Electronic Colloquium on Computational Complexity, Report No. 152 (2017)

One-way Communication and Non-adaptive Decision Tree

Swagato Sanyal *

October 9, 2017

Abstract

Let f be a Boolean function on n-bits, and IP the inner-product function on $2b$ bits. Let $f^{\langle\rangle} := f \circ \mathsf{IP}^n$ be the two party function obtained by composing f with IP. In this work we bound the one-way communication complexity of $f^{\langle\rangle}$ in terms of the non-adaptive query complexity of f, from below. Similar results are known for general communication protocols and adaptive decision trees, when the arity of the inner function (inner-product in our case) is at least logarithmic in n. We prove analogous results for one-way communication as long as b is a large enough constant. Let $\mathsf{R}^{\rightarrow}_{\mathrm{cc},\epsilon}(\cdot)$ and $\mathsf{D}^{\rightarrow}_{\mathrm{cc}}(\cdot)$ denote the randomized ϵ-error and deterministic one-way communication complexity respectively. Let $\mathsf{D}^{\rightarrow}_{\mathrm{dt}}(\cdot)$ denote the deterministic non-adaptive query complexity. Let $\mathrm{H}_{\mathsf{bin}}(\cdot)$ denote the binary entropy function. We prove that

- If f is a *total* Boolean function and $b \geq 2$, then $\mathsf{R}^{\rightarrow}_{\mathrm{cc},\epsilon}(f^{\langle\rangle}) \geq (1 - \mathrm{H}_{\mathsf{bin}}(\epsilon))n(b-1)$.
- If f is a partial Boolean function and $b \geq 4$, then $\mathsf{D}^{\rightarrow}_{\mathrm{cc}}(f^{\langle\rangle}) = \Omega(b \cdot \mathsf{D}^{\rightarrow}_{\mathrm{dt}}(f))$.

1 Introduction

Communication complexity of composed functions has been a topic of active research. In this setting, there is an *outer function* f on n-bits and an *inner function* $g : \{0,1\}^r \times \{0,1\}^s \to \{0,1\}$ (also referred to as the *gadget*). Alice and Bob get strings $x = (x^{(1)}, \ldots, x^{(n)}) \in (\{0,1\}^r)^n$, $y = (y^{(1)}, \ldots, y^{(n)}) \in (\{0,1\}^s)^n$ as inputs. The task for them is to compute $f \circ g^n(x,y) := f(g(x^{(1)}, y^{(1)}), \ldots, g(x^{(n)}, y^{(n)}))$. There is a direct upper bound to the amount of communication required to accomplish this task in terms of the decision tree complexity of f. If f has a decision tree which makes d queries, the parties can simulate the decision tree; each query to f can be implemented by computing the corresponding copy of g through communication. Thus if g admits a communication protocol of complexity c, $f \circ g^n$ has a communication protocol of complexity cd.

The above bound holds for both deterministic and randomized versions of the two models. A natural question is whether the above bound is optimal. In 2015, Göös Pitassi and Watson settled this question in the affirmative for

*Centre for Quantum Technologies, Singapore and Nanyang Technological University, Singapore. ssanyal@ntu.edu.sg

ISSN 1433-8092

the deterministic case, when g is the index-function with number of index bits logarithmic in n [GPW15]. They showed this result by using techniques developed by Raz and McKenzie [RM99]. Their result was later generalized by Chattopadhyay et al. [CKLM17]. Afterwards in 2017, the same group of authors [GPW17] as well as Anshu et al. [AGJ$^+$17] settled the randomized case in the affirmative. In both of these works, the number of inputs to the gadgets are logarithmic in n. In all the above results, f can be an arbitrary partial Boolean function.

The above results established that the task of determining the communication complexity of $f \circ g^n$ is the same as the task of determining the decision tree complexity of f. In other words, a lower bound in the simpler model of decision tree is *lifted* by these theorems to a lower bound in the richer and more powerful model of communication; this justifies the usage of the term *lifting theorem* for referring to these results. Many important functions studied in communication complexity are indeed composed functions. Notable examples include Disjointness, gap-Hamming and Equality. Research in lifting theorems contributes in an important way to acquiring a unified understanding of the power and limitations of communication protocols for composed functions. Further research showed lifting theorems for other measures of complexity. Examples include lifting *approximate junta degree* to *smooth corruption bound* [GLM$^+$15], *approximate degree* to *approximate rank* [She11], and *conical junta degree* to *non-negative rank* [KMR17]. When the gadget is the two-bit XOR function, Hatami et al. [HHL16] lifted a lower bound on the parity decision tree complexity of the outer function to a lower bound on the deterministic communication complexity of the composed function, with a polynomial gap.

As mentioned in [GPW17] one limitation of many of the above results is that the gadget size is logarithmic in the arity of the outer function. This limits their applicability, as for several important composed functions, including the ones listed above, the gadget size is constant. In this paper, we make progress in this front. We prove some of the results mentioned above for the special case of one-way protocols, with an inner-product gadget whose number of inputs is at least some large enough constant. Our proofs make use of the simplicity and structure of one-way communication protocols, and avoid the machinery used in the earlier lifting theorems for general communication protocols. To the best of our knowledge our results are new.

Let $\mathsf{R}^{\rightarrow}_{\mathrm{cc},\epsilon}(\cdot)$ and $\mathsf{D}^{\rightarrow}_{\mathrm{cc}}(\cdot)$ denote the randomized ϵ-error and deterministic one-way communication complexity respectively. Let $\mathsf{R}^{\rightarrow}_{\mathrm{dt},\epsilon}(\cdot)$ and $\mathsf{D}^{\rightarrow}_{\mathrm{dt}}(\cdot)$ denote the randomized ϵ-error and the deterministic non-adaptive query complexity respectively. Let f denote the outer function, and $f^{\langle\rangle}$ denote the function $f \circ \mathsf{IP}^n$, where IP stands for the inner product function on $2b$ bits. Our first result shows that if f is a total function, the randomized bounded-error one-way communication complexity of $f^{\langle\rangle}$ is $\Omega(bn)$. Let $\mathrm{H}_{\mathsf{bin}}(\cdot)$ stand for the binary entropy function.

Theorem 1. *Let $f : \{0,1\}^n \to \{0,1\}$ be a total Boolean function that depends on all its inputs (i.e., it is not a junta on a strict subset of its inputs), and let $\epsilon \in (0, 1/2)$. Then,*

$$\mathsf{R}^{\rightarrow}_{\mathrm{cc},\epsilon}(f^{\langle\rangle}) \geq (1 - \mathrm{H}_{\mathsf{bin}}(\epsilon))n(b-1).$$

Coupled with the upper bound $\mathsf{R}^{\rightarrow}_{\mathrm{cc},\epsilon}(f^{\langle\rangle}) \leq b \cdot \mathsf{R}^{\rightarrow}_{\mathrm{dt},\epsilon}(\cdot)$, Theorem 1 yields a

proof of the fact that for a total Boolean function $f : \{0,1\}^n \to \{0,1\}$ depending on all input bits and $\epsilon \leq 1/3$, $\mathsf{R}^{\rightarrow}_{\mathrm{dt},\epsilon}(f) = \Omega(n)$. However, it can also be verified directly.

Theorem 1 is proved by bounding the information complexity of any correct randomized one-way protocol with respect to a suitably chosen distribution on Alice's inputs. Our proof relies crucially on the totality of f. In particular, our proof makes use of the observation that if a total function depends on an input variable, then there is an input on which the variable is sensitive, i.e., flipping the variable flips the value of the function. This is not necessarily true for partial functions; if all the valid inputs of a partial function are far from one another in Hamming distance, then flipping a single variable at a valid input always leads to an invalid input.

Note that Theorem 1 is useful only when $b > 1$. Indeed, the statement is false for $b = 1$ when f is the AND function on n bits.

Our second result relates the deterministic one-way communication complexity of $f^{\langle\rangle}$ to the deterministic non-adaptive query complexity of f, where f is an arbitrary partial Boolean function.

Theorem 2. *Let $\mathsf{S} \subseteq \{0,1\}^n$ be arbitrary, and $f : \mathsf{S} \to \{0,1\}$ be a* partial *Boolean function. Let $b \geq 4$. Then,*

$$\mathsf{D}^{\rightarrow}_{\mathrm{cc}}(f^{\langle\rangle}) = \Omega(b \cdot \mathsf{D}^{\rightarrow}_{\mathrm{dt}}(f)).$$

The proof of Theorem 2 is combinatorial. The following claim which follows directly from the work of Frankl and Tokushige [FT99] is a crucial ingredient in our proof.

Claim 1 (Theorem 2 in [FT99])**.** *Let $q \geq 8$. Let $\mathcal{A} \subseteq [q]^n$ be such that $\forall x^{(1)} = (x^{(1)}_1, \ldots, x^{(1)}_n)$, $x^{(2)} = (x^{(2)}_1, \ldots, x^{(2)}_n) \in \mathcal{A}$, $|\{i \in [n] \mid x^{(1)}_i = x^{(2)}_i\}| \geq d$. Then, $|\mathcal{A}| < q^{n-\frac{d}{2}}$.*

We give the details of the derivation of Claim 1 from the result of Frankl and Tokushige in Appendix A.

Claim 1 admits simple proofs when q is large compared to n. See [GMWW17] for a proof when there is a finite field of size q, and $q \geq n$. Their proof is based on polynomials. We give a different proof for all $q > (\frac{en}{d})^2$ in Appendix B[1]. However, such statements will only enable us to prove a lifting theorem for a gadget of size $b = \Omega(\log n)$. To prove Theorem 2 for constant-sized gadgets we need to set q to $O(1)$.

Given a protocol Π, our proof extracts a set of variables of cardinality linear in the complexity of Π, whose values always determine the value of f.

Remark: An analogous lifting theorem for deterministic one-way protocols for total outer functions follows as a special case of both Theorem 1 and Theorem 2. However, the statement admits a simple and direct proof based on a fooling set argument.

2 Preliminaries

Let $\mathsf{S} \subseteq \{0,1\}^n$ be an arbitrary subset of the Boolean hypercube, and let $f : \mathsf{S} \to \{0,1\}$ be a partial Boolean function. If $\mathsf{S} = \{0,1\}^n$, f is said to be a

[1]For every $\delta > 0$, the proof can be extended to work for all $q > \Omega(n/d)^{1+\delta}$

total Boolean function. For $b > 1$ and $x, y \in \{0,1\}^b$ define $\mathsf{IP}(x,y)$ to be the inner-product between x and y, which is $\sum_{i=1}^{b} x_i y_i \mathsf{mod} 2$. For $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n) \in \left(\{0,1\}^b\right)^n$, define $f^{\langle\rangle}(x,y) := f(\langle x_1, y_1\rangle, \ldots, \langle x_n, y_n\rangle)$. If f is a partial function, so is $f^{\langle\rangle}$. $[n]$ denotes the set $\{1, \ldots, n\}$.

2.1 Query and Communication Complexity

Let $\mathcal{U} \subseteq \{0,1\}^n \times \{0,1\}^n$. Let $F : \mathcal{U} \to \{0,1\}$ be a partial Boolean function. Let the inputs to Alice and Bob be $x \in \{0,1\}^n$ and $y \in \{0,1\}^n$ respectively. In a deterministic one-way communication protocol, Alice sends a message $m(x)$ to Bob. Then Bob outputs a bit depending on $m(x)$ and y. If Π is a correct protocol for F, then for every $(x,y) \in \mathcal{U}$, the output of Bob equals $F(x,y)$. The complexity of the protocol is the maximum number of bits a message contains for any input x to Alice. In a randomized one-way protocol, the parties share some common random bits $\mathcal{R}$. Alice's message is a function of x and $\mathcal{R}$. Bob's output is a function of $m(x), y$ and $\mathcal{R}$. Π is said to compute F with error $\epsilon \in (0, 1/2)$ if for every $(x,y) \in \mathcal{U}$, the probability over $\mathcal{R}$ of the event that Bob's output equals $F(x,y)$ is at least $1 - \epsilon$. The complexity of the protocol is the maximum number of bits contained in Alice's message for any x and $\mathcal{R}$.

The deterministic (resp. randomized-bounded error) one-way communication complexity of F, denoted by $\mathsf{R}^{\rightarrow}_{\mathrm{cc},\epsilon}(\cdot)$ (resp. $\mathsf{D}^{\rightarrow}_{\mathrm{cc}}(\cdot)$), is the minimum complexity of any deterministic (resp. ϵ-error randomized) one-way communication protocol for F.

For a partial Boolean function $f : \mathsf{S} \to \{0,1\}^n$, the deterministic non-adaptive query complexity $\mathsf{D}^{\rightarrow}_{\mathrm{dt}}(f)$ is the minimum integer k such that the following is true: there exist k indices $i_1, \ldots, i_k \in [n]$, such that for every Boolean assignment $a_{i_1}, \ldots, a_{i_k}$ to the input variables $x_{i_1}, \ldots, x_{i_k}$, f is constant on $\mathsf{S} \cap \{x \in \{0,1\}^n \mid \forall j = 1, \ldots, k, x_{i_j} = a_{i_j}\}$. It is easy to see that if f is a total function that depends on all input variables, then $\mathsf{D}^{\rightarrow}_{\mathrm{dt}}(f) = n$.

The ϵ-error randomized non-adaptive query complexity $\mathsf{R}^{\rightarrow}_{\mathrm{dt},\epsilon}(f)$ of f is the minimum integer k such that the following is true: There exists a distribution $\mathcal{D}$ on k-tuples of indices in $[n]^k$ and a function $h : [n]^k \times \{0,1\}^k \to \{0,1\}$ such that for each x in S,
$\mathbb{P}_{(i_1,\ldots,i_k)\sim\mathcal{D}}[h(i_1, \ldots, i_k, x_{i_1}, \ldots, x_{i_k}) = f(x)] \geq 1 - \epsilon$. It can be verified that if f is a total Boolean function depending on all its variables and $\epsilon \leq 1/3$, then $\mathsf{R}^{\rightarrow}_{\mathrm{dt},\epsilon}(f) = \Omega(n)$.

2.2 Information Theory

Let X be a random variable supported on a finite set $\{x_1, \ldots, x_s\}$. Let $\mathcal{E}$ be any event in the same probability space. Let $\mathbb{P}[\cdot]$ denote the probability of any event. The *conditional entropy* $\mathrm{H}(X \mid \mathcal{E})$ of X conditioned on $\mathcal{E}$ is defined as follows.

Definition 1 (Conditional entropy)**.**

$$\mathrm{H}(X \mid \mathcal{E}) := \sum_{i=1}^{s} \mathbb{P}[X = x_i \mid \mathcal{E}] \log_2 \frac{1}{\mathbb{P}[X = x_i \mid \mathcal{E}]}.$$

An important special case is when $\mathcal{E}$ is the entire sample space. In that case the above conditional entropy is referred to as the entropy $\mathrm{H}(X)$ of X.

Definition 2 (Entropy)**.**

$$\mathrm{H}(X) := \sum_{i=1}^{s} \mathbb{P}[X = x_i] \log_2 \frac{1}{\mathbb{P}[X = x_i]}.$$

Let Y be another random variable in the same probability space as X, taking values from a finite set $\{y_1, \ldots, y_t\}$. Then the conditional entropy of X conditioned on Y, $\mathrm{H}(X \mid Y)$, is defined as follows.

Definition 3.

$$\mathrm{H}(X \mid Y) = \sum_{i=1}^{t} \mathbb{P}[Y = y_i] \cdot \mathrm{H}(X \mid Y = y_i).$$

Definition 4 (Binary entropy)**.** *For $p \in (0, 1)$, the binary entropy of p, $\mathrm{H}_{\mathsf{bin}}(p)$, is defined to be the Shannon entropy of a random variable taking two distinct values with probabilities p and $1 - p$.*

$$\mathrm{H}_{\mathsf{bin}}(p) := p \log_2 \frac{1}{p} + (1 - p) \log \frac{1}{1 - p}.$$

$\mathrm{H}_{\mathsf{bin}}(p)$ is a concave function of p. When p is a random parameter, Jensen's inequality implies that

$$\mathsf{E}[\mathrm{H}_{\mathsf{bin}}(p)] \leq \mathrm{H}_{\mathsf{bin}}(\mathsf{E}[p]). \tag{1}$$

The following properties of entropy and conditional entropy will be useful.

Fact 5. (1) *Let X be a random variable supported on a finite set $\mathcal{A}$, and let Y be another random variable in the same probability space. Then $0 \leq \mathrm{H}(X \mid Y) \leq \mathrm{H}(X) \leq \log_2 |\mathcal{A}|$.*

(2) (Sub-additivity of conditional entropy). *Let $X_1, \ldots, X_n$ be n jointly distributed random variables in some probability space, and let Y be another random variable in the same probability space, all taking values in finite domains. Then,*

$$\mathrm{H}(X_1, \ldots, X_n \mid Y) \leq \sum_{i=1}^{n} \mathrm{H}(X_i \mid Y).$$

(3) *Let the random variables $X_1, \ldots, X_n$ are independent conditioned on each value of another random variable Y. Then,*

$$\mathrm{H}(X_1, \ldots, X_n \mid Y) = \sum_{i=1}^{n} \mathrm{H}(X_i \mid Y).$$

Definition 6 (Mutual information)**.** *Let X, Y and Z be two random variables in the same probability space, taking values from finite sets. The mutual information between X and Y conditioned on Z, $I(X; Y \mid Z)$, is defined as follows.*

$$\mathrm{I}(X; Y \mid Z) := \mathrm{H}(X \mid Z) - \mathrm{H}(X \mid Y, Z).$$

It can be shown that $\mathrm{I}(X; Y \mid Z)$ is symmetric in X and Y: $\mathrm{I}(X; Y \mid Z) = \mathrm{I}(Y; X \mid Z) = \mathrm{H}(Y \mid Z) - \mathrm{H}(Y \mid X, Z)$.

The following observation follows immediately from Fact 5 (1).

Observation 7. *Let X, Y and Z be random variables, and let $\mathcal{A}$ be the support of X. Then* $\mathrm{I}(X;Y \mid Z) \leq \log_2 |\mathcal{A}|$.

Theorem 8 (Fano's inequality)**.** *Let X and Y be random variables in a probability space. Let f be a function, and define $\hat{X} := f(Y)$. Let $\mathbb{P}[\hat{X} \neq X] = \epsilon$. Then,*

$$\mathrm{H}[X \mid Y] \leq \mathrm{H}_{\mathsf{bin}}(\epsilon) + \epsilon \cdot \log_2(|\mathcal{X}| - 1)$$

Where $\mathcal{X}$ is the support of X. If X is a binary random variable, then $\mathrm{H}(X \mid Y) \leq H_{\mathsf{bin}}(\epsilon)$.

3 Randomized one-way communication for total outer function

In this section, we prove Theorem 1 (restated below).

Theorem 1. *Let $f : \{0,1\}^n \to \{0,1\}$ be a total Boolean function that depends on all its inputs (i.e., it is not a junta on a strict subset of its inputs), and let $\epsilon \in (0, 1/2)$. Then,*

$$\mathsf{R}^{\rightarrow}_{\mathrm{cc},\epsilon}(f^{\langle\rangle}) \geq (1 - \mathrm{H}_{\mathsf{bin}}(\epsilon))n(b-1).$$

Proof. Let Π be an optimal randomized one-way protocol for $f^{\langle\rangle}(\cdot,\cdot)$ with error probability at most ϵ. Thus the cost of Π is $c_\Pi = \mathsf{R}^{\rightarrow}_{\epsilon}(f^{\langle\rangle})$. Π will also denote the random message sent by Alice. We will denote the public randomness of Π by $\mathcal{R}$. Let μ stand for the following distribution of Alice's input x: For each $i \in [n]$ independently, pick $x^{(i)}$ uniformly at random from $\{1\} \times \{0,1\}^{b-1}$.

For $i = 1, \ldots, n$, let $x^{(i)} = (x^{(i)}_1, \ldots, x^{(i)}_b)$. In the next lemma, we show that when Alice's input $x = (x^{(1)}, \ldots, x^{(n)})$ is sampled from μ, Π contains $\Omega(1)$ bits of information about each bit $x^{(i)}_j$ of x with $j > 1$ (recall that for each i, $x^{(i)}_1$ is 1 with probability 1).

Lemma 9. *Let x be distributed as per μ. Then for each $i = 1, \ldots, n$ and $j = 2, \ldots, b$,*

$$\mathrm{I}(\Pi; x^{(i)}_j \mid \mathcal{R}) \geq (1 - \mathrm{H}_b(\epsilon)).$$

We first finish the proof assuming Lemma 9. We have that,

$$\begin{aligned}
\mathsf{R}^{\rightarrow}_{\epsilon}(f^{\langle\rangle}) &= c_\Pi \quad \text{(By the optimality of } \Pi\text{)} \\
&\geq \mathrm{I}(x : \Pi \mid \mathcal{R}) \quad \text{(By Observation 7)} \\
&= \mathrm{H}(x \mid R) - \mathrm{H}(x \mid \Pi, \mathcal{R}) \\
&= \sum_{\substack{i=1,\ldots,n \\ j=2,\ldots,b}} \mathrm{H}(x^{(i)}_j \mid \mathcal{R}) - \mathrm{H}(x \mid \Pi, \mathcal{R}) \\
&\text{(since } x^{(i)}_j\text{'s are independent conditioned on } \mathcal{R} \text{ (see Fact 5(3)), and } x^{(i)}_1\text{'s are 0)} \\
&\geq \sum_{\substack{i=1,\ldots,n \\ j=2,\ldots,b}} \mathrm{H}(x^{(i)}_j \mid \mathcal{R}) - \sum_{\substack{i=1,\ldots,n \\ j=2,\ldots,b}} \mathrm{H}(x^{(i)}_j \mid \Pi, \mathcal{R}) \\
&\text{(By sub-additivity of entropy (see fact 5(2)) and since } \mathrm{H}(x^{(1)}_i) = 0 \text{ for each } i\text{)}
\end{aligned}$$

$$= \sum_{\substack{i=1,\ldots,n \\ j=2,\ldots,b}} \mathrm{I}(\Pi; x_j^{(i)} \mid \mathcal{R})$$
$$\geq (1 - \mathrm{H}_b(\epsilon))n(b-1). \quad \text{(By lemma 9)}$$

We now prove Lemma 9.

Proof of Lemma 9. Fix $i \in \{1, \ldots, n\}, j \in \{2, \ldots, n\}$. Recall that f depends on all its n inputs. Thus there is an assignment to all inputs other than the i-th input, on which f is sensitive to the i-th input bit. Formally, there exists $z^{(-i)} \in \{0,1\}^{[n]\setminus\{i\}}$ such that $f(0, z^{(-i)}) \neq f(1, z^{(-i)})$, i.e., $f(a, z^{(-i)})$ is the function a or $\overline{a}$. Without loss of generality, assume that $f(a, z^{(-i)}) = a$ (the other case is similar).

We will analyse Π when Bob is given the following input y.

$$y_\ell^{(k)} = \begin{cases} 0 & (\text{if } \ell > 1 \text{ and } k \neq i), \\ z_k^{(-i)} & (\text{if } \ell = 1 \text{ and } k \neq i), \\ 0 & (\text{if } \ell \neq j \text{ and } k = i), \\ 1 & (\text{if } \ell = j \text{ and } k = i). \end{cases}$$

Since Π outputs correctly with probability at least $1 - \epsilon$ on every input to Alice and Bob, Π outputs correctly with probability at least $1 - \epsilon$ when Alice's input x is sampled from μ and Bob's input y is as described above. However, under this distribution, $f^{\langle\rangle}(x, y) = f(z^{(-i)}, x_j^{(i)}) = x_j^{(i)}$. Note that Bob's output depends only on Bob's input, Alice's message and the randomness $\mathcal{R}$. For a fixing r of $\mathcal{R}$, let ϵ_r be the probability that Π is correct when $\mathcal{R} = r$, x is sampled from μ, and y is as described above. In that case a fixed function of the message Π equals $f^{\langle\rangle}(x, y) = x_j^{(i)}$ with probability at least $1 - \epsilon_r$. *Fano's inequality* (Theorem 8) implies that,

$$\mathrm{H}(x_j^{(i)} \mid \Pi, \mathcal{R} = r]) \leq \mathrm{H}_{\mathsf{bin}}(\epsilon_r).$$

Taking expectation over r on both sides we have,

$$\begin{aligned} \mathrm{H}(x_j^{(i)} \mid \Pi, \mathcal{R}]) &\leq \mathsf{E}_r[\mathrm{H}_{\mathsf{bin}}(\epsilon_r)] \\ &\leq \mathrm{H}_{\mathsf{bin}}(\mathsf{E}_r[\epsilon_r]) \quad \text{(By (1))} \\ &\leq \mathrm{H}_{\mathsf{bin}}(\epsilon). \quad \text{(Since } \mathbb{E}_r[\epsilon_r] \leq \epsilon < 1/2) \end{aligned} \tag{2}$$

The Lemma follows from (2) and the observation that $\mathrm{H}(x_j^{(i)} \mid \mathcal{R}) = 1$. □

□

4 Deterministic one-way communication for partial outer function

In this section we prove Theorem 2 (restated below).

Theorem 2. *Let $\mathsf{S} \subseteq \{0,1\}^n$ be arbitrary, and $f : \mathsf{S} \to \{0,1\}$ be a* partial *Boolean function. Let $b \geq 4$. Then,*

$$\mathsf{D}^{\rightarrow}_{\mathrm{cc}}(f^{\langle\rangle}) = \Omega(b \cdot \mathsf{D}^{\rightarrow}_{\mathrm{dt}}(f)).$$

We restate Claim 1 here.

Claim 1 (Theorem 2 in [FT99])**.** *Let $q \geq 8$. Let $\mathcal{A} \subseteq [q]^n$ be such that $\forall x^{(1)} = (x_1^{(1)}, \ldots, x_n^{(1)})$, $x^{(2)} = (x_1^{(2)}, \ldots, x_n^{(2)}) \in \mathcal{A}$, $|\{i \in [n] \mid x_i^{(1)} = x_i^{(2)}\}| \geq d$. Then, $|\mathcal{A}| < q^{n-\frac{d}{2}}$.*

We show how to derive Claim 1 from the result of Frankl and Tokushige in Appendix A. Now we proceed to the proof of Theorem 2.

Proof of Theorem 2. Leq $q := 2^b - 1$. Let Π be an optimal one-way deterministic protocol for $f^{\langle\rangle}$ of complexity $\mathsf{D}^{\rightarrow}_{\mathrm{cc}}(f^{\langle\rangle}) := c \log_2 q$. Π induces a partition of $\{0,1\}^{nb}$ into at most q^c parts; each part corresponds to a distinct message. There are $(2^b - 1)^n = q^n$ inputs $(x_1, \ldots, x_n)$ to Alice such that for each i, $x_i \neq 0^b$. Let $\mathcal{Z}$ be the set of those inputs. Identify $\mathcal{Z}$ with $[q]^n$. By the *pigeon-hole principle* there exists one part P in the partition induced by Π that contains at least q^{n-c} strings in $\mathcal{Z}$. Claim 1 (note that the assumption $b \geq 4$ implies that $q \geq 8$) implies that there are two strings $x_1 = (x_1^{(1)}, \ldots, x_1^{(n)}), x_2 = (x_1^{(2)}, \ldots, x_2^{(n)}) \in \mathsf{P} \cap \mathcal{Z}$ such that $|\{i \in [n]\} \mid x_1^{(i)} = x_2^{(i)}\}| < 2c$. Let $\mathcal{I} := \{i \in [n]\} \mid x_1^{(i)} = x_2^{(i)}\}$. Let $z = (z^{(1)}, \ldots, z^{(n)})$ denote a generic input to f. We claim that for each Boolean assignment $(a^{(i)})_{i \in \mathcal{I}}$ to the variables in $\mathcal{I}$, f is constant on $\mathsf{S} \cap \{z : \forall i \in \mathcal{I}, z^{(i)} = a^{(i)}\}$. This will prove the theorem, since querying the variables $\{z^{(i)} \mid i \in \mathcal{I}\}$ determines f; thus $D^{\rightarrow}_{\mathrm{dt}}(f) \leq |\mathcal{I}| < 2c$. Towards a contradiction, assume that there exist $z_1, z_2 \in \mathsf{S} \cap \{z : \forall i \in \mathcal{I}, z^{(i)} = a^{(i)}\}$ such that $f(z_1) \neq f(z_2)$. We will construct a string $y = (y^{(1)}, \ldots, y^{(n)}) \in \{0,1\}^{nb}$ in the following way:

$i \in \mathcal{I}$: Choose $y^{(i)}$ such that $\mathsf{IP}(y^{(i)}, x_1^{(i)}) = \mathsf{IP}(y^{(i)}, x_2^{(i)}) = a^{(i)}$.

$i \neq \mathcal{I}$: Choose $y^{(i)}$ such that $\mathsf{IP}(y^{(i)}, x_1^{(i)}) = z_1^{(i)}$ and $\mathsf{IP}(y^{(i)}, x_2^{(i)}) = z_2^{(i)}$.

Note that we can always choose a y as above since for each $i \in [n]$, $x_1^{(i)}, x_2^{(i)} \neq 0^b$, and for each $i \notin \mathcal{I}$, $x_1^{(i)} \neq x_2^{(i)}$. By the above construction, $f^{\langle\rangle}(x_1, y) = f(z_1)$ and $f^{\langle\rangle}(x_2, y) = f(z_2)$. Since by assumption $f(z_1) \neq f(z_2)$, we have that $f^{\langle\rangle}(x_1, y) \neq f^{\langle\rangle}(x_2, y)$. But since Alice sends the same message on inputs x_1 and x_2, Π produces the same output on (x_1, y) and (x_2, y). This contradicts the correctness of Π. □

Acknowledgements: I thank Prahladh Harsha and Jaikumar Radhakrishnan for pointing out the reference [FT99].

This material is based on research supported by the Singapore National Research Foundation under NRF RF Award No. NRF-NRFF2013-13.

References

[AGJ+17] Anurag Anshu, Naresh B. Goud, Rahul Jain, Srijita Kundu, and Priyanka Mukhopadhyay. Lifting randomized query complexity to randomized communication complexity. *arXiv:1703.07521v4 [cs.CC]*, 2017.

[CKLM17] Arkadev Chattopadhyay, Michal Koucký, Bruno Loff, and Sagnik Mukhopadhyay. Simulation theorems via pseudorandom properties. *CoRR*, abs/1704.06807, 2017.

[FT99] Peter Frankl and Norihide Tokushige. The erdos-ko-rado theorem for integer sequences. *Combinatorica*, 19(1):55–63, 1999.

[GLM+15] Mika Göös, Shachar Lovett, Raghu Meka, Thomas Watson, and David Zuckerman. Rectangles are nonnegative juntas. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 257–266, 2015.

[GMWW17] Dmitry Gavinsky, Or Meir, Omri Weinstein, and Avi Wigderson. Toward better formula lower bounds: The composition of a function and a universal relation. *SIAM J. Comput.*, 46(1):114–131, 2017.

[GPW15] Mika Göös, Toniann Pitassi, and Thomas Watson. Deterministic communication vs. partition number. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 1077–1088, 2015.

[GPW17] Mika Göös, Toniann Pitassi, and Thomas Watson. Query-to-communication lifting for BPP. *CoRR*, abs/1703.07666, 2017.

[HHL16] Hamed Hatami, Kaave Hosseini, and Shachar Lovett. Structure of protocols for XOR functions. In *IEEE 57th Annual Symposium on Foundations of Computer Science, FOCS 2016, 9-11 October 2016, Hyatt Regency, New Brunswick, New Jersey, USA*, pages 282–288, 2016.

[KMR17] Pravesh K. Kothari, Raghu Meka, and Prasad Raghavendra. Approximating rectangles by juntas and weakly-exponential lower bounds for LP relaxations of csps. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 590–603, 2017.

[RM99] Ran Raz and Pierre McKenzie. Separation of the monotone NC hierarchy. *Combinatorica*, 19(3):403–435, 1999.

[She11] Alexander A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011.

A Derivation of Claim 1

Let q be as in the statement of the claim. For $x \in \{0,1\}^n$ and $\mathsf{S} \subseteq [n]$, let x_{S} denote the restriction of x to the indices in S. Let $|x|$ denote the Hamming weight of x, which is $|\{i \in [n] \mid x_i = 1\}|$.

For an arbitrary alphabet L, a set $\mathcal{H} \subseteq L^n$ is called d-intersecting if for each $x = (x_i)_{i\in[n]}, x' = (x'_i)_{i\in[n]} \in \mathcal{H}$, $|\{i \in [n] \mid x_i = x'_i\}| \geq d$. Let $\mathsf{agr}(d,q,n)$ denote the size of a largest d-intersecting set in $[q]^n$. Frankl and Tokushige determined $\mathsf{agr}(d,q,n)$ in their work.

For an integer $r \leq (n-d)/2$, Let $\mathcal{A}_r$ be the following d-intersecting family in $\{0,1\}^n$.

$$A_r := \{x \in \{0,1\}^n \mid |x_{\{1,\ldots,d+2r\}}| \geq d + r\}.$$

Now consider the following d-intersecting family $\mathcal{B}_r$ in $[q]^n$: A string $x \in [q]^n$ belongs to $\mathcal{B}_r$ iff there exists a string $z \in \mathcal{A}_r$ such that for each $i \in [n]$, $z_i = 1 \Rightarrow x_i = 1$. $\mathcal{B}_r$ is easily seen to be d-intersecting. Hence for each such r, $\mathsf{agr}(d, q, n) \geq |\mathcal{B}_r|$.

Frankl and Tokushige showed that in fact there is a choice of r for which $\mathsf{agr}(d, q, n) = |\mathcal{B}_r|$. In other words, there exists a choice of r such that $\mathcal{B}_r$ is a largest d-intersecting family in $[q]^n$.

Theorem 10 (Frankl and Tokushige [FT99])**.** *Let $q \geq 3, r = \lfloor \frac{d-1}{q-2} \rfloor$ and $n \geq d + 2r$. Then, $\mathsf{agr}(d, q, n) = |\mathcal{B}_r|$.*

Proving Claim 1 now amounts to estimating $|\mathcal{B}_r|$. A string in $\mathcal{B}_r$ can be generated as follows.

- choose a subset $T \in [d + 2r]$ of size $d + r$.
- For each $i \in T$, set $x_i = 1$.
- For each $i \notin T$, set x_i arbitrarily.

There are $\binom{d+2r}{d+r}$ choices of T. For each choice of T, there are q^{n-d-r} ways of assigning variables with indices outside T. We thus have,

$$\begin{aligned}
|\mathcal{B}_r| &\leq \binom{d+2r}{d+r} \cdot q^{n-d-r} \\
&\leq \left(\frac{e(d+2r)}{d+r}\right)^{d+r} \cdot q^{n-d-r} \\
&= e^{d+r} \cdot \left(1 + \frac{r}{d+r}\right)^{d+r} \cdot q^{n-d-r} \\
&\leq e^{d+2r} \cdot q^{n-d-r} \quad \text{(Since } 1 + z \leq e^z \text{ for all real } z\text{)} \\
&= q^{n-d(1-\frac{1}{\log_e q})-r(1-\frac{2}{\log_e q})} \qquad (3)
\end{aligned}$$

By the assumption $q \geq 8$, we have that $1 - \frac{2}{\log_e q} > 0$ and $1 - \frac{1}{\log_e q} > \frac{1}{2}$. Thus from (3) we have,

$$|\mathcal{B}_r| < q^{n-\frac{d}{2}}.$$

B Proof of Claim 1 for $q = \Omega(\frac{n}{d})^2$

In this section we give a self-complete and simple proof of the statement of Claim 1 for the special case of $q > (en/d)^2$ (with a worse constant).

Let $\mathcal{X} \subseteq [q]^n$ be such that every $x, x' \in \mathcal{X}$ agree in at least d locations. Observe that each pair (x, x') can be uniquely specified by,

- A set $P_{x,x'} \subseteq [n]$ of indices of size d such that $x_i = x'_i$ for each $i \in P_{x,x'}$.
- A vector $\mathbf{a} = (a_i)_{i \in P_{x,x'}} \in [q]^d$. $\mathbf{a}$ represents $x_{P_{x,x'}} = x'_{P_{x,x'}}$.
- Vectors $x_{\overline{P_{x,x'}}}$ and $x'_{\overline{P_{x,x'}}}$.

Thus the number of pairs (x, x') is at most the number of such representations, which is upper bounded by $\binom{n}{d} \cdot q^d \cdot q^{2(n-d)} \leq (en/d)^d \cdot q^{2n-d} < q^{2n-\frac{d}{2}}$ (since $q > (\frac{en}{d})^2$). Thus $|X|^2 < q^{2n-\frac{d}{2}} \Rightarrow |X| < q^{n-\frac{d}{4}}$.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il