Electronic Colloquium on Computational Complexity, Report No. 55 (2015)

How to Compress Asymmetric Communication

Sivaramakrishnan Natarajan Ramamoorthy [*] Anup Rao[†]

April 11, 2015

Abstract

We study the relationship between communication and information in 2-party communication protocols when the information is asymmetric. If I^A denotes the number of bits of information revealed by the first party, I^B denotes the information revealed by the second party, and C is the number of bits of communication in the protocol, we show that

- one can simulate the protocol using order $I^A + \sqrt[4]{C^3 \cdot I^B} \cdot \log C + \sqrt{C \cdot I^B} \cdot \log C$ bits of communication,
- one can simulate the protocol using order $I^A \cdot 2^{O(I^B)}$ bits of communication

The first result gives the best known bound on the complexity of a simulation when $I^A \gg I^B, C^{3/4}$. The second gives the best known bound when $I^B \ll \log C$. In addition we show that if a function is computed by a protocol with asymmetric information complexity, then the inputs must have a large, nearly monochromatic rectangle of the right dimensions, a fact that is useful for proving lower bounds on lopsided communication problems.

1 Introduction

Can one compress the communication in 2-party communication protocols when the information revealed by the messages of the protocol is small? This has been a central question in communication complexity in recent years. Interest in the question is fueled by applications to proving lower bounds in communication complexity, which in turn yield lower bounds for streaming algorithms and data structures among other models. In this work, we obtain stronger results when one party reveals much less information than the other, a case that often arises when studying data structures.

Throughout this discussion we assume that there is a known distribution on inputs to a communication protocol. The question of compressing protocols was posed by Chakrabarti, Shi, Wirth and Yao [CSWY01] (see also [BYCKO93, Abl93, SS02]), who defined the information cost of the protocol as the mutual information between the inputs and the messages of the protocol. Barak, Braverman, Chen and Rao [BBCR10] called this measure the *external* information cost of a protocol, and proved that if the protocol π has external information cost I_π^{ext} and communication C_π, then one can simulate the protocol with $O(I_\pi^{\mathsf{ext}} \log C_\pi)$ bits of communication, which is optimal

[*] Computer Science and Engineering, University of Washington, sivanr@cs.washington.edu. Supported by the National Science Foundation under agreement CCF-1016565. Also partially supported by CCF-1016509

[†] Computer Science and Engineering, University of Washington, anuprao@cs.washington.edu. Supported by an Alfred P. Sloan Fellowship, the National Science Foundation under agreement CCF-1016565, an NSF Career award, and by the Binational Science Foundation under agreement 2010089.

ISSN 1433-8092

upto the factor of $\log C_\pi$. In addition, [BBCR10] identified another measure of information called the *internal* information cost of the protocol. This is the amount of information that is revealed by the parties of the protocol: let I_π^A denote the information revealed by the first party, and I_π^B denote the information revealed by the second party. Then the internal information cost is defined to be $I_\pi = I_\pi^A + I_\pi^B$. Since each of the parties already knows one of the inputs, the internal information cost is never more than the external information cost, with equality when the inputs to the parties are independent of each other. This quantity turns out to be the most interesting for applications to lower bounds. Indeed, Braverman and Rao [BR11] showed that the internal information cost required to compute a function is exactly equal to the amortized communication complexity of the function, so this quantity has a very natural interpretation, seemingly independent of information theory.

[BBCR10] showed that any protocol π can be simulated in $O(\sqrt{I_\pi C_\pi}\log C_\pi)$ bits of communication. If we wish the simulation to not have a dependence on the communication of the original protocol, Braverman [Bra11] showed that one can carry out the simulation using communication complexity $2^{O(I_\pi)}$ (see also [BW12, KLL$^+$12]), a result that was subsequently proven to be tight by Ganor, Kol and Raz [GKR14]. In addition, Braverman and Weinstein [BW12] (see also [KLL$^+$12]) showed that if a function is computed by π, then the space of inputs must contain a nearly monochromatic rectangle of density $2^{-O(I_\pi)}$, a fact that can be used to prove lower bounds on the information complexity of computing functions.

In the setting of bounded round communication, Braverman and Rao [BR11] showed that a single message can be compressed to its internal information, giving a protocol that can simulate any r-round protocol with internal information cost I using $I + O(r)$ bits of communication.

In this work, we generalize and strengthen several of these results, in the case that $I^B \lll I^A, C$. This case is interesting in part because many lower bounds for data structures involve proving lower bounds on so called *lopsided* problems, problems where the optimal lower bound on communication for one party is much smaller than for the other party(see [MNSW98], [Pǎt11], [AIP06], [Mil99], [Mil94], [BOR99], [JKKR04], [PT]). Indeed, our techniques allow us to reprove an important and well known theorem of Patrascu [Pǎt11] giving a tight lower bound on the communication complexity of lopsided disjointness.

Our first theorem is somewhat analogous to the result of [BBCR10]. We show

Theorem 1. *Every protocol π can be ϵ-simulated by a protocol with expected communication $O\left(I_\pi^A + \sqrt[4]{\|\pi\|^3 \cdot I_\pi^B} \cdot \log(1/\epsilon) + \sqrt[4]{\|\pi\|^3 \cdot I_\pi^B} \cdot \log\|\pi\| + \sqrt{\|\pi\| \cdot I_\pi^B} \cdot \log\|\pi\|\right)$.*

We prove Theorem 1 in Section 3. At a high level, we first show that π can be simulated by a bounded round protocol, and then use the ideas of [BR11] to get the final simulation.

Our second result is an analogue of [Bra11]:

Theorem 2. *π can be simulated in communication complexity $I_\pi^A \cdot 2^{O(I_\pi^B)}$.*

Theorem 2 shows that when the information revealed by one of the parties is a constant, the communication is within a constant factor of the information. We prove Theorem 2 in Section 4. As a corollary to Theorem 2, we show the following,

Corollary 3. *If π computes $f(x,y)$, then there exists a rectangle $S \times T$ such that*

$$\Pr[x \in S] \geq 2^{-O(I_\pi^A)}, \qquad \Pr[y \in T | x \in S] \geq 2^{-O(I_\pi^B)},$$

and there exists a constant $c \in \{0, 1\}$ *such that*

$$\Pr[f(x, y) = c | (x, y) \in S \times T] \geq 2/3.$$

One can view Corollary 3 as defining an asymmetric notion of discrepancy, and showing that if the information complexity is small, then the discrepancy must be large. Corollary3 is a useful tool to prove lower bounds on lopsided problems. We illustrate this by using the ideas going into Corollary 3 to give optimal lower bounds on the communication complexity of lopsided disjointness (a bound first proved by Patrascu [Pǎt11]).

Table 1 summarizes all of the simulation results discussed in this introduction.

Reference	Communication Complexity of the Simulation
[BBCR10]	$O\left(\sqrt{(I^A + I^B)C} \log C\right)$
[BBCR10]	$O\left((I^A + I^B) \log C\right)$ (when inputs are independent)
[Bra11]	$2^{O\left(I^A + I^B\right)}$
[BR11]	$I^A + I^B + O\left(\sqrt{r \cdot (I^A + I^B)} + 1\right) + r \log(1/\epsilon)$
Theorem 1	$O\left(I^A_\pi + \sqrt[4]{\lVert\pi\rVert^3 \cdot I^B_\pi} \cdot \log(1/\epsilon) + \sqrt[4]{\lVert\pi\rVert^3 \cdot I^B_\pi} \cdot \log \lVert\pi\rVert + \sqrt{\lVert\pi\rVert \cdot I^B_\pi} \cdot \log \lVert\pi\rVert\right)$
Theorem 2	$I^A \cdot 2^{O(I^B)}$

Table 1: Known bounds on the complexity of simulating r-round protocols with communication C and information I^A, I^B

2 Preliminaries

Unless otherwise stated, logarithms in this text are computed base two. Random variables are denoted by capital letters and values they attain are denoted by lower-case letters. For example, A may be a random variable and then a denotes a value A may attain and we may consider the event $A = a$. Given $a = a_1, a_2, \ldots, a_n$, we write $a_{\leq i}$ to denote $a_1, \ldots, a_i$. We define $a_{>i}$ and $a_{\leq i}$ similarly. $[\ell]$ denotes the set $\{1, 2, \ldots, \ell\}$.

We use the notation $p(a)$ to denote both the distribution on the variable a, and the number $\Pr_p[A = a]$. The meaning will be clear from context. We write $p(a|b)$ to denote either the distribution of A conditioned on the event $B = b$, or the number $\Pr[A = a|B = b]$. Again, the meaning will be clear from context. Given a distribution $p(a, b, c, d)$, we write $p(a, b, c)$ to denote the marginal distribution on the variables a, b, c (or the corresponding probability). We often write $p(ab)$ instead of $p(a, b)$ for conciseness of notation. If W is an event, we write $p(W)$ to denote its probability according to p. We denote by $\mathbb{E}_{p(a)}\left[g(a)\right]$ the expected value of $g(a)$ with respect to a distributed according to p.

For two distributions p, q, we write $|p(a) - q(a)|$ to denote the ℓ_1 distance between the distributions p and q. We write $p \overset{\epsilon}{\approx} q$ if $|p - q| \leq \epsilon$.

Proposition 4. *Let* $p(x), q(x)$ *be two distributions and* F *be an event such that* $p(x|F) \overset{\epsilon}{\approx} q(x)$*. Then if* $p(F) \geq 1 - \gamma$*, we have* $p(x) \overset{\epsilon+2\gamma}{\approx} q(x)$*.*

Proof. The ℓ_1 distance between p, q can be expressed as $2\max_T(p(T) - q(T))$, where the maximum is taken over all subsets of the support of $p(x)$. Let T be the maximizer. Then

$$\begin{aligned} |p(x) - q(x)| &= 2(p(T) - q(T)) \\ &= (2(p(T|F)p(F) + p(\neg F)p(T|\neg F)) - q(T)) \\ &\leq 2(p(T|F) - q(T)) + 2p(\neg F) \\ &\leq \epsilon + 2\gamma, \end{aligned}$$

as required. □

Proposition 5. *Let $p(x), q(x)$ be two distributions and F be an event such that $p(x) \overset{\epsilon}{\approx} q(x)$. Then if $p(F) \geq 1 - \gamma \geq 3/4$, we have $p(x|F) \overset{\epsilon+4\gamma}{\approx} q(x)$.*

Proof. The ℓ_1 distance between $p(x|F), q(x)$ can be expressed as $2\max_T p(T|F) - q(T)$, where the maximum is taken over all subsets of the support of $p(x)$. Let T be the maximizer. Then

$$\begin{aligned} |p(x|F) - q(x)| &= 2(p(T|F) - q(T)) \\ &= 2(p(T, F)/p(F) - q(T)) \\ &\leq 2(p(T)/p(F) - q(T)) \\ &\leq 2(p(T)/(1-\gamma) - q(T)) \\ &\leq 2(p(T)(1+2\gamma) - q(T)) \\ &\leq \epsilon + 4\gamma, \end{aligned}$$

as required. □

The *entropy* of a random variable A, conditioned on B is defined to be

$$\mathsf{H}_p(A|B) = \sum_{a,b} p(ab) \log \frac{1}{p(a|b)}$$

For a binary random variable A, we denote the entropy of A to be

$$\mathsf{h}(p(0)) = -\left[p(0)\log p(0) + (1 - p(0))\log(1 - p(0))\right]$$

The *divergence* between two distributions is defined to be

$$\mathbb{D}\left(\frac{p}{q}\right) = \sum_a p(a) \log \frac{p(a)}{q(a)}$$

The *mutual information* between two random variables A, B, conditioned on C is defined to be

$$I_p(A; B|C) = \sum_{a,b,c} p(abc) \log \frac{p(abc)}{p(a|c)p(b|c)}.$$

This is always a non-negative quantity, and is at most $\log |\mathsf{Supp}(A)|$. When the underlying distribution p is clear from the context, we sometimes omit it from the notation. The mutual information satisfies the chain rule:

Proposition 6 (Chain Rule)**.** $I(A_1A_2;B|C) = I(A_1;B|C) + I(A_2;B|A_1C)$.

Pinsker's inequality bounds the ℓ_1 distance in terms of the divergence:

Proposition 7 (Pinsker)**.** $\mathbb{D}\left(\frac{p}{q}\right) \geq \cdot|p-q|^2$.

An alternate formulation is as follows:

Proposition 8 (Alternate Pinsker)**.** $\mathbb{E}_{p(bc)}\left[|p(a|bc) - p(a|c)|\right] \leq \sqrt{I(A;B|C)}$.

The chain rule easily gives the following inequality:

Proposition 9 (Data Processing Inequality)**.** *If the random variable A determines B, then $I(A;C) \geq I(B;C)$.*

Proposition 10 ([GKR14])**.** *Let $p(ab)$ be a distribution and $q(a)$ be another. Then*

$$\mathop{\mathbb{E}}_{p(b)}\left[\mathbb{D}\left(\frac{p(a|b)}{p(a)}\right)\right] \leq \mathop{\mathbb{E}}_{p(b)}\left[\mathbb{D}\left(\frac{p(a|b)}{q(a)}\right)\right].$$

We shall sometimes deal with distribution on strings of variable length. We have the following proposition, which follows from Shannon's source coding theorem:

Proposition 11. *Suppose A is a random variable supported on binary strings of length up to n, such that no string in the support of A is a prefix of another string in the support of A. Then $I(A;B|C) \leq \mathbb{E}\left[|A|\right]$.*

2.1 Communication Complexity

For a more involved introduction to communication complexity, we refer the reader to the book [KN97]. Given a protocol π that operates on inputs x, y drawn from a distribution μ using public randomness[1] r and messages m, we write $\pi(xymr)$ to denote the joint distribution of these variables. We write $\|\pi\|$ to denote the *communication complexity* of π, namely the maximum number of bits that may be exchanged by the protocol in any execution. The maximum number of alternations between messages sent by Alice and those sent by Bob is called the number of *rounds* of the protocol.

Let $q(x, y, a)$ be an arbitrary distribution. We say that a protocol π δ*-simulates* q, if there is a function g and a function h such that

$$\pi(x, y, g(x, r, m), h(y, r, m)) \overset{\delta}{\approx} q(x, y, a, a),$$

where $q(x, y, a, a)$ is the distribution on 4-tuples (x, y, a, a) where (x, y, a) are distributed according to q. Thus if π δ-simulates q, the protocol allows the parties to sample a according to $q(a|xy)$.

[1] In our paper we define protocols where the public randomness is sampled from a continuous (i.e. non-discrete) set. Nevertheless, we often treat the randomness as if it were supported on a discrete set, for example by taking the sum over the set rather than the integral. This simplifies notation throughout our proofs, and does not affect correctness in any way, since all of our public randomness can be approximated to arbitrary accuracy by sufficiently dense finite sets.

If λ is a protocol with inputs x, y, public randomness r' and messages m', we say that π δ-simulates λ if π δ-simulates $\lambda(x, y, (r', m'))$. We say that π simulates λ if π 0-simulates λ. We say that π computes a function $f(x, y)$ with success probability $1 - \delta$, if π δ-simulates $\pi(x, y, f(x, y))$.

We shall sometimes refer to the *expected communication*, or *expected number of rounds* of a protocol π. We note here that one can always use a bound on the expected communication or number of rounds to get a bound on the worst case communication, via the following proposition:

Proposition 12. *If π has expected communication c, then it can be γ-simulated by a protocol with communication c/γ.*

Our work relies on ways to measure the information complexity of a protocol (see [BBCR10, Bra12] and references within for a more detailed overview). The *internal information cost* [BBCR10] of π is defined to be $I_\pi(X; M|YR) + I_\pi(Y; M|XR)$. This quantity is the sum of the information learnt by Alice about Bob's input, $I_\pi^B = I_\pi(Y; M|XR)$, and the information learnt by Bob about Alice's input, $I_\pi^A = I_\pi(X; M|YR)$. We will sometimes say that the internal information is (I^A, I^B) when we want to consider the values of both quantities instead of the sum.

2.1.1 Results from prior work

Theorem 13 ([BR11])**.** *For every $\epsilon > 0$, if π is an a protocol with internal information cost I and r rounds in expectation, then π can be ϵ-simulated with expected communication $I + O(\sqrt{r \cdot I} + 1) + r \log(1/\epsilon)$.*

Theorem 14 ([BR11])**.** *For any f, μ, ϵ, let π be the protocol computing f on n independent pairs of inputs, each drawn from the distribution μ and probability of error is at most ϵ on each pair, then there exists a protocol τ computing f on a single input pair with communication $||\tau|| = ||\pi||$ and information $I_\tau^A \leq \frac{I_\pi^A}{n}$, $I_\tau^B \leq \frac{I_\pi^B}{n}$*

3 Compresing Protocols with Asymmetric Information - I

In this section, we show how to compress protocols to take advantage of situations where the information learnt by one party is significantly larger than the information revealed by the other party. We shall prove Theorem 1.

We compress the given protocol in two steps. In the first step, we convert the protocol into a bounded round protocol, while controlling its internal information cost. In the second step, we apply Theorem 13 to conclude the proof. The first step is captured by the following theorem:

Theorem 15 (Bounded round simulation)**.** *Given any protocol π and a parameter k, there exists a protocol that simulates π with $\sqrt{I_\pi^B \cdot \|\pi\|} + \|\pi\|/k$ number of rounds in expectation, and internal information at most $\frac{\|\pi\| \log \|\pi\|}{k} + k\sqrt{I_\pi^B \cdot \|\pi\|} + I_\pi^A + 2 \log \|\pi\| \sqrt{\|\pi\| \cdot I_\pi^B} + 3$.*

We use the protocol τ given in figure 1[2] to simulate π.

Let M denote the output of τ. Then we claim that the distribution of m is correct:

Lemma 16. *$\tau(xyrm) = \pi(xyrm)$.*

[2] Here we do not bother optimizing the communication of τ, since the communication will be eventually optimized via Theorem 13.

```
Input: $x, y$, the inputs to $\pi$. A parameter $k$.
Output: $m, r$ distributed according to $\pi(mr|xy)$.
Public Randomness: The public randomness $r$ of $\pi$, as well as an additional sequence of
                   uniformly random numbers $r' = \rho_1, \dots, \rho_{\|\pi\|} \in [0, 1]$.

Let $m$ be the empty string;
while $|m| < \|\pi\|$ do
    Set $t = |m|$;
    for $i = t + 1, \dots, \min\{k + t, \|\pi\|\}$ do
        if $m_i$ is sent by Bob in $\pi$ then Alice checks if $\rho_i < \pi(m_i = 1|xrm_{<i})$, and sets
        $m_i = 1$ if this is the case. Otherwise she sets $m_i = 0$;
        else Alice samples $m_i$ privately according to the distribution $\pi(m_i|xrm_{<i})$;
    end
    Alice sends the current transcript $m$ to Bob;
    Bob computes the smallest index $j \in [\min\{k + t, \|\pi\|\}]$ such that $m_j$ would have been
    sent by Bob in $\pi$ and $\rho_j$ lies in the interval between $\pi(m_j = 1|xrm_{<j})$ and
    $\pi(m_j = 1|yrm_{<j})$. Bob can check this using $\rho_j, m, y, r$;
    Bob sends $j$ to Alice, or reports that there is no such $j$;
    Alice corrects $m$ if such $j$ is found, by flipping the bit $m_j$, and truncating $m = m_{\leq j}$;
end
return $m$;
```

Figure 1: Protocol τ simulating π

Proof. It is clear that $\tau(xyr) = \pi(xyr)$. For each $i \in [\|\pi\|]$, if m_i is to be sent by Alice in π, then $m_i = 1$ exactly when $\rho_i < \pi(m_i|xrm_{<i})$, and if m_i is to be sent by Bob in π, then $m_i = 1$ exactly when $\rho_i < \pi(m_i|yrm_{<i})$. Thus, if m_i is to be sent by Alice in π, $\tau(m_i|xyrm_{<i}) = \pi(m_i|xrm_{\leq i}) = \pi(m_i|xyrm_{\leq i})$. On the other hand, if m_i is to be sent by Bob in π, then $\tau(m_i|xyrm_{<i}) = \pi(m_i|yrm_{<i}) = \pi(m_i|xyrm_{<i})$. Thus

$$\tau(xyrm) = \tau(xyr) \cdot \prod_{i=1}^{\|\pi\|} \tau(m_i|xyrm_{<i}) = \pi(xyr) \cdot \prod_{i=1}^{\|\pi\|} \pi(m_i|xyrm_{<i}) = \pi(xyrm).$$

□

Let L denote the number of *mistake* indices j reported to Alice by Bob in τ. Then we have:

Lemma 17. *The number of rounds in τ is at most $\|\pi\|/k + L = \sqrt[4]{\|\pi\|^3 \cdot I_\pi^B} + L$.*

Proof. There are L rounds where Alice needs to truncate m. In every other round, at least k new bits of the messages of π are sampled, so there can be at most $\|\pi\|/k$ additional rounds of communication. □

Given the last lemma, we can bound the number of rounds in the protocol by bounding L:

Lemma 18. $\mathbb{E}[L] \leq \sqrt{I_\pi^B \cdot \|\pi\|}$.

Proof. Let L_i denote the indicator random variable for the event that the i'th index of the message is corrected by Bob in τ, so $L = \sum_{i=1}^{\|\pi\|} L_i$. If the i'th message is sent by Alice, then $L_i = 0$. On the other hand, if it is sent by Bob, by Proposition 8, we get

$$\begin{aligned}
\tau(L_i = 1) &= \mathop{\mathbb{E}}_{xyrm_{<i}} \left[|\pi(m_i|xrm_{<i}) - \pi(m_i|yrm_{<i})| \right] \\
&= \mathop{\mathbb{E}}_{xyrm_{<i}} \left[|\pi(m_i|xrm_{<i}) - \pi(m_i|xyrm_{<i})| \right] \\
&\le \sqrt{I(M_i; Y|XRM_{<i})}.
\end{aligned}$$

The penultimate inequality is true since m_i is sampled by Bob in π, hence is independent of x conditioned on $y, r, m_{<i}$.

Thus by linearity of expectation and the Cauchy Schwartz inequality, we get

$$\begin{aligned}
\mathbb{E}[L] &\le \sum_{i=1}^{\|\pi\|} \sqrt{I(M_i; Y|XRM_{<i})} \\
&\le \sqrt{\|\pi\| \cdot \sum_{i=1}^{\|\pi\|} I(M_i; Y|XRM_{<i})} \\
&= \sqrt{\|\pi\| \cdot I(M; Y|XR)},
\end{aligned}$$

where here we used the chain rule (Proposition 6) in the last step. □

Lemma 18 and Lemma 17 together imply that the expected number of rounds in τ is at most $\sqrt{I_\pi^B \cdot \|\pi\|} + \|\pi\|/k$. It only remains to bound the internal information of τ:

Lemma 19. *The internal information of τ is at most*

$$\frac{\|\pi\| \log \|\pi\|}{k} + k\sqrt{I_\pi^B \cdot \|\pi\|} + I_\pi^A + 2 \log \|\pi\| \sqrt{\|\pi\| \cdot I_\pi^B} + 3$$

Proof. Recall that R' denotes the sequence of numbers $\rho_1, \ldots, \rho_{\|\pi\|}$. The public randomness of τ consists of R, R'. Let Z denote the messages exchanged in the protocol τ. Let Z_A denote the bits sent by Alice, and Z_B denote the bits sent by Bob. Then the information learnt by Alice can be expressed as

$$I_\tau(Z_A Z_B; Y|XRR') = I_\tau(Z_B; Y|XRR') + I_\tau(Z_A; Y|XRR'Z_B), \tag{1}$$

by the chain rule. The the second term of (1) is 0, since Alice's messages are independent of Y, given Bob's messages and Alice's inputs. For the first term, we use Proposition 11 to bound it by $\mathbb{E}[|Z_B|]$. The total number of rounds of the protocol is at most $L + \|\pi\|/k$, since every round where there is no mistake must simulate at least k messages from the protocol. Thus $\mathbb{E}[|Z_B|] \le (\mathbb{E}[L] + \|\pi\|/k) \log \|\pi\| \le \sqrt{I_\pi^B \cdot \|\pi\|} \log \|\pi\| + \frac{\|\pi\| \log \|\pi\|}{k}$, by Lemma 18.

Next we bound the information learnt by Bob in τ. This can be written

$$I_\tau(Z; X|YRR') = \sum_{i=1}^{|Z|} I_\tau(Z_i; X|Z_{<i}YRR') = \mathop{\mathbb{E}}_{xyrr'z} \left[\sum_{i=1}^{|z|} \mathbb{D}\left(\frac{z_i|xyrr'z_{<i}}{z_i|yrr'z_{<i}} \right) \right] \tag{2}$$

by the chain rule.

Claim 20. *For $x \in [0, 1/2]$ $\log(1/(1-x)) \leq 3x$*

Proof. Let $T = \mathsf{ln}(1/(1-x))$. The Taylor expansion of $\mathsf{ln}(1/(1-x))$ gives,

$$\begin{aligned} T = \mathsf{ln}(1/(1-x)) &= x + x^2/2 + x^3/3 + \cdots \\ &= x + x(x/2 + x^2/3 + x^3/4 + \cdots) \\ &\leq x + x \cdot T. \end{aligned}$$

This implies, $T \leq x/(1-x)$. Since, $x \leq 1/2$, $T \leq 2x$. Now,

$$\log(1/(1-x)) = \mathsf{ln}(1/(1-x))/\mathsf{ln}(2) \leq 1.5\mathsf{ln}(1/(1-x)) \leq 3x,$$

which follows from the fact that $1/\mathsf{ln}(2) \leq 1.5$ □

Claim 21. *If $m_{<j}$ represents the messages of π sampled by the simulation τ at the point the messages $z_{<i}$ were sent, then*

$$\mathop{\mathbb{E}}_{r'|xyrz_{<i}}\left[\mathbb{D}\left(\frac{\tau(z_i|xyrr'z_{<i})}{\tau(z_i|yrr'z_{<i})}\right)\right]\begin{cases} = 0 & \textit{if Bob sends } z_i, \\ \leq 1 & \textit{if } z_i \textit{ is discarded,} \\ \leq \mathbb{D}\left(\frac{\pi(m_j|xyrm_{<j})}{\pi(m_j|yrm_{<j})}\right) & \textit{if Alice sends } m_j, \\ \leq \sqrt{\mathbb{D}\left(\frac{\pi(m_j|xyrm_{<j})}{\pi(m_j|xrm_{<j})}\right)\log\|\pi\| + \frac{3}{\|\pi\|}} & \textit{if Bob sends } m_j \end{cases}$$

Proof. When z_i is sent by Bob, both distributions are the same, so the divergence is 0.

To prove the remaining cases, we apply Proposition 10. When z_i is to be discarded, set $q(z_i|yrr'z_{<i})$ to be the uniform distribution on bits. When q is the uniform distribution on the bits, $\mathbb{D}\left(\frac{p}{q}\right) = 1 - \mathsf{h}(p(0)) \leq 1$. Since $\mathbb{D}\left(\frac{p}{q}\right) \leq 1$ for any p, the bound follows using Proposition 10. When m_j is sent by Alice in π, observe that the distribution of $\tau(z_i|xyrr'z_{<i})$ is exactly the same as the distribution of $\pi(m_j|xyrm_{<j})$. Set $q(z_i|yrr'z_{<i}) = \pi(m_j|yrm_{<j})$. The bound follows by Proposition 10. For the last case, observe that in τ, z_i is determined by $xyrr'm_{<j}$, since $z_i = 1$ exactly when $\rho_i < \pi(m_j = 1|xrm_{<j})$. Set

$$b(\rho_i, y, r, r', m_{<j}) = \begin{cases} 1 & \text{if } \rho_i < \pi(m_j = 1|yrm_{<j}), \\ 0 & \text{otherwise} \end{cases}$$

and

$$q(z_i|yrr'z_{<i}) = \begin{cases} 1 - 1/\|\pi\| & \text{if } b(\rho_i, y, r, r', m_{<j}) = z_i, \\ 1/\|\pi\| & \text{otherwise.} \end{cases}$$

When ρ_i is in between $\pi(m_j = 1|yrm_{<j})$ and $\pi(m_j = 1|xrm_{<j})$,

$$\mathbb{D}\left(\frac{\tau(z_i|xyrr'z_{<i})}{q(z_i|yrr'z_{<i}))}\right) = \log(1/(1/\|\pi\|)) = \log\|\pi\|,$$

for all z_i with positive probability. When ρ_i is not in between those two quantities,

$$\mathbb{D}\left(\frac{\tau(z_i|xyrr'z_{<i})}{q(z_i|yrr'z_{<i}))}\right) \leq \log\frac{1}{1-1/\|\pi\|} \leq 3/\|\pi\|.$$

which follows from Claim 20 and the assumption that $\|\pi\| > 2$. It is safe to assume that $\|\pi\| > 2$, as the compression is trivial for protocols with communication at most 2.

The probability of the first case is at most $\sqrt{\mathbb{D}\left(\frac{\pi(m_j|xyrm_{<j})}{\pi(m_j|yrm_{<j})}\right)}$, by Proposition 7. □

Now given Z, call i good if the message Z_i does not correspond to a mistake and is not discarded by the simulation. Let G denote the set of good indices. We have,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{xyrr'z}\left[\sum_{i=1}^{|z|}\mathbb{D}\left(\frac{\tau(z_i|xyrr'z_{<i})}{\tau(z_i|yrr'z_{<i})}\right)\right]\\
&= \mathop{\mathbb{E}}_{xyrz}\left[\sum_{i=1}^{|z|}\mathop{\mathbb{E}}_{r'|xyrz_{<i}}\left[\mathbb{D}\left(\frac{\tau(z_i|xyrr'z_{<i})}{\tau(z_i|yrr'z_{<i})}\right)\right]\right]\\
&= \mathop{\mathbb{E}}_{xyrz}\left[\sum_{i\in G}\mathop{\mathbb{E}}_{r'|xyrz_{<i}}\left[\mathbb{D}\left(\frac{\tau(z_i|xyrr'z_{<i})}{\tau(z_i|yrr'z_{<i})}\right)\right]\right] + \mathop{\mathbb{E}}_{xyrz}\left[\sum_{i\notin G}\mathop{\mathbb{E}}_{r'|xyrz_{<i}}\left[\mathbb{D}\left(\frac{\tau(z_i|xyrr'z_{<i})}{\tau(z_i|yrr'z_{<i})}\right)\right]\right]
\end{aligned}$$

For every Z, at most $k\cdot L$ indices are discarded. This is because, at most k indices are discarded for every round with a mistake. Then by Claim 21,

$$\mathop{\mathbb{E}}_{xyrz}\left[\sum_{i\notin G}\mathop{\mathbb{E}}_{r'|xyrz_{<i}}\left[\mathbb{D}\left(\frac{\tau(z_i|xyrr'z_{<i})}{\tau(z_i|yrr'z_{<i})}\right)\right]\right] \leq k\,\mathbb{E}\left[L\right]$$

For every index $i \in G$, by Claim 21

$$\begin{aligned}
&\mathop{\mathbb{E}}_{xyrz}\left[\sum_{i\in G}\mathop{\mathbb{E}}_{r'|xyrz_{<i}}\left[\mathbb{D}\left(\frac{\tau(z_i|xyrr'z_{<i})}{\tau(z_i|yrr'z_{<i})}\right)\right]\right]\\
&\leq \mathop{\mathbb{E}}_{\pi(mxyr)}\left[\mathbb{D}\left(\frac{\pi(m_j|xyrm_{<j})}{\pi(m_j|yrm_{<j})}\right)\right] + \mathop{\mathbb{E}}_{\pi(mxyr)}\left[\sum_{j=1}^{\|\pi\|}\sqrt{\mathbb{D}\left(\frac{\pi(m_j|xyrm_{<j})}{\pi(m_j|xrm_{<j})}\right)}\log\|\pi\| + \frac{3}{\|\pi\|}\right]\\
&= \mathop{\mathbb{E}}_{\pi(mxyr)}\left[\mathbb{D}\left(\frac{\pi(m_j|xyrm_{<j})}{\pi(m_j|yrm_{<j})}\right)\right] + \log\|\pi\|\mathop{\mathbb{E}}_{\pi(mxyr)}\left[\sum_{j=1}^{\|\pi\|}\sqrt{\mathbb{D}\left(\frac{\pi(m_j|xyr)}{\pi(m_j|xr)}\right)}\right] + 3\\
&\leq \mathop{\mathbb{E}}_{\pi(mxyr)}\left[\mathbb{D}\left(\frac{\pi(m_j|xyrm_{<j})}{\pi(m_j|yrm_{<j})}\right)\right] + \log\|\pi\|\sqrt{\|\pi\|\cdot\mathop{\mathbb{E}}_{\pi(mxyr)}\left[\sum_{j=1}^{\|\pi\|}\mathbb{D}\left(\frac{\pi(m_j|xyr)}{\pi(m_j|xr)}\right)\right]} + 3\\
&= I_\pi^A + \log\|\pi\|\sqrt{\|\pi\|\cdot I_\pi^B} + 3,
\end{aligned}$$

where the penultimate inequality follows from an application of Cauchy Schwartz inequality, and the last inequality follows from the definition of I_π^A, I_π^B. □

Corollary 22. *Given any protocol π, there exists a protocol that simulates π with $2 \cdot \sqrt[4]{\|\pi\|^3 \cdot I_\pi^B}$ number of rounds in expectation, and internal information at most $I_\pi^A + \sqrt[4]{\|\pi\|^3 \cdot I_\pi^B} + \sqrt[4]{\|\pi\|^3 \cdot I_\pi^B} \cdot$ $\log \|\pi\| + 2 \cdot \sqrt{\|\pi\| \cdot I_\pi^B} \cdot \log \|\pi\| + 3$.*

Proof. Set $k = \sqrt[4]{\|\pi\|/I_\pi^B}$. By Theorem 15, we get that the expected number of rounds of the simulation is at most $\sqrt{I_\pi^B \cdot \|\pi\|} + \sqrt[4]{\|\pi\|^3 \cdot I_\pi^B} \leq 2\sqrt[4]{\|\pi\|^3 \cdot I_\pi^B}$(since $I_\pi^B \leq \|\pi\|$) and the internal information is at most $I_\pi^A + \sqrt[4]{\|\pi\|^3 \cdot I_\pi^B} + \sqrt[4]{\|\pi\|^3 \cdot I_\pi^B} \cdot \log \|\pi\| + 2 \cdot \sqrt{\|\pi\| \cdot I_\pi^B} \cdot \log \|\pi\| + 3$ □

Applying Theorem 13 to the simulation guaranteed by Corollary 22, gives a simulation with communication bounded by

$$O\left(I_\pi^A + \sqrt[4]{\|\pi\|^3 \cdot I_\pi^B} \cdot \log(1/\epsilon) + \sqrt[4]{\|\pi\|^3 \cdot I_\pi^B} \cdot \log \|\pi\| + \sqrt{\|\pi\| \cdot I_\pi^B} \cdot \log \|\pi\|\right),$$

as required in Theorem 1.

4 Compresing Protocols with Asymmetric Information - II

In this section, we prove Theorem 2, (i.e) we show how to simulate π with communication $I_\pi^A \cdot 2^{O(I_\pi^B)}$.

Theorem 23. *Let U be a finite set. Let $p_A, p_B, q_A, q_B : U \to [0,1]$ be such that $\forall z \in U$, $\mu(z) = p_A(z)q_B(z)$, $p(z) = p_A(z)p_B(z)$ and $q(z) = q_A(z)q_B(z)$ are distributions. There exists a randomized protocol with inputs p_A, p_B to Alice and q_A, q_B to Bob, such that*

- *Both Alice and Bob either accept and compute (possibly different) samples $z \in U$, or abort the protocol.*
- $\Pr[$*Both parties accept*$] \geq 2^{-O\left(\mathbb{D}\left(\frac{\mu}{q}\right)+1\right)}$.
- *Given that both parties accept, the distribution of their samples is 0.35-close in ℓ_1 distance to the distribution where both parties sample the same sample from $\mu(z)$.*

Let $I^B = \mathbb{D}\left(\frac{\mu}{p}\right)$ and $I^A = \mathbb{D}\left(\frac{\mu}{q}\right)$. Figure 2 describes the randomized protocol promised by the lemma. Let

$$\mathcal{G} = \left\{z \middle| 2^{20(I^B+1)} \cdot p(z) \geq \mu(z), 2^{20(I^A+1)} \cdot q(z) \geq \mu(z)\right\}.$$

We need the following simple claim, which was proved in [Bra11].

Claim 24. $\mu(G) \geq \frac{9}{10}$.

We proceed with the analysis of the simulation.

Lemma 25. $\Pr[i^*$*is defined*$] \geq 1 - e^{-11}$.

Simulation π

Public randomness: A sequence of $L = 10\cdot|U|\cdot 2^{20(I^A+1)}$ tuples $(z_i, a_i, b_i) \in U \times \left[0, 2^{20(I^A+1)}\right] \times \left[0, 2^{20(I^B+1)}\right]$, for $i = 1, 2, \ldots, L$, and a random function $h : [L] \to \left[2^{40(I^A+1)}\right]$.

1. Alice computes the set $\mathcal{A} = \left\{ i \middle| a_i \leq p_A(z_i), b_i \leq p_B(z_i) \cdot 2^{20(I^B+1)} \right\}$, and Bob computes the set $\mathcal{B} = \left\{ i \middle| a_i \leq q_A(z_i) \cdot 2^{20(I^A+1)}, b_i \leq q_B(z_i) \right\}$.
2. Alice computes i^*, the the smallest element of $\mathcal{A}$.
3. Alice sends $h(i^*)$ to Bob.
4. If there is a unique $i \in \mathcal{B}$ such that $h(i) = h(i^*)$, Bob accepts and assumes that the outcome of the protocol is z_i. Otherwise Bob aborts.

Figure 2: The sampling procedure

Proof. For each i, we have

$$\begin{aligned}
\Pr[i \in \mathcal{A}] &= \sum_{z \in U} \frac{p_A(z) p_B(z)}{|U| \cdot 2^{20(I^A+1)}} \\
&= \frac{1}{|U| \cdot 2^{20(I^A+1)}} \sum_{z \in U} p_A(z) p_B(z) \\
&= \frac{1}{|U| \cdot 2^{20(I^A+1)}}. && \text{(since } p \text{ is a distribution)}
\end{aligned}$$

The probability that i^* is not defined is exactly equal to the probability that $i \notin \mathcal{A}$ for all $1 \leq i \leq L$. Thus,

$$\begin{aligned}
\Pr[i^* \text{ not defined}] &= \left(1 - \frac{1}{|U| \cdot 2^{20(I^A+1)}}\right)^L \\
&\leq e^{-L/|U| \cdot 2^{20(I^A+1)}}. && \text{(using } (1-x)^n \leq e^{-xn}, x \geq 0\text{)} \\
&\leq e^{-10}.
\end{aligned}$$

□

Lemma 26. *For $z \in U$,*

$$\Pr[z_{i^*} = z \ \&\ i^* \in \mathcal{B} | i^* \textit{ is defined}] \leq \frac{\mu(z)}{2^{20(I^B+1)}},$$

with equality when $z \in \mathcal{G}$.

Proof.

$$\Pr[z_{i^*} = z \ \&\ i^* \in \mathcal{B} | i^* \text{ is defined}] = \Pr[z_{i^*} = z | i^* \text{ is defined}] \cdot \Pr[i^* \in \mathcal{B} | z_{i^*} = z]. \tag{3}$$

We have

$$\begin{aligned} \Pr[z_{i^*} = z | i^* \text{ is defined}] &= \frac{p_A(z) p_B(z) 2^{-20(I^A+1)}}{\sum_{z \in U} p_A(z) p_B(z) 2^{-20(I^A+1)}} \\ &= p_A(z) p_B(z). \end{aligned} \tag{4}$$

Let us now analyze $\Pr[i^* \in \mathcal{B} | z_{i^*} = z]$. If $z_{i^*} = z$, we have $a_{i^*} \leq p_A(z), b_{i^*} \leq p_B(z) 2^{20(I^B+1)}$. Thus $\Pr[i^* \in \mathcal{B} | z_{i^*} = z]$ is exactly

$$\begin{aligned} \Pr[i^* \in \mathcal{B} | z_i^* = z] &= \min\left\{ \frac{q_B(z)}{2^{20(I^B+1)} p_B(z)}, 1 \right\} \cdot \min\left\{ \frac{2^{20(I^A+1)} q_A(z)}{p_A(z)}, 1 \right\} \\ &\leq \frac{q_B(z)}{2^{20(I^B+1)} p_B(z)}. \end{aligned} \tag{5}$$

Equality holds in (5) when $z \in \mathcal{G}$, since for such z, $\frac{q_B(z)}{2^{20(I^B+1)} p_B(z)} \leq 1$ and $\frac{2^{20(I^A+1)} q_A(z)}{p_A(z)} \geq 1$. Therefore, using (4),(5),(3)

$$\Pr[z_{i^*} = z \ \&\ i^* \in \mathcal{B}] \leq p_A(z) p_B(z) \cdot \frac{q_B(z)}{2^{20(I^B+1)} p_B(z)} = \mu(z) / 2^{20(I^B+1)},$$

with equality for $z \in \mathcal{G}$. □

When i^* is defined, let E denote the event that i^* is the only possible index that is in $\mathcal{B}$ and consistent with the message Bob receives, namely: $\forall i \neq i^*, i \in \mathcal{B} \Rightarrow h(i) \neq h(i^*)$. Then we have

Lemma 27. $\Pr[\neg E | i^*$ *is defined*$] \leq \frac{L 2^{-40(I^A+1) - 20(I^B+1)}}{|U| \Pr[i^* \textit{ is defined}]}$.

Proof. Given that i^* is defined, probability that $i \in \mathcal{B}$ and $h(i) = h(i^*)$ is at most $\frac{\Pr[i \in \mathcal{B}] \cdot 2^{-40(I^A+1)}}{\Pr[i^* \text{ is defined}]}$. Thus,

$$\begin{aligned} \frac{\Pr[i \in \mathcal{B}] \cdot 2^{-40(I^A+1)}}{\Pr[i^* \text{ is defined}]} &= \frac{1}{|U| \cdot \Pr[i^* \text{ is defined}]} \cdot \sum_{z \in U} q_A(z) \cdot 2^{-20(I^B+1)} q_B(z) 2^{-40(I^A+1)} \\ &\leq \frac{2^{-40(I^A+1) - 20(I^B+1)}}{|U| \cdot \Pr[i^* \text{ is defined}]}. \end{aligned}$$

Thus, by the union bound, the probability than any such i is accepted by Bob is at most $\frac{L 2^{-40(I^A+1) - 20(I^B+1)}}{|U| \Pr[i^* \text{ is defined}]}$
□

Lemma 28. $\Pr[i^* \in \mathcal{B} | i^*$ *is defined*$] \geq \mu(\mathcal{G}) \cdot 2^{-20(I^B+1)}$

Proof.

$$\begin{aligned}\Pr[i^* \in \mathcal{B} | i^* \text{ is defined}] &= \sum_{z \in |U|} \Pr[z_{i^*} = z, i^* \in \mathcal{B} | i^* \text{ is defined}] \\ &\geq \sum_{z \in \mathcal{G}} \mu(z)/2^{20(I^B+1)} && \text{(by Lemma 26)} \\ &= \mu(\mathcal{G}) \cdot 2^{-20(I^B+1)}. && \text{(by Lemma 25)}\end{aligned}$$

□

Lemma 29. $\Pr[\textit{Both parties accept} | i^* \textit{is defined}] \geq \frac{8}{10} \cdot 2^{-20(I^B+1)}$

Proof.

$$\begin{aligned}&\Pr[\text{Both parties accept} | i^* \text{is defined}] \\ &\geq \Pr[i^* \in \mathcal{B} | i^* \text{ is defined}] - \Pr[\neg E | i^* \text{is defined}] \\ &\geq \mu(\mathcal{G})/2^{20(I^B+1)} - \Pr[\neg E | i^* \text{is defined}] && \text{(by Lemma 28)} \\ &\geq \frac{9}{10} \cdot 2^{-20(I^B+1)} - \frac{10}{1-e^{-10}} \cdot 2^{-20(I^A+1)-20(I^B+1)} && \text{(by Lemma 27, Claim 24)} \\ &> \frac{8}{10} \cdot 2^{-20(I^B+1)}.\end{aligned}$$

where the last inequality follows from $I^A \geq 0$. □

Lemma 30. $\Pr[\textit{Both parties accept}] \geq \frac{7}{10} \cdot 2^{-20(I^B+1)}$.

Proof.

$$\begin{aligned}\Pr[\text{Both parties accept}] &= \Pr[i^* \text{is defined}] \Pr[\text{Both parties accept} | i^* \text{is defined}] \\ &\geq \frac{8(1-e^{-10})}{10 \cdot 2^{20(I^B+1)}} > \frac{7}{10} \cdot 2^{-20(I^B+1)},\end{aligned}$$

which follows from Lemma 29 and Lemma 25. □

Lemma 31. $|\pi(z_{i^*} = z | i^* \in \mathcal{B}) - \mu(z)| \leq 2(1-\mu(\mathcal{G}))/\mu(\mathcal{G})$.

Proof. By Lemma 26,

$$\Pr[i^* \in \mathcal{B} | i^* \text{ is defined}] \cdot \Pr[z_{i^*} = z | i^* \in \mathcal{B}] \leq \mu(z) \cdot 2^{-20(I^B+1)}.$$

Combining the above inequality and Lemma 28,

$$\Pr[z_{i^*} = z | i^* \in \mathcal{B}] \leq \mu(z)/\mu(\mathcal{G}).$$

For any set $T \subseteq U$,

$$\sum_{z \in T} \pi(z_{i^*} = z | i^* \in \mathcal{B}) - \mu(z) \leq \sum_{z \in T} \mu(z)/\mu(\mathcal{G}) - \mu(z) \leq (1-\mu(\mathcal{G}))/\mu(\mathcal{G}),$$

where the last inequality follows from μ being a distribution. Therefore,

$$|\pi(z_{i^*} = z | i^* \in \mathcal{B}) - \mu(z)| = 2 \max_T \left(\sum_{z \in T} \pi(z_{i^*} = z | i^* \in \mathcal{B}) - \mu(z) \right) \leq 2(1-\mu(\mathcal{G}))/\mu(\mathcal{G}).$$

□

Lemma 32.

$$\Pr[\neg E | i^* \in \mathcal{B}] < 0.01, \qquad \Pr[\neg E | \textit{Both parties accept}] < 0.01$$

Proof. We have,

$$\begin{aligned}
&\Pr[\neg E | i^* \in \mathcal{B}] \\
&\Pr[\neg E | i^* \in \mathcal{B}, i^* \text{ is defined}] \\
&= \frac{\Pr[\neg E \ \&\ i^* \in \mathcal{B} | i^* \text{ is defined}]}{\Pr[i^* \in \mathcal{B} | i^* \text{ is defined}]} \\
&\leq \frac{\Pr[\neg E | i^* \text{is defined}]}{\Pr[i^* \in \mathcal{B} | i^* \text{ is defined}]} \\
&\leq \frac{10}{\mu(\mathcal{G})(1 - e^{-10})} \cdot 2^{-20(I^A+1)} \\
&< 0.01.
\end{aligned}$$

The penultimate inequality follows from Lemmas 27, 28.

Simultaneously,

$$\begin{aligned}
&\Pr[\neg E | \text{Both parties accept}] \\
&= \Pr[\neg E | \text{Both parties accept}, i^* \text{ is defined}] \\
&= \frac{\Pr[\neg E \ \&\ \text{Both parties accept} | i^* \text{is defined}]}{\Pr[\text{Both parties accept} | i^* \text{is defined}]} \\
&\leq \frac{\Pr[\neg E | i^* \text{is defined}]}{\Pr[\text{Both parties accept} | i^* \text{is defined}]} \\
&\leq \frac{100}{8(1 - e^{-10})} \cdot 2^{-20(I^A+1)} \\
&< 0.01.
\end{aligned}$$

The penultimate inequality follows from Lemmas 27, 29.

□

Lemma 33. $|\mu(z) - \pi(z | \textit{Both parties accept})| < 0.35$

Proof. Applying Proposition 5 twice gives,

$$\begin{aligned}
&|\mu(z) - \pi(z | \text{Both parties accept})| \\
&\leq |\pi(z_{i^*} = z | i^* \in \mathcal{B}, E) - \mu(z)| + 4\Pr[\neg E | \text{Both parties accept}] \\
&\leq |\pi(z_{i^*} = z | i^* \in \mathcal{B}) - \mu(z)| + 4\Pr[\neg E | i^* \in \mathcal{B}] + 4\Pr[\neg E | \text{Both parties accept}].
\end{aligned}$$

Applying Lemmas 31, 32 give,

$$|\mu(z) - \pi(z | \text{Both parties accept})| < 2(1 - \mu(\mathcal{G}))/\mu(\mathcal{G}) + 0.04 + 0.04 < 0.35,$$

as required. □

In Theorem 23, property 2 is guaranteed by Lemma 30 and property 3 is guaranteed by Lemma 33.

4.1 Proof of Theorem 2

Define U to be the set of all transcripts of protocol π. For every $m \in U$, define $\pi_x(m) = \pi(m|x)$, $\pi_y(m) = \pi(m|y)$, $\pi_{xy}(m) = \pi(m|xy)$. We have,

$$I_\pi(X; M | YR) = \mathbb{E}_{xyr}\left[\mathbb{D}\left(\frac{\pi_{xy}(m|r)}{\pi_y(m|r)}\right)\right].$$

Define $\Gamma = \left\{(x,y,r) \middle| \mathbb{D}\left(\frac{\pi_{xy}(m|r)}{\pi_y(m|r)}\right) \leq 50 \cdot I_\pi^A, \mathbb{D}\left(\frac{\pi_{xy}(m|r)}{\pi_x(m|r)}\right) \leq 50 \cdot I_\pi^B\right\}$. By an union bound and Markov's inequality,

$$\Pr[(x,y,r) \in \Gamma] \geq \frac{24}{25} \tag{6}$$

By Theorem 23, with $\mu(m) = \pi_{xy}(m|r)$, $p(m) = \pi_x(m|r)$, $q(m) = \pi_y(m|r)$, implies that for every $(x,y,r) \in \Gamma$, there exists a constant c and a randomized protocol τ with communication $O(I_\pi^A)$ that samples a a transcript m such that

$$|\pi(m) - \tau(m|\text{Both parties accept})| < 0.35 \qquad \Pr[\text{Both parties accept in } \tau] \geq 2^{-c(I_\pi^B+1)} \tag{7}$$

Lemma 34. $\mathbb{E}_{\pi(xyr)}|\pi(m|xyr) - \tau(m|xyr, \textit{Both parties accept})| < \frac{1}{25} + 0.35$

Proof. τ guarantees that for $(x,y,r) \in \Gamma$, $|\pi(m|xyr) - \tau(m|xyr, \text{Both parties accept})| < 0.35$. Therefore,

$$\begin{aligned}
&\mathbb{E}_{\pi(xyr)}|\pi(m|xyr) - \tau(m|xyr, \text{Both parties accept})| \\
&\leq \Pr[(x,y,r) \in \Gamma] \cdot \max_{(x,y,r)\in\Gamma} |\pi(z|xyr) - \tau(z|xyr, \text{Both parties accept})| + \Pr[(x,y,r) \notin \Gamma] \\
&< 0.35 + \frac{1}{25}, && \text{(by (6))}
\end{aligned}$$

□

```
Input: $x, y$, the inputs to $\pi$. A parameter $t$.
Public Randomness: Sequence of strings $L_1, \cdots, L_t, r$. A random transcript $m'$.
i=1;
while $i \leq t$ do
    Run protocol $\tau$ with $L_i$ as the public random tape;
    if $\tau$ Accepts then return the output of $\tau$;
    else i=i+1;
end
return $m'$;
```

Figure 3: Protocol Σ simulating π

We now show a simulation of π. Run Σ shown in Figure 3 with parameter $t = 10 \cdot 2^{c(I^B+1)}$. We have,

$$CC(\Sigma) \leq t \cdot CC(\tau) = 10 \cdot I^A 2^{c(I^B+1)}.$$

Let J be the value of i at the time of termination of Σ.

Lemma 35. $\Pr[J = t+1] \leq 1/25 + e^{-10}$

Proof. Conditioned on $(x, y) \in \Gamma$, the probability that τ does not accept in iteration i equals $1 - \Pr[\text{Both parties accepts}]$. Therefore,

$$\begin{aligned}
&\Pr[J = t+1 | (x, y, r) \in \Gamma] \\
&= (1 - \Pr[\text{Both party accepts in } \tau])^t \\
&\le \left(1 - 2^{-c(I^B+1)}\right)^t && \text{(by (7))} \\
&\le e^{-10} && ((1-x)^n \le e^{-xn}, x > 0). \qquad (8)
\end{aligned}$$

Now,

$$\begin{aligned}
&\Pr[J = t+1] \\
&= \Pr[(x, y, r) \in \Gamma] \cdot \Pr[J = t+1 | (x, y, r) \in \Gamma] + \Pr[(x, y, r) \notin \Gamma] \cdot \Pr[J = t+1 | (x, y, r) \notin \Gamma] \\
&\le e^{-10} + 1/25,
\end{aligned}$$

where the last inequality follows from (6), (8). □

Let us now analyze the ℓ_1 distance between Σ and π.

Lemma 36. $|\Sigma - \pi| < 3/25 + 2e^{-10} + 0.35$

Proof. Conditioning on $J \le t$, we know that Σ's output corresponds to the output of τ. Therefore,

$$|\pi(m) - \Sigma(m | J \le t)| = \mathbb{E}_{\pi(xyr)} |\pi(m|xyr) - \tau(m|xyr, \text{Both parties accept})|$$

By Proposition 4 and Lemma 35,

$$\begin{aligned}
|\Sigma - \pi| &\le 2\Pr[J = t+1] + |\pi(z) - \Sigma(z | J \le t)| \\
&= 2\Pr[J = t+1] + \mathbb{E}_{\pi(xyr)} |\pi(m|xyr) - \tau(m|xyr, \text{Both parties accept}) \\
&< 2/25 + 2e^{-10} + 1/25 + 0.35,
\end{aligned}$$

where the last inequality follows from Lemma 35 and Lemma 34. □

This concludes the proof of the theorem.

4.2 A Rectangle Lower Bound

In this subsection, we show a corollary to Theorem 2. The simulation of Theorem 2 shows the existence of almost monochromatic rectangles of the right dimension.

Corollary 37. *Given a protocol π with internal information (I_A, I_B) and inputs drawn from μ, there exists a zero communication protocol for sampling a message m such that,*

- $\mu(\textit{Alice Accepts}) = 2^{-O(I_A)}$
- $\mu(\textit{Bob Accepts} | \textit{Alice Accepts}) \ge 2^{-O(I_B)}$

Moreover, given that both parties accept, the distribution of their samples is $\delta-$close in ℓ_1 distance to the distribution where both parties sample consistently from $\pi(m)$.

Proof. First we wish to fix the public randomness of π, such that the internal information and the error bound are within limit. We have,

$$I_A = \mathop{\mathbb{E}}_{r}\left[I(M;X|Yr)\right], \qquad I_B = \mathop{\mathbb{E}}_{r}\left[I(M;Y|Xr)\right], \qquad \mathop{\mathbb{E}}_{r}\left[\mu(\pi(x,y) \neq f(x,y)|r)\right] \leq \epsilon$$

By Markov's inequality,

$$\begin{aligned} \Pr_r[I(M;X|Yr) > 3I_A] &< 1/3 \\ \Pr_r[I(M;Y|Xr) > 3I_B] &< 1/3 \\ \Pr_r[\mu(\pi(x,y) \neq f(x,y)|r) > 3\epsilon] &< 1/3. \end{aligned}$$

Therefore, by an union bound, there exits an r such that

$$I(M;X|Yr) \leq 3I_A, \qquad I(M;Y|Xr) \leq 3I_B, \qquad \mu(\pi(x,y) \neq f(x,y)|r) \leq 3\epsilon$$

After fixing r, we now have a protocol with internal information at most $(3I_A, 3I_B)$ and error at most 3ϵ and no public randomness.

The following observations are useful.

$$\begin{aligned} \pi(m|xy) &= \prod_{i:m_i \text{ sent by Alice}} \pi(m_i|xym_{<i}) \cdot \prod_{i:m_i \text{ sent by Bob}} \pi(m_i|xym_{<i}) \\ &= \prod_{i:m_i \text{ sent by Alice}} \pi(m_i|xm_{<i}) \cdot \prod_{i:m_i \text{ sent by Bob}} \pi(m_i|ym_{<i}), \end{aligned}$$

where the last inequality follows from the fact that Alice's messages depend only on x, the public randomness and the previous messages and Bob's messages depend only on y, public randomness and the previous messages.

Define,

$$\pi_A(m|x) = \prod_{i:m_i \text{ sent by Alice}} \pi(m_i|xm_{<i}); \qquad \pi_B(m|y) = \prod_{i:m_i \text{ sent by Bob}} \pi(m_i|ym_{<i})$$

In a similar fashion define,

$$\pi'_A(m|x) = \prod_{i:m_i \text{ sent by Bob}} \pi(m_i|ym_{<i}); \qquad \pi'_B(m|y) = \prod_{i:m_i \text{ sent by Alice}} \pi(m_i|ym_{<i})$$

We are now in a position to describe the simulation. Consider the simulation τ in Figure 4(take c to be a large constant):

Let $\mathcal{L}$ denote the sequence of L tuples in the public tape. Theorem 2 implies,

- $\mathbb{E}_{\mathcal{L},h}\left[\mu(\text{Alice Accepts})\right] \geq 2^{-O(I_A)}$ (this follows from h being a random hash function)
- $\mathbb{E}_{\mathcal{L},h}\left[\mu(\text{Bob Accepts}|\text{Alice Accepts})\right] \geq 2^{-O(I_B)}$
- Given both parties accept, the distribution of the samples is $\delta/3-$close in ℓ_1 distance to the distribution where both parties sample the same from $\pi(x,y)$, for all constant $\delta > 0$.

Simulation

Public randomness: A sequence of $L = 10 \cdot |U| \cdot 2^{c(I^A+1)}$ tuples $(z_i, a_i, b_i) \in \mathcal{M} \times \left[0, 2^{c(I^A+1)}\right] \times \left[0, 2^{c(I^B+1)}\right]$, for $i = 1, 2, \ldots, L$, r and a random function $h : [L] \to \left[2^{2c(I^A+1)}\right]$, where $\mathcal{M}$ is the set of all transcripts of π.

1. Alice computes the set $\mathcal{A} = \left\{i \middle| a_i \leq \pi_A(z_i|xr), b_i \leq \pi'_A(z_i|xr) \cdot 2^{c(I^B+1)}\right\}$, and Bob computes the set $\mathcal{B} = \left\{i \middle| a_i \leq \pi'_B(z_i|yr) \cdot 2^{c(I^A+1)}, b_i \leq \pi_B(z_i|yr)\right\}$.
2. Alice computes i^*, the the smallest element of $\mathcal{A}$.
3. Alice accepts if $h(i^*) = 0^{2c(I^A+1)}$.
4. If there is a unique $i \in \mathcal{B}$ such that $h(i) = 0^{2c(I^A+1)}$, Bob accepts and assumes that the outcome of the protocol is z_i. Otherwise Bob aborts.

Figure 4: The sampling procedure

The third condition translates to, $\mathbb{E}_{\mathcal{L},h}\left[|\pi(m|xy) - \tau(m|xy, \text{Both parties accept})|\right] \leq \delta/3$. This implies that there exists a fixing of $\mathcal{L}, h$ such that

$$\mu(\text{Alice Accepts}) \geq 2^{-O(I_A)}, \qquad \mu(\text{Bob Accepts}|\text{Alice Accepts}) \geq 2^{-O(I_B)}$$

$$|\pi(m|xy) - \tau(m|xy, \text{Both parties accept})| \leq \delta.$$

(the argument is similar to the one used for fixing of public randomness of π. One applies 3 Markov inequalities followed by an union bound) This completes the proof of the corollary. □

Corollary 38 (Restated)**.** *Given any randomized protocol π with internal information (I_A, I_B), there exists sets S, T and $z \in \{0, 1\}$ such that*

$$\mu(x \in S) \geq 2^{-O(I_A)}, \qquad \mu(y \in T|x \in S) \geq 2^{-O(I_B)}$$

$$\mu(f(x,y) \neq z|S \times T) \leq 3\epsilon + \delta,$$

where ϵ is the error incurred by π under μ and $\delta > 0$, a constant.

Proof. The protocol in Corollary 37 is deterministic. Any transcript in a deterministic protocol corresponds to a rectangle. Therefore, the state of both parties accepting, corresponds to a rectangle $S \times T$ with

$$\mu(x \in S) \geq 2^{-O(I_A)}, \qquad \mu(y \in T|x \in S) \geq 2^{-O(I_B)}.$$

In addition, conditioning on the event that both parties accept, the output $z \in \{0, 1\}$ of the protocol (the value of the function that both parties agree on) has the property that $\mu(f(x,y) \neq z) \leq 3(\epsilon + \delta)$, where ϵ is the error incurred by the the protocol π under μ. □

4.2.1 Application - Lower Bounds for Lopsided Set Disjointness

In this subsection, we use the rectangle lower bound to reprove the well known lower bound for *lopsided* set disjointness by Pătraşcu in [Păt11]. The problem of lopsided set disjointness is defined as follows,

Definition 39. *The set disjointness function on sets $x, y \subseteq [NB]$ is*

$$\mathsf{SD}(x,y) = \begin{cases} 1 & \textit{if } x \cap y = \emptyset, \\ 0 & \textit{otherwise.} \end{cases}$$

Lopsided set disjointness(LSD) is a restricted version of the problem where we are promised that $x, y \subseteq [NB]$ and $|x| = N$. The following bound proved by Patrascu [Păt11] has found many applications to proving data structure lower bounds. The bound was shown to be tight by Saglam [Sag14].

Theorem 40 ([Păt11])**.** *For any protocol computing* LSD *with error probability ϵ, one of the following holds,*

- *Alice communicates at least $\gamma N \log B$ bits.*
- *Bob communicates at least $NB^{1-c\gamma}$ bits.*

where $c = c_1 + 1 + \frac{1}{\gamma \log B}\left(\log 2c_1 - \log\left(1 - \mathsf{h}(12\epsilon + \delta)\right)\right)$, for a constant c_1 and $B = \Omega(1)$.

Here we give a slightly different proof of Theorem 40, using Corollary 37. The universe is taken to be $\left(2^{[B]}\right)^N$, the cartesian product of N power sets of $[B]$. $(x, y) \in \left(2^{[B]}, 2^{[B]}\right)$ is restricted tuples with $|x| = 1$ and y takes exactly one element from the pair $(2k, 2k+1)$, for all $2k, 2k+1 \in [B]$. We define two distributions ψ and μ on (x, y) as follows,

- ψ is a uniform distribution on all such pairs (x, y), with $\mathsf{LSD}(x, y) = 1$.
- μ is a uniform distribution on all such pairs (x, y).

The hard distribution for disjointness μ_h is one where $i \in [N]$ chosen at random, with (x_i, y_i) drawn from distribution μ and rest of the coordinates (x_j, y_j), for $j \in [N] \setminus \{i\}$ is drawn i.i.d from ψ.

Having described the hard distribution, we are all set to describe the proof of Theorem 40.

Proof. We assume the contrary that there exists a protocol π computing $\mathsf{LSD}(x, y)$ on the distribution μ_h with Alice communicating $a < \gamma N \log B$ bits and Bob communicating $b < NB^{1-c\gamma}$ bits.

We now use protocol π to compute LSD on a single block. Consider the case when the inputs (x, y) is drawn according to the distribution ψ. We know that protocol π computes LSD on inputs drawn i.i.d in ψ with information $\left(I^A, I^B\right) < \left(\gamma N \log B, NB^{1-c\gamma}\right)$. By Theorem 14, there exist a protocol τ computing $\mathsf{LSD}(x, y)$ on ψ with information $\left(I^A_\tau, I^B_\tau\right) \leq \left(\gamma \log B, B^{1-c\gamma}\right)$.

Define $\mathcal{M} = \{m | \mu(\mathsf{LSD}(x, y) \neq \tau(x, y) | m) \leq 4\epsilon\}$, a subset of all transcripts of τ. Note that $\mu(\mathcal{M}) \geq 1 - \frac{1}{4} = \frac{3}{4}$, using the fact that $\mu(\mathsf{LSD}(x, y) \neq \tau(x, y)) \leq \epsilon$. Therefore, $\psi(\mathcal{M}) \geq 1 - \frac{2}{4} = \frac{1}{2}$, since density of $\mathsf{Supp}(\psi)$ under μ is one half.

First observe that Corollary 37 holds when the transcripts are restricted to the set $\mathcal{M}$. Now, Corollary 38 shows the existence of constant c_1 and sets S, T such that $\psi(x \in S) \geq 2^{-c_1(I^A)} \geq$

$2^{-c_1(\gamma \log B)}$ and $\psi(y \in T | x \in S) \geq 2^{-c_1(I^B)} \geq 2^{-c_1(B^{1-4\gamma})}$ We used with upper bounds on information (I^A, I^B) under ψ.

Since restricted to $m \in \mathcal{M}$, $\mu(\mathsf{LSD}(x,y) \neq 1 | S \times T) \leq 3 \times 4\epsilon + \delta$, where the factor 3 is an outcome of an averaging argument(see Corollary 38) and δ corresponds to the ℓ_1 distance between the simulation in Corollary 37 and the actual protocol.

The marginal distribution in x being the same on μ and ψ imply,

$$\mu(x \in S) \geq 2^{-c_1(\gamma \log B)}$$

Also, $\mu(y) \geq \frac{1}{2} \cdot \psi(y)$ since $\psi(x,y) = \mu(x, y | x \cap y = \emptyset)$ and $\mu(x \cap y = \emptyset) = \frac{1}{2}$. Therefore,

$$\mu(y \in T) \geq \frac{1}{2} \cdot \psi(y \in T) \geq 2^{-c_1(\gamma \log B + B^{1-c\gamma})-1}.$$

From here on, we work only with the distribution μ. The bounds on probabilities imply the following bounds on the corresponding entropy,

$$\mathsf{H}(X|S) \geq (1 - c_1\gamma) \log B \tag{9}$$

$$\mathsf{H}(Y|T) \geq \frac{B}{2} - c_1 \gamma \log B - c_1 B^{1-c\gamma} - 1. \tag{10}$$

The error bound implies,

$$\forall x \in S, \mu(Y_x = 1 | T) \leq \mu(\pi(x,y) \neq 1 | S \times T) \leq 12\epsilon + \delta. \tag{11}$$

Note that Y_x is the projection of the vector Y(the indicator random variable for the subset in $\{0,1\}^B$) onto the coordinate indexed by x. This implies, $\forall x \in S$ $H(Y_x|T) \leq \mathsf{h}(12\epsilon + \delta)$.

Using subadditivity of entropy, we upper bound $\mathsf{H}(Y|T)$ by $\mathsf{H}(Y_S|T) + \mathsf{H}(Y_{\bar{S}}|T)$, where $\bar{S}$ is the complement of the set S. $Y_S, Y_{\bar{S}}$ are projections of vector Y onto coordinates indexed by elements of sets S and $\bar{S}$ The first term in the expression can be simplified(using subadditivity of entropy) as follows,

$$\mathsf{H}(Y_S|T) \leq \mathsf{h}\left(12\epsilon + \delta\right) \cdot |S|$$

where the inequality follows from subadditivity of entropy and (11).
The second term yields,

$$\mathsf{H}(Y_{\bar{S}}|T) \leq \frac{B}{2} - |S|,$$

by an application of subadditivity of entropy and upper bounding binary entropy by 1.
(9) implies $|S| \geq B^{1-c_1\gamma}$. Therefore

$$\mathsf{H}(Y|T) \leq \frac{B}{2} - \left(1 - [\mathsf{h}(12\epsilon + \delta)]\right) B^{1-c_1\gamma}.$$

Equation (10) implies,

$$\mathsf{H}(Y|T) \geq \frac{B}{2} - c_1\gamma \log B - c_1 B^{1-c\gamma} - 1$$

This yields a contradiction, as $c > c_1 + \frac{1}{\gamma \log B}\left(\log 2c_1 - \log\left(1 - \mathsf{h}(12\epsilon + \delta)\right)\right)$ and $B = \Omega(1)$ imply,

$$\frac{B}{2} - \left(1 - [\mathsf{h}(12\epsilon + \delta)]\right) B^{1-c_1\gamma} < \frac{B}{2} - c_1\gamma \log B - c_1 B^{1-c\gamma} - 1.$$

This concludes the proof of the theorem. □

Acknowledgements

We thank Mert Saglam and Makrand Sinha for useful discussions. We also thank the anonymous reviewer for pointing out an error in the previous version of Lemma 19.

References

[Abl93] F. Ablayev. Lower bounds for one-way probabilistic communication complexity. In A. Lingas, R. Karlsson, and S. Carlsson, editors, *Proceedings of the 20th International Colloquium on Automata, Languages, and Programming*, volume 700 of *LNCS*, pages 241–252. Springer-Verlag, 1993.

[AIP06] A. Andoni, P. Indyk, and M. Pătraşcu. On the optimality of the dimensionality reduction method. In *Proc. 47th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 449–458, 2006.

[BBCR10] B. Barak, M. Braverman, X. Chen, and A. Rao. How to compress interactive communication. In *Proceedings of the 2010 ACM International Symposium on Theory of Computing*, pages 67–76, 2010.

[BOR99] A. Borodin, R. Ostrovsky, and Y. Rabani. Lower bounds for high dimensional nearest neighbor search and related problems. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing (STOC'99)*, pages 312–321, New York, May 1999. Association for Computing Machinery.

[BR11] M. Braverman and A. Rao. Information equals amortized communication. In R. Ostrovsky, editor, *FOCS*, pages 748–757. IEEE, 2011.

[Bra11] M. Braverman. Interactive information complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:123, 2011.

[Bra12] M. Braverman. Interactive information complexity. In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, pages 505–524, New York, NY, USA, 2012. ACM.

[BW12] M. Braverman and O. Weinstein. A discrepancy lower bound for information complexity. In A. Gupta, K. Jansen, J. D. P. Rolim, and R. A. Servedio, editors, *APPROX-RANDOM*, volume 7408 of *Lecture Notes in Computer Science*, pages 459–470. Springer, 2012.

[BYCKO93] R. Bar-Yehuda, B. Chor, E. Kushilevitz, and A. Orlitsky. Privacy, additional information and communication. *IEEE Transactions on Information Theory*, 39(6):1930–1943, 1993. Preliminary version in CCC '90.

[CSWY01] A. Chakrabarti, Y. Shi, A. Wirth, and A. Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 270–278, 2001.

[GKR14] A. Ganor, G. Kol, and R. Raz. Exponential separation of information and communication for boolean functions. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:113, 2014.

[JKKR04] T. S. Jayram, S. Khot, R. Kumar, and Y. Rabani. Cell-probe lower bounds for the partial match problem. *J. Comput. Syst. Sci*, 69(3):435–447, 2004.

[KLL$^+$12] I. Kerenidis, S. Laplante, V. Lerays, J. Roland, and D. Xiao. Lower bounds on information complexity via zero-communication protocols and applications. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:38, 2012.

[KN97] E. Kushilevitz and N. Nisan. *Communication complexity*. Cambridge University Press, Cambridge, 1997.

[Mil94] P. B. Miltersen. Lower bounds for union-split-find related problems on random access machines. In *Proceedings of the 26th Annual Symposium on the Theory of Computing*, pages 625–634, New York, May 1994. ACM Press.

[Mil99] P. B. Miltersen. Cell probe complexity — a survey. In V. Raman, C. P. Rangan, and R. Ramanujam, editors, *Foundations of Software Technology and Theoretical Computer Science: 19th Conference, Chennai, India, December 13–15, 1999: Proceedings*, volume 1738 of *Lecture Notes in Computer Science*, pub-SV:adr, 1999. Springer-Verlag.

[MNSW98] P. Miltersen, N. Nisan, S. Safra, and A. Wigderson. On data structures and asymmetric communication complexity. *Journal of Computer and System Sciences*, 57:37–49, 1 1998.

[Păt11] M. Pătraşcu. Unifying the landscape of cell-probe lower bounds. *SIAM Journal on Computing*, 40(3):827–847, 2011. See also FOCS'08, arXiv:1010.3783.

[PT] M. Pătraşcu and M. Thorup. Higher lower bounds for near-neighbor and further rich problems. *SIAM Journal on Computing*, 39(2):730–741. See also FOCS'06.

[Sag14] M. Saglam. Private communication. 2014.

[SS02] M. E. Saks and X. Sun. Space lower bounds for distance approximation in the data stream model. In *STOC*, pages 360–369. ACM, 2002.