Communication Complexity of Permutation-Invariant Functions

Badih Ghazi[*]   Pritish Kamath[†]   Madhu Sudan[‡]

May 30, 2015

Abstract

Motivated by the quest for a broader understanding of communication complexity of simple functions, we introduce the class of "permutation-invariant" functions. A partial function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1,?\}$ is permutation-invariant if for every bijection $\pi : \{1,\dots,n\} \to \{1,\dots,n\}$ and every $\mathbf{x}, \mathbf{y} \in \{0,1\}^n$, it is the case that $f(\mathbf{x},\mathbf{y}) = f(\mathbf{x}^\pi, \mathbf{y}^\pi)$. Most of the commonly studied functions in communication complexity are permutation-invariant. For such functions, we present a simple complexity measure (computable in time polynomial in n given an implicit description of f) that describes their communication complexity up to polynomial factors and up to an additive error that is logarithmic in the input size. This gives a coarse taxonomy of the communication complexity of simple functions. Our work highlights the role of the well-known lower bounds of functions such as SET-DISJOINTNESS and INDEXING, while complementing them with the relatively lesser-known upper bounds for GAP-INNER-PRODUCT (from the sketching literature) and SPARSE-GAP-INNER-PRODUCT (from the recent work of Canonne et al. [ITCS 2015]). We also present consequences to the study of communication complexity with imperfectly shared randomness where we show that for total permutation-invariant functions, imperfectly shared randomness results in only a polynomial blow-up in communication complexity after an additive $O(\log \log n)$ overhead.

Contents

1 Introduction

Communication complexity, introduced by Yao [Yao79], has been a central object of study in complexity theory. In the *two-way model*, two players, Alice and Bob, are given private inputs $\mathbf{x}$ and $\mathbf{y}$ respectively, along with some shared randomness, and they exchange bits according to a predetermined protocol and produce an output. The protocol computes a function $f(\cdot, \cdot)$ if the output equals $f(\mathbf{x}, \mathbf{y})$ with high probability over the randomness [1]. The communication complexity of f is the minimum over all protocols computing f of the maximum, over inputs $\mathbf{x}$ and $\mathbf{y}$, of the number of bits exchanged by the protocol. The one-way communication model is defined similarly except that all the communication is from Alice to Bob and the output is produced by Bob. For an overview of communication complexity, we refer the reader to the book [KN97] and the survey [LS09].

While communication complexity of functions has been extensively studied, the focus typically is on lower bounds. Lower bounds on communication complexity turn into lower bounds on Turing machine complexity, circuit depth, data structures, streaming complexity, just to name a few. On the other hand, communication complexity is a very natural notion to study on its own merits and indeed positive results in communication complexity can probably be very useful in their own rights, by suggesting efficient communication mechanisms and paradigms in specific settings. For this perspective to be successful, it would be good to have a compact picture of the various communication protocols that are available, or even the ability to determine, given a function f, the best, or even a good, communication protocol for f. Of course such a goal is overly ambitious. For example, the seminal work of Karchmer and Wigderson [KW90] implies that finding the best protocol for f is as hard as finding the best (shallowest) circuit for some related function $\tilde{f}$.

Given this general barrier, one way to make progress is to find a restrictive, but natural, subclass of all functions and to characterize the complexity of all functions within this class. Such approaches have been very successful in the context of non-deterministic computation by restricting to satisfiability problems [Sch78], in optimization and approximation by restricting to constraint satisfaction problems [Cre95, KSTW00], in the study of decision tree complexity by restricting to graph properties [Ros73], or in the study of property testing by restricting to certain symmetric properties (see the survey [Sud10, Gol10] and the references therein). In the above cases, the restrictions have led to characterizations (or conjectured characterizations) of the complexity of all functions in the restricted class. In this work, we attempt to bring in a similar element of unification to communication complexity.

In this work, we introduce the class of "permutation-invariant" (total or partial) functions. Let $[n]$ denote the set $\{1, \cdots, n\}$. A function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1,?\}$ is permutation invariant if for every bijection $\pi : [n] \to [n]$ and every $\mathbf{x}, \mathbf{y} \in \{0,1\}^n$ it is the case that $f(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}^\pi, \mathbf{y}^\pi)$. We propose to study the communication complexity of this class of functions.

To motivate this class, we note that most of the commonly studied functions in communication complexity including EQUALITY [Yao79], (GAP) HAMMING DISTANCE [Woo04, JKS08, CR12, Vid11, She12, PEG86, Yao03, HSZZ06, BBG14], (GAP) INNER PRODUCT, (SMALL-SET) DISJOINTNESS [KS92, Raz92, HW07, ST13], SMALL-SET INTERSECTION [BCK$^+$14] are all permutation-invariant functions. Other functions, such as INDEXING [JKS08], can be expressed without changing the input length significantly, as permutation-invariant functions. Permutation-invariant functions also include as subclasses, several classes of functions that have been well-studied in communication complexity, such as (AND)-symmetric functions [BdW01, Raz03, She11] and XOR-symmetric functions [ZS09]. It is worth noting that permutation-invariant functions are completely expres-

---

[1]In this work, we also consider partial functions where $f(\mathbf{x}, \mathbf{y})$ may sometimes be undetermined, denoted $f(\mathbf{x}, \mathbf{y}) = ?$. In such cases, the protocol can output anything when $f(\mathbf{x}, \mathbf{y})$ is undetermined.

sive if one allows an exponential blow-up in input size, namely, for every function $f(\mathbf{x}, \mathbf{y})$ there are functions F, A, B s.t. F is permutation-invariant and $f(\mathbf{x}, \mathbf{y}) = F(A(\mathbf{x}), B(\mathbf{y}))$. So results on permutation-invariant functions that don't depend on the input size apply to all functions. Finally, we point out that permutation-invariant functions have an important standpoint among functions with small communication complexity, as permutation-invariance often allows the use of hashing/bucketing based strategies, which would allow us to get rid of the dependence of the communication complexity on the input length n. We also note that functions on non-Boolean domains that are studied in the literature on sketching such as distance estimation (given $\mathbf{x}, \mathbf{y} \in \mathrm{R}^n$, decide if $\|\mathbf{x} - \mathbf{y}\|_p \leq d$ or if $\|\mathbf{x} - \mathbf{y}\|_p > d(1 + \epsilon)$) are also permutation-invariant. In particular, the resulting sketching/communication protocols are relevant to (some functions in) our class.

1.1 Coarse characterization of Communication Complexity

Permutation-invariant functions on n-bits are naturally succinctly described (by $O(n^3)$ bits). Given this natural description, we introduce a simple combinatorial measure $m(f)$ (which is easy to compute, in particular in time $\mathrm{poly}(n)$ given f) which produces a coarse approximation of the communication complexity of f. We note that our objective is different from that of the standard objectives in the study of communication complexity lower bounds, where the goal is often to come up with a measure that has nice mathematical properties, but may actually be more complex to compute than communication complexity itself. In particular, this is true of the Information Complexity measure introduced by [CSWY01] and [BYJKS02], and used extensively in recent works, and which, until recently, was not even known to be approximately computable [BS15] (whereas communication complexity can be computed exactly, albeit in doubly exponential time). Nevertheless, our work does rely on known bounds on the information complexity of some well-studied functions and our combinatorial measure $m(f)$ also coarsely approximates the information complexity for all the functions that we study.

To formally state our first theorem, let $\mathrm{R}(f)$ denote the randomized communication complexity of a function f and $\mathrm{IC}(f)$ denote its information complexity. Our result about our combinatorial measure $m(f)$ (see Definition 3.2) is summarized below.

Theorem 1.1. *Let* $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1,?\}$ *be a (total or partial) permutation-invariant function. Then,*
$$m(f) \leq \mathrm{IC}(f) \leq \mathrm{R}(f) \leq \mathrm{poly}(m(f)) + O(\log n).$$

In other words, the combinatorial measure $m(f)$ approximates communication complexity to within a polynomial factor, up to an additive $O(\log n)$ factor. Our result is constructive — given f it gives a communication protocol whose complexity is bounded from above by $\mathrm{poly}(m(f)) + O(\log n)$. It would be desirable to get rid of the $O(\log n)$ factor but this seems hard without improving the state of the art vis-a-vis communication complexity and information complexity. To see this, first note that our result above also implies that information complexity provides a coarse approximator for communication complexity. Furthermore, any improvement to the additive $O(\log n)$ error in this relationship would imply improved relationship between information complexity and communication complexity for general functions (better than what is currently known). Specifically, we note:

Proposition 1.2. *Let* $G(\cdot)$ *be a function such that* $\mathrm{R}(f) \leq \mathrm{poly}(\mathrm{IC}(f)) + G(\log n)$ *for every permutation-invariant partial function* f *on* $\{0,1\}^n \times \{0,1\}^n$. *Then, for every (general) partial function* g *on* $\{0,1\}^n \times \{0,1\}^n$, *we have* $\mathrm{R}(g) \leq \mathrm{poly}(\mathrm{IC}(g)) + G(n)$.

Thus, even an improvement from an additive $O(\log n)$ to additive $o(\log n)$ would imply new relationships between information complexity and communication complexity.

1.2 Communication with imperfectly shared randomness

Next, we turn to communication complexity when the players only share randomness imperfectly, a model introduced by [BGI14, CGMS14]. Specifically, we consider the setting where Alice gets a sequence of bits $r = (r_1, \ldots, r_t)$ and Bob gets a sequence of bits $s = (s_1, \ldots, s_t)$ where the pairs (r_i, s_i) are identically and independently distributed according to distribution $\mathrm{DSBS}(\rho)$, which means, the marginals of r_i and s_i are uniformly distributed in $\{0, 1\}$ and r_i and s_i are ρ-correlated (i.e., $\Pr[r_i = s_i] = 1/2 + \rho/2$).

The question of what can interacting players do with such a correlation has been investigated in many different contexts including information theory [GK73, Wit75], probability theory [MO05, BM11, CMN14, MOR$^+$06], cryptography [BS94, Mau93, RW05] and quantum computing [BBP$^+$96]. In the context of communication complexity, however, this question has only been investigated recently. In particular, Bavarian et al. [BGI14] study the problem in the Simultaneous Message Passing (SMP) model and Canonne et al. [CGMS14] study it in the standard one-way and two-way communication models. Let $\mathrm{ISR}_\rho(f)$ denote the communication complexity of a function f when Alice and Bob have access to ρ-correlated bits. The work of [CGMS14] shows that for any total or partial function f with communication complexity $R(f) \le k$ it is the case that $\mathrm{ISR}_\rho(f) \le \min\{O(2^k), k + O(\log n)\}$. They also give a partial function f with $R(f) \le k$ for which $\mathrm{ISR}_\rho(f) = \Omega(2^k)$. Thus, imperfect sharing leads to an exponential slowdown for low-communication promise problems.

One of the motivations of this work is to determine if the above result is tight for total functions. Indeed, for most of the common candidate functions with low-communication complexity such as SMALL-SET-INTERSECTION and SMALL-HAMMING-DISTANCE, we show (in Section 4.1) that $\mathrm{ISR}_\rho(f) \le \mathrm{poly}(\mathrm{R}(f))$. [2] This motivates us to study the question more systematically and we do so by considering permutation-invariant total functions. For this class, we show that the communication complexity with imperfectly shared randomness is within a polynomial of the communication complexity with perfectly shared randomness up to an additive $O(\log \log n)$ factor; this is a tighter connection than what is known for general functions. Interestingly, we achieve this by showing that the same combinatorial measure $m(f)$ also coarsely captures the communication complexity under imperfectly shared randomness. Once again, we note that the $O(\log \log n)$ factor is tight unless we can improve the upper bound of [CGMS14].

Theorem 1.3. *Let $f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$ be a permutation-invariant total function. Then, we have*

$$\mathrm{ISR}_\rho(f) \le \mathrm{poly}(\mathrm{R}(f)) + O(\log \log n)$$

Furthermore, $\mathrm{ISR}_\rho^{1\text{-}way}(f) \le \mathrm{poly}(\mathrm{R}^{1\text{-}way}(f)) + O(\log \log n)$.

1.3 Overview of Proofs

Our proof of Theorem 1.1 starts with the simple observation that for any permutation-invariant partial function $f(\cdot, \cdot)$, its value $f(\mathbf{x}, \mathbf{y})$ is determined completely by $|\mathbf{x}|$, $|\mathbf{y}|$ and $\Delta(\mathbf{x}, \mathbf{y})$ (where $|\mathbf{x}|$ denotes the Hamming weight of $\mathbf{x}$ and $\Delta(\mathbf{x}, \mathbf{y})$ denotes the (non-normalized) Hamming distance between $\mathbf{x}$ and $\mathbf{y}$). By letting Alice and Bob exchange $|\mathbf{x}|$ and $|\mathbf{y}|$ (using $O(\log n)$ bits of communication), the problem now reduces to being a function only of the Hamming distance

---

[2]In fact, this polynomial relationship holds for a broad subclass of permutation-invariant functions that we call "strongly permutation-invariant". A function $f(x, y)$ is strongly permutation invariant if there exists $h : \{0, 1\}^2 \to \{0, 1\}$ and symmetric function $\sigma : \{0, 1\}^n \to \{0, 1\}$ such that $f(x, y) = \sigma(h(x_1, y_1), \ldots, h(x_n, y_n))$. Theorem 4.12 shows a polynomial relationship between $\mathrm{R}(f)$ and $\mathrm{ISR}(f)$ for all strongly-permutation-invariant total functions f.

$\Delta(\mathbf{x}, \mathbf{y})$. To understand the remaining task, we introduce a multiparameter version of the Hamming distance problem $\mathrm{GHD}^n_{a,b,c,g}(\cdot, \cdot)$ where $\mathrm{GHD}^n_{a,b,c,g}(\mathbf{x}, \mathbf{y})$ is undefined if $|\mathbf{x}| \neq a$ or $|\mathbf{y}| \neq b$ or $c - g < \Delta(\mathbf{x}, \mathbf{y}) < c + g$. The function is 1 if $\Delta(\mathbf{x}, \mathbf{y}) \geq c + g$ and 0 if $\Delta(\mathbf{x}, \mathbf{y}) \leq c - g$.

This problem turns out to have different facets for different choices of the parameters. For instance, if $a \approx b \approx c$, then the communication complexity of this problem is roughly $O((c/g)^2)$ and the optimal lower bound follows from the lower bound on Gap Hamming Distance [CR12, Vid11, She12] whereas the upper bound follows from simple hashing. However, when $a \ll b$, $c \approx b$ and $g \approx a$ different bounds and protocols kick in. In this range, the communication complexity turns out to be $O(\log(c/g))$ with the upper bound coming from the protocol for Sparse-Gap-Inner-Product given in [CGMS14], and a lower bound that we give based on a reduction from Set Disjointness. In this work, we start by giving a complete picture of the complexity of GHD for all parameter settings. The lower bound for communication complexity, and even information complexity, of general permutation-invariant functions f follows immediately - we just look for the best choice of parameters of GHD that can be found in f. The upper bound requires more work in order to ensure that Alice and Bob can quickly narrow down the Hamming distance $\Delta(\mathbf{x}, \mathbf{y})$ to a range where the value of f is clear. To do this, we need to verify that f does not change values *too quickly* or *too often*. The former follows from the fact that hard instances of GHD cannot be embedded in f, and the latter involves some careful accounting, leading to a full resolution.

Turning to the study of communication with imperfectly shared randomness, we hit an immediate obstacle when extending the above strategy since Alice and Bob cannot afford to exchange $|\mathbf{x}|$ and $|\mathbf{y}|$ anymore, since this would involve $\Omega(\log n)$ bits of communication and we only have an additional budget of $O(\log \log n)$. Instead, we undertake a partitioning of the "weight-space", i.e., the set of pairs $(|\mathbf{x}|, |\mathbf{y}|)$, into a finite number of regions. For most of the regions, we reduce the communication task to one of the SMALL-SET-INTERSECTION or SMALL-HAMMING-DISTANCE problems. In the former case, the sizes of the sets are polynomially related to the randomized communication complexity, whereas in the latter case, the Hamming distance threshold is polynomially related to the communication complexity. A naive conversion to protocols for the imperfectly shared setting using the results of [CGMS14] would result in an exponential blow-up in the communication complexity. We give new protocols with imperfectly shared randomness for these two problems (which may be viewed as extensions of protocols in [BGI14] and [CGMS14]) that manage to reduce the communication blow-up to just a polynomial. This manages to take care of most regions, but not all. To see this, note that any total function $h(|\mathbf{x}|, |\mathbf{y}|)$ can be encoded as a permutation-invariant function $f(\mathbf{x}, \mathbf{y})$ and such functions cannot be partitioned into few classes. Our classification manages to eliminate all cases except such functions, and in this case, we apply Newman's theorem to conclude that the randomness needed in the perfectly shared setting is only $O(\log \log n)$ bits (since the inputs to h are in the range $[n] \times [n]$). Communicating this randomness and then executing the protocol with perfectly shared randomness gives us in this case a private-randomness protocol with communication $R(f) + O(\log \log n)$.

1.4 Roadmap of this paper

In Section 2, we give some of the basic definitions and introduce the background material necessary for understanding the contributions of this paper. In Section 3, we introduce our measure $m(f)$ and prove Theorem 1.1. In Section 4, we show the connections between communication complexity with imperfectly shared randomness and that with perfectly shared randomness and prove Theorem 1.3. We end with a summary and some future directions in Section 5.

2 Preliminaries

In this section, we provide all the necessary background needed to understand the contributions in this paper.

2.1 Notations and Definitions

Throughout this paper, we will use bold letters such as $\mathbf{x}$, $\mathbf{y}$, etc. to denote strings in $\{0, 1\}^n$, where the i-th bit of $\mathbf{x}$ will be accessed as x_i. We denote by $|\mathbf{x}|$ the *Hamming weight* of binary string $\mathbf{x}$, i.e., the number of non-zero coordinates of $\mathbf{x}$. We will also denote by $\Delta(\mathbf{x}, \mathbf{y})$ the *Hamming distance* between binary strings $\mathbf{x}$ and $\mathbf{y}$, i.e., the number of coordinates in which $\mathbf{x}$ and $\mathbf{y}$ differ. We also denote $[n] \stackrel{\text{def}}{=} \{1, \cdots, n\}$ for every positive integer n.

Very significant for our body of work is the definition of *permutation-invariant* functions, which we define as follows,

Definition 2.1 (Permutation-Invariant functions). *A (total or partial) function $f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1, ?\}$ is permutation-invariant if for all $\mathbf{x}, \mathbf{y} \in \{0, 1\}^n$ and every bijection $\pi : [n] \to [n]$, $f(\mathbf{x}^\pi, \mathbf{y}^\pi) = f(\mathbf{x}, \mathbf{y})$ (where $\mathbf{x}^\pi$ is such that $x_i^\pi = x_{\pi(i)}$).*

We note the following simple observation about permutation-invariant functions.

Observation 2.2. *Any permutation-invariant function f depends only on $|\mathbf{x} \wedge \mathbf{y}|$, $|\mathbf{x} \wedge \neg\mathbf{y}|$, $|\neg\mathbf{x} \wedge \mathbf{y}|$ and $|\neg\mathbf{x} \wedge \neg\mathbf{y}|$. Since these numbers add up to n, f really depends on any three of them, or in fact any three linearly independent combinations of them. Thus, we have that for some appropriate functions g, h,*

$$f(\mathbf{x}, \mathbf{y}) = g(|\mathbf{x}|, |\mathbf{y}|, |\mathbf{x} \wedge \mathbf{y}|) = h(|\mathbf{x}|, |\mathbf{y}|, \Delta(\mathbf{x}, \mathbf{y}))$$

We will use these 3 representations of f interchangeably throughout this paper. We will often refer to the slices of f obtained by fixing $|\mathbf{x}| = a$ and $|\mathbf{y}| = b$ for some a and b, in which case we will denote the sliced h by either $h_{a,b}(\cdot)$ or $h(a, b, \cdot)$, and similarly for g.

2.2 Communication Complexity

We define the standard notions of two-way (resp. one-way) randomized commmunication complexity[3] $R(f)$ (resp. $R^{\text{1-way}}(f)$), that is studied under shared/public randomness model (cf. [KN97]).

Definition 2.3 (Randomized communication complexity R(f)). *For any function $f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1, ?\}$, the randomized communication complexity R(f) is defined as the cost of the smallest randomized protocol, which has access to public randomness, that computes f correctly on any input with probability at least $2/3$. In particular,*

$$\text{R}(f) = \min_{\substack{\Pi \\ \forall \mathbf{x}, \mathbf{y} \in \{0,1\}^n \; s.t. \; f(\mathbf{x},\mathbf{y}) \neq \; ?: \\ \Pr[\Pi(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y})] \geq 2/3}} \text{CC}(\Pi)$$

where the minimum is taken over all randomized protocols Π, where Alice and Bob have access to public randomness.

The one-way randomized communication complexity $R^{1\text{-}way}(f)$ is defined similarly, with the only difference being that we allow only protocols Π where only Alice communicates to Bob, but not other way round.

---

[3] we will often abbreviate "communication complexity" by CC

Another notion of randomized communication complexity that is studied, is under private randomness model. The work of [CGMS14] sought out to study an intermediate model, where the two parties have access to i.i.d. samples from a correlated random source $\mu(r,s)$, that is, Alice has access to r and Bob has access to s. In their work, they considered the *doubly symmetric binary source*, parametrized by ρ, defined as follows,

Definition 2.4 (Doubly Symmetric Binary Source DSBS(ρ))**.** DSBS(ρ) *is a distribution on* $\{0,1\} \times \{0,1\}$*, such that for* $(r,s) \sim$ DSBS(ρ)*,*

$$\Pr[r=1, s=1] = \Pr[r=0, s=0] = (1+\rho)/4$$

$$\Pr[r=1, s=0] = \Pr[r=0, s=1] = (1-\rho)/4$$

Note that $\rho = 1$ corresponds to the standard notion of public randomness, and $\rho = 0$ corresponds to the standard notion of private randomness.

Definition 2.5 (Communication complexity with imperfectly shared randomness [CGMS14])**.** *For any function* $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1,?\}$*, the ISR-communication complexity* $\text{ISR}_\rho(f)$ *is defined as the cost of the smallest randomized protocol, where Alice and Bob have access to samples from* DSBS(ρ)*, that computes* f *correctly on any input with probability at least* $2/3$*. In particular,*

$$\text{ISR}_\rho(f) \quad = \quad \min_{\substack{\Pi \\ \forall \mathbf{x},\mathbf{y} \in \{0,1\}^n \ s.t. \ f(\mathbf{x},\mathbf{y}) \neq \ ?: \\ \Pr[\Pi(\mathbf{x},\mathbf{y}) = f(\mathbf{x},\mathbf{y})] \geq 2/3}} \text{CC}(\Pi)$$

where the minimum is taken over all randomized protocols Π*, where Alice and Bob have access to samples from* DSBS(ρ)*.*

For ease of notation, we will often drop the subscript ρ *and denote* $\text{ISR}(f) \stackrel{\text{def}}{=} \text{ISR}_\rho(f)$*.*

We use the term ISR as abbreviation for "Imperfectly Shared Randomness" and ISR-CC for "ISR-Communication Complexity". To emphasize the contrast, we will use PSR and PSR-CC for the classical case of (perfectly) shared randomness. It is clear that if $\rho > \rho'$, then $\text{ISR}_\rho(f) \leq \text{ISR}_{\rho'}(f)$.

An extreme case of ISR is when $\rho = 0$. This corresponds to communication complexity with *private* randomness, denoted by $R_{\text{priv}}(f)$. Note that $\text{ISR}_\rho(f) \leq R_{\text{priv}}(f)$ for any $\rho > 0$. A theorem (due to Newman [New91]) shows that any communication protocol using public randomness can be simulated using only private randomness with an extra communication of additive $O(\log n)$ (both in the 1-way and 2-way models). We state the theorem here for the convenience of the reader.

Theorem 2.6 (Newman's theorem [New91])**.** *For any function* $f : \{0,1\}^n \times \{0,1\}^n \to \Omega$ *(any range* Ω*), the following hold,*

$$R_{\text{priv}}(f) \quad \leq \quad R(f) + O(\log n)$$
$$R_{\text{priv}}^{1\text{-}way}(f) \quad \leq \quad R^{1\text{-}way}(f) + O(\log n)$$

here, $R_{\text{priv}}(f)$ *is also* $\text{ISR}_0(f)$ *and* $R(f)$ *is also* $\text{ISR}_1(f)$*.*

2.3 Information Complexity

Information complexity[4] is an interactive analogue of *Shannon's information theory* [Sha48]. Informally, information complexity is defined as the minimum number of bits of *information* that the two parties have to reveal to each other, when the inputs $(\mathbf{x}, \mathbf{y}) \in \{0,1\}^n \times \{0,1\}^n$ are coming from the 'worst' possible distribution μ.

---

[4]we will often abbreviate "information complexity" by IC

Definition 2.7 ((Prior-Free) Interactive Information Complexity; [Bra12])**.** *For any $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, the* (prior-free) interactive information complexity *of f, denoted by* IC(f), *is defined as,*

$$\mathrm{IC}(f) \quad = \quad \inf_{\Pi} \ \sup_{\mu} \ I(X; \Pi|Y) + I(Y; \Pi|X)$$

where, the infimum is over all randomized protocols Π such that for all $\mathbf{x}, \mathbf{y} \in \{0,1\}^n$ such that $f(\mathbf{x}, \mathbf{y}) \neq ?$, $\Pr[\Pi(\mathbf{x}, \mathbf{y}) \neq f(\mathbf{x}, \mathbf{y})] \leq 1/3$ and the supremum is over all distributions $\mu(\mathbf{x}, \mathbf{y})$ over $\{0,1\}^n \times \{0,1\}^n$. [$I(A; B|C)$ is the mutual information between A and B conditioned on C]

We refer the reader to the survey by Weinstein [Wei15] for a more detailed understanding of the definitions and the role of information complexity in communication complexity.

A general question of interest is: what is the relationship between IC(f) and R(f)? It is straightforward to show R(f) $\geq$ IC(f). Upper bounding R(f) as a function of IC(f) has been investigated in several works including the work of Barak et al. [BBCR13]. The cleanest relation known is that R(f) $\leq 2^{O(\mathrm{IC}(f))}$ [Bra12]. Our first result, namely Theorem 1.1, shows that for permutation-invariant functions, R(f) is *not much larger* than IC(f).

2.4 Some Useful Communication Problems

Central to our proof techniques is a multi-parameter version of GAP-HAMMING-DISTANCE, which we define as follows.

Definition 2.8 (GAP-HAMMING-DISTANCE, $\mathrm{GHD}^n_{a,b,c,g}$, $\overline{\mathrm{GHD}}^n_{a,b,c,g}$)**.** *We define $\mathrm{GHD}^n_{a,b,c,g}$ as the following partial function,*

$$\mathrm{GHD}^n_{a,b,c,g}(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } |\mathbf{x}| = a, |\mathbf{y}| = b \text{ and } \Delta(\mathbf{x}, \mathbf{y}) \geq c + g \\ 0 & \text{if } |\mathbf{x}| = a, |\mathbf{y}| = b \text{ and } \Delta(\mathbf{x}, \mathbf{y}) \leq c - g \\ ? & \text{otherwise} \end{cases}$$

Additionally we define $\overline{\mathrm{GHD}}^n_{a,b,c,g}$ as the following partial function,

$$\overline{\mathrm{GHD}}^n_{a,b,c,g}(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } |\mathbf{x}| = a, |\mathbf{y}| = b \text{ and } \Delta(\mathbf{x}, \mathbf{y}) = c + g \\ 0 & \text{if } |\mathbf{x}| = a, |\mathbf{y}| = b \text{ and } \Delta(\mathbf{x}, \mathbf{y}) = c - g \\ ? & \text{otherwise} \end{cases}$$

Informally, computing $\mathrm{GHD}^n_{a,b,c,g}$ is equivalent to the following problem:

- Alice is given $\mathbf{x} \in \{0,1\}^n$ such that $|\mathbf{x}| = a$
- Bob is given $\mathbf{y} \in \{0,1\}^n$ such that $|\mathbf{y}| = b$
- They wish to distinguish between the cases $\Delta(\mathbf{x}, \mathbf{y}) \geq c + g$ and $\Delta(\mathbf{x}, \mathbf{y}) \leq c - g$.

In the $\overline{\mathrm{GHD}}^n_{a,b,c,g}$ problem, they wish to distinguish between the cases $\Delta(\mathbf{x}, \mathbf{y}) = c + g$ or $\Delta(\mathbf{x}, \mathbf{y}) = c - g$. Arguably, $\overline{\mathrm{GHD}}^n_{a,b,c,g}$ is an 'easier' problem than $\mathrm{GHD}^n_{a,b,c,g}$. However, in this work we will show that in fact $\mathrm{GHD}^n_{a,b,c,g}$ is not much harder than $\overline{\mathrm{GHD}}^n_{a,b,c,g}$.

We will use certain known lower bounds on the information complexity and one-way communication complexity of $\overline{\mathrm{GHD}}^n_{a,b,c,g}$ for some settings of the parameters. The two main settings of parameters that we will be using correspond to the problems of UNIQUE-DISJOINTNESS and SPARSE-INDEXING (a variant of the more well-known INDEXING problem).

2.4.1 IC lower bound for UNIQUE-DISJOINTNESS

Definition 2.9 (UNIQUE-DISJOINTNESS, UDISJ_t^n). UDISJ_t^n *is given by the following partial function,*

$$\text{UDISJ}_t^n(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } |\mathbf{x}| = t, |\mathbf{y}| = t \text{ and } |\mathbf{x} \wedge \mathbf{y}| = 1 \\ 0 & \text{if } |\mathbf{x}| = t, |\mathbf{y}| = t \text{ and } |\mathbf{x} \wedge \mathbf{y}| = 0 \\ ? & \text{otherwise} \end{cases}$$

Note that UDISJ_t^n is an instance of $\overline{\text{GHD}}_{t,t,2t-1,1}^n$.

Informally, UDISJ_t^n is the problem where the inputs $\mathbf{x}, \mathbf{y} \in \{0,1\}^n$ satisfy $|\mathbf{x}| = |\mathbf{y}| = t$ and Alice and Bob wish to decide whether $|\mathbf{x} \wedge \mathbf{y}| = 1$ or $|\mathbf{x} \wedge \mathbf{y}| = 0$ (promised that it is one of them is the case).

Lemma 2.10. *For all $n \in \mathbb{N}$, $\text{IC}(\text{UDISJ}_n^{3n}) = \Omega(n)$.*

Proof. Bar-Yossef et al. [BJKS04] proved that GENERAL-UNIQUE-DISJOINTNESS, that is, unique disjointness without restrictions on $|\mathbf{x}|$, $|\mathbf{y}|$ on inputs of length n, has information complexity $\Omega(n)$. We convert the general UNIQUE-DISJOINTNESS instance into an instance of UDISJ_n^{3n} by a simple padding argument as follows. Given an instance of general UNIQUE-DISJOINTNESS $(\mathbf{x}', \mathbf{y}') \in \{0,1\}^n \times \{0,1\}^n$. Alice constructs $\mathbf{x} = \mathbf{x}' \circ 1^{(n-|\mathbf{x}'|)} \circ 0^{(n+|\mathbf{x}'|)}$ and Bob constructs $\mathbf{y} = \mathbf{y}' \circ 0^{(n+|\mathbf{y}'|)} \circ 1^{(n-|\mathbf{y}'|)}$. Note that $|\mathbf{x} \wedge \mathbf{y}| = |\mathbf{x}' \wedge \mathbf{y}'|$. Also, we have that $\mathbf{x}, \mathbf{y} \in \{0,1\}^{3n}$, and $|\mathbf{x}| = |\mathbf{y}| = n$. Thus, we have reduced GENERAL-UNIQUE-DISJOINTNESS to UDISJ_n^{3n}, and thus the lower bound of [BJKS04] implies that $\text{IC}(\text{UDISJ}_n^{3n}) = \Omega(n)$. $\qquad \square$

On top of the above lemma, we apply a simple *padding* argument in addition to the above lower bound, to get a more general lower bound for UNIQUE-DISJOINTNESS as follows.

Proposition 2.11 (Unique Disjointness IC Lower Bound). *For all $t, w \in \mathbb{N}$,*

$$\text{IC}(\text{UDISJ}_t^{2t+w}) = \Omega(\min\{t, w\})$$

Proof. We look at two cases, namely $w \leq t$ and $w > t$.
Case 1. $[w \leq t]$: We have from Lemma 2.10 that $\text{IC}(\text{UDISJ}_w^{3w}) \geq \Omega(w)$. We map the instance $(\mathbf{x}', \mathbf{y}')$ of UDISJ_w^{3w} to an instance $(\mathbf{x}, \mathbf{y})$ of UDISJ_t^{2t+w} by the following reduction, $\mathbf{x} = \mathbf{x}' \circ 1^{(t-w)} \circ 0^{(t-w)}$ and $\mathbf{y} = \mathbf{y}' \circ 0^{(t-w)} \circ 1^{(t-w)}$. This implies that $\text{IC}(\text{UDISJ}_t^{2t+w}) = \Omega(w)$.

Case 2. $[w > t]$: We have from Lemma 2.10 that $\text{IC}(\text{UDISJ}_t^{3t}) = \Omega(t)$. As before, we map the instance $(\mathbf{x}', \mathbf{y}')$ of UDISJ_t^{3t} to an instance $(\mathbf{x}, \mathbf{y})$ of UDISJ_t^{2t+w} by the following reduction, $\mathbf{x} = \mathbf{x}' \circ 0^{(w-t)}$ and $\mathbf{y} = \mathbf{y}' \circ 0^{(w-t)}$. This implies that $\text{IC}(\text{UDISJ}_t^{2t+w}) = \Omega(t)$.

Combining the above two lower bounds, we get that $\text{IC}(\text{UDISJ}_t^{2t+w}) = \Omega(\min\{t, w\})$. $\qquad \square$

2.4.2 1-way CC lower bound for SPARSE-INDEXING

Definition 2.12 (SPARSE-INDEXING, SPARSEIND_t^n). SPARSEIND_t^n *is given by the following partial function,*

$$\text{SPARSEIND}_t^n(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } |\mathbf{x}| = t, |\mathbf{y}| = 1 \text{ and } |\mathbf{x} \wedge \mathbf{y}| = 1 \\ 0 & \text{if } |\mathbf{x}| = t, |\mathbf{y}| = 1 \text{ and } |\mathbf{x} \wedge \mathbf{y}| = 0 \\ ? & \text{otherwise} \end{cases}$$

Note that SPARSEIND_t^n is an instance of $\overline{\text{GHD}}_{t,1,t,1}^n$.

Informally, $\textsc{SparseInd}_t^n$ is the problem where the inputs $\mathbf{x}, \mathbf{y} \in \{0,1\}^n$ satisfy $|\mathbf{x}| = t$ and $|\mathbf{y}| = 1$ and Alice and Bob wish to decide whether $|\mathbf{x} \wedge \mathbf{y}| = 1$ or $|\mathbf{x} \wedge \mathbf{y}| = 0$ (promised that one of them is the case).

Lemma 2.13. *For all* $a \in \mathbb{N}$, $\mathrm{R}^{\text{1-way}}(\textsc{SparseInd}_n^{2n}) = \Omega(n)$.

Proof. Jayram et al. [JKS08] proved that if Alice is given $\mathbf{x} \in \{0,1\}^n$, Bob is given $i \in [n]$, and Bob needs to determine x_i upon receiving a single message from Alice, then Alice's message should consist of $\Omega(n)$ bits, even if they are allowed shared randomness. Using their result, we deduce that $\mathrm{R}^{\text{1-way}}(\textsc{SparseInd}_n^{2n}) = \Omega(n)$ via the following simple padding argument: Alice and Bob double the length of their strings from n to $2n$, with Alice's new input consisting of $(\mathbf{x}, \overline{\mathbf{x}})$ while Bob's new input consists of $(\mathbf{e}_i, 0)$, where $\overline{\mathbf{x}}$ is the bitwise complement of $\mathbf{x}$ and $\mathbf{e}_i$ is the indicator vector for location i. Note that the Hamming weight of Alice's new string is equal to n while its length is $2n$, as desired. $\qquad\square$

On top of the above lemma, we apply a simple *padding* argument in addition to the above lower bound, to get a more general lower bound for SPARSE-INDEXING as follows.

Proposition 2.14 (SPARSE-INDEXING 1-way CC Lower Bound)**.** *For all* $t, w \in \mathbb{N}$,

$$\mathrm{R}^{\text{1-way}}(\textsc{SparseInd}_t^{t+w}) = \Omega(\min\{t, w\})$$

Proof. We look at two cases, namely $w \leq t$ and $w > t$.
Case 1. $[w \leq t]$: We have from Lemma 2.13 that $\mathrm{R}^{\text{1-way}}(\textsc{SparseInd}_w^{2w}) \geq \Omega(w)$. We map the instance $(\mathbf{x}', \mathbf{y}')$ of $\textsc{SparseInd}_w^{2w}$ to an instance $(\mathbf{x}, \mathbf{y})$ of $\textsc{SparseInd}_t^{t+w}$ by the following reduction, $\mathbf{x} = \mathbf{x}' \circ 1^{(t-w)}$ and $\mathbf{y} = \mathbf{y}' \circ 0^{(t-w)}$. This implies that $\mathrm{R}^{\text{1-way}}(\textsc{SparseInd}_t^{t+w}) \geq \Omega(w)$.

Case 2. $[w > t]$: We have from Lemma 2.13 that $\mathrm{R}^{\text{1-way}}(\textsc{SparseInd}_t^{2t}) = \Omega(t)$. We map the instance $(\mathbf{x}', \mathbf{y}')$ of $\textsc{SparseInd}_t^{2t}$ to an instance $(\mathbf{x}, \mathbf{y})$ of $\textsc{SparseInd}_t^{t+w}$ by the following reduction, $\mathbf{x} = \mathbf{x}' \circ 0^{(w-t)}$ and $\mathbf{y} = \mathbf{y}' \circ 0^{(w-t)}$. This implies that $\mathrm{R}^{\text{1-way}}(\textsc{SparseInd}_t^{t+w}) \geq \Omega(t)$.

Combining the above two lower bounds, we get that $\mathrm{R}^{\text{1-way}}(\textsc{SparseInd}_t^{t+w}) \geq \Omega(\min\{t, w\})$. $\qquad\square$

3 Coarse Characterization of Information Complexity

In this section, we prove the first of our results, namely Theorem 1.1, which we restate below for convenience of the reader.

Theorem 1.1. *Let* $f : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1,?\}$ *be a (total or partial) permutation-invariant function. Then,*
$$m(f) \leq \mathrm{IC}(f) \leq \mathrm{R}(f) \leq \mathrm{poly}(m(f)) + O(\log n).$$
where $m(f)$ *is the combinatorial measure we define in Definition 3.2.*

3.1 Overview of proof

We construct a measure $m(f)$ such that $m(f) \leq \mathrm{IC}(f) \leq \mathrm{R}(f) \leq \tilde{O}(m(f)^4) + O(\log n)$. In order to do this, we look at the *slices* of f obtained by restricting $|\mathbf{x}|$ and $|\mathbf{y}|$. As in Observation 2.2, let $h_{a,b}(\Delta(\mathbf{x}, \mathbf{y}))$ be the restriction of h to $|\mathbf{x}| = a$ and $|\mathbf{y}| = b$. We define the notion of a *jump* in $h_{a,b}$ as follows.

Definition 3.1 (Jump in $h_{a,b}$). *(c,g) is a* jump *in $h_{a,b}$ if $h_{a,b}(c+g) \neq h_{a,b}(c-g)$, both $h_{a,b}(c+g)$, $h_{a,b}(c-g)$ are in $\{0,1\}$ and $h_{a,b}(r)$ is undefined for $c-g < r < c+g$.*

Thus, any protocol that computes f with low error will in particular be able to solve a GAP-HAMMING-DISTANCE problem $\overline{\text{GHD}}^n_{a,b,c,g}$ as in Definition 2.8. Thus, if (c,g) is a *jump* for $h_{a,b}$, then $\text{IC}(\overline{\text{GHD}}^n_{a,b,c,g})$ is a lower bound on $\text{IC}(f)$. We will prove lower bounds on $\text{IC}(\overline{\text{GHD}}^n_{a,b,c,g})$ for any value of a, b, c and g by obtaining a variety of reductions from UNIQUE-DISJOINTNESS, and then our measure $m(f)$ will be obtained by taking the largest of these lower bounds for $\text{IC}(\overline{\text{GHD}}^n_{a,b,c,g})$ over all choices of a and b and jumps (c,g) in $h_{a,b}$.

Suppose $m(f) = k$. We construct a randomized communication protocol with cost $\widetilde{O}(k^4)+O(\log n)$ that computes f correctly with low constant error. The protocol works as follows: First, Alice and Bob exchange the values $|\mathbf{x}|$ and $|\mathbf{y}|$, which requires $O(\log n)$ communication (say, $|\mathbf{x}| = a$ and $|\mathbf{y}| = b$). Now, all they need to figure out is the range in which $\Delta(\mathbf{x},\mathbf{y})$ lies (note that finding $\Delta(\mathbf{x},\mathbf{y})$ exactly can require $\Omega(n)$ communication!). Let $\mathscr{J}(h_{a,b}) = \{(c_1,g_1),(c_2,g_2),\cdots,(c_m,g_m)\}$ be the set of all *jumps* in $h_{a,b}$. Note that the intervals $[c_i - g_i, c_i + g_i]$ are all pairwise disjoint. To compute $h_{a,b}(\Delta(\mathbf{x},\mathbf{y}))$, it suffices for Alice and Bob need to *resolve* each jump, that is, for each $i \in [m]$, they need to figure out whether $\Delta(\mathbf{x},\mathbf{y}) \geq c+g$ or $\Delta(\mathbf{x},\mathbf{y}) \leq c-g$. We will show that any particular jump can be resolved with a constant probability of error using $O(k^3)$ communication, and the number of jumps m is at most $2^{O(k)}\log n$. Although the number of jumps is large, it suffices for Alice and Bob to do a binary search through the jumps, which will require them to resolve only $O(k\log\log n)$ jumps each requiring $\widetilde{O}(k^3)$ communication. Thus, the total communication cost will be $\widetilde{O}(k^4)+O(\log n)$.[5]

3.2 Proof of Theorem 1.1

As outlined earlier, we define the measure $m(f)$ as follows.

Definition 3.2 (Measure $m(f)$). *Given a permutation-invariant function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1,?\}$ and integers a, b, s.t. $0 \leq a,b \leq n$, let $h_{a,b} : \{0,\cdots,n\} \to \{0,1,?\}$ be the function given by $h_{a,b}(d) = f(\mathbf{x},\mathbf{y})$ if there exist $\mathbf{x},\mathbf{y}$ with $|\mathbf{x}| = a$, $|\mathbf{y}| = b$, $\Delta(\mathbf{x},\mathbf{y}) = d$ and ? otherwise. (Note. by permutation invariance of f, $h_{a,b}$ is well-defined.) Let $\mathscr{J}(h_{a,b})$ be the set of jumps in $h_{a,b}$, defined as follows,*

$$\mathscr{J}(h_{a,b}) \stackrel{\text{def}}{=} \left\{ (c,g) : \begin{array}{c} h_{a,b}(c-g),\ h_{a,b}(c+g) \in \{0,1\} \\ h_{a,b}(c-g) \neq h_{a,b}(c+g) \\ \forall i \in (c-g, c+g)\ :\ h_{a,b}(i) = ? \end{array} \right\}$$

Then, we define $m(f)$ as follows.

$$m(f) \stackrel{\text{def}}{=} \frac{1}{C} \cdot \max_{\substack{a,b\in[n] \\ (c,g)\in\mathscr{J}(h_{a,b})}} \max\left\{ \frac{\min\{a,b,c,n-a,n-b,n-c\}}{g}, \log\left(\frac{\min\{c,n-c\}}{g}\right) \right\}$$

where C is a suitably large constant (which does not depend on n).

We will need the following lemma to show that the measure $m(f)$ is a lower bound on $\text{IC}(f)$.

---

[5]We will need to resolve each jump correctly with error probability of at most $1/\Omega(k\log\log n)$. And for that we will actually require $O(k^3 \log k \log\log\log n)$ communication. So the total communication is really $O(k^4 \cdot \log k \cdot \log\log n \cdot \log\log\log n)$ which we write as $\widetilde{O}(k^4)+O(\log n)$ in short. See Remark 3.7.

Lemma 3.3. *For all n, a, b, c, g for which $\overline{\mathrm{GHD}}_{a,b,c,g}^n$ is a meaningful problem[6], the following lower bounds hold,*

$$\mathrm{IC}(\overline{\mathrm{GHD}}_{a,b,c,g}^n) \geq \frac{1}{C} \cdot \frac{\min\{a, b, c, n-a, n-b, n-c\}}{g}$$

$$\mathrm{IC}(\overline{\mathrm{GHD}}_{a,b,c,g}^n) \geq \frac{1}{C} \cdot \log\left(\frac{\min\{c, n-c\}}{g}\right)$$

Next, we obtain randomized communication protocols to solve $\mathrm{GHD}_{a,b,c,g}^n$.

Lemma 3.4. *Let $a \leq b \leq n/2$. Then, the following upper bounds hold*

$$\mathrm{R}(\mathrm{GHD}_{a,b,c,g}^n) = O\left(\left(\frac{a}{g}\right)^2 \left(\log\left(\frac{b}{a}\right) + \log\left(\frac{a}{g}\right)\right)\right)$$

$$\mathrm{R}(\mathrm{GHD}_{a,b,c,g}^n) = O\left(\min\left\{\left(\frac{c}{g}\right)^2, \left(\frac{n-c}{g}\right)^2\right\}\right)$$

We defer the proofs of the Lemmas 3.3 and 3.4 to Section 3.3. For now, we will use these lemmas to prove Theorem 1.1. First, we show that each *jump* can be resolved using $O(m(f)^3)$ communication.

Lemma 3.5. *Suppose $m(f) = k$, and let $|\mathbf{x}| = a$ and $|\mathbf{y}| = b$. Let (c, g) be any jump in $h_{a,b}$. Then the problem of $\mathrm{GHD}_{a,b,c,g}^n$, that is, deciding whether $\Delta(\mathbf{x}, \mathbf{y}) \geq c + g$ or $\Delta(\mathbf{x}, \mathbf{y}) \leq c - g$ can be solved, with a constant probability of error, using $O(k^3)$ communication.*

Proof. We can assume without loss of generality that $a \leq b \leq n/2$. This is because both Alice and Bob can flip their respective inputs to ensure $a, b \leq n/2$. Also, if $a > b$, then we can flip the role of Alice and Bob to get $a \leq b \leq n/2$.

Since $a \leq b \leq n/2$, from the definition of $m(f)$ in Lemma 3.3 we have that,

$$\min\left\{\frac{a}{g}, \frac{c}{g}, \frac{n-c}{g}\right\} \leq O(k)$$

We consider two cases as follows.

Case 1. $(c/g) \leq O(k)$ or $((n-c)/g) \leq O(k)$: In this case we have, from part (ii) of Lemma 3.4, a randomized protocol with cost $O\left(\min\{(c/g)^2, ((n-c)/g)^2\}\right) = O(k^2)$.

Case 2. $(a/g) \leq O(k)$: In this case, we will show that $\left((a/g)^2 (\log(b/a) + \log(a/g))\right) \leq O(k^3)$. Clearly, $(a/g)^2 \leq O(k^2)$ and $\log(a/g) \leq O(\log k)$. We will now show that in fact $\log(b/a) \leq O(k)$. From part (ii) of Lemma 3.3 we know that either $\log(c/g) \leq O(k)$ or $\log((n-c)/g) \leq O(k)$. Thus, it suffices to show that $(b/a) \leq O(\min\{(c/g), ((n-c)/g)\})$. We know that $b - a + g \leq c \leq b + a - g$. The left inequality gives $(b/a) \leq (c+a-g)/a = (c/g).(g/a) + 1 - (g/a) \leq O(c/g)$ (since $g/a \leq 1$). The right inequality gives $((n-b)/a) \leq ((n-c)/g).(g/a) + 1 - g/a \leq O((n-c)/g)$. Since $b \leq n/2$, we have $(b/a) \leq ((n-b)/a) \leq O((n-c)/g)$. $\square$

Next, we obtain an upper bound on $|\mathscr{J}(h_{a,b})|$, that is, the number of jumps in $h_{a,b}$.

Lemma 3.6. *For any function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ with $m(f) = k$, the number of jumps in $h_{a,b}$ is at most $2^{O(k)} \log n$.*

---

[6]that is, $a, b \leq n$ and $c + g$ and $c - g$ are achievable hamming distances when $|\mathbf{x}| = a$ and $|\mathbf{y}| = b$.

Proof. Let $\mathscr{J} = \{(c_1, g_1), \cdots, (c_m, g_m)\}$ be the set of all jumps in $h_{a,b}$. Partition $\mathscr{J}$ into $\mathscr{J}_1 \cup \mathscr{J}_2$, where $\mathscr{J}_1 = \{(c, g) \in \mathscr{J} : c \leq n/2\}$ and $\mathscr{J}_2 = \{(c, g) \in \mathscr{J} : c > n/2\}$. From second part in Lemma 3.3 we know the following:

$$\forall (c, g) \in \mathscr{J}_1 \quad : \quad \log\left(\frac{c}{g}\right) \leq O(k) \quad \text{that is,} \quad g \geq c 2^{-O(k)}$$

$$\forall (c, g) \in \mathscr{J}_2 \quad : \quad \log\left(\frac{n-c}{g}\right) \leq O(k) \quad \text{that is,} \quad g \geq (n-c) 2^{-O(k)}$$

Let $\mathscr{J}_1 = \{(c_1, g_1), \cdots, (c_p, g_p)\}$, where the c_i's are sorted in increasing order. We have that $c_i + g_i \leq c_{i+1} - g_{i+1}$ for all i. That is, we have that $c_i(1 + 2^{-O(k)}) \leq c_{i+1}(1 - 2^{-O(k)})$, which gives that $c_{i+1} \geq c_i(1 + 2^{-O(k)})$. Thus, $n/2 \geq c_p \geq c_1(1 + 2^{-O(k)})^p$, which gives that $|\mathscr{J}_1| = p \leq 2^{O(k)} \log n$. Similarly, by looking at $n - c_i$'s in $\mathscr{J}_2$, we get that $|\mathscr{J}_2| \leq 2^{O(k)} \log n$, and thus, $|\mathscr{J}| \leq 2^{O(k)} \log n$. $\quad\square$

We now complete the proof of Theorem 1.1.

Proof of Theorem 1.1. Any protocol to compute f also computes $\overline{\text{GHD}}^n_{a,b,c,g}$ for any a, b and any jump $(c, g) \in \mathscr{J}(h_{a,b})$. Consider the choice of a, b and a jump $(c, g) \in \mathscr{J}(h_{a,b})$ such that the lower bound obtained on $\text{IC}(\overline{\text{GHD}}^n_{a,b,c,g})$ through Lemma 3.3 is maximized, which by definition is equal to $m(f)$ (after choosing an appropriate C in the definition of $m(f)$). Thus, we have $m(f) \leq \text{IC}(\overline{\text{GHD}}^n_{a,b,c,g}) \leq \text{IC}(f)$.

We also have a protocol to solve f, which works as follows: First Alice and Bob exchange $|\mathbf{x}| = a$ and $|\mathbf{y}| = b$, requiring $O(\log n)$ communication. From Lemma 3.6, we know that the number of jumps in $h_{a,b}$ is at most $2^{O(m(f))} \log n$, and so Alice and Bob need to do a binary search through the jumps, resolving only $O(m(f) \log \log n)$ jumps, each with an error probability of at most $O(1/m(f) \log \log n)$. This can be done using $\widetilde{O}(m(f)^3)$ communication[7] (using Lemma 3.5). Thus, the total amount of communication is $\text{R}(f) \leq \widetilde{O}(m(f)^4) + O(\log n)$[8]. All together, we have shown that,

$$m(f) \leq \text{IC}(f) \leq \text{R}(f) \leq \widetilde{O}(m(f)^4) + O(\log n)$$

$\square$

Remark 3.7. *It is a valid concern that we are hiding a $\log \log n$ factor in the $\widetilde{O}(m(f)^4)$ term. But this is fine because of the following reason: If $m(f) \leq (\log n)^{1/5}$, then the $O(\log n)$ term is the dominating term and thus the overall communication is $O(\log n)$. And if $m(f) \geq (\log n)^{1/5}$, then $\log m(f) = \Omega(\log \log n)$, in which case the $\widetilde{O}(m(f)^4)$ is truly hiding only $\text{poly} \log m(f)$ factors.*

In the following section, we prove the main technical lemmas used, namely Lemmas 3.3 and 3.4.

3.3 Lower and upper bounds on Gap Hamming Distance

In this section, we prove lower bounds on information complexity (Lemma 3.3), and upper bounds on the randomized communication complexity (Lemma 3.4) of GAP-HAMMING-DISTANCE.

Lower bounds

We will prove Lemma 3.3 by getting certain reductions from UNIQUE-DISJOINTNESS (namely Proposition 2.11). In order to do so, we first prove lower bounds on information complexity of two problems, namely, SET-INCLUSION (Definition 3.8) and SPARSE-INDEXING (Definition 2.12). We do this by obtaining reductions from UNIQUE-DISJOINTNESS.

---

[7] Here, $\widetilde{O}(m(f)^3) = O(m(f)^3 \log m(f) \log \log \log n)$.
[8] Here $\widetilde{O}(m(f)^4) = O(m(f)^4 \log m(f) \log \log n \log \log \log n)$

Definition 3.8 ($\text{SETINC}^n_{p,q}$). *Let $p \leq q \leq n$. Alice is given $\mathbf{x} \in \{0,1\}^n$ such that $|\mathbf{x}| = p$ and Bob is given $\mathbf{y} \in \{0,1\}^n$ such that $|\mathbf{y}| = q$ and they wish to distinguish between the cases $|\mathbf{x} \wedge \mathbf{y}| = p$ and $|\mathbf{x} \wedge \mathbf{y}| = p - 1$. Note that $\text{SETINC}^n_{p,q}$ is the same as $\overline{\text{GHD}}^n_{p,q,q-p+1,1}$.*

Proposition 3.9 (SET-INCLUSION lower bound). *$\forall\, t, w \in \mathbb{N}$, $\text{IC}(\text{SETINC}^{2t+w}_{t,t+w}) \geq \Omega(\min(t, w))$*

Proof. We know that $\text{IC}(\text{UDISJ}^{2t+w}_t) \geq \Omega(\min(t,w))$ from Proposition 2.11. Note that UDISJ^{2t+w}_t is same as the problem of $\overline{\text{GHD}}^{2t+w}_{t,t,2t-1,1}$. If we instead think of Bob's input as complemented, we get that solving $\overline{\text{GHD}}^{2t+w}_{t,t,2t-1,1}$ is equivalent to solving $\overline{\text{GHD}}^{2t+w}_{t,t+w,w+1,1}$, which is same as $\text{SETINC}^{2t+w}_{t,t+w}$. Thus, we conclude that $\text{IC}(\text{SETINC}^n_{t,t+w}) \geq \Omega(\min(t,w))$. $\square$

Proposition 3.10 (SPARSE-INDEXING lower bound). *$\forall\, t \in \mathbb{N}$, $\text{IC}(\text{SPARSEIND}^{2^{t+1}}_{2^t}) \geq \Omega(t)$*

Proof. We know that $\text{IC}(\text{UDISJ}^{t+1}_{t/3}) \geq \Omega(t)$ from Proposition 2.11. Recall that $\text{SPARSEIND}^{2^{t+1}}_{2^t}$ is an instance of $\overline{\text{GHD}}^{2^{t+1}}_{2^t,1,2^t,1}$. Alice uses $\mathbf{x}$ to obtain the Hadamard code of $\mathbf{x}$, which is $\mathbf{X} \in \{0,1\}^{2^{t+1}}$ such that $\mathbf{X}(\mathbf{a}) = \mathbf{a} \cdot \mathbf{x}$ (for $\mathbf{a} \in \{0,1\}^{t+1}$). On the other hand, Bob uses $\mathbf{y}$ to obtain the indicator vector of $\mathbf{y}$, which is $\mathbf{Y} \in \{0,1\}^{2^{t+1}}$ such that $\mathbf{Y}(\mathbf{a}) = 1$ iff $\mathbf{a} = \mathbf{y}$. Clearly $|\mathbf{Y}| = 1$. Observe that, $|\mathbf{X}| = 2^t$ and $\mathbf{X}(\mathbf{y}) = \mathbf{y} \cdot \mathbf{x}$ is 1 if $|\mathbf{x} \wedge \mathbf{y}| = 1$ and 0 if $|\mathbf{x} \wedge \mathbf{y}| = 0$. Thus, we end up with an instance of $\text{SPARSEIND}^{2^{t+1}}_{2^t}$. Hence $\text{IC}(\text{SPARSEIND}^{2^{t+1}}_{2^t}) \geq \text{IC}(\text{UDISJ}^{t+1}_{t/3}) \geq \Omega(t)$. $\square$

We now state and prove a technical lemma that will help us prove Lemma 3.3.

Lemma 3.11. *Let $a \leq b \leq n/2$. Then, the following lower bounds hold,*

(i) $\text{IC}(\overline{\text{GHD}}^n_{a,b,c,g}) \geq \Omega\left(\min\left\{\frac{c-b+a}{g}, \frac{n-c}{g}\right\}\right)$

(ii) $\text{IC}(\overline{\text{GHD}}^n_{a,b,c,g}) \geq \Omega\left(\min\left\{\frac{a+b-c}{g}, \frac{c+b-a}{g}\right\}\right)$

(iii) $\text{IC}(\overline{\text{GHD}}^n_{a,b,c,g}) \geq \Omega\left(\min\left\{\log\left(\frac{c}{g}\right), \log\left(\frac{n-c}{g}\right)\right\}\right)$

Proof. We prove the three parts of the above lemma using reductions from UDISJ, SETINC and SPARSEIND respectively. Note that once we fix $|\mathbf{x}| = a$ and $|\mathbf{y}| = b$, a jump (c, g) is meaningful only if $b - a + g \leq c \leq b + a - g$ (since $b - a \leq \Delta(\mathbf{x}, \mathbf{y}) \leq b + a$). We will assume that $c \equiv b + a - g \pmod 2$, so that $c + g$ and $c - g$ will be achievable Hamming distances.

Proof of (i). We obtain a reduction from UDISJ^{3t}_t for which we know from Proposition 2.11 that $\text{IC}(\text{UDISJ}^{3t}_t) \geq \Omega(t)$. Recall that UDISJ^{3t}_t is same as $\overline{\text{GHD}}^{3t}_{t,t,2t-1,1}$. Given any instance of $\overline{\text{GHD}}^{3t}_{t,t,2t-1,1}$, we first repeat the instance g times to get an instance of $\overline{\text{GHD}}^{3gt}_{gt,gt,g(2t-1),g}$. Now, we need to append $(a - gt)$ 1's to $\mathbf{x}$ and $(b - gt)$ 1's to $\mathbf{y}$. This will increase the Hamming distance by a fixed amount which is at least $(b - a)$ and at most $(b + a - 2gt)$. Also, the number of inputs we need to add is at least $((a - gt) + (b - gt) + (c - g(2t - 1)))/2$ [9]. Thus, we can get a reduction to $\overline{\text{GHD}}^n_{a,b,c,g}$ if and only if,

$$(b - a) \leq c - g(2t - 1) \leq b + a - 2gt$$

$$n \geq 3gt + \frac{(a - gt) + (b - gt) + (c - g(2t - 1))}{2}$$

---

[9] We will be repeatedly using this idea in several proofs. The reason we obtain the said constraints is as follows: Suppose Alice has to add A 1's to her input and Bob has to add B 1's to his input. Then the hamming distance increases by an amount C such that $|A - B| \leq C \leq A + B$. Also, the minimum number of coordinates that need to be added to achieve this is at least $(A + B + C)/2$

The constraints on c give us that $2gt \le c - (b - a) + g$ and $c \le b + a - g$ (recall that this is always true). The constraint on n gives that $gt \le n - (a + b + c + g)/2$, which is equivalent to

$$t \le \frac{n - a - b}{2g} + \frac{n - c - g}{2g}$$

Thus, we can have the reduction work by choosing t to be

$$t = \min\left\{\frac{c - b + a + g}{2g}, \frac{n - c - g}{2g}\right\} = \Omega\left(\min\left\{\frac{c - b + a}{g}, \frac{n - c}{g}\right\}\right)$$

(since $n \ge a + b$) and thus we obtain

$$\text{IC}(\overline{\text{GHD}}^n_{a,b,c,g}) \ge \text{IC}(\text{UDISJ}^{3t}_t) \ge \Omega\left(\min\left\{\frac{c - b + a}{g}, \frac{n - c}{g}\right\}\right)$$

Proof of (ii). We obtain a reduction from $\text{SETINC}^m_{t,w}$ (where $m = 2t + w$) for which we know from Proposition 3.9 that $\text{IC}(\text{SETINC}^m_{t,w}) \ge \Omega(\min\{t, w\})$. Recall that $\text{SETINC}^m_{t,w}$ is same as $\overline{\text{GHD}}^m_{t,t+w,w+1,1}$. Given an instance of $\overline{\text{GHD}}^m_{t,t+w,w+1,1}$, we first repeat the instance g times to get an instance of $\overline{\text{GHD}}^{gm}_{gt,gt+gw,gw+g,g}$. Now, we need to append $(a - gt)$ 1's to $\mathbf{x}$ and $(b - gt - gw)$ 1's to $\mathbf{y}$. This will increase the Hamming distance by a fixed amount which is at least $|b - a - gw|$ and at most $(b - gt - gw) + (a - gt)$. Also, the number of inputs we need to add is at least $((a - gt) + (b - gt - gw) + (c - g(w + 1)))/2$. Thus, we can get a reduction to $\overline{\text{GHD}}^n_{a,b,c,g}$ if and only if,

$$|b - a - gw| \le c - g(w + 1) \le b + a - 2gt - gw$$
$$n \ge 2gt + gw + \frac{(b - gt - gw) + (a - gt) + (c - gw - g)}{2}$$

The left constraint on c requires $c \ge \max\{b - a + g, 2gw - (b - a) + g\}$. We know that $c \ge b - a + g$, so the only real constraint is $c \ge 2gw - (b - a) + g$, which gives us that,

$$w \le \frac{c + b - a - g}{2g}$$

The right constraint on c requires $c \le b + a - 2gt + g$, which gives us that,

$$t \le \frac{a + b - c + g}{2g}$$

Suppose we choose $t = \frac{a+b-c+g}{2g}$. Then the constraint on n is giving us that,

$$n \ge gt + \frac{a + b + c - g}{2} = \frac{a + b - c + g}{2} + \frac{a + b + c - g}{2} = a + b$$

We already assumed that $a \le b \le n/2$, and hence this is always true.

Thus, we choose $t = \frac{a+b-c+g}{2g}$ and $w = \frac{c+b-a-g}{2g}$, and invoking Proposition 3.9, we get,

$$\text{IC}(\overline{\text{GHD}}^n_{a,b,c,g}) \ge \text{IC}(\text{SETINC}^{2t+w}_{t,w}) \ge \min(\{t, w\}) \ge \Omega\left(\min\left\{\frac{a + b - c}{g}, \frac{c + b - a}{g}\right\}\right)$$

Proof of (iii). We obtain a reduction from $\text{SPARSEIND}^{2^{t+1}}_{2^t}$ for which we know from Proposition 3.10 that $\text{IC}(\text{SPARSEIND}^{2^{t+1}}_{2^t}) \ge \Omega(t)$. Recall that $\text{SPARSEIND}^{2^{t+1}}_{2^t}$ is same as $\overline{\text{GHD}}^{2^{t+1}}_{2^t,1,2^t,1}$, which is equivalent to $\overline{\text{GHD}}^{2^{t+1}}_{1,2^t,2^t,1}$ (if we flip roles of Alice and Bob). Given an instance of $\overline{\text{GHD}}^{2^{t+1}}_{1,2^t,2^t,1}$, we first

repeat the instance g times to get an instance of $\overline{\text{GHD}}^{g2^{t+1}}_{g,g2^t,g2^t,g}$. Now, we need to append $(a-g)$ 1's to $\mathbf{x}$ and $(b-g2^t)$ 1's to $\mathbf{y}$. This will increase the Hamming distance by a fixed amount which is at least $|b-g2^t-a+g|$ and at most $(b-g2^t+a-g)$. Also, the number of inputs we need to add is at least $((a-g)+(b-g2^t)+(c-g2^t))/2$. Thus, we can get a reduction to $\overline{\text{GHD}}^{n}_{a,b,c,g}$ if and only if,

$$|b-g2^t-a+g| \le c-g2^t \le (b-g2^t+a-g)$$

$$n \ge g2^{t+1} + \frac{(a-g)+(b-g2^t)+(c-g2^t)}{2}$$

The left constraint on c requires $c \ge \max\{b-a+g, 2g2^t-b+a-g\}$. Since $c \ge b-a+g$ anyway, this only requires $2g \cdot 2^t \le c+b-a+g$. The right constraint on c requires $c \le b+a-g$ which is also true anyway. The constraint on n is equivalent to,

$$g2^t \le n - \frac{a+b+c-g}{2} = \frac{n-a-b}{2} + \frac{n-c-g}{2}$$

Thus, we choose t such that,

$$t = \min\left\{\log_2\left(\frac{c+b-a+g}{2g}\right), \log_2\left(\frac{n-c-g}{2g}\right)\right\} \ge \Omega\left(\min\frac{c}{g}, \frac{n-c}{g}\right)$$

and invoking Proposition 3.10, we get,

$$\text{IC}(\overline{\text{GHD}}^{n}_{a,b,c,g}) \ge \text{IC}(\text{SPARSEIND}^{2^{t+1}}_{2^t}) \ge \Omega(t) \ge \Omega\left(\min\left\{\log_2\left(\frac{c}{g}\right), \log_2\left(\frac{n-c}{g}\right)\right\}\right)$$

$\square$

We are now finally able to prove Lemma 3.3.

Proof of Lemma 3.3. Assume for now that $a \le b \le n/2$. From parts (i) and (ii) of Lemma 3.11, we know the following,

$$\text{IC}(\overline{\text{GHD}}^{n}_{a,b,c,g}) \ge \Omega\left(\min\left\{\frac{c-b+a}{g}, \frac{n-c}{g}\right\}\right)$$

$$\text{IC}(\overline{\text{GHD}}^{n}_{a,b,c,g}) \ge \Omega\left(\min\left\{\frac{a+b-c}{g}, \frac{c+b-a}{g}\right\}\right)$$

Adding these up, we get that,

$$\text{IC}(\overline{\text{GHD}}^{n}_{a,b,c,g}) \ge \Omega\left(\min\left\{\frac{c-b+a}{g}, \frac{n-c}{g}\right\} + \min\left\{\frac{a+b-c}{g}, \frac{c+b-a}{g}\right\}\right)$$

Since $\min\{A,B\} + \min\{C,D\} = \min\{A+C, A+D, B+C, B+D\}$, we get that,

$$\text{IC}(\overline{\text{GHD}}^{n}_{a,b,c,g}) \ge \Omega\left(\min\left\{\frac{2a}{g}, \frac{2c}{g}, \frac{n+a+b-2c}{g}, \frac{n+b-a}{g}\right\}\right)$$

For the last two terms, note that, $n+a+b-2c \ge n-c$ (since $a+b \ge c$) and $n+b-a \ge n$ (since $b \ge a$). Thus, overall we get,

$$\text{IC}(\overline{\text{GHD}}^{n}_{a,b,c,g}) \ge \Omega\left(\min\left\{\frac{a}{g}, \frac{c}{g}, \frac{n-c}{g}\right\}\right)$$

Note that this was assuming $a \le b \le n/2$. In general, we get,

$$\text{IC}(\overline{\text{GHD}}^{n}_{a,b,c,g}) \ge \Omega\left(\frac{\min\{a,b,c,n-a,n-b,n-c\}}{g}\right)$$

17

[We get b, $(n-b)$, $(n-a)$ terms because we could have flipped inputs of either or both of Alice and Bob. Moreover, to get $a \leq b$, we might have flipped the role of Alice and Bob.]

The second lower bound of $\mathrm{IC}(\overline{\mathrm{GHD}}_{a,b,c,g}^{n}) \geq \Omega\left(\log\left(\frac{\min\{c,\, n-c\}}{g}\right)\right)$ follows immediately from part (iii) of Lemma 3.11.
We choose C to be a large enough constant, so that the desired lower bounds hold. $\qquad\square$

Upper bounds

We will now prove Lemma 3.4.

Proof of Lemma 3.4. We use different protocols to prove the two parts of the lemma.
Proof of Part 1. The main idea is similar to that of Proposition 5.7 of [CGMS14], except that we first hash into a small number of buckets. The details are as follows.

Alice and Bob have an instance $(\mathbf{x}, \mathbf{y})$ of $\mathrm{GHD}_{a,b,c,g}^{n}$, that is, Alice has $\mathbf{x} \in \{0,1\}^{n}$ such that $|\mathbf{x}| = a$, and Bob has $\mathbf{y} \in \{0,1\}^{n}$ such that $|\mathbf{y}| = b$, and they wish to distinguish between the cases $\Delta(\mathbf{x}, \mathbf{y}) \geq c + g$ and $\Delta(\mathbf{x}, \mathbf{y}) \leq c - g$.

Bob defines $\widetilde{\mathbf{y}} \in \{1, -1\}^{n}$ such that $\widetilde{y}_i = 1 - 2y_i$ for every $i \in [n]$. Then, the number of -1 coordinates in $\widetilde{\mathbf{y}}$ is exactly equal to b. It is easy to see that $\langle \mathbf{x}, \widetilde{\mathbf{y}} \rangle = (\Delta(\mathbf{x}, \mathbf{y}) - b)$, and hence computing $\mathrm{GHD}_{a,b,c,g}^{n}(\mathbf{x}, \mathbf{y})$ is equivalent to distinguishing between the cases $\langle \mathbf{x}, \widetilde{\mathbf{y}} \rangle \geq \alpha a$ and $\langle \mathbf{x}, \widetilde{\mathbf{y}} \rangle \leq \beta a$, where $\alpha \overset{\mathrm{def}}{=} (c - b + g)/a$ and $\beta \overset{\mathrm{def}}{=} (c - b - g)/a$. Note that $\alpha, \beta \in [-1, +1]$.

Alice and Bob use their shared randomness to generate a uniformly random hash function $h : [n] \to [B]$ where $B \overset{\mathrm{def}}{=} 100(a + b)(a/g)^2$. Basically, each coordinate $i \in [n]$ is mapped to one of the B 'buckets' uniformly and independently at random. Let $\mathrm{supp}(\mathbf{x}) := \{i \in [n] : x_i = 1\}$. We say that a coordinate $i \in \mathrm{supp}(\mathbf{x})$ is *bad* if there is a coordinate $j \in [n]$, $j \neq i$ such that $h(j) = h(i)$ and at least one of $x_j = 1$ or $y_j = 1$. For any $i \in \mathrm{supp}(\mathbf{x})$, the probability that i is *bad* is at most $\left(1 - (1 - 1/B)^{(a+b)}\right) \leq (a + b)/B = g^2/100a^2$. Thus, the expected number of *bad* coordinates is at most $(g^2/100a)$, and hence by Markov's inequality, we have that with probability at least $1 - g/(10a)$, there are at least $a(1 - g/(10a))$ coordinates in $\mathrm{supp}(x)$ that are not *bad*. Suppose we have chosen an h such that at least $a(1 - g/(10a))$ coordinates in $\mathrm{supp}(\mathbf{x})$ that are not *bad*. Let $\ell \overset{\mathrm{def}}{=} (2B/a)\ln(20a/g)$ and consider the following atomic protocol $\tau_h(\mathbf{x}, \mathbf{y})$:

- Alice and Bob use shared randomness to sample a sequence of ℓ indices $b_1, \ldots, b_\ell \in_R [B]$.

- Alice picks the smallest index $j \in [\ell]$ such that $h^{-1}(b_j) \cap \mathrm{supp}(\mathbf{x})$ is non-empty and sends j to Bob. If there is no such j, then the protocol aborts and outputs $+1$.

- Bob outputs $\prod\limits_{i \in h^{-1}(b_j)} \widetilde{y}_i$.

We first show that the difference in the probability of $\tau_h(\mathbf{x}, \mathbf{y})$ outputting $+1$, in the two cases $\langle \mathbf{x}, \widetilde{\mathbf{y}} \rangle \geq \alpha a$ and $\langle \mathbf{x}, \widetilde{\mathbf{y}} \rangle \leq \beta a$, is at least $\Omega(g/a)$. In particular, we will show the following,

$$\Pr[\tau_h(\mathbf{x}, \mathbf{y}) = +1 \mid \langle \mathbf{x}, \widetilde{\mathbf{y}} \rangle \geq \alpha a] \geq 1/2 + \alpha/2 - 3g/20a \tag{1}$$
$$\Pr[\tau_h(\mathbf{x}, \mathbf{y}) = +1 \mid \langle \mathbf{x}, \widetilde{\mathbf{y}} \rangle \leq \beta a] \leq 1/2 + \beta/2 + 3g/20a \tag{2}$$

Before we prove Inequalities 1 and 2, we first show how to use these to obtain our desired protocol. We observe that the difference in the two probabilities is at least $(\alpha - \beta)/2 - 3g/10a \geq \Omega(g/a)$. We repeat the above atomic procedure T times and declare the input to satisfy $\langle \mathbf{x}, \widetilde{\mathbf{y}} \rangle \geq \alpha a$, if the number of times the output is $+1$, is at least $((1 + \alpha + \beta)/2)T$, and $\langle \mathbf{x}, \widetilde{\mathbf{y}} \rangle \leq \beta a$ otherwise. A Chernoff bound implies that we will have the correct value the given $\mathrm{GHD}_{a,b,c,g}^{n}$ instance with

18

probability at least $1 - e^{-\Omega(g^2 T/a^2)}$. Setting $T = \Theta((a/g)^2)$ gives us that our protocol gives the right answer with probability at least 3/4. And the overall communication complexity of this protocol is

$$O((a/g)^2 \log_2(\ell)) = O((a/g)^2 \log_2(2B/a \log_2(10a/g))) = O((a/g)^2 (\log(b/a) + \log(a/g)))$$

We now prove Inequalities 1 and 2. Note that there are three possible situations that can arise in $\tau_h(\mathbf{x}, \mathbf{y})$,

1. the protocol aborts (that is, for all $j \in [\ell]$, $h^{-1}(b_j) \cap \text{supp}(\mathbf{x}) = \emptyset$)

2. the index j picked by Alice is such that $|h^{-1}(b_j) \cap \text{supp}(\mathbf{x})| > 1$

3. the index j picked by Alice is such that $|h^{-1}(b_j) \cap \text{supp}(\mathbf{x})| = 1$

We will say that an index $b \in [B]$ is 'good' if $|h^{-1}(b) \cap \text{supp}(\mathbf{x})| = 1$, 'bad' if $|h^{-1}(b) \cap \text{supp}(\mathbf{x})| > 1$, and 'empty' if $|h^{-1}(b) \cap \text{supp}(\mathbf{x})| = 0$

For Inequality 1, we have that $\langle \mathbf{x}, \widetilde{\mathbf{y}} \rangle \geq \alpha a$, and we wish to lower bound the probability that the protocol $\tau_h(\mathbf{x}, \mathbf{y})$ outputs $+1$. Notice that, when the protocol aborts, it always outputs $+1$. And conditioned on not aborting, the probability that we are in situation (3) and not (2), is at least $(1 - g/10a)$. This is because the number of non-'empty' b's is at most a, but the number of 'good' b's is at least $a(1 - g/10a)$. Thus, overall, we get that,

$$
\begin{aligned}
\Pr[\tau_h(\mathbf{x}, \mathbf{y}) = +1 \mid \langle \mathbf{x}, \widetilde{\mathbf{y}} \rangle \geq \alpha a] &\geq \left(1 - \frac{g}{10a}\right)\left(\frac{1}{2} + \frac{\alpha}{2}\right) \\
&\geq \left(\frac{1}{2} + \frac{\alpha}{2} - \frac{g}{10a}\right) \quad [\because \alpha \leq 1] \\
&\geq \left(\frac{1}{2} + \frac{\alpha}{2} - \frac{3g}{20a}\right)
\end{aligned}
$$

For Inequality 2, we have that $\langle \mathbf{x}, \widetilde{\mathbf{y}} \rangle \leq \beta a$, and we wish to upper bound the probability that the protocol $\tau_h(\mathbf{x}, \mathbf{y})$ outputs $+1$. The probability that we are in situation (1) is at most $\left(1 - \frac{(a - g/10)}{B}\right)^\ell \leq \left(1 - \frac{a}{2B}\right)^\ell \leq e^{-a\ell/2B} = g/20a$. Conditioned on not aborting, the probability that we are in situation (3) and not (2), is at least $(1 - g/10a)$ as before. Thus overall, we get that,

$$
\begin{aligned}
\Pr[\tau_h(\mathbf{x}, \mathbf{y}) = +1 \mid \langle \mathbf{x}, \widetilde{\mathbf{y}} \rangle \leq \beta a] &\leq \left(\frac{g}{20a}\right) + \left(\frac{g}{10a}\right) + \left(1 - \frac{g}{10a}\right)\left(\frac{1}{2} + \frac{\beta}{2}\right) \\
&\leq \left(\frac{1}{2} + \frac{\beta}{2} + \frac{3g}{10a}\right)
\end{aligned}
$$

Proof of Part 2. Kushilevitz, Ostrovsky and Rabani [KOR00] gave a protocol for distinguishing between the cases $\Delta(\mathbf{x}, \mathbf{y}) \geq c + g$ and $\Delta(\mathbf{x}, \mathbf{y}) \leq c - g$ using $O((c/g)^2)$ communication (without requiring any knowledge about $|\mathbf{x}|$ and $|\mathbf{y}|$). The upper bound of $O(((n - c)/g)^2)$ follows by Alice flipping her input so that the task is of distinguishing between the cases $\Delta(\mathbf{x}, \mathbf{y}) \geq n - c + g$ and $\Delta(\mathbf{x}, \mathbf{y}) \leq n - c - g$, and then the upper bound of [KOR00] applies again. $\qquad \square$

4 Communication with Imperfectly Shared Randomness

In this section, we prove our second result, namely Theorem 1.3, which we restate below for convenience of the reader.

Theorem 1.3. *Let $f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$ be a permutation-invariant total function. Then, we have*

$$\text{ISR}_\rho(f) \leq \text{poly}(\text{R}(f)) + O(\log \log n)$$

Furthermore, $\text{ISR}_\rho^{1\text{-way}}(f) \leq \text{poly}(\text{R}^{1\text{-way}}(f)) + O(\log \log n)$.

The outline of this section is as follows: In Section 4.1, we prove upper bounds on the ISR-CC of two basic problems: SMALL-SET-INTERSECTION (in Section 4.1.1) and SMALL-HAMMING-DISTANCE (in Section 4.1.2). As an aside, in Section 4.1.3, we introduce a new class of functions, called *strongly permutation-invariant functions*, which is a generalization of both SET-INTERSECTION and HAMMING-DISTANCE, and show that for every strongly permutation-invariant function f there is a polynomial relationship between R(f) and ISR(f) (Theorem 4.12). In Section 4.2 we give an overview of the proof for Theorem 1.3. The proof of the 2-way part of Theorem 1.3 appears in Section 4.3, and that of the 1-way part appears in Section 4.4. In Section 4.5, we prove a technical lemma needed in the proof of the 1-way part of Theorem 1.3.

4.1 ISR Protocols for Basic Problems

In this section, we prove that ISR-CC and PSR-CC are polynomially related for some specific functions (note that this is stronger than Theorem 1.3, in the sense that the additive $O(\log\log n)$ factor is not required). In particular, we give ISR protocols for two basic problems: SMALL-SET-INTERSECTION (in Section 4.1.1) and SMALL-HAMMING-DISTANCE (in Section 4.1.2), such that the communication costs of these protocols are polynomially related to the respective PSR-CC of these functions. Our motivation in doing so is two-fold: firstly, to give techniques for designing efficient ISR protocols, and secondly because these protocols are at the heart of our proof of Theorem 1.3. In addition to these, we also give ISR-protocols for the class of *strongly permutation-invariant functions* which we describe in Section 4.1.3.

4.1.1 Small Set Intersection

The SMALL-SET-INTERSECTION problem is defined as follows.

Definition 4.1 (SMALL-SET-INTERSECTION). $\text{SSI}_{a,b}^n : \{0,1\}^n \times \{0,1\}^n \to \mathbb{Z} \cup \{?\}$ *is defined as follows,*

$$\text{SSI}_{a,b}^n(\mathbf{x}, \mathbf{y}) = \begin{cases} |\mathbf{x} \wedge \mathbf{y}| & \text{if } |\mathbf{x}| = a, |\mathbf{y}| = b \\ ? & \text{otherwise} \end{cases}$$

Essentially, Alice is given $\mathbf{x} \in \{0,1\}^n$ such that $|\mathbf{x}| = a$, Bob is given $\mathbf{y} \in \{0,1\}^n$ such that $|\mathbf{y}| = b$, and they wish to compute $|\mathbf{x} \wedge \mathbf{y}|$.

The next two lemmas show that ISR-CC and PSR-CC are polynomially related for HD_k^n (for 1-way and 2-way models respectively).

Lemma 4.2 (1-way ISR Protocol for $\text{SSI}_{a,b}^n$). *Let $n, a, b \in \mathbb{N}$, such that, $a, b \le n/2$. Then,*

$$\Omega(\max\{a, \log b\}) \le \text{R}^{1\text{-}way}(\text{SSI}_{a,b}^n) \le \text{ISR}_\rho^{1\text{-}way}(\text{SSI}_{a,b}^n) = O(a\log(ab))$$

Proof. We first describe the 1-way ISR-protocol for $\text{SSI}_{a,b}^n$.

Let $\mathbf{x}$ be Alice's string and $\mathbf{y}$ be Bob's string, with $a = |\mathbf{x}|$ and $b = |\mathbf{y}|$. First, Alice and Bob use their correlated randomness to sample hash functions $h_A, h_B : [n] \to \{0,1\}^r$ such that for any i, $h_A(i)$ and $h_B(i)$ are ρ-correlated strings, but for $i \ne j$, $h_A(i)$ and $h_B(j)$ are independent. Now, Alice sends $\{h_A(i) : x_i = 1\}$, which Bob sees as $h_1, h_2, \ldots, h_a$. Then, Bob computes the size of the set $\{j \in \text{supp}(\mathbf{y}) : \Delta(h_B(j), h_i) \le \frac{1}{2} - \frac{\rho}{4} \text{ for some } i \in [a]\}$ and outputs it.

By the Chernoff bound, we have that for any i, $\Delta(h_A(i), h_B(i)) \le \frac{1}{2} - \frac{\rho}{4}$ with probability $1 - 2^{-\Omega(r)}$. Also, for any $i \ne j$, $\Delta(h_A(i), h_B(j)) \ge \frac{1}{2} - \frac{\rho}{4}$ with probability $1 - 2^{-\Omega(r)}$. Thus, the probability that for every i such that $x_i = y_i = 1$, $\Delta(h_A(i), h_B(i)) \le \frac{1}{2} - \frac{\rho}{4}$ and for every $i \ne j$ such that $x_i = y_j = 1$, $\Delta(h_A(i), h_B(i)) \ge \frac{1}{2} - \frac{\rho}{4}$ is at least $1 - ab2^{-\Omega(r)}$ (by a union bound). Thus, with

probability at least $1 - ab2^{-\Omega(r)}$, Bob is able to correctly determine the exact value of $|\mathbf{x} \wedge \mathbf{y}|$. Choosing $r = \Theta(\log(ab))$ yields a 1-way ISR protocol with $O(a \log(ab))$ bits of communication from Alice to Bob.

The lower bound $\mathrm{R}^{1\text{-way}}(\mathrm{SSI}^n_{a,b}) \geq \Omega(\max\{a, \log b\})$ will follow from Lemma 4.18 which we proved in Section 4.5, because any protocol for $\mathrm{SSI}^n_{a,b}$ can be used to compute $\overline{\mathrm{GHD}}^n_{a,b,c,1}$, where we can choose c to be anything, in particular, we choose $c \approx \max\{a, b\}$. $\qquad\square$

Lemma 4.3 (2-way ISR Protocol for $\mathrm{SSI}^n_{a,b}$). *Let $n, a, b \in \mathbb{N}$. Let $a, b \leq n/2$. Additionally, assume wlog that $a \leq b$ (since the roles of Alice and Bob can be flipped). Then,*

$$\Omega(\max\{a, \log b\}) \leq \mathrm{R}(\mathrm{SSI}^n_{a,b}) \leq \mathrm{ISR}_\rho(\mathrm{SSI}^n_{a,b}) = O(a \log(ab))$$

Proof. The ISR protocol is same as in proof of Lemma 4.3, with the difference that we flip the roles of Alice and Bob if $a \geq b$. The lower bound of $\mathrm{R}^{1\text{-way}}(\mathrm{SSI}^n_{a,b}) \geq \Omega(\max\{a, \log b\})$ follows from Lemma 3.3 as any protocol for $\mathrm{SSI}^n_{a,b}$ also solves $\overline{\mathrm{GHD}}^n_{a,b,b,1}$. $\qquad\square$

Remark 4.4. *The protocol given in proof of Lemma 4.2 can be viewed as a generalization of the protocol of [BGI14] for the* EQUALITY *function. More precisely, the* EQUALITY *function on n bit strings is equivalent to the $\mathrm{SSI}^{2^n}_{1,1}$ problem. This is because Alice and Bob can keep a list of all 2^n elements of $\{0,1\}^n$ (e.g., in lexicographic order) and then view their input strings as subsets of cardinality 1 of this list.*

We will repeatedly use the following corollary which follows from Lemma 4.2 by setting $a = 1$ and $b = 2^k$. Note that $\mathrm{SSI}^n_{1,2^k}$ is like the reverse direction of SPARSE-INDEXING, in which Alice had a large set and Bob had a singleton set.

Corollary 4.5 (Reverse SPARSE-INDEXING). *$\forall\ n, k \in \mathbb{N}$, $\mathrm{ISR}^{1\text{-way}}(\mathrm{SSI}^n_{1,2^k}) = O(k)$.*

4.1.2 Small Hamming Distance

The SMALL-HAMMING-DISTANCE problem is defined as follows.

Definition 4.6 (SMALL-HAMMING-DISTANCE). *Let $n, k \in \mathbb{N}$ and $0 \leq k \leq n-1$. Then $\mathrm{HD}^n_k : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is defined as follows,*

$$\mathrm{HD}^n_k(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{if } \Delta(\mathbf{x}, \mathbf{y}) \leq k \\ 0 & \text{if } \Delta(\mathbf{x}, \mathbf{y}) > k \end{cases}$$

Essentially, Alice is given $\mathbf{x} \in \{0,1\}^n$ and Bob is given $\mathbf{y} \in \{0,1\}^n$ and they wish to distinguish between the cases $\Delta(\mathbf{x}, \mathbf{y}) \leq k$ and $\Delta(\mathbf{x}, \mathbf{y}) > k$.

The following lemma shows that ISR-CC and PSR-CC are polynomially related for HD^n_k (for both the 1-way and 2-way models).

Lemma 4.7 (ISR Protocol for SMALL-HAMMING-DISTANCE). *Let $n, k \in \mathbb{N}$. Additionally, assume wlog that $k \leq n/2$ (since Bob can flip his input, and thus computing HD^n_k is equivalent to computing HD^n_{n-k}). Then,*

$$\Omega(k) \leq \mathrm{R}(\mathrm{HD}^n_k) \leq \mathrm{ISR}^{1\text{-way}}(\mathrm{HD}^n_k) \leq O(k^2)$$

In order to prove Lemma 4.7, we will use the following protocol (from [CGMS14]) twice, namely in the proofs of Lemmas 4.9 and 4.10.

Lemma 4.8 (ISR protocol for GAP-INNER-PRODUCT [CGMS14]). *Let $-1 \leq s < c \leq 1$ be real numbers. Assume that Alice is given a vector $\mathbf{u} \in \mathbb{R}^n$ such that $||\mathbf{u}||_2 = 1$, and that Bob is given a vector $\mathbf{v} \in \mathbb{R}^n$ such that $||\mathbf{v}||_2 = 1$. Then, there is a 1-way ISR protocol that distinguishes between the cases $\langle \mathbf{u}, \mathbf{v} \rangle \geq c$ and $\langle \mathbf{u}, \mathbf{v} \rangle \leq s$ using $O(1/(c-s)^2)$ bits of communication from Alice to Bob.*

Lemma 4.9. *Assume that $k < n/20$. Then, there is a 1-way ISR protocol that distinguishes between the cases $\Delta(x, y) \leq k$ and $\Delta(x, y) > n/10$ using $O(1)$ bits of communication.*

Proof. Let Alice construct the vector $\mathbf{u} \in \mathbb{R}^n$ by setting $\mathbf{u}_i = (-1)^{x_i}/\sqrt{n}$ for every $i \in [n]$ and let Bob construct the vector $\mathbf{v} \in \mathbb{R}^n$ by setting $\mathbf{v}_i = (-1)^{y_i}/\sqrt{n}$ for every $i \in [n]$. Then, we have that $||\mathbf{u}||_2 = ||\mathbf{v}||_2 = 1$. Furthermore, $\langle \mathbf{u}, \mathbf{v} \rangle = 1 - 2\Delta(x, y)/n$. Therefore, $\Delta(\mathbf{x}, \mathbf{y}) \leq k$ implies that $\langle \mathbf{u}, \mathbf{v} \rangle \geq 1 - 2k/n$ and $\Delta(\mathbf{x}, \mathbf{y}) > n/10$ implies that $\langle \mathbf{u}, \mathbf{v} \rangle < 4/5$. Setting $c := 1 - 2k/n$ and $s := 4/5$ and using the assumption that $k < n/20$, Lemma 4.8 yields a 1-way ISR protocol with $O(1)$ bits of communication from Alice to Bob (since $(c - s) \geq 1/10$). $\square$

Lemma 4.10. *Assume that $k < n/20$. Then, there is a 1-way ISR protocol that distinguishes between the cases $\Delta(x, y) \leq k$ and $k < \Delta(x, y) \leq n/10$ using $O(k^2)$ bits of communication.*

Proof. As in the proof of Lemma 4.9, we let Alice and Bob construct unit vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ by setting $\mathbf{u}_i = (-1)^{x_i}/\sqrt{n}$ and $\mathbf{v}_i = (-1)^{y_i}/\sqrt{n}$ (respectively) for every $i \in [n]$. Then, $\langle \mathbf{u}, \mathbf{v} \rangle = 1 - 2\Delta(\mathbf{x}, \mathbf{y})/n$. Let $t := n/(10k)$. Alice tensorizes her vector t times to obtain the vector $\mathbf{u}^{\otimes t} \in \mathbb{R}^{n^t}$, namely, for every $i_1, i_2, \ldots, i_t \in [n]$, she sets $\mathbf{u}^{\otimes t}_{(i_1, i_2, \ldots, i_t)} = \prod_{j=1}^{t} u_{i_j}$. Similarly, Bob tensorizes his vector t times to obtain the vector $\mathbf{v}^{\otimes t} \in \mathbb{R}^{n^t}$. Observe that $||\mathbf{u}^{\otimes t}||_2 = ||\mathbf{u}||_2^t = 1$, $||\mathbf{v}^{\otimes t}||_2 = ||\mathbf{v}||_2^t = 1$ and

$$\langle \mathbf{u}^{\otimes t}, \mathbf{v}^{\otimes t} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle^t = (1 - 2\Delta(\mathbf{x}, \mathbf{y})/n)^t.$$

Therefore, $\Delta(\mathbf{x}, \mathbf{y}) \leq k$ implies that $\langle \mathbf{u}^{\otimes t}, \mathbf{v}^{\otimes t} \rangle \geq (1 - 2k/n)^t := c$, and $k < \Delta(x, y) \leq n/10$ implies that $\langle \mathbf{u}^{\otimes t}, \mathbf{v}^{\otimes t} \rangle \leq (1 - 2(k+1)/n)^t := s$. The inner product gap $c - s$ is at least

$$
\begin{aligned}
\left(1 - \frac{2k}{n}\right)^t - \left(1 - \frac{2(k+1)}{n}\right)^t
&= \left(1 - \frac{2k}{n}\right)^t \left[1 - \left(1 - \frac{1}{n/2 - k}\right)^t\right] \\
&\geq \left(1 - \frac{2kt}{n}\right)\left(1 - e^{-t/(n/2 - k)}\right) \\
&\geq \frac{4}{5}\left(1 - e^{-2/(9k)}\right) \quad [\because t = n/10k \text{ and } k < n/20] \\
&\geq \Omega\left(\frac{1}{k}\right)
\end{aligned}
$$

where the first inequality above follows from the fact that $(1 + x)^r \leq e^{rx}$ for every $x, r \in \mathbb{R}$ with $r > 0$, as well as the fact that $(1 - x)^r \geq 1 - xr$ for every $0 \leq x \leq 1$ and $r \geq 1$. Moreover, the last equality above follows from the fact that $(1 - e^{-x})/x \to 1$ as $x \to 0$. Therefore, applying Lemma 4.8 with $c - s = \Omega(1/k)$ yields a 1-way ISR protocol with $O(k^2)$ bits of communication from Alice to Bob. $\square$

We are now ready to prove of Lemma 4.7.

Proof of Lemma 4.7. Assume without loss of generality that $k < n/20$, since otherwise, Alice can simply send her entire input (n bits) to Bob, requiring only $O(k)$ communication. Run the protocols from Lemmas 4.9 and 4.10 in sequence[10] and declare that $\Delta(x, y) \leq k$ if and only if both

---

[10]More precisely, we first repeat each of the two protocols a constant number of times and take a majority-vote of the outcomes. This allows us to reduce the error probability to a small enough constant and thereby apply a union bound.

protocols say so; otherwise, declare that $\Delta(x,y) > k+1$. This gives a 1-way ISR protocol with $O(k^2)$ bits of communication from Alice to Bob.

The lower bound $R(\mathrm{HD}_k^n) \geq \Omega(k)$ follows from Lemma 3.3 as any protocol the computes HD_k^n can be used to compute $\overline{\mathrm{GHD}}_{k,k,k,1}^n$ as well. $\qquad\square$

4.1.3 Strongly Permutation-Invariant functions

In this section, we show that the ISR-CC is polynomially related to the PSR-CC – without any additive dependence on n – for a natural subclass of permutation-invariant functions that we call "strongly permutation-invariant functions". We point out that this section is not needed for proving Theorem 1.3, but we include it because it highlights some of the proof ideas that we eventually use. We start by defining strongly permutation-invariant functions.

Definition 4.11 ((Total) Strongly Permutation-Invariant functions)**.** *A (total) function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is strongly permutation-invariant if there exists a symmetric function $\sigma : \{0,1\}^n \to \{0,1\}$ and a function $h : \{0,1\}^2 \to \{0,1\}$ such that for every $\mathbf{x}, \mathbf{y} \in \{0,1\}^n$,*

$$f(\mathbf{x}, \mathbf{y}) = \sigma(h(x_1, y_1), h(x_2, y_2), \cdots, h(x_n, y_n))$$

Note that strongly permutation-invariant functions include as subclasses, (AND)-symmetric functions (studied, e.g., by [BdW01, Raz03, She11]) and XOR-symmetric functions (studied, e.g., by [ZS09]).

The following theorem shows that ISR-CC of any strongly permutation-invariant function is polynomially related to its PSR-CC, with no dependence on n.

Theorem 4.12. *For any total strongly permutation-invariant function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, if $R(f) = k$ then,*

$$
\begin{aligned}
\mathrm{ISR}(f) &\leq& \widetilde{O}(k^2) \\
\mathrm{ISR}^{1\text{-}way}(f) &\leq& \widetilde{O}(k^3)
\end{aligned}
$$

Proof. Depending on h, any such function depends only on the sum of some subset of the quantities $\{|\mathbf{x} \wedge \mathbf{y}|, |\mathbf{x} \wedge \neg\mathbf{y}|, |\neg\mathbf{x} \wedge \mathbf{y}|, |\neg\mathbf{x} \wedge \neg\mathbf{y}|\}$. There are three main cases to consider (the remaining cases being similar to these three):

(i) *f depends only on $|\mathbf{x} \wedge \mathbf{y}| + |\mathbf{x} \wedge \neg\mathbf{y}|$:* In this case, f depends only on $|\mathbf{x}|$, and hence $R(f)$, $\mathrm{ISR}(f)$ and $\mathrm{ISR}^{1\text{-}way}(f)$ are all 1.

(ii) *f depends only on $|\mathbf{x} \wedge \mathbf{y}| + |\neg\mathbf{x} \wedge \neg\mathbf{y}|$:* In this case, f depends only on $|\mathbf{x} \oplus \mathbf{y}|$. Let $R(f) = k$, and suppose i is such that $f(\mathbf{x}, \mathbf{y}) = 0$ for $|\mathbf{x} \oplus \mathbf{y}| = i - 1$, and $f(\mathbf{x}, \mathbf{y}) = 1$ for $|\mathbf{x} \oplus \mathbf{y}| = i + 1$. If $i \leq n/2$ then any protocol to compute f can be used to compute $\overline{\mathrm{GHD}}_{i,i,i,1}^n$. Applying Lemma 3.3 we get that, $k = R(f) \geq R(\overline{\mathrm{GHD}}_{i,i,i,1}^n) \geq \mathrm{IC}(\overline{\mathrm{GHD}}_{i,i,i,1}^n) \geq i/C$. If $i > n/2$, then any protocol to compute f can be used to compute $\overline{\mathrm{GHD}}_{i,i,i,1}^n$. Applying Lemma 3.3 again, we get that, $k = R(f) \geq R(\overline{\mathrm{GHD}}_{i,n-i,i,1}^n) \geq \mathrm{IC}(\overline{\mathrm{GHD}}_{i,n-i,i,1}^n) \geq \Omega(n-i)$. Thus, we get that for any such i, it must be the case that either $i \leq Ck$ or $i \geq n - Ck$.

Alice and Bob now use the 1-way ISR protocol given by Lemma 4.7 to solve HD_i^n for every i such that $i \leq Ck$ or $i \geq n - Ck$, and for each such problem, they repeat the protocol $O(\log k)$ times to make the error probability down to $O(1/k)$. This yields a 1-way ISR protocol with $\widetilde{O}(k^3)$ bits of communication from Alice to Bob. This protocol can be modified into a 2-way ISR protocol with only $\widetilde{O}(k^2)$ bits of communication by letting Alice and Bob binary-search over the $O(k)$ Hamming distance problems that they need to solve, instead of solving all of them in parallel.

23

(iii) *f depends only on $|\mathbf{x} \wedge \mathbf{y}|$:* Suppose j is such that $f(\mathbf{x}, \mathbf{y}) = 0$ when $|\mathbf{x} \wedge \mathbf{y}| = j$ and $f(\mathbf{x}, \mathbf{y}) = 1$ when $|\mathbf{x} \wedge \mathbf{y}| = j + 1$. Then, if we restrict to only $\mathbf{x}$ and $\mathbf{y}$ such that $|\mathbf{x}| = |\mathbf{y}| = (n + 2j)/3$, then any protocol to compute f can be used to compute $\overline{\mathrm{GHD}}_{a,a,c,1}^{n}$, where $a = (n + 2j)/3$ and $c = (2n - 2j)/3$. Applying Lemma 3.3 we get that, $k = \mathrm{R}(f) \geq \mathrm{R}(\overline{\mathrm{GHD}}_{a,a,c,1}^{n}) \geq \mathrm{IC}(\overline{\mathrm{GHD}}_{a,a,c,1}^{n}) \geq 2(n - j)/3C$. This implies that $j \geq n - 2Ck$. In particular, we deduce that there are at most $O(k)$ such values of j. On input pair $(\mathbf{x}, \mathbf{y})$, Alice checks whether $|\mathbf{x}| \geq n - 2Ck$ and Bob checks whether $|\mathbf{y}| \geq n - 2Ck$. If one of these inequalities fail, then it is the case that $|\mathbf{x} \wedge \mathbf{y}| < n - 2Ck$ and the function value can be deduced. Suppose that $|\mathbf{x}| \geq n - 100k$ and $|\mathbf{y}| \geq n - 100k$. In this case, $|\mathbf{x}|$ takes one of $O(k)$ possibilities and hence Alice can send $|\mathbf{x}|$ to Bob using $O(\log k)$ bits. At this point, Bob knows both $|\mathbf{x}|$ and $|\mathbf{y}|$. Thus, using the identity $\Delta(\mathbf{x}, \mathbf{y}) = |\mathbf{x}| + |\mathbf{y}| - 2|\mathbf{x} \wedge \mathbf{y}|$, the problem gets reduced to a collection $O(k)$-Hamming Distance problems as in case (ii) above. The same protocols as in case (ii) imply a 1-way ISR protocol with $\widetilde{O}(k^3)$ bits of communication from Alice to Bob, and a 2-way ISR protocol with $\widetilde{O}(k^2)$ bits of communication. □

4.2 Overview of Proofs

The proofs of the 1-way and 2-way parts of Theorem 1.3 follow the same general framework, which we describe next. Let $(\mathbf{x}, \mathbf{y})$ be the input pair, $a := |\mathbf{x}|$ and $b := |\mathbf{y}|$. We partition the (a, b)-plane into a constant number of regions such that:

(i) Using a small amount of communication, Alice and Bob can distinguish in which region their combined input lies.

(ii) For each of these regions, there is an ISR protocol with small communication that computes the function value on any input in this region.

Some of the region-specific protocols (in (ii) above) will be based on low-communication ISR protocols for two "atomic" problems: SMALL-SET-INTERSECTION and SMALL-HAMMING-DISTANCE described in Section 4.1.

We point out again that both of our protocols for SMALL-SET-INTERSECTION and SMALL-HAMMING-DISTANCE have ISR-CC that is polynomial in the underlying PSR-CC, which is crucial for our purposes. In particular, one cannot instead use the generic exponential simulation of [CGMS14].

The additive $O(\log \log n)$ factor in the ISR-CC upper bound of Theorem 1.3 is due to the fact that for one region (other than the ones mentioned above), we show that in order for Alice and Bob to compute the function value $f(\mathbf{x}, \mathbf{y})$, it is enough that they compute some other low PSR-CC function $f'(|\mathbf{x}|, |\mathbf{y}|)$ of the *Hamming weights* of $(\mathbf{x}, \mathbf{y})$. Since the Hamming weight of an n-bit string can be expressed using $O(\log n)$ bits, we have effectively reduced the "dimension" of the problem from n to $O(\log n)$. At this point, we can apply Newman's theorem (Theorem 2.6) to obtain a private-coin protocol computing $f'(|\mathbf{x}|, |\mathbf{y}|)$ (and hence $f(\mathbf{x}, \mathbf{y})$) while increasing the communication cost by at most an additive $O(\log \log n)$.

4.3 2-way ISR Protocol for Permutation-Invariant Functions

In this section, we prove the 2-way part of Theorem 1.3. We again use the measure $m(f)$ (introduced in Definition 3.2) when restricted to *total* functions. For the sake of clarity, we describe the resulting specialized expression of $m(f)$ again in the following proposition.

Proposition 4.13 (Measure $m(f)$ for total functions)**.** *Given a total permutation-invariant function $f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$, and integers a, b, s.t. $0 \leq a, b \leq n$, let $h_{a,b} : \{0, 1, \cdots, n\} \to \{0, 1, ?\}$ be the function given by $h_{a,b}(d) = f(\mathbf{x}, \mathbf{y})$ if there exist $\mathbf{x}$, $\mathbf{y}$ with $|\mathbf{x}| = a$, $|\mathbf{y}| = b$, $\Delta(\mathbf{x}, \mathbf{y}) = d$ and*

? *otherwise. (Note. by permutation invariance of f, $h_{a,b}$ is well-defined.) Let $\mathscr{J}(h_{a,b})$ be the set of jumps in $h_{a,b}$, defined as follows,*

$$\mathscr{J}(h_{a,b}) \quad \overset{\text{def}}{=} \quad \left\{ c : \begin{array}{c} h_{a,b}(c-1) \neq h_{a,b}(c+1) \\ h_{a,b}(c-g),\ h_{a,b}(c+g) \in \{0,1\} \end{array} \right\}$$

Then, we define $m(f)$ as follows.

$$m(f) \quad \overset{\text{def}}{=} \quad \frac{1}{C} \cdot \max_{\substack{a,b \in [n] \\ (c,g) \in \mathscr{J}(h_{a,b})}} \max\{\min\{a,b,c,n-a,n-b,n-c\}, \log(\min\{c,n-c\})\}$$

where C is a suitably large constant (which does not depend on n).

We will now prove the following theorem, which immediately implies the 2-way part of Theorem 1.3.

Theorem 4.14. *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be any (total) permutation-invariant function. Then,*

$$m(f) \leq R(f) \leq ISR(f) \leq \widetilde{O}(m(f)^3) + R(f) + O(\log\log n)$$

Proof. Any protocol to compute f also computes $\overline{\text{GHD}}^n_{a,b,c,g}$ for any a, b and any jump $c \in \mathscr{J}(h_{a,b})$. Consider the choice of a, b and a jump $c \in \mathscr{J}(h_{a,b})$ such that the lower bound obtained on $\text{IC}(\overline{\text{GHD}}^n_{a,b,c,1})$ through Lemma 3.3 is maximized, which by definition is equal to $m(f)$ (after choosing an appropriate C in the definition of $m(f)$). Thus, we have $m(f) \leq \text{IC}(\overline{\text{GHD}}^n_{a,b,c,g}) \leq \text{IC}(f) \leq R(f)$.

Let $k \overset{\text{def}}{=} m(f)$. The main part of the proof is to show that $ISR(f) \leq \widetilde{O}(k^3) + R(f) + O(\log\log n)$. We first divide the input space into a constant number of regions, such that Alice can send $O(\log k)$ number of bits to Bob with which he can decide in which of the regions does their combined input lie (with high probability). Thus, once we break down the input space into these regions, it will suffice to give 2-way protocols with small ISR-CC for computing the function over each of these regions; as Alice and Bob can first determine in which region their combined input lies, and then run the corresponding protocol to compute the function value.

Let $a = |\mathbf{x}|$ and $b = |\mathbf{y}|$. We divide the (a,b)-plane into 4 regions, (I), (II), (III) and (IV), based on the values of a and b as follows. First let $A = \min\{a, n-a\}$ and $B = \min\{b, n-b\}$. Then the regions are given by,

(I) $(A \leq Ck$ and $B \leq 2^{Ck})$ or $(A \leq 2^{Ck}$ and $B \leq Ck)$

(II) $(A \leq Ck$ and $B > 2^{Ck})$ or $(A > 2^{Ck}$ and $B \leq Ck)$

(III) $A > Ck$ and $B > Ck$ and $|A - B| < Ck$

(IV) $A > Ck$ and $B > Ck$ and $|A - B| \geq Ck$

where C comes from Definition 3.2. Note that regions (I), (II), (III) and (IV) form a *partition* of the (a,b)-plane. This division is shown pictorially in Figure 1.

First, note that if $|\mathbf{x}| > n/2$, then Alice can flip all her input bits and convey that she did so to Bob using one bit of communication. Similarly, if $|\mathbf{y}| > n/2$, then Bob can flip all his input bits and convey that he did so to Alice using one bit of communication. After these flipping operations, Alice and Bob will look at the appropriately modified version of f based on who all flipped their input. Note that flipping all the bits of Alice and/or Bob preserves the permutation-invariance of the function. We will henceforth assume w.l.o.g. that $a = |\mathbf{x}| \leq n/2$ and $b = |\mathbf{y}| \leq n/2$. [This is also the reason that the regions described above are in terms of $A = \min\{a, n-a\}$ and $B = \min\{b, n-b\}$ and $A, B \leq n/2$.]
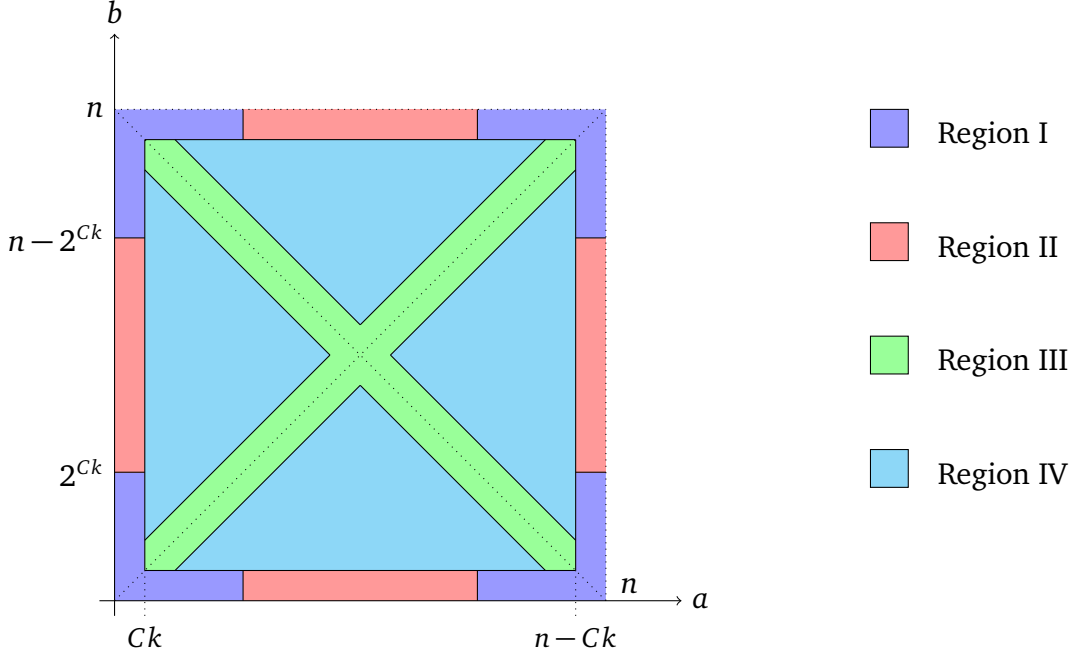
25

Figure 1: Division of the (a, b) plane into four regions in the 2-way setting

Next, we show that determining the region to which the input pair $(\mathbf{x}, \mathbf{y})$ belongs can be done using $O(\log k)$ bits of (ISR) communication from Alice to Bob. First, Alice will send Bob two bits indicating whether $a \leq Ck$ and whether $a \leq 2^{Ck}$ respectively. With this information Bob can determine in which of regions $\{(\text{I}), (\text{II}), (\text{III}) \cup (\text{IV})\}$ the combined input lies. To distinguish between regions (III) and (IV), Alice and Bob can first check whether $|a - b| < 100k$ by setting up an instance of SPARSE-INDEXING. Namely, Alice will translate the value a into a string s_a where $s_a(i) = 1$ iff $i = a$. And Bob will translate b into a string s_b such that $s_b(i) = 1$ iff $b - Ck < i < b + Ck$. This is an instance of SPARSE-INDEXING which can be solved with $O(\log k)$ bits of ISR-CC, by Corollary 4.5.

We now show how to compute the value of the function f in each of the 4 individual regions (I), (II), (III) and (IV) using ISR-CC of at most $\widetilde{O}(k^3) + R(f) + O(\log \log n)$ bits. Since f is a permutation-invariant function, we use the following 2 interchangeable representations of f using Observation 2.2,

$$f(\mathbf{x}, \mathbf{y}) = g(|\mathbf{x}|, |\mathbf{y}|, |\mathbf{x} \wedge \mathbf{y}|) = h(|\mathbf{x}|, |\mathbf{y}|, \Delta(\mathbf{x}, \mathbf{y}))$$

(I) (Main idea: SMALL-SET-INTERSECTION) We have that either ($a \leq Ck$ and $b \leq 2^{Ck}$) or ($a \leq 2^{Ck}$ and $b \leq Ck$). Since we can interchange the roles of Alice and Bob if required, we can assume w.l.o.g. that $a \leq Ck$ and $b \leq 2^{Ck}$. In this case, Alice first sends the value of $a = |\mathbf{x}|$ to Bob. They can then apply the protocol from Lemma 4.2 in order to compute $|\mathbf{x} \wedge \mathbf{y}|$ using $O(a \log(ab)) = O(k^2)$ bits of 1-way ISR communication from Alice to Bob. Hence, Bob can determine $|\mathbf{x} \wedge \mathbf{y}|$ correctly with high probability, and hence deduce $g(|\mathbf{x}|, |\mathbf{y}|, |\mathbf{x} \wedge \mathbf{y}|) = f(\mathbf{x}, \mathbf{y})$.

(II) (Main idea: *No dependence on* $\Delta(\mathbf{x}, \mathbf{y})$) We have that either ($a \leq Ck$ and $b > 2^{Ck}$) or ($a > 2^{Ck}$ and $b \leq Ck$). Since we can interchange the roles of Alice and Bob if required, we can assume w.l.o.g. that $a \leq Ck$ and $b > 2^{Ck}$. Then, the definition of the measure $m(f)$ implies that for this range of values of a and b, the function h cannot depend on $\Delta(\mathbf{x}, \mathbf{y})$ (because in this case $\Delta(\mathbf{x}, \mathbf{y}) \geq (b - a)$). Since h depends only on $|\mathbf{x}|$ and $|\mathbf{y}|$, Alice can simply send the value of a (which takes only $O(\log k)$ bits), with which Bob can compute $h(a, b, c)$ for any valid c,

26

that is, $h_{a,b}(c) \neq ?$.

(III) (Main idea: SMALL-HAMMING-DISTANCE) We have that $|a - b| < 100k$. In this case, Alice sends the value of $a \pmod{2Ck}$ to Bob, requiring $O(k)$ 1-way communication. Since Bob knows b, and so he can figure out the exact value of a. Next, they need to determine the Hamming distance $\Delta(\mathbf{x}, \mathbf{y})$. The definition of our measure $m(f)$ (along with the fact that $k := m(f)$) implies that if $h(a, b, c-1) \neq h(a, b, c+1)$, then c must be either $\leq Ck$ or $\geq n - Ck$. That is, given a and b, $h(a, b, c)$ must equal a constant for all valid c such that $Ck < c < n - Ck$.

Since Bob knows both a and b exactly, Alice and Bob can run the 1-way ISR-protocol for HD_i^n (from Lemma 4.7) for every $i < Ck$ and $i > n - Ck$. This requires $\widetilde{O}(k^3)$ communication.[11] If the Hamming distance is either $\leq Ck$ or $\geq n - Ck$, then Bob can determine it exactly and output $h(a, b, c)$. If the Hamming distance is not in this range, then the Hamming distance does not influence the value of the function, in which case Bob can output $h(a, b, c)$ for any valid c such that $Ck < c < n - Ck$.

(IV) (Main idea: *Newman's theorem on* (a, b)) We claim that for this range of values of a and b, the function $h(a, b, c)$ cannot depend on c. Suppose for the sake of contradiction that the function depends on c. Hence, there exists a value of c for which $h(a, b, c-1) \neq h(a, b, c+1)$. Since $|a - b| \geq Ck$, we have that $c \geq Ck$. And since $|n - a - b| \geq Ck$, we also have that $c \leq \min\{a + b, 2n - a - b\} \leq n - Ck$. Thus, we get that $Ck \leq c \leq n - Ck$, which contradicts that $m(f) = k$ (see Proposition 4.13).

Since $f(\mathbf{x}, \mathbf{y})$ only depends on $|\mathbf{x}|$ and $|\mathbf{y}|$ in this region, we have converted the problem which depended on inputs of size n, into a problem which depends on inputs of size $O(\log n)$ only. Since the original problem had a PSR-protocol with $\mathrm{R}(f)$ bits of communication, applying Newman's theorem (Theorem 2.6), we conclude that, with private randomness itself, the problem can be solved using $\mathrm{R}(f) + O(\log \log n)$ bits of 2-way communication. □

Note 4.15. *We point that the proof of Theorem 4.14 shows, more strongly, that for any function $G(.)$, the following two statements are equivalent:*

(1) *Any total function f with $\mathrm{R}(f) = k$ has $\mathrm{ISR}(f) \leq \mathrm{poly}(k) + G(n)$*

(2) *Any total permutation-invariant function f with $\mathrm{R}(f) = k$ has $\mathrm{ISR}(f) \leq \mathrm{poly}(k) + G(O(\log n))$.*

Newman's theorem (Theorem 2.6) implies that (1) holds for $G(n) = O(\log n)$, and hence we have that (2) also holds for $G(n) = O(\log n)$; thereby yielding the bound implied by Theorem 4.14. On the other hand, improving the $O(\log \log n)$ term in Theorem 4.14 to, say $o(\log \log n)$ will imply that for all total functions we can have $G(n)$ in (1) to be $o(\log n)$. We note that currently such a statement is unknown.

4.4 1-way ISR Protocol for Permutation-Invariant Functions

In this section, we prove the 1-way part of Theorem 1.3. On a high level, the proof differs from that of the 2-way part in two aspects:

1. The underlying measure that is being used.

2. The partition of the (a, b)-plane into regions, which is no longer symmetric with respect to Alice and Bob as can be seen in Figure 2.

We introduce a new measure $m^{\text{1-way}}(f)$ as follows (this is the 1-way analog of Proposition 4.13).

---

[11] The ISR-protocol for HD_i^n takes $O(i^2) \leq O(k^2)$ communication each. But we can only afford to make error with probability $O(1/k)$, and thus for sake of amplification, the overall communication is $O(k^3 \log k)$.

Definition 4.16 (Measure $m^{1\text{-way}}(f)$ for total functions). *Given a total permutation-invariant function $f : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}$, and integers a, b, s.t. $0 \leq a, b \leq n$, let $h_{a,b} : \{0, 1, \cdots, n\} \rightarrow \{0, 1, ?\}$ be the function given by $h_{a,b}(d) = f(\mathbf{x}, \mathbf{y})$ if there exist $\mathbf{x}, \mathbf{y}$ with $|\mathbf{x}| = a$, $|\mathbf{y}| = b$, $\Delta(\mathbf{x}, \mathbf{y}) = d$ and $?$ otherwise. (Note. by permutation invariance of f, $h_{a,b}$ is well-defined.) Let $\mathscr{J}(h_{a,b})$ be the set of* jumps *in $h_{a,b}$, defined as follows,*

$$\mathscr{J}(h_{a,b}) \quad \overset{\text{def}}{=} \quad \left\{ c : \begin{array}{c} h_{a,b}(c-1) \neq h_{a,b}(c+1) \\ h_{a,b}(c-g),\, h_{a,b}(c+g) \in \{0,1\} \end{array} \right\}$$

Then, we define $m^{1\text{-way}}(f)$ as follows.

$$m^{1\text{-way}}(f) \quad \overset{\text{def}}{=} \quad \frac{1}{C} \cdot \max_{\substack{a,b \in [n] \\ (c,g) \in \mathscr{J}(h_{a,b})}} \max\left\{\min\{a, c, n-a, n-c\}, \log\left(\min\{c, n-c\}\right)\right\}$$

where C is a suitably large constant (which does not depend on n).

We point out that the only difference between Definition 4.16 and Proposition 4.13 is that the term $\min\{a, b, c, n-a, n-b, n-c\}$ in Proposition 4.13 is replaced by $\min\{a, c, n-a, n-c\}$ in Definition 4.16. In particular, Definition 4.16 is not symmetric with respect to Alice and Bob. We will now prove the following theorem, which immediately implies the 1-way part of Theorem 1.3.

Theorem 4.17. *Let $f : \{0,1\}^n \times \{0,1\}^n \rightarrow \{0,1\}$ be any (total) permutation-invariant function. Then,*

$$m^{1\text{-way}}(f) \leq \mathrm{R}^{1\text{-way}}(f) \leq \mathrm{ISR}^{1\text{-way}}(f) \leq \widetilde{O}(m^{1\text{-way}}(f)^3) + \mathrm{R}^{1\text{-way}}(f) + O(\log\log n)$$

We will need the following lemma to show that the measure $m^{1\text{-way}}(f)$ is a lower bound on $\mathrm{R}^{1\text{-way}}(f)$. This is analogous to Lemma 3.3, which was used to show that $m(f)$ is a lower bound on $\mathrm{R}(f)$ and in particular, $\mathrm{IC}(f)$.

Lemma 4.18. *For all n, a, b, c, g for which $\overline{\mathrm{GHD}}^n_{a,b,c,g}$ is a meaningful problem, the following lower bounds hold,*

$$\mathrm{R}^{1\text{-way}}(\overline{\mathrm{GHD}}^n_{a,b,c,g}) \geq \frac{1}{C} \cdot \left(\frac{\min\{a, n-a, c, n-c\}}{g}\right)$$

$$\mathrm{R}^{1\text{-way}}(\overline{\mathrm{GHD}}^n_{a,b,c,g}) \geq \frac{1}{C} \cdot \left(\log\left(\frac{\min\{c, n-c\}}{g}\right)\right)$$

We defer the proof of Lemma 4.18 to Section 4.5. For now, we will use this lemma to prove Theorem 4.17.

Proof. Any protocol to compute f also computes $\overline{\mathrm{GHD}}^n_{a,b,c,1}$ for any a, b and any jump $c \in \mathscr{J}(h_{a,b})$. Consider the choice of a, b and a jump $c \in \mathscr{J}(h_{a,b})$ such that the lower bound obtained on $\mathrm{IC}(\overline{\mathrm{GHD}}^n_{a,b,c,1})$ through Lemma 4.18 is maximized, which by definition is equal to $m^{1\text{-way}}(f)$ (after choosing an appropriate C in the definition of $m^{1\text{-way}}(f)$). Thus, we have $m^{1\text{-way}}(f) \leq \mathrm{R}^{1\text{-way}}(\overline{\mathrm{GHD}}^n_{a,b,c,g}) \leq \mathrm{R}^{1\text{-way}}(f)$.

Let $k \overset{\text{def}}{=} m^{1\text{-way}}(f)$. The main part of the proof is to show that $\mathrm{ISR}^{1\text{-way}}(f) \leq \widetilde{O}(k^3) + \mathrm{R}^{1\text{-way}} + O(\log\log n)$. We first divide the input space into a constant number of regions, such that Alice can send $O(\log k)$ bits to Bob with which he can decide in which of the regions does their combined input lie (with high probability). Thus, once we break down the input space into these regions, it will suffice to give 1-way protocols with small ISR-CC for computing the function over each of these regions; as Alice can then send the 1-way messages corresponding to all of these regions, and

Bob will first determine in which region their combined input lies and then use the corresponding messages of Alice to compute the function value.

Suppose that we have a function f with $m^{\text{1-way}}(f) = k$. Let $a = |\mathbf{x}|$ and $b = |\mathbf{y}|$. We partition the (a, b)-plane into 4 regions based on the values of a and b as follows. First let $A = \min\{a, n-a\}$ and $B = \min\{b, n-b\}$. Then the regions are given by,

(I) $A \leq Ck$ and $B \leq 2^{Ck}$

(II) $A \leq Ck$ and $B > 2^{Ck}$

(III) $A > Ck$ and $|A - B| < Ck$

(IV) $A > Ck$ and $|A - B| \geq Ck$

where C comes from Definition 4.16. Note that the regions (I), (II), (III) and (IV) form a partition of the (a, b) plane. This partition is shown pictorially in Figure 2.

First, note that if $|\mathbf{x}| > n/2$, then Alice can flip all her input bits and convey that she did so to Bob using one bit of communication. Similarly, if $|\mathbf{y}| > n/2$, then Bob can flip all his input bits and convey that he did so to Alice using one bit of communication. After these flipping operations, Alice and Bob will look at the appropriately modified version of f based on who all flipped their input. Note that flipping all the bits of Alice and/or Bob preserves the permutation-invariance of the function. We will henceforth assume w.l.o.g. that $a = |\mathbf{x}| \leq n/2$ and $b = |\mathbf{y}| \leq n/2$. [This is also the reason that the regions described above are in terms of $A = \min\{a, n-a\}$ and $B = \min\{b, n-b\}$ and $A, B \leq n/2$.]

We now show that determining the region to which the input pair $(\mathbf{x}, \mathbf{y})$ belongs can be done using $O(\log k)$ bits of (ISR) communication from Alice to Bob. First, to distinguish between $\{$(I), (II)$\}$ and $\{$(III), (IV)$\}$, Alice needs to send one bit to Bob indicating whether $a \leq 100k$. Moreover, in the case of $\{$(I), (II)$\}$, Bob can easily differentiate between (I) and (II) because he knows the value of b. To distinguish between regions (III) and (IV), Alice and Bob can first check whether $|a - b| < 100k$ by setting up an instance of SPARSE-INDEXING. Namely, Alice will translate the value a into a string s_a where $s_a(i) = 1$ iff $i = a$. And Bob will translate b into a string s_b such that $s_b(i) = 1$ iff $b - Ck < i < b + Ck$. This is an instance of SPARSE-INDEXING which can be solved with $O(\log k)$ bits of ISR-CC, by Corollary 4.5.

We now show how to compute the value of the function f in each of the 4 individual regions (I), (II), (III) and (IV) using 1-way ISR-CC of at most $\widetilde{O}(k^3) + R^{\text{1-way}}(f) + O(\log\log n)$ bits. Since f is a permutation-invariant function, we use the following 2 interchangeable representations of f using Observation 2.2,

$$f(\mathbf{x}, \mathbf{y}) = g(|\mathbf{x}|, |\mathbf{y}|, |\mathbf{x} \wedge \mathbf{y}|) = h(|\mathbf{x}|, |\mathbf{y}|, \Delta(\mathbf{x}, \mathbf{y}))$$

(I) (Main idea: SMALL-SET-INTERSECTION) We have that $a \leq Ck$ and $b \leq 2^{Ck}$. Alice can first send the value of $a = |\mathbf{x}|$ to Bob. They can then apply the protocol from Lemma 4.2 in order to compute $|\mathbf{x} \wedge \mathbf{y}|$ using $O(a \log(ab)) = O(k^2)$ bits of 1-way ISR communication from Alice to Bob. Hence, Bob can determine $|\mathbf{x} \wedge \mathbf{y}|$ correctly with high probability, and hence deduce $g(|\mathbf{x}|, |\mathbf{y}|, |\mathbf{x} \wedge \mathbf{y}|) = f(\mathbf{x}, \mathbf{y})$.

(II) (Main idea: *No dependence on $\Delta(\mathbf{x}, \mathbf{y})$*) We have that $a \leq Ck$ and $b > 2^{Ck}$. In this case, the definition of our measure $m^{\text{1-way}}(f)$ implies that for this range of values of a and b, the function h cannot depend on $\Delta(\mathbf{x}, \mathbf{y})$ (because in this case $\Delta(\mathbf{x}, \mathbf{y}) \geq (b - a)$). Since h depends only on $|\mathbf{x}|$ and $|\mathbf{y}|$, Alice can simply send the value of a (which takes only $O(\log k)$ bits), with which Bob can compute $h(a, b, c)$ for any valid c, that is, $h(a, b, c) \neq ?$.

(III) (Main idea: SMALL-HAMMING-DISTANCE) We have that that $|a - b| < Ck$. Then, Alice sends the value of $a \pmod{2Ck}$ to Bob, requiring $O(k)$ 1-way communication. Since Bob knows b, he can figure out the exact value of a. Next, they need to determine the Hamming distance
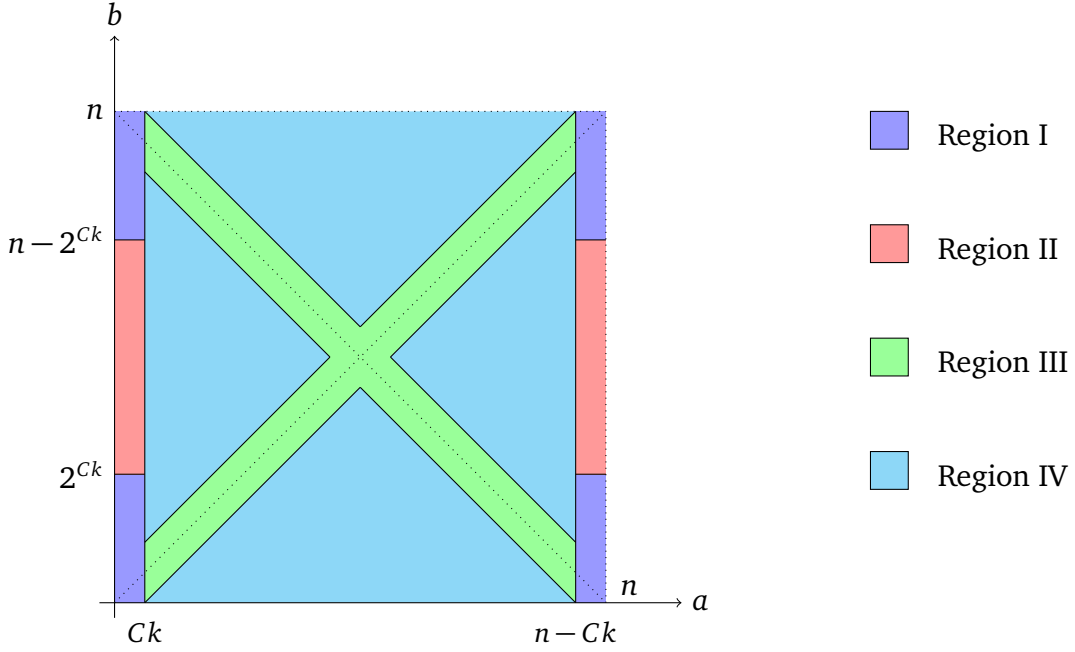
Figure 2: Division of the (a, b) plane into four regions in the 1-way setting

$\Delta(\mathbf{x}, \mathbf{y})$. The definition of our measure $m^{\text{1-way}}(f)$ (along with the fact that $k := m^{\text{1-way}}(f)$) implies that if $h(a, b, c-1) \neq h(a, b, c+1)$, then c must be either $\leq Ck$ or $\geq n-Ck$. That is, given a and b, $h(a, b, c)$ must equal a constant for all valid c such that $Ck < c < n-Ck$. Since Bob knows both a and b exactly, Alice and Bob can run the 1-way ISR-protocol for HD_i^n (from Lemma 4.7) for every $i < Ck$ and $i > n-Ck$. This requires $\widetilde{O}(k^3)$ communication.[12] If the Hamming distance is either $\leq Ck$ or $\geq n-Ck$, then Bob can determine it exactly and output $h(a, b, c)$. If the Hamming distance is not in this range, then the Hamming distance does not influence the value of the function, in which case Bob can output $h(a, b, c)$ for any valid c such that $Ck < c < n-Ck$.

(IV) (Main idea: *Newman's theorem* on (a, b)) We claim that for this range of values of a and b, the function $h(a, b, c)$ cannot depend on c. Suppose for the sake of contradiction that the function depends on c. Hence, there exists a value of c for which $h(a, b, c-1) \neq h(a, b, c+1)$. Since $|a - b| \geq Ck$, we have that $c \geq Ck$. And since $|n - a - b| \geq Ck$, we also have that $c \leq \min\{a + b, 2n - a - b\} \leq n - Ck$. Thus, we get that $Ck \leq c \leq n - Ck$, which contradicts that $m^{\text{1-way}}(f) = k$ (see Definition 4.16).

Since $f(\mathbf{x}, \mathbf{y})$ only depends on $|\mathbf{x}|$ and $|\mathbf{y}|$ in this region, we have converted the problem which depended on inputs of size n, into a problem which depends on inputs of size $O(\log n)$ only. Since the original problem had a 1-way PSR-protocol with $R^{\text{1-way}}(f)$ bits of communication, applying Newman's theorem (Theorem 2.6), we conclude that, with private randomness itself, the problem can be solved using $R^{\text{1-way}}(f) + O(\log \log n)$ bits of 1-way communication.

$\square$

Note 4.19. *We point that the proof of Theorem 4.17 shows, more strongly, that for any function $G(.)$, the following two statements are equivalent:*

---

[12]The ISR-protocol for HD_i^n takes $O(i^2) \leq O(k^2)$ communication each. But we can only afford to make error with probability $O(1/k)$, and thus for sake of amplification, the overall communication is $O(k^3 \log k)$.

(1) *Any total function f with $R^{1\text{-}way}(f) = k$ has $ISR^{1\text{-}way}(f) \leq \text{poly}(k) + G(n)$*

(2) *Any total permutation-invariant function f with $R^{1\text{-}way}(f) = k$ has $ISR^{1\text{-}way}(f) \leq \text{poly}(k) + G(O(\log n))$.*

Newman's theorem (Theorem 2.6) implies that (1) holds for $G(n) = O(\log n)$, and hence we have that (2) also holds for $G(n) = O(\log n)$; thereby yielding the bound implied by Theorem 4.17. On the other hand, improving the $O(\log\log n)$ term in Theorem 4.17 to, say $o(\log\log n)$ will imply that for all total functions we can have $G(n)$ in (1) to be $o(\log n)$. We note that currently such a statement is unknown.

4.5   1-way CC lower bounds on Gap Hamming Distance

In this section, we prove lower bounds on 1-way randomized communication complexity of GAP-HAMMING-DISTANCE (Lemma 4.18).

We will prove Lemma 4.18 by getting certain reductions from SPARSE-INDEXING (namely Proposition 2.13). Similar to our approach in Section 3.3, we first prove lower bounds on 1-way PSR-CC of SET-INCLUSION (Definition 3.8). We do this by obtaining a reduction from SPARSE-INDEXING.

Proposition 4.20 (SET-INCLUSION 1-way PSR-CC lower bound). *For all $t, w \in \mathbb{N}$,*

$$R^{1\text{-}way}(\text{SETINC}^{t+w}_{t,t+w-1}) \geq \Omega(\min(t, w))$$

Proof. We know that $R^{1\text{-}way}(\text{SPARSEIND}^{t+w}_t) \geq \Omega(\min(t, w))$ from Proposition 2.14. Note that SPARSEIND$^{t+w}_t$ is same as the problem of $\overline{\text{GHD}}^{t+w}_{t,1,t,1}$. If we instead think of Bob's input as complemented, we get that solving $\overline{\text{GHD}}^{t+w}_{t,t+w-1,w,1}$ is equivalent to solving $\overline{\text{GHD}}^{t+w}_{t,1,t,1}$, which is same as SETINC$^{t+w}_{t,t+w-1}$. Thus, we conclude that $R^{1\text{-}way}(\text{SETINC}^{t+w}_{t,t+w-1}) \geq \Omega(\min(t, w))$. $\qquad\square$

We now state and prove a technical lemma that will help us prove Lemma 4.18. Note that this is a 1-way analogue of Lemma 3.11.

Lemma 4.21. *Let $a, b \leq n/2$. Then, the following lower bounds hold,*

(i) $R^{1\text{-}way}(\overline{\text{GHD}}^n_{a,b,c,g}) \geq \Omega\left(\min\left\{\frac{c-b+a}{g}, \frac{n-c}{g}\right\}\right)$

(ii) $R^{1\text{-}way}(\overline{\text{GHD}}^n_{a,b,c,g}) \geq \Omega\left(\min\left\{\frac{a+b-c}{g}, \frac{c+b-a}{g}\right\}\right)$

(iii) $R^{1\text{-}way}(\overline{\text{GHD}}^n_{a,b,c,g}) \geq \Omega\left(\min\left\{\log\left(\frac{c}{g}\right), \log\left(\frac{n-c}{g}\right)\right\}\right)$

Proof. We first note that part (iii) follows trivially from Lemma 3.11 because $IC(f) \leq R(f) \leq R^{1\text{-}way}(f)$. But we require slightly different proofs for parts (i) and (ii). The main technical difference in these proofs and the corresponding ones in Lemma 3.11 is that now we cannot assume that $a \leq b \leq n/2$. We note that this assumption of $a \leq b$ was extremely crucial for Lemma 3.11, without which parts (i) and (ii) would not have been possible (even though they look exactly same in form!). The reason why we are able to succeed here without making the assumption because we derive reductions from a much "simpler" problem, but which is still hard in the 1-way model, namely the SPARSE-INDEXING problem.

Proof of (i). We obtain a reduction from SPARSEIND$^{2t}_t$ for which we know from Proposition 2.14 that $R^{1\text{-}way}(\text{SPARSEIND}^{2t}_t) = \Omega(t)$. Recall that SPARSEIND$^{2t}_t$ is same as $\overline{\text{GHD}}^{2t}_{t,1,t,1}$. We repeat the instance g times to get an instance of $\overline{\text{GHD}}^{2gt}_{gt,g,gt,g}$. Now, we need to append $(a - gt)$ 1's to $\mathbf{x}$ and $(b - g)$ 1's to $\mathbf{y}$. This will increase the Hamming distance by a fixed amount which is at least

$|a - b - gt + g|$ and at most $b + a - gt - g$. Also, the number of inputs we need to add is at least $((a - gt) + (b - g) + (c - gt))/2$. Thus, we can get a reduction to $\overline{\text{GHD}}^n_{a,b,c,g}$ if and only if,

$$|a - b - gt + g| \leq c - gt \leq b + a - gt - g$$

$$n \geq 2gt + \frac{(a - gt) + (b - g) + (c - gt)}{2}$$

The left condition on c gives us that $2gt \leq c + a - b + g$ and $b - a + g \leq c$ (which is always true). The right condition on c gives $c \leq b + a - g$ (which is always true). The condition on n gives that $gt \leq n - (a + b + c - g)/2$, which is equivalent to

$$t \leq \frac{n - a - b}{2g} + \frac{n - c - g}{2g}$$

Thus, we can have the reduction work by choosing t to be

$$t = \min\left\{\frac{c - b + a + g}{2g}, \frac{n - c - g}{2g}\right\} = \Omega\left(\min\left\{\frac{c - b + a}{g}, \frac{n - c}{g}\right\}\right)$$

(since $n \geq a + b$) and thus we obtain

$$\text{R}^{\text{1-way}}(\overline{\text{GHD}}^n_{a,b,c,g}) \geq \text{R}^{\text{1-way}}(\text{SPARSEIND}^{2t}_t) \geq \Omega\left(\min\left\{\frac{c - b + a}{g}, \frac{n - c}{g}\right\}\right)$$

Proof of (ii). We obtain a reduction from $\text{SETINC}^m_{t,t+w-1}$ (where $m = t + w$) for which we know from Proposition 4.20 that $\text{R}^{\text{1-way}}(\text{SPARSEIND}^m_{t,t+w-1}) = \Omega(\min\{t, w\})$. Recall that $\text{SETINC}^m_{t,t+w-1}$ is same as $\overline{\text{GHD}}^m_{t,t+w-1,w,1}$. Given an instance of $\overline{\text{GHD}}^m_{t,t+w-1,w,1}$, we first repeat the instance g times to get an instance of $\overline{\text{GHD}}^{gm}_{gt,g(t+w-1),gw,g}$. Now, we need to append $(a - gt)$ 1's to $\mathbf{x}$ and $(b - gt - gw + g)$ 1's to $\mathbf{y}$. This will increase the Hamming distance by a fixed amount which is at least $|b - a - gw + g|$ and at most $(b - gt - gw + g) + (a - gt)$. Also, the number of inputs we need to add is at least $((a - gt) + (b - gt - gw + g) + (c - gw))/2$. Thus, we can get a reduction to $\overline{\text{GHD}}^n_{a,b,c,g}$ if and only if,

$$|b - a - gw + g| \leq c - gw \leq b + a - 2gt - gw + g$$

$$n \geq gt + gw + \frac{(b - gt - gw + g) + (a - gt) + (c - gw)}{2}$$

The left constraint on c requires $c \geq \max\{b - a + g, a + 2gw - b - g\}$. We know that $c \geq b - a + g$, so the only real constraint is $c \geq 2gw - (b - a) + g$, which gives us that,

$$w \leq \frac{c + b - a - g}{2g}$$

The right constraint on c requires $c \leq b + a - 2gt + g$, which gives us that,

$$t \leq \frac{a + b - c + g}{2g}$$

Suppose we choose t to be $\frac{a+b-c+g}{2g}$. Then the constraint on n is giving us that,

$$n \geq gt + \frac{a + b + c - g}{2} = \frac{a + b - c + g}{2} + \frac{a + b + c - g}{2} = a + b$$

We already assumed that $a, b \leq n/2$, and hence this is always true.

Thus, we can choose $t = \frac{a+b-c+g}{2g}$ and $w = \frac{c+b-a-g}{2g}$, and from Proposition 4.20, we get,

$$\text{R}^{\text{1-way}}(\overline{\text{GHD}}^n_{a,b,c,g}) \geq \text{R}^{\text{1-way}}(\text{SETINC}^{t+w}_{t,w}) \geq \min(\{t, w\}) \geq \Omega\left(\min\left\{\frac{a + b - c}{g}, \frac{c + b - a}{g}\right\}\right)$$

$\square$

We are now finally able to prove Lemma 4.18.

Proof of Lemma 4.18. Assume for now that $a, b \leq n/2$. From parts (i) and (ii) of Lemma 4.21, we know the following,

$$R^{1\text{-way}}(\overline{\text{GHD}}_{a,b,c,g}^n) \geq \Omega\left(\min\left\{\frac{c-b+a}{g}, \frac{n-c}{g}\right\}\right)$$

$$R^{1\text{-way}}(\overline{\text{GHD}}_{a,b,c,g}^n) \geq \Omega\left(\min\left\{\frac{a+b-c}{g}, \frac{c+b-a}{g}\right\}\right)$$

Adding these up, we get that,

$$R^{1\text{-way}}(\overline{\text{GHD}}_{a,b,c,g}^n) \geq \Omega\left(\min\left\{\frac{c-b+a}{g}, \frac{n-c}{g}\right\} + \min\left\{\frac{a+b-c}{g}, \frac{c+b-a}{g}\right\}\right)$$

Since $\min\{A, B\} + \min\{C, D\} = \min\{A+C, A+D, B+C, B+D\}$, we get that,

$$R^{1\text{-way}}(\overline{\text{GHD}}_{a,b,c,g}^n) \geq \Omega\left(\min\left\{\frac{2a}{g}, \frac{2c}{g}, \frac{n+a+b-2c}{g}, \frac{n+b-a}{g}\right\}\right)$$

For the last two terms, note that, $n+a+b-2c \geq n-c$ (since $a+b \geq c$) and $n+b-a \geq n-a \geq a$ (since $n/2 \geq a$). Thus, overall we get,

$$R^{1\text{-way}}(\overline{\text{GHD}}_{a,b,c,g}^n) \geq \Omega\left(\min\left\{\frac{a}{g}, \frac{c}{g}, \frac{n-c}{g}\right\}\right)$$

Note that this was assuming $a, b \leq n/2$. In general, we get,

$$R^{1\text{-way}}(\overline{\text{GHD}}_{a,b,c,g}^n) \geq \Omega\left(\frac{\min\{a, c, n-a, n-c\}}{g}\right)$$

[We get the $(n-a)$ term because we might have flipped Alice's input to make sure $a \leq n/2$ But unlike in the proof of Lemma 3.3, we don't get b or $(n-b)$ in the above lower bound because while restricting to $a, b \leq n/2$, we never flipped the role of Alice and Bob.]

The second part follows trivially from the corresponding part of Lemma 3.3, since $\text{IC}(f) \leq \text{R}(f) \leq \text{R}^{1\text{-way}}(f)$ for any f.

We choose C to be a large enough constant, so that the desired lower bounds hold. □

5 Conclusion

In this work, we initiated the study of the communication complexity of permutation-invariant functions. We gave a coarse characterization of their information complexity and communication complexity (Theorem 1.1). We also showed for total permutation-invariant functions that the communication complexity with imperfectly shared randomness is not much larger than the communication complexity with perfectly shared randomness (Theorem 1.3). Our work points to several possible future directions.

- [**Generalized Permutation-Invariance**] Is it possible to generalize our results for any larger class of functions? One candidate might be classes of functions that are invariant under natural subgroups of the permutation group S_n, or more generally any group of actions on the input spaces of Alice and Bob. For example, once choice of a subgroup of permutations is the set of all permutations on $[n]$, that map $[n/2]$ to $[n/2]$ and map $[n]\setminus[n/2]$ to $[n]\setminus[n/2]$. Or more generally, the subgroup of ℓ-part permutation-symmetric functions, which consists of functions f for which there is a partition $I = \{I_1, \ldots, I_\ell\}$ of $[n]$ such that f is invariant under any permutation $\pi \in S_n$ where $\pi(I_i) = I_i$ for every $i \in \{1, \ldots, \ell\}$.

- [**Permutation-Invariance over higher alphabets**] Another interesting question is to generalize our results to larger alphabets, i.e., to permutation-invariant functions of the form $f : \mathcal{X}^n \times \mathcal{Y}^n \to R$ where $\mathcal{X}$, $\mathcal{Y}$ and R are not necessarily binary sets.

- [**Tight lower bounds for Gap-Hamming-Distance**] What are the tightest lower bounds on $\text{GHD}^n_{a,b,c,g}$ for all choices of parameters, a, b, c, g? Our lower bounds on $\text{GHD}^n_{a,b,c,g}$ are not tight for all choices of parameters a, b, c and g. For example, when $a = b = c = n/2$ and $g = \sqrt{n}$, our lower bound in Lemma 3.3 only implies $\text{IC}(\text{GHD}^n_{a,b,c,g}) \geq \Omega(\sqrt{n})$. By slightly modifying the proof techniques in [CR12, Vid11, She12], one can obtain that $\text{R}(\text{GHD}^n_{a,b,c,g}) = \Omega(n)$. Sherstov's proof [She12] is based on the *corruption bound*. Combining this with the recent result due to [KLL$^+$12] which showed (by studying a relaxed version of the partition bound of [JK10]) that many known lower bound methods for randomized communication complexity – including the corruption bound – are also lower bounds for the information complexity. This implies that $\text{IC}(\text{GHD}^n_{a,b,c,g}) = \Omega(n)$ for $a = b = c = n/2$ and $g = \sqrt{n}$.

- [**Hierarchy within ISR**] The work of [CGMS14] shows an exponential separation between $\text{ISR}_\rho(f)$ and $\text{R}(f)$. However, it is unclear if some *strong separation* could be shown between $\text{ISR}_\rho(f)$ and $\text{ISR}_{\rho'}(f)$ for some function f (where $\rho < \rho' < 1$).

- [**Limits of separation in [CGMS14]**] Canonne et al showed that for some *unspecified k*, there is a *partial* permutation-invariant function with communication at most k under perfect sharing of randomness, but with communication at least $\Omega(2^k)$ under imperfect sharing of randomness. Can their separation be made to hold for $k = \Theta(\log\log n)$? Answering this question would shed some light on the possibility of proving an analogue of Theorem 1.3 for partial permutation-invariant functions.

References

[BBCR13] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013. 9

[BBG14] Eric Blais, Joshua Brody, and Badih Ghazi. The Information Complexity of Hamming Distance. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*, volume 28 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 465–489, 2014. 3

[BBP$^+$96] Charles H Bennett, Gilles Brassard, Sandu Popescu, Benjamin Schumacher, John A Smolin, and William K Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Physical review letters*, 76(5):722, 1996. 5

[BCK$^+$14] Joshua Brody, Amit Chakrabarti, Ranganath Kondapally, David P Woodruff, and Grigory Yaroslavtsev. Beyond set disjointness: the communication complexity of finding the intersection. In *Proceedings of the 2014 ACM symposium on Principles of distributed computing*, pages 106–113. ACM, 2014. 3

[BdW01] Harry Buhrman and Ronald de Wolf. Communication complexity lower bounds by polynomials. In *Computational Complexity, 16th Annual IEEE Conference on, 2001.*, pages 120–130. IEEE, 2001. 3, 23

[BGI14] Mohammad Bavarian, Dmitry Gavinsky, and Tsuyoshi Ito. On the role of shared randomness in simultaneous communication. In *Automata, Languages, and Programming*, pages 150–162. Springer, 2014. 5, 6, 21

[BJKS04]   Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. Syst. Sci.*, 68(4):702–732, 2004. 10

[BM11]   Andrej Bogdanov and Elchanan Mossel. On extracting common random bits from correlated sources. *Information Theory, IEEE Transactions on*, 57(10):6351–6355, 2011. 5

[Bra12]   Mark Braverman. Interactive information complexity. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 505–524. ACM, 2012. 9

[BS94]   Gilles Brassard and Louis Salvail. Secret-key reconciliation by public discussion. In *advances in Cryptology—EUROCRYPT'93*, pages 410–423. Springer, 1994. 5

[BS15]   Mark Braverman and Jon Schneider. Information complexity is computable. *arXiv preprint arXiv:1502.02971*, 2015. 4

[BYJKS02]   Ziv Bar-Yossef, TS Jayram, Ravi Kumar, and D Sivakumar. An information statistics approach to data stream and communication complexity. In *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, pages 209–218. IEEE, 2002. 4

[CGMS14]   Clement Canonne, Venkat Guruswami, Raghu Meka, and Madhu Sudan. Communication with imperfectly shared randomness. *ITCS*, 2014. 5, 6, 8, 18, 21, 22, 24, 34

[CMN14]   Siu On Chan, Elchanan Mossel, and Joe Neeman. On extracting common random bits from correlated sources on large alphabets. *Information Theory, IEEE Transactions on*, 60(3):1630–1637, 2014. 5

[CR12]   Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. *SIAM Journal on Computing*, 41(5):1299–1317, 2012. 3, 6, 34

[Cre95]   Nadia Creignou. A dichotomy theorem for maximum generalized satisfiability problems. *J. Comput. Syst. Sci.*, 51(3):511–522, 1995. 3

[CSWY01]   Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 270–278. IEEE, 2001. 4

[GK73]   Peter Gács and János Körner. Common information is far less than mutual information. *Problems of Control and Information Theory*, 2(2):149–162, 1973. 5

[Gol10]   Oded Goldreich. *Property testing: current research and surveys*, volume 6390. Springer Science & Business Media, 2010. 3

[HSZZ06]   Wei Huang, Yaoyun Shi, Shengyu Zhang, and Yufan Zhu. The communication complexity of the hamming distance problem. *Information Processing Letters*, 99(4):149–153, 2006. 3

[HW07]   Johan Håstad and Avi Wigderson. The randomized communication complexity of set disjointness. *Theory of Computing*, 3(1):211–219, 2007. 3

[JK10]   Rahul Jain and Hartmut Klauck. The partition bound for classical communication complexity and query complexity. In *Computational Complexity (CCC), 2010 IEEE 25th Annual Conference on*, pages 247–258. IEEE, 2010. 34

[JKS08]   TS Jayram, Ravi Kumar, and D Sivakumar. The one-way communication complexity of hamming distance. *Theory of Computing*, 4(1):129–135, 2008. 3, 11

[KLL⁺12] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 500–509. IEEE, 2012. 34

[KN97] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997. 3, 7

[KOR00] Eyal Kushilevitz, Rafail Ostrovsky, and Yuval Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. *SIAM J. Comput.*, 30(2):457–474, 2000. 19

[KS92] Bala Kalyanasundaram and Georg Schintger. The probabilistic communication complexity of set intersection. *SIAM Journal on Discrete Mathematics*, 5(4):545–557, 1992. 3

[KSTW00] Sanjeev Khanna, Madhu Sudan, Luca Trevisan, and David P. Williamson. The approximability of constraint satisfaction problems. *SIAM J. Comput.*, 30(6):1863–1920, 2000. 3

[KW90] Mauricio Karchmer and Avi Wigderson. Monotone circuits for connectivity require super-logarithmic depth. *SIAM Journal on Discrete Mathematics*, 3(2):255–265, 1990. 3

[LS09] Troy Lee and Adi Shraibman. *Lower bounds in communication complexity*. Now Publishers Inc, 2009. 3

[Mau93] Ueli M Maurer. Secret key agreement by public discussion from common information. *Information Theory, IEEE Transactions on*, 39(3):733–742, 1993. 5

[MO05] Elchanan Mossel and Ryan O'Donnell. Coin flipping from a cosmic source: On error correction of truly random bits. *Random Structures & Algorithms*, 26(4):418–436, 2005. 5

[MOR⁺06] Elchanan Mossel, Ryan O'Donnell, Oded Regev, Jeffrey E Steif, and Benny Sudakov. Non-interactive correlation distillation, inhomogeneous markov chains, and the reverse bonami-beckner inequality. *Israel Journal of Mathematics*, 154(1):299–336, 2006. 5

[New91] Ilan Newman. Private vs. common random bits in communication complexity. *Information processing letters*, 39(2):67–71, 1991. 8

[PEG86] King F Pang and Abbas El Gamal. Communication complexity of computing the hamming distance. *SIAM Journal on Computing*, 15(4):932–947, 1986. 3

[Raz92] Alexander A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992. 3

[Raz03] Alexander A Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145, 2003. 3, 23

[Ros73] A.L. Rosenberg. On the time required to recognize properties of graphs: A problem. *SIGACT News*, 5, 1973. 3

[RW05] Renato Renner and Stefan Wolf. Simple and tight bounds for information reconciliation and privacy amplification. In *Advances in Cryptology-ASIACRYPT 2005*, pages 199–216. Springer, 2005. 5

[Sch78] Thomas J. Schaefer. The complexity of satisfiability problems. In Richard J. Lipton, Walter A. Burkhard, Walter J. Savitch, Emily P. Friedman, and Alfred V. Aho, editors, *Proceedings of the 10th Annual ACM Symposium on Theory of Computing, May 1-3, 1978, San Diego, California, USA*, pages 216–226. ACM, 1978. 3

[Sha48]   Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 1948. 8

[She11]   Alexander A Sherstov. The unbounded-error communication complexity of symmetric functions. *Combinatorica*, 31(5):583–614, 2011. 3, 23

[She12]   Alexander A Sherstov. The communication complexity of gap hamming distance. *Theory of Computing*, 8(1):197–208, 2012. 3, 6, 34

[ST13]   Mert Saglam and Gábor Tardos. On the communication complexity of sparse set disjointness and exists-equal problems. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on*, pages 678–687. IEEE, 2013. 3

[Sud10]   Madhu Sudan. Invariance in property testing. In Oded Goldreich, editor, *Property Testing - Current Research and Surveys [outgrow of a workshop at the Institute for Computer Science (ITCS) at Tsinghua University, January 2010]*, volume 6390 of *Lecture Notes in Computer Science*, pages 211–227. Springer, 2010. 3

[Vid11]   Thomas Vidick. A concentration inequality for the overlap of a vector on a large set. 2011. 3, 6, 34

[Wei15]   Omri Weinstein. Information complexity and the quest for interactive compression. (TR15-060), 2015. 9

[Wit75]   Hans S Witsenhausen. On sequences of pairs of dependent random variables. *SIAM Journal on Applied Mathematics*, 28(1):100–113, 1975. 5

[Woo04]   David Woodruff. Optimal space lower bounds for all frequency moments. In *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 167–175. Society for Industrial and Applied Mathematics, 2004. 3

[Yao79]   Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 209–213. ACM, 1979. 3

[Yao03]   Andrew Chi-Chih Yao. On the power of quantum fingerprinting. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 77–81. ACM, 2003. 3

[ZS09]   Zhiqiang Zhang and Yaoyun Shi. Communication complexities of symmetric xor functions. *Quantum Information & Computation*, 9(3):255–263, 2009. 3, 23