Communication Complexity and Quasi-randomness [1]

Fan R.K. Chung [2] and **Prasad Tetali**[3]

1 Introduction

Many problems arising in interactive and distributive computation share the general framework that a number of processors wish to collaboratively evaluate a Boolean function while each processor has only partial information. The question of interest is to determine the minimum amount of information transfer required under the assumption that each processor has unlimited computational power and the messages are transferred by a "blackboard", viewed by all processors.

One of the most interesting examples is the *round-table* model, proposed by Chandra, Furst and Lipton [CFL], involving k players each having a number X_i on his/her forehead; (so that the i-th player knows all numbers except for X_i). For $k = 3$, they proved a tight lower bound for the minimum number of bits to be exchanged to compute the sum of X_i's. For general k, the lower bounds were further improved by Babai, Nisan and Szegedy [BNS] who gave a lower bound of $\Omega(m2^{-k})$ for computing some explicit functions on k strings m-bits each.

When only two players are involved, it is just the usual model for communication complexity, which was first proposed by Yao [Y1] and has been studied extensively by many researchers [HMT, LS, MS, PS, Th]. In this paper we consider the following model generalizing both the round-table model and Yao's model:

> A number of players wish to cooperatively determine a Boolean function $f(x_1, \ldots, x_k)$ which accepts k inputs each m bits long. Suppose each player knows at most t inputs. The question of interest is to minimize the number of bits $C_{k,t}(f)$ to be exchanged in order to compute f.

To determine the communication complexity $C_{k,t}(f)$ could be a difficult problem for a general function f. The main thrust of this paper is to demonstrate the relation of communication complexity to several hypergraph properties. Consequently, lower bounds for $C_{k,t}$ can then be established. These hypergraph properties arise in the study of random-like graph properties, so called *quasi-random*.

[1]This paper has appeared in SIAM J. Discrete Math. 6 (1993), 110-123
[2]Bell Communications Research, Morristown, NJ 07960.
[3]Visitor, DIMACS center, Rutgers University, P.O. Box 1179–Busch Campus, Piscataway, NJ 08855

Quasi-randomness was first introduced in [CGW] by showing a large number of disparate graph properties are mutually equivalent in the sense that any graph satisfying one of the properties must of necessity satisfy all of them. More recently, in [C] it was shown that several equivalence classes $\mathcal{A}_i$ form a hierarchy of classes of properties for k-uniform hypergraphs (or k-graphs, for short) and for Boolean functions with k input arguments (also called k-functions). The quasi-random class $\mathcal{A}_k$, introduced in [CG1], consists of graph properties such as : "For any fixed $s \geq 2k$ all k-graphs on s vertices appear almost equally often as induced subgraph of G." On the other hand, in $\mathcal{A}_0$ there is the property that the number of edges in G is approximately the same as the number of non-edges in G. The detailed description of the equivalence classes $\mathcal{A}_i$ and the hierarchy

$$\mathcal{A}_0 \supset \mathcal{A}_1 \supset \cdots \supset \mathcal{A}_k$$

will be described in Section 2.

Among various properties in the equivalence class $\mathcal{A}_i$, there are two interesting invariants– the i-discrepancy and the i-deviation (see Section 2 for definition). Intuitively, the i-deviation provides a quantitative indication as to how much the graph deviates from random graphs. *Discrepancy* is useful in various contexts, in particular, corresponding to various statistical tests arising in complexity analysis. Roughly speaking, *discrepancy* is a "global" property that is often hard to compute, while *deviation* is a "local" property that is easy to compute. The quasi-randomness results imply that the i-discrepancy of a function is small if and only if its i-deviation is small. Furthermore, the i-discrepancy can be used to characterize the communication complexity $C_{k,i}$. Using the results of [BNS], this further leads to explicit construction of functions $f_{k,t}$ with communication complexity $C_{k,t}$ lower bounded by $\Omega(mc^{-t})$. One of the consequences is a simple proof of the lower bound of $\Omega(m2^{-k})$ on the communication complexity of the "generalized inner product" function as described in Section 3.

The communication complexity $C_{k,t}$ corresponds in a natural way to the complexity of a t-head Turing machine that computes Boolean functions with k inputs (as discussed in Section 3). As an immediate consequence, lower bounds for time-space tradeoffs can be obtained. We prove that for any fixed t, any $(t-1)$-head TM computing the function $f_{k,t}$ on m-bit strings requires a time-space tradeoff of $TS \geq \Omega(m^2)$.

Discrepancy can also be interpreted in terms of a game of *switches* and *lights* (also discussed in Section 3). Apart from being interesting in its own right, this interpretation yields an short proof that the communication complexity $C_{k,i}$ of a *random* k-function f is at least $\frac{(k-i+1)}{2}m$.

In Section 4, we conclude with some open problems and remarks about the relations of communication complexity to other complexity issues. The *quantitative* quasi-random classes for k-graphs with edge density α and various *expansion* properties are also mentioned.

2 Quasi-random Classes

2.1 notation

We use $\binom{X}{k}$ to denote the set of k-element subsets of a set X of cardinality $\geq k$. A k-graph $G = (V, E)$ consists of a set $V = V(G)$, called the *vertices* of G, and a subset $E = E(G)$ of the set $\binom{V}{k}$ called the *edges* of G. Throughout this paper, G denotes a k-graph on n vertices unless otherwise specified.

For $X \subseteq V$, $G[X]$ denotes the subgraph of G induced by X, i.e. $G[X] = \left(X, E \cap \binom{X}{k}\right)$. Let H denote an l-graph where $l < k$ and $V(H) = V(G)$. The set $E(G, H)$ of edges of G induced by H is defined to be:

$$E(G, H) = \{x \in E(G) : \binom{x}{l} \subseteq E(H)\}$$

For $l = 1$, the edge set of H is just a subset of $V(G)$ and $E(G, H) = E(G[H])$. We denote $e(G) =\mid E(G) \mid$ and $e(G, H) =\mid E(G, H) \mid$.

Discrepancy. For $i \geq 2$, the $i - discrepancy$ of G, denoted by $disc_i(G)$, is defined as follows:

$$disc_i(G) = \underset{H:(i-1)-graph}{max} \frac{\mid e(G, H) - e(\bar{G}, H) \mid}{\mid V(G) \mid^k}$$

where $\bar{G}$ denotes the complement of G with edge set $\left\{x \in \binom{V}{k} : x \notin E(G)\right\}$.
We remark that $disc_2$ is often called *discrepancy* in the literature (see [ES]). $disc_i$ can be viewed as a natural generalization of *discrepancy*.

We let $\mu_G : \binom{V}{k} \to \{-1, 1\}$ denote the edge function of G, i.e. for $x \in \binom{V}{k}$,

$$\mu_G(x) = \begin{cases} -1 & \text{if } x \in E \\ 1 & otherwise \end{cases}$$

Let V^k denote the set of k-tuples $(v_1, \ldots, v_k)$, $v_i \in V$, where the v's are not necessarily distinct. Let $\prod_G^{(i)} : V^{k+i} \to \{-1, 1\}$ denote the following function of G.

$$\prod\nolimits_G^{(i)}(u_1, \ldots, u_{2i}, v_{i+1}, \ldots, v_k) = \prod_{\epsilon_1} \cdots \prod_{\epsilon_i} \mu_G(\epsilon_1, \ldots, \epsilon_i, v_{i+1}, \ldots, v_k)$$

where $\epsilon_j \in \{u_{2j-1}, u_{2j}\}$ for $j \leq i$. Note that $\prod_G^{(i)}$ is a product of 2^i terms each of which is an edge function. For $i = 0$, we define $\prod_G^0 = \mu_G$.

Deviation. The $i - deviation$ of G, denoted by $dev_i(G)$, is defined as follows:

$$dev_i(G) = \frac{1}{n^{k+i}} \sum_{u_1, \ldots u_{k+i}} \prod\nolimits_G^{(i)}(u_1, \ldots, u_{k+i})$$

Thus $dev_i(G)$ assumes a value between -1 and 1. (Another interpretation is that n^{k+i} dev_i is the difference of the number of "even partial (squashed) octahedrons" and the "odd partial (squashed) octahedrons" as described in [CG1] and [CG2].)

2.2 quasi-randomness

We will use the following convention. Suppose we have two classes $P = P(o(1))$ and $P' = P'(o(1))$, each with occurrences of the asymptotic $o(1)$ notation. By the implication "$P \Rightarrow P'$", we mean that for each $\epsilon > 0$ there is a $\delta > 0$ (a function of ϵ and k but independent of n) such that if $G(n)$ satisfies $P(\delta)$ then it also satisfies $P'(\epsilon)$, provided $n > n_0(\epsilon)$. Two properties P and P' are said to be equivalent if $P \Rightarrow P'$ and $P' \Rightarrow P$.

Here we define several classes of properties for k-graphs.

For $i = 0$ and 1, define the properties

R_0 : $e(G) - e(\bar{G}) = o(n^k)$ where $\bar{G}$ denotes the complement of G.

R_1 : G is almost regular. That is,

$$\sum_{u_1,\ldots,u_{k-1}} \left(d^+(u_1,\ldots,u_{k-1}) - d^-(u_1,\ldots,u_{k-1})\right)^2 = o(n^{k+1})$$

where $d^+(u_1,\ldots,u_{k-1}) = \mid \{v \in V : \{u_1,\ldots,u_{k-1},v\} \in E(G)\} \mid$, and
$d^-(u_1,\ldots,u_{k-1}) = \mid \{v \in V : \{u_1,\ldots,u_{k-1},v\} \notin E(G)\} \mid$.
For $i \geq 2$, define

$$R_i : \text{ For every } (i-1)\text{-graph } H \ , \quad e(G,H) - e(\bar{G},H) = o(n^k)$$

In [CG1] it was shown that the property $dev_k(G) = o(1)$ for a hypergraph G is equivalent to a number of properties, among which are :

Q : For all k-graphs G' on $2k$ vertices, the number of (labelled) occurences of G' in G as an induced subgraph is $(1 + o(1))n^{2k}2^{-\binom{2k}{k}}$.

Let s denote a fixed integer and $s \geq 2k$.

$Q(s)$: For all k-graphs $G'(s)$ on s vertices the number of (labelled) occurrences of G' in G as an induced subgraph is $(1 + o(1))n^s 2^{-\binom{s}{k}}$.

In [C] the deviation property is further generalized to the following property (denoted P_i)
For $i \geq 0$,

P_i : $dev_i(G) = o(1)$.

The main results of [C] can be summarized in the following two theorems.

Theorem 1 *Properties P_i and R_i are equivalent for $i = 0, \ldots, k$. In particular for $i \geq 2$, we have*

$$\begin{aligned} (i) \quad disc_i(G) \quad &= \max_{H:(i-1)-graph} \frac{\mid e(G,H) - e(\bar{G},H) \mid}{\mid V(G) \mid^k} < (dev_i(G))^{1/2^i} \\ (ii) \quad dev_i(G) \quad &< 4^i \, (disc_i(G))^{1/2^i} \end{aligned}$$

Theorem 1, in fact, has interesting computational implications. It is easy to see that computing $disc_i$ for general G (naively) takes time $O(2^{n^i}.n^k)$, since the number of i-graphs is $O(2^{n^i})$ and for each i-graph H, computing $\mid e(G,H) - e(\bar{G},H) \mid$ takes $O(n^k)$ time. On the other hand, dev_i can be computed in time $O(n^{k+i})$ since dev_i is a sum of n_{k+i} terms, each term in turn is a product of 2^i subterms each of which is an edge function. Thus Theorem 1 leads to the following conclusion: Although it takes exponential time to compute $disc_i$ exactly, an approximation can be obtained by using dev_i in only polynomial-time. We remark that it would be of interest if the power $1/2^i$ on the right-hand sides of the inequalities could be improved.

Theorem 2 *Let $\mathcal{A}_i$ denote the equivalence class of k-graphs for which P_i holds. Then,*

$$\mathcal{A}_0 \supset \mathcal{A}_1 \supset \mathcal{A}_2 \cdots \supset \mathcal{A}_k$$

The family $\mathcal{A}_i = \mathcal{A}_i^{(k)}$ of k-graphs is said to be (k,i)-quasi-random, or sometimes i-quasi-random if there is no confusion. The term, "k-quasi-random" for k-graphs is the same as "quasi-random" as in previous papers.

Here we describe the constructions of k-graphs G_i, separating class $\mathcal{A}_i$ from $\mathcal{A}_{i+1}$, for it is used in a later section on lower bounds for communication complexity. Since $P_i \Rightarrow P_{i+1}$ for any i, we have $\mathcal{A}_i \supseteq \mathcal{A}_{i+1}$. To show $\mathcal{A}_i \supset \mathcal{A}_{i+1}$, for $i = 0, \ldots, k-1$, the idea is to construct k-graphs G_i with the property that $G_i \in \mathcal{A}_i$ and $G_i \notin \mathcal{A}_{i+1}$ using quasi-random graphs as the basic building blocks. In [CG1], two families of quasi-random k-graphs are given, one of which is the Paley k-graph P_k with $V(P_k) = \{1, 2, \ldots, n\}$ (n is a prime) and $\mu_{P_k}(u_1, \ldots, u_k) = 1$ if and only if $u_1 + \cdots + u_k$ is a quadratic residue modulo n.

For each i, we define the k-graph G_i as follows:

$$\begin{aligned} V(G_i) \quad &= \quad V(P_i) = V \\ E(G_i) \quad &= \quad \{x \in \binom{V}{k} : \mid \binom{x}{i} \cap E(P_i) \mid \equiv 0 \pmod 2\} \end{aligned}$$

Claim $G_i \in \mathcal{A}_i \setminus \mathcal{A}_{i+1}$

Proof: **Part 1** $G_i \in \mathcal{A}_i$:

It is shown in [C] (by making use of the character sum inequality of Burgess [B]) that

$$dev_i(G_i) = O(n^{-1/2})$$

Therefore G_i satisfies Property P_i and hence is in $\mathcal{A}_i$.

Part 2 $G_i \notin \mathcal{A}_{i+1}$:

Consider the set $E(G_i, P_i)$ of edges of G_i induced by the Paley graph P_i. An edge x is in $E(G_i, P_i)$ means every i-subset of x has a sum which is a quadratic nonresidue. By definition, x contains an even number of i-sets each of which has a sum which is quadratic nonresidue. This can happen only when $\binom{k}{i} \equiv 0 \pmod 2$. Therefore either $E(G_i, P_i)$ is empty or $E(\bar{G}_i, P_i)$ is empty. Since k and i are all fixed integers,

$$
\begin{aligned}
\mid E(G_i, P_i) - E(\bar{G}_i, P_i) \mid &= \mid E\left(\binom{V}{k}, P_i\right) \mid \\
&= (1 + o(1))\, \frac{n^k}{2^{\binom{k}{i}}} \\
&\neq o(n^k)
\end{aligned}
$$

Thus $G_i \notin \mathcal{A}_{i+1}$.

We now describe a more general construction of k-functions G_i using any quasi-random graph in $\mathcal{A}_k$ as the basic building block.

General construction for $G_i \in \mathcal{A}_i \setminus \mathcal{A}_{i+1}$. Note that the proof of Part 2 is quite general– does not make use of the fact that the basic building block was the Paley k-graph P_k. We show here that, in fact, any quasi-random graph in $\mathcal{A}_k$ serves the purpose as well. (For example, the family of "even intersection" k-graphs defined in [CG1] is an equally good choice.) First we need the following definition of the "neighborhood graph" of a k-graph. Given a k-graph G, the *neighborhood graph* $G(v)$ of a vertex v is the graph having vertex set $G \setminus \{v\}$ and edge set $E(G(v)) = \left\{x \in \binom{V}{k-1} : x \cup \{v\} \in E(G)\right\}$.

Let H_i be a quasi-random i-graph on n vertices. Then we define the k-graph G_i as follows:

$$
\begin{aligned}
V(G_i) &= V(H_i) = V \\
E(G_i) &= \{x \in \binom{V}{k} : \mid \binom{x}{i} \cap E(H_i) \mid \equiv 0 \pmod 2\}
\end{aligned}
$$

We outline the proof of $G_{k-1} \in \mathcal{A}_{k-1} \setminus \mathcal{A}_k$.

Part 1 $G_{k-1} \in \mathcal{A}_{k-1}$:

As a direct consequence of the definition of a neighborhood graph, we have

$$dev_i(G) = \frac{1}{n} \sum_{v \in V} dev_i(G(v))$$

For a fixed vertex v, consider the neighborhood graph $G_{k-1}(v)$ of the k-graph G_{k-1}. The edge set of $G_{k-1}(v)$ can be characterized as follows.

$$E\left(G_{k-1}(v)\right) = E_1 \cup E_2$$

where

$$E_1 = \left\{ y \in \tbinom{V}{k-1} : y \in H_{k-1} \text{ and } E(H_{k-1}(v)) \cap \tbinom{y}{k-2} \equiv 0 \pmod 2 \right\} \quad \text{and}$$
$$E_2 = \left\{ y \in \tbinom{V}{k-1} : y \notin H_{k-1} \text{ and } E(H_{k-1}(v)) \cap \tbinom{y}{k-2} \equiv 1 \pmod 2 \right\}$$

Thus

$$\mu_{G_{k-1}(v)} = \mu_{H_{k-1}} . \mu_{\delta(H_{k-1}(v))} \qquad (*)$$

where $\delta(H_{k-1}(v))$ is defined to be :

$$\delta(H_{k-1}(v)) = \left\{ y \in \binom{V}{k-1} : E(H_{k-1}(v)) \cap \binom{y}{k-2} \equiv 0 \pmod 2 \right\}$$

It is not very hard to verify that $(*)$ implies that

$$dev_{k-1}(G_{k-1}(v)) \quad = \quad dev_{k-1}\,(H_{k-1})$$

Thus

$$dev_{k-1}(G_{k-1}) = \sum_v \frac{dev_{k-1}\,(G_{k-1}(v))}{n} = o(1), \quad \text{since} \quad H_{k-1} \in A_{k-1}$$

This shows $G_{k-1} \in \mathcal{A}_{k-1}$.

Part 2 The proof of $G_{k-1} \notin A_k$ is identical to the proof of Part 2 with the Paley graph construction.

3 Communication Complexity

3.1 Quasi-random classes of functions

A k-function is a function f from V^k to $\{-1, 1\}$. We note that k-functions can be viewed as ordered k-graphs and k-graphs can be regarded as symmetric k-functions. In fact, most known lower bound constructions for k-functions are symmetric and thus can be reduced to hypergraphs. We shall see in the following that the notions of discrepancy and deviation extend to k-functions as well. For convenience, we use the same notation (*disc* and *dev*) for discrepancy and deviation of k-functions, and we warn the reader to interpret appropriately depending on the context. Thus, for example, $disc(f)$ refers to the deviation of a k-function f, whereas $disc(G)$ stands for that of a k-graph G.

Let I denote a subset of size i of $\{1, \ldots, k\} = [k]$. For a k-tuple $x = (x_1, \ldots, x_k)$, we define x_I to be an i-tuple $(x_{a_1}, \ldots, x_{a_i})$ where $a_1 < \ldots < a_i$ and $a_i \in I$.

Discrepancy. Let $\mathcal{H}_i$ denote a family of i-functions where $i < k$ and the members of $\mathcal{H}_i$ are indexed by $\binom{[k]}{i}$, denoted by h_I. We define $E(f, \mathcal{H}_i)$ as follows:

$$E(f, \mathcal{H}_i) = \left\{ x \in V^k : f(x) = -1 \text{ and for every } h_I \in \mathcal{H}_i,\ h_I(x_I) = -1 \right\}.$$

We denote the cardinality of $E(f, \mathcal{H}_i)$ by $e(f, \mathcal{H}_i)$. The i-discrepancy of f is defined as follows

$$disc_i(f) = \max_{\mathcal{H}_{i-1}} \frac{\mid e(f, \mathcal{H}_{i-1}) - e(-f, \mathcal{H}_{i-1}) \mid}{\mid V \mid^k}$$

Deviation. Define $\prod_{f,I}^{(i)} : V^{k+i} \to \{-1,1\}$ by

$$\textstyle\prod_{f,I}^{(i)}(x_1,\ldots,x_{k+i}) \ = \prod_{\epsilon_1}\cdots\prod_{\epsilon_k} f(\epsilon_1,\ldots,\epsilon_k)$$

where $\epsilon_j \in \{x_{j+m-1}, x_{j+m}\}$ if $j \in I$ and $m = \mid I \cap [1,j] \mid$; and $\epsilon_j = x_{i+m}$ if $j \notin I$. The i-deviation of f is defined to be:

$$dev_i(f) = \max_I \ \frac{1}{n^{k+i}} \sum_{x_1,\ldots,x_{k+i}} \textstyle\prod_{f,I}^{(i)}(x_1,\ldots,x_{k+i})$$

where I ranges over all subsets of $[k]$ of size i.

For fixed i, we consider the following properties for a k-function:

$$\begin{aligned}
\tilde{R}_i \quad &: \quad \text{For } i \geq 2, \text{ for every family } \mathcal{H}_{i-1} \text{ of } i-1\text{-functions} \\
& \quad\; e(f,\mathcal{H}_{i-1}) - e(-f,\mathcal{H}_{i-1}) = o(n^k) \\
\tilde{P}_i \quad &: \quad dev_i(f) = o(1).
\end{aligned}$$

It can be shown that the properties $\tilde{R}_i$ and $\tilde{P}_i$ are equivalent. In fact, the analogs of Theorems 1 and 2 for k-functions also hold (see [C]).

3.2 Multiparty communication games

In ([BNS]), Babai, Nisan and Szegedy considered the communication complexity for k-functions where each of the k players knows exactly $k-1$ inputs. Let $x = (x_1,\ldots,x_k)$ denote an input chosen uniformly over all k-tuples. Then the communication complexity is bounded by $\log \frac{1}{\Gamma(f)}$ where

$$\Gamma(f) \quad = \quad \max_S \left(\Pr[x \in S \text{ and } f(x) = -1] - \Pr[x \in S \text{ and } f(x) = 1]\right)$$

where S ranges over so-called "cylinder intersections". The theorem below generalizes the result of [BNS].

We first extend the notion of "cylinders" and "cylinder intersections" for functions in class A_i. A subset of $S^{(i-1)}$ of k-tuples is called a *cylinder* if membership in $S^{(i-1)}$ depends only on $i-1$ coordinates. Thus, based on which $i-1$ of the coordinates the k-tuple depends on, there will be $\binom{k}{i-1}$ types of $S^{(i-1)}$ in A_i. Furthermore, a subset of k-tuples is a *cylinder intersection* if it can be represented as an intersection of cylinders. Let $\cap S^{(i-1)}$ represent a subset which is an intersection of all $\binom{k}{i-1}$ types of cylinders. We define $\Gamma_i(f)$ of f to be

$$\Gamma_i(f) = \max_{\cap S^{(i-1)}} \left(\Pr[x \in \cap S^{(i-1)} \text{ and } f(x) = -1] - \Pr[x \in \cap S^{(i-1)} \text{ and } f(x) = 1]\right)$$

Let I denote the subset of i coordinates that $S^{(i)}$ depends on. Then we have the following natural correspondence between cylinders $S^{(i)}$ and i-functions h_I, for $i = 1,\ldots,k-1$:

$$x \in S^{(i)} \quad \Leftrightarrow \quad h_I(x_I) = -1$$

and

$$x \in \cap S^{(i)} \quad \Leftrightarrow \quad \text{for every } h_I \in \mathcal{H}_i,\ h_I(x_I) = -1.$$

This enables us to prove the following.

Theorem 3 *For $i = 2, \ldots, k$,*

$$\begin{aligned} \Gamma_i(f) &= disc_i(f) \\ C_i(f) &\geq \log \frac{1}{disc_i(f)} \end{aligned}$$

where $C_i(f)$ denotes the communication complexity of f in class A_i

Proof. Since x is chosen uniformly over all 2^{mk} possible k-tuples, we have

$$\begin{aligned} \Gamma_i(f) &= \max_{\cap S^{(i-1)}} \left(\Pr[x \in \cap S^{(i-1)} \text{ and } f(x) = -1] - \Pr[x \in \cap S^{(i-1)} \text{ and } f(x) = 1]\right) \\ &= \max_{\cap S^{(i-1)}} \frac{1}{2^{mk}} \left[\mid \{x : x \in \cap S^{(i-1)} \text{ and } f(x) = -1\} \mid - \mid \{x : x \in \cap S^{(i-1)} \text{ and } f(x) = 1\} \mid\right] \\ &= \max_{\mathcal{H}_{i-1}} \frac{1}{n^k} \left[e(f, \mathcal{H}_{i-1}) - e(-f, \mathcal{H}_{i-1})\right] \\ &= disc_i(f) \end{aligned}$$

The second part of the proof is similar to that of Lemma 2.2 in [BNS]; we include it here for the sake of completeness. Let P be any valid protocol for the given function f. We denote by $P(x)$, the value of $f(x)$ as computed by the protocol P. Let N be the number of different possible strings that may be written on the board by P. We want to prove that $N \geq 1/\Gamma_i(f)$. With each string s we associate $X_{P,s}$, the set of inputs for which s gets written on the board by P . It is not hard to see that $X_{P,s}$ is a *cylinder intersection* $\cap S^{(i-1)}$.

Let x be chosen uniformly over all k-tuples. Since P is a valid protocol,

$$|\Pr[P(x) = f(x)] - \Pr[P(x) \neq f(x)]| = 1$$

We can estimate the same by summing over different $X_{P,s}$:

$$\begin{aligned} &|\Pr[P(x) = f(x)] - \Pr[P(x) \neq f(x)]| \\ &\qquad \leq \sum_s |\Pr\left[P(x) = f(x) \text{ and } x \in X_{P,s}\right] - \Pr\left[P(x) \neq f(x) \text{ and } x \in X_{P,s}\right] \end{aligned}$$

where s ranges over all possible strings that may be written.

Thus

$$\begin{aligned} 1 &\leq \sum_s |\Pr\left[P(x) = f(x) \text{ and } x \in X_{P,s}\right] - \Pr\left[P(x) \neq f(x) \text{ and } x \in X_{P,s}\right] \\ &= \sum_s \Pr\left[f(x) = 1 \text{ and } x \in X_{P,s}\right] - \Pr\left[f(x) = -1 \text{ and } x \in X_{P,s}\right] \\ &\leq \sum_s \Gamma_i(f), \text{ since } X_{P,s} \text{ is a cylinder intersection} \\ &= N\Gamma_i(f) \end{aligned}$$

This proves

$$C_i(f) = \log N \geq \log \left[\frac{1}{\Gamma_i(f)} \right].$$

□

We remark that we do not restrict the number of players. Suppose we consider the minimum number $C_{k,i}(p)$ of bits required to be exchanged for some p players, each knowing at most $i-1$ inputs of a k-function. It is easy to see that $C_{k,i}(p) = C_{k,i}(p')$ if $p' > p$. Moreover, $C_{k,i}(p'') > C_{k,i}(p)$ if $p'' < p$.

Fact. For any k-function f, $\quad C_i(f) \leq (k-i+1)m$.
Proof. If $(k-i+1)$ inputs get written on the board, then *some* player would know all k inputs. This could be done, trivially, if a player always writes an input that is not already present on the board.

Theorem 4 *For a random k-function f,* $\quad C_i(f) \geq \frac{(k-i+1)}{2} m.$

Proof. For a random k-function f, it is not hard to verify that with probability approaching 1, we have $|e(f,H) - e(-f,H)| = O(n^{(k+i-1)/2})$ for every $(i-1)$-function H and this is best possible. Using similar methods as in [ESp], this implies,

$$\begin{aligned} disc_i(f) &= \max_{\mathcal{H}_{i-1}} \frac{|e(f,\mathcal{H}_{i-1}) - e(-f,\mathcal{H}_{i-1})|}{n^k} \\ &= O(n^{(-k+i-1)/2}) \\ &= O(2^{(-k+i-1)m/2}) \end{aligned}$$

Hence

$$C_i(f) = \Omega \left(\frac{(k-i+1)}{2} m \right).$$

In [BNS] examples of functions f with $C_k(f) = \Omega\left(\frac{m}{2^k}\right)$ are given. Here we give a short proof for the following "Box-product" of functions.

Box-product of k-functions and Deviation. Let $f : V^k \to \{-1,1\}$ and $g : W^k \to \{-1,1\}$ be two k-functions. We define $f \Box g : (V \times W)^k \to \{-1,1\}$ to be the following k-function

$$f \Box g \left((x_1,y_1),\ldots,(x_k,y_k)\right) = f(x_1,\ldots,x_k).g(y_1,\ldots,y_k)$$

It can be shown that (also see [CG2])

$$dev_i(f \Box g) = dev_i(f).dev_i(g)$$

Example 1. Consider the graph G on three vertices v_1, v_2, v_3, with the edges $\{v_1,v_2\}$ and $\{v_2,v_3\}$; let $V = \{v_1,v_2,v_3\}$ and f denote the edge function of G. It is easy to check that

$dev_0(f) = dev_1(f) = 1/9$. Taking the Box-product of f with itself gives us the function $f' = f \Box f$ with the properties: $dev_0(f') = dev_1(f') = 1/81$.

Example 2. Consider the following "generalized inner product function" f_m, defined on subsets S_i of a set of size m.

$$f_m(S_1, \ldots, S_k) = \begin{cases} 1 & \text{if } S_1 \cap \ldots \cap S_k \text{ is even} \\ -1 & \text{otherwise} \end{cases}$$

For the special case $m = 1$, f_1, each S_i is a singleton or empty. It is easy to verify, by induction on m, that

$$f_m = f_1 \Box \cdots \Box f_1 \qquad \text{(m times)}$$

Since $dev_i(f_1) = 1 - 2^{-k-i+1} = c < 1$, we readily obtain $dev_i(f_m) < c^m$. In particular, $dev_k(f_m) < c^m$, where $c < 1$.

This implies that $disc_k(f_m) < c^{m/2^k}$. And by Theorem 3,

$$C_k(f_m) \geq log \frac{1}{disc_k(f_m)} = \Omega\left(\frac{m}{2^k}\right)$$

Therefore, we prove the following.

Theorem 5 *The generalized inner product function f_m has $C_k(f_m) = \Omega\left(\frac{m}{2^k}\right)$.*

One of the main results in [BNS] is to establish an upper bound for $disc_k f_m$ and thereby obtain a lower bound for $C_k(f_m)$. Independently, an upper bound for $disc_k f_m$ is also proved in [CG1]. However, both the proofs are more complicated in comparison to the one we described above. The significance of the Box-product is thus apparent. Starting with a function with $dev_i < 1$, we can construct functions with exponentially small dev_i by repeatedly considering Box-product of the original function with itself.

The following result shows that Theorem 5 is an instance in a more general setting.

Theorem 6 *There are explicit k-functions f satisfying*

$$C_i(f) = \Omega\left(\frac{m}{2^i}\right).$$

Proof. Recall from Section 2.2, we constructed k-graphs $G_i \in \mathcal{A}_i \setminus \mathcal{A}_{i+1}$ for which

$$dev_i(G_i) = O(n^{-1})$$

In terms of k-functions, this implies that

$$dev_i(f_{k,i}) = O(2^{-m})$$

So

$$\begin{aligned} disc_i(f_{k,i}) &\leq (dev_i)^{1/2^i} \\ &= O(2^{-m/2^i}) \end{aligned}$$

This implies $\quad C_i(f_{k,i}) = \Omega\left(\frac{m}{2^i}\right)$.

Remark. One of the important questions is to find communication complexity lower bounds that do not decrease exponentially in k for some explicit k-function. This would improve results [BNS] on pseudorandom sequences, time-space tradeoffs for multi-head Turing machines, and length-width tradeoffs for oblivious branching programs. Improving the relation (Theorem 1) between $disc_i$ and dev_i would be significant for the same reason.

3.3 Application to Turing machines

Let f be a k-function. Under our general communication model, we have the following analog of the result of Babai et al [BNS] for the time-space tradeoff of Turing machines and we omit the proofs here.

Lemma 1 *Any i-head Turing machine that computes a k-function f from the following input:*

$$< x_1 > * * ** < x_2 > * * * * \cdots * * * * < x_k >$$

(where $* * **$ means l spaces on the input tape) requires a time-space tradeoff of $TS \geq lC_{i+1}(f)/i$.

And hence

Theorem 7 *For any fixed i, any i-head Turing machine computing the k-function $f_{k,i}$ requires a time-space tradeoff of $TS \geq \Omega(m^2)$.*

3.4 Discrepancy and the switching lights model

There is yet another interpretation for $disc_i$ in terms of the *switching lights* model, first described in [Sp] for the two dimensional case. The game consists of an $n \times n$ array A of lights and $2n$ switches, one for each row x_i and column y_j. Each switch when thrown changes each light in its line from *off* to *on* or from *on* to *off*. The *difference* is defined as the absolute value of the number of lights on minus the number of lights off ranging over all possible settings of the switches. Given an initial configuration, the object is to maximize the difference. Mathematical formulation of this problem shows that maximizing this difference corresponds to computing the discrepancy (the Γ function) in the multiparty communication model in a sense made precise in the theorem below.

Consider a k-dimensional array of n^k lights. Imagine each switch controlling an i-dimensional hyperplane of n^i lights; i.e. each switch when thrown changes each light in the particular hyperplane from *off* to *on* or from *on* to *off*. There are $(i+1)n^{k-i}$ such switches and the aim is to maximize the difference between the number of lights *on* and *off*. We denote this by D_k^i.

Thus, in k-dimensions, we formulate $k-1$ discrepancy problems associated with the switching game.

In 3-dimensions we have two problems: D_3^2 and D_3^1. The distinction is that each switch controls a plane of lights in one case and a line of lights in the other. Intuitively, we would expect D_3^1 to be higher than D_3^2, and the intuition is right. The mathematical formulation of this case (D_3^2) is as follows.

Let the array of n^3 lights be represented by $A(ijk) = \pm 1$, for $i, j, k = 1, \ldots, n$. Thus 1 represents a light *on* and -1 a light *off*. Further we let x_i, y_j, z_k represent the $3n$ switches. "Throwing" a switch x_i corresponds to setting $x_i = -1$. Given an initial setting of $A(i,j,k) = \pm 1$ we define the *discrepancy* of A to be

$$D(A) = \underset{x_i, y_j, z_k = \pm 1}{max} A(i,j,k) \cdot x_i y_j z_k$$

i.e. the maximum difference between the number of lights *on* and *off* that one can obtain by throwing the switches. Further we define

$$D_3^2 = \underset{A}{min} \underset{x_i y_j z_k}{max} A(i,j,k) x_i y_j z_k$$

to be the maximum ranging over all possible initial configurations of A. The case D_k^i for general i have a similar mathematical formulation.

The following theorem establishes the equivalence between D_k^i and the "discrepancy" Γ_i in the context of multiparty communication complexity. Firstly, we associate with a given k–input function f, the k–dimensional array A_f of size $2^m \times \cdots \times 2^m$ where

$$A_f(i_1, \ldots, i_k) = f(x_1 = i_1, \ldots, x_k = i_k)$$

Thus we are assuming (without loss of generality) that each input x_j ranges from 1 to 2^m. We then have the following:

Theorem 8

$$\Gamma_i(f(m)) = \frac{1}{2^{mk}} D_k^{k-i}(A_f)$$

Proof. Basically, the number of inputs each player knows corresponds to the number of coordinates required to specify a switch, and the possible bit sequences by the players correspond to the switch settings. We describe the proof for $i = k-1$. The general case is quite similar and will be omitted. It is not difficult to see that Γ_{k-1} can be rewritten as follows (see [BNS]).

$$\Gamma_{k-1}(f(m)) = \underset{\phi_1, \ldots, \phi_k}{max} \left| E\left[f(x_1, \ldots, x_k)\phi_1(x_1, \ldots, x_k) \cdots \phi_k(x_1, \ldots, x_k)\right] \right|$$

where the expectation is over all possible 2^{mk} choices of $x_1, \ldots, x_k$, and the maximum is taken over all functions $\phi_j : (\{0,1\}^m)^k \to \{0,1\}$ such that ϕ_j does not depend on x_j. (Intuitively, ϕ_j corresponds to possible messages communicated by player P_i.) Thus

$$\Gamma_{k-1}(f(m)) \quad = \quad \frac{1}{2^{mk}} \max_{\phi_1,\ldots,\phi_k} |\sum_{x_1} \cdots \sum_{x_k} [f(x_1,\ldots,x_k)\phi_1(x_1,\ldots,x_k)\cdots\phi_k(x_1,\ldots,x_k)]\,|$$

Whereas *discrepancy* of A_f in the switching game is defined as

$$D_k^1(A_f) = \max_{s_{i_1},\ldots,s_{i_k}} \sum_{i_1=1}^{2^m} \cdots \sum_{i_k=1}^{2^m} A(i_1,\ldots,i_k) s_{i_1} \cdots s_{i_k}$$

where the switch $s_{i_j} : \{1,\ldots,2^m\}^k \to \{0,1\}$ depends on all but index i_j. It is now easy to see that the functions ϕ_j correspond to the switches s_{i_j}.

Thus $\Gamma_{k-1}(f(m)) = \dfrac{1}{2^{mk}} D_k^1(A_f)$ □

The following theorem appears in [ESp] in the form of a result on a hypergraph-coloring problem.

Theorem 9 *There exist arrays A of n^k lights such that*

$$D_k^i(A) \le c(k,i) n^{(k+i-1)/2}$$

where $c(k,i)$ is an explicit constant depending on k and i.

Proof. The proof is straight forward using the probabilistic method, and can be found in [T].

Remark 1. Theorem 7 shows that for a random k-function f, $disc_i(f) = O\left(n^{(k+i-1)/2}\right)$. Thus this yields a simple proof of

$$\begin{aligned} C_i(f) &\ge \log\left(n^{(k+i-1)/2}\right) \\ &= \log\left(2^{(k+i-1)m/2}\right) \\ &= \frac{(k+i-1)}{2} m. \end{aligned}$$

Remark 2. Note that Theorem 9 guarantees the *existence* of an array A such that $D_k^i(A) \le cn^{(2k-i)/2}$. Can we, in fact, construct such an array? The question is open for $k > 2$. For $k = 2$ it is known that an $n \times n$ *Hadamard matrix* H works! That is,

$$D_2^1(H) \le n^{3/2}$$

However, it is not clear how to generalize the notion of Hadamard matrices for the case of $k > 2$. Apart from being an interesting derandomization question by itself, this has the following implications. In view of Theorem 8, upper bounds on D_k^i yield, in turn, upper bounds on

Γ_i, and further give lower bounds on the communication complexity of multiparty protocols. Thus, making Theorem 5.1 constructive seems to be an interesting open problem.

Remark 3. The inequality in Theorem 7 is the best possible. That is, given any arbitrary initial configuration for the array of lights, one can set the switches such that the maximum difference is $\Omega(n^{(k+i-1)/2})$. In fact, the random configuration achieves the bound which can be proved by generalizing the result in [ESp]. In fact, the method of conditional expectations can be used in derandomizing the algorithm and a sequential as well as a parallel algorithm is described in [T] to achieve the optimal setting of the switches.

4 Problems and Remarks

In addition to various problems that were mentioned in previous sections, many problems and directions remain to be explored. It would be of interest to establish relations and connections with other complexity problems. For example, an interesting relation between circuit complexity and quasi-randomness has been demonstrated through some recent work of Hastad and Goldmann [HG]. Using the results of [BNS], Hastad and Goldmann show that (inter alia), evaluating the generalized inner product function on $k+1$ inputs by a depth 3 unweighted threshold circuit with bottom fanin at most k would require size $2^{\Omega(n/k4^k)}$. One way to improve these lower bounds is to come up with explicit hypergraphs or k-functions with smaller discrepancy or higher communication complexity.

Although we deal with hypergraphs with the edge density 1/2, the results can easily be generalized to hypergraphs or functions with any fixed edge density α, for $0 < \alpha < 1$. For a function f from V^k to $\{-1, 1\}$, we define $f_\alpha(x) = 1-\alpha$ if $f(x) = -1$ and $f_\alpha(x) = -\alpha$ if $f(x) = 1$. In [C], $dev_i f_\alpha$, $disc_i f_\alpha$ and the class $\mathcal{A}_{i,\alpha}$ are defined analogous to dev_i, $disc_i$, and $\mathcal{A}_i$. In particular, the 2-discrepancy $disc_{2,\alpha}$ is described as follows:

$$disc_{2,\alpha}(f) = \underset{X \subseteq V}{max} \frac{e(f,X) - \alpha \mid X \mid^k}{\mid X \mid^k}$$

where $e(f,X) =\mid \{x \in \binom{X}{k} : f(x) = -1\} \mid$. Suppose we choose α to be $e(f,X) =\mid \{x \in \binom{V}{k} : f(x) = -1\} \mid / \mid V \mid^k$ (which can be viewed as the density of "ordered" hyper-edges). Then $disc_{2,\alpha}(f)$ associates with the maximum quantity that the number of ordered-edges in a subset X can differ from the average. If we can use $dev_{2,\alpha}$ to (upper) bound $disc_{2,\alpha}(f)$, then we can (lower) bound the number of edges leaving X from every $X \subseteq V$ and thus assert the expanding property of the hypergraphs.

References

[B] D.A. Burgess, On character sums and primitive roots, *Proc. London Math. Soc.* **12** (1962), 179-192.

[BFL] L. Babai, P. Frankl and J. Simon, Complexity classes in communication complexity theory, *27th FOCS* (1986), 337-347.

[BNS] L. Babai, N. Nisan, and M. Szegedy, Multiparty protocols and logspace-hard pseudorandom sequences, *21st STOC* (1989), 1-11.

[C] F.R.K. Chung, Quasi-random classes of hypergraphs, *Random Structures and Algorithms*, **1** (1990), 363-382.

[C2] F.R.K. Chung, Regularity lemmas for hypergraphs and quasi-randomness, *Random Structures and Algorithms*, **2** (1991), 241-252.

[CFL] A.K. Chandra, M.L. Furst and R.J. Lipton, Multiparty protocols, *24th FOCS* (1983), 94-99.

[CGW] F.R.K. Chung, R.L. Graham and R.M. Wilson, Quasi-random graphs, *Combinatorica* **9** (1989), 345-362.

[CG1] F.R.K. Chung and R.L. Graham, Quasi-random hypergraphs, *Random Structures and Algorithms*, **1** (1990), 105-124.

[CG2] F.R.K. Chung and R.L. Graham, Quasi-random set systems, *J. of AMS*, **4** (1991) 151-196.

[ES] P. Erdős and V.T. Sós, On Ramsey-Turán type theorems for hypergraphs, *Combinatorica* **2** (1982), 289-295.

[ESp] P. Erdős and J. Spencer, Imbalances in k-colorations, *Networks* **1** (1971), 379-385.

[HG] J. Hastad and M. Goldmann, On the Power of Small-Depth Threshold Circuits, *31st FOCS* (1990), 610-618.

[HMT] A. Hajnal, W. Maass, and G. Turán, On the Communication Complexity of graph properties, *20th STOC* (1988), 186-191.

[MS] K.Melhorn and E.M. Schmidt, Las Vegas is better than determinism in VLSI and distributed computing, *14th STOC* (1982), 330-337.

[L] L. Lovász, Computational Complexity : A Survey, in Paths, Flows, and VLSI-Layout, (B. Korte et al eds.), Springer-Verlag (1990), 235-266.

[LS] L. Lovász and M. Saks, Lattices, Möbius functions and communication complexity, *29th FOCS* (1988), 81-90.

[PS] C.H. Papadimitriou and M. Sipser, Communication Complexity, *14th STOC* (1983) 196-200.

[T] P. Tetali, *Analysis and Applications of Probabilistic Techniques*, Ph.D. Thesis, New York University (October 1991).

[SS] M. Simonovits and V.T. Sós, Szemerédi-partition and Quasi-randomness, *Random Structures and Algorithms* **2** (1991), 1-10.

[Sp] J. Spencer, *Ten Lectures on the Probabilistic method*, SIAM Publications, Philadelphia (1987).

[Sz] E. Szemerédi, Regular partitions of graphs, *Problémes combinatoires et théorie des graphs*, Coll. CNRS (1976), 399-401.

[Th] C.D. Thompson, Area-time complexity for VLSI, *11th STOC* (1979), 81-88.

[Y1] A.C.C. Yao, "Some complexity questions related to distributive computing", *11th STOC* (1979), 209-213.