vec{x})]| = |\mathbf{E}_{\vec{x}}\left[f(\vec{x})\mathbf{E}_g[g(\vec{x})]\right]|$$

$$= |\mu\{\vec{x} : \vec{x} \in T, f(\vec{x}) = 1\} - \mu\{\vec{x} : \vec{x} \in T, f(\vec{x}) = -1\}| = \mathrm{Disc}_T(f),$$

and by convexity

$$\mathbf{E}_g |\Delta(fg)| \geq |\mathbf{E}_g \Delta(fg)| = \mathrm{Disc}_T(f) = \mathrm{Disc}(f).$$

At the other hand, since $g_1, \ldots, g_k$ are cylindrical, we have by the previous claim

$$\mathcal{E}(fg) = \mathcal{E}(f).$$

Therefore with non-zero probability $|\Delta(fg)| \geq \mathrm{Disc}(f)$, and $\mathcal{E}(fg) = \mathcal{E}(f)$. So, if we set $h = fg$ the lemma will hold for h. □

Proof of Theorem 3.2:
The theorem follows immediately as a consequence of the previous two lemmas. Take h from Lemma 5.1. Then by Lemma 4.1, and Lemma 5.1 we have

$$\mathcal{E}(f) = \mathcal{E}(h) \geq |\Delta(h)|^{2^k} \geq \mathrm{Disc}(f)^{2^k}.$$

□

6 Example: Matrix Multiplication

Some examples for lower bounds of the type $\Omega(n/c^k)$ on the multi-party communication complexity of explicit functions f were given in [BNS, Chu, CT]. Here we give a new example for a function that has not been analyzed before.

Let $X = \{0,1\}^{n \times n}$ be the set of all boolean matrices of size $n \times n$. Think of the elements of X as matrices over the field $GF[2]$. For $x_1 \in X_1, x_2 \in X_2, \ldots, x_k \in X_k$, denote by $x_1 \cdot x_2 \cdots x_k$ the product of $x_1, \ldots, x_k$, as matrices over the field $GF[2]$. Define

$$F(x_1, x_2, \ldots, x_k) = (x_1 \cdot x_2 \cdots x_k)_{1,1},$$

i.e. the element in the first line and the first column of the product $x_1 \cdot x_2 \cdots x_k$. Obviously F is linear in each one of its input variables. To be consistent with the previous notations (i.e. to get a function to the range $\{-1, 1\}$), define

$$f(x_1, x_2, \ldots, x_k) = -1^{F(x_1, x_2, \ldots, x_k)}.$$

Clearly, we can also write f as

$$f(x_1, x_2, \ldots, x_k) = 1 - 2F(x_1, x_2, \ldots, x_k).$$

For every cube $D = \{a_1, b_1\} \times \{a_2, b_2\} \times \cdots \times \{a_k, b_k\}$, the sign $S_f(D)$ satisfies

$$S_f(D) = \prod_{d_1 \in \{a_1, b_1\}} \prod_{d_2 \in \{a_2, b_2\}} \cdots \prod_{d_k \in \{a_k, b_k\}} f(d_1, \ldots, d_k)$$

$$= -1^{\bigoplus_{d_1 \in \{a_1,b_1\}} \bigoplus_{d_2 \in \{a_2,b_2\}} \cdots \bigoplus_{d_k \in \{a_k,b_k\}} F(d_1,\ldots,d_k)},$$

where $\oplus$ denotes addition over $GF[2]$. By the linearity of F in each of its input variables, we have

$$S_f(D) = -1^{F(a_1 \oplus b_1, a_2 \oplus b_2, \ldots, a_k \oplus b_k)} = 1 - 2F(a_1 \oplus b_1, a_2 \oplus b_2, \ldots, a_k \oplus b_k),$$

where $a_i \oplus b_i$ denotes the sum of these two matrices over $GF[2]$. If we choose D at random, according to the uniform distribution then $(a_1 \oplus b_1, a_2 \oplus b_2, \ldots, a_k \oplus b_k)$ is a random variable uniformly distributed over $X_1 \times \cdots \times X_k$. Therefore

$$\mathcal{E}(f) = \mathbf{E}_D[S_f(D)] = \mathbf{E}_D[1 - 2F(a_1 \oplus b_1, a_2 \oplus b_2, \ldots, a_k \oplus b_k)]$$

$$= \mathbf{E}_{\vec{x}}[1 - 2F(x_1, x_2, \ldots, x_k)] = \mathbf{E}_{\vec{x}}[f(x_1, x_2, \ldots, x_k)] = \Delta(f).$$

To estimate $\Delta(f)$, let $(x_1, x_2, \ldots, x_k)$ be a random variable, uniformly distributed over $X_1 \times \cdots \times X_k$. For $1 \le d \le k$, denote by E_d the following event: $\{$*the first line of the matrix* $x_1 \cdot x_2 \cdots x_d$ *is all* $0\}$. Denote $p_d = Prob(E_d)$. Since E_1 is determined by x_1 and since x_1 is uniformly distributed, we have $p_1 = Prob(E_1) = 2^{-n}$. Clearly we also have $Prob(E_{d+1}|E_d) = 1$. Also, since x_{d+1} is uniformly distributed $Prob(E_{d+1}|\neg E_d) = 2^{-n}$. Therefore, for $1 \le d < k$,

$$p_{d+1} = p_d + (1 - p_d)2^{-n} \le p_d + 2^{-n},$$

and by induction, for $1 \le d \le k$,

$$p_d \le d2^{-n}.$$

If E_{k-1} occurs then $F(x_1, x_2, \ldots, x_k)$ is always 0 (and therefore $f(x_1, x_2, \ldots, x_k)$ is always 1). Otherwise, since x_k is uniformly distributed, $F(x_1, x_2, \ldots, x_k)$ is uniformly distributed over $\{0, 1\}$ (and therefore $f(x_1, x_2, \ldots, x_k)$ is uniformly distributed over $\{-1, 1\}$). Therefore,

$$\mathcal{E}(f) = \Delta(f) = p_{k-1} \le (k-1)2^{-n}.$$

By Theorem 3.1, we get

$$C_{\frac{1}{4}}(f) = \Omega(n/2^k).$$

We believe that this lower bound is not tight, and that f is an example for a hard function even when k is much larger than $\log n$.

Acknowledgments

I would like to thank Avi Wigderson for helpful discussions.

References

[BNS] L. Babai, N. Nisan, and M. Szegedy, "Multiparty protocols and logspace-hard pseudorandom sequences," In *Proc. 21st Ann. ACM Symp. Theor. Comput.*, 1989, 1–11.

[CFL] A. K. Chandra, M. L. Furst, and R. J. Lipton, “Multy-Party Protocols”, In *Proc. 15th Ann. ACM Symp. Theor. Comput.*, 1983, 94–99.

[Chu] F. R. K. Chung, “Quasi-Random Classes of Hypergraphs”, *Random Structures and Algorithms*, Vol 1., No. 4, pp. 363–382, 1990.

[CT] F. R. K. Chung, and P. Tetali, “Communication Complexity and Quasi Randomness”, *SIAM J. Discrete Math.*, Vol 6., No 1., pp.110–123, 1993.

[HG] J. Håstad, and M. Goldmann, “On the power of small-depth threshold circuits”, *Comput. Complexity*, 1:113–129, 1991.

[KN] E. Kushilevitz, and N. Nisan, *Communication Complexity.* Cambridge University Press, to appear.

[NW] N. Nisan, and A. Wigderson, “Rounds in communication complexity revisited,” *SIAM J. of computing*, 22(1):211–219, 1993.

[RR] A. Razborov, S. Rudich, “Natural Proofs,” In *Proc. 26th Ann. ACM Symp. Theor. Comput.*, 1994, 204–213.

[Yao1] A. C. C. Yao, “Some Complexity Questions Related to Distributive Computing,” In *Proc. 11th Ann. ACM Symp. Theor. Comput.*, 1979, 209–213.

[Yao2] A. C. C. Yao, “On ACC and threshold circuits”, In *Proc. of the 31st FOCS*, pages 619–627. IEEE, 1990.