Electronic Colloquium on Computational Complexity, Report No. 132 (2014)

Information Complexity for Multiparty Communication*

Diptarka Chakraborty† Elazar Goldenberg‡ Michal Koucký §

October 23, 2014

Abstract

In this paper, we have studied the information complexity for the communication model involving more than two parties. A lot of work has already been done on the information complexity in two party communication model and the question of extending the definition of information complexity to the multiparty communication model was posed in [Bra12]. In this paper, we first give a definition of internal information cost for a protocol involving more than two parties and our definition matches the definition known for two party model. Our definition is valid for both in NIH model and NOF model. We also extend several results known for information complexity of two party model to multiparty communication setting. One of them is the additivity property, which eventually gives us the direct sum theorem for both information cost and distributional information cost. We give a lower bound for the information complexity of a function involving more that two parties in terms of communication complexity and this lower bound matches with the bound known for the function evaluated by only two parties. We also show that the amortized communication complexity of a function computed by k parties is lower bounded by the information complexity and upper bounded by $(k-1)$ times the information complexity and this relation is true for both distributional and non-distributional case.

1 Introduction

The study of information theory was initiated by Shannon [Sha48] to answer the questions in the areas of data compression and transmission. It is known from Shannon's noiseless coding theorem that over a noiseless channel, the cost of transmission of a message X is closely related to the entropy $H(X)$. This result establishes a connection between the communication cost of sending a message from one party to another with the information content of that message. Later, Slepian-Wolf [SW73] showed that *amortized* communication cost of transmitting a message is equal to the conditional entropy. These two results act as initial stepping stones towards the connection between the domain of communication complexity and information theory .

Over the last decade, information-theoretic techniques have been appeared as one of the main tolls to prove lower bounds of various problems in different communication settings. In case of one

*Research supported in part by Research-I Foundation and (FP7/2007-2013)/ERC Consolidator grant LBCAD no. 616787. A part of the research was done while the first author was visiting Charles University in Prague.

†Indian Institute of Technology Kanpur diptarka@cse.iitk.ac.in

‡Charles University in Prague elazargo@iuuk.mff.cuni.cz

§Charles University in Prague koucky@iuuk.mff.cuni.cz

ISSN 1433-8092

simultaneous round of communication model, to prove a *direct sum* theorem, which is to determine the relationship between the amortized communication complexity and the communication complexity of the function, Chakrabarti *et. al.* [CSWY01] developed the notion of *external information cost* of a protocol. Later that was used in [BJKS04] to show a linear lower bound of the two-party disjointness function. Similar type of information-theoretic techniques can also be found in several recent results [DW07], [JKR09], [LS09], [JKZ10]. The information cost of a protocol over an a-priori input distribution for two-party communication model was mentioned explicitly in [BBCR10], which also provides us the definition of the information complexity of a function computed by two parties. It was shown that for two-party communication model, *internal information cost* of a protocol over an input distribution is always upper bounded by the external cost of that protocol and for product distribution, these two costs are equal. In [BR11], authors proved that in *distributional* case, the internal information complexity of a function involving two parties is equal to the amortized communication complexity of computing independent copies of the same function. The result was shown by providing a way to compress one round communication protocol depending on the internal information cost. Later, this result was extended in non-distributional setting [Bra12]. By [Bra12], we also have a lower bound for information complexity of a function in terms of communication complexity. Recently, Ganor *et. al.* [GKR14] showed an exponential gap between the communication complexity and information complexity by providing an explicit example and according to [Bra12], this is the largest possible.

In [Bra12], the author posed the question whether it is possible to extend the definition of internal information cost of a protocol to the multiparty communication model, in both number-in-hand and number-on-forehead model. To the best of our knowledge, the question is still open. In this paper, we give an positive answer to this question by providing with an definition of internal information cost of a communication protocol involving more than two parties. This immediately gives us the definition of information complexity of a function computed by more that two parties. Our definition is valid for the number-in-hand as well as the number-on-forehead model and is equivalent to the definition known for two party communication model. In this paper, we consider the general public and private randomness, i.e., all the parties have access to a shared tape of randomness and along with that, each party has access to its own private randomness. We show that our definition of internal information cost will always be bounded by external information cost just like two party model. Our definition obeys the *additivity* property and that leads to the direct sum theorem for both the information cost and distributional information cost. By extending the sampling lemma mentioned in [Bra12], we provide a lower bound for the information complexity of a function involving more that two parties in terms of communication complexity and this lower bound matches with the bound known for the function evaluated by only two parties. We also establish a connection between the information complexity and the amortized complexity, for distributional as well as non-distributional case. However, unlike two party communication model, for both the cases, we have only been able to show that amortized complexity will be lower bounded by information complexity and upper bounded by $(k-1)$ times the information complexity, where k denotes the number of parties involved. It is clear from the expression that for $k = 2$, the relation will become equality. Our proof technique is a straight forward generalization of that used in [BR11] and [Bra12] to show the equality between these two quantity in two party communication model.

2 Preliminaries

2.1 Information Theory

Definition 2.1 (Entropy)**.** *The* entropy *of a discrete random variable X is defined as*

$$H(X) := -\sum_x Pr[X = x] \log Pr[X = x] = -\mathbf{E}_{x\sim X}[\log Pr[X = x]].$$

The joint entropy $H(X, Y)$ *is defined to be* $-\mathbf{E}_{x\sim X, y\sim Y}[\log Pr[X = x, Y = y]]$ *and the* conditional entropy $H(Y \mid X)$ *is defined to be* $\mathbf{E}_{x\sim X}[H(Y \mid X = x)]$.

Proposition 2.1 (Chain Rule of Entropy)**.**

$$H(X, Y) = H(X) + H(Y \mid X).$$

Definition 2.2 (Mutual Information)**.** *The* mutual information *between two random variables X and Y is*

$$I(X; Y) := H(X) - H(X \mid Y) = H(Y) - H(Y \mid X).$$

Likewise, the conditional mutual information $I(X; Y \mid Z)$ *is* $H(X \mid Z) - H(X \mid YZ)$.

Similar to the Chain Rule of Entropy, we have

Proposition 2.2 (Chain Rule of Mutual Information)**.**

$$I(X_1X_2; Y \mid Z) = I(X_1; Y \mid Z) + I(X_2; Y \mid X_1Z).$$

Definition 2.3 (Relative Entropy)**.** *The* relative entropy *or* Kullback-Leibler distance *or* divergence *between two distributions P and Q is defined as*

$$\mathbf{D}(P\|Q) := \sum_x P(x) \log \frac{P(x)}{Q(x)}.$$

Proposition 2.3.

$$\mathbf{D}(P_1 \times P_2\|Q_1 \times Q_2) = \mathbf{D}(P_1\|Q_1) + \mathbf{D}(P_2\|Q_2).$$

Proposition 2.4. *Suppose X, Y and Z are three random variables in the same probability space. For every x in the support of X and z in the support of Z, let Y_z denotes $Y \mid Z = z$ and Y_{xz} denotes $Y \mid X = x, Z = z$. Then, $I(X; Y \mid Z) = \mathbf{E}_{x\sim X, z\sim Z}[\mathbf{D}(Y_{xz}\|Y_z)]$.*

In this paper, we will extensively use the following two propositions which are just the corollaries of the chain rule of mutual information and are taken from [Bra12].

Proposition 2.5. *Suppose A, B, C and D are four random variables such that $I(B; D \mid AC) = 0$, then*

$$I(A; B \mid C) \geq I(A; B \mid CD).$$

Proposition 2.6. *Suppose A, B, C and D are four random variables such that $I(B; D \mid C) = 0$, then*

$$I(A; B \mid C) \leq I(A; B \mid CD).$$

2.2 Multiparty Communication Complexity

Consider a function $f : X_1 \times X_2 \times \cdots \times X_k \to \mathbb{Z}_K$. There are k parties $P_1, P_2, \cdots, P_k$, each having unbounded computation power and their task is to evaluate f by some sort of collaboration among themselves. The communication between the parties is through broadcast which can be thought of as writing on a board, i.e., any bit sent by any party is visible to all other parties. This exchange of messages between the parties is done according to a previously fixed protocol. The protocol's task is the following:

- To determine whether to continue or not and if yes then the protocol should return the value computed by the protocol and the value should be solely determined by the information written on the board.
- If continue then the protocol should specify the party that will write down the next bit and this as well should completely be determined by the information written on the board so far.
- Whatever a party writes down on the board should be a function of the input possessed by it, the information written on the board so far. If the protocol under consideration is a private coin protocol then a party can also use the private randomness available to it to write down the next bit.

A public coin protocol is a protocol where there is a shared randomness available along with the private coin randomness accessible to each party. This shared randomness can be thought of as a random bit string written on the board before the protocol starts.

There are several models for multiparty communication. They can be broadly split into the number-in-hand (NIH) model and the number-on-forehead (NOF) model. In NIH model, the i-th party is given the input x_i and in NOF model, the i-th party knows all the inputs except x_i, which is denoted as $\overline{x_i}$. In this paper, sometimes we will use x to denote the concatenation of all the inputs, i.e., $x_1 x_2 \cdots x_k$.

Given a protocol π involving k parties, $\pi(x_1, x_2, \cdots, x_k)$ the *transcript* of that protocol, which is the concatenation of public randomness and all the messages communicated during the protocol. When rather than a specific transcript, we refer a random variable denoting a transcript, we denote it by $\Pi(x_1, x_2, \cdots, x_k)$ or just Π.

Definition 2.4 (Communication Cost)**.** *The* communication cost *of a protocol π is the number of bits written on the board for worst case input and is denoted by $CC(\pi)$.*

Definition 2.5 (Distributional Complexity)**.** *Let μ be a probability distribution on $X_1 \times X_2 \times \cdots \times X_k$. For a function $f : X_1 \times X_2 \times \cdots \times X_k \to \mathbb{Z}_K$, the* distributional complexity *of f, $D^{\mu}_{\rho}(f)$ is the communication cost of the best deterministic protocol that outputs the correct value of f on at least $1 - \rho$ fraction of all inputs in $X_1 \times X_2 \times \cdots \times X_k$.*

Definition 2.6. *The cost of the best randomized public coin protocol for computing the function f with error at most ρ on every inputs is denoted by $R_\rho(f)$.*

Theorem 2.1 (Yao's Min-Max Theorem)**.**

$$R_\rho(f) = \max_{\mu} D^{\mu}_{\rho}(f).$$

3 Information Cost for Multiparty Communication

In this section, we try to extend the definition of information complexity for two party communication protocol to multiparty setting. The notion of information complexity for two party communication was implicitly mentioned in [BJKS04] and was defined explicitly in [BBCR10].

Definition 3.1 (Internal Information Cost)**.** *Suppose there are k parties $P_1, P_2, \cdots, P_k$, where i-th party holds the input X_{P_i} and there is a function $f : X \to \mathbb{Z}_K$, where the input X is distributed according to the probability distribution μ. Then the information complexity of a protocol π is defined as*

$$IC^k_\mu(\pi) := \frac{1}{k-1} \sum_{i=1}^{k} I(\Pi; \overline{X_{P_i}} \mid X_{P_i})$$

where $\overline{X_{P_i}} = X \setminus X_{P_i}$.

It can be noted that the above definition matches the definition of internal information complexity for two party system (when $k = 2$) defined in [BBCR10]. Consider the function $f : X \to \mathbb{Z}_K$, where $X = X_1 \times X_2 \times \cdots \times X_k$. Then note that, for NIH model, $X_{P_i} = X_i$ and for NOF model, $X_{P_i} = \overline{X_i}$.
We now mention a lemma which is just an easy extension of a lemma from [BR11] and the proof is also same, hence omitted.

Lemma 3.1. *Let R be the public randomness available to all the k parties. Then*

$$IC^k_\mu(\pi) = \mathbf{E}_R[IC^k_\mu(\pi_R)].$$

The next lemma will upper bound the internal information complexity by the communication cost of a protocol. The proof is also similar to the two party case, but for the sake of clarity, we provide the proof here.

Lemma 3.2. *For any distribution μ, $IC^k_\mu(\pi) \leq CC(\pi)$.*

Proof. Let us first consider that the protocol π is a private coin protocol and let π_n be the n-th bit written on the board. Then,

$$\begin{aligned}
IC^k_\mu(\pi) &= \frac{1}{k-1} \sum_{i=1}^{k} I(\pi(X); \overline{X_{P_i}} \mid X_{P_i}) \\
&= \frac{1}{k-1} \sum_{n=1}^{CC(\pi)} \sum_{i=1}^{k} I(\pi_n; \overline{X_{P_i}} \mid \pi_1 \pi_2 \cdots \pi_{n-1} X_{P_i}) \quad \text{by Proposition 2.2} \\
&= \frac{1}{k-1} \sum_{n=1}^{CC(\pi)} \sum_{i=1}^{k} \mathbf{E}_{\gamma \in_R \pi_1 \pi_2 \cdots \pi_{n-1}} [I(\pi_n; \overline{X_{P_i}} \mid E_\gamma X_{P_i})].
\end{aligned}$$

where E_γ denotes the event that the first $n-1$ bits of communication is equal to γ. If γ is such that it is the i-th party's turn to write on board and thus $I(\pi_n; \overline{X_{P_i}} \mid X_{P_i}) = 0$ and for all $j \neq i$, $I(\pi_n; \overline{X_{P_j}} \mid X_{P_j}) \leq 1$ as π_n contains only one bit. Hence, $IC^k_\mu(\pi) \leq CC(\pi)$.
If π is allowed to use public randomness, then by Lemma 3.1, we can write

$$IC^k_\mu(\pi) = \mathbf{E}_R[IC^k_\mu(\pi_R)] \leq CC(\pi)$$

where R denotes the public randomness of π. □

3.1 Relationship with External Information Cost

We adopt the definition of *external information complexity* of a problem from [Bra12].

Definition 3.2 (External Information Cost)**.** *The external information cost of a protocol π with respect to a probability distribution μ on input X as*

$$IC_{\mu}^{ext}(\pi) := I(X;\Pi).$$

In this section, we establish a relation between it and internal information cost of a protocol π. The relation is same as that for two party communication setting, however the proof differs in essential details.

Theorem 3.1. *For any function f and distribution μ,*

$$IC_{\mu}^{k}(\pi) \leq IC_{\mu}^{ext}(\pi).$$

Proof. First consider the internal information cost,

$$\begin{aligned} IC_{\mu}^{k}(\pi) &= \frac{1}{k-1}\sum_{i=1}^{k} I(\Pi;\overline{X_{P_i}} \mid X_{P_i}) \\ &= \frac{1}{k-1}\sum_{n=1}^{CC(\pi)}\sum_{i=1}^{k} I(\Pi_n;\overline{X_{P_i}} \mid \Pi_1\Pi_2\cdots\Pi_{n-1}X_{P_i}) \quad \text{by Proposition 2.2} \end{aligned}$$

Whereas the external information cost is

$$\begin{aligned} IC_{\mu}^{ext}(\pi) &= I(X;\Pi) \\ &= \sum_{n=1}^{CC(\pi)} I(X;\Pi_n \mid \Pi_1\Pi_2\cdots\Pi_{n-1}) \quad \text{by Proposition 2.2} \end{aligned}$$

Suppose at the n-th step of communication, it was j-th party's turn. Then,

$$\begin{aligned} I(X;\Pi_n \mid \Pi_1\Pi_2\cdots\Pi_{n-1}) &= I(X_{P_j};\Pi_n \mid \Pi_1\Pi_2\cdots\Pi_{n-1}) + \\ &\quad I(\overline{X_{P_j}};\Pi_n \mid \Pi_1\Pi_2\cdots\Pi_{n-1}X_{P_j}) \quad \text{by Proposition 2.2} \\ &= I(X_{P_j};\Pi_n \mid \Pi_1\Pi_2\cdots\Pi_{n-1}). \end{aligned}$$

Now observe the following,

$$\begin{aligned} \sum_{i=1}^{k} I(\overline{X_{P_i}};\Pi_n \mid \Pi_1\Pi_2\cdots\Pi_{n-1}X_{P_i}) &= \sum_{i=1}^{k}[I(X_{P_j};\Pi_n \mid \Pi_1\Pi_2\cdots\Pi_{n-1}X_{P_i}) + \\ &\quad I(\overline{X_{P_i}} \setminus X_{P_j};\Pi_n \mid \Pi_1\Pi_2\cdots\Pi_{n-1}X_{P_i}X_{P_j})] \quad \text{by Proposition 2.2} \\ &= \sum_{i=1}^{k} I(X_{P_j};\Pi_n \mid \Pi_1\Pi_2\cdots\Pi_{n-1}X_{P_i}) \\ &\leq (k-1)I(X_{P_j};\Pi_n \mid \Pi_1\Pi_2\cdots\Pi_{n-1}) \\ &= (k-1)I(X;\Pi_n \mid \Pi_1\Pi_2\cdots\Pi_{n-1}). \end{aligned}$$

The last inequality is due to an application of Proposition 2.5, by considering $A = X_{P_j}, B = \Pi_n, C = \Pi_1\Pi_2\cdots\Pi_{n-1}$ and $D = X_{P_i}$ and also note that $k-1$ factor comes in the inequality because $I(X_{P_j}; \Pi_n \mid \Pi_1\Pi_2\cdots\Pi_{n-1}X_{P_i}) = 0$ when $i = j$. Hence, we get that,

$$IC^k_\mu(\pi) \leq IC^{ext}_\mu(\pi).$$

□

4 The prior-free Information Complexity

One of our main motivation is to make the information cost independent of any prior distribution, just like in the two party case and also we want to investigate the information cost of a given function instead of a protocol that evaluates that function. We define the *information complexity* of a function for mutiparty setting same as that for two party setting.

Definition 4.1. *Consider a function $f : X \to \{0,1\}$ and an error parameter $\epsilon \geq 0$.*

- *The k-party max-distributional information complexity is defined as*

$$IC^k_D(f,\epsilon) := \max_{\mu \text{ a distribution on } X} IC^k_\mu(f,\epsilon)$$

  where $IC^k_\mu(f,\epsilon) := \inf_{\pi: Pr_{x\sim\mu}[\pi(x)\neq f(x)]\leq\epsilon} IC^k_\mu(\pi)$.

- *The k-party prior-free informational complexity is*

$$IC^k(f,\epsilon) := \inf_{\pi \text{ is a protocol s.t. } \forall x,\, Pr[\pi(x)\neq f(x)]\leq\epsilon} \max_\mu IC^k_\mu(\pi).$$

Note that in the definition of $IC^k(f,\epsilon)$, the probability is over the public and private coin randomness used in the protocol. Clearly, $IC^k(f,\epsilon) \geq IC^k_D(f,\epsilon)$. For the opposite direction, we have the following two theorems.

Theorem 4.1. *Consider a function $f : X \to \{0,1\}$ and an error parameter $\epsilon \geq 0$. For any α, $0 < \alpha < 1$,*

$$IC^k(f, \frac{\epsilon}{\alpha}) \leq \frac{IC^k_D(f,\epsilon)}{1-\alpha}.$$

For zero-error case, the above two notions will coincide.

Theorem 4.2. *Consider a function $f : X \to \{0,1\}$. Then,*

$$IC^k(f,0) = IC^k_D(f,0).$$

The proofs of above two theorems are same as that for two party setting [Bra12] and hence we omit the details here.

Another important property of information complexity is convexity and the proof does not depend on the number of parties and hence is same as that for two party system.

Theorem 4.3 ([Bra12])**.** *For any f, the functions $IC^k(f,\epsilon)$ and $IC^k_D(f,\epsilon)$ are convex on the interval $\epsilon \in [0,1]$.*

5 The additivity of Information Complexity

A task $T(x_1, \cdots, x_k)$ is a relation $R(x_1, \cdots, x_k, O_{P_1}, \cdots, O_{P_k})$ along with a required success criterion, where $x = x_1 \cdots x_k$ is the input and O_{P_i} corresponds to the output of i-th party. Informally a task is anything that can be solved by a k-party communication protocol. The information cost of a task is defined similarly to the information cost of a function [Bra12]. Here we assume that each $x_i \in \{0,1\}^n$ and $\mathcal{D}$ be a set of distributions on $\{0,1\}^n \times \cdots \times \{0,1\}^n$.

Definition 5.1 ([Bra12])**.** *The cost of a task T with respect to a set of distributions $\mathcal{D}$ is defined as*

$$IC^k(T, \mathcal{D}) := \inf_{\pi \text{ succeeds to perform } T} \sup_{\mu \in \mathcal{D}} IC^k_\mu(\pi).$$

Let $T_1(x_1^1, \cdots, x_k^1)$ and $T_2(x_1^2, \cdots, x_k^2)$ be two tasks and $T_1 \times T_2 := T(x_1^1, \cdots, x_k^1, x_1^2, \cdots, x_k^2)$ be the task to perform both t_1 and T_2 in parallel on two set of inputs. A protocol is said to be successful on $T_1 \times T_2$ if it succeeds both on T_1 and T_2 separately.
Now lets define two products of sets of distributions. Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be two sets of distributions on $(x_1^1, \cdots, x_k^1)$ and $(x_1^2, \cdots, x_k^2)$ respectively. Denote

- $\mathcal{D}_1 \times \mathcal{D}_2 := \{\mu_1 \times \mu_2 : \mu_1 \in \mathcal{D}_1,\ \mu_2 \in \mathcal{D}_2\}$.
- $\mathcal{D}_1 \otimes \mathcal{D}_2 := \{\mu : \mu\mid_{(X_1^1, \cdots, X_k^1)} \in \mathcal{D}_1,\ \mu\mid_{(X_1^2, \cdots, X_k^2)} \in \mathcal{D}_2\}$.

Theorem 5.1. *Let $T_1(x_1^1, \cdots, x_k^1)$ and $T_2(x_1^2, \cdots, x_k^2)$ be two tasks and $\mathcal{D}_1$ and $\mathcal{D}_2$ be two sets of distributions on $(x_1^1, \cdots, x_k^1)$ and $(x_1^2, \cdots, x_k^2)$ respectively. Then for $T = T_1 \times T_2$,*

$$IC^k(T, \mathcal{D}_1 \times \mathcal{D}_2) = IC^k(T, \mathcal{D}_1 \otimes \mathcal{D}_2) = IC^k(T_1, \mathcal{D}_1) + IC^k(T_2, \mathcal{D}_2).$$

Proof. $IC^k(T, \mathcal{D}_1 \times \mathcal{D}_2) \leq IC^k(T, \mathcal{D}_1 \otimes \mathcal{D}_2)$. This follows from Definition 5.1 since $\mathcal{D}_1 \times \mathcal{D}_2 \subset \mathcal{D}_1 \otimes \mathcal{D}_2$.

$IC^k(T, \mathcal{D}_1 \otimes \mathcal{D}_2) \leq IC^k(T_1, \mathcal{D}_1) + IC^k(T_2, \mathcal{D}_2)$. Consider $\epsilon > 0$ and let π_1 and π_2 be two protocol succeeding on tasks T_1 and T_2 respectively such that for all $\mu_1 \in \mathcal{D}_1$ and $\mu_2 \in \mathcal{D}_2$,

$$IC^k_{\mu_1}(\pi_1) < IC^k(T_1, \mathcal{D}_1) + \epsilon \text{ and } IC^k_{\mu_2}(\pi_2) < IC^k(T_2, \mathcal{D}_2) + \epsilon.$$

Now consider a protocol π that on random inputs $X_1^1, \cdots, X_k^1, X_1^2, \cdots, X_k^2$ runs π_1 on $X_1^1, \cdots, X_k^1$ and π_2 on $X_1^2, \cdots, X_k^2$ independently. Clearly π succeeds on T. Let us consider a distribution $\mu \in \mathcal{D}_1 \otimes \mathcal{D}_2$. We will show that

$$IC^k_\mu(\pi) < IC^k(T_1, \mathcal{D}_1) + IC^k(T_2, \mathcal{D}_2) + 2\epsilon.$$

Suppose $\mu_1 := \mu\mid_{(X_1^1, \cdots, X_k^1)} \in \mathcal{D}_1$ and $\mu_2 := \mu\mid_{(X_1^2, \cdots, X_k^2)} \in \mathcal{D}_2$. We have that

$$\frac{1}{k-1} \sum_{i=1}^{k} I(\overline{X_{P_i}^1}; \Pi_1 \mid X_{P_i}^1) = IC^k_{\mu_1}(\pi_1) < IC^k(T_1, \mathcal{D}_1) + \epsilon.$$

and

$$\frac{1}{k-1} \sum_{i=1}^{k} I(\overline{X_{P_i}^2}; \Pi_2 \mid X_{P_i}^2) = IC^k_{\mu_2}(\pi_2) < IC^k(T_2, \mathcal{D}_2) + \epsilon.$$

Now consider the following

$$
\begin{aligned}
IC^k_\mu(\pi) &= \frac{1}{k-1}\sum_{i=1}^{k} I(\Pi; \overline{X^1_{P_i}}\,\overline{X^2_{P_i}} \mid X^1_{P_i}X^2_{P_i}) \\
&= \frac{1}{k-1}\sum_{i=1}^{k} I(\Pi_1\Pi_2; \overline{X^1_{P_i}}\,\overline{X^2_{P_i}} \mid X^1_{P_i}X^2_{P_i}) \\
&= \frac{1}{k-1}\sum_{i=1}^{k} [I(\Pi_1; \overline{X^1_{P_i}}\,\overline{X^2_{P_i}} \mid X^1_{P_i}X^2_{P_i}) + I(\Pi_2; \overline{X^1_{P_i}}\,\overline{X^2_{P_i}} \mid X^1_{P_i}X^2_{P_i}\Pi_1)] \quad \text{by Proposition 2.2} \\
&= \frac{1}{k-1}\sum_{i=1}^{k} [I(\Pi_1; \overline{X^1_{P_i}} \mid X^1_{P_i}X^2_{P_i}) + I(\Pi_1; \overline{X^2_{P_i}} \mid X^1_{P_i}X^2_{P_i}\overline{X^1_{P_i}}) \\
&+ I(\Pi_2; \overline{X^2_{P_i}} \mid X^1_{P_i}X^2_{P_i}\Pi_1) + I(\Pi_2; \overline{X^1_{P_i}} \mid X^1_{P_i}X^2_{P_i}\Pi_1\overline{X^2_{P_i}})] \quad \text{by Proposition 2.2} \\
&= \frac{1}{k-1}\sum_{i=1}^{k} [I(\Pi_1; \overline{X^1_{P_i}} \mid X^1_{P_i}X^2_{P_i}) + I(\Pi_2; \overline{X^2_{P_i}} \mid X^1_{P_i}X^2_{P_i}\Pi_1)] \\
&\leq \frac{1}{k-1}\sum_{i=1}^{k} [I(\Pi_1; \overline{X^1_{P_i}} \mid X^1_{P_i}) + I(\Pi_2; \overline{X^2_{P_i}} \mid X^2_{P_i})] \\
&= IC^k_{\mu_1}(\pi_1) + IC^k_{\mu_2}(\pi_2) < IC^k(T_1, \mathcal{D}_1) + IC^k(T_2, \mathcal{D}_2) + 2\epsilon.
\end{aligned}
$$

The first inequality is due to the application of Proposition 2.5 on two terms: in the first term, by assuming $A = \overline{X^1_{P_i}}, B = \Pi_1, C = X^1_{P_i}$ and $D = X^2_{P_i}$ and in second term, by assuming $A = \overline{X^2_{P_i}}, B = \Pi_2, C = X^2_{P_i}$ and $D = X^1_{P_i}\Pi_1$.

$IC^k(T_1, \mathcal{D}_1) + IC^k(T_2, \mathcal{D}_2) \leq IC^k(T, \mathcal{D}_1 \times \mathcal{D}_2)$. Let $\mu_1 \in \mathcal{D}_1$ and $\mu_2 \in \mathcal{D}_2$ be two distributions and $\epsilon > 0$ be an error parameter. We will show that there are protocols π_1 and π_2 that succeeds to complete tasks T_1 and T_2 respectively such that

$$IC^k_{\mu_1}(\pi_1) + IC^k_{\mu_2}(\pi_2) < IC^k(T, \mathcal{D}_1 \times \mathcal{D}_2) + \epsilon.$$

By the definition of $IC^k(T, \mathcal{D}_1 \times \mathcal{D}_2)$, there is a protocol π that succeeds to complete the task $T_1 \times T_2$ such that

$$IC^k_{\mu_1 \times \mu_2}(\pi) < IC^k(T, \mathcal{D}_1 \times \mathcal{D}_2) + \epsilon.$$

Now define the protocols π_1 and π_2 as follows:
Protocol $\pi_1(x^1_1, \cdots x^1_k)$**:**

1. Using the public randomness, the parties together sample $X^2 = X^2_1, \cdots, X^2_k$ according to μ_2.
2. The parties run $\pi(x^1_1, \cdots x^1_k, X^2_1, \cdots, X^2_k)$ and output the value of the task T_1.

Protocol $\pi_2(x^2_1, \cdots x^2_k)$**:**

1. Using the public randomness, the parties together sample $X^1 = X^1_1, \cdots, X^1_k$ according to μ_1.
2. The parties run $\pi(X^1_1, \cdots X^1_k, x^2_1, \cdots, x^2_k)$ and output the value of the task T_2.

By the definition of $T_1 \times T_2$, the protocols π_1 and π_2 succeed to complete the tasks T_1 and T_2 respectively. Now we have

$$\begin{aligned}
IC^k_{\mu_1}(\pi_1) &= \frac{1}{k-1}\sum_{i=1}^{k} I(\Pi_1; \overline{X^1_{P_i}} \mid X^1_{P_i}) \\
&= \frac{1}{k-1}\sum_{i=1}^{k} I(\Pi X^2; \overline{X^1_{P_i}} \mid X^1_{P_i}) \\
&= \frac{1}{k-1}\sum_{i=1}^{k} [I(X^2; \overline{X^1_{P_i}} \mid X^1_{P_i}) + I(\Pi; \overline{X^1_{P_i}} \mid X^1_{P_i} X^2)] \quad \text{by Proposition 2.2} \\
&= \frac{1}{k-1}\sum_{i=1}^{k} I(\Pi; \overline{X^1_{P_i}} \mid X^1_{P_i} X^2).
\end{aligned}$$

and

$$\begin{aligned}
IC^k_{\mu_2}(\pi_2) &= \frac{1}{k-1}\sum_{i=1}^{k} I(\Pi_2; \overline{X^2_{P_i}} \mid X^2_{P_i}) \\
&= \frac{1}{k-1}\sum_{i=1}^{k} I(\Pi X^1; \overline{X^2_{P_i}} \mid X^2_{P_i}) \\
&= \frac{1}{k-1}\sum_{i=1}^{k} [I(X^1; \overline{X^2_{P_i}} \mid X^2_{P_i}) + I(\Pi; \overline{X^2_{P_i}} \mid X^2_{P_i} X^2)] \quad \text{by Proposition 2.2} \\
&= \frac{1}{k-1}\sum_{i=1}^{k} I(\Pi; \overline{X^2_{P_i}} \mid X^2_{P_i} X^1).
\end{aligned}$$

Thus we get

$$\begin{aligned}
IC^k_{\mu_1}(\pi_1) + IC^k_{\mu_2}(\pi_2) &= \frac{1}{k-1}\sum_{i=1}^{k} [I(\Pi; \overline{X^1_{P_i}} \mid X^1_{P_i} X^2) + I(\Pi; \overline{X^2_{P_i}} \mid X^2_{P_i} X^1)] \\
&= \frac{1}{k-1}\sum_{i=1}^{k} [I(\Pi; \overline{X^1_{P_i}} \mid X^1_{P_i} X^2_{P_i} \overline{X^2_{P_i}}) + I(\Pi; \overline{X^2_{P_i}} \mid X^2_{P_i} X^1_{P_i} \overline{X^1_{P_i}})] \\
&\leq \frac{1}{k-1}\sum_{i=1}^{k} [I(\Pi; \overline{X^1_{P_i}} \mid X^1_{P_i} X^2_{P_i}) + I(\Pi; \overline{X^2_{P_i}} \mid X^2_{P_i} X^1_{P_i} \overline{X^1_{P_i}})] \\
&= \frac{1}{k-1}\sum_{i=1}^{k} I(\Pi; \overline{X^1_{P_i}}\, \overline{X^2_{P_i}} \mid X^1_{P_i} X^2_{P_i}) \quad \text{by Proposition 2.2} \\
&= IC^k_{\mu_1 \times \mu_2}(\pi) < IC^k(T, \mathcal{D}_1 \times \mathcal{D}_2) + \epsilon.
\end{aligned}$$

The first inequality is due to the application of Proposition 2.5 on the first term, by assuming $A = \Pi, B = \overline{X^1_{P_i}}, C = X^1_{P_i} X^2_{P_i}$ and $D = \overline{X^2_{P_i}}$.

□

Now using the last theorem, we can show the following theorem on exact direct sum of information cost and distributional information cost. The reader may refer to [Bra12] for the proof.

Theorem 5.2 ([Bra12]). *Let $f(x_1, \cdots, x_k)$ be any function and $\epsilon > 0$ an error parameter and f^n be the problem of computing f on n sets of inputs such that in each coordinate the error is bounded by ϵ. Then the following holds:*

1. $IC^k_D(f^n, \epsilon) = n.IC^k_D(f, \epsilon)$.
2. $IC^k(f^n, \epsilon) = n.IC^k(f, \epsilon)$.

6 Information Complexity vs. Communication Complexity

6.1 Extended Sampling Lemma

In this section, we prove an extended version of sampling lemma which is a strict generalization of the sampling lemma proved in [Bra12] and then using that we establish a connection between information complexity and communication complexity for multiparty setting. The proof idea is same as that for two party case. We start with a claim proved in [Bra12].

Claim 6.1 ([Bra12]). *For any two distributions μ and ν and an error parameter $\epsilon > 0$, if $D(\mu\|\nu) \leq I$, then the following holds*

$$\mu\{x : \frac{2^{I+1}}{\epsilon}.\nu(x) < \mu(x)\} < \epsilon.$$

We are now ready to state and prove the extended version of the sampling lemma.

Lemma 6.1. *Let μ be any distribution over a universe $\mathcal{U}$ and $I \geq 1$ be a parameter known to all the parties $P_1, \cdots, P_k$. Further let $\nu_{P_1}, \cdots, \nu_{P_k}$ be the distributions over $\mathcal{U}$ such that $D(\mu\|\nu_{P_i}) \leq I$, for all i, $1 \leq i \leq k$. The i-th party is given real valued functions $g_{P_i}, h_{P_i,P_j} : \mathcal{U} \to [0,1]$, for all $1 \leq j \leq k$ and $j \neq i$ such that for all $x \in \mathcal{U}$, $\mu(x) = \prod_{j=1}^k g_{P_j}(x)$ and $\nu_{P_i}(x) = g_{P_i}(x) \prod_{j \neq i} h_{P_i,P_j}(x)$. Let $\epsilon > 0$ be an error parameter. Then there is a sampling protocol $\pi(g_{P_1}, \cdots, g_{P_k}, h_{P_1,P_2}, \cdots, h_{P_k,P_{k-1}}, I, \epsilon)$ that communicates total $2^{O(1+kI/\epsilon)}$ bits such that the following holds:*

1. *at the end of the protocol, the parties output $x_i \in \mathcal{U}$;*
2. *there is an event ξ such that $\overline{\xi} \implies x_1 = x_2 = \cdots = x_k$ and $Pr[\xi] < \epsilon$;*
3. *let μ' is the distribution of x_1 conditioned on $\overline{\xi}$, then $\mid \mu - \mu' \mid < \epsilon$.*

Proof. The parties interpret the shared tape as a source of points of the form $(x_i, \alpha_{i,1}, \alpha_{i,2}, \cdots, \alpha_{i,k})$ uniformly distributed over $\mathcal{U} \times [0,1]^k$ and they consider $N = 2 \mid \mathcal{U} \mid \ln \frac{1}{\epsilon}$ such points. Their goal is to find out an index τ such that for all i, $\alpha_{\tau,i} \leq g_{P_i}(x_\tau)$. The probability that each x_i to be such x_τ is $\prod_{i=1}^k g_{P_i}(x_\tau) = \mu(x_\tau)$. Now we denote $B_{P_i} := \{x : 2^{8k(I+1)/\epsilon}.\nu_{P_i} < \mu(x)\}$ for all $1 \leq i \leq k$ and then by Claim 6.1, $\mu(B_{P_i}) < \frac{\epsilon}{8k}$ for all i. Next, note that an index t satisfies $\alpha_{t,i} \leq g_{P_i}(x_t)$ for all i is exactly $\frac{1}{|\mathcal{U}|}$. Hence,

$$Pr[\tau > N] \leq (1 - \frac{1}{|\mathcal{U}|})^N < e^{-N/|\mathcal{U}|} = e^{-2\ln 1/\epsilon} = \epsilon^2 < \epsilon/16.$$

Now consider the set of indices that are candidates to be τ for a party P_i's viewpoint

$$S_{P_i} := \{t \leq N : \alpha_{t,i} \leq g_{P_i}(x_t) \text{ and } \prod_{j \neq i} \alpha_{t,j} \leq 2^{8k(I+1)/\epsilon}.\prod_{j \neq i} h_{P_i,P_j}(x_t)\}.$$

Assuming $x_\tau \notin \bigcup_i B_{P_i}$, we claim that $\tau \in \bigcap_i S_{P_i}$. This is true because $x_\tau \notin B_{P_i}$ implies

$$\frac{\prod_{j \neq i} g_{P_j}(x_\tau)}{\prod_{j \neq i} h_{P_i,P_j}(x_\tau)} = \frac{\mu(x_\tau)}{\nu_{P_i}(x_\tau)} \leq 2^{8k(I+1)/\epsilon}.$$

Now as $\forall_j,\ \alpha_{\tau,j} \leq g_{P_j}(x_\tau)$, we have that,

$$\prod_{j \neq i} \alpha_{\tau,j} \leq \prod_{j \neq i} g_{P_j}(x_\tau) \leq 2^{8k(I+1)/\epsilon}.\prod_{j \neq i} h_{P_i,P_j}(x_\tau)$$

and hence, for all i, $\tau \in S_{P_i}$. In fact, τ is the first element in $\bigcap_i S_{P_i}$. Note that for each t, $Pr[t \in S_{P_i}] \leq \frac{2^{8k(I+1)/\epsilon}}{|\mathcal{U}|}$. Thus $\mathbf{E}[|S_{P_i}|] \leq 2^{8k(I+1)/\epsilon}.2\ln\frac{1}{\epsilon} < 2^{9Ik/\epsilon}$. Thus by Markov inequality, $Pr[|S_{P_i}| > 2^{10Ik/\epsilon}] < 2^{-Ik/\epsilon} \ll \frac{\epsilon}{16k}$.
Now consider the event $\xi_1 := [x_\tau \in \bigcup_i B_{P_i}]$ and $\xi_2 := [\tau > N \text{ or } |S_{P_i}| > 2^{10Ik/\epsilon} \text{ for some } i]$ and let $\xi := \xi_1 \bigcup \xi_2$. then by union bound, $Pr[\xi] < k.\frac{\epsilon}{8k} + \frac{\epsilon}{16} + k.\frac{\epsilon}{16k} < \frac{\epsilon}{2}$. The distribution μ' conditioned on $\bar{\xi}$ satisfies $|\mu - \mu'| < \epsilon$ as it is the distribution on $\mathcal{U} \setminus \bigcup_i B_{P_i}$. Next, we describe the sampling protocol.
Sampling Protocol:

1. Each P_i computes the set S_{P_i} and if for some i, $|S_{P_i}| > 2^{10Ik/\epsilon}$, the protocol fails.
2. For each $s \in S_{P_1}$, P_1 computes $d = \lceil \frac{10Ik^2}{\epsilon} + \log\frac{1}{\epsilon} + k + 2 \rceil$ random hash values $h'_1(s), h'_2(s), \cdots, h'_d(s)$, where the hash functions are evaluated using the public randomness.
3. P_1 writes down the values $\{h_j(s_i)\}_{s_i \in S_{P_1}, 1 \leq j \leq d}$.
4. P_2 finds the set of indices I_2 of set S_{P_1} such that for each index $i \in I_2$, there is a $s \in S_{P_2}$ for which $h'_j(s) = h'_j(s_i)$, $\forall_{j=1,2,\cdots,d}$.
5. P_2 writes down the set I_2 as a characteristic sequence of length $|S_{P_1}|$.
6. $P_3, \cdots, P_{k-1}$ do the same as P_2 such that we get $I_{k-1} \subseteq I_{k-2} \subseteq \cdots \subseteq I_2$.
7. P_k finds the first index $i \in I_{k-1}$ such that there is a $s \in S_{P_2}$ for which $h'_j(s) = h'_j(s_i)$, $\forall_{j=1,2,\cdots,d}$ and if such i exists, then writes down i and output x_s.
8. $P_1, \cdots, P_{k-1}$ output the corresponding values of x.

In the above protocol, the total number of bits communicated is bounded by $2^{10Ik/\epsilon}.d + \frac{10Ik^2}{\epsilon} = 2^{O(1+Ik/\epsilon)}$.
In the above protocol, we use the hash functions that are selected from k-universal family of hash functions constructed using the public randomness available to the protocol. For the definition and the construction of k-universal family of hash functions, interested readers may refer to [WC81]. To analyze the correctness of the above protocol, lets first observe that for $s_1 \in S_{P_1}, s_2 \in S_{P_2} \cdots, s_k \in$

S_{P_k} such that they are not all equal but still $h_j(s_1) = h_j(s_2) = \cdots = h_j(s_k)$, for all $1 \leq j \leq d$ is bounded by $(2^k - 1)2^{-d} < \frac{\epsilon}{4\prod_{i=1}^{k} |S_{P_i}|}$ for the specified value of d. Thus by union bound, there exists some non equal $s_i \in S_{P_i}$, but all the hash values match is bounded by $\frac{\epsilon}{4}$. Assuming that there is no such s_i and there is a $\tau \in \bigcap_i S_{P_i}$, the above protocol is guaranteed to find it and this completes the proof. □

6.2 Information vs. Communication

Lemma 6.1 implies the following connection between information and communication complexity.

Theorem 6.1. *Let $f : X \to \mathbb{Z}_K$ be any function where $X = X_1 \times X_2 \times \cdots \times X_k$ and let $\rho, \epsilon > 0$ be error parameters. Then,*

1. *For any distribution μ over X, $D^{\mu}_{\rho+\epsilon}(f) \leq 2^{O(1+k^3 IC^k_{\mu}(f,\rho)/\epsilon^2)}$.*
2. *$R_{\rho+\epsilon}(f) \leq 2^{O(1+k^3 IC^k(f,\rho)/\epsilon^2)}$.*

Proof. Let μ be any distribution and π be the protocol that realizes the value $I_{\mu} := IC^k_{\mu}(f, \rho)$. Now by Proposition 2.4,

$$I_{\mu} = \frac{1}{k-1} \sum_{i=1}^{k} I(\overline{X_{P_i}}; \pi_X \mid X_{P_i}) = \frac{1}{k-1} \sum_{i=1}^{k} \mathbf{E}_{x\sim\mu}[\mathbf{D}(\pi_x \| \pi_{x_{P_i}})].$$

Thus for any $1 \leq i \leq k$, $\mathbf{E}_{x\sim\mu}[\mathbf{D}(\pi_x \| \pi_{x_{P_i}})] \leq (k-1)I_{\mu}$ and so by Markov inequality,

$$Pr[\mathbf{D}(\pi_x \| \pi_{x_{P_i}}) > \frac{2k(k-1)I_{\mu}}{\epsilon}] \leq \frac{\epsilon}{2k}.$$

Now by union bound,

$$Pr[\exists i : \mathbf{D}(\pi_x \| \pi_{x_{P_i}}) > \frac{2k(k-1)I_{\mu}}{\epsilon}] \leq \frac{\epsilon}{2}.$$

Next, we run the sampling protocol from Lemma 6.1 by considering the following in Lemma 6.1: $\mu = \pi_x$, $\nu_{P_i} = \pi_{x_{P_i}}$, $I = \frac{2k(k-1)I_{\mu}}{\epsilon}$ and error parameter be $\frac{\epsilon}{4}$.
In the protocol tree, at a node v owned by a party P_i, let $g_{P_i,0}(v)$ and $g_{P_i,1}(v) = 1 - g_{P_i,0}(v)$ denote the probabilities that the next bit sent is 0 or 1 respectively. For the node w owned by some other party P_j, let $h_{P_i,P_j,0}(w)$ and $g_{P_i,P_j,1}(w) = 1 - g_{P_i,P_j,0}(w)$ denote the probabilities that the next bit sent is 0 or 1 respectively, estimated by the party P_i. Now for each leaf node l, $g_{P_i}(l)$ be the product of all the values $g_{P_i,b}(v)$ where v is owned by the party P_i and h_{P_i,P_j} be the product of all the values $h_{P_i,P_j,b}(v)$ where v is owned by the party P_j. Then, $Pr[\pi_x = l] = \prod_{i=1}^{k} g_{P_i}(l)$ and $Pr[\pi_{P_i} = l] = g_{P_i}(l) \prod_{j\neq i} h_{P_i,P_j}(l)$.

Thus we can apply Lemma 6.1 to obtain a sample transcript T using only $2^{O(1+\frac{k^2(k-1)I_{\mu}}{\epsilon^2})} = 2^{O(1+\frac{k^3 I_{\mu}}{\epsilon^2})}$ bits of communication, such that $|T - \pi_x| < \frac{\epsilon}{2}$. Hence, if T_{out} be the final output of the transcript T and π_{out} be the final output of the original protocol π_x, then $Pr[T_{out} \neq \pi_{out}] < \frac{\epsilon}{2}$.

Now let us bound the error probability of the new protocol.

$$\begin{aligned} Pr[T_{out} \neq f(x)] &\leq Pr[\exists i : \mathbf{D}(\pi_x \| \pi_{x_{P_i}}) > \frac{2k(k-1)I_\mu}{\epsilon}] \\ &+ Pr[T_{out} \neq \pi_{out} \mid \forall_i, \mathbf{D}(\pi_x \| \pi_{x_{P_i}}) \leq \frac{2k(k-1)I_\mu}{\epsilon}] + Pr[\pi_{out} \neq f(x)] \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} + \rho = \rho + \epsilon. \end{aligned}$$

Note that in the expression $Pr[T_{out} \neq f(x)]$, the probability is calculated over $x \sim \mu$ and the randomness used by the protocol. So by averaging argument, we can fix some "good" random bits such that probability of failure will be bounded by $\rho+\epsilon$, while calculating the probability only over $x \sim \mu$. This completes the first part of the proof.
The second part follows from the first part along with Yao's Min-max Theorem and let μ' be the distribution such that $D^{\mu'}_{\rho+\epsilon}(f) = R_{\rho+\epsilon}(f)$ and since by definition, $IC^k_{\mu'}(f,\rho) \leq IC^k(f,\rho)$, hence,

$$R_{\rho+\epsilon}(f) = D^{\mu'}_{\rho+\epsilon}(f) \leq 2^{O(1+k^3 IC^k_{\mu'}(f,\rho)/\epsilon^2)} \leq 2^{O(1+k^3 IC^k(f,\rho)/\epsilon^2)}.$$

□

Note that for $k = 2$, the above two relations coincide with the similar relations shown for two party communication setting [Bra12].

7 Information Complexity and Amortized Communication

7.1 The distributional case

Let $f : X \to \mathbb{Z}_K$ be any function, where $X = X_1 \times X_2 \times \cdots \times X_k$. We shall consider the problem of computing n copies of f with error ρ in each coordinate of the computation.

Definition 7.1 ([BR11])**.** *Let μ be any distribution on $X_1 \times X_2 \times \cdots \times X_k$ and consider any $1 < \rho < 1$. The distributional complexity of computing f on each of n independent set of inputs drawn according to μ, with failure probability at most ρ on each of the inputs is denoted by $D^{\mu,n}_\rho(f^n)$.*

Note that just by running the n copies of protocol that solves f, we can get $D^{\mu,n}_\rho(f^n) \leq n.D^\mu_\rho(f)$. The following theorem establish a relation between information complexity and amortized communication complexity for multiparty communication setting. The similar theorem for two party case was given in [BR11].

Theorem 7.1. *For every μ, f, ρ, there exists a protocol π that computes the value of f with failure probability ρ and communicates at most $D^{\mu,n}_\rho(f^n)$ such that $IC^k_\mu(\pi) \leq \frac{D^{\mu,n}_\rho(f^n)}{n}$ and thus $IC^k_\mu(f,\rho) \leq \frac{D^{\mu,n}_\rho(f^n)}{n}$.*

Proof. Suppose τ be a protocol realizing $D^{\mu,n}_\rho(f^n)$ and without loss of generality we can assume that τ only uses private coin randomness. Now define the protocol $\pi(x)$ as follows:

1. The parties publicly sample $J \in \{1, 2, \cdots, n\}$.
2. The parties publicly sample $X^1, X^2, \cdots, X^{J-1}, X^{J+1}, \cdots, X^n$.

3. The parties set $X_{P_i}^J = x_{P_i}$ and run the protocol τ on $X^1, X^2, \cdots, X^n$. They output the result computed for the J-th coordinate.

Note that $CC(\pi) = CC(\tau)$ and the failure probability of π is ρ. Now let us bound $IC_\mu^k(\pi)$.

$$\begin{aligned}
IC_\mu^k(\pi) &= \frac{1}{k-1}\sum_{i=1}^{k} I(\overline{X_{P_i}}; \Pi \mid X_{P_i}) \\
&= \frac{1}{k-1}\sum_{i=1}^{k} I(\overline{X_{P_i}}; JX^1X^2\ldots X^{J-1}X^{J+1}\ldots X^n\tau \mid X_{P_i}) \\
&= \frac{1}{k-1}\sum_{i=1}^{k} [I(\overline{X_{P_i}}; JX^1X^2\ldots X^{J-1}X^{J+1}\ldots X^n \mid X_{P_i}) + \\
&I(\overline{X_{P_i}}; \tau \mid X_{P_i}JX^1X^2\ldots X^{J-1}X^{J+1}\ldots X^n)] \quad \text{by Proposition 2.2} \\
&= \frac{1}{k-1}\sum_{i=1}^{k} I(\overline{X_{P_i}}^J; \tau \mid X_{P_i}^1X_{P_i}^2\ldots X_{P_i}^n J\overline{X_{P_i}}^1\overline{X_{P_i}}^2\ldots\overline{X_{P_i}}^{J-1}\overline{X_{P_i}}^{J+1}\ldots\overline{X_{P_i}}^n) \\
&\leq \frac{1}{k-1}\sum_{i=1}^{k} I(\overline{X_{P_i}}^J; \tau \mid X_{P_i}^1X_{P_i}^2\ldots X_{P_i}^n J\overline{X_{P_i}}^1\overline{X_{P_i}}^2\ldots\overline{X_{P_i}}^{J-1}) \\
&= \frac{1}{k-1}\sum_{i=1}^{k}\frac{1}{n}\sum_{j=1}^{n} I(\overline{X_{P_i}}^j; \tau \mid X_{P_i}^1X_{P_i}^2\ldots X_{P_i}^n\overline{X_{P_i}}^1\overline{X_{P_i}}^2\ldots\overline{X_{P_i}}^{j-1}) \\
&= \frac{1}{n}\frac{1}{k-1}\sum_{i=1}^{k} I(\overline{X_{P_i}}^1\overline{X_{P_i}}^2\ldots\overline{X_{P_i}}^n; \tau \mid X_{P_i}^1X_{P_i}^2\ldots X_{P_i}^n) \quad \text{by Proposition 2.2} \\
&= \frac{1}{n}IC_{\mu^n}^k(\tau) \leq \frac{CC(\tau)}{n} \quad \text{by Lemma 3.2}
\end{aligned}$$

The first inequality is due to an application of Proposition 2.5, by assuming $A = \tau$, $B = \overline{X_{P_i}}^J$, $C = X_{P_i}^1X_{P_i}^2\ldots X_{P_i}^n J\overline{X_{P_i}}^1\overline{X_{P_i}}^2\ldots\overline{X_{P_i}}^{J-1}$ and $D = \overline{X_{P_i}}^{J+1}\ldots\overline{X_{P_i}}^n$.
If τ uses public coin randomness R, then by Lemma 3.1,

$$IC_\mu^k(\pi) = \mathbf{E}_R[IC_\mu^k(\pi_R)] \leq \frac{1}{n}\mathbf{E}_R[IC_{\mu^n}^k(\tau_R)] = \frac{1}{n}\mathbf{E}_R[IC_{\mu^n}^k(\tau)] \leq \frac{CC(\tau)}{n}.$$

This completes the proof. □

Theorem 7.2. *The parties $P_1, P_2, \cdots, P_k$ are given distributions $Q_1, Q_2, \cdots, Q_k$ respectively, where all the distributions are over the universe $\mathcal{U}$. Then for any $\epsilon > 0$, there is a protocol such that:*

- *at the end of the protocol, P_1 will output an element x_1 distributed according to the distribution Q_1.*
- *$P_2, P_3, \cdots, P_k$ will output the elements $x_2, x_3, \cdots, x_k$ respectively so that for each x,*

$$Pr[x_1 = x_2 = \cdots = x_k \mid x_1 = x] > 1 - \epsilon.$$

- *the number of bits communicated in this protocol is bounded by*

$$\sum_{i=1}^{k} \log(Q_1(x)/Q_i(x)) + \log\frac{1}{\epsilon} + \log\log\frac{1}{\epsilon} + (k+3)\sqrt{\sum_{i=1}^{k} \log(Q_1(x)/Q_i(x))} + 2k + \log k + 7.$$

Proof. The protocol we designed use public coin randomness and runs as follows. All the parties interpret the shared random tape as a sequence of uniformly selected elements $\{a_j\}_{j=1}^{\infty} = \{(x_j, q_j)\}_{j=1}^{\infty}$ from the set $\mathcal{A} := \mathcal{U} \times [0,1]$. For every $1 \le i \le k$, define the subsets $\mathcal{Q}_i := \{(x,q) : Q_i(x) > q\}$. For a constant $C \ge 1$ and for $2 \le i \le k$, also define the subsets $C.\mathcal{Q}_i := \{(x,q) \in \mathcal{A} : (x, \frac{q}{C}) \in \mathcal{Q}_i\}$. We use the other part of the shared randomness to obtain a sequence of random hash functions $h_m : \mathcal{U} \to \{0,1\}$ so that for any $x \ne y \in \mathcal{U}$, $Pr[h_m(x) = h_m(y)] = \frac{1}{2}$.
Now let us present the protocol:

1. P_1 selects the first index j such that $a_j = (x_j, q_j) \in \mathcal{Q}_1$ and outputs x_j.
2. P_1 uses $1 + \log\log\frac{1}{\epsilon}$ bits to communicate the binary encoding of $l := \lceil \frac{j}{|\mathcal{U}|} \rceil$ (if l is too long, then P_1 communicate some arbitrary string).
3. For all t, set $C_t := 2^{t^2}$ and $s_t := 1 + \lceil \log\frac{1}{\epsilon} \rceil + \lceil \log k \rceil + (t+1)^2$.
4. Repeat until every player P_i produces an output, beginning with $t = 0$:

   (a) P_1 communicates the values of all the hash values $h_m(x_j)$, for $1 \le m \le s_t$, that have not been communicated previously.

   (b) if there is an $a_r = (x_r, q_r)$ such that $r \in \{(l-1).|\mathcal{U}| + 1, \cdots, l.|\mathcal{U}|\}$ in $\mathcal{C}_t.\mathcal{Q}_i$ and $h_m(x_r) = h_m(x_j)$, for $1 \le m \le s_t$, then P_i responds "success" and outputs x_r; if there is more than one such a_r, P_i selects the first one.

   (c) if there exists a party P_i which have not responded "success" so far, increment t and repeat.

The output by P_1 is distributed according to the distribution Q_1. Now fix a choice of j and the pair (x_j, q_j) by P_1. Step 4 is guaranteed to terminate when $t^2 \ge \sum_{i=1}^{k} \log(Q_1(x_j)/Q_i(x_j))$ since for all $2 \le i \le k$, $a_j \in \frac{Q_1(x_j)}{Q_i(x_j)}.\mathcal{Q}$. Denote $T := \sqrt{\sum_{i=1}^{k} \log(Q_1(x_j)/Q_i(x_j))}$. By iteration T, P_1 will have sent s_T bits in step 4 and each party P_i, $2 \le i \le k$ will have sent $T+1$ bits. Thus total number of bits communicated in step 4 is bounded by

$$\begin{aligned}
s_T + (k-1)(T+1) &= 1 + \lceil \log(1/\epsilon) \rceil + \lceil \log k \rceil + (T+1)^2 + (k-1)(T+1) \\
&\le (\sqrt{\sum_{i=1}^{k} \log(Q_1(x_j)/Q_i(x_j))} + 2)^2 + (k-1)\sqrt{\sum_{i=1}^{k} \log(Q_1(x_j)/Q_i(x_j))} \\
&\quad + \log\frac{1}{\epsilon} + \log k + 2k + 1 \\
&= \sum_{i=1}^{k} \log(Q_1(x_j)/Q_i(x_j)) + (k+3)\sqrt{\sum_{i=1}^{k} \log(Q_1(x_j)/Q_i(x_j))} \\
&\quad + \log\frac{1}{\epsilon} + 2k + \log k + 5
\end{aligned}$$

So, total communication is bounded by

$$\sum_{i=1}^{k} \log(Q_1(x_j)/Q_i(x_j)) + (k+3)\sqrt{\sum_{i=1}^{k} \log(Q_1(x_j)/Q_i(x_j))} + \log\frac{1}{\epsilon} + \log\log\frac{1}{\epsilon} + 2k + \log k + 7.$$

Note that for any L,

$$Pr[l > L] = Pr[a_j \in \mathcal{Q}_1, \text{ for } i = 1, 2, \cdots, L.|\mathcal{U}|] = (1 - 1/|\mathcal{U}|)^{L.|\mathcal{U}|} < e^{-L}.$$

Thus the probability that the binary encoding of l exceeds $1 + \lceil \log\log(1/\epsilon) \rceil$ bits is less than $e^{-2.2^{\lceil \log\log(1/\epsilon) \rceil}} \leq \epsilon/2$.
Now let us analyze the protocol. We say that an element $a = (x, q)$ *survives* iteration t, with respect to a party P_i if $a \in 2^{t^2}.\mathcal{Q}_i$ and it satisfies $h_m(x) = h_m(x_j)$, for all $1 \leq m \leq s_t$ for this t. Observe that "correct" element a_j survives iteration t with respect to a party P_i if and only if $2^{t^2} \geq Q_1(x_j)/Q_i(x_j)$.

Claim 7.1 ([BR11])**.** *Let E_{a_j} be the event that the element selected by P_1 is a_j, which is the j-th element on the shared tape of random bits and let $l := \lceil j/|\mathcal{U}| \rceil$. Conditioned on E_{a_i}, the probability that a different element a_r with $r \in \{(l-1).|\mathcal{U}| + 1, \cdots, l.|\mathcal{U}|\}$ survives iteration t with respect to a party P_i is bounded by $\epsilon/k2^{t+1}$.*

Thus for any E_{a_j}, the probability that P_i, for any $2 \leq i \leq k$, will output an element other than x_j conditioned on E_{a_j} is bounded by $\sum_{t=0}^{\infty} \epsilon/k2^{t+1} = \epsilon/k$. Hence, by union bound, some party P_i will output an element other than x_j conditioned on E_{a_j} is bounded by ϵ and this completes the proof. □

7.1.1 Multiparty Correlated Pointer Jumping

In this subsection, we first define the k-party correlated pointer jumping problem, the solution of which will be the main ingredient of establishing the connection between information and amortized communication complexity. The input of this problem is a rooted tree, where

- each non-leaf node is owned by exactly one of the parties P_i, for $1 \leq i \leq k$,
- each non-leaf node owned by a particular party has a set of children that are owned by other parties,
- each node v is associated with k distributions on its children: $child_v^{P_i}$ known to the party P_i,
- the leaf nodes are labeled with the output values.

For every non-leaf vertex w in the tree, whose parent is v and v is owned by P_j, define the *divergence cost* of w as $D(w) = \sum_{i=1}^{k} \log(child_v^{P_j}(w)/child_v^{P_i}(w))$. The divergence cost of the root is set to be 0. Now given a path T from the root to the leaf, the divergence cost of the path, denoted by $D(T)$ is the sum of the divergence costs of all all the nodes present in that path. The divergence cost of an instance F, denoted by $D(F)$, is the expected sum of divergence costs of the nodes present in the correct distribution on paths.

Theorem 7.3. *Given a k-party correlated pointer jumping instance F having depth l, there is a protocol that can sample a path T such that there is an event E, with $Pr[E] > 1 - l\epsilon$ and conditioned on E,*

- *the parties output the same sampled path T that has the correct distribution,*
- *the number of bits communicated is bounded by $D(T) + 2l\log(1/\epsilon) + (k+3)\sqrt{lD(T)} + (2k + \log k + 7)l$.*

Proof. We obtain the protocol to sample the correct path by repeatedly applying the protocol from Theorem 7.2. For each step j, let E_j be the event that the corresponding party sample the correct child. By Theorem 7.2, $Pr[E_j] > 1 - \epsilon$. Let $E := \cap_j E_j$ and thus by union bound, $Pr[E] > 1 - l\epsilon$. Conditioned on E, the sampled path has the correct distribution.
Now, suppose the sampled path is $T = v_0, v_1, \cdots, v_l$. Then by Theorem 7.2, the total number of bits communicate dis bounded by

$$\begin{aligned}
&\sum_{j=1}^{l}(D(v_j) + \log(1/\epsilon) + \log\log(1/\epsilon) + (k+3)\sqrt{D(v_j)} + 2k + \log k + 7) \\
&\leq \sum_{j=1}^{l} D(v_j) + l\log(1/\epsilon) + l\log\log(1/\epsilon) + (k+3)\sqrt{l\sum_{j=1}^{l} D(v_j)} + (2k + \log k + 7)l \\
&\leq D(T) + 2l\log(1/\epsilon) + (k+3)\sqrt{lD(T)} + (2k + \log k + 7)l
\end{aligned}$$

where the first inequality is by the Cauchy-Schwartz inequality. □

Given a public coin multiparty protocol with inputs $X = X_1, X_2, \cdots, X_k$ and public randomness R, for every fixing of x and r, we obtain an instance of k-party correlated pointer jumping problem. The tree is same as the protocol tree with the public randomness r. For a node v at depth d is owned by P_i, for $1 \leq i \leq k$, define $child_v^{P_i, x_{P_i}}$ so that it has the same distribution as $M|X_{P_i} = x_{P_i}, \pi(X)_{\leq d} = rv$. We denote the instance of k-party correlated pointer jumping by $F_\pi(x, r)$ and let μ be the distribution on X.

Lemma 7.1. $\mathbf{E}_{X,R}[D(F_\pi(x,r))] = (k-1)IC_\mu^k(\pi)$.

Proof. We first fix a r and then denote the corresponding protocol as π_r. We then use the induction argument to prove that $\mathbf{E}_X[D(F_{\pi_r}(x,r))] = (k-1)IC_\mu^k(\pi_r)$. Then we use Lemma 3.1 to complete the proof.
When the depth is 0, then terms in both the sides are 0. Now without loss of generality, lets assume that the root node is owned by the party P_1 and let M denotes the child of the root sampled during the protocol. Let $F(x,r)_m$ denotes the divergent cost of the subtree rooted at m. Then

$$\mathbf{E}_X[D(F_{\pi_r}(x,r))] = \mathbf{E}_{X,M}[\sum_{j=1}^{k} \log(child_v^{P_1,x_{P_1}}(m)/child_v^{P_j,x_{P_j}}(m)) + \mathbf{D}(F(x,r)_m)] \tag{1}$$

Note that for every x, $M|X = x$ has the same distribution as of $M|X_{P_1} = x_{P_1}$ and thus by Proposition 2.4, the expectation of the first term in Equation 1 is equal to $\sum_{j=1}^{k} I(\overline{X_{P_j}}; M \mid X_{P_j})$. By the induction hypothesis, for every fix $M = m$, the second term is equal to $(k-1)IC_\mu^k(\pi_r \mid$

$m) = \sum_{j=1}^{k} I(\overline{X_{P_j}}; \pi_r(X) \mid X_{P_j} m)$, where $\pi_r \mid m$ denotes the protocol π_r given that the child of the root node of the protocol tree for π_r is m. Then, these two terms together give

$$\begin{aligned}
\mathbf{E}_X[D(F_{\pi_r}(x,r))] &= \sum_{j=1}^{k} I(\overline{X_{P_j}}; M \mid X_{P_j}) + \sum_{j=1}^{k} I(\overline{X_{P_j}}; \pi_r(X) \mid X_{P_j} M) \\
&= \sum_{j=1}^{k} I(\overline{X_{P_j}}; \pi_r(X) \mid X_{P_j}) \quad \text{by Proposition 2.2} \\
&= (k-1) IC_{\mu}^{k}(\pi_r)
\end{aligned}$$

□

Now Theorem 7.3 and Lemma 7.1 together lead to the following corollary.

Corollary 7.1. *Let X be the input to a l round multiparty communication protocol π, where the number of parties involved is k and let I be the internal information cost of π. Then for every $\epsilon > 0$, there exists a protocol τ, such that at the end of this protocol each party outputs a transcript for π. Moreover, there is a event E with $Pr[E] > 1 - l\epsilon$ such that conditioned on E, all the parties output the same transcript distributed according to $\pi(X)$ and the expected communication of τ is $(k-1)I + (k+3)\sqrt{l(k-1)I} + 2l\log(1/\epsilon) + (2k + \log k + 7)l$.*

7.1.2 Information and amortized communication

Here we establish a tight connection between the amount of information revealed by a protocol computing a function f and the amortized communication complexity of computing many copies of the function f. We first recall a simple observation from [BR11].

Claim 7.2 ([BR11])**.** *For each f, ρ and μ, $\lim_{\alpha \to \rho} IC_{\mu}^{k}(f, \alpha) = IC_{\mu}^{k}(f, \rho)$.*

Another simple but important observation required for our purpose is given below. Before stating the observation, we need the following definition.

Definition 7.2 (Layered Multiparty Protocol)**.** *A mulitparty protocol π involving k parties $P_1, \cdots, P_k$ is said to be layered if the communication is done by the parties in the following order: first P_1's turn, then P_2's turn and so on up to P_k's turn and then again P_1's turn and so on.*

Lemma 7.2. *For every I and input distribution μ, if there is a multiparty protocol τ involving k parties, having $IC_{\mu}^{k}(\tau) = I$, then there exists a layered protocol π. Moreover, $IC_{\mu}^{k}(\pi) = IC_{\mu}^{k}(\tau) = I$ and $CC(\pi) \leq 2k.CC(\tau)$.*

Proof. Without loss of generality we can assume that τ is a private coin protocol as otherwise we can fix a random string r and prove the same result and after that apply Lemma 3.1. We modify the protocol τ to get the desired protocol π, in the following way:

- If τ selects the party P_i to communicate, but according to the ordering mentioned in the statement of the lemma, it is the turn of the party P_j, for $j \neq i$, then the party P_j communicate an idle symbol ϕ.

Clearly, $CC(\pi) \leq 2k.CC(\tau)$. Note that as here the communication involves a symbol other than 0 or 1, thus we get the term $2k$. It remains to bound the internal information cost of π.

$$\begin{aligned} IC^k_\mu(\pi) &= \frac{1}{k-1}\sum_{i=1}^{k} I(\pi(X); \overline{X_{P_i}} \mid X_{P_i}) \\ &= \frac{1}{k-1}\sum_{i=1}^{k} [I(\tau(X); \overline{X_{P_i}} \mid X_{P_i}) + I(\Phi; \overline{X_{P_i}} \mid X_{P_i}\tau(X))] \quad \text{by Proposition 2.2} \\ &= \frac{1}{k-1}\sum_{i=1}^{k} I(\tau(X); \overline{X_{P_i}} \mid X_{P_i}) = IC^k_\mu(\tau). \end{aligned}$$

where Φ denotes the string of all ϕ symbols. □

Now define the amortized communication complexity of a function f.

Definition 7.3. *For any function f, the amortized communication complexity with respect to a distribution μ on the input, is defined as*

$$AC(f^\mu_\rho) := \lim_{n\to\infty} \frac{D^{\mu,n}_\rho(f)}{n}.$$

We are now ready to state the main theorem of this section.

Theorem 7.4. *For $\rho > 0$,*

$$IC^k_\mu(f,\rho) \leq AC(f^\mu_\rho) \leq (k-1)IC^k_\mu(f,\rho).$$

Note that for $k = 2$, the above theorem results in equality between the two terms and that matches with the theorem stated in [BR11].

Proof. $\underline{IC^k_\mu(f,\rho) \leq AC(f^\mu_\rho)}$. This direction directly follows from Theorem 7.1.
$\underline{AC(f^\mu_\rho) \leq (k-1)IC^k_\mu(f,\rho)}$. Let $\delta > 0$. we will show that for sufficiently large n, $D^{\mu,n}_\rho(f)/n < (k-1)IC^k_\mu(f,\rho) + \delta$.
By Claim 7.2, there is an $\alpha < \rho$, such that $IC^k_\mu(f,\alpha) < IC^k_\mu(f,\rho) + \delta/4$. Thus there is a protocol τ that computes f with error bounded by α with respect to the input distribution μ and the internal information cost is bounded by $I := IC^k_\mu(f,\rho) + \delta/4$. By Lemma 7.2, there is a layered protocol π computing f with error bounded by α and internal information cost if same as that of τ, where $CC(\pi) \leq 2k.CC(\tau)$.
For every n, π^n denotes the protocol that executes the protocol π independently in parallel with different sets of inputs from X^n. Thus π^n has $CC(\pi)$ rounds and communication complexity $nCC(\pi)$. Moreover, the error at each coordinate is bounded by α. Now we obtain our desired protocol by compressing π^n.
Let j_π denote the transcript for input X^j and observe that for every j, (X^j, j_π) are mutually independent. This implies $IC^k_{\mu^n}(\pi^n) = nIC^k_\mu(\pi)$. Let T^n denotes the random variable of the path sampled in π^n and $T_1, \cdots, T_n$ denote the random variables of the n paths sampled in the individual copies of π. As the protocols run independently, $\mathbf{E}[D(T^n)] = \sum_{j=1}^{n} \mathbf{E}[D(T_j)]$. Each vertex in the

protocol tree of π^n corresponds to an n-tuple of vertices of π. If w corresponds to the vertices $(w_1, \cdots, w_n)$ having parents $(v_1, \cdots, v_n)$ owned by say P_i, then

$$\begin{aligned} D(w) &= \sum_{j=1}^{k} \log(child_v^{P_i}(w)/child_v^{P_j}(w)) \\ &= \sum_{j=1}^{k} \log(\prod_{m=1}^{n} child_{v_m}^{P_i}(w_m)/\prod_{m=1}^{n} child_{v_m}^{P_j}(w_m)) \\ &= \sum_{j=1}^{k} \sum_{m=1}^{n} \log(child_{v_m}^{P_i}(w_m)/child_{v_m}^{P_j}(w_m)) \\ &= \sum_{m=1}^{n} D(w_m). \end{aligned}$$

By Lemma 7.1, $\mathbf{E}[D(T_j)] = IC_\mu^k(\pi)$. Thus, by the central limit theorem, for large enough n,

$$Pr[D(T^n) \geq n((k-1)IC_\mu^k(\pi) + \delta/4)] < (\rho - \delta)/2.$$

Now we use Corollary 7.1 to simulate π^n, with error parameter $\epsilon = (\rho - \delta)/2CC(\pi)$ and get the desired protocol having error bounded by $\alpha + (\rho - \alpha) = \rho$ and truncate the protocol after $n.((k-1)IC_\mu^k(\tau) + \delta/4) + (k+3)\sqrt{2kCC(\tau).n.((k-1)IC_\mu^k(\tau) + \delta/4)} + 4kCC(\tau)\log(1/\epsilon) + (2k + \log k + 7)2kCC(\tau)$ bits of communication. For large enough n, the per copy communication of this protocol is at most $(k-1)IC_\mu^k(\tau) + \delta/2$, as required. □

7.2 The non-distributional case

In this subsection, we establish a relation between prior-free information complexity and amortized communication complexity for non-distributional case. This relation is similar to the relation we have shown for distributional case in the last subsection.

Theorem 7.5. *For $\rho > 0$,*

$$IC^k(f, \rho) \leq \lim_{n\to\infty} \frac{R_\rho^n(f^n)}{n} \leq (k-1)IC^k(f, \rho).$$

Above relation will lead to equality for $k = 2$ and this matches the result sated in [Bra12]. We adapt the proof technique from [BR11], to prove this relation. Before proving the above theorem, we first state another theorem.

Theorem 7.6. *Let $f : X \to \{0,1\}$, where $X = X_1 \times X_2 \times \cdots \times X_k$, then for each $\rho, \delta_1, \delta_2 > 0$ there is an N such that $n \geq N$ there is a protocol $\pi_n(x^1, x^2, \cdots, x^n)$ for computing n instances of f. The protocol π_n will have communication complexity at most $n(k-1).IC^k(f, \rho).(1 + \delta_1)$ and the error will be bounded by a quantity slightly greater that ρ on each copy. Furthermore, the error vector of the protocol will behave as if n evaluations were independent, except with a probability at most δ_2.*

More precisely, let $Q : \{0,1\}^n \to \{0,1\}$ be any monotone function and let $\mathbf{p} = (p_1, \cdots, p_n)$ be the random variable representing π_n's output on input $x^1, \cdots, x^n$. Suppose $\mathbf{e} = (e_1, \cdots, e_n)$ be the

error vector, where e_i be a indicator random variable representing $p_i \neq f(x^i)$ and $\mathbf{b} = (b_1, \cdots, b_n)$ be a vector of independent Barnoulli variable $b_i \sim B_\rho$. Then,

$$Pr[Q(\mathbf{e}) = 1] \leq Pr[Q(\mathbf{b}) = 1] + \delta_2.$$

Assuming the above theorem, we can now prove Theorem 7.5.

Proof of Theorem 7.5. Second part of the Theorem 5.2 implies the following

$$IC^k(f, \rho) = \frac{IC^n(f^n, \rho)}{n} \leq \frac{R^n_\rho(f^n)}{n}$$

where the last inequality is due to the fact that information cost of a protocol is always upper bounded by the communication cost (by Lemma 3.2). This shows that $IC^k(f, \rho) \leq \lim_{n\to\infty} \frac{R^n_\rho(f^n)}{n}$.

The remaining part follows from Theorem 7.6. Consider Q as a one-coordinate indicator function and this implies $R^n_\rho(f^n) \leq n(k-1).IC^k(f, \rho - \delta_2).(1 + \delta_1)$. As $\delta_1 \to 0$ and $n \to \infty$, for each $\delta_2 > 0$, we have

$$\lim_{n\to\infty} \frac{R^n_\rho(f^n)}{n} \leq IC^k(f, \rho - \delta_2).$$

Now as a consequence of Theorem 4.3, we get that $IC^k(f, \rho)$ is continuous in $\rho \in [0, 1]$ and this implies that $IC^k(f, \rho) = \lim_{\delta_2 \to 0} IC^k(f, \rho - \delta_2)$. Hence,

$$\lim_{n\to\infty} \frac{R^n_\rho(f^n)}{n} \leq (k-1)IC^k(f, \rho)$$

and this completes the proof. □

Now it remains to prove Theorem 7.6 and we provide it in the next subsection.

7.2.1 Proof of Theorem 7.6

The proof follows the idea used in [Bra12]. We first state two main lemmas that will be used in the proof. Proofs of the following lemmas are same as that given in [Bra12], but for the sake of clarity, we are providing the detailed proofs.

Lemma 7.3. *Consider any $\delta_1, \delta_4 > 0$. Let a layered $l' \leq 2kl$, for some l, round protocol π for computing a function $f : X_1 \times \cdots \times X_k \to \{0, 1\}$ (denote $X := X_1 \times \cdots \times X_k$) involving k parties, that satisfies $IC^k_\mu(\pi) \leq IC^k(f, \rho)(1+\delta_1/3)$, is given. Then for sufficiently large n, there is a protocol π_n that takes n instance of f as an input and the following holds:*

1. *For every input $\mathbf{x}$, $|\pi_n(\mathbf{x}) - \pi^n(\mathbf{x})| < \delta_4/(2n^2)$.*
2. *The expected communication cost is at most $n(k-1).IC^k(f, \rho).(1 + 2\delta_1/3)$.*
3. *The worst case communication cost is bounded by $100nkM/\delta_1$, where $M = \max_i \log |X_i|$.*

Proof. Let $\alpha > 0$ be a parameter. Now consider a zero-sum game $\mathcal{G}$, where first player A selects a distribution on n sets of inputs and the second player B produces a randomized protocol τ. The payoff for A is

$$P_A(\mu, \tau) := (1 - \alpha)\frac{\mathbf{E}_{\mathbf{x}\sim\mu}|\tau(\mathbf{x})|}{n(k-1)IC^k(f, \rho)(1 + \delta_1/2)} + \frac{\mathbf{E}_{\mathbf{x}\sim\mu}|\tau(\mathbf{x}) - \pi^n(\mathbf{x})|}{\delta_4/(2n^2)}$$

where $\mathbf{x} := (x^1, \cdots, x^n)$, i.e., n sets of inputs and each $x^j = x_1^j, \cdots, x_k^j$.

Let ν be any mixed strategy for A and we represent $\overline{\mu}(\mathbf{x}) = \mathbf{E}_{\mu\sim\nu}\mu(\mathbf{x})$. Now observe that, for any protocol τ, $\mathbf{E}_{\mu\sim\nu}P_A(\mu,\tau) = P_A(\overline{\mu},\tau)$. Thus to prove that the value of the game is bounded by 1, it is sufficient show that for each distribution μ, there is a protocol τ such that $P_A(\mu,\tau) < 1$.

Fix a distribution μ and consider its projection on n coordinates, denoted as $\mu_1, \cdots, \mu_n$. Observe that the proof of the second part of Theorem 5.1 provides us the following

$$IC_\mu^k(\pi^n) \leq \sum_{j=1}^{n} IC_{\mu_j}^k(\pi) \leq n.IC^k(f,\rho).(1+\delta_1/3).$$

The protocol π^n is also of l round. Let $\epsilon := \frac{\alpha.\delta_4}{4n^2kl}$. By Lemma 7.1, the protocol π^n can be simulated by a protocol τ such that the expected communication is bounded by $(k-1).n.IC^k(f,\rho)(1+\delta_1/3)+(k+3)\sqrt{2kl(k-1).n.IC^k(f,\rho)(1+\delta_1/3)}+4kl\log(1/\epsilon)+(2k+\log k+7)2kl < (k-1).n.IC^k(f,\rho)(1+\delta_1/2)$ and $|\tau - \pi^n| < 2kl\epsilon$. Hence, we get that $P_A(\mu,\tau) < 1$.

By Yao's Min-max Theorem, there exists a distribution κ on protocols, such that for each distribution μ on inputs, $\mathbf{E}_{\tau\sim\kappa}[P_A(\mu,\tau)] < 1$. Now consider π_n as a randomized protocol that executes first selects a protocol according to the distribution κ and then executes it. Clearly, $P_A(\mu,\pi_n) < 1$, for all distributions μ.

A particular input $\mathbf{x}$ can be thought as a singleton distribution, denoted by $\mu_{\mathbf{x}}$. Let $\alpha < 1-(1+\delta_1/2)/(1+3\delta_1/5)$. Then for every input $\mathbf{x}$, $P_A(\mu_{\mathbf{x}},\pi_n) < 1$ and this implies that

$$|\tau(\mathbf{x}) - \pi^n(\mathbf{x})| < \frac{\delta_4}{2n^2}$$

and the expected number of bits of communication of π_n on input $\mathbf{x}$ is bounded by

$$\frac{n(k-1)IC^k(f,\rho)(1+\delta_1/2)}{1-\alpha} < n(k-1)IC^k(f,\rho)(1+3\delta_1/5).$$

Now to get the worst case bound on the number of bits communicated, we use the most common technique. We modify the protocol π_n in such a way that it runs as usual for $80nkM/\delta_1$ bits and if does not terminate within this limit then just communicates all the inputs, which takes total nkM bits. By Markov inequality, the probability that the modified protocol crosses $80nkM/\delta_1$ bits of communication is bounded by $\frac{2nIC^k(f,\rho)}{80nkM/\delta_1} = \frac{\delta_1 IC^k(f,\rho)}{40kM}$. Thus, the expected number of bits of communication for this modified protocol is at most

$$n(k-1)IC^k(f,\rho)(1+3\delta_1/5) + \frac{\delta_1 IC^k(f,\rho)}{40kM}.(nkM) < n(k-1)IC^k(f,\rho)(1+2\delta_1/3).$$

□

The above lemma gives us an upper bound on the expected number of bits communicated and now the following lemma helps us to get an upper bound on the worst case communication cost.

Lemma 7.4. *Consider any $\delta_1, \delta_4 > 0$. Let a layered $l' \leq 2kl$, for some l, round protocol π for computing a function $f : X_1 \times \cdots \times X_k \to \{0,1\}$ involving k parties, that satisfies $IC_\mu^k(\pi) \leq IC^k(f,\rho)(1+\delta_1/3)$, is given. Suppose τ be a protocol that runs on n^3 sets of inputs by dividing them into n^2 blocks of n sets each and then executing π_n (as in Lemma 7.3) on each block. If τ is truncated after $n^3(k-1)IC(f,\rho)(1+\delta_1)$, then for each set of inputs $\mathbf{x}$, $|\tau(\mathbf{x}) - \pi^{n^3}(\mathbf{x})| < \delta_4$.*

Proof. First fix a set of n^3 inputs. Now by union bound along with the second condition of Lemma 7.3, the probability that there exists a block among n^2 blocks, that is different under π_n than under π^n is bounded by $\delta_4/2$. Let T_j, for $1 \leq j \leq n^2$ denote the random variable representing the communication cost by the j-th copy during the execution of τ. Denote $T := \sum_{j=1}^{n^2} T_j$.

Note that T_j are i.i.d. and $\mathbf{E}[T_j] < n(k-1).IC^k(f,\rho).(1+2\delta_1/3)$ and thus

$$Var(T_j) < \mathbf{E}[T_j].\frac{100nkM}{\delta_1} < 200n^2kM.IC^k(f,\rho)/\delta_1.$$

So, $\mathbf{E}[T] < n^3(k-1).IC^k(f,\rho).(1+2\delta_1/3)$ and $Var(T) < 200n^4kM.IC^k(f,\rho)/\delta_1$. Now by Chebyshev's inequality, we get

$$Pr[T > n^3(k-1).IC^k(f,\rho).(1+\delta_1)] < \frac{200n^4kM.IC^k(f,\rho)/\delta_1}{(n^3(k-1).IC^k(f,\rho).\delta_1/3)^2} < \delta_4/2$$

where the last inequality holds for large enough n. □

It only remains to combine Lemma 7.3 and Lemma 7.4 to get Theorem 7.6.

Proof of Theorem 7.6. Suppose for all $1 \leq i \leq k$, $|X_i| \leq 2^M$. As a consequence of Theorem 4.3, we get that $IC^k(f,\rho)$ is continuous in $\rho \in [0,1]$ and this implies that there exists a $\delta_3 > 0$ such that

$$IC^k(f,\rho-\delta_3) < IC^k(f,\rho).(1+\delta_1/3).$$

By the definition of $IC^k(f,\rho-\delta_3)$, there is a protocol τ that on each set of inputs succeeds except with probability at most $\rho-\delta_3$ and for each distribution μ,

$$IC^k_\mu(\tau) \leq IC^k(f,\rho).(1+\delta_1/3)$$

where τ is a l round protocol. Now take the layered protocol π according to Lemma 7.2 and thus number of rounds of π is bounded by $2kl$. Then by setting $\delta_4 := \min(\delta_2,\delta_3)/2$, we apply Lemma 7.4 and this completes the proof. □

References

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 67–76, 2010.

[BJKS04] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004.

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 748–757, 2011.

[Bra12] Mark Braverman. Interactive information complexity. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 505–524, 2012.

[CSWY01] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, pages 270–278, 2001.

[DW07] Martin Dietzfelbinger and Henning Wunderlich. A characterization of average case communication complexity. *Inf. Process. Lett.*, 101(6):245–249, March 2007.

[GKR14] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:49, 2014.

[JKR09] T. S. Jayram, Swastik Kopparty, and Prasad Raghavendra. On the communication complexity of read-once acˆ0 formulae. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity, CCC 2009, Paris, France, 15-18 July 2009*, pages 329–340, 2009.

[JKZ10] Rahul Jain, Hartmut Klauck, and Shengyu Zhang. Depth-independent lower bounds on the communication complexity of read-once boolean formulas. In *Computing and Combinatorics, 16th Annual International Conference, COCOON 2010, Nha Trang, Vietnam, July 19-21, 2010. Proceedings*, pages 54–59, 2010.

[LS09] Nikos Leonardos and Michael Saks. Lower bounds on the randomized communication complexity of read-once functions. *2012 IEEE 27th Conference on Computational Complexity*, 0:341–350, 2009.

[Sha48] Claude Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, July, October 1948.

[SW73] David Slepian and Jack K. Wolf. Noiseless coding of correlated information sources. *Information Theory, IEEE Transactions on*, 19(4):471–480, 1973.

[WC81] Mark N. Wegman and Larry Carter. New hash functions and their use in authentication and set equality. *J. Comput. Syst. Sci.*, 22(3):265–279, 1981.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de