Electronic Colloquium on Computational Complexity, Report No. 177 (2012)

Information lower bounds via self-reducibility

Mark Braverman * Ankit Garg † Denis Pankratov ‡ Omri Weinstein §

December 19, 2012

Abstract

We use self-reduction methods to prove strong information lower bounds on two of the most studied functions in the communication complexity literature: Gap Hamming Distance (GHD) and Inner Product (IP). In our first result we affirm the conjecture that the information cost of GHD is linear even under the *uniform* distribution, which strengthens the $\Omega(n)$ bound recently shown by [15], and answering an open problem from [10]. In our second result we prove that the information cost of IP_n is arbitrarily close to the trivial upper bound n as the permitted error tends to zero, again strengthening the $\Omega(n)$ lower bound recently proved by [9].

Our proofs demonstrate that self-reducibility makes the connection between information complexity and communication complexity lower bounds a two-way connection. Whereas numerous results in the past [12, 3, 4] used information complexity techniques to derive new communication complexity lower bounds, we explore a generic way in which communication complexity lower bounds imply information complexity lower bounds *in a black-box manner*.

1 Introduction

The primary objective of this paper is to continue the investigation of the information complexity vs. communication complexity problem. Informally, in a two-party setting, communication complexity (CC) measures the number of bits two parties need to exchange to solve a certain problem. Information complexity (IC) measures the average amount of information the parties need to reveal each other about their inputs in order to solve it. IC is always bounded by CC from above. A key open problem surrounding information complexity is actually understanding the gap between the two:

Problem 1.1. *Is it true that for all functions f it holds that $IC(f) = \Omega(CC(f))$?*

The problem, and where it fits more broadly within communication complexity is discussed in [5]. The above question is a natural question in the context of coding theory, where it can be re-interpreted as asking whether an analogue of Huffman coding holds for interactive computation. Shannon's original insight [19] was that the (amortized) number of bits one needs to send in order to transmit a message X equals to the amount of information it coveys – its entropy $H(X)$. Huffman coding [14] can be viewed as a one-copy version of this result: even when sending one instance of the message, we can guarantee expected cost of $\leq H(X) + 1$ – i.e. messages can be compressed into their information content plus at most one bit. Problem 1.1 can be viewed as a quest for an interactive analogue of Huffman coding: can a long (interactive) communication protocol that solves f but only conveys $IC(f)$ information be compressed in a way that only requires $O(IC(f))$ communication?

*Department of Computer Science, Princeton University. Research supported in part by an Alfred P. Sloan Fellowship, an NSF CAREER award, and a Turing Centenary Fellowship.
†Department of Computer Science, Princeton University
‡Department of Computer Science, University of Chicago
§Department of Computer Science, Princeton University

ISSN 1433-8092

Another direction which motivates Problem 1.1 are *direct sum* problems in randomized communication complexity [12, 3, 4, 8]. It turns out that the analogue of the Shannon's amortized coding theorem does in fact hold for interactive computation [8], asserting that $\lim_{k\to\infty} CC(f^k)/k = IC(f)$. Thus, understanding the relationship between $IC(f)$ and $CC(f)$ is equivalent to understanding the relationship between computing one copy of f and the amortized cost of computing many copies of f in parallel, which is the essence of the direct sum problem.

Yet another motivation for considering the information complexity of tasks comes from the study private two-party computation [16, 18, 1]. In this setting Alice and Bob want to compute a function $f(x, y)$ on their private inputs x and y respectively without "leaking" too much information to each other. This can be accomplished using cryptography, assuming Alice and Bob are computationally bounded. Without this assumption, the amount of information that Alice and Bob must reveal to each other is exactly $IC(f)$. In this context, an affirmative answer for Problem 1.1 would mean that, up to a constant, a protocol minimizing communication (with no special consideration for privacy) will reveal the same amount of information as the most "private" protocol. Moreover, if for a family $\{f_n\}$ of functions we have that $IC(f_n)/CC(f_n) \to 1$ as $n \to \infty$, it means that as n grows, there is *nothing* the parties can do to perform the computation more privately, and the most efficient protocol is also the most private. In this paper we show, for example, that this is the case for the Inner Product function IP_n, whose communication complexity is n, and whose information complexity we show to be $n - o(n)$ (for negligible error).

In this paper we develop a new self-reducibility technique for deriving information complexity lower bounds from communication complexity lower bounds. The technique works for functions that have a "self-reducible structure". Informally speaking f has a self-reducible structure, if for large enough inputs, solving f_{nk} essentially amounts to solving f_n^k (f_{nk} denotes the function f under inputs of length nk, while f_n^k denotes k independent copies of f under inputs of size n). Our departing point is a communication complexity lower bound for f_{nk} (that may be obtained by any means). Assuming self-reducibility, the same bound applies to f_n^k, which through the connection between information complexity and amortized communication complexity [8], implies a lower bound on the information complexity of f_n. In this paper we develop tools to make this reasoning go through.

Ideas of self-reducibility are central in applications of information complexity to communication complexity lower bounds, starting with the work of Bar-Yossef et al. [3]. These argument start with an information complexity lower bound for a (usually very simple) problem, and derive a communication complexity bound on many copies of the problem. The logic of this paper is reversed: we start with a communication complexity lower bound, which we use as a black-box, and use self-reducibility to derive an amortized communication complexity bound, which translates into an information complexity lower bound. An additional conceptual take-away from the present paper, is that to look for a counterexample for Problem 1.1, one would likely need to consider problems that are highly non-self-reducible.

1.1 Results

We use the self-reducibility technique to prove results about the information complexity of Gap Hamming Distance and Inner Product. We prove that the information complexity of the Gap Hamming Distance problem with respect to the uniform distribution is linear. This was explicitly stated as an open problem by Chakrabarti et al. [10]. Formally, let $\mathsf{IC}_\mu(GHD_{n,t,g}, \varepsilon)$ denote the information cost of the Gap Hamming promise problem, where inputs x, y are n-bit strings distributed according to μ, and the players need to determine whether the Hamming distance between x and y is at least $t + g$, or at most $t - g$, with error at most ε under μ. We prove

Theorem 1.2. *There exists an absolute constant $\varepsilon > 0$ for which*

$$\mathsf{IC}_{\mathcal{U}}(GHD_{n,n/2,\sqrt{n}}, \varepsilon) = \Omega(n)$$

where $\mathcal{U}$ is the uniform distribution.

For the Inner Product, we prove a stronger bound on its information complexity. Formally

Theorem 1.3. *For every constant $\delta > 0$, there exists a constant $\epsilon > 0$, and n_0 such that $\forall\ n \geq n_0$, $IC_{\mathcal{U}_n}(IP_n, \epsilon) \geq (1-\delta)n$. Here $\mathcal{U}_n$ is the uniform distribution over $\{0,1\}^n \times \{0,1\}^n$.*

Note that $IC_{\mathcal{U}_n}(IP_n, \epsilon) \leq (1-2\epsilon)(n+1)$, since the parties can always give a random output with probability ϵ, and exchange their inputs with probability $1-\epsilon$. Also it is known that $IC_{\mathcal{U}_n}(IP_n, \epsilon) \geq \Omega(n)$, for all $\epsilon \in [0, 1/2)$ [9]. We prove that the information complexity of IP_n can be arbitrarily close to the trivial upper bound n as we keep decreasing the error (though keeping it a constant).

1.2 Discussion and open problems

Although in Complexity Theory we often don't care about the constants (and often it is not necessary), proving theorems with the right constants can often lead to deeper insights into the mathematical structure of the problem [6, 7]. There are few techniques that allow us to find the right constants and there are fewer problems for which we can. We believe that answering the following problem will lead to development of new techniques and also reveal interesting insights into the problem of computing the XOR of n copies of a function.

Open Problem 1.4. *Is it true that for small constants ϵ and sufficiently large n, $IC_{\mathcal{U}_n}(IP_n, \epsilon) \geq (1-2\epsilon-o(\epsilon))n$? As before $\mathcal{U}_n$ is the uniform distribution. If this is false, is there a different constant $\alpha > 2$ such that as $\epsilon \to 0$ we get $IC_{\mathcal{U}_n}(IP_n, \epsilon) \geq (1-\alpha \cdot \epsilon)n$?*

Solving this problem may require shedding new light on the rate of convergence of the $IC_\mu(\bullet, \epsilon)$ to $IC_\mu(\bullet, 0)$ as $\epsilon \to 0$, and better understanding the role error plays in information complexity.

It is somewhat difficult to define the exact meaning of the "right" constant for the Gap Hamming Distance problem, since it is a promise problem defined by two parameters (gap and error). Nonetheless, there is a very natural regime in which understanding the exact information complexity of GHD_n is a natural and interesting problem. Namely:

Open Problem 1.5. *Is it true that for all $\varepsilon > 0$, there is a $\delta > 0$ and a distribution μ such that $\mathsf{IC}_\mu(GHD_{n,n/2,\delta\sqrt{n}}, \delta) > (1-\varepsilon)n$?*

In other words, does the information complexity of GHD_n tends to the trivial upper bound as we tighten the gap and error parameters? This is related to the same (but weaker) question one can ask about the communication complexity of GHD_n in this regime.

2 Preliminaries

In this section we briefly survey the necessary background for this paper on information theory and communication complexity. For a more thorough treatment of these subjects see [8] and references therein.

Notation. We use capital letters for random variables, calligraphic letters for sets, and small letters for elements of sets. For random variables A and B and an element b we write A_b to denote the random variable A conditioned on the event $B = b$.

2.1 Information Theory

Definition 2.1. The *entropy* of a random variable X, denoted by $H(X)$, is defined as $H(X) = \sum_x \Pr[X = x] \log(1/\Pr[X = x])$. The *conditional entropy* of X given Y, denoted by $H(X|Y)$, is $\mathbf{E}_{\mathrm{y}}[\mathrm{H(X|Y=y)}]$.

Definition 2.2. The *mutual information* between two random variables A, B, denoted $I(A;B)$, is defined to be the quantity $H(A) - H(A|B) = H(B) - H(B|A)$. The *conditional mutual information* $I(A;B|C)$ is $H(A|C) - H(A|BC)$.

Fact 2.3 (Chain Rule)**.** *Let A_1, A_2, B, C be random variables. Then $I(A_1A_2;B|C) = I(A_1;B|C) + I(A_2;B|A_1C)$.*

Definition 2.4. *Kullback-Leibler Divergence* between probability distributions A and B is defined as $\mathbb{D}(A||B) = \sum_x A(x) \log \frac{A(x)}{B(x)}$.

Fact 2.5. *For random variables $A, B,$ and C we have $I(A;B|C) = \mathbb{E}_{b,c}(\mathbb{D}(A_{bc}||A_c))$.*

Fact 2.6. *Let X and Y be random variables. Then for any random variable Z we have $\mathbb{E}_x[\mathbb{D}(Y_x||Y)] \leq \mathbb{E}_x[\mathbb{D}(Y_x||Z)]$.*

Fact 2.7. *Let A, B, C, D be four random variables such that $I(B; D|AC) = 0$. Then $I(A; B|C) \geq I(A; B|CD)$*

Fact 2.8. *Let A, B, C, D be four random variables such that $I(A; C|BD) = 0$. Then $I(A; B|D) \geq I(A; C|D)$*

Definition 2.9. The *statistical distance (total variation)* between random variables D and F taking values in a set $\mathcal{S}$ is defined as $|D - F| \stackrel{\text{def}}{=} \max_{\mathcal{T} \subseteq \mathcal{S}}(|\Pr[D \in \mathcal{T}] - \Pr[F \in \mathcal{T}]|) = \frac{1}{2} \sum_{s \in \mathcal{S}} |\Pr[D = s] - \Pr[F = s]|$.

2.2 Communication Complexity

We use standard definitions of the two-party communication model that was introduced by Yao in [20]:

Definition 2.10. The *distributional communication complexity* of $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ with respect to a distribution μ on $\mathcal{X} \times \mathcal{Y}$ and *error tolerance* $\epsilon > 0$ is the least cost of a deterministic protocol computing f with error probability at most ϵ when the inputs are sampled according to μ. It is denoted by $\mathcal{D}_\mu(f, \epsilon)$.

Definition 2.11. The *randomized communication complexity* of $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ with error tolerance $\epsilon > 0$, denoted by $R_\epsilon(f)$, is the least cost of a public-coin protocol computing f with error at most ϵ on *every* input.

For a thorough treatment of pre-1997 results in communication complexity see an excellent monograph by Kushilevitz and Nisan [17].

2.3 Information + Communication: The Information Cost

We consider protocols with *both private and public randomness*. Let $\Pi(X, Y)$ (random variable) denote *the transcript*, i. e., the concatenation of the public randomness with all the messages sent during the execution of π on (X, Y). When $X = x, Y = y$, we write $\Pi(x, y)$. When (X, Y) or (x, y) are clear from the context, we shall omit them and simply write Π for the transcript.

The notion of *internal information cost* was implicit in [3] and was explicitly defined in [4] as follows:

Definition 2.12. The *internal information cost* of a protocol π over inputs drawn from a distribution μ on $\mathcal{X} \times \mathcal{Y}$, is given by:

$$\mathsf{IC}_\mu(\pi) := I(\Pi; X|Y) + I(\Pi; Y|X).$$

Intuitively, the information cost captures the amount of information the two parties learn about each others' inputs during communication. Note that the information cost of a protocol π depends on the prior distribution μ. Naturally, the information cost of a protocol over *any* distribution is a lower bound on the communication cost.

Lemma 2.13. *[8] For any distribution μ we have $\mathsf{IC}_\mu(\pi) \leq CC(\pi)$.*

Definition 2.14. The *information complexity* of f with respect to distribution μ and error tolerance $\epsilon \geq 0$ is defined as

$$\mathrm{IC}_\mu(f, \epsilon) = \inf_\pi \mathrm{IC}_\mu(\pi),$$

where the infimum ranges over all randomized protocols π solving f with error at most ϵ when inputs are sampled according to μ.

3 Information complexity of Gap Hamming Distance

Given two strings $x, y \in \{0,1\}^n$, the hamming distance x and y is defined to be $HAM(x,y) = |\{i \,|\, x_i \neq y_i\}|$. In the Gap Hamming Distance (GHD) problem, Alice gets a string $x \in \{0,1\}^n$ and Bob gets a string $y \in \{0,1\}^n$. They are promised that either $HAM(x,y) \geq n/2 + \sqrt{n}$ or $HAM(x,y) \leq n/2 - \sqrt{n}$, and they have to find which is the case. We can define a general version $GHD_{n,t,g}$, where Alice and Bob have to determine if $HAM(x,y) \geq t+g$ or $HAM(x,y) \leq t-g$, but the parameters $t = n/2$ and $g = \sqrt{n}$ are the most natural as discussed in [11]. In a technical tour-de-force, it was proved in [11] that the randomized communication complexity of the Gap Hamming Distance problem is linear. Formally,

Theorem 3.1. *For all constants $\gamma > 0$, and $\epsilon \in [0, 1/2)$, $R_\epsilon(GHD_{n,n/2,\gamma\sqrt{n}}) \geq \Omega(n)$.*

One can extend the formulation of GHD beyond the promise-problem setting. This particularly makes sense in a distributional-complexity setting. In this setting, we allow f to take the value $\star$, which means that we don't care about the output. The error in this model is aggregated only over points on which the value of f is not $\star$. Chakrabarti and Regev [11] also prove a distributional version of the linear lower bound over the *uniform* distribution $\mathcal{U}$. Specifically, they prove

Theorem 3.2. *[11] There exists an absolute constant $\varepsilon > 0$ for which*

$$\mathcal{D}_{\mathcal{U}}(GHD_{n,n/2,\sqrt{n}}, \varepsilon) = \Omega(n).$$

Kerenidis et al. [15] proved that the information complexity of Gap Hamming Distance is also linear, at least with respect to some distribution. The proof of Kerenidis et al. relies on a reduction that shows that a large class of communication complexity lower bound techniques also translate into information complexity lower bounds – including the lower bound for GHD:

Theorem 3.3. *[15] There exists a distribution μ on $\{0,1\}^n \times \{0,1\}^n$ and an absolute constant $\varepsilon > 0$ such that*

$$\mathsf{IC}_\mu(GHD_{n,n/2,\sqrt{n}}, \varepsilon) = \Omega(n).$$

Interestingly, while this approach yields an analogue of Theorem 3.1 for information complexity, it does not seem to yield an analogue of the stronger Theorem 3.2.

We give an alternate proof of the linear information complexity lower bound for GHD using the self reducibility technique. Unlike the proof in [15] we do not need to dive into the details of the proof of the communication complexity lower bound for GHD. Rather, our starting point is Theorem 3.2, which we use as a black-box.

In fact, we will prove a slightly weaker lemma, with Theorem 1.2 following by a reduction. The reduction is conceptually very simple, but the details are somewhat tedious.

Lemma 3.4. *There exists absolute constants $\varepsilon > 0$ and $\gamma > 0$ for which*

$$\mathsf{IC}_{\mathcal{U}}(GHD_{n,n/2,\gamma\sqrt{n}}, \varepsilon) = \Omega(n).$$

4 Proof of Theorem 1.2

4.1 Proof Idea

We use the self-reducibility argument. Assume that for some $\epsilon > 0$, $IC_{\mathcal{U}}(GHD_n, \epsilon) = o(n)$. Then using information = amortized communication, we can get a protocol τ that solves N copies of GHD_n with $o(nN)$ communication. The heart of the argument is to use this to solve GHD_{nN} with $o(nN)$ *communication*, which is a contradiction. Say that Alice and Bob are given $x, y \in \{0,1\}^{nN}$ respectively. They sample $c \cdot nN$ random coordinates (for some constant c) and then divide these into cN blocks and run GHD_n on them all in parallel using $o(nN)$ communication. If $HAM(x,y) = nN/2 + \sqrt{nN}$, then the expected hamming distance of each block is $n/2 + \sqrt{n/N}$. Although the gain over $n/2$ is small, the hamming distance is still biased towards being $> n/2$. We will see that on each instance the protocol for GHD_n must gain an advantage of $\Omega(1/\sqrt{N})$ over random guessing. This in turn implies that cN copies suffice to get the correct answer with high probability.

4.2 Formal Proof of Lemma 3.4

Assume that for some ρ sufficiently small (to be specified later), $\mathsf{IC}_{\mathcal{U}}(GHD_{n,n/2,\sqrt{n}}, \rho) = o(n)$. Thus $\forall \alpha > 0$, for n sufficiently large $\mathsf{IC}_{\mathcal{U}}(GHD_{n,n/2,\sqrt{n}}, \rho) \leq \alpha n$. We will need the following theorem from [8]:

Theorem 4.1. *[8] Let $f : X \times Y \to \{0, 1\}$ be a (possibly partial) function, let μ be any distribution on $X \times Y$, and let $I = \mathsf{IC}_\mu(f, \rho)$, then for each $\delta_1, \delta_2 > 0$, there is an $N = N(f, \rho, \mu, \delta_1, \delta_2)$ such that for each $n \geq N$, there is a protocol π_n for computing n instances of f over μ^n such that error on each copy is $\leq \rho$. The protocol has communication complexity $< nI(1 + \delta_1)$. Moreover, if we let π be any protocol for computing f with information cost $\leq I(1 + \delta_1/3)$ w.r.t. μ, then we can design π_n so that for each set of inputs, the statistical distance between the output of π_n and π^n is $< \delta_2$, where π^n denotes n independent executions of π.*

In other words, Theorem 4.1 allows us to take a low-information protocol for f and turn it into a low-communication protocol for (sufficiently) many copies of f.

Step 1: From GHD to a tiny advantage.

In the first step we show that a protocol for GHD over the uniform distribution has a small but detectable advantage in distinguishing inputs from two distributions that are very close to each other. Denote by μ_η the distribution where $X \in \{0, 1\}^n$ is chosen uniformly, and Y is chosen so that $X_i \oplus Y_i \sim B_{1/2+\eta}$ is an i.i.d. Bernoulli random variable with bias η. Note that in this language the GHD problem is essentially about distinguishing $\mu_{-1/\sqrt{n}}$ from $\mu_{1/\sqrt{n}}$.

Lemma 4.2. *There exists absolute constants $\tau > 0$, $\gamma > 0$ and $\rho > 0$ with the following property. Suppose that for all n large enough there is a protocol π_n such that π_n solves $GHD_{n,n/2,\gamma\sqrt{n}}$ with error ρ w.r.t the uniform distribution. Then for all n large enough for all $\varepsilon < 1/n^2$ we have*

$$Pr_{(x,y)\sim\mu_\varepsilon}[\pi_n(x, y) = 1] - Pr_{(x,y)\sim\mu_0}[\pi_n(x, y) = 1] > \tau \cdot \varepsilon \cdot \sqrt{n}, \tag{1}$$

and

$$Pr_{(x,y)\sim\mu_{-\varepsilon}}[\pi_n(x, y) = 0] - Pr_{(x,y)\sim\mu_0}[\pi_n(x, y) = 0] > \tau \cdot \varepsilon \cdot \sqrt{n}, \tag{2}$$

Proof. Note that we can assume that the protocol π_n is symmetric w.r.t the hamming distance, i.e. its behavior depends just on the hamming distance between x and y. This is because Alice and Bob can start with applying a random permutation and a random XOR on their inputs i.e. they sample (using public randomness) a permutation $\pi \in S_n$ and $r \in \{0, 1\}^n$ and change their inputs to $\pi(x \oplus r)$ and $\pi(y \oplus r)$. Note that the information cost of the protocol remains the same.

We will establish (1), with (2) established identically. We first focus on the region where $HAM(x, y) \geq n/2$ and show that its contribution to (1) is at least $\Omega(\varepsilon\sqrt{n})$. We break the region into two further regions: (I) (x, y) with $n/2 < H(x, y) < n/2 + \gamma\sqrt{n}$; (II) (x, y) with $n/2 + \gamma\sqrt{n} \leq H(x, y)$ for appropriately chosen γ. We show that the contribution of region (II) is $\Omega(\varepsilon\sqrt{n})$, while the fact that the contribution of regions (I) being positive is easy to see.

Denote by p_i the probability that π_n returns 1 on an input of hamming distance $n/2+i$. The contribution of the region where $H(x, y) = n/2 + i$ is equal to

$$p_i \cdot (Pr_{\mu_\varepsilon}[H(x, y) = n/2 + i] - Pr_{\mu_0}[H(x, y) = n/2 + i]) =$$
$$p_i \cdot Pr_{\mu_0}[H(x, y) = n/2 + i] \cdot \left((1 - 4\varepsilon^2)^{n/2-i}(1 + 2\varepsilon)^{2i} - 1\right).$$

Now $(1-4\varepsilon^2)^{n/2-i} \geq 1-2\varepsilon/n$ and $(1+2\varepsilon)^{2i} \leq e^2$ ($\varepsilon < 1/n^2$). Thus $\sum_{i=0}^{n/2} p_i \cdot Pr_{\mu_0}[H(x,y) = n/2+i] \cdot (2\varepsilon/n) \cdot (1+2\varepsilon)^{2i} \leq O(\varepsilon/n)$ and therefore we will ignore the term $(1-4\varepsilon^2)^{n/2-i}$. After ignoring the term, the contribution in region (I) is positive .

This leaves us with region (II), where we need to show that we actually get a non-negligible advantage. Let T be an appropriately chosen constant. The advantage

$$\sum_{i=\gamma\sqrt{n}}^{n/2} p_i \cdot Pr_{\mu_0}[H(x,y) = n/2+i] \cdot \big((1+2\varepsilon)^{2i} - 1\big) \geq$$

$$\sum_{i=\gamma\sqrt{n}}^{T\sqrt{n}} p_i \cdot Pr_{\mu_0}[H(x,y) = n/2+i] \cdot 4i\varepsilon = \sum_{i=\gamma\sqrt{n}}^{T\sqrt{n}} Pr_{\mu_0}[H(x,y) = n/2+i] \cdot 4i\varepsilon$$

$$- \sum_{i=\gamma\sqrt{n}}^{T\sqrt{n}} (1-p_i) \cdot Pr_{\mu_0}[H(x,y) = n/2+i] \cdot 4i\varepsilon \geq \Theta(\varepsilon\sqrt{n}) - 4T\rho\varepsilon\sqrt{n}$$

since $(1-p_i)$ is the probability that protocol errs when hamming distance is $n/2+i$ and average error $\leq \rho$. By making ρ small enough we can get noticeable advantage $(\Theta(\varepsilon\sqrt{n}))$ in this region.

We now consider the region $HAM(x,y) \leq n/2$ and show that the absolute value of the contribution of this region can be made arbitrarily small w.r.t. $\varepsilon\sqrt{n}$ by appropriate choices of ρ, γ and T which will complete the proof. Let us break this region into three further regions : (I) (x,y) with $n/2 - \gamma\sqrt{n} < HAM(x,y) \leq n/2$; (II) (x,y) with $n/2 - T\sqrt{n} \leq HAM(x,y) < n/2 - \gamma\sqrt{n}$; (III) (x,y) with $HAM(x,y) < n/2 - T\sqrt{n}$ for appropriately chosen T and γ. Denote by q_i the probability that π_n returns 1 on an input of hamming distance $n/2 - i$. The absolute value of the contribution of the region where $HAM(x,y) = n/2 - i$ is equal to

$$q_i \cdot (Pr_{\mu_0}[HAM(x,y) = n/2-i] - Pr_{\mu_{-\varepsilon}}[HAM(x,y) = n/2-i]) =$$
$$q_i \cdot Pr_{\mu_0}[HAM(x,y) = n/2-i] \cdot (1 - (1-4\varepsilon^2)^{n/2-i}(1-2\varepsilon)^{2i})$$

As before, we can ignore the term $(1-4\varepsilon^2)^{n/2-i}$. In region (I) the negative contribution is bounded in absolute terms by:

$$1 - (1-2\varepsilon)^{2\gamma\sqrt{n}} < 4\gamma\varepsilon\sqrt{n}.$$

In region (III) the contribution is again bounded by

$$\sum_{i=T\sqrt{n}}^{n/2} Pr_{\mu_0}[HAM(x,y) = n/2-i] \cdot (1-(1-2\varepsilon)^{2i}) < \sum_{i=T\sqrt{n}}^{n/2} Pr_{\mu_0}[HAM(x,y) = n/2-i] \cdot 4i\varepsilon.$$

By a standard Chernoff bound[1], the probability $Pr_{\mu_0}[HAM(x,y) = n/2-i]$ is dominated by $e^{-\Omega(i^2/n)}$, and thus the sum can be made into an arbitrarily small multiple of $\varepsilon\sqrt{n}$ by choosing T large enough. For region (II) the advantage

$$\sum_{i=\gamma\sqrt{n}}^{T\sqrt{n}} q_i \cdot Pr_{\mu_0}[HAM(x,y) = n/2-i] \cdot (1-(1-2\varepsilon)^{2i}) \leq$$

$$\sum_{i=\gamma\sqrt{n}}^{T\sqrt{n}} q_i \cdot Pr_{\mu_0}[HAM(x,y) = n/2-i] \cdot 4i\varepsilon \leq 4T\varepsilon\sqrt{n} \sum_{i=\gamma\sqrt{n}}^{T\sqrt{n}} q_i \cdot Pr_{\mu_0}[HAM(x,y) = n/2-i]$$
$$\leq 4T\rho\varepsilon\sqrt{n}.$$

[1] See e,g [2].

By making ρ small enough we can make the absolute contribution of this region small relative to $\varepsilon\sqrt{n}$. This completes the proof. $\square$

Step 2: From tiny advantage to low-communication GHD.

We can now apply Lemma 4.2 together with Theorem 4.1 to show that a low-information solution to $GHD_{n,n/2,\gamma\sqrt{n}}$ with respect to the uniform distribution contradicts the communication complexity lower bound of Theorem 3.2.

Proof. (of Lemma 3.4). Assume for the sake of contradiction that for each α there is an n and a protocol π_n with $\mathsf{IC}_{\mathcal{U}}(\pi_n) < \alpha n$ and which solves $GHD_{n,n/2,\gamma\sqrt{n}}$ with error ρ, where the parameters γ and ρ are from Lemma 4.2. Let $N > \max(n^7, N(GHD_{n,n/2,\gamma\sqrt{n}}, \rho, \mathcal{U}, \delta_1, \delta_2))$, where $\delta_1 = 1$ and $\delta_2 = \varepsilon/2$, where ε is the error parameter in Theorem 3.2. Then using Theorem 4.1, for each $c > 1$, cN copies of π_n can be executed with communication $< 2\alpha cn \cdot N$ (as long as the inputs to each π_n are distributed according to $\mathcal{U}$) such that on each copy the error is at most ρ w.r.t $\mathcal{U}$. Also for each set of inputs, the statistical distance between the output of the execution and $\pi_n^{cN} \leq \varepsilon/2$.

Let $t = Pr_{(x,y)\sim\mathcal{U}}[\pi_n(x,y) = 1]$. W.l.o.g. we assume $t = 1/2$. We solve $GHD_{n\cdot N,n\cdot N/2,\sqrt{n\cdot N}}$ over the uniform distribution with a small constant error ε using the protocol depicted in Figure 1.

Input: A pair $x, y \in \{0,1\}^{nN}$.
Output: $GHD_{n\cdot N,n\cdot N/2,\sqrt{n\cdot N}}$.

1. Create cN instances of GHD_n by sampling n random coordinates each time (with replacement): $(x_1, y_1), \ldots, (x_{cN}, y_{cN}) \in \{0,1\}^n \times \{0,1\}^n$.
2. Use compression (Theorem 4.1) to run $\pi_n(x_1, y_1), \ldots, \pi_n(x_{cN}, y_{cN})$ in communication $2\alpha cNn$.
3. Return $MAJORITY(\pi_n(x_1, y_1), \ldots, \pi_n(x_{cN}, y_{cN}))$.

Protocol 1: The protocol $\Pi_{nN}(x, y)$

The communication cost upper bound follows from the way the protocol $\Pi_{nN}(x, y)$ is constructed. To finish the proof we need to analyze its success probability. Suppose that the hamming distance between x and y is $nN/2 + \ell\sqrt{nN}$, where $\ell > 1$. Note that $\ell < n$ except with probability $e^{-\Omega(n^2)}$. The samples (x_i, y_i) are drawn iid according to the distribution $\mu_{\ell\cdot\sqrt{1/(nN)}}$. Since $N > n^7$ we have $\ell \cdot \sqrt{1/nN} < 1/n^2$. By Lemma 4.2, the output of π_n on each copy is thus $\tau \cdot \ell/\sqrt{N}$-biased towards 1. An application of the Chernoff bounds along with the fact that, for each set of inputs, the statistical distance between the output of the execution and $\pi_n^{cN} \leq \varepsilon/2$, implies that the probability that the protocol Π_{nN} outputs 1 is at least $1 - e^{-2\tau^2\ell^2 c} - \varepsilon/2$. For constant τ, we can make this expression as close to $1 - \varepsilon/2$ as we like by letting c be a sufficiently large constant. But this means that for an arbitrarily small constant $\alpha > 0$, $\Pi_{nN}(x, y)$ will solve $GHD_{n\cdot N,n\cdot N/2,\sqrt{n\cdot N}}$ with error $\leq \varepsilon$ (the case when the hamming distance between x and y is $nN/2 - \ell\sqrt{nN}$ is symmetric) in communication $O(\alpha cNn)$, which can be made arbitrarily small relatively to Nn, leading to a contradiction. Note that we got a randomized protocol for solving $GHD_{n\cdot N,n\cdot N/2,\sqrt{n\cdot N}}$ but we can fix the randomness to get a deterministic algorithm. $\square$

4.3 The reduction from a small-gap instance to a large-gap instance

Now we complete the proof of Theorem 1.2 by providing the details of the reduction. We will start by proving a few technical lemmas.

Lemma 4.3. *Let $\alpha > 1$ be an integer. Let $\mathcal{U}_n$ be the uniform distribution over $\{0,1\}^n \times \{0,1\}^n$. Let $X, Y \sim \mathcal{U}_n$. Define a distribution μ over $\{0,1\}^{\alpha n} \times \{0,1\}^{\alpha n}$ by picking αn random coordinates of X, Y and then taking a XOR with a random string $r \in_R \{0,1\}^{\alpha n}$ (let U', V' be the strings obtained by sampling αn*

random coordinates of X, Y. Then $U = U' \oplus r, V = V' \oplus r$ are the final strings sampled). Then for all $\epsilon > 0$ and n large enough, there exists a constant M_ϵ and a distribution μ_ϵ such that

1. $|\mu - \mu_\epsilon| \le \epsilon$

2. $\mu_\epsilon \le M_\epsilon \cdot \mathcal{U}_{\alpha n}$

Proof. It is easy to see that the distribution μ is symmetric w.r.t the hamming distance i.e. if $x, y \in \{0,1\}^{\alpha n} \times \{0,1\}^{\alpha n}$, and $x', y' \in \{0,1\}^{\alpha n} \times \{0,1\}^{\alpha n}$ such that $HAM(x,y) = HAM(x',y')$, then $\mu(x,y) = \mu(x',y')$. This is because μ is invariant under the application of a random permutation and a random XOR i.e. if $\pi \in_R S_n$ and $r' \in_R \{0,1\}^n$, then $\mu(x,y) = \mu(\pi(x \oplus r'), \pi(y \oplus r'))$. With a slight abuse of notation let $\mu(d)$ denote the probability mass on strings of hamming distance d, and let $\mathcal{U}_{\alpha n}(d)$ denote the probability mass w.r.t the uniform distribution. Let $N = \alpha n$.

The distributions μ_ϵ will be the truncations of the distribution μ to the interval $[N/2 - C\sqrt{N}, N/2 + C\sqrt{N}]$. Let $Pr[t]$ denote the probability that the strings X, Y have hamming distance t. Then by Chernoff bounds $Pr[t \notin [n/2 - \beta\sqrt{n}, n/2 + \beta\sqrt{n}]] \le 2e^{-2\beta^2}$. Now if we pick N random coordinates distributed according to $B_{\frac{1}{2}+p}$, where $|p| \le \beta/\sqrt{n}$, then the expected number of 1's $\in [N/2 - \beta\sqrt{\alpha}\sqrt{N}, N/2 + \beta\sqrt{\alpha}\sqrt{N}]$. Thus by another application of Chernoff bound and C large enough, we can make the statistical distance between μ_ϵ and μ small enough.

Now let μ' be the distribution μ restricted to the interval $[N/2 - C\sqrt{N}, N/2 + C\sqrt{N}]$, for some constant C, with a slight scaling, which we can ignore. We will show that there exists a constant M such that $\mu' \le M \cdot \mathcal{U}_{\alpha n}$. Note that by the symmetry properties of μ, it suffices to prove that for all d, $\mu'(d) \le M \cdot \mathcal{U}_{\alpha n}(d)$. Now

$$\mu'(d)/\mathcal{U}_{\alpha n}(d) = \sum_{k=0}^{n} \binom{n}{k} \cdot 2^{-n} \cdot \left(\frac{2k}{n}\right)^d \cdot \left(\frac{2(n-k)}{n}\right)^{N-d}$$

Let $d = N/2 + T$, where $|T| \le C\sqrt{N}$. Also we will just concentrate on the sum for $k \ge n/2$. The lower half is analogous. Also it is easy to see that the sum from $k = 3n/4$ to $k = n$ is small. So we consider

$$\begin{aligned}
&\sum_{k=n/2}^{3n/4} \binom{n}{k} \cdot 2^{-n} \cdot \left(\frac{2k}{n}\right)^d \cdot \left(\frac{2(n-k)}{n}\right)^{N-d} \\
&= \sum_{k=n/2}^{3n/4} \binom{n}{k} \cdot 2^{-n} \cdot \left(\frac{2k}{n}\right)^T \cdot \left(\frac{2(n-k)}{n}\right)^{-T} \cdot \left(\frac{4k(n-k)}{n^2}\right)^{N/2} \\
&\le \sum_{k=n/2}^{3n/4} \binom{n}{k} \cdot 2^{-n} \cdot \left(\frac{k}{n-k}\right)^T
\end{aligned}$$

If $T < 0$, then we are done. So assume $T > 0$. For $n/2 \le k \le 3n/4$, $\frac{k}{n-k} = 1 + \frac{2k-n}{n-k} \le 1 + \frac{8(k-n/2)}{n}$. For $k \le n/2 + T$, the sum is small as $\frac{k}{n-k}$ is small. Otherwise $(1 + \frac{8(k-n/2)}{n})^T \lesssim (1 + \frac{8T}{n})^{k-n/2}$. Then the sum

$$\begin{aligned}
&\le 2^{-n} \sum_{k=n/2+T}^{3n/4} \binom{n}{k} \cdot \left(1 + \frac{8T}{n}\right)^{k-n/2} \\
&\le 2^{-n} \sum_{k=n/2+T}^{3n/4} \binom{n}{k} \cdot \left(1 + \frac{8T}{n}\right)^{k-n/2} \left(1 - \frac{8T}{n}\right)^{n/2-k} \\
&=\le 2^{-n} \sum_{k=n/2+T}^{3n/4} \binom{n}{k} \cdot \left(1 + \frac{8T}{n}\right)^{k} \left(1 - \frac{8T}{n}\right)^{n-k} \left(1 + \frac{8T}{n}\right)^{-n/2} \left(1 - \frac{8T}{n}\right)^{n/2}
\end{aligned}$$

Now $\sum_{k=n/2+T}^{3n/4} \binom{n}{k} \cdot \left(1+\frac{8T}{n}\right)^k \left(1-\frac{8T}{n}\right)^{n-k} \leq 2^n$ by binomial theorem, and $\left(1+\frac{8T}{n}\right)^{-n/2} \left(1-\frac{8T}{n}\right)^{n/2} = \left(1-\frac{64T^2}{n^2}\right)^{-n/2}$ is a constant, since $T \leq C\sqrt{N}$ for some constant C. This completes the proof. $\square$

The next lemma relates the information cost of a protocol w.r.t two distributions that are close in statistical distance. We haven't seen the lemma in this specific form elsewhere. Nevertheless it is not hard to prove.

Lemma 4.4. *Let μ_1 and μ_2 be distributions on $\{0,1\}^N \times \{0,1\}^N$ such that $|\mu_1 - \mu_2| \leq \epsilon$. Also let $\epsilon < 1/2$. Let π be a protocol for solving a function (possibly partial) with domain $\{0,1\}^N \times \{0,1\}^N$. Then $|IC(\pi,\mu_1) - IC(\pi,\mu_2)| \leq 4N\epsilon + 2H(2\epsilon)$. If ϵ is a constant and N large enough, then $|IC(\pi,\mu_1) - IC(\pi,\mu_2)| \leq 5N\epsilon$. In general for distributions over $\mathcal{X} \times \mathcal{Y}$, we get $|IC(\pi,\mu_1) - IC(\pi,\mu_2)| \leq 2\epsilon \log(|\mathcal{X}| \cdot |\mathcal{Y}|) + 2H(2\epsilon)$.*

Proof. We will design random variables X, Y, E such that $X, Y \in \{0,1\}^N$ and $E \in \{0,1,2\}$, $X,Y|E \in \{0,1\} \sim \mu_1$, $X,Y|E \in \{0,2\} \sim \mu_2$ and $Pr[E=1] = Pr[E=2] \leq \epsilon$. First let us see how this helps. Let Π denote the random variable for the transcript of the protocol when the inputs are X, Y. Let $X_1Y_1 \sim \mu_1$ and $X_2Y_2 \sim \mu_2$. Also let Π_1 and Π_2 denote the random variables for the transcript in these cases respectively.

$$\begin{aligned} I(\Pi; X|YE) &= Pr[E=0] \cdot I(\Pi; X|Y, E=0) + Pr[E=1] \cdot I(\Pi; X|Y, E=1) + Pr[E=2] \cdot I(\Pi; X|Y, E=2) \\ &= Pr[E \in \{0,1\}] \cdot I(\Pi; X|Y, E_{\{0,1\}}) + Pr[E=2] \cdot I(\Pi; X|Y, E=2) \end{aligned}$$

Here conditioning on $E_{\{0,1\}}$ means that $E \in \{0,1\}$ and that both Alice and Bob know the value of E. Now $I(\Pi; X|Y, E \in \{0,1\}) \leq I(\Pi; X|Y, E_{\{0,1\}}) + H(E|E \in \{0,1\}) = I(\Pi; X|Y, E_{\{0,1\}}) + C_1$, where $C_1 \leq H(\epsilon/(1-\epsilon) \leq H(2\epsilon)$. Also $I(\Pi; X|Y, E=2) \leq N$ and $I(\Pi; X|Y, E \in \{0,1\}) = I(\Pi_1; X_1|Y_1)$. Thus

$$I(\Pi; X|YE) = (1 - Pr[E=2]) \cdot (I(\Pi_1; X_1|Y_1) + C_1) + Pr[E=2] \cdot C_2$$

where $C_1 \leq 1$ and $C_2 \leq N$. Similarly

$$I(\Pi; X|YE) = (1 - Pr[E=1]) \cdot (I(\Pi_2; X_2|Y_2) + C_3) + Pr[E=1] \cdot C_4$$

where $C_3 \leq H(2\epsilon)$ and $C_4 \leq N$. Equating the two we get that

$$(1 - Pr[E=1]) \cdot (I(\Pi_1; X_1|Y_1) - I(\Pi_2; X_2|Y_2)) = Pr[E=1] \cdot (C_4 - C_3) + (1 - Pr[E=1]) \cdot (C_2 - C1)$$

Since $Pr[E=1] \leq \epsilon \leq 1/2$, we get that

$$|I(\Pi_1; X_1|Y_1) - I(\Pi_2; X_2|Y_2)| \leq 2N\epsilon + H(2\epsilon)$$

and hence $|IC(\pi,\mu_1) - IC(\pi,\mu_2)| \leq 4N\epsilon + 2H(2\epsilon)$.

Now let us see how to design random variables X, Y, E satisfying the given conditions. Let U, V, P denote the random variables obtained by sampling uniformly from $\{0,1\}^N \times \{0,1\}^N \times [0,1]$. Let G denote the event that $P < \max(\mu_1(U,V), \mu_2(U,V))$. Let $X, Y = U, V|G$. Also define a random variable $F \in \{0,1,2\}$ as follows :

- $F = 0$, if $P < \min(\mu_1(U,V), \mu_2(U,V))$
- $F = 1$, if $\mu_2(U,V) \leq P < \mu_1(U,V)$
- $F = 2$, if $\mu_1(U,V) \leq P < \mu_2(U,V)$

Now define $E = F|G$. Let us verify that X, Y, E satisfy the conditions.

$$\begin{aligned} Pr[X=x, Y=y|E \in \{0,1\}] &= \frac{Pr[U=x, V=y, F \in \{0,1\}, G]}{Pr[F \in \{0,1\}, G]} \\ &= \frac{\frac{1}{2^{2N}} \mu_1(x,y)}{\sum_{x,y} \frac{1}{2^{2N}} \mu_1(x,y)} = \mu_1(x,y) \end{aligned}$$

Thus $X, Y|E \in \{0,1\} \sim \mu_1$. Similarly $X, Y|E \in \{0,2\} \sim \mu_2$. Also

$$\begin{aligned} Pr[E=1] = Pr[F=1|G] &= \sum_{x,y} Pr[U=x, V=y|G] Pr[F=1|G, U=x, V=y] \\ &= \sum_{x,y \text{ s.t. } \mu_1(x,y) > \mu_2(x,y)} \frac{\frac{1}{2^{2N}} \max(\mu_1(x,y), \mu_2(x,y))}{\frac{1}{2^{2N}} \sum_{x,y} \max(\mu_1(x,y), \mu_2(x,y))} \cdot \frac{\mu_1(x,y) - \mu_2(x,y)}{\max(\mu_1(x,y), \mu_2(x,y))} \\ &= \frac{\sum_{x,y \text{ s.t. } \mu_1(x,y) > \mu_2(x,y)} (\mu_1(x,y) - \mu_2(x,y))}{\sum_{x,y} \max(\mu_1(x,y), \mu_2(x,y))} \end{aligned}$$

Thus $Pr[E=1] = \frac{|\mu_1 - \mu_2|}{\sum_{x,y} \max(\mu_1(x,y), \mu_2(x,y))} \leq |\mu_1 - \mu_2| \leq \epsilon$. Similarly $Pr[E=2] = \frac{|\mu_1 - \mu_2|}{\sum_{x,y} \max(\mu_1(x,y), \mu_2(x,y))}$. Hence $Pr[E=1] = Pr[E=2] \leq \epsilon$. This completes the proof. The general form can be proved in a similar manner.

□

We also need a lemma which relates the information cost of distributions which are not very skewed w.r.t to each other. Formally

Lemma 4.5. *Let μ_1 and μ_2 be distributions over $\{0,1\}^N \times \{0,1\}^N$ such that $\mu_1 \leq M \cdot \mu_2$ for some constant M. Let f be a function (possibly partial) with domain $\{0,1\}^N \times \{0,1\}^N$ and let π be a protocol for solving it. Then $IC(\pi, \mu_1) \leq M \cdot IC(\pi, \mu_2)$.*

Proof. Let $X_1, Y_1 \sim \mu_1$ and Π_1 denote the random variable for the transcript when inputs are X_1, Y_1. Let $X_2, Y_2 \sim \mu_2$ and define Π_2 similarly. Now

$$I(\Pi_1; X_1|Y_1) = \mathbb{E}_{x,y \sim \mu_1} D[\Pi_1|_{x,y} || \Pi_1|_y] = \mathbb{E}_y(\mathbb{E}_x D[\Pi_1|_{x,y} || \Pi_1|_y])$$

By Fact 2.6, $\mathbb{E}_x D[\Pi_1|_{x,y} || \Pi_1|_y] \leq \mathbb{E}_x D[\Pi_1|_{x,y} || \Pi_2|_y]$. Also $\Pi_1|_{x,y} = \Pi_2|_{x,y}$. Thus

$$\begin{aligned} I(\Pi_1; X_1|Y_1) \leq \mathbb{E}_{x,y \sim \mu_1} D[\Pi_2|_{x,y} || \Pi_2|_y] &\leq M \cdot \mathbb{E}_{x,y \sim \mu_2} D[\Pi_2|_{x,y} || \Pi_2|_y] \\ &= M \cdot I(\Pi_2; X_2|Y_2) \end{aligned}$$

Hence $IC(\pi, \mu_1) \leq M \cdot IC(\pi, \mu_2)$. □

The next lemma says that if the information cost w.r.t the distribution μ from Lemma 4.3 is high, then the information cost w.r.t the uniform distribution is high too.

Lemma 4.6. *Let $f : \{0,1\}^N \times \{0,1\}^N \to \{0,1\}$ be a function (possibly partial). Let μ be a distribution over $\{0,1\}^N \times \{0,1\}^N$, as defined in Lemma 4.3. Then if $IC(f, \mu, \delta) \geq \Omega(N)$, for some $\delta > 0$, then $IC(f, \mathcal{U}_N, \eta) \geq \Omega(N)$, for some $\eta > 0$.*

Proof. Let π be a protocol for computing f with error η w.r.t. the distribution $\mathcal{U}_N$, and information cost $IC(\pi, \mathcal{U}_N) = I$. Let $\epsilon > 0$. Then by Lemma 4.3, for N large enough, there exists a distribution μ_ϵ over $\{0,1\}^N \times \{0,1\}^N$ such that $|\mu - \mu_\epsilon| \leq \epsilon$ and $\mu_\epsilon \leq M_\epsilon \cdot \mathcal{U}_N$ for some constant M_ϵ. Then error of the protocol π w.r.t. μ is $\leq M_\epsilon \eta + \epsilon$. Also the information cost of π w.r.t. μ is $\leq M_\epsilon I + 5N\epsilon$ (using Lemmas 4.4 and 4.5). Now if $M_\epsilon \eta + \epsilon \leq \delta$, then $M_\epsilon I + 5N\epsilon \geq c \cdot N$, for some constant c. Take $\epsilon = \min(\delta/2, c/10)$ and $\eta = (\delta - \epsilon)/M_\epsilon$. Then $I \geq cN/2M_\epsilon$. Thus $IC(f, \mathcal{U}_N, \eta) \geq \Omega(N)$. □

Proof. (of Theorem 1.2) Note that because of Lemma 4.6, we just need to prove that $IC(GHD_{N,N/2,\sqrt{N}}, \mu, \epsilon) \geq \Omega(N)$ for some $\epsilon > 0$ for the distribution μ in Lemma 4.3. Assume that for all $\epsilon > 0$, $IC(GHD_{N,N/2,\sqrt{N}}, \mu, \epsilon) = o(N)$. That is for all β, ϵ, and for N sufficiently large, $IC(GHD_{N,N/2,\sqrt{N}}, \mu, \epsilon) \leq \beta \cdot N$. By Lemma 3.4, there exist constants $\epsilon' > 0$, $\gamma > 0$ and $c > 0$ such that $IC(GHD_{n,n/2,\gamma\sqrt{n}}, \mathcal{U}, \epsilon') \geq c \cdot n$.

Let α be a large integer to be determined later. Set $N = \alpha \cdot n$. Let π_N be a protocol that solves $GHD_{N,N/2,\sqrt{N}}$ with error $\leq \epsilon$ w.r.t μ, and let the information cost of π_N w.r.t μ be $\leq \beta \cdot N$. Consider the following protocol $\pi_n(x,y)$ for $GHD_{n,n/2,\gamma\sqrt{n}}$: Pick N random coordinates of x, y, call them u', v'. Now pick a random string $r \in_R \{0,1\}^N$ and set $u = u' \oplus r$ and $v = v' \oplus r$. Run π_N on u, v. Let $X, Y \sim \mathcal{U}_n$ be the inputs for π_n. Let U, V denote the random variables denoting the sampled coordinates. Note that $U, V \sim \mu$. Let Π denote the random variable for the transcript of running π_N on U, V. Then the transcript of running π_n on X, Y is ΠR, where R denotes the public randomness involved in sampling u, v from x, y. Now

$$I(\Pi R; X|Y) = I(R; X|Y) + I(\Pi; X|YR) = I(\Pi; X|YR) = I(\Pi; X|VYR)$$

The last equality follows from the fact that V is a deterministic function of YR. Now Π is a probabilistic function of U, V, and the internal randomness of the protocol Π is independent of X, Y and R. Thus $I(\Pi; XYR|UV) = 0$. Since

$$I(\Pi; XYR|UV) = I(\Pi; YR|UV) + I(\Pi; X|UVYR)$$

and $I(\Pi; YR|UV) = 0$, $I(\Pi; X|UVYR) = 0$. Applying Fact 2.8, with $A = \Pi$, $B = U$, $C = X$ and $D = VYR$, we get that $I(\Pi; X|VYR) \leq I(\Pi; U|VYR)$. Also $I(\Pi; YR|UV) = 0$. Applying Fact 2.7 with $A = U$, $B = \Pi$, $C = V$ and $D = YR$, we get $I(\Pi; U|V) \geq I(\Pi; U|VYR)$. This implies that $I(\Pi R; X|Y) \leq I(\Pi; U|V)$. A similar argument shows that $I(\Pi R; Y|X) \leq I(\Pi; V|U)$ and hence $IC(\pi_n, \mathcal{U}_n) \leq IC(\pi_N, \mu)$.

Now let us calculate the error of the protocol π_n. If $HAM(x,y) \geq n/2 + \gamma\sqrt{n}$, then for a random coordinate I, $Pr[x_I \oplus y_I = 1] \geq 1/2 + \gamma/\sqrt{n}$. Then the expected hamming distance of N random coordinates is $N/2 + \gamma\sqrt{\alpha}\sqrt{N}$. Probability that the hamming distance is $\leq N/2 + \frac{\gamma\sqrt{\alpha}}{2}\sqrt{N}$ is bounded by $e^{-\frac{\alpha\gamma^2}{2}}$. Similarly for the lower case. Choose α so that $\gamma\sqrt{\alpha} \geq 2$ and $e^{-\frac{\alpha\gamma^2}{2}} \leq \epsilon'/2$. Then

$$\begin{aligned}\text{error}(\pi_n) = &\sum_{x,y \text{ s.t. } HAM(x,y) \geq n/2 + \gamma\sqrt{n}} \mathcal{U}_n(x,y) \cdot Pr[\pi_n \text{ outputs 0 on input } x,y] \\ &+ \sum_{x,y \text{ s.t. } HAM(x,y) \leq n/2 - \gamma\sqrt{n}} \mathcal{U}_n(x,y) \cdot Pr[\pi_n \text{ outputs 1 on input } x,y]\end{aligned}$$

Now

$$Pr[\pi_n \quad \text{outputs 0 on input } x,y] \quad = \quad \sum_{u,v} \mu(u,v|x,y) \quad \cdot \quad Pr[\pi_N \quad \text{outputs 0 on input } u,v]$$

where $\mu(u,v|x,y)$ the probability of getting u, v when coordinates are sampled from x, y. For x, y s.t. $HAM(x,y) \geq n/2 + \gamma\sqrt{n}$,

$$\begin{aligned}&\sum_{u,v} \mu(u,v|x,y) \cdot Pr[\pi_N \text{ outputs 0 on input } u,v] \leq \\ &\qquad \sum_{u,\,v \text{ s.t. } HAM(u,\,v) \,\geq\, N/2 + \sqrt{N}} \mu(u,v|x,y) \cdot Pr[\pi_N \text{ outputs 0 on input } u,v] + \epsilon'/2\end{aligned}$$

Doing a similar exercise for the other half, we get that

$$\begin{aligned}\text{error}(\pi_n) \leq &\sum_{u,\,v \text{ s.t. } HAM(u,\,v) \,\geq\, N/2 + \sqrt{N}} \mu(u,v) \cdot Pr[\pi_N \text{ outputs 0 on input } u,v] + \\ &\sum_{u,\,v \text{ s.t. } HAM(u,\,v) \,\leq\, N/2 - \sqrt{N}} \mu(u,v) \cdot Pr[\pi_N \text{ outputs 1 on input } u,v] + \epsilon'/2 \\ &= \text{error}(\pi_N) + \epsilon'/2\end{aligned}$$

Choosing $\epsilon = \epsilon'/2$, and $\beta = c/2\alpha$, we get a protocol π_n with error $\leq \epsilon'$ and information cost $\leq \beta\alpha n \leq cn/2$, which is a contradiction. $\square$

5 Information Complexity of Inner Product

The inner product function $IP_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is defined as follows:

$$IP_n(x,y) = \sum_{i=0}^{n} x_i y_i \pmod 2$$

The proof exploits the self-reducible structure of the inner-product function. But since, IP_n is such a sensitive function, we will first prove a statement about the 0-error information cost, and then use continuity of information cost to argue about non-zero errors.

We will need the following lemma from [8]. It is essentially the same as Theorem 4.1, just that when dealing with 0 error, we cannot ensure that error on each copy is 0. We just have an overall error which is the error introduced if compression fails.

Lemma 5.1. *Let $f : X \times Y \to \{0,1\}$ be a function, and let μ be a distribution over the inputs. Let π be a protocol computing f with error 0 w.r.t μ, and internal information cost $IC_\mu(\pi) = I$. Then for all $\delta > 0$, $\epsilon > 0$, there is a protocol π_n for computing f^n with error ϵ w.r.t μ^n, with worst case communication cost*

$$\begin{aligned} &= n(I+\delta/4) + O(\sqrt{CC(\pi)\cdot n \cdot (I+\delta/4)}) + O(\log(1/\epsilon)) + O(CC(\pi)) \\ &\leq n(I+\delta) \text{ (for } n \text{ sufficiently large)} \end{aligned}$$

The following lemma from [4] relates the information cost of computing XOR of n copies of a function f to the information cost of a single copy.

Lemma 5.2. *Let f be a function, and let μ be a distribution over the inputs. Then $IC_{\mu^n}(\oplus_n f, \epsilon) \geq n(IC_\mu(f,\epsilon) - 2)$*

The next lemma says that there is no 0-error protocol for IP_n which conveys slightly less information than the trivial protocol.

Lemma 5.3. *$\forall n, IC_{\mathcal{U}_n}(IP_n, 0) \geq n$, where $\mathcal{U}_n$ is the uniform distribution over $\{0,1\}^n \times \{0,1\}^n$*

Proof. It is known that $D_\epsilon^{\mathcal{U}_n}(IP_n) \geq n - c_\epsilon$, for all constant $\epsilon \in (0, 1/2)$, where c_ϵ is a constant depending just on ϵ [17, 13]. Assume that for some n, $IC_{\mathcal{U}_n}(IP_n, 0) \leq n - c$. Then using, Lemma 5.1 with $\delta = c/2$ and $\epsilon = 1/3$, we can get a protocol π for solving N copies of IP_n with overall error 1/3 w.r.t $\mathcal{U}_n^N$, and $CC(\pi) \leq N(n - c + c/2)$. This gives us a protocol π' for solving IP_{Nn} with error 1/3 w.r.t the uniform distribution, and $CC(\pi') \leq Nn - Nc/2$ (divide the inputs into N chunks, solve the N chunks using π and XOR the answers). But $CC(\pi') \geq Nn - c_{1/3}$, a contradiction. ☐

Proof. (of Theorem 1.3) We use the continuity of (internal) information cost in the error parameter at $\epsilon = 0$:

Theorem 5.4. *([6]) For all $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and $\mu \in \Delta(\mathcal{X} \times \mathcal{Y})$ we have*

$$\lim_{\epsilon \to 0} \mathrm{IC}_\mu(f, \epsilon) = \mathrm{IC}_\mu(f, 0) \tag{3}$$

Given $\delta > 0$, let $l = \lceil \frac{3}{\delta} \rceil$. Then

$$IC_{\mathcal{U}_l}(IP_l, 0) \geq l \geq (1-\delta)l + 3$$

Since $\lim_{\epsilon \to 0} IC_{\mathcal{U}_l}(IP_l, \epsilon) = IC_{\mathcal{U}_l}(IP_l, 0)$, $\exists\ \epsilon(l,\delta) = \epsilon(\delta)$ s.t.

$$IC_{\mathcal{U}_l}(IP_l, \epsilon) \geq (1-\delta)l + 2$$

Now using Lemma 5.2, we get that $IC_{\mathcal{U}_l^N}(\oplus_N IP_l, \epsilon) \geq (1-\delta)Nl$. Thus $IC_{\mathcal{U}_{Nl}}(IP_{Nl}, \epsilon) \geq (1-\delta)Nl$. Thus for sufficiently large n, $IC_{\mathcal{U}_n}(IP_n, \epsilon) \geq (1-\delta)n$. ☐

References

[1] A. Ada, A. Chattopadhyay, S. Cook, L. Fontes, M. Koucky, and T. Pitassi. The hardness of being private. In *Computational Complexity (CCC), 2012 IEEE 27th Annual Conference on*, pages 192–202. IEEE, 2012.

[2] Noga Alon and Joel Spencer. *The Probabilistic Method.* John Wiley, 1992.

[3] Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004.

[4] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *STOC*, pages 67–76, 2010.

[5] Mark Braverman. Interactive information complexity. In *STOC*, pages 505–524, 2012.

[6] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. From information to exact communication. *Electronic Colloquium on Computational Complexity (ECCC)*, 19(171), 2012.

[7] Mark Braverman and Ankur Moitra. An information complexity approach to extended formulations. *Electronic Colloquium on Computational Complexity (ECCC)*, 19:131, 2012.

[8] Mark Braverman and Anup Rao. Information equals amortized communication. *CoRR*, abs/1106.3595, 2010.

[9] Mark Braverman and Omri Weinstein. A discrepancy lower bound for information complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 18:164, 2011.

[10] Amit Chakrabarti, Ranganath Kondapally, and Zhenghui Wang. Information complexity versus corruption and applications to orthogonality and gap-hamming. *CoRR*, abs/1205.0968, 2012.

[11] Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. In *STOC*, pages 51–60, 2011.

[12] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In Bob Werner, editor, *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 270–278, Los Alamitos, CA, October 14–17 2001. IEEE Computer Society.

[13] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.

[14] D.A. Huffman. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9):1098–1101, 1952.

[15] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. *CoRR*, abs/1204.1505, 2012.

[16] Hartmut Klauck. Quantum and approximate privacy. *Theory Comput. Syst.*, 37(1):221–246, 2004.

[17] Eyal Kushilevitz and Noam Nisan. *Communication complexity.* Cambridge University Press, Cambridge, 1997.

[18] A. McGregor, I. Mironov, T. Pitassi, O. Reingold, K. Talwar, and S. Vadhan. The limits of two-party differential privacy. In *51st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 81–90, 2010.

[19] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27, 1948. Monograph B-1598.

[20] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *STOC*, pages 209–213, 1979.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de