THE UNIVERSITY OF CHICAGO

COMMUNICATION COMPLEXITY AND INFORMATION COMPLEXITY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF COMPUTER SCIENCE

BY
DENIS PANKRATOV

CHICAGO, ILLINOIS
JUNE 2015

ABSTRACT

Shannon introduced information theory in 1948. In Shannon's model, the central question is to minimize the number of bits required to transmit a message from one party to another over a (possibly noisy) communication channel. Yao introduced communication complexity in 1979. In Yao's model, the central question is to minimize the amount of communication required to compute a function of an input distributed between two parties that communicate over a perfect channel. In spite of the fact that communication complexity and information theory try to quantify communication in various contexts, communication complexity developed without the influence of information theory. This changed recently when the notion of information complexity was introduced. The current definition of *internal information complexity* is due to Bar-Yossef et al. (2004), but similar definitions in restricted communication settings can be traced back to Chakrabarti et al. (2001) and Ablayev (1996).

Information complexity enabled the use of information-theoretic tools in communication complexity theory. Prior to the results presented in this thesis, information complexity was mainly used for proving lower bounds and direct-sum theorems in the setting of communication complexity. We present three results that demonstrate new connections between information complexity and communication complexity.

In the first contribution we thoroughly study the information complexity of the smallest nontrivial two-party function: the AND function. While computing the communication complexity of AND is trivial, computing its exact information complexity presents a major technical challenge. In overcoming this challenge, we reveal that information complexity gives rise to rich geometrical structures. Our analysis of information complexity relies on new analytic techniques and new characterizations of communication protocols. We also uncover a connection of information complexity to the theory of elliptic partial differential equations. Once we compute the exact information complexity of AND, we can compute *exact communication complexity* of

several related functions on n-bit inputs with some additional technical work. For instance, we show that communication complexity of the disjointness function on n-bit inputs is $C_{\mathrm{DISJ}}n + o(n)$, where $C_{\mathrm{DISJ}} \approx 0.4827$. This level of precision, i.e., revealing the actual coefficient in front of n, is unprecedented in communication complexity. Previous combinatorial and algebraic techniques could only prove bounds of the form $\Theta(n)$. Interestingly, this level of precision is typical in the area of information theory, so our result demonstrates that this meta-property of precise bounds carries over to information complexity and in certain cases even to communication complexity. Our result does not only strengthen the lower bound on communication complexity of disjointness by making it more exact, but it also shows that information complexity provides the exact upper bound on communication complexity. In fact, this result is more general and applies to a whole class of communication problems.

In the second contribution, we use self-reduction methods to prove strong lower bounds on the information complexity of two of the most studied functions in the communication complexity literature: Gap Hamming Distance (GHD) and Inner Product mod 2 (IP). In our first result we affirm the conjecture that the information complexity of GHD is linear even under the *uniform distribution*. This strengthens the $\Omega(n)$ bound shown by Kerenidis et al. (2012) and answers an open problem by Chakrabarti et al. (2012). We also prove that the information complexity of IP is arbitrarily close to the trivial upper bound n as the permitted error tends to zero, again strengthening the $\Omega(n)$ lower bound proved by Braverman and Weinstein (2011). More importantly, our proofs demonstrate that self-reducibility makes the connection between information complexity and communication complexity lower bounds a two-way connection. Whereas numerous results in the past used information complexity techniques to derive new communication complexity lower bounds, we explore a generic way, in which communication complexity lower bounds imply information complexity lower bounds *in a black-box manner*.

In the third contribution we consider the roles that private and public randomness play in the definition of information complexity. In communication complexity, private randomness can be trivially simulated by public randomness. Moreover, the communication cost of simulating public randomness with private randomness is well

understood due to Newman's theorem (1991). In information complexity, the roles of public and private randomness are reversed: public randomness can be trivially simulated by private randomness. However, the information cost of simulating private randomness with public randomness is not understood. We show that protocols that use only public randomness admit a rather strong compression. In particular, efficient simulation of private randomness by public randomness would imply a version of a direct sum theorem in the setting of communication complexity. This establishes a yet another connection between the two areas.

The first and second contributions are the result of collaboration with Braverman, Garg, and Weinstein. The third contribution is my work alone.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor László (Laci) Babai for guiding me through my PhD program. I am grateful to Laci for his valuable support, advice and feedback regarding all aspects of my PhD work from writing papers to applying for postdocs. I am also thankful to Laci for giving me the freedom to explore many different projects in theoretical computer science.

I feel extremely fortunate to have had the opportunity of an extended collaboration with Mark Braverman and his students Omri Weinstein and Ankit Garg. Mark is a leading expert in information complexity, and I was lucky to learn about this fascinating subject from him firsthand. A lot of the technical work in my thesis is based on the joint work with Mark, Ankit, and Omri. Mark is an inspiring figure for me with his infinite capacity to work and his generosity in sharing his ideas with us. He continued to care for me even when he was diagnosed with frightening illness (from which he is, to our great relief, recovering). I am also grateful to Mark for including me in many discussions that he had with his students long after our initial projects had been completed.

I would like to thank Sasha Razborov for introducing me to Mark Braverman and inviting Mark to give a talk at the theory seminar at the University of Chicago. This started my fruitful collaboration with Mark and his students. This would not have been possible without the attention and care that Sasha pays to all theory students at the University of Chicago. As an example, on another occasion Sasha took me to a proof complexity workshop, which I enjoyed tremendously.

I am grateful to Madhur Tulsiani for our ongoing collaboration. Madhur has been an excellent teacher and a wonderful collaborator. He made me feel at home at the Toyota Technological Institute at Chicago (TTIC).

During my undergraduate studies at the University of Toronto I had the good fortune of meeting Allan Borodin, who became my undergraduate research advisor.

Working with Allan sparked my interests in the area of theoretical computer science. When I was choosing graduate schools, Allan recommended that I apply and go to the University of Chicago. I followed Allan's advice and never regretted it.

During my PhD studies I have made friends with fellow graduate students Pooya Hatami and Mrinalkanti (Mrinal) Ghosh. We had many wonderful discussions over the years. Some of these discussions led to improvements in this thesis.

I am grateful to the Natural Sciences and Engineering Research Council of Canada for providing additional financial support through Postgraduate Scholarship during my first year of studies at the University of Chicago. I would like to express my gratitude to Bryan and Catherine Daniels whose Outstanding Student Fellowship program provided extra funding during three quarters of my PhD studies. I am also grateful to the Physical Sciences Division at the University of Chicago for providing extra support through William Rainey Harper Dissertation Fellowship during the last year of my PhD studies.

I am happy to acknowledge the constant support and encouragement I received from my family during the arduous years of my struggle with information complexity. The unconditional love of my dear parents, Galina Pankratova and Andrey Pankratov, as well as their unshakable belief in my own potential helped me immensely throughout my PhD program. My wife, Altynai (Aya) Omurbekova, has been with me during the last three years of my studies, and she has made them the best years of my life.

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGEMENTS iv

1 INTRODUCTION 1
- 1.1 Background and Our Contributions . . . 1
- 1.2 Notation . . . 4
- 1.3 Communication Complexity . . . 5
- 1.4 Information Theory . . . 8
- 1.5 Information Complexity . . . 10
- 1.6 Discussion of Generalizations of Communication Complexity and Information Complexity . . . 13

2 COMMUNICATION COMPLEXITY BOUNDS VIA INFORMATION COMPLEXITY 15
- 2.1 Introduction . . . 15
- 2.2 Main Results of this Chapter . . . 18
- 2.3 Characterization of Information Complexity via Local Concavity Constraints . . . 21
- 2.4 Continuity of $\mathrm{IC}_\mu(f,\epsilon)$ at $\epsilon = 0$ . . . 25
- 2.5 Information Complexity of AND with Zero Error Tolerance . . . 34
  - 2.5.1 Summary of Results for AND . . . 34
  - 2.5.2 Random Walk View of a Protocol: Distribution on Distributions and Splitting Lemmas . . . 37
  - 2.5.3 Information-Optimal Protocol for AND . . . 40
  - 2.5.4 Regions of $\Delta(\{0,1\}\times\{0,1\})$ for the AND Function . . . 42
  - 2.5.5 Internal Information Cost: Upper Bound . . . 46
  - 2.5.6 Internal Information Cost: Lower Bound . . . 49
  - 2.5.7 External Information Cost: Upper Bound . . . 50
  - 2.5.8 External Information Cost: Lower Bound . . . 52
- 2.6 Partial Differential Equation Formulation of Information Complexity 54
  - 2.6.1 Information Complexity of AND under Product Distributions via PDEs . . . 54
  - 2.6.2 System of PDEs for Information Complexity of General Functions 59
- 2.7 Rate of Convergence of $\mathrm{IC}^r_\mu(\mathrm{AND},0)$ to $\mathrm{IC}_\mu(\mathrm{AND},0)$ . . . 61
  - 2.7.1 Lower Bound on the Rate of Convergence . . . 63

2.7.2 Informational Wastage 65
2.7.3 Distance Traveled in the Wrong Region 68
2.7.4 Upper Bound on the Rate of Convergence 72
2.8 Communication Complexity of $\bigvee$-type Functions 76
2.8.1 Lower Bound on Communication Complexity of $\bigvee$-type Functions 81
2.8.2 Upper Bound on Communication Complexity of $\bigvee$-type Functions 83
2.8.3 Application: Exact Communication Complexity of DISJ_n . . . 87
2.9 Exact Communication Complexity of DISJ_n^k 88
2.9.1 Lower Bound 88
2.9.2 Upper Bound 95

3 INFORMATION COMPLEXITY BOUNDS VIA COMMUNICATION COMPLEXITY 103
3.1 Introduction 103
3.2 Main Results of this Chapter 104
3.3 Information Complexity of Gap Hamming Distance 105
3.3.1 Information Complexity of Small-Gap Instances 106
3.3.2 The Reduction from a Small-Gap instance to a Large-Gap instance 111
3.4 Information Complexity of Inner Product 115

4 PUBLIC VS. PRIVATE RANDOMNESS IN INFORMATION COMPLEXITY 117
4.1 Introduction 117
4.2 Efficient Simulation of Private Randomness with Public Randomness Leads to Strong Compression 119

5 CONCLUSIONS 125
5.1 Open Problems 125
5.2 Conclusions 126

REFERENCES 128

LIST OF FIGURES

1.1 Typical public-coin protocol 6

2.1 Partition of the space of product distributions into Alice's region, Bob's region, and the diagonal. 43
2.2 Graph of Alice's extension $\mathcal{I}_{\mathcal{A}}$. 58
2.3 Graph of Bob's extension $\mathcal{I}_{\mathcal{B}}$. 58
2.4 The combination of both extensions gives the actual information complexity. 58
2.5 Better view of the actual information complexity. 58
2.6 Empirical evidence that rate of convergence is $\Theta(1/r^2)$. The log-log scale figure shows the graph of $\max_{\mu \text{ - product}} \mathrm{IC}^r_\mu(f) - \mathrm{IC}_\mu(f)$ for a range of values r together with the line $1/(16r^2)$. The x-axis is the number of rounds r. The y-axis is the change in the information cost $\max_{\mu \text{ - product}} \mathrm{IC}^r_\mu(f) - \mathrm{IC}_\mu(f)$. 62

LIST OF TABLES

1 Notation x

Table 1: Notation

$[n]$	$\{1, 2, \ldots, n\}$
$\mathbb{R}^{\geq 0}$	the set of non-negative real numbers
$\mathrm{CC}(\pi)$	communication cost of protocol π
$\mathrm{R}(f, \epsilon)$	randomized (public randomness) communication complexity of f with error tolerance ϵ
$\mathrm{R}^{\mathrm{priv}}(f, \epsilon)$	randomized (private randomness) communication complexity of f with error tolerance ϵ
$\mathrm{R}^{r}(f, \epsilon)$	r-round randomized (private randomness) communication complexity of f with error tolerance ϵ
$\mathrm{D}_{\mu}(f, \epsilon)$	distributional communication complexity of f with error tolerance ϵ when inputs are sampled from μ
$\Pi(x, y)$	random variable equal to the concatenation of public randomness with a transcript of Π on input (x, y)
$\mathrm{IC}_{\mu}(\pi)$	internal information cost of protocol π with respect to distribution μ
$\mathrm{IC}_{\mu}^{\mathrm{ext}}(\pi)$	external information cost of protocol π with respect to distribution μ
$\mathrm{IC}_{\mu}(f, \epsilon)$	internal information complexity of f with respect to distribution μ and with error tolerance ϵ
$\mathrm{IC}_{\mu}^{\mathrm{ext}}(\pi)$	external information complexity of f with respect to distribution μ and with error tolerance ϵ
$\mathrm{IC}_{\mu}^{r}(f, \epsilon)$	r-round information complexity of f with respect to distribution μ and with error tolerance ϵ
$\mathrm{IC}_{\mu}^{\mathrm{priv}}(f, \epsilon)$	private-coin information complexity of f with respect to distribution μ and with error tolerance ϵ
$\mathrm{IC}_{\mu}^{\mathrm{pub}}(f, \epsilon)$	public-coin information complexity of f with respect to distribution μ and with error tolerance ϵ
$\Delta(S)$	family of all probability distributions on set S
$H(\mu)$	Shannon entropy (base 2) of a distribution μ
$I(\mu_1; \mu_2)$	mutual information between distributions μ_1 and μ_2
$\mathbb{D}(\mu_1 \| \mu_2)$	Kullback-Leibler divergence between distributions μ_1 and μ_2
$\|\mu_1 - \mu_2\|$	total variation distance between distributions μ_1 and μ_2

CHAPTER 1
INTRODUCTION

1.1 Background and Our Contributions

Shannon introduced information theory in the late 1940's [48]. Shannon was interested in *the transmission problem*: "given a (possibly noisy) channel between two players, what is the minimum amount of communication required to transmit a message X from one player to another reliably?" In this setting, the message X comes from a distribution known to both players, i.e., X is a random variable. Shannon introduced the entropy function $H(X)$ to measure the amount of information contained in the random variable X. In the case of a noiseless channel, Shannon proved that $H(X)$ is the exact solution to the transmission problem in the limit. In other words, Shannon's source-coding theorem states that for transmitting a sequence of messages $x_1, x_2, \ldots$, where the x_i are independently distributed according to X, $H(X)$ is the necessary and sufficient per-message amount of communication in the limit. Shannon also introduced the mutual information function $I(X;Y)$ that measures the amount of information shared between two random variables X and Y. The action of a noisy channel can be interpreted as modifying the sender's distribution X into the receiver's distribution Y. In the case of a noisy channel, Shannon showed that the best rate of a reliable transmission through a noisy channel is $I(X;Y)$ in the limit. In this thesis we shall work exclusively with noiseless channels, but the mutual information function can be applied to models other than noisy channel transmission. In particular, mutual information will play a central role in our work.

Since its introduction in the 1940s, information theory has been applied to and studied in many areas of the natural and social sciences and technology. With regards to the transmission problem, one of the early results of great theoretical and practical significance is Huffman coding [27], which establishes the single-copy version of the

source-coding theorem. Huffman showed a method for encoding a single copy of a message $x \sim X$ with at most $H(x)+1$ bits on average. Among other notable achievements, the Slepian-Wolf theorem [51] proves the analogue of the source-coding theorem in the presence of shared information between the two players. In spite of some remaining open problems, the transmission problem between two players is fairly well understood with information-theoretic quantities giving the exact bounds in many settings. While, in principle, the same results could have been obtained by combinatorial arguments, information-theoretic arguments often lead to especially elegant solutions. This suggests that information theory is "the right" tool for analyzing the transmission problem.

Yao introduced communication complexity in the late 1970s [52]. Yao was interested in the minimum amount of communication that two players have to perform in order to compute a function of a distributed input. Communication complexity has two main features that make it especially important in complexity theory in general. First of all, many problems of communication complexity turn out to be at the heart of problems in a variety of models of computation. For example, such connections allow carrying over lower bounds from communication complexity to streaming algorithms [2], data structures [39], property testing [6], circuit complexity [31, 5], and extended formulations [11]. Secondly, communication complexity is amenable to various strong and unconditional lower bound techniques [34, 33].

In spite of the fact that communication complexity and information theory try to quantify communication in various contexts, communication complexity developed without the influence of information theory until the late 1990s and the early 2000s. Most of the lower bound techniques on communication complexity of functions $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ were based on analyzing various combinatorial and analytic measures of communication matrices. The communication matrix associated with f is an $|\mathcal{X}| \times |\mathcal{Y}|$ matrix M defined by $M_{x,y} = f(x, y)$. The modern notion of internal information complexity was introduced by Bar-Yossef et al. [3]. Earlier versions of information complexity in restricted communication models appear in [1, 20]. Information complexity is a lower bound on communication complexity and yields an approach that is drastically different from the previous lower bound methods. In

particular, it does not refer to properties of the communication matrix; instead, information complexity is defined purely in terms of information-theoretic quantities. More specifically, information complexity is defined as the least amount of information (as measured by the Shannon's mutual information) that a communication protocol leaks about players' inputs.

Information complexity is a particularly useful tool for proving lower bounds on communication complexity, because it has the direct sum property [12, 7], and it is among the strongest lower bound methods on communication complexity [32, 25, 35]. The direct sum property asserts that the information complexity of n independent copies of a communication problem is exactly equal to n times the information complexity of a single copy of the problem. Often, communication problems have the structure that allows the direct sum property to be applied [3, 29, 4]. This leads to a surprising reduction: in order to prove a lower bound on the communication complexity of a size-n problem, one has to prove a lower bound on the information complexity of a related *constant size problem*. In addition, information complexity is increasingly recognized as an interesting measure in its own right [7, 35, 14]. In this thesis we explore three new connections between information complexity and communication complexity, which we describe next. The first and second contributions are based on the joint work with Braverman, Garg, and Weinstein.

One interesting feature of information theory is that it gives tight precise bounds on rates and capacities for the transmission problems. In fact, unlike computational complexity, where we often ignore constant, and sometimes even polylogarithmic, factors, a large fraction of results in information theory give precise answers up to additive lower-order terms. For example, we know that sending a sequence of random digits would take exactly $\log_2 10 \approx 3.322$ bits per digit, and that the capacity of a binary symmetric channel with substitution probability 0.2 is exactly $1 - H(0.2) \approx 0.278$ bits per symbol. Prior to this work, this benefit has not been fully realized in the setting of information complexity. In our first contribution, we explore and develop analytic machinery needed to bring tight bounds into the realm of information and communication complexity. We show that information complexity provides exact matching upper and lower bounds on the communication complexity

of $\bigvee$-type functions (see Section 2.8). As an application, we use these tools to calculate the tight communication complexity of the set disjointness function and several other related functions. The main technical ingredient is the exact computation of information complexity of the smallest non-trivial communication problem – the two-bit AND : $\{0,1\} \times \{0,1\} \to \{0,1\}$ function. While computing the communication complexity of AND is trivial, computing its exact information complexity is a major technical challenge. We overcome this challenge, and in the process we uncover an interesting geometric structure of the information complexity function, we describe new characterizations of communication protocols, and uncover connections of information complexity with the theory of elliptic partial differential equations.

In our second contribution, we explore generic techniques for deriving information complexity lower bounds from communication complexity lower bounds for self-reducible functions. Informally, f has a self-reducible structure if f on inputs of length nk (which we denote by f_{nk}) reduces to k independent copies of f under inputs of size n (which we denote by f_n^k). Our technique can be summarized in the following argument by contradiction. Let C_{nk} be a lower bound on the communication complexity of f_{nk}. From self-reducibility if follows that C_{nk} is a lower bound on f_n^k. If information complexity of f_n is $I_n << C_{nk}/k$ then, by "information=amortized communication" (see [12]), we can construct a protocol for f_n^k with communication $\approx kI_n << C_{nk}$ – contradiction. We apply this reasoning to two functions: Gap Hamming Distance, and Inner Product mod 2. Chakrabarti and Regev [19] proved that the randomized communication complexity of the Gap Hamming Distance problem is linear and that the distributional communication complexity is linear under the uniform distribution. Kerenidis et al. [32] proved that the information complexity of Gap Hamming Distance is also linear with respect to some implicitly defined distribution. Our techniques allow us to use the result of Chakrabarti and Regev [19] in a black-box manner to prove that the information complexity of the Gap Hamming Distance problem is linear with respect to the uniform distribution – this was explicitly stated as an open problem by Chakrabarti et al. [18]. We also show that the information complexity of the Inner Product function gets arbitrarily close to the trivial upper bound as the error tolerance goes to zero. Ideas of self-reducibility play

a central role in applications of information complexity to communication complexity lower bounds, starting with the work of Bar-Yossef et al. [3]. These arguments start with an information complexity lower bound for a (usually very simple) problem, and derive a communication complexity bound on many copies of the problem. Here, we start with a communication complexity lower bound, which we use as a black-box, and use self-reducibility to derive an information complexity lower bound.

In our third contribution, we explore the role played by private randomness in the definition of information complexity. In communication complexity, private randomness can always be simulated by private randomness without increasing communication, thus public randomness is more powerful. Moreover, the best general simulation of public randomness by private randomness is well understood due to Newman [40]. In information complexity, the situation is reversed – public randomness can always be simulated by private randomness without increasing information, thus private randomness is more powerful. We show that private randomness plays an important role in the compression problem: "given a protocol π with information cost I and communication cost C, what is the least communication cost of a protocol that simulates π?" In [4], it was shown that it is possible to compress protocols that use public and private randomness to $\widetilde{O}(\sqrt{IC})$ communication. In the third contribution, we show that if a protocol does not use private randomness, then it can be compressed to $\widetilde{O}(I)$ communication. In particular, this implies that any result separating information complexity from communication complexity, such as [25], has to crucially rely on using private randomness. Another implication is that an efficient simulation of private randomness with public randomness in terms of information complexity would imply a version of a direct sum theorem in communication complexity.

The rest of the thesis is organized as follows. This chapter introduces the notation and the main definitions of communication complexity, information theory, and information complexity that are used throughout the rest of the thesis. Chapter 2 contains all the technical details of our first contribution, as described above. Chapter 3 describes our second contribution, and Chapter 4 describes the third contribution. We conclude with open problems and final remarks in Chapter 5.

1.2 Notation

Capital letters $(A, B, C, \ldots)$ denote random variables. The corresponding lower-case letters $(a, b, c, \ldots)$ denote specific values attained by the random variables. For $n \in \mathbb{N}$ define $[n] = \{1, 2, \ldots, n\}$.

Let μ be a distribution on a product of two sets $\mathcal{X} \times \mathcal{Y}$. We shall think of μ in two ways:

1. as a function $\mu : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$,

2. as a matrix μ with rows indexed by $\mathcal{X}$ and columns indexed by $\mathcal{Y}$.

Due to the matrix interpretation of the distribution μ, we shall use linear algebra terms with respect to μ whenever it is convenient to do so. For instance, we shall sometimes refer to the transpose of μ, denoted by μ^T, understanding it to be the distribution given by the transpose of the matrix corresponding to μ.

Example 1.2.1. Consider $\mathcal{X} = \mathcal{Y} = \{0, 1\}$. Viewing μ as a function, it is specified by four numbers $\mu(0,0) = \alpha, \mu(0,1) = \beta, \mu(1,0) = \gamma, \mu(1,1) = \delta$, where $\alpha, \beta, \gamma, \delta \geq 0$ and $\alpha + \beta + \gamma + \delta = 1$. We can also think of μ as a matrix by arranging these four numbers as follows:

$$\mu = \begin{array}{|c|c|} \hline \alpha & \beta \\ \hline \gamma & \delta \\ \hline \end{array}.$$

Let $t \in \{0,1\}^n$ and $i \in [n]$ define $t_{\leq i} = t_1 t_2 \ldots t_i$, $t_{<i} = t_1 t_2 \ldots t_{i-1}$, $t_{\geq i} = t_i t_{i+1} \ldots t_n$, and $t_{>i} = t_{i+1} t_{i+2} \ldots t_n$. Out of bound indexes correspond to empty strings. The same notation is used with random variables whose values are binary strings.

1.3 Communication Complexity

The two-party communication model was introduced by Andrew Yao [52] in 1979. In this model, two players, Alice and Bob, attempt to compute the value of $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ at (x, y). Each player is computationally unbounded and each player knows f; however, Alice only knows $x \in \{0,1\}^n$ and Bob only knows

$y \in \{0,1\}^n$. In order to compute $f(x,y)$, Alice and Bob communicate according to a protocol π, which they agree upon before seeing the input. Protocol π specifies as a function of transmitted bits whether the communication is over and, if not, who sends the next bit. Moreover, π specifies as a function of transmitted bits and x the value of the next bit to be sent by Alice. Similarly for Bob. The communication is over when *both parties* know the value of $f(x,y)$.

Definition 1.3.1. *The transcript* of a protocol π on input (X,Y) is the concatenation of all bits exchanged during the execution of π on input (X,Y).

Definition 1.3.2. *The communication cost* of a protocol π, denoted by $\mathrm{CC}(\pi)$, is the maximum length of a transcript of π, where the maximum is taken over all possible inputs.

Definition 1.3.3. *The number of rounds* of a protocol is the maximum number of alternations between Alice and Bob, where the maximum is taken over all the transcripts.

This completes the description of the *deterministic 0-error communication model*.

The above model can be extended to include randomness in several ways. In the *public-coin model*, Alice and Bob have access to a shared random string R. Now the protocol π specifies the next bit to be sent by Alice as a function of x, the already transmitted bits, and the random string R. Similarly for Bob. The communication cost of a public-coin protocol is defined as the maximum total number of bits sent by the players, where the maximum is taken over the choice of inputs and the choice of randomness. A public-coin protocol can be viewed as a distribution on deterministic protocols. The players may terminate prior to knowing the exact value of $f(x,y)$ and output a consistent guess. In this case we measure the probability (over public randomness) of players outputting an incorrect value. Figure 1.1 depicts a typical public-coin protocol.

Definition 1.3.4. Fix a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and a parameter $\epsilon \geq 0$. *The randomized communication complexity* of f with error tolerance ϵ, denoted by

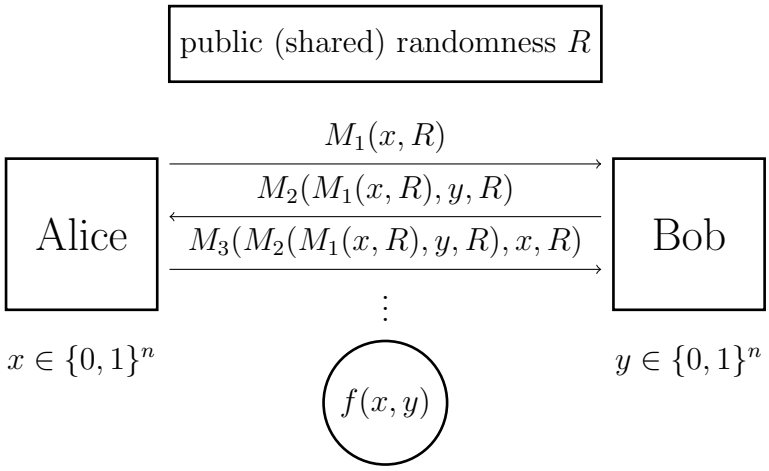

Figure 1.1: Typical public-coin protocol

$\mathrm{R}(f, \epsilon)$, is the cost of a least-cost public-coin protocol that computes f with error at most ϵ on every input.

Definition 1.3.5. Fix a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and a parameter $\epsilon \geq 0$. *The r-round randomized communication complexity* of f with error tolerance ϵ, denoted by $\mathrm{R}^r(f, \epsilon)$, is the cost of a least-cost r-round public-coin protocol that computes f with error at most ϵ on every input.

In the *private-coin model*, Alice has access to a random string $R_{\mathcal{A}}$ hidden from Bob, and Bob has access to a random string $R_{\mathcal{B}}$ hidden from Alice. As before, private-coin protocol π specifies as a function of transmitted bits whether the communication is over and, if not, who sends the next bit. Now, π specifies as a function of transmitted bits, x, and $R_{\mathcal{A}}$ next bit to be sent by Alice. Similarly for Bob. Players may terminate prior to knowing the exact value of $f(x, y)$ and output a consistent guess. In this case we measure the probability (over private randomness) of players outputting an incorrect value.

Definition 1.3.6. Fix a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and a parameter $\epsilon \geq 0$. *The randomized communication complexity with private randomness* of f with error tolerance ϵ, denoted by $\mathrm{R}^{\mathrm{priv}}(f, \epsilon)$, is the cost of a least-cost private-coin protocol that computes f with error at most ϵ on every input.

Private-coin protocols are in one-to-one correspondence with binary trees that have additional structure, which we describe next. Each node u of T has an owner - Alice or Bob. Each node u of T has an associated function $p_u : \{0,1\}^n \to [0,1]$. Each leaf ℓ of T has a label $o_\ell \in \{0,1\}$. Algorithm 1 describes how to turn such a tree into a private-coin protocol. The key observation is that functions p_u specify the probability of a player sending 0 as the next bit, conditioned on the transcript so far. It is easy to see that this correspondence is bijective.

Algorithm 1 Converting a tree into a private-coin protocol

Require:

$x \in \{0,1\}^n, R_{\mathcal{A}}$ - known to Alice
$y \in \{0,1\}^n, R_{\mathcal{B}}$ - known to Bob
T - known to Alice and Bob

Both players set $u \leftarrow$ root of T
while u is not a leaf **do**
    if owner of u is Alice **then**
        Alice privately samples $r \in [0,1]$
        if $r \leq p_u(x)$ **then**
            Alice sends 0
            Both players update $u \leftarrow$ left child of u
        else
            Alice sends 1
            Both players update $u \leftarrow$ right child of u
    else
        Bob privately samples $r \in [0,1]$
        if $r \leq p_u(y)$ **then**
            Bob sends 0
            Both players update $u \leftarrow$ left child of u
        else
            Bob sends 1
            Both players update $u \leftarrow$ right child of u
Both players output o_u

From the point of view of communication complexity, once we allow public randomness, it makes no difference whether players have access to private random strings or not. This is because disjoint parts of the public random string can be designated for simulation of private randomness. However, we shall see later that for informa-

tion complexity it is crucial to consider protocols that use *both private and public* randomness. A protocol with both public and private randomness is a distribution on protocols with private randomness only.

Lastly, we consider the scenario when inputs are sampled from a probability distribution μ on $\{0,1\}^n \times \{0,1\}^n$.

Definition 1.3.7. Fix a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, a distribution μ on $\{0,1\}^n \times \{0,1\}^n$, and a parameter $\epsilon \geq 0$. *The distributional communication complexity* of f with respect to μ and error tolerance ϵ, denoted by $\mathrm{D}_\mu(f,\epsilon)$, is the cost of a least-cost deterministic protocol computing f with probability of error at most ϵ, where the probability is measured over inputs $(x,y) \sim \mu$.

The distributional and randomized communication complexities are related via Yao's minimax principle.

Theorem 1.3.1 (Yao's minimax principle)**.** $\forall f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}, \epsilon \geq 0$ *we have*

$$\mathrm{R}(f,\epsilon) = \max_\mu \mathrm{D}_\mu(f,\epsilon).$$

For more background on communication complexity we refer the interested reader to the excellent monograph by Kushilevitz and Nisan [33].

1.4 Information Theory

In this section we briefly provide the essential information-theoretic concepts. For a thorough introduction to the area of information theory, the reader should consult a classical textbook by Cover and Thomas [23]. Unless stated otherwise, all log's are to the base 2.

Let S be a set. We use $\Delta(S)$ to denote *the family of all probability distributions* on S.

Definition 1.4.1. Let μ be a probability distribution on a sample space Ω. *Shannon entropy* (or just *entropy*) of μ, denoted by $H(\mu)$, is defined as $H(\mu) := \sum_{\omega \in \Omega} \mu(\omega) \log \frac{1}{\mu(\omega)}$.

For a random variable A we shall write $H(A)$ to denote the entropy of the induced distribution on the range of A. The same also holds for other information-theoretic quantities appearing later in this section.

Definition 1.4.2. For the Bernoulli distribution with probability of success p we write $H(p) = -p \log p - (1-p)\log(1-p)$.

Definition 1.4.3. *Conditional entropy* of a random variable A conditioned on B is defined as

$$H(A|B) = \mathbb{E}_b(H(A|B=b)).$$

Fact 1.4.1. $H(AB) = H(A) + H(B|A)$.

Definition 1.4.4. The *mutual information* between two random variable A and B, denoted by $I(A;B)$ is defined as

$$I(A;B) := H(A) - H(A|B) = H(B) - H(B|A).$$

The *conditional mutual information* between A and B given C, denoted by $I(A;B|C)$, is defined as

$$I(A;B|C) := H(A|C) - H(A|BC) = H(B|C) - H(B|AC).$$

Fact 1.4.2 (Chain Rule)**.** *Let A_1, A_2, B, C be random variables. Then*

$$I(A_1A_2;B|C) = I(A_1;B|C) + I(A_2;B|A_1C).$$

Fact 1.4.3. *Let A, B, C, D be four random variables such that $I(B;D|AC) = 0$. Then*

$$I(A;B|C) \geq I(A;B|CD)$$

Fact 1.4.4. *Let A, B, C, D be four random variables such that $I(A;C|BD) = 0$. Then $I(A;B|D) \geq I(A;C|D)$*

Definition 1.4.5. Given two probability distributions μ_1 and μ_2 on the same sample space Ω such that $(\forall \omega \in \Omega)(\mu_2(\omega) = 0 \Rightarrow \mu_1(\omega) = 0)$, the *Kullback-Leibler divergence*

between is defined as

$$\mathbb{D}(\mu_1||\mu_2) = \sum_{\omega\in\Omega} \mu_1(\omega) \log \frac{\mu_1(\omega)}{\mu_2(\omega)}.$$

For random variables A and B we shall write $\mathbb{D}(A||B)$ for Kullback-Leibler divergence between the induced probability distributions on the range of A and B, respectively.

The connection between the mutual information and the Kullback-Leibler divergence is provided by the following fact.

Fact 1.4.5. *For random variables $A, B,$ and C we have*

$$I(A; B|C) = \mathbb{E}_{B,C}(\mathbb{D}(A_{BC}||A_C)).$$

Fact 1.4.6. *Let X and Y be random variables. Then for any random variable Z we have $\mathbb{E}_X[\mathbb{D}(Y_X||Y)] \leq \mathbb{E}_X[\mathbb{D}(Y_X||Z)]$.*

Definition 1.4.6. Let μ_1 and μ_2 be two probability distributions on the same sample space Ω. *Total variation distance* is defined as

$$\|\mu_1 - \mu_2\| := \frac{1}{2} \sum_{\omega\in\Omega} |\mu_1(\omega) - \mu_2(\omega)|.$$

For random variables A and B we shall write $\|A - B\|$ for the total variation distance between the induced probability distributions on the range of A and B, respectively.

Fact 1.4.7. $\|\mu_1 - \mu_2\| = \max_{\mathcal{S}\subseteq\Omega} |\mu_1(\mathcal{S}) - \mu_2(\mathcal{S})|$.

Fact 1.4.8 (Data Processing Inequality)**.** *Let A, B, C be random variables on the same sample space, and let D be a probabilistic function of B only. Then we have*

$$I(A; D|C) \leq I(A; B|C).$$

The above concepts were defined for the *discrete probability distributions*. In this work we shall also encounter continuous probability distributions. There are some

subtleties in going from discrete case to continuous case in the area of information theory; however, we shall not encounter these subtleties. For our purposes, the above definitions and facts generalize to the continuous case in a straightforward way.

For instance, Kullback-Leibler divergence between two continuous distributions over $\mathbb{R}$ given by their probability density functions (PDFs) p and q is defined as

$$\mathbb{D}(p||q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)}\, dx.$$

1.5 Information Complexity

In this section we consider protocols that have both public and private randomness. We also consider inputs X and Y being sampled from a joint distribution μ. Let $\Pi(X, Y)$ denote the random variable that is the concatenation of public randomness with a transcript of π on a random input (X, Y). Due to private randomness, $\Pi(x, y)$ remains a random variable even after particular inputs $X = x$ and $Y = y$ are fixed.

Definition 1.5.1. Fix a communication protocol π on inputs $\{0,1\}^n \times \{0,1\}^n$ and a distribution $\mu \in \Delta(\{0,1\}^n \times \{0,1\}^n)$. *The (internal) information cost* of π with respect to μ, denoted by $\mathrm{IC}_\mu(\pi)$, is defined as

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

Definition 1.5.2. Fix a communication protocol π on inputs $\{0,1\}^n \times \{0,1\}^n$ and a distribution $\mu \in \Delta(\{0,1\}^n \times \{0,1\}^n)$. *The external information cost* of π with respect to μ, denoted by $\mathrm{IC}^{\mathrm{ext}}_\mu(\pi)$, is defined as

$$\mathrm{IC}^{\mathrm{ext}}_\mu(\pi) := I(\Pi(X,Y); XY).$$

Definition 1.5.3. Fix a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, a distribution $\mu \in \Delta(\{0,1\}^n \times \{0,1\}^n)$, and a parameter $\epsilon \geq 0$. *The (internal) information complexity* of f with respect to μ and error tolerance ϵ, denoted by $\mathrm{IC}_\mu(f, \epsilon)$, is defined as

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

where the infimum ranges over all protocols π with public and private randomness solving f with error at most ϵ when inputs are sampled according to μ.
The external information complexity of f with respect to μ and error tolerance ϵ is defined analogously.

Definition 1.5.4. Fix a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, a distribution $\mu \in \Delta(\{0,1\}^n \times \{0,1\}^n)$, and a parameter $\epsilon \geq 0$. *The (internal) private-coin information complexity* of f with respect to μ and error tolerance ϵ, denoted by $\mathrm{IC}_\mu^{\mathrm{priv}}(f,\epsilon)$, is defined as

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

where the infimum ranges over all protocols π with private randomness only (no public randomness) solving f with error at most ϵ when inputs are sampled according to μ. *The external private-coin information complexity* of f with respect to μ and error tolerance ϵ is defined analogously.

Definition 1.5.5. Fix a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, a distribution $\mu \in \Delta(\{0,1\}^n \times \{0,1\}^n)$, and a parameter $\epsilon \geq 0$. *The (internal) public-coin information complexity* of f with respect to μ and error tolerance ϵ, denoted by $\mathrm{IC}_\mu^{\mathrm{pub}}(f,\epsilon)$, is defined as

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

where the infimum ranges over all protocols π with public randomness only (no private randomness) solving f with error at most ϵ when inputs are sampled according to μ. *The external public-coin information complexity* of f with respect to μ and error tolerance ϵ is defined analogously.

Definition 1.5.6. Fix a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, a distribution $\mu \in \Delta(\{0,1\}^n \times \{0,1\}^n)$, and parameters $\epsilon \geq 0$ and $r \in \mathbb{N}$. *The r-round information complexity* of a function f with respect to μ and error tolerance ϵ, denoted by $\mathrm{IC}_\mu^r(f,\epsilon)$, is defined as

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

where the infimum ranges over all r-round protocols π solving f with error at most ϵ

when inputs are sampled according to μ.
The r-round external information complexity is defined analogously.

The above notions of information complexity depend on the input distribution μ. We shall sometimes refer to the input distribution as *the prior distribution*. A straightforward way of defining *the prior-free* versions of information complexity notions is to take the maximum over distributions μ. For instance, the following is the definition of *the prior-free (internal) information complexity*

Definition 1.5.7. Fix a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and $\epsilon \in [0, 1/2]$. *The prior-free (internal) information complexity* of f with error tolerance ϵ is defined as

$$\mathrm{IC}(f, \epsilon) = \max_{\mu} \mathrm{IC}_{\mu}(f, \epsilon).$$

Prior-free notions of external information complexity, bounded-round information complexity, information complexity with public/private randomness only are defined analogously.

Fact 1.5.1. *For all $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}, \epsilon \geq 0, r \in \mathbb{N}$ we have*

$$\mathrm{IC}^{priv}_{\mu}(f, \epsilon) \geq \mathrm{IC}_{\mu}(f, \epsilon)$$

$$\mathrm{IC}^{pub}_{\mu}(f, \epsilon) \geq \mathrm{IC}_{\mu}(f, \epsilon)$$

$$\mathrm{IC}^{r}_{\mu}(f, \epsilon) \geq \mathrm{IC}_{\mu}(f, \epsilon)$$

We shall also need the following definition:

Definition 1.5.8. Let $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ be a function, μ be a distribution on $\{0,1\}^k \times \{0,1\}^k$ and $\epsilon \in [0,1]$ be a parameter. Define

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

where the infimum is taken over all protocols π such that for all inputs $(x, y) \in \{0,1\}^k \times \{0,1\}^k$ the probability that π makes a mistake on (x, y) is at most ϵ.

In particular the main difference of the above definition from $\mathrm{IC}_\mu(f,\epsilon)$ is that in the latter definition the error of the protocol is measured with respect to μ.

Next we describe how a protocol can be viewed as a random walk on the space of distributions for the purpose of information cost. Let π be a protocol on the input space $\{0,1\}^n \times \{0,1\}^n$. Let $\mu \in \Delta(\{0,1\}^n \times \{0,1\}^n)$ be a distribution. With each partial transcript $t \in \bigcup_{i=0}^{\mathrm{CC}(\pi)} \{0,1\}^i$ we can associate a distributions μ_t, where

$$\mu_t(x,y) = P_{(X,Y)\sim\mu}(X = x, Y = y | \Pi(X,Y)_{\leq |t|} = t).$$

Both players can compute the distribution μ_t given the partial transcript t. A particular μ_t represents the belief of the players about the distribution on the inputs after the communication transcript t. In defining μ_t we do not condition on any input of the players, since otherwise the players would not be able to consistently update their beliefs about the inputs. Thus, for the purpose of updating their beliefs, the players "forget" their actual input. The probability that players end up in a particular μ_t is equal to $P(\Pi(x,y)_{\leq |t|} = t)$. Then π can be viewed as a random walk on $\Delta(\{0,1\}^n \times \{0,1\}^n)$ as follows: both players set μ as the current location in $\Delta(\{0,1\}^n \times \{0,1\}^n)$. After each bit of communication, the players update their current location to μ_t, where t is the partial transcript so far. Once the players reach $|t| = \mathrm{CC}(\pi)$, they terminate the random walk.

The walk defined above is random for the following reason: suppose that t is the partial transcript so far, and Alice is about to send the next bit B. Then with probability $P(B = 0 | X = x, \Pi(X,Y)_{\leq |t|} = t)$ the players move from μ_t to μ_{t0} and with probability $P(B = 0 | X = x, \Pi(X,Y)_{\leq |t|} = t)$ the players move from μ_t to μ_{t1}. Note that if π solves some function f with 0 error, it means that the random walk must always terminate in a distribution μ_t such that $\mathrm{R}(f|_{\mathrm{supp}(\mu_t)}, 0) = 0$. In later sections, we shall rely heavily on this view of a protocol. In particular, we shall prove that the information cost of a protocol is a function of the distribution on final distributions μ_t only (where t is the entire transcript of π). In other words, it does not matter how the players perform the walk on the space $\Delta(\{0,1\}^n \times \{0,1\}^n)$ – as long as they end up in the same final distributions with the same probability – the information cost

remains the same.

1.6 Discussion of Generalizations of Communication Complexity and Information Complexity

In Sections 1.3 and 1.5 various notions of information and communication complexities were defined for Boolean functions on Boolean cubes. One obvious generalization is to consider functions of the form $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ with an arbitrary range $\mathcal{Z}$ and an arbitrary domain $\mathcal{X} \times \mathcal{Y}$. The definitions for this general case are obtained via trivial modifications of the definitions given in the above sections. We gave the definitions for Boolean functions on Boolean cubes for two reasons: (1) to keep the notation to the bare minimum, and (2) to concentrate on this setting, as it is the main setting for the applications of the information complexity theory. At times it becomes necessary to consider the more general setting in this thesis. For instance, the direct sum of a Boolean function f has a multiple-bit output and therefore is not Boolean. Working with Boolean functions on Boolean cubes is essentially without loss of generality, as all the definitions and results carry over to the general setting in a trivial manner. We shall continue to work in the Boolean setting whenever it is convenient.

Another generalization is to consider partial functions $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and (partial) relations $R \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. This generalization is much more subtle than the one described in the previous paragraph. A complexity measure can be significantly smaller for a partial function versus its total extension. Thus, when proving lower bounds, one has to be extra careful with partial functions. Therefore, we shall explicitly mention when we deal with partial functions in this thesis. Similar considerations apply to relations.

Another generalization is to consider a communication problem with input distributed among $k \geq 3$ players, a.k.a. multiparty communication. There are two main communication models for $k \geq 3$ players – "number-in-hand" (NIH) and "number-on-forehead" (NOF) models. For the definitions and general background we refer the interested reader to the book by Kushilevitz and Nisan [33] and references therein. NOF model is notoriously difficult with regards to obtaining strong lower bounds in

this model. In fact, one of the major open problems in information complexity theory is to *even define a reasonable notion* of information complexity in the NOF setting. We shall not discuss these models in this thesis beyond this. Other modifications of Yao's communication model, such as quantum communication complexity, are also beyond the scope of this thesis.

CHAPTER 2
COMMUNICATION COMPLEXITY BOUNDS VIA INFORMATION COMPLEXITY

2.1 Introduction

The results of this chapter are based on the joint work of the author with Braverman, Garg, and Weinstein and have appeared in [9].

In this chapter we shall study communication complexity and information complexity of several explicit functions. For easy reference, we list these functions here.

Definition 2.1.1. *The two-bit* AND *function*, denoted by $\mathrm{AND} : \{0,1\} \times \{0,1\} \to \{0,1\}$, is defined as

$$\mathrm{AND}(x, y) = x \wedge y.$$

Definition 2.1.2. *The (set) disjointness function*, denoted by $\mathrm{DISJ}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, is defined as

$$\mathrm{DISJ}_n(x, y) = \neg \bigvee_{i=1}^{n} x_i \wedge y_i.$$

Definition 2.1.3. *The k-set disjointness partial function* is denoted by DISJ_n^k. Let $S = \{x \in \{0,1\}^n \mid \text{ the Hamming weight of } x \text{ is at most } k\}$. The k-set disjointness is the partial function $\mathrm{DISJ}_n^k : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ such that $\mathrm{DISJ}_n^k|_{S\times S} = \mathrm{DISJ}_n|_{S\times S}$.

Definition 2.1.4. *The set intersection function*, denoted by $\mathrm{SETINT}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^n$, is the function with multiple bit output defined as follows

$$\mathrm{SETINT}_n(x, y) = (x_1 \wedge y_1, x_2 \wedge y_2, \ldots, x_n \wedge y_n).$$

The set disjointness function DISJ_n is one of the oldest and most studied functions in communication complexity [33]. In the set disjointness problem (see Definition 2.1.2), Alice and Bob are given $x, y \in \{0,1\}^n$, respectively, where x and y are characteristic vectors of subsets of $[n]$. The players need to output 1 if the sets corresponding to x and y do not intersect, and 0 otherwise. In the deterministic communication complexity model, it is easy to show that DISJ_n has communication complexity $n+1$. In the randomized communication complexity model – which is the focus of this chapter – an $\Omega(n)$ lower bound was first proved by Kalyanasundaram and Schnitger [30]. The proof was combinatorial in its nature. A much simpler combinatorial proof was given by Razborov a few years later [45]. In terms of upper bounds on the randomized communication complexity of disjointness, an $n+1$ bound is trivial. No better bound was known prior to [9]; although by examining the problem, one can directly convince oneself that there is a protocol for DISJ_n that uses only $(1-\epsilon)n$ communication for some small $\epsilon > 0$ – so that the deterministic algorithm is suboptimal. Another set of techniques which were successfully applied to show lower bounds on communication complexity of versions of disjointness, especially in the quantum and multiparty settings [46, 21, 50], are analytic techniques. Analytic techniques such as the pattern matrix method [49], allow one to further extend the reach of combinatorial techniques.

The first information-theoretic proof of the $\Omega(n)$ lower bound on the randomized communication complexity of DISJ_n was given by Bar-Yossef et al. [3]. While not materially improving the lower bound, the information-theoretic approach was extended to the multiparty number-in-hand setting [17, 29] with applications to tight lower bounds on streaming algorithms. At the core of the proof is a direct-sum reduction of proving an $\Omega(n)$ bound on DISJ_n to proving an $\Omega(1)$ bound on the information complexity of AND. The direct sum in this and other proofs follows from an application of the chain rule for mutual information – one of the primary information-theoretic tools. An information complexity view of disjointness lead to tight bounds on the ability of extended formulations by linear programs to approximate the CLIQUE problem [11]. This suggests that information complexity and a better understanding of the disjointness problem may have other interesting implications within computational

complexity.

The set intersection function SETINT_n (see Definition 2.1.4) is closely related to the disjointess function. Now instead of determining whether sets corresponding to x and y intersect, Alice and Bob need to learn the characteristic vector of the whole intersection $(x_1 \wedge y_1, x_2 \wedge y_2, \ldots, x_n \wedge y_n)$. For this problem, a lower bound of n bits on the communication is trivial even in the randomized setting. Fix $x = (1, 1, \ldots, 1)$ and observe that in this special case the set intersection problem will amount to Bob sending his input to Alice, which clearly requires $\geq n$ bits. Thus the randomized communication complexity of this problems lies somewhere between n and $2n$ – the trivial upper and lower bounds. Note that the intersection problem is nothing but n copies of the two-bit AND function. Therefore, determining the communication complexity of SETINT_n is equivalent to determining the information complexity of the two-bit AND function by the "information = amortized communication" connection [12].

Essentially independently of the communication complexity line of work described above, a study of the AND/intersection problem has recently originated in the information theory community. A series of papers by Ma and Ishwar [36, 38] develops techniques and characterizations which allow one to rigorously calculate tight bounds on the communication complexity of SETINT_n and other amortized functions on the condition that one only considers protocols restricted to r rounds of communication. These techniques allow one to numerically (and sometimes analytically) compute the information complexity of the two-bit AND function – although the numerical computation is not provably correct for the most general unbounded-round case since the rate of convergence of r-round information complexity down to the true information complexity is unknown. Furthermore, their results about the AND function are non-constructive in the sense that they do not exhibit a protocol achieving their bounds. Nonetheless, numerical calculations produced by Ma and Ishwar do point at convergence to 1.4923 bits for the AND function [28]. As discussed below, our tight upper and lower bounds are consistent with this evidence.

The main result of this chapter and [9] is giving tight bounds on the information and communication complexity of the AND, SETINT_n, and DISJ_n functions. Being able to obtain tight bounds is a benefit provided by information theory – one

that has been largely untapped by the communication complexity community. We give a (provably) information-theoretic optimal protocol for the two-bit AND function. This optimality gives a tight optimal randomized protocol for SETINT_n that uses $C_\wedge n \pm o(n)$ bits of communication and fails with a vanishing probability. Here $C_\wedge \approx 1.4923$ is an explicit constant given as a maximum of a concave analytic function. We then apply the same optimal result to obtain the optimal protocol for set disjointness, showing that the best vanishing error randomized protocol for DISJ_n will take $C_{\text{DISJ}} n \pm o(n)$ bits of communication, where $C_{\text{DISJ}} \approx 0.4827$ is another explicit constant (which we found to be surprisingly low). The fact that we need the bounds to be exact throughout requires us to develop some technical tools for dealing with information complexity in this context. For example, we show that unlike communication complexity, the randomized ϵ-error information complexity converges to the 0-error information complexity as $\epsilon \to 0$.

Applying what we have learned about the AND function to the *sparse sets* regime, we are able to determine the precise communication complexity of disjointness DISJ_n^k where the sets are restricted to be of size at most k (see Definition 2.1.3). Håstad and Wigderson [26] showed that the randomized communication complexity of this problem is $\Theta(k)$. We sharpen this result by showing that for vanishing error the communication complexity of DISJ_n^k is $\frac{2}{\ln 2} k \pm o(k) \approx 2.885k \pm o(k)$.

Interestingly the optimal protocol we obtain for AND is not an actual protocol in the strict sense of communication protocols definitions. One way to visualize it is as a game show where Alice and Bob both have access to a "buzzer" and the game stops when one of them "buzzes in". The exact time of the "buzz in" matters. If we wanted to simulate this process with a conventional protocol, we'd need the time to be infinitely quantized, with Alice and Bob exchanging messages of the form "no buzz in yet", until the buzz in finally happens. Thus the optimal information complexity of AND is obtained by an infimum of a sequence of conventional protocols rather than by a single protocol.

It turns out that the unlimited number of rounds is necessary, both for the AND function and for DISJ_n. Our understanding of information complexity in the context of the AND function allows us to lower bound the amount of communication needed

for DISJ_n if we restrict the number of rounds of interaction between players to r. $\mathrm{R}^r(\mathrm{DISJ}_n, \epsilon) \geq (C_{\mathrm{DISJ}} + \Omega(1/r^2)) \cdot n$. In particular, any constant bound on the number of rounds means a linear loss in communication complexity. There are well-known examples in communication complexity where adding even a single round causes an exponential reduction in the amount of communication needed [41]. There are also examples of very simple transmission problems where it can be shown that two rounds are much better than one, and more than two are better yet [42, 43]. However, to our knowledge, together with a very recent independently obtained result on rounds in the communication complexity of small set intersection [16, 47], this is the first example of a "natural" function where an arbitrary number of additional rounds is provably helpful.

2.2 Main Results of this Chapter

Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function. Our first contribution is a characterization of $\mathrm{IC}_\mu(f, 0)$ in terms of local concavity constrains as follows.

Definition 2.2.1 ([9])**.** Fix a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. Define $\mathfrak{C}(f)$ to be the family of functions $C : \Delta(\{0,1\}^n \times \{0,1\}^n) \to \mathbb{R}^{\geq 0}$ that satisfy the following constraints:

- for all $\mu \in \Delta(\{0,1\}^n \times \{0,1\}^n)$ if $\mathrm{R}(f|_{\mathrm{supp}(\mu)}, 0) = 0$ then $C(\mu) = 0$,
- for all $\mu, \mu_0^A, \mu_1^A \in \Delta(\{0,1\}^n \times \{0,1\}^n)$ if there exists a bit B that Alice can send starting from μ such that $P(B = 0) = P(B = 1) = 1/2$, $\mu_0^A(x, y) = P(X = x, Y = y | B = 0)$, and $\mu_1^A(x, y) = P(X = x, Y = y | B = 1)$ then

$$C(\mu) \leq C(\mu_0^A)/2 + C(\mu_1^A)/2 + I(X; B|Y),$$

- for all $\mu, \mu_0^B, \mu_1^B \in \Delta(\{0,1\}^n \times \{0,1\}^n)$ if there exists a bit B that Bob can send starting from μ such that $P(B = 0) = P(B = 1) = 1/2$, $\mu_0^B(x, y) = P(X =$

$x, Y = y|B = 0)$, and $\mu_1^B(x, y) = P(X = x, Y = y|B = 1)$ then

$$C(\mu) \leq C(\mu_0^B)/2 + C(\mu_1^B)/2 + I(Y; B|X).$$

- for all μ, $C(\mu) \leq 2n$.

Remark 2.2.1. The condition $\mathrm{R}(f|_{\mathrm{supp}(\mu)}, 0) = 0$ means that both parties can determine the function's output under μ by considering their own input only. This does not mean that the function is determined under μ from an external point of view. The example $f(0,0) = 0$, $f(1,1) = 1$, $\mu(0,0) = \mu(1,1) = 1/2$ illustrates this point.

Lemma 2.2.1 ([9])**.** *For all* $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *and for all* $\mu \in \Delta(\{0,1\}^n \times \{0,1\}^n)$ *we have*

$$\mathrm{IC}_\mu(f, 0) = \max_{C \in \mathfrak{C}(f)} C(\mu).$$

In particular, this lemma says that every $C \in \mathfrak{C}(f)$ is a lower bound on the zero-error information complexity of f. Of course, for a general function f it is not obvious how to find even a single nontrivial element of $\mathfrak{C}(f)$. However, if one guesses such an element C, it is easy to verify that $C \in \mathfrak{C}(f)$ by checking the local concavity constraints from Definition 2.2.1. In this chapter we consider $f = \mathrm{AND}$ and successfully apply the "guess-and-verify" approach as just described. In fact, the element $C \in \mathfrak{C}(\mathrm{AND})$ that we guess is defined by the zero-error information complexity of a certain protocol for AND. Thus, we immediately conclude that C is also an upper bound on the information complexity, so C must be equal to the information complexity of AND. This also implies that the protocol that gives rise to C is optimal for AND. Since we derive an explicit formula for C, we are able to prove the following results.

Theorem 2.2.2 ([9])**.**

$$\mathrm{IC}(\mathrm{AND}, 0) = C_\wedge \approx 1.4923$$
$$\mathrm{IC}^{ext}(\mathrm{AND}, 0) = \log_2 3 \approx 1.58496$$

We analyze the rate of convergence of $\mathrm{IC}^r_\mu(\mathrm{AND}, 0)$ to $\mathrm{IC}_\mu(\mathrm{AND}, 0)$, as the number of rounds r increases, and derive the following tight bound on the rate of convergence.

Theorem 2.2.3 ([9]). *Let μ be a distribution on $\{0,1\} \times \{0,1\}$ with full support. Then we have*

$$\mathrm{IC}^r_\mu(\mathrm{AND}, 0) = \mathrm{IC}_\mu(\mathrm{AND}, 0) + \Theta_\mu\left(\frac{1}{r^2}\right).$$

Moreover, the lower bound holds even for μ such that $\mu(1,1) = 0$.

The above exact information complexity results lead to exact communication complexity bounds for the disjointness function and its variants in the regime of error tending to 0. Suppose that we have a protocol π that computes $\mathrm{DISJ}_n = \neg \bigvee_{i \in [n]} x_i \wedge y_i$ correctly on all inputs. Then we can extract a protocol for the two-bit $\mathrm{AND}(x, y)$ from π by sampling inputs for coordinates $j \in [n] - \{i\}$ randomly, embedding (x, y) into coordinate i, and running π on thus created input. The output of π on this input will be exactly $\mathrm{AND}(x, y)$ provided that the sampled parts do not influence the output of DISJ_n. We can guarantee this by sampling inputs for coordinates $j \in [n] - \{i\}$ from a distribution μ such that $\mathrm{supp}(\mu) \subseteq f^{-1}(0)$. This reduction allows us to apply direct-sum type arguments to compute communication complexity of DISJ_n from $\mathrm{IC}_\mu(\mathrm{AND}, 0)$ where μ is the worst distributions such that $\mu(1,1) = 0$. This approach is more general and motivates the following definition.

Definition 2.2.2 ([9]). Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function. Define

$$\mathrm{IC}^{\mathrm{zero}}(f, 0) = \max_\mu \inf_\pi \mathrm{IC}_\mu(\pi),$$

where the maximum ranges over distributions μ with $\mathrm{supp}(\mu) \subseteq f^{-1}(0)$ and the infimum ranges over all protocols π that compute f with 0 error on *every input*. Note that we could also write $\mathrm{IC}^{\mathrm{zero}}(f, 0) = \max_\mu \mathrm{IC}^{\mathrm{all}}_\mu(f, 0)$, where the maximum is over all μ with $\mathrm{supp}(\mu) \subseteq f^{-1}(0)$.

The following theorem characterizes the exact randomized communication complexity of $\bigvee$-type functions with error tolerance tending to zero in terms of $\mathrm{IC}^{\mathrm{zero}}(f, 0)$.

Theorem 2.2.4 ([9]). *Let $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ and $g_n : \{0,1\}^{nk} \times \{0,1\}^{nk} \to \{0,1\}$ be functions, such that $g_n(x,y) = \vee_{i=1}^n f(x_i, y_i)$, where $x = \{x_i\}_{i=1}^n, y = \{y_i\}_{i=1}^n$ and $x_i, y_i \in \{0,1\}^k$. Then for all $\epsilon > 0$, there exists $\delta = \delta(f,\epsilon) > 0$ such that $\delta \to 0$ as $\epsilon \to 0$ and*

$$(\mathrm{IC}^{zero}(f,0) - \delta)n \leq \mathrm{R}(g_n, \epsilon) \leq \mathrm{IC}^{zero}(f,0)n + o(n)k.$$

We apply the above to DISJ_n, DISJ_n^k, and SETINT_n functions.

Theorem 2.2.5 ([9]). *For all $\epsilon > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that $\delta \to 0$ as $\epsilon \to 0$ and*

$$(C_{\mathrm{DISJ}} - \delta)n \leq \mathrm{R}(\mathrm{DISJ}_n, \epsilon) \leq C_{\mathrm{DISJ}}n + o(n).$$

where $C_{\mathrm{DISJ}} \approx 0.4827$ bits.

Theorem 2.2.6 ([9]). *Let n, k be such that $k = \omega(1)$ and $n/k = \omega(1)$. Then for all constant $\epsilon > 0$,*

$$\left(\frac{2}{\ln 2} - O(\sqrt{\epsilon})\right)k - o(k) \leq \mathrm{R}(\mathrm{DISJ}_n^k, \epsilon) \leq \frac{2}{\ln 2}k + o(k).$$

Theorem 2.2.7 ([9]). *For all $\epsilon > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that $\delta \to 0$ as $\epsilon \to 0$ and*

$$(C_\wedge - \delta)n \leq \mathrm{R}(\mathrm{SETINT}_n, \epsilon) \leq C_\wedge n + o(n).$$

In the process of proving the above results, we derive properties of information complexity that may be of independent interest. One such property is the continuity of the information complexity function at $\epsilon = 0$:

Theorem 2.2.8 ([9]). *For all $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and $\mu \in \Delta(\mathcal{X} \times \mathcal{Y})$ we have*

$$\lim_{\epsilon \to 0} \mathrm{IC}_\mu(f,\epsilon) = \mathrm{IC}_\mu(f,0)$$
$$\lim_{\epsilon \to 0} \mathrm{IC}_\mu^{ext}(f,\epsilon) = \mathrm{IC}_\mu^{ext}(f,0).$$

2.3 Characterization of Information Complexity via Local Concavity Constraints

In this section we prove Lemma 2.2.1, a local characterization of the zero-error information complexity. Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function. We shall prove that each member of $\mathfrak{C}(f)$ (see Definition 2.2.1) is a lower bound on the zero-error information complexity $I(\mu) := \mathrm{IC}_\mu(f, 0)$. It will be evident that $I(\mu)$ itself satisfies the local concavity constraints, i.e., $I \in \mathfrak{C}(f)$. Thus we obtain a new characterization of the zero-error information complexity of f as the point-wise maximum over all functions in the family $\mathfrak{C}(f)$.

In the definition of $\mathfrak{C}(f)$ each bit B sent by a player is uniformly distributed from the external point of view, i.e., $P(B = 0) = P(B = 1) = 1/2$. We say that a protocol is in *the normal form* if every bit sent by a player in the protocol is uniformly distributed from the external point of view. More precisely:

Definition 2.3.1 ([9])**.** Let π be a protocol. We say that π is in *the normal form* if for each fixing r of public randomness and for each node u in the protocol tree of π_r

$$P(\text{owner of } u \text{ sends } 0 | \text{players reach } u \text{ in } \pi_r) = 1/2.$$

We show that an arbitrary protocol can be accurately simulated by a protocol in the normal form of the same information cost.

Lemma 2.3.1 ([9])**.** *Let π be a protocol with input space $\{0,1\}^n \times \{0,1\}^n$. Let $\mu \in \Delta(\{0,1\}^n \times \{0,1\}^n)$. For every $\delta > 0$ there exists a protocol π_δ and a function χ mapping transcripts of π_δ to transcripts of π such that:*

1. *π_δ is in the normal form,*
2. *$P\big(\Pi(x,y) \neq \chi(\Pi_\delta(x,y))\big) \leq \delta$,*
3. *$\mathrm{IC}_\mu(\pi_\delta) \leq \mathrm{IC}_\mu(\pi)$.*

Proof. Let ℓ be such that $\mathrm{CC}(\pi)2^{-\ell} \leq \delta$. Let π_δ be the π_ℓ constructed from π via Algorithm 2. Let u be the current node in the simulation. Let q be the probability

density function of ρ_2. We shall analyze q when Alice is the owner of u. When Bob is the owner of u the analysis is similar. Let

$$p_{\text{priv}}(x) = P(\text{Alice sends 0 in } T_r | \text{players reached } u, X = x)$$
$$\nu(x) = P(X = x | \text{players reached } u).$$

From Algorithm 2 it follows that

$$q(t) = \begin{cases} \sum_x \nu(x) p_{\text{priv}}(x) \frac{1}{p} & \text{if } t \in [0, p] \\ \sum_x \nu(x)(1 - p_{\text{priv}}(x)) \frac{1}{1-p} & \text{if } t \in (p, 1] \end{cases}$$

Since $p = P(\text{Alice sends 0 in } T_r | \text{players reached } u) = \sum_x p_{\text{priv}}(x)\nu(x)$, we have that $q(t) = 1$ for $t \in [0, 1]$. Thus P_2 is uniform and bits of $\widetilde{\mathrm{P}_2}$ transmitted by Alice are uniform too. This proves that π_δ is in the normal form.

The simulation of π fails at node u if the first ℓ bits of ρ_2 are equal to the first ℓ bits of p. This happens with probability $2^{-\ell}$. By union bound, the entire simulation of π fails with probability at most $\mathrm{CC}(\pi)2^{-\ell} \leq \delta$. If the simulation does not fail then it is obvious how to reconstruct the transcript of π from the transcript of π_δ. This defines the desired function χ such that $P\big(\Pi(x, y) \neq \chi(\Pi_\delta(x, y))\big) \leq \delta$.

Let B denote the bit (random variable) sent by Alice to Bob in π at node u. Since $\widetilde{\mathrm{P}_2}$ is a probabilistic function of B we can apply data processing inequality to conclude that

$$I(\widetilde{\mathrm{P}_2}; X | Y, \text{ players reached } u) \leq I(B; X | Y, \text{ players reached } u).$$

We can prove a similar inequality for the nodes owned by Bob. Hence $\mathrm{IC}_\mu(\pi_\delta) \leq \mathrm{IC}_\mu(\pi)$. □

Remark 2.3.1. Similar result holds for the *external information cost*.

Remark 2.3.2. We say that π_δ *δ-simulates* π due to Condition 2 in Lemma 2.3.1.

Algorithm 2 Constructing π_ℓ in the normal form out of arbitrary π

Require:
$x \in \{0,1\}^n$ - known to Alice
$y \in \{0,1\}^n$ - known to Bob
μ, ℓ - known to Alice and Bob

Players publicly sample r and construct T_r - tree representation of π_r
Both players set the current node u to the root of T_r
while u is not a leaf **do**
    if owner of u is Alice **then**
        Both players set $p = P(\text{Alice sends 0 in } T_r | \text{players reached } u)$
        Alice sets $p_{\text{priv}} = P(\text{Alice sends 0 in } T_r | \text{players reached } u, X = x)$
    else
        Both players set $p = P(\text{Bob sends 0 in } T_r | \text{players reached } u)$
        Bob sets $p_{\text{priv}} = P(\text{Bob sends 0 in } T_r | \text{players reached } u, Y = y)$
    Owner of u privately samples $\rho_1 \in [0,1]$ uniformly at random
    if $\rho_1 \leq p_{\text{priv}}$ **then**
        Owner of u privately samples $\rho_2 \in [0,p]$ uniformly at random
    else
        Owner of u privately samples $\rho_2 \in (p,1]$ uniformly at random
    Owner of u sends $\widetilde{\rho_2}$ = the first ℓ bits of binary expansion of ρ_2
    Both players know $\widetilde{p}$ = the first ℓ bits of binary expansion of p
    if $\widetilde{\rho_2} < \widetilde{p}$ **then**
        Both players update u to the left child of current u
    else if $\widetilde{\rho_2} > \widetilde{p}$ **then**
        Both players update u to the right child of current u
    else
        Both players terminate the simulation and report an error

Lemma 2.3.2 ([9]). *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function. Let $C \in \mathfrak{C}(f)$ and $\mu \in \Delta(\{0,1\}^n \times \{0,1\}^n)$. Let π be a protocol in the normal form that solves $f|_{\text{supp}(\mu)}$ with 0 error. Then we have*

$$C(\mu) \leq \text{IC}_\mu(\pi) + \mathbb{E}_{R,T \sim L(\pi_R)}(C(\mu_{R,T})),$$

where $L(\pi_r)$ is the distribution on the leaves of π_r and $\mu_{r,t}(x,y) = P(X = x, Y = y |$ players reach t in $\pi_r)$.

Proof by induction on $c = \text{CC}(\pi)$. When $c = 0$ the claim is clearly true, since there

is only one leaf t and $\mu_t = \mu$.

Assume the claim holds for all c-bit protocols where $c \geq 0$. Consider a $c+1$-bit protocol π. Without loss of generality, assume that Alice sends the first bit B. Let $\mu_i^{\mathcal{A}}(x,y) = P(X = x, Y = y | B = i)$ for $i \in \{0,1\}$. Let π_i denote the protocol obtained from π by letting Alice send the first bit as i, $i \in \{0,1\}$. Since π is in the normal form, we have

$$I(\Pi; X|Y) = I(\Pi; X|Y, B=0)/2 + I(\Pi; X|Y, B=1)/2 + I(B; X|Y)$$
$$I(\Pi; Y|X) = I(\Pi; Y|X, B=0)/2 + I(\Pi; Y|X, B=1)/2$$

Thus we have

$$\begin{aligned}
\mathrm{IC}_\mu(\pi) &= \mathrm{IC}_{\mu_0^{\mathcal{A}}}(\pi_0)/2 + \mathrm{IC}_{\mu_1^{\mathcal{A}}}(\pi_1)/2 + I(X; B|Y) \\
&\geq C(\mu_0^{\mathcal{A}})/2 - \mathbb{E}_{R,T\sim L(\pi_{0,R})}(C(\mu_{0,R,T}^{\mathcal{A}})/2) + \\
&+ C(\mu_0^{\mathcal{A}})/2 - \mathbb{E}_{R,T\sim L(\pi_{1,R})}(C(\mu_{1,R,T}^{\mathcal{A}})/2) + I(X; B|Y) \text{ (by induction)} \\
&= C(\mu_0^{\mathcal{A}})/2 + C(\mu_0^{\mathcal{A}})/2 + I(X; B|Y) - \mathbb{E}_{r,t\sim L(\pi_r)}(C(\mu_{r,t})) \\
&\geq C(\mu) - \mathbb{E}_{R,T\sim L(\pi_R)}(C(\mu_{R,T})) \text{ (by properties of } C)
\end{aligned}$$

□

Lemma 2.3.3 ([9])**.** *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function. Let $C \in \mathfrak{C}(f)$ and $\mu \in \Delta(\{0,1\}^n \times \{0,1\}^n)$. Let π be a protocol that solves $f|_{\mathrm{supp}(\mu)}$ with 0 error. Then we have*

$$C(\mu) \leq \mathrm{IC}_\mu(\pi).$$

Proof. Fix arbitrary $\delta > 0$. Let G_{π_r} denote the set of leaves t of π_r such that $\mathrm{R}(f|_{\mathrm{supp}(\mu_{r,t})}, 0) = 0$ (see the notation introduced in Lemma 2.3.2). By Lemma 2.3.1 there exists a protocol π_δ in the normal form, such that it δ-simulates π and $\mathrm{IC}_\mu(\pi_\delta) \leq \mathrm{IC}_\mu(\pi)$. Therefore:

$$\sum_{t \in G_{\pi_{\delta,r}}} P(\text{players reach } t \text{ in } \pi_{\delta,r}) \geq 1 - \delta.$$

Moreover, by the definition of $\mathfrak{C}(f)$ we have $C(\mu_{r,t}) = 0$ for all $t \in G_{\pi_{\delta,r}}$ and $C(\mu) \leq 2n$ for all μ. Thus by Lemma 2.3.3 it follows that that

$$C(\mu) \leq \mathrm{IC}_\mu(\pi_\delta) + 2n\delta \leq \mathrm{IC}_\mu(\pi) + 2n\delta.$$

As $\delta > 0$ was arbitrary, we have $C(\mu) \leq \mathrm{IC}_\mu(\pi)$. □

From Lemma 2.3.1 it immediately follows that $\mathrm{IC}_\mu(f, 0) \in \mathfrak{C}(f)$. This observation together with Lemma 2.3.3 prove Lemma 2.2.1. For convenience we list the main conclusions in a single place in the following corollary:

Corollary 2.3.4 ([9])**.** *For all* $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *we have*

$$\begin{aligned} &\mathrm{IC}_\mu(f,0) \in \mathfrak{C}(f) \\ &\mathrm{IC}_\mu(f,0) \geq C(\mu) && \forall \mu \forall C \in \mathfrak{C}(f) \\ &\mathrm{IC}_\mu(f,0) = \max_{C \in \mathfrak{C}(f)} C(\mu) && \forall \mu \end{aligned}$$

Remark 2.3.3. The above definitions and claims can be repeated for *the external information cost*. Replacing $I(X; B|Y)$ and $I(Y; B|X)$ with $I(XY; B)$ in Definition 2.3.1, we obtain a class $\mathfrak{C}^{\mathrm{ext}}(f)$ of lower bounds on the $I^{\mathrm{ext}}(\mu) := \mathrm{IC}^{\mathrm{ext}}_\mu(f, 0)$. Repeating the steps of Lemma 2.3.2 and Lemma 2.3.3 but replacing the internal information cost with the external information cost, we arrive at similar conclusions as in Corollary 2.3.4:

$$\begin{aligned} &\mathrm{IC}^{\mathrm{ext}}_\mu(f,0) \in \mathfrak{C}^{\mathrm{ext}}(f) \\ &\mathrm{IC}^{\mathrm{ext}}_\mu(f,0) \geq C(\mu) && \forall \mu \forall C \in \mathfrak{C}^{\mathrm{ext}}(f) \\ &\mathrm{IC}^{\mathrm{ext}}_\mu(f,0) = \max_{C \in \mathfrak{C}^{\mathrm{ext}}(f)} C(\mu) && \forall \mu \end{aligned}$$

2.4 Continuity of $\mathrm{IC}_\mu(f,\epsilon)$ at $\epsilon = 0$

In [7] it was shown that the information complexity function $\mathrm{IC}_\mu(f,\epsilon)$ is convex in ϵ on the interval $[0,1]$. An immediate corollary is that the information complexity is continuous in ϵ on the open interval $(0,1)$. This left open whether $\mathrm{IC}_\mu(f,\epsilon)$ is continuous at $\epsilon = 0$. In this section we prove that it is. This property is essential for the rest of this chapter. We arrive at the result in two steps: (1) we describe a matrix view of message transmission in communication protocols, which (2) lets us exploit the rectangular nature of protocols. We show that protocols solving f with small probability of error must terminate with a distribution on the inputs that with high probability has almost all weight on monochromatic rectangles. To turn such a protocol into a zero-error protocol, the players simply verify that their inputs belong to such a rectangle. If so, the players know the answer; otherwise, they exchange the inputs.

Let π be a communication protocol on the input space $\mathcal{X}\times\mathcal{Y}$. Let μ be a distribution on $\mathcal{X}\times\mathcal{Y}$. Let $t \in \{0,1\}^{\mathrm{CC}(\pi)}$ be a particular transcript (without loss of generality, we can assume that the transcript is of length $\mathrm{CC}(\pi)$). For each $i \in [\mathrm{CC}(\pi)]$ define μ_i as follows:

$$\mu_i(x,y) = P(X = x, Y = y | \Pi(X,Y)_{\leq i} = t_{\leq i}),$$

where the probability is taken over $(X,Y) \sim \mu$, public and private randomness of π. Thus, we may associate the sequence of distributions $\mu_1, \ldots, \mu_{\mathrm{CC}(\pi)}$ with each transcript t of π. As described in Section 1.2, each distribution can be viewed as a matrix. The following lemma asserts that in the sequence of matrices arising out of a communication transcript, the matrix μ_{i+1} can be obtained from the matrix μ_i only by either multiplying entire rows or entire columns by specific numbers.

Lemma 2.4.1 ([9])**.** *Let $\mu, \mu_0, \mu_1 \in \Delta(\mathcal{X}\times\mathcal{Y})$. The following two statements are equivalent:*

1. *There exists a signal B that Bob can send starting from μ such that $\mu_i(x,y) = P(X = x, Y = y \mid B = i)$ for $i \in \{0,1\}$.*

2. *There exists $t \in (0,1)$ and $\delta_0^y \in [0, 1/t], \delta_1^y \in [0, 1/(1-t)]$ $(y \in \mathcal{Y})$ such that*

- $\mu = t\mu_0 + (1-t)\mu_1$
- $(\forall i \in \{0,1\})(\forall (x,y) \in \mathcal{X} \times \mathcal{Y})(\mu_i(x,y) = \delta_i^y \mu(x,y))$.

Similarly for Alice, but with rows.

Proof. ($\Rightarrow$) By definition, $\mu_i(x,y) = P(X = x, Y = y | B = i)$, which by Bayes' rule is equivalent to $\mu_i(x,y) = P(B = i | X = x, Y = y)P(X = x, Y = y)/P(B = i)$. Since Bob is the speaker, $P(B = i | X = x, Y = y) = P(B = i | Y = y)$. Thus we have

$$\mu_i(x,y) = \frac{P(B = i | Y = y)}{P(B = i)} \mu(x,y).$$

Defining $\delta_i^y = P(B = i | Y = y)/P(B = i)$ and $t = P(B = 0)$ finishes the proof of (1) $\Rightarrow$ (2).

($\Leftarrow$) Define signal B by $P(B = 0 | Y = y) := t\delta_0^y \in [0,1]$ and $P(B = 1 | Y = y) := (1-t)\delta_1^y \in [0,1]$. For each y such that $\sum_x \mu(x,y) > 0$ this defines a valid distribution on $\{0,1\}$, because $\mu(x,y) = t\mu_0(x,y) + (1-t)\mu_1(x,y) = t\delta_0^y \mu(x,y) + (1-t)\delta_1^y \mu(x,y)$ therefore $t\delta_0^y + (1-t)\delta_1^y = 1$.

Next, observe that

$$P(B = 0) = \sum_y P(B = 0 | Y = y)P(Y = y) = \sum_{x,y} t\delta_0^y \mu(x,y) = \sum_{x,y} t\mu_0(x,y) = t.$$

Thus, we have defined the signal B in such a way that $\delta_i^y = P(B = i | Y = y)/P(B = i)$, and consequently we have $\mu_i(x,y) = P(X = x, Y = y | B = i)$ (following the steps of ($\Rightarrow$) direction in reverse). □

Corollary 2.4.2 ([9])**.** *Let π be a protocol on the input space $\mathcal{X} \times \mathcal{Y}$, let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$, and $t \in \{0,1\}^{CC(\pi)}$ – a transcript of π. Then there exist vectors $V_r^t \in \left(\mathbb{R}^{\geq 0}\right)^{|\mathcal{X}|}$ and $V_c^t \in \left(\mathbb{R}^{\geq 0}\right)^{|\mathcal{Y}|}$ such that for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$ we have*

$$P(X = x, Y = y | \Pi(X,Y) = t) = V_r^t(x) V_c^t(y) \mu(x,y).$$

In the rest of this section we prove Theorem 2.2.8, which asserts that the information complexities $\mathrm{IC}_\mu(f,\epsilon)$ and $\mathrm{IC}_\mu^{\mathrm{ext}}(f,\epsilon)$ are continuous at $\epsilon = 0$. We shall prove

this theorem for the internal information complexity only, because the proof for the external information complexity is similar.

Clearly, by definition we have $\mathrm{IC}_\mu(f,\epsilon) \leq \mathrm{IC}_\mu(f,0)$. We need to show that the reverse inequality holds up to a small additive error, i.e., $\mathrm{IC}_\mu(f,0) \leq \mathrm{IC}_\mu(f,\epsilon) + q(\epsilon)$ where $q(\epsilon)\to 0$ as $\epsilon \to 0$.

We first prove the theorem for full-support distributions μ. Later we will show how to reduce the case of general distributions to the case of distributions with full support. To facilitate the reduction, it will be useful to prove the full-support case for the more general setting of *(complete) relations* rather than just functions.

Definition 2.4.1. A relation $R \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ is *complete* if for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$ there exists $z \in \mathcal{Z}$ such that $(x,y,z) \in R$.

Definition 2.4.2. Let $R \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. We say that a combinatorial rectangle $G = A \times B$, where $A \subseteq \mathcal{X}$ and $B \subseteq \mathcal{Y}$, is *z-monochromatic* with respect to R if $\forall (x,y) \in G$ we have $(x,y,z) \in R$. We say that G is *monochromatic* if there exists $z \in \mathcal{Z}$ such that G is z-monochromatic.

Definition 2.4.3. Let $R \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. Fix an ordering on $\mathcal{Z}$. Let G be a monochromatic rectangle with respect to R. *The color of* G is defined to be the first $z \in \mathcal{Z}$ (in our fixed ordering) such that G is z-monochromatic.

Lemma 2.4.3 ([9])**.** *Let $R \subseteq \mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$ be a complete relation. Let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$ with full support. Let $\rho = \min_{(x,y)} \mu(x,y)$. Then for all sufficiently small $\epsilon > 0$ we have*

$$\mathrm{IC}_\mu(R,0) \leq \mathrm{IC}_\mu(R,\epsilon) + 2\left(H\left(1 - \frac{3|\mathcal{X}||\mathcal{Y}|\epsilon^{1/4}}{\rho}\right) + \log(4|\mathcal{X}||\mathcal{Y}|)\frac{3|\mathcal{X}||\mathcal{Y}|\epsilon^{1/4}}{\rho}\right).$$

Proof. Fix $\beta > 0$. Let π be a protocol solving R with error tolerance ϵ with respect to μ such that $\mathrm{IC}_\mu(\pi) \leq \mathrm{IC}_\mu(R,\epsilon)+\beta$. By Corollary 2.4.2, for all $t \in \{0,1\}^{\mathrm{CC}(\pi)}$ there exists $V_r^t \in \left(\mathbb{R}^{\geq 0}\right)^{|\mathcal{X}|}$ and $V_c^t \in \left(\mathbb{R}^{\geq 0}\right)^{|\mathcal{Y}|}$ such that for all (x,y) we have $\mu_t(x,y) := P(X = x, Y = y | \Pi(X,Y) = t) = V_r^t(x)V_c^t(y)\mu(x,y)$. Algorithm 3 shows how to construct the 0-error protocol τ out of π using these vectors V_c^t and V_r^t.

Algorithm 3 Converting an ϵ-error protocol π into a 0-error protocol τ

Require:

$x \in \mathcal{X}$ - known to Alice

$y \in \mathcal{Y}$ - known to Bob

$R \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ (complete relation), $\mu \in \Delta(\mathcal{X} \times \mathcal{Y})$ (with full support), $\rho = \min_{(x,y)} \mu(x,y)$, an ordering on $\mathcal{Z}$ - known to Alice and Bob

1: Players run π on (x, y). Let t be the resulting transcript.
2: Both players compute $L = \{x' \mid V_r^t(x') > \epsilon^{1/4}/\rho^{1/2}\}$ and $M = \{y' \mid V_c^t(y') > \epsilon^{1/4}/\rho^{1/2}\}$.
3: **if** $L \times M$ is a monochromatic rectangle **then**
4: Alice sends a bit A indicating if her input is in L.
5: Bob sends a bit B indicating if his input is in M.
6: **if** $A = B = 1$ **then**
7: Players output the color of $L \times M$.
8: **else**
9: Players exchange inputs.
10: **else**
11: Players exchange the inputs

The intuition behind Algorithm 3 is that with high probability μ_t is concentrated on monochromatic rectangles, thus the verification step (lines 4-5) does not reveal much information about the inputs. All other events in the algorithm happen with small probability, thus do not contribute much to the information cost of τ.

Now, we formalize the intuition. Let $\mathcal{E}$ be the event that π makes a mistake, and let $\mathcal{E}_t$ denote the event that π makes a mistake given that transcript is t. We have $P(\mathcal{E}) = \mathbb{E}_T(P(\mathcal{E}_T)) \leq \epsilon$ and by Markov's inequality it follows that

$$P_T(P(\mathcal{E}_T) > \epsilon^{1/2}) \leq \epsilon^{1/2}.$$

For the remainder of the argument, consider a transcript t such that $P(\mathcal{E}_t) \leq \epsilon^{1/2}$. We begin with the following claim which upper bounds the maximal entry in V_c^t, V_r^t:

Claim 2.4.4 ([9])**.** *Without loss of generality, we may assume that V_c^t and V_r^t satisfy the following:*

1. $\|V_r^t\|_\infty = \|V_c^t\|_\infty$,

2. $\|V_r^t\|_\infty, \|V_c^t\|_\infty \leq 1/\sqrt{\rho}$.

Proof. (1) Let $m_c = \max_y V_c^t(y)$ and $m_r = \max_x V_r^t(x)$, and suppose that $m_c > m_r$. Define $d = \sqrt{m_r/m_c}$. Define $(V_r^t)' = V_r^t/d$ and $(V_c^t)' = dV_c^t$. Note that $(V_r^t)'$ and $(V_c^t)'$ satisfy the first condition, and we may use these vectors instead of V_r^t and V_c^t without affecting the distribution μ_t (indeed, recall that $\mu_t(x,y) = V_r^t(x)V_c^t(y)\mu(x,y)$).

(2) Let x^* and y^* be some entries achieving the maximum values in V_r^t and V_c^t, respectively. By (1), $V_c^t(y^*) = V_r^t(x^*)$. If this common value is larger than $1/\sqrt{\rho}$ then $\mu_t(x^*,y^*) = V_c^t(y^*)V_r^t(x^*)\mu(x^*,y^*) > \mu(x^*,y^*)/\rho \geq 1$, since $\rho = \min_{(x,y)} \mu(x,y)$ and μ has full support. Contradiction. □

Claim 2.4.5 ([9])**.** *For a transcript t with $P(\mathcal{E}_t) \leq \epsilon^{1/2}$, the rectangle $L \times M$ defined in line 2 of Algorithm 3 satisfies the following properties:*

1. *$L \times M$ is monochromatic,*

2. $\mu_{t,\mathcal{A}}(L), \mu_{t,\mathcal{B}}(M) \geq 1 - \frac{|\mathcal{X}||\mathcal{Y}|\epsilon^{1/4}}{\rho}$,

where $\mu_{t,\mathcal{A}}(x) = \sum_y \mu_t(x,y)$ and $\mu_{t,\mathcal{B}}(y) = \sum_x \mu_t(x,y)$ denote the marginal distributions of μ_t.

Proof. Suppose that $L \times M$ is not monochromatic, then there exists some input $(x_0, y_0) \in L \times M$ such that $(x_0, y_0, \tau_t(x_0,y_0)) \notin R$. Therefore $P(\mathcal{E}_t) \geq \mu_t(x_0,y_0) = V_r^t(x_0)V_c^t(y_0)\mu(x_0,y_0) > (\epsilon^{1/2}/\rho)\rho = \epsilon^{1/2}$ (by the definition of L and M). This contradicts the assumption that $P(\mathcal{E}_t) \leq \epsilon^{1/2}$. This proves the first part of the claim.

Let $x^* \notin L$. Then we have

$$\begin{aligned}
\mu_{t,\mathcal{A}}(x^*) &= \sum_y \mu_t(x^*,y) \\
&= \sum_y V_r^t(x^*)V_c^t(y)\mu(x,y) \\
&\leq \sum_y \frac{\epsilon^{1/4}}{\rho^{1/2}} \frac{1}{\rho^{1/2}} \mu(x,y) \quad \text{by the definition of } L \text{ and Claim 2.4.4} \\
&= \frac{\epsilon^{1/4}}{\rho} \mu_{\mathcal{A}}(x)
\end{aligned}$$

Therefore $\mu_{t,\mathcal{A}}(\mathcal{X} - L) \leq \frac{|\mathcal{X}|\epsilon^{1/4}}{\rho} \leq \frac{|\mathcal{X}||\mathcal{Y}|\epsilon^{1/4}}{\rho}$. Similar argument shows that $\mu_{t,\mathcal{B}}(\mathcal{Y} - M) \leq \frac{|\mathcal{X}||\mathcal{Y}|\epsilon^{1/4}}{\rho}$. □

From Claim 2.4.5 it follows that for t with $P(\mathcal{E}_t) \leq \epsilon^{1/2}$ we have

$$\begin{aligned} P_{(X,Y)\sim\mu_t}((X,Y) \in L \times M) &= \mu_{t,\mathcal{A}}(L) + \mu_{t,\mathcal{B}}(M) - P_{(X,Y)\sim\mu_t}(X \in L \text{ or } Y \in M) \\ &\geq \mu_{t,\mathcal{A}}(L) + \mu_{t,\mathcal{B}}(M) - 1 \\ &\geq 1 - 2\frac{|\mathcal{X}||\mathcal{Y}|\epsilon^{1/4}}{\rho} \end{aligned}$$

Let τ_1 denote the part of the transcript of τ that corresponds to running π. Let τ_2 denote the remaining part of the transcript of τ. Let S be an indicator random variable of the event "players do not exchange inputs in τ_2". For a transcript t of τ_1 let L_t and M_t denote the sets constructed in line 2 of Algorithm 3. Then we have

$$\begin{aligned} P(S=1) &\geq P_{T,(X,Y)\sim\mu}(L_T \times M_T \text{ is monochromatic and } (X,Y) \in L_T \times M_T) \\ &\geq P_T(P(\mathcal{E}_T) \leq \epsilon^{1/2}) P_{T,(X,Y)\sim\mu_T}((X,Y) \in L \times M | P(\mathcal{E}_T) \leq \epsilon^{1/2}) \\ &\geq (1-\epsilon^{1/2})(1 - 2|\mathcal{X}||\mathcal{Y}|\epsilon^{1/4}/\rho) \\ &\geq 1 - 3|\mathcal{X}||\mathcal{Y}|\epsilon^{1/4}/\rho \end{aligned}$$

for all sufficiently small $\epsilon > 0$. Since S is determined by τ_2 we have

$$\begin{aligned} H(\tau_2) &= H(\tau_2 S) = H(S) + H(\tau_2|S) \\ &= H(S) + H(\tau_2|S=0)p(S=0) + H(\tau_2|S=1)p(S=1) \\ &= H(S) + H(\tau_2|S=0)p(S=0) \\ &\leq H\left(1 - \frac{3|\mathcal{X}||\mathcal{Y}|\epsilon^{1/4}}{\rho}\right) + (\log|\mathcal{X}| + \log|\mathcal{Y}| + 2)\frac{3|\mathcal{X}||\mathcal{Y}|}{\rho}\epsilon^{1/4} \end{aligned}$$

where (1) $H(\tau_2|S=1) = 0$, since when players do not exchange inputs τ_2 ="11", (2) $H(p)$ is a decreasing function for $p \in [1/2, 1]$, and (3) $H(\tau_2|S=0) \leq \log|\mathcal{X}| + \log|\mathcal{Y}| + 2$, since when players exchange inputs they send messages of sizes $\lceil \log|\mathcal{X}| \rceil \leq \log|\mathcal{X}| + 1$ and $\lceil \log|\mathcal{Y}| \rceil \leq \log|\mathcal{Y}| + 1$.

Now, we can relate information cost of τ to that of π:

$$\begin{aligned}
\mathrm{IC}_\mu(\tau) &= I(\tau; X|Y) + I(\tau; Y|X) \\
&= I(\tau_1\tau_2; X|Y) + I(\tau_1\tau_2; Y|X) \\
&= I(\tau_1; X|Y) + I(\tau_2; X|Y\tau_1) + I(\tau_1; Y|X) + I(\tau_2; Y|X\tau_1) \\
&= \mathrm{IC}_\mu(\pi) + I(\tau_2; X|Y\tau_1) + I(\tau_2; Y|X\tau_1) \\
&\leq \mathrm{IC}_\mu(f, \epsilon) + \beta + 2H(\tau_2).
\end{aligned}$$

The above inequality holds for all $\beta > 0$ and therefore the 0-error information complexity of R is

$$\begin{aligned}
\mathrm{IC}_\mu(R, 0) &\leq \mathrm{IC}_\mu(R, \epsilon) + 2H(\tau_2) \\
&\leq \mathrm{IC}_\mu(R, \epsilon) + 2\left(H\left(1 - \frac{3|\mathcal{X}||\mathcal{Y}|\epsilon^{1/4}}{\rho}\right) + \log(4|\mathcal{X}||\mathcal{Y}|)\frac{3|\mathcal{X}||\mathcal{Y}|}{\rho}\epsilon^{1/4}\right).
\end{aligned}$$

as claimed. □

Corollary of Lemma 2.4.3 is Theorem 2.2.8 for the case of distributions μ with full support. We need the following lemma to complete the proof of Theorem 2.2.8 for the case of general distributions.

Lemma 2.4.6 ([9])**.** *Let $\epsilon \in [0, 1/2)$ be a parameter. Let μ_1 and μ_2 be distributions on $\{0,1\}^n \times \{0,1\}^n$ such that $|\mu_1 - \mu_2| \leq \epsilon$. Let π be a protocol on the input space $\{0,1\}^n \times \{0,1\}^n$. Then we have*

$$|\mathrm{IC}_{\mu_1}(\pi) - \mathrm{IC}_{\mu_2}(\pi)| \leq 10\epsilon n + 4H(2\epsilon).$$

This means that $\mathrm{IC}_\mu(\pi)$ is continuous in μ for all protocols π.

Proof. Let $(\tilde{X}, \tilde{Y}, V)$ be distributed uniformly over the space $\{0,1\}^n \times \{0,1\}^n \times [0,1]$. Define the random variable $F \in \{0,1,2,3\}$ as follows

- $F = 0$, if $V < \min(\mu_1(\tilde{X}, \tilde{Y}), \mu_2(\tilde{X}, \tilde{Y}))$

- $F = 1$, if $\mu_2(\tilde{X}, \tilde{Y}) \leq V < \mu_1(\tilde{X}, \tilde{Y})$
- $F = 2$, if $\mu_1(\tilde{X}, \tilde{Y}) \leq V < \mu_2(\tilde{X}, \tilde{Y})$
- $F = 3$, otherwise

Define $E = F|F \in \{0, 1, 2\}$, and $(X, Y) = (\tilde{X}, \tilde{Y})|F \in \{0, 1, 2\}$. Let μ be the distribution of (X, Y). Next, we list properties of X, Y and E.

$$\begin{aligned} P(X = x, Y = y|E \in \{0, 1\}) &= \frac{P(X = x, Y = y, E \in \{0, 1\})}{P(E \in \{0, 1\})} \\ &= \frac{\mu_1(x, y)/2^{2n}}{\sum_{x,y} \mu_1(x, y)/2^{2n}} \\ &= \mu_1(x, y) \end{aligned}$$

Thus, $(X, Y)|(E \in \{0, 1\}) \sim \mu_1$. Similarly $(X, Y)|(E \in \{0, 2\}) \sim \mu_2$. This means that $I_{\mu_1}(X; \Pi|Y) = I_\mu(X; \Pi|Y, E \in \{0, 1\})$ and $I_{\mu_2}(X; \Pi|Y) = I_\mu(X; \Pi|Y, E \in \{0, 2\})$.

$$\begin{aligned} P(E = 1) &= \frac{\sum_{x,y} \max(\mu_1(x, y) - \mu_2(x, y), 0)/2^{2n}}{\sum_{x,y} \max(\mu_1(x, y), \mu_2(x, y))/2^{2n}} \\ &\leq ||\mu_1 - \mu_2|| \leq \epsilon. \end{aligned}$$

Similarly, we have $P(E = 2) \leq \epsilon$. Define $E_{\{0,1\}} = E|(E \in \{0, 1\})$ and $E_{\{0,2\}} = E|(E \in \{0, 2\})$. Note that

$$H(E_{\{0,1\}}) = H(P(E = 1)/(1 - P(E = 2)) \leq H\left(\frac{\epsilon}{1 - \epsilon}\right) \leq H(2\epsilon)$$

The idea of the rest of the proof is to show that $I_\mu(X; \Pi|Y E_{\{0,1\}})$ is close to $I_{\mu_1}(X; \Pi|Y)$ and that $I_\mu(X; \Pi|Y E_{\{0,2\}})$ is close to $I_{\mu_2}(X; \Pi|Y)$. Then we can expand the expression $I_\mu(X; \Pi|Y E)$ over different values of E and rewrite it in two different ways: one involving $I_{\mu_1}(X; \Pi|Y)$ and small additional terms, and one involving $I_{\mu_2}(X; \Pi|Y)$ and small additional terms. This demonstrates that the two quantities of interest are equal up to small additional terms that vanish as $\epsilon \to 0$. We proceed with this program.

Start by expressing $I_\mu(X;\Pi E_{\{0,1\}}|Y)$ in two ways as follows:

$$\begin{aligned} I_\mu(X;\Pi E_{\{0,1\}}|Y) &= I_\mu(X;E_{\{0,1\}}|Y) + I_\mu(X;\Pi|YE_{\{0,1\}}) \quad \text{first way} \\ &= I_\mu(X;\Pi|Y, E\in\{0,1\}) + I_\mu(X;E_{\{0,1\}}|Y\Pi) \\ &= I_{\mu_1}(X;\Pi|Y) + I_\mu(X;E_{\{0,1\}}|Y\Pi) \quad \text{second way} \end{aligned}$$

Therefore, we have

$$\begin{aligned} |I_{\mu_1}(X;\Pi|Y) - I_\mu(X;\Pi|YE_{\{0,1\}})| &\le |I_\mu(X;E_{\{0,1\}}|Y) - I_\mu(X;E_{\{0,1\}}|Y\Pi)| \\ &\le H(E_{\{0,1\}}) \le H(2\epsilon) \end{aligned}$$

This shows that $I_{\mu_1}(X;\Pi|Y)$ is close to $I_\mu(X;\Pi|YE_{\{0,1\}})$. Similarly, we can derive that

$$|I_{\mu_2}(X;\Pi|Y) - I_\mu(X;\Pi|YE_{\{0,2\}})| \le H(2\epsilon)$$

Define $p_i = P(E=i)$ for notational convenience. Now we proceed to the second part of the plan and write $I_\mu(X;\Pi|YE)$ in two ways. The first way of writing it is as follows:

$$\begin{aligned} I_\mu(X;\Pi|YE) &= \sum_{i=0}^{2} p_i I_\mu(X;\Pi|Y,E=i) \\ &= p_2 I_\mu(X;\Pi|Y,E=2)+ \\ &+ (p_0+p_1)\sum_{i\in\{0,1\}} P(E_{\{0,1\}}=i) I_\mu(X;\Pi|Y,E_\{0,1\}=i) \\ &= p_2 I_\mu(X;\Pi|Y,E=2) + (p_0+p_1) I_\mu(X;\Pi|YE_{\{0,1\}}) \end{aligned}$$

Similarly, we can rewrite $I_\mu(X;\Pi|YE)$ as follows

$$I_\mu(X;\Pi|YE) = p_1 I_\mu(X;\Pi|Y,E=1) + (p_0+p_2) I_\mu(X;\Pi|YE_{\{0,2\}})$$

Subtracting the right hand sides of the above two expressions and rearranging we get

$$\begin{aligned} p_0(I_\mu(X;\Pi|YE_{\{0,1\}}) - I_\mu(X;\Pi|YE_{\{0,2\}})) = p_1 I_\mu(X;\Pi|Y,E=1)+ \\ + p_2 I_\mu(X;\Pi|YE_{\{0,2\}})- \\ - p_2 I_\mu(X;\Pi|Y,E=2)- \\ - p_1 I_\mu(X;\Pi|YE_{\{0,1\}}) \end{aligned}$$

Since each of the informational terms on the right hand side is trivially bounded by n and $p_1, p_2 \leq \epsilon$ we conclude

$$|I_\mu(X;\Pi|YE_{\{0,1\}}) - I_\mu(X;\Pi|YE_{\{0,2\}})| \leq 4\epsilon n/p_0 \leq 4\epsilon n/(1-2\epsilon) \leq 5\epsilon n$$

Combining this with our estimates on $I_{\mu_1}(X;\Pi|Y)$ and $I_{\mu_2}(X;\Pi|Y)$ from above we obtain

$$|I_{\mu_1}(X;\Pi|Y) - I_{\mu_2}(X;\Pi|Y)| \leq 2H(2\epsilon) + 5\epsilon n$$

We can repeat the entire process exchanging X and Y to bound the difference in the other term of the information cost. Overall, we get

$$|\operatorname{IC}_{\mu_1}(\pi) - \operatorname{IC}_{\mu_2}(\pi)| \leq 4H(2\epsilon) + 10\epsilon n.$$

□

Now, we finish the proof of Theorem 2.2.8 for the case of general distributions.

Proof. We show how to prove the first part of the theorem – the continuity of the internal information complexity at error tolerance 0. Fix $\epsilon > 0$. Let μ be a distribution over $\mathcal{X} \times \mathcal{Y}$ (not necessarily with full support). Let π be an ϵ-error protocol for f under μ.

Let U be the uniform distribution on $\mathcal{X} \times \mathcal{Y}$, and define $\mu'(x,y) := pU(x,y) + (1-p)\mu(x,y)$, where $p = \epsilon^{1/8}$. Then μ' has full support, and for small enough ϵ, we have $\rho = \min_{(x,y)} \mu'(x,y) = p = \epsilon^{1/8}$. Define the relation $R_f \subseteq \mathcal{X} \times \mathcal{Y} \times \{0,1\}$ as follows: $(x,y,f(x,y)) \in R_f$ for all $(x,y) \in \operatorname{supp}(\mu)$, and $(x,y,z) \in R_f$ where $z = \{0,1\}$ for all

$(x, y) \notin \text{supp}(\mu)$. Then R_f is trivially satisfied outside the support of μ, and it agrees with f on the support of μ.

Clearly, π solves R_f under μ' with error tolerance at most ϵ. By Lemma 2.4.3, there is a zero-error protocol τ for R_f under μ' such that

$$\text{IC}_{\mu'}(\tau) \leq \text{IC}_{\mu'}(\pi) + \alpha,$$

where $\alpha = 2\left(H\left(1 - \frac{3|\mathcal{X}||\mathcal{Y}|\epsilon^{1/4}}{\rho}\right) + \log(4|\mathcal{X}||\mathcal{Y}|)\frac{3|\mathcal{X}||\mathcal{Y}|}{\rho}\epsilon^{1/4}\right)$. Observe that α goes to 0 as ϵ goes to 0, since $\rho = \epsilon^{1/8}$. Also, note that τ is a zero-error protocol for f under μ.

Since $\|\mu - \mu'\| \leq p$, Lemma 2.4.6 implies that

$$\text{IC}_{\mu'}(\pi) \leq \text{IC}_{\mu}(\pi) + 5p \log |\mathcal{X}||\mathcal{Y}| + 4H(2p)$$

and therefore $\text{IC}_{\mu'}(\tau) \leq \text{IC}_{\mu}(\pi) + 5p \log |\mathcal{X}||\mathcal{Y}| + 4H(2p) + \alpha$. Using Lemma 2.4.6 again, we have

$$\begin{aligned}
\text{IC}_{\mu}(\tau) &\leq \text{IC}_{\mu'}(\tau) + 5p \log |\mathcal{X}||\mathcal{Y}| + 4H(2p) \\
&\leq \text{IC}_{\mu}(\pi) + 10p(\log |\mathcal{X}||\mathcal{Y}| + 8H(2p) + \alpha \\
&\leq \text{IC}_{\mu}(\pi) + 10\epsilon^{1/8} \log |\mathcal{X}||\mathcal{Y}| + 8H(2\epsilon^{1/8}) + \alpha,
\end{aligned}$$

and clearly all the terms except $\text{IC}_{\mu}(\pi)$ in the above expression tend to 0 when $\epsilon \to 0$. This finishes the proof of the first part of the statement.

The second part of the statement that deals with the external information complexity is proved analogously. □

2.5 Information Complexity of AND with Zero Error Tolerance

In this section we shall compute *the exact internal and the external information complexities* of the 2-bit AND function. We summarize our findings about the AND function in Section 2.5.1. In Section 2.5.3 we present a *clocked protocol* π for the

AND function, in which the parties use a continuous clock in an asynchronous fashion (this will become clearer later). The protocol π is infeasible in the sense that no finite-round protocol can simulate it; however, we may still analyze its information cost as a function of the input distribution μ. We use the machinery developed in the previous sections to demonstrate that the information cost function of π gives a lower bound on the information complexity (Sections 2.5.6 and 2.5.8) of AND. Hence, the information cost of π is precisely the information complexity of the AND function. The infeasibility of π is an expected side effect. The information complexity of a function is the infimum over protocols, and thus may not be achievable by any finite-round protocol. In Section 2.7 we describe a natural finite-round discretization of π and analyze its rate of convergence (as a function of the number of rounds) to the true unbounded-round information complexity of AND.

The protocol π suggests that the space of distributions on $\{0,1\} \times \{0,1\}$ is partitioned into three regions - "Alice's region", "Bob's region", and a "diagonal" region (corresponding to symmetric distributions). Section 2.5.4 describes the regions and how together with the results from Section 2.5.2 they reduce the number of cases necessary to consider in the analysis of the information cost function of π.

2.5.1 Summary of Results for AND

In Sections 2.5.5, 2.5.6, 2.5.7, and 2.5.8 we shall derive the exact closed-form formulas for the internal and external 0-error information complexities of the AND function. In this section we present the main results.

Theorem (Theorem 2.2.2 restated)**.**

1. $\mathrm{IC}(\mathrm{AND}, 0) = C_\wedge = 1.49238\ldots$
2. $\mathrm{IC}^{ext}(\mathrm{AND}, 0) = \log 3 = 1.58396\ldots$

Proof. 1. The precise number $C_\wedge$ is obtained via numerical optimization of the specific concave function obtained in Sections 2.5.5 and 2.5.6, using Wolfram

Mathematica. The distribution that achieves this maximum is

$$\mu = \begin{array}{|c|c|}\hline 0.0808931\ldots & 0.264381\ldots \\ \hline 0.264381\ldots & 0.390346\ldots \\ \hline \end{array}.$$

2. The external information complexity is concave, so the distribution that achieves the maximum has to be symmetric. We first show an upper bound. That is for every distribution $\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \beta & \delta \\ \hline \end{array}$ we have $\mathrm{IC}^{\mathrm{ext}}_{\mu}(\mathrm{AND}, 0) \leq \log 3$. Consider a trivial protocol, in which Alice sends her bit X. Then if X turns out to be 1, Bob sends his bit. The information cost of this protocol is

$$\begin{aligned} &H(X) + p(X=1)H(Y|X=1) = (\alpha+\beta)\log\frac{1}{\alpha+\beta} + (\beta+\delta)\log\frac{1}{\beta+\delta} + \\ &\quad + (\beta+\delta)\left(\frac{\beta}{\beta+\delta}\log\frac{\beta+\delta}{\beta} + \frac{\delta}{\beta+\delta}\log\frac{\beta+\delta}{\delta}\right) \\ &= (\alpha+\beta)\log\frac{1}{\alpha+\beta} + \beta\log\frac{1}{\beta} + \delta\log\frac{1}{\delta} = H(\mu'), \end{aligned}$$

where μ' is a distribution on a sample space with three elements $1, 2, 3$ and $\mu'(1) = \alpha+\beta, \mu'(2) = \beta, \mu'(3) = \delta$. Since Shannon entropy is maximized for a uniform distribution, we immediately get that the information cost of the above protocol is at most $\log 3$.

Now we turn to the lower bound. Consider the distribution

$$\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \beta & \gamma \\ \hline \end{array}.$$

From Sections 2.5.8 and 2.5.7, we have

$$\mathrm{IC}^{\mathrm{ext}}_{\mu}(\mathrm{AND}, 0) = (1-\alpha-2\beta)\log\frac{1}{1-\alpha-2\beta} + \frac{\beta}{\ln 2} + \frac{\beta}{\alpha}\log\beta - \frac{(\alpha+\beta)^2}{\alpha}\log(\alpha+\beta).$$

By taking the limit of the above expression as α approaches 0, we obtain the following formula for the external information complexity of AND for distributions

$\nu =$ | 0 | β |
|---|---|
| β | $1-2\beta$ |:

$$\mathrm{IC}^{\mathrm{ext}}_{\nu}(\mathrm{AND}, 0) = (1-2\beta)\log\frac{1}{1-2\beta} - 2\beta\log\beta.$$

Consider the particular $\nu_{1/3} =$ | 0 | 1/3 |
|---|---|
| 1/3 | 1/3 |:

$$\mathrm{IC}^{\mathrm{ext}}_{\nu_{1/3}}(\mathrm{AND}, 0) = \frac{1}{3}\log 3 - \frac{2}{3}\log\frac{1}{3} = \log 3.$$

This shows that $\log 3$ is also a lower bound on $\mathrm{IC}^{\mathrm{ext}}(\mathrm{AND}, 0)$.

□

Remark 2.5.1. Observe that there is a symmetric distribution that achieves the maximum of $\mathrm{IC}(\mathrm{AND}, 0)$. This holds for all symmetric functions. Let f be a symmetric function and μ be an arbitrary distribution on the inputs of f. Then $\mathrm{IC}_{\mu}(f, 0) = \mathrm{IC}_{\mu^T}(f, 0)$ and it is easy to see that the information complexity is a concave function in μ. Thus for $\mu' = \mu/2 + \mu^T/2$, which is symmetric, we have $\mathrm{IC}_{\mu'}(f, 0) \geq \mathrm{IC}_{\mu}(f, 0)/2 + \mathrm{IC}_{\mu^T}(f, 0)/2 = \mathrm{IC}_{\mu}(f, 0)$. The same holds for the external information complexity.

When in later sections we consider the disjointness function, distributions μ that place 0 mass on $(1, 1)$ entry will play a crucial role. Note that for such distributions we still insist that the protocol solving AND has 0 *error on all inputs*. The following two claims describe the information complexity of AND with respect to such distributions. These claims follow immediately from Sections 2.5.6 and 2.5.5.

Claim 2.5.1 ([9])**.** *For symmetric distributions*

$\mu =$ | α | β |
|---|---|
| β | 0 |

we have

$$\mathrm{IC}_{\mu}^{all}(\mathrm{AND}, 0) = \frac{\beta}{\ln 2} + \frac{\beta^2}{\alpha} \log \frac{\beta}{\alpha + \beta} + \alpha \log \frac{\alpha + \beta}{\alpha}.$$

Claim 2.5.2 ([9])**.** *For distributions*

$$\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \gamma & 0 \\ \hline \end{array}$$

we have

$$\mathrm{IC}_{\mu}^{all}(\mathrm{AND}, 0) = (\alpha + \beta) H\left(\frac{\beta}{\gamma} \frac{\alpha + \gamma}{\alpha + \beta}\right) - \alpha H\left(\frac{\beta}{\gamma}\right) + t\, \mathrm{IC}_{\nu}^{all}(\mathrm{AND}, 0),$$

where

$$t = 2\beta + \frac{\alpha\beta}{\gamma}$$

and

$$\nu = \begin{array}{|c|c|}\hline \frac{\beta\alpha}{\gamma t} & \frac{\beta}{t} \\ \hline \frac{\beta}{t} & 0 \\ \hline \end{array}$$

Now, the following theorem follows via numerical optimization of the above formulas.

Theorem 2.5.3.

$$\lim_{\epsilon \to 0} \max_{\mu : \mu(1,1) \leq \epsilon} \mathrm{IC}_{\mu}(\mathrm{AND}, 0) = 0.482702\ldots.$$

We will also need the following claim about the information complexity of AND with respect to symmetric distributions with non-zero mass on $(1, 1)$.

Claim 2.5.4 ([9])**.** *For a symmetric distribution* $\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \beta & \delta \\ \hline \end{array}$ *we have*

$$\mathrm{IC}_{\mu}(\mathrm{AND}, 0) = \frac{\beta}{\ln 2} + 2\delta \log \frac{\beta + \delta}{\delta} + + 2\beta \log \frac{\beta + \delta}{\beta} + \frac{\beta^2}{\alpha} \log \frac{\beta}{\beta + \alpha} + \alpha \log \frac{\alpha + \beta}{\alpha}.$$

In Section 2.7 we prove Theorem 2.2.3 - a tight bound on the rate of convergence of the r-round information complexity of the AND function to the unbounded-round information complexity:

Theorem (Theorem 2.2.3 restated)**.** *Let μ be a distribution on $\{0,1\} \times \{0,1\}$ with full support. Then we have*

$$\mathrm{IC}^r_\mu(\mathrm{AND}, 0) = \mathrm{IC}_\mu(\mathrm{AND}, 0) + \Theta_\mu\left(\frac{1}{r^2}\right).$$

Moreover, the lower bound holds even for μ such that $\mu(1,1) = 0$.

2.5.2 Random Walk View of a Protocol: Distribution on Distributions and Splitting Lemmas

A natural question, which arises when we view a communication protocol as a random walk on $\Delta(\mathcal{X} \times \mathcal{Y})$ (see Section 1.5), is whether the amount of information revealed in a single step of the random walk depends on how that step was performed. For instance, suppose that in a single step the random walk moves from μ to μ_0 with probability p and μ_1 with probability $1 - p$. Furthermore, suppose that there is a different random walk, such that after an entire sequence of steps it moves from μ to μ_0 with probability p and μ_1 with probability $1 - p$. Can we conclude that the information revealed by the second random walk is the same as the information revealed by the first random walk (when viewed as protocols)? The answer is yes, and it is the first result of this section.

It will be convenient to collect all bits sent by the same player in one round into a single message. Thus, each step of a protocol can be viewed as follows: starting from a commonly known prior distribution $\mu \in \Delta(\mathcal{X} \times \mathcal{Y})$, the current speaker transmits a message $M \in \{0,1\}^\ell$ where ℓ is the largest length of a message for this step. When a certain instance m of the message is communicated, the players update their common prior distribution to $\mu_m(x,y) = P(X = x, Y = y | M = m)$.

Definition 2.5.1. For a message M we define the *distribution on distributions for* M as follows: the sample space is $\Omega = \{\mu_m \mid m \in \mathrm{range}(M)\}$ and the distribution p on Ω is $p(\mu_m) = P(\mu_M = \mu_m) = \sum_{\widetilde{m}:\mu_{\widetilde{m}}=\mu_m} P(M = \widetilde{m})$, where the probability is over the private randomness of the speaker and conditioned on the speaker's input.

We shall use notation $(\{\mu_1, \mu_2, \ldots\}, \{p_1, p_2, \ldots\})$ to denote a particular distribution

on distributions.

Now, we are ready to prove the first result of this section.

Lemma 2.5.5 (Distribution on Distributions Lemma, [9])**.** *Let μ be a prior on inputs $\mathcal{X} \times \mathcal{Y}$. Suppose that in one protocol starting with μ Bob transmits B such that $\mu_b(x, y) = P(X = x, Y = y \mid B = b)$ for $b \in \{0, 1\}$. Suppose that in another protocol starting with μ Bob transmits a sequence of bits M such that*

- $\mu_m(x, y) := P(X = x, Y = y | M = m)$,
- $(\forall m \in \text{range}(M))(\mu_m \in \{\mu_0, \mu_1\})$,
- $P(\mathcal{M}_b) = P(B = b)$, *where* $\mathcal{M}_b = \{m | \mu_m = \mu_b\}$ *for* $b \in \{0, 1\}$.

Then we have

$$I(Y; M|X) = I(Y; B|X).$$

Proof. For all $b \in \{0, 1\}$ and for all $m \in \mathcal{M}_b$ we have $\mu_m = \mu_b$, i.e., $P(X = x, Y = y|M = m) = P(X = x, Y = y|B = b)$. Hence $P(X = x|M = m) = P(X = x|B = b)$ and consequently $P(Y = y|X = x, M = m) = P(Y = y|X = x, B = b)$. We have

$$\begin{aligned}
&I(Y; M|X) = \\
&= \mathbb{E}_{x,m}(\mathbb{D}(Y_{xm}||Y_x) \\
&= \textstyle\sum_{x,y,m} P(X = x, Y = y, M = m) \log \frac{P(Y=y|X=x,M=m)}{P(Y=y|X=x)} \\
&= \textstyle\sum_{x,y,b} \sum_{m \in \mathcal{M}_b} P(X = x, Y = y, M = m) \log \frac{P(Y=y|X=x,B=b)}{P(Y=y|X=x)} \\
&= \textstyle\sum_{x,y,b} \sum_{m \in \mathcal{M}_b} \mu_m(x, y) P(M = m) \log \frac{P(Y=y|X=x,B=b)}{P(Y=y|X=x)} \\
&= \textstyle\sum_{x,y,b} \mu_b(x, y) P(\mathcal{M}_b) \log \frac{P(Y=y|X=x,B=b)}{P(Y=y|X=x)} \\
&= \textstyle\sum_{x,y,b} P(X = x, Y = y|B = b) P(B = b) \log \frac{P(Y=y|X=x,B=b)}{P(Y=y|X=x)} \\
&= \mathbb{E}_{x,b}(\mathbb{D}(Y_{xb}||Y_x)) \\
&= I(Y; B|X).
\end{aligned}$$

□

The tools introduced in this section shall be used later to reduce the number of cases necessary to consider in the analysis of the information complexity of AND

function. One such tool is the distribution on distributions. Another tool is the Splitting Lemma, which is the second result of this section: if a player can "split" prior μ into μ_0 and μ_1 by transmitting a bit, then the same player can split any prior ρ into any $\rho_0, \rho_1 \in [\mu_0, \mu_1]$ satisfying $\rho \in [\rho_0, \rho_1]$ by transmitting a bit. Here, "splitting μ into μ_0 and μ_1" means that there exists a message consisting of a single bit B such that the distribution on distributions of B is $(\{\mu_0, \mu_1\}, \{p_0, p_1\})$ for some p_0, p_1. We formalize this below.

The proof of the Splitting Lemma uses the matrix view of message transmission (Lemma 2.4.1). Since the transmitted bit B satisfies the assumptions of Lemma 2.4.1, we may express μ_0 and μ_1 as μ with its columns (or rows, depending on the speaker) scaled by certain *scaling coefficients* (direction (1) $\Rightarrow$ (2) of Lemma 2.4.1). Every distribution in the interval $[\mu_0, \mu_1]$ is a linear combination of column-scaled (or row-scaled) versions of μ, and thus is a column-scaled (or row-scaled) μ itself. Finding scaling coefficients for ρ_0, ρ_1 and ρ we observe that ρ_0 and ρ_1 are, in fact, column-scaled (or row-scaled) versions of ρ. Applying direction (2) $\Rightarrow$ (1) of Lemma 2.4.1 we arrive at the desired conclusion.

Lemma 2.5.6 (Splitting Lemma, [9])**.** *] Suppose that starting with $\mu \in \Delta(\mathcal{X} \times \mathcal{Y})$ Bob sends signal B such that $\mu_i(x, y) = P(X = x, Y = y \mid B = i)$. Let $\rho_0, \rho_1 \in [\mu_0, \mu_1]$ and $\rho \in [\rho_0, \rho_1]$. Then there exists signal B' that Bob can send starting at distribution ρ such that $\rho_i(x, y) = P(X = x, Y = y \mid B' = i)$. Similarly, when Alice sends bit B.*

Proof. Since $\rho_0, \rho_1 \in [\mu_0, \mu_1]$ there exist numbers $t_0, t_1 \in [0, 1]$ such that $\rho_0 = t_0\mu_0 + (1-t_0)\mu_1$ and $\rho_1 = t_1\mu_0 + (1-t_1)\mu_1$. Also, since $\rho \in [\rho_0, \rho_1]$ we have $\rho = t\rho_0 + (1-t)\rho_1$ for some $t \in [0, 1]$. By direction (1) $\Rightarrow$ (2) of Lemma 2.4.1 we have $\mu_i(x, y) = \delta_i^y \mu(x, y)$ for some δ_i^y, $i \in \{0, 1\}, y \in \mathcal{Y}$. Then we can express ρ_0 and ρ_1 in terms of μ as follows:

$$\begin{aligned} \rho_0(x, y) &= (t_0\delta_0^y + (1 - t_0)\delta_1^y)\mu(x, y), \\ \rho_1(x, y) &= (t_1\delta_0^y + (1 - t_1)\delta_1^y)\mu(x, y). \end{aligned}$$

Define $C_0^y := t_0\delta_0^y + (1 - t_0)\delta_1^y$ and $C_1^y := t_1\delta_0^y + (1 - t_1)\delta_1^y$. Then we have

$$\rho(x, y) = (tC_0^y + (1 - t)C_1^y)\mu(x, y).$$

Now, it is easy to see that ρ_0 and ρ_1 are "column-scaled" versions of ρ with scaling coefficients defined by

$$\widetilde{\delta}_i^y := \frac{C_i^y}{tC_0^y + (1-t)C_1^y}.$$

Overall, we have

1. $\rho = t\rho_0 + (1-t)\rho_1$,
2. $\rho_i(x,y) = \widetilde{\delta}_i^y \rho(x,y)$,
3. $\widetilde{\delta}_0^y = \frac{C_0^y}{tC_0^y + (1-t)C_1^y}$,
4. $\widetilde{\delta}_1^y = \frac{C_1^y}{tC_0^y + (1-t)C_1^y}$

Thus by Lemma 2.4.1 there exists a signal B' with the desired properties. □

2.5.3 Information-Optimal Protocol for AND

In this section we present Protocol 4 - a zero-error protocol for AND : $\{0,1\}^2 \to \{0,1\}$, which achieves both the internal and the external information complexities of AND[1]. The inputs (X,Y) to AND are distributed according to $\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \gamma & \delta \\ \hline \end{array}$. Protocol 4 gives a formal description of our protocol. Below we provide a verbal explanation of each of the steps. In the process, we try to describe some intuition as to how information is revealed in this protocol, and why this protocol is a good candidate for being information-optimal for AND.

Protocol 4 consists of two parts. In the first part (lines 1-8), Alice and Bob check to see if μ is symmetric. If μ is not symmetric, the appropriate player sends a single bit such that μ becomes symmetric conditioned on the value of that bit. We shall refer to the first part of π as its *non-symmetric part*. In the second part (lines 9-18), Alice and Bob start with a symmetric distribution and observe the clock as it increases from 0 to 1. As the time passes, the distribution on the inputs is continuously being updated by the players. During the update process, the distribution remains

[1]In general, there is no reason to believe that the protocol achieving the internal information complexity of a function should also achieve its external information complexity.

Protocol 4 Protocol for AND with optimal information complexity

Require:

$x \in \{0,1\}$ - known to Alice

$y \in \{0,1\}$ - known to Bob

$\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \gamma & \delta \\ \hline \end{array}$ - known to Alice and Bob

if $\beta < \gamma$ **then**

  Bob sends bit B as follows $B = \begin{cases} 1 & \text{if } y = 1 \\ 0 & \text{with probability } 1 - \beta/\gamma \text{ if } y = 0 \\ 1 & \text{with probability } \beta/\gamma \text{ if } y = 0 \end{cases}$

  if $B = 0$ **then**

    The protocol terminates, the players output 0

if $\beta > \gamma$ **then**

  Alice sends bit B as follows $B = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{with probability } 1 - \gamma/\beta \text{ if } x = 0 \\ 1 & \text{with probability } \gamma/\beta \text{ if } x = 0 \end{cases}$

  if $B = 0$ **then**

    The protocol terminates, the players output 0.

if $x = 0$ **then**

  Alice samples $N^A \in_R [0,1)$ uniformly at random

else

  Alice sets $N^A = 1$

if $y = 0$ **then**

  Bob samples $N^B \in_R [0,1)$ uniformly at random

else

  Bob sets $N^B = 1$

Alice and Bob monitor the value of the clock C, which starts at 0.

The clock continuously increases to 1. If $\min(N^A, N^B) < 1$, when the clock reaches $\min(N^A, N^B)$ the corresponding player sends 0 to the other player, the protocol ends, the players output 0. If $\min(N^A, N^B) = 1$, once the clock reaches 1, Alice sends 1 to Bob, the protocol ends, and the players output 1.

symmetric throughout. Intuitively, as the time passes, each player becomes more and more convinced that the other player has 1 as their input. The presence of the clock enables "continuous leakage" of information, but makes this protocol infeasible in the strict sense of communication protocols - no finite-round protocol can simulate it. A finite-round protocol necessarily leaks bounded-from-zero amount of information in each (non-redundant) round. In π when a player's private number (N^A or N^B) is

reached by the clock, the player "raises the hand" to indicate the end of the protocol. The rules for picking the private numbers N^A and N^B can be intuitively justified by the following two observations:

1. When a player has input 0, that player does not need to know the other player's input. However, the other player must become aware that the first player has input 0, so that both players agree on the output of AND being 0.

2. When both players have 0 as input, their roles are completely symmetric, because AND is a symmetric function.

We shall refer to the second part of π as its *symmetric part*.

Remark 2.5.2. In a well-defined protocol, the order in which the players communicate should depend solely on the previous communication. For our "clocked" protocol, it is natural to require that the order should depend on the previous communication and the value of the clock. This presents a small problem: in case $N^A = N^B < 1$ both players transmit 0 simultaneously. However, the event "$N^A = N^B < 1$" happens with probability 0, thus without loss of generality we may assume that it never happens.

From the description of Protocol 4, it is clear that it correctly solves AND on all inputs. Analyzing its information cost requires careful calculations. This is what the remaining part of this section is devoted to. The separation of the protocol into *non-symmetric* and *symmetric* parts makes our calculations more modular. This separation will be reflected in subsection structure of the remainder of this section.

2.5.4 Regions of $\Delta(\{0,1\} \times \{0,1\})$ for the AND *Function*

Protocol 4 suggests that the space $\Delta(\{0,1\} \times \{0,1\})$ of distributions $\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \gamma & \delta \\ \hline \end{array}$ on $\{0,1\} \times \{0,1\}$ is partitioned into three regions for the AND function:

1. **Bob's region** consisting of all distributions μ with $\beta < \gamma$,

2. **Alice's region** consisting of all distributions μ with $\beta > \gamma$,

3. **Diagonal region** consisting of all symmetric distributions μ with $\beta = \gamma$.

Bob's region consists of all priors, for which Bob is more likely to have 0 as his input than Alice, i.e.,

$$P(Y = 0) = \alpha + \gamma > \alpha + \beta = P(X = 0).$$

Similarly, Alice's region consists of all priors, for which Alice is more likely to have 0 as her input. The diagonal region consists of all distributions, for which both players are equally likely to have 0 as their inputs.

Recall that a communication protocol can be viewed as a random walk on the space of distributions $\Delta(\{0,1\} \times \{0,1\})$ (see Section 1.5), where at each step one of the players notifies the other player of the new location in the space of distributions chosen randomly based on the speaker's input. Next, we interpret Protocol 4 as such a random walk: if the random walk starts in Bob's region, Bob makes a step that places the current position either on the diagonal, or terminates the walk. Similarly, if the random walk starts in Alice's region, Alice makes a step that places the current position either on the diagonal, or terminates the walk. If the current position is on the diagonal, then the random walk proceeds along the diagonal to position $\begin{array}{|c|c|}\hline 0 & 0 \\ \hline 0 & 1 \\ \hline \end{array}$, and at each step there is a chance that the protocol terminates when one of the players declares that they have 0. Later we show that if in some protocol Alice speaks in Bob's region, then that particular step releases non-optimal amount of information and may be improved by changing the speaker (see Sections 2.5.6 and 2.5.8). Bob speaking in Alice's region reveals non-optimal amount of information too.

In a feasible (i.e., finite-round) protocol, it is impossible to perform the random walk perfectly along the diagonal – once the prior is on the diagonal, the next bit of communication necessarily moves the prior off the diagonal with probability 1/2 (assuming normal form) into Alice's region and probability 1/2 into Bob's region, making that step non-optimal no matter who the speaker is. If the players could transmit infinitesimal amount of information at each step, they would be able to maintain the prior on the diagonal. This is exactly what the clock in Protocol 4 achieves.

Example 2.5.2. In this example we shall consider *product distributions* parameter-

ized by $a := P(X = 0)$ and $b := P(Y = 0)$, i.e.,

$$\mu_{a,b} := \begin{array}{|c|c|} \hline ab & a(1-b) \\ \hline (1-a)b & (1-a)(1-b) \\ \hline \end{array}$$

Since product distributions are parameterized by two values, we can actually plot the partition of the space of distributions in 2 dimensions. Figure 2.1 shows the space of all product distributions partitioned into the three regions: Alice's region, Bob's region, and the diagonal.

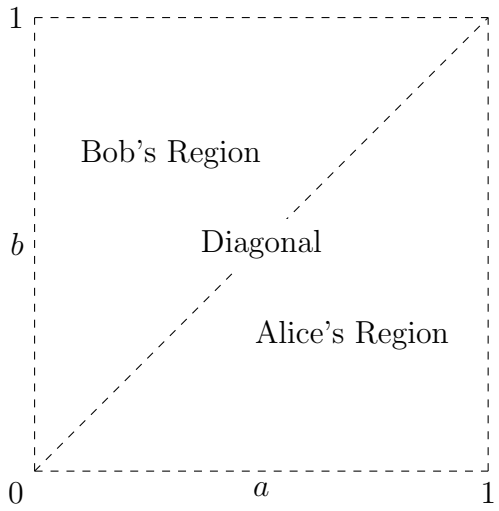

Figure 2.1: Partition of the space of product distributions into Alice's region, Bob's region, and the diagonal.

In Sections 2.5.6 and 2.5.8 we prove that the information cost of Protocol 4 satisfies Definition 2.2.1 and therefore is a lower bound on the information complexity of AND. Recall that to prove that a function satisfies Definition 2.2.1, we need to prove that the function satisfies certain concavity constraints arising out of all possible signals that could be sent by players in a communication protocol. We claim that among all possible signals that either Alice or Bob can send in Definition 2.2.1, it suffices to consider just three cases for Bob and three cases for Alice. In the rest of the section we shall only talk about the cases when Bob is the speaker, understanding that the

analogous considerations immediately follow for the case when Alice is the speaker by exchanging the roles of two players. Assume that the players start with a prior μ. The three cases for Bob are as follows:

1. the prior μ is in Bob's region, Bob sends a bit, each of the possible resulting distributions either *remains* in Bob's region or belongs to the diagonal,

2. the prior μ is in Alice's region, Bob sends a bit, each of the possible resulting distributions either *remains* in Alice's region or belongs to the diagonal,

3. the prior μ is in the diagonal region, Bob sends a bit, one of the possible resulting distributions falls in Alice's region and the other possible resulting distribution falls in Bob's region.

The cases missing above are when Bob sends a bit and one of the possible resulting distributions "crosses the diagonal" (i.e., if the players start in Bob's region and end up in Alice's region or start in Alice's region and end up in Bob's region). We refer to such bits as *crossing bits*, and bits of one of the forms above (1-3) as *non-crossing bits*. In this section we show that we can omit checking crossing bits for the purpose of Definition 2.2.1. At the end of the section we also show that we can omit checking case (1) of the non-crossing bits too. The following claim establishes that we can simulate every crossing bit with a sequence of non-crossing bits without increasing the information cost of such a step.

Claim 2.5.7 ([9])**.** *Any crossing bit B sent by Bob in an execution of a normal-form protocol may be replaced by a sequence $(B_1, B_2, \ldots)$ of non-crossing bits (in normal form) such that the distribution on distributions of $(B_1, B_2, \ldots)$ is the same as the distribution on distributions of B.*

Proof. Suppose that Bob's signal B starts at μ and has a distribution on distributions $(\{\mu_0, \mu_1\}, \{1/2, 1/2\})$ and moreover $[\mu_0, \mu_1]$ contains a symmetric distribution μ_D. We shall replace B with a sequence $(B_1, B_2, \ldots)$ representing the random walk on $[\mu_0, \mu_1]$ where each step is as large as possible under a constraint of not crossing μ_D, μ_0, and μ_1. If the random walk reaches μ_0 or μ_1 it terminates. Formally this simulation is described in Protocol 5.

Protocol 5 Protocol for simulating a crossing bit with non-crossing bits

Set $\mu_c \leftarrow \mu$
while $\mu_c \neq \mu_0$ and $\mu_c \neq \mu_1$ **do**
    if $(2\mu_c - \mu_D) \in [\mu_0, \mu_1]$ **then**
        Bob sends B_i (by Lemma 2.5.6) splitting μ_c into $2\mu_c - \mu_D$ and μ_D
    else if $(2\mu_c - \mu_0) \in [\mu_0, \mu_1]$ **then**
        Bob sends B_i (by Lemma 2.5.6) splitting μ_c into $2\mu_c - \mu_0$ and μ_0
    else
        Bob sends B_i (by Lemma 2.5.6) splitting μ_c into $2\mu_c - \mu_1$ and μ_1
    Update μ_c to the current distribution

Each bit sent in Protocol 5 is in normal form, hence the random walk on $[\mu_0, \mu_1]$ is unbiased. The optional stopping theorem from the theory of martingales implies that the probability of random walk reaching μ_0 is $1/2$. Hence the distribution on distributions is preserved. □

Remark 2.5.3.

- By Distribution on Distributions Lemma 2.5.5, the message $(B_1, B_2, \ldots)$ in Protocol 5 carries exactly the same information as the crossing bit B.
- Protocol 5 may not terminate, but this happens with probability 0. This bad behavior of our simulation can be handled via a standard argument – truncate the protocol after a sufficiently large number of steps have been performed.
- The same argument holds for the external information cost.

Now we turn to showing that case (1) of the non-crossing bits can also be omitted. Suppose that starting from μ in Bob's regions Bob sends a non-crossing bit B and then executes Protocol 4. The information about inputs revealed by these two steps is exactly the same as if the players executed Protocol 4 from μ right away. We prove this in the rest of this section. Let π denote Protocol 4. First we need the following lemma.

Lemma 2.5.8 ([9])**.** *Let μ be a non-symmetric distribution $\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \gamma & \delta \\ \hline \end{array}$ such that at least one symmetric distribution is reachable from μ if only Bob speaks. Then there*

exists a unique symmetric distribution μ_D such that for any message M that Bob can send we have $(\forall m \in \text{range}(M))(\mu_m \textit{is symmetric} \Rightarrow \mu_m = \mu_D)$.

Proof. Suppose that $\gamma < \beta$. Bob sending a message is equivalent to multiplying the columns of the matrix for μ by nonnegative numbers c_0, c_1. In order for Bob to arrive at a symmetric distribution he must achieve $c_0\gamma = c_1\beta$. There are two possibilities:

1. $\gamma = 0$ then $\beta \neq 0$ (μ is not symmetric). There is only one possibility for the resulting symmetric distribution $\begin{array}{|c|c|}\hline 1 & 0 \\ \hline 0 & 0 \\ \hline \end{array}$, which uniquely determines $(c_0, c_1) = (1/\alpha, 0)$.

2. $\gamma > 0$ then we must have $c_0 > 1$. But since the resulting matrix still has to correspond to a valid distribution we have $c_0(\alpha+\gamma)+c_1(\beta+\delta) = 1$. This forces $c_1 < 1$. Moreover, as c_0 decreases, c_1 increases. Thus, by continuity there is only one solution (c_0, c_1) satisfying $c_0\gamma = c_1\beta$.

□

Claim 2.5.9 ([9])**.** *If Bob sends a non-crossing signal B in normal form starting from prior μ in Bob's region and having a distribution on distributions $(\{\mu_0, \mu_1\}, \{1/2, 1/2\})$ then*

$$\text{IC}_\mu(\pi) = \text{IC}_{\mu_0}(\pi)/2 + \text{IC}_{\mu_1}(\pi)/2 + I(B; Y|X).$$

In particular, constraint in Definition 2.2.1 is satisfied for such signals.

Proof. Define τ to be the following protocol:

1. Bob sends signal B as in the statement of the claim, resulting in distribution μ_B

2. The players run π starting at μ_B

Observe that expanding the information cost of τ after step 1 above we obtain

$$\text{IC}_\mu(\tau) = \text{IC}_{\mu_0}(\pi)/2 + \text{IC}_{\mu_1}(\pi)/2 + I(Y; B|X)$$

It is left to show that $\mathrm{IC}_\mu(\tau) = \mathrm{IC}_\mu(\pi)$.

Let π_1 denote the non-symmetric part of π when it is executed on μ and π_2 denote the remaining part of π. Let τ_1 denote the part of τ corresponding to step 1 above together with the non-symmetric part of π from step 2. Let τ_2 denote the remaining part of τ. To finish the proof it suffices to show that Π_1 and T_1 have the same distribution on distributions, because then the information content of messages Π_1 and T_1 would be the same by Lemma 2.5.5, and $\Pi_2|\Pi_1$ would have the same distribution as $T_2|T_1$ implying that $\mathrm{IC}_\mu(\tau) = \mathrm{IC}_\mu(\pi)$.

Suppose that the message Π_1 has a distribution on distributions $(\{\nu_0, \nu_1\}, \{t, 1-t\})$, i.e., $\mu = t\nu_0 + (1-t)\nu_1$, where ν_0 is the distribution after Bob sent 0 in the non-symmetric part of π (note: $P_{\nu_0}(Y = 1) = 0$) and ν_1 is the distribution on the diagonal.

Define random variables $X_0 = \mu$, $X_1 = \mu_B$ - the updated distribution after Bob sent bit B, and $X_2 = \mu_{T_1}$ - the updated distribution after τ_1 was executed. We have

1. $\mathbb{E}(X_2) = X_0 = \mu$, because X_0, X_1, X_2 is a martingale by definition of μ_B and μ_{T_1}, and

2. $X_2 \in \{\nu_0, \nu_1\}$ by Lemma 2.5.8 and a simple observation that there is a unique distribution $\widetilde{\nu}$ reachable by Bob from μ such that $P_{\widetilde{\nu}}(Y = 1) = 0$.

The above two facts imply that $P(X_2 = \nu_0) = t$. So T_1 has the same distribution on distributions as Π_1. □

2.5.5 Internal Information Cost: Upper Bound

We start by analyzing **the symmetric part** of Protocol 4, i.e., we shall compute $\mathrm{IC}_\nu(\pi)$ where

$$\nu = \begin{array}{|c|c|} \hline \alpha & \beta \\ \hline \beta & 1-\alpha-2\beta \\ \hline \end{array}$$

is a symmetric distribution.

Since ν is symmetric and the roles of Alice and Bob in Protocol 4 are symmetric,

we have

$$\mathrm{IC}_\nu(\pi) = I(X;\Pi|Y) + I(Y;\Pi|X) = 2I(X;\Pi|Y).$$

Working from first principles, we obtain

$$\begin{aligned} I(X;\Pi|Y) &= (\alpha+\beta)I(X;\Pi|Y=0) + (1-\alpha-\beta)I(X;\Pi|Y=1) \\ &= (\alpha+\beta)I(X;\Pi|Y=0) + (1-\alpha-\beta)H(X|Y=1) \\ &= \alpha\mathbb{D}(\Pi_{X=0,Y=0}||\Pi_{Y=0}) + \beta\mathbb{D}(\Pi_{X=1,Y=0}||\Pi_{Y=0}) \\ &+ (1-\alpha-\beta)H(X|Y=1). \end{aligned}$$

The second step follows from

$$I(X;\Pi|Y=1) = H(X|Y=1) - H(X|\Pi, Y=1)$$

and $H(X|\Pi, Y=1) = 0$, since given $Y=1$ the transcript Π determines X.

A transcript of Π on (x,y) can be represented uniquely by the value $c \in [0,1]$ of the clock when the protocol terminates together with a name of a player $\mathcal{P} \in \{A,B\}$, whose random number is reached by the clock first. For $x, y \in \{0,1\}$ we have

$$\mathbb{D}(\Pi_{xy}||\Pi_y) = \sum_{\mathcal{P}\in\{A,B\}} \int_0^1 f_{x,y}(c,\mathcal{P}) \log \frac{f_{x,y}(c,\mathcal{P})}{f_y(c,\mathcal{P})} dc,$$

where $f_{x,y}(c,\mathcal{P})$ is the probability density function (PDF) for Π_{xy} and $f_y(c,\mathcal{P})$ is the PDF for Π_y.

We have

- $f_{0,0}(c,A) = f_{0,0}(c,B) = 1-c$ for $c \in [0,1]$
- $f_{1,0}(c,A) = 0$ for $c \in [0,1)$ and $f_{1,0}(c,B) = 1$ for $c \in [0,1)$
- $f_0(c,A) = \frac{\alpha}{\alpha+\beta}(1-c)$ for $c \in [0,1]$ and $f_0(c,B) = \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\beta+\alpha}(1-c)$ for $c \in [0,1)$

Overall we obtain

$$I(X;\Pi \mid Y) = \alpha \int_0^1 (1-c)\log\frac{\alpha+\beta}{\alpha} + (1-c)\log\frac{(1-c)(\alpha+\beta)}{\beta+(1-c)\alpha}\,\mathrm{d}c + \\ + \beta\int_0^1 \log\frac{\alpha+\beta}{\beta+(1-c)\alpha}\,\mathrm{d}c + (1-\alpha-\beta)H\left(\frac{\beta}{1-\alpha-\beta}\right).$$

After using Wolfram Mathematica to simplify the expressions, we obtain:

$$\mathrm{IC}_\nu(\pi) = \frac{\beta}{\ln 2} + 2(1-\alpha-2\beta)\log\frac{1-\alpha-\beta}{1-\alpha-2\beta} + \\ + 2\beta\log\frac{1-\alpha-\beta}{\beta} + \frac{\beta^2}{\alpha}\log\frac{\beta}{\beta+\alpha} + \alpha\log\frac{\alpha+\beta}{\alpha} \tag{2.1}$$

Now, we consider **the non-symmetric part** of Protocol 4 for the prior

$$\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \gamma & 1-\alpha-\beta-\gamma \\ \hline \end{array},$$

where $\beta < \gamma$. Recall that Bob sends bit B with distribution

$$B = \begin{cases} 1 & \text{if } y = 1 \\ 0 & \text{with probability } 1-\beta/\gamma \text{ if } y = 0 \\ 1 & \text{with probability } \beta/\gamma \text{ if } y = 0 \end{cases}$$

The contribution of this bit to the *internal* information cost is

$$\begin{aligned} I(Y;B|X) &= H(B|X) - H(B|XY) \\ &= (\alpha+\beta)H\left(\frac{\beta}{\alpha+\beta} + \frac{\beta}{\gamma}\cdot\frac{\alpha}{\alpha+\beta}\right) + \\ &+ (\gamma+\delta)H\left(\frac{\delta}{\gamma+\delta} + \frac{\beta}{\gamma}\cdot\frac{\gamma}{\gamma+\delta}\right) - \\ &- (\alpha+\gamma)H\left(\frac{\beta}{\gamma}\right). \end{aligned} \tag{2.2}$$

Bob sends bit 1 with probability $t = 1 - \alpha - \gamma + \beta + \alpha\beta/\gamma$. In that case the protocol continues on distribution

$$\widetilde{\nu} = \begin{array}{|c|c|} \hline (\beta\alpha)/(\gamma t) & \beta/t \\ \hline \beta/t & (1-\alpha-\beta-\gamma)/t \\ \hline \end{array}.$$

If Bob sends 0 the protocol terminates. Thus the overall internal information cost of π for the case $\beta \leq \gamma$ is

$$\mathrm{IC}_{\mu}(\pi) = I(Y; B|X) + t\,\mathrm{IC}_{\widetilde{\nu}}(\pi). \tag{2.3}$$

Closed-form formula for the above equation may be obtained from (2.1) and (2.2). Since the roles of Alice and Bob are symmetric, we have

$$\mathrm{IC}_{\mu}(\pi) = \mathrm{IC}_{\mu^T}(\pi).$$

This completes the analysis of $\mathrm{IC}_{\mu}(\pi)$ for all three cases $\beta < \gamma, \beta = \gamma, \beta > \gamma$.

2.5.6 Internal Information Cost: Lower Bound

In this section we shall show that Expression (2.3) is a lower bound on $\mathrm{IC}_{\mu}(\mathrm{AND}, 0)$. Let

$$\mu = \begin{array}{|c|c|} \hline \alpha & \beta \\ \hline \gamma & 1-\alpha-\beta-\gamma \\ \hline \end{array}$$

and suppose that Bob sends signal B with properties

- $P(B = 1) = P(B = 0) = 1/2$,
- $P(B = 1 \mid Y = 1) = 1/2 + \epsilon_1/2$,
- $P(B = 0 \mid Y = 0) = 1/2 + \epsilon_0/2$.

The resulting distributions are

- $\mu_0 = \begin{array}{|c|c|} \hline (1+\epsilon_0)\alpha & (1-\epsilon_1)\beta \\ \hline (1+\epsilon_0)\gamma & (1-\epsilon_1)(1-\alpha-\beta-\gamma) \\ \hline \end{array}$ if Bob sends 0, and

- $\mu_1 = \begin{array}{|c|c|}\hline (1-\epsilon_0)\alpha & (1+\epsilon_1)\beta \\ \hline (1-\epsilon_0)\gamma & (1+\epsilon_1)(1-\alpha-\beta-\gamma) \\ \hline \end{array}$ if Bob sends 1.

Also note that $\epsilon_1 = \epsilon_0 \frac{\alpha+\gamma}{1-\alpha-\gamma}$.

Corollary 2.3.4 says that to demonstrate that $\mathrm{IC}_\mu(\pi)$ is a lower bound on $\mathrm{IC}_\mu(\mathrm{AND}, 0)$ it suffices to prove the following concavity constraint:

$$\mathrm{IC}_\mu(\pi) \le \mathrm{IC}_{\mu_0}(\pi)/2 + \mathrm{IC}_{\mu_1}(\pi)/2 + I(B;Y|X),$$

where

$$\begin{aligned} I(B;Y|X) &= H(B|X) - H(B|XY) \\ &= (\alpha+\beta)H(B \mid X=0) + (\gamma+\delta)H(B \mid X=1) - \\ &- \sum_{i,j\in\{0,1\}} H(B \mid X=i, Y=j) \\ &= (\alpha+\beta)H\left(\frac{\alpha}{\alpha+\beta}(1/2-\epsilon_0/2) + \frac{\beta}{\alpha+\beta}(1/2+\epsilon_1/2)\right) + \\ &+ (\gamma+\delta)H\left(\frac{\gamma}{\gamma+\delta}(1/2-\epsilon_0/2) + \frac{\delta}{\gamma+\delta}(1/2+\epsilon_1/2)\right) - \\ &- (\alpha+\gamma)H(1/2+\epsilon_0/2) - (\beta+\delta)H(1/2+\epsilon_1/2) \end{aligned}$$

By Claims 2.5.7 and 2.5.9, to demonstrate that $\mathrm{IC}_\mu(\pi)$ is a lower bound on $I(\mathrm{AND}) := \mathrm{IC}_\mu(\mathrm{AND}, 0)$ it suffices to consider only two types of *non-crossing* signals B that are sent by Bob:

1. The prior μ is in Alice's region, i. e., $\beta > \gamma$. Using Wolfram Mathematica we obtain

$$\mathrm{IC}_{\mu_0}(\pi)/2 + \mathrm{IC}_{\mu_1}(\pi)/2 + I(Y;B|X) - \mathrm{IC}_\mu(\pi) = \frac{\alpha(\beta-\gamma)}{(\alpha+\beta)(1-\alpha-\gamma)^2 \ln 4}\epsilon_0^2 + O(\epsilon_0^3),$$

   which is > 0 for small enough ϵ_0.

2. The prior μ is in the diagonal region, i. e., $\beta = \gamma$. Using Wolfram Mathematica

we obtain

$$\mathrm{IC}_{\mu_0}(\pi)/2 + \mathrm{IC}_{\mu_1}(\pi)/2 + I(Y;B|X) - \mathrm{IC}_{\mu}(\pi) = \frac{\alpha\beta}{12(\alpha+\beta)(1-\alpha-\beta)^3 \ln 2}\epsilon_0^3 + O(\epsilon_0^4),$$

which is > 0 for small enough ϵ_0.

Also, note that trivially $\mathrm{IC}_{\mu}(\pi) \leq 2$, as the players learn at most each others bits during the execution of π. Hence Expression (2.3) satisfies all the constraints of Definition 2.2.1 and thus is a lower bound on $\mathrm{IC}_{\mu}(\mathrm{AND}, 0)$ by Corollary 2.3.4.

2.5.7 External Information Cost: Upper Bound

We start by analyzing **the symmetric part** of Protocol 4, i.e., we shall compute $\mathrm{IC}^{\mathrm{ext}}_{\nu}(\pi)$ where

$$\nu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \beta & 1-\alpha-2\beta \\ \hline \end{array}$$

is a symmetric distribution.

Working from first-principles, we obtain

$$\begin{aligned}\mathrm{IC}^{\mathrm{ext}}_{\nu}(\pi) &= I(XY;\Pi) \\ &= \mathbb{E}_{x,y}(\mathbb{D}(\Pi_{xy}||\Pi)) \\ &= \alpha\mathbb{D}(\Pi_{X=0,Y=0}||\Pi) + \beta\mathbb{D}(\Pi_{X=0,Y=1}||\Pi) + \\ &+ \beta\mathbb{D}(\Pi_{X=1,Y=0}||\Pi) + (1-\alpha-2\beta)\mathbb{D}(\Pi_{X=1,Y=1}||\Pi).\end{aligned}$$

A transcript of Π on x, y is determined by the value $c \in [0,1]$ of the clock when the protocol is terminated together with a name of a player $\mathcal{P} \in \{A, B\}$, whose random number is reached by a counter first. For $x, y \in \{0,1\}$ we have

$$\mathbb{D}(\Pi_{xy}||\Pi) = \sum_{\mathcal{P}\in\{A,B\}} \int_0^1 f_{x,y}(c,\mathcal{P}) \log \frac{f_{x,y}(c,\mathcal{P})}{f(c,\mathcal{P})} dc,$$

where $f_{x,y}(c,\mathcal{P})$ is the pdf for Π_{xy} and $f(c,\mathcal{P})$ is the PDF for Π.

We have

- $f_{0,0}(c, A) = f_{0,0}(c, B) = 1 - c$ for $c \in [0, 1]$
- $f_{0,1}(c, A) = 1$ for $c \in [0, 1)$ and $f_{0,1}(c, B) = 0$ for $c \in [0, 1]$
- $f_{1,1}(c, A) = f_{1,1}(c, B) = 0$ for $c \in [0, 1)$ and $P(\Pi_{X=1,Y=1} = (1, A)) = 1$
- $f(c, A) = f(c, B) = \alpha(1 - c) + \beta$ for $c \in [0, 1)$ and $P(\Pi = (1, A)) = 1 - \alpha - 2\beta$

After plugging in the above PDFs in the expression for $\mathrm{IC}_\nu^{\mathrm{ext}}(\pi)$ and using Wolfram Mathematica to simplify the expressions, we obtain:

$$\begin{aligned}\mathrm{IC}_\nu^{\mathrm{ext}}(\pi) &= 2\alpha \int_0^1 (1-c) \log \frac{(1-c)}{\alpha(1-c)+\beta} dc + 2\beta \int_0^1 \log \frac{1}{\alpha(1-c)+\beta} dc + \\ &\quad + (1-\alpha-2\beta) \log \frac{1}{1-\alpha-2\beta} \\ &= (1-\alpha-2\beta) \log \frac{1}{1-\alpha-2\beta} + \frac{\beta}{\ln 2} + \frac{\beta^2}{\alpha} \log \beta - \frac{(\alpha+\beta)^2}{\alpha} \log(\alpha+\beta).\end{aligned}$$

Now, we consider **the non-symmetric part** of Protocol 4 for the prior

$$\mu = \begin{array}{|c|c|} \hline \alpha & \beta \\ \hline \gamma & 1-\alpha-\beta-\gamma \\ \hline \end{array},$$

where $\beta < \gamma$. Bob sends bit B with distribution

$$B = \begin{cases} 1 & \text{if } y = 1 \\ 0 & \text{with probability } 1 - \beta/\gamma \text{ if } y = 0 \\ 1 & \text{with probability } \beta/\gamma \text{ if } y = 0 \end{cases}$$

The contribution of this bit to the external information cost is

$$\begin{aligned} I(XY; B) &= H(B) - H(B \mid XY) \\ &= H(B) - H(B \mid Y) \\ &= H((1-\alpha-\gamma) + (\beta/\gamma)(\alpha+\gamma)) - (\alpha+\gamma) H(\beta/\gamma). \end{aligned}$$

Bob sends bit 1 with probability $t = 1 - \alpha - \gamma + \beta + \alpha\beta/\gamma$. In that case the protocol continues on distribution

$$\widetilde{\nu} = \begin{array}{|c|c|} \hline \beta\alpha/(\gamma t) & \beta/t \\ \hline \beta/t & (1-\alpha-\beta-\gamma)/t \\ \hline \end{array}.$$

If Bob sends 0 the protocol terminates. Thus the overall external information cost of π for the case $\beta \leq \gamma$ is as follows (once again, Wolfram Mathematica was used to simplify the expressions):

$$\begin{aligned} \mathrm{IC}_{\mu}^{\mathrm{ext}}(\pi) &= I(XY;B) + t\,\mathrm{IC}_{\widetilde{\nu}}^{\mathrm{ext}}(\pi) \\ &= \frac{\beta}{\ln 2} + \beta \log \frac{1}{\beta} + (1-\alpha-\beta-\gamma)\log\frac{1}{1-\alpha-\beta-\gamma} + \\ &+ \frac{\beta(\alpha+\gamma)}{\alpha}\log\gamma + \frac{(\alpha+\beta)(\alpha+\gamma)}{\alpha}\log\frac{1}{\alpha+\gamma}. \end{aligned} \tag{2.4}$$

Since the roles of Alice and Bob are symmetric, we have

$$\mathrm{IC}_{\mu}^{\mathrm{ext}}(\pi) = \mathrm{IC}_{\mu^T}^{\mathrm{ext}}(\pi).$$

This completes the analysis of $\mathrm{IC}_{\mu}^{\mathrm{ext}}(\pi)$ for all three cases $\beta < \gamma, \beta = \gamma, \beta > \gamma$.

Remark 2.5.4. Observe that if $\alpha = 0$, i. e.,

$$\mu = \begin{array}{|c|c|} \hline 0 & \beta \\ \hline \gamma & 1-\beta-\gamma \\ \hline \end{array},$$

the expression of $\mathrm{IC}_{\mu}^{\mathrm{ext}}(\pi)$ simplifies to

$$\mathrm{IC}_{\mu}^{\mathrm{ext}}(\pi) = \beta\log\frac{1}{\beta} + \gamma\log\frac{1}{\gamma} + (1-\beta-\gamma)\log\frac{1}{1-\beta-\gamma} = H(\mu).$$

2.5.8 External Information Cost: Lower Bound

In this section we shall show that Expression (2.4) is a lower bound on $\mathrm{IC}_{\mu}^{\mathrm{ext}}(\mathrm{AND}, 0)$. Let

$$\mu = \begin{array}{|c|c|} \hline \alpha & \beta \\ \hline \gamma & 1-\alpha-\beta-\gamma \\ \hline \end{array}$$

and suppose that Bob sends signal B with properties

- $P(B=1) = P(B=0) = 1/2$,
- $P(B=1 \mid Y=1) = 1/2 + \epsilon_1/2$,
- $P(B=0 \mid Y=0) = 1/2 + \epsilon_0/2$.

The resulting distributions are

- $\mu_0 = \begin{array}{|c|c|} \hline (1+\epsilon_0)\alpha & (1-\epsilon_1)\beta \\ \hline (1+\epsilon_0)\gamma & (1-\epsilon_1)(1-\alpha-\beta-\gamma) \\ \hline \end{array}$ if Bob sends 0, and
- $\mu_1 = \begin{array}{|c|c|} \hline (1-\epsilon_0)\alpha & (1+\epsilon_1)\beta \\ \hline (1-\epsilon_0)\gamma & (1+\epsilon_1)(1-\alpha-\beta-\gamma) \\ \hline \end{array}$ if Bob sends 1.

Also note that $\epsilon_1 = \epsilon_0 \frac{\alpha+\gamma}{1-\alpha-\gamma}$.

Remark 2.3.3 says that to demonstrate that $\mathrm{IC}_{\mu}^{\mathrm{ext}}(\pi)$ is a lower bound on $\mathrm{IC}_{\mu}^{\mathrm{ext}}(\mathrm{AND}, 0)$ it suffices to prove the following concavity constraint:

$$\mathrm{IC}_{\mu}^{\mathrm{ext}}(\pi) \leq \mathrm{IC}_{\mu_0}^{\mathrm{ext}}(\pi)/2 + \mathrm{IC}_{\mu_1}^{\mathrm{ext}}(\pi)/2 + I(XY; B),$$

where

$$\begin{aligned} & I(XY; B) \\ &= H(B) - H(B \mid XY) \\ &= H(B) - H(B \mid Y) \\ &= 1 - ((\alpha+\gamma)H(1/2+\epsilon_0/2) + (1-\alpha-\gamma)H(1/2+\epsilon_1/2)) \,. \end{aligned}$$

By Claims 2.5.7 and 2.5.9, to demonstrate that $\mathrm{IC}_{\mu}^{\mathrm{ext}}(\pi)$ is a lower bound on $I^{\mathrm{ext}}(\mathrm{AND}) := \mathrm{IC}_{\mu}^{\mathrm{ext}}(\mathrm{AND}, 0)$ it suffices to consider only two types of *non-crossing* signals B that are sent by Bob:

1. The prior μ is in Alice's region, i. e., $\beta > \gamma$. Using Wolfram Mathematica we obtain

$$\mathrm{IC}_{\mu_0}^{\mathrm{ext}}(\pi)/2 + \mathrm{IC}_{\mu_1}^{\mathrm{ext}}(\pi)/2 + I(XY; B) - \mathrm{IC}_{\mu}^{\mathrm{ext}}(\pi) = \frac{\alpha(\beta-\gamma)}{(\alpha+\beta)(1-\alpha-\gamma)^2 \ln 4}\epsilon_0^2 + O(\epsilon_0^3),$$

   which is > 0 for small enough ϵ_0.

2. The prior μ is in the diagonal region, i. e., $\beta = \gamma$. Using Wolfram Mathematica we obtain

$$\mathrm{IC}_{\mu_0}^{\mathrm{ext}}(\pi)/2 + \mathrm{IC}_{\mu_1}^{\mathrm{ext}}(\pi)/2 + I(XY; B) - \mathrm{IC}_{\mu}^{\mathrm{ext}}(\pi) = \frac{\alpha\beta}{12(\alpha+\beta)(1-\alpha-\beta)^3 \ln 2}\epsilon_0^3 + O(\epsilon_0^4),$$

   which is > 0 for small enough ϵ_0.

Also, note that trivially $\mathrm{IC}_{\mu}^{\mathrm{ext}}(\pi) \leq 2$, as the players learn at most each others bits during the execution of π. Hence Expression (2.4) satisfies all the constraints of Definition 2.2.1 and thus is a lower bound on $\mathrm{IC}_{\mu}(\mathrm{AND}, 0)$ by Remark 2.3.3.

2.6 Partial Differential Equation Formulation of Information Complexity

In this section we show how to derive a system of partial differential equations, such that information complexity is a local solution to this system. This gives an alternative way of deriving formulas for the information complexity of AND function. In Subsection 2.6.1 we derive the system of PDEs for the information complexity of AND

function with respect to product distributions. Moreover, we solve this system of differential equations and plot the solution in 3-dimensional space. In Subsection 2.6.2 we generalize the system of PDEs from Subsection 2.6.1 to general distributions and arbitrary functions.

2.6.1 Information Complexity of AND *under Product Distributions via PDEs*

Notation: in this section we shall only consider product distributions parameterized by $a := P(X = 0)$ and $b := P(Y = 0)$, i.e.,

$$\mu_{a,b} := \begin{array}{|c|c|} \hline ab & a(1-b) \\ \hline (1-a)b & (1-a)(1-b) \\ \hline \end{array}$$

For convenience we shall write $\mathcal{I}(a, b) := \mathrm{IC}_{\mu_{a,b}}(\mathrm{AND})$.

Definition 2.6.1. For $\epsilon \in [-1, 1]$ define $B_{\mathcal{A}}(\epsilon)$ to be the random variable such that

- $P(B_{\mathcal{A}}(\epsilon) = 0|X = 0) = \frac{1}{2} + \frac{\epsilon}{2}$,
- $P(B_{\mathcal{A}}(\epsilon) = 0) = P(B_{\mathcal{A}}(\epsilon) = 1) = \frac{1}{2}$.

Define $B_{\mathcal{B}}(\epsilon)$ similarly, that is $P(B_{\mathcal{B}}(\epsilon) = 0|Y = 0) = \frac{1}{2} + \frac{\epsilon}{2}$, and $P(B_{\mathcal{B}}(\epsilon) = 0) = P(B_{\mathcal{B}}(\epsilon) = 1) = \frac{1}{2}$.

By Lemma 2.3.1 we may assume that in a nearly-optimal protocol Alice and Bob send bits of the form $B_{\mathcal{A}}(\epsilon)$ and $B_{\mathcal{B}}(\epsilon)$, respectively. Previously we referred to such protocols as being in normal form. We collect a few simple facts about $B_{\mathcal{A}}(\epsilon)$ in the following lemma:

Lemma 2.6.1. *Let $\mu_{a,b}$ be the prior distribution and let Alice send bit $B_{\mathcal{A}}(\epsilon)$ then*

1. $P(X = 0|B_{\mathcal{A}}(\epsilon) = 0) = (1 + \epsilon)a$, $P(Y = 0|B_{\mathcal{A}}(\epsilon) = 0) = b$, *and*
2. $P(X = 0|B_{\mathcal{A}}(\epsilon) = 1) = (1 - \epsilon)a$, $P(Y = 0|B_{\mathcal{A}}(\epsilon) = 1) = b$,
3. $\lim_{\epsilon \to 0} \frac{-2I(B_{\mathcal{A}}(\epsilon);X|Y)}{\epsilon^2 a^2} = \frac{1}{a(a-1)\ln 2}$.

Analogous statements hold for Bob's bits.

Proof.

1. Straightforward computation via Bayes' theorem.

2. Same as above.

3. First compute

$$\begin{aligned} I(B_{\mathcal{A}}(\epsilon); X|Y) &= I(B_{\mathcal{A}}(\epsilon); X) && X,Y \text{ indep.} \\ &= H(B_{\mathcal{A}}(\epsilon)) - H(B_{\mathcal{A}}(\epsilon)|X) \\ &= 1 - aH\left(\tfrac{1+\epsilon}{2}\right) - (1-a)H\left(\tfrac{1-(1+\epsilon)a}{2(1-a)}\right) \end{aligned}$$

   Repeated application of L'Hôpital's rule gives the result.

□

Note that analogous lemma holds for Bob sending bit $B_{\mathcal{B}}(\epsilon)$ by exchanging the roles of Alice and Bob. Henceforth, we shall only work with Alice's bits, tacitly assuming analogous results for Bob's bits.

Definition 2.6.2. $B_{\mathcal{A}}(\epsilon)$ is optimal at $\mu_{a,b}$ if the following equality holds:

$$\mathcal{I}(a,b) = \frac{1}{2}\mathcal{I}((1+\epsilon)a, b) + \frac{1}{2}\mathcal{I}((1-\epsilon)a, b) + I(B_{\mathcal{A}}(\epsilon); X|Y).$$

Optimality for $B_{\mathcal{B}}(\epsilon)$ is defined analogously.

In other words, $B_{\mathcal{A}}(\epsilon)$ is optimal at $\mu_{a,b}$ if for every $\delta > 0$ there exists a protocol π such that $\mathrm{IC}_{\mu_{a,b}}(\pi) - \mathcal{I}(a,b) < \delta$ and the first bit of π is $B_{\mathcal{A}}(\epsilon)$, or, intuitively, we can assume that Alice sends $B_{\mathcal{A}}(\epsilon)$ in an "optimal" protocol.

Lemma 2.6.2. *If for some $\epsilon > 0$ $B_{\mathcal{A}}(\epsilon)$ is optimal at $\mu_{a,b}$ then $B_{\mathcal{A}}(\epsilon')$ is optimal at $\mu_{a,b}$ for all $\epsilon' \in [0, \epsilon]$. Analogous statement holds for Bob's bits.*

Proof. Straightforward verification. □

Lemma 2.6.3. *If $\frac{\partial^2 \mathcal{I}}{\partial a^2}$ exists at (a, b) and for some $\epsilon > 0$ $B_{\mathcal{A}}(\epsilon)$ is optimal at $\mu_{a,b}$ then*

$$\left.\frac{\partial^2 \mathcal{I}}{\partial a^2}\right|_{a,b} = \frac{1}{a(a-1)\ln 2}.$$

Analogous statement holds for Bob's optimal bits.

Proof. Using Definition 2.6.2 we have

$$I(B_{\mathcal{A}}(\epsilon); X|Y) = \mathcal{I}(a, b) - \frac{\mathcal{I}((1+\epsilon)a, b) + \mathcal{I}((1-\epsilon)a, b)}{2}$$

Let $f(x) = \mathcal{I}(x, b)$. Then by Taylor's theorem we have

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + o((x-a)^2),$$

Therefore

$$f(a \pm \epsilon a) = f(a) + f'(a)(\pm\epsilon a) + \frac{f''(a)}{2}(\pm\epsilon a)^2 + o((\pm\epsilon a)^2)$$

Thus

$$f(a) - \frac{f(a+\epsilon a) + f(a-\epsilon a)}{2} = -\epsilon^2 a^2 \frac{f''(a)}{2} + o(\epsilon^2 a^2)$$

It follows that

$$f''(a) = \lim_{\epsilon \to 0} \frac{-2}{\epsilon^2 a^2} \left(f(a) - \frac{f(a+\epsilon a) + f(a-\epsilon a)}{2} \right)$$

Note that Lemma 2.6.2 justifies taking the limit above. Translating the above statement in terms of $\mathcal{I}(a, b)$ and applying Lemma 2.6.1, we finally obtain:

$$\left.\frac{\partial^2 \mathcal{I}}{\partial a^2}\right|_{a,b} = \lim_{\epsilon \to 0} \frac{-2I(B_{\mathcal{A}}(\epsilon); X|Y)}{\epsilon^2 a^2} = \frac{1}{a(a-1)\ln 2}$$

□

Definition 2.6.3. We define Alice's regions $\mathcal{R}_{\mathcal{A}}$ to be the following set of distributions:

$$\mathcal{R}_{\mathcal{A}} := \{\mu_{a,b} \mid \exists \epsilon > 0 \; B_{\mathcal{A}}(\epsilon) \text{ is optimal at } \mu_{a,b}\}$$

We define Bob's region $\mathcal{R}_{\mathcal{B}}$ similarly. We also define the diagonal region $\Delta = \{\mu_{a,a}\}$.

In words, $\mathcal{R}_{\mathcal{A}}$ consists of distributions, starting at which Alice sends a small signal in some "optimal" protocol. This definition of regions matches the definitions given in Section 2.5.4. See Figure 2.1 for a pictorial depiction of the regions.

Definition 2.6.4. We say that $\mathcal{I}_{\mathcal{A}} : [0,1]^2 \to \mathbb{R}$ is Alice's extension function if

1. $\mathcal{I}_{\mathcal{A}}|_{\mathcal{R}_{\mathcal{A}}} = \mathcal{I}|_{\mathcal{R}_{\mathcal{A}}}$
2. $\forall a, b \in (0,1)$ we have $\left.\frac{\partial^2 \mathcal{I}_{\mathcal{A}}}{\partial a^2}\right|_{(a,b)} = \frac{1}{a(a-1)\ln 2}$

Bob's extension function is defined similarly.

Assuming that $\mathcal{I}(a,b)$ has second partial derivatives on $(0,1)^2$ in order to find $\mathcal{I}(a,b)$ we need to find a pair of extension functions $\mathcal{I}_{\mathcal{A}}$ and $\mathcal{I}_{\mathcal{B}}$ that satisfy the following conditions:

$$\mathcal{I}_{\mathcal{A}} \circ q(t) = \mathcal{I}_{\mathcal{B}} \circ q(t) \tag{2.5}$$

$$\left.\frac{\partial^i \mathcal{I}_{\mathcal{A}}}{\partial x^i}\right|_{q(t)} = \left.\frac{\partial^i \mathcal{I}_{\mathcal{B}}}{\partial x^i}\right|_{q(t)} \qquad i \in \{1,2\}, x \in \{a,b\} \tag{2.6}$$

$$\left.\frac{\partial^2 \mathcal{I}_{\mathcal{A}}}{\partial a^2}\right|_{a,b} = \frac{1}{a(a-1)\ln 2} \tag{2.7}$$

$$\left.\frac{\partial^2 \mathcal{I}_{\mathcal{B}}}{\partial b^2}\right|_{a,b} = \frac{1}{b(b-1)\ln 2} \tag{2.8}$$

$$\mathcal{I}_{\mathcal{A}}(1,b) = \mathcal{I}_{\mathcal{B}}(a,1) = 0 \tag{2.9}$$

$$\mathcal{I}_{\mathcal{A}}(0,0) = \mathcal{I}_{\mathcal{B}}(0,0) = 0 \tag{2.10}$$

In the above $q(t) = (t,t)$.

Theorem 2.6.4. *The above system has a unique solution:*

$$\mathcal{I}_{\mathcal{A}}(a,b) = \left(b \log \frac{1-b}{b} + \frac{1}{\ln 2} b\right)(1-a) + H(a),$$

$$\mathcal{I}_{\mathcal{B}}(a,b) = \left(a \log \frac{1-a}{a} + \frac{1}{\ln 2} a\right)(1-b) + H(b).$$

Proof. Integrating (3) twice with respect to a we get

$$(\exists C_1, C_2 : [0,1] \to \mathbb{R})\ (\mathcal{I}_{\mathcal{A}}(a,b) = C_1(b) + C_2(b)a + H(a))$$

Using (5) we have $C_1(b) = -C_2(b)$. Thus we can write

$$\mathcal{I}_{\mathcal{A}}(a,b) = C(b)(1-a) + H(a)$$

for some $C : [0,1] \to \mathbb{R}$. Similarly, we get $\mathcal{I}_{\mathcal{B}}(a,b) = D(a)(1-b) + H(b)$ for some $D : [0,1] \to \mathbb{R}$. In fact, by (1) we have $C(t) = D(t)$ for $t \in [0,1)$ and hence $C(t) = D(t)$ on $[0,1]$ by continuity. By (2) and (4) we have

$$C''(t) = \frac{-1}{t(1-t)^2 \ln 2}$$

Solving this ODE, we get

$$(\exists c_0, c_1 \in \mathbb{R})\ \ \left(C(t) = t \log \frac{1-t}{t} + tc_0 + c_1\right)$$

It follows from (6) that $c_1 = 0$. Then by (2) we get

$$-t \log \frac{1-t}{t} - tc_0 + \log \frac{1-t}{t} = \left(\frac{-1}{(1-t)\ln 2} + \log \frac{1-t}{t} + c_0\right)(1-t)$$

Upon simplification the above gives $c_0 = \frac{1}{\ln 2}$. □

Corollary 2.6.5.

$$\mathcal{I}(a,b) = \begin{cases} \left(b \log \frac{1-b}{b} + \frac{1}{\ln 2} b\right)(1-a) + H(a) & \text{if } a \geq b \\ \left(a \log \frac{1-a}{a} + \frac{1}{\ln 2} a\right)(1-b) + H(b) & \text{if } a < b \end{cases}$$

Since the input to $\mathcal{I}_{\mathcal{A}}$, $\mathcal{I}_{\mathcal{B}}$, and $\mathcal{I}$ is 2-dimensional, we can plot graphs of these functions in Euclidean 3-dimensional space. See Figures 2.2, 2.3, 2.4, and 2.5.

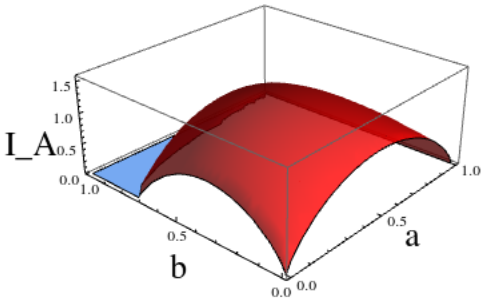

Figure 2.2: Graph of Alice's extension $\mathcal{I}_{\mathcal{A}}$.

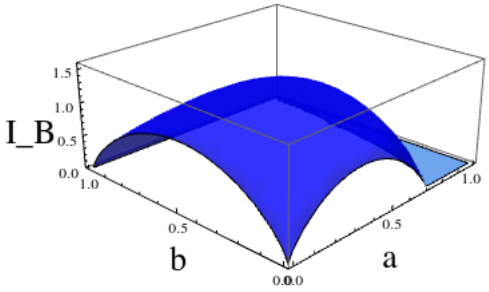

Figure 2.3: Graph of Bob's extension $\mathcal{I}_{\mathcal{B}}$.

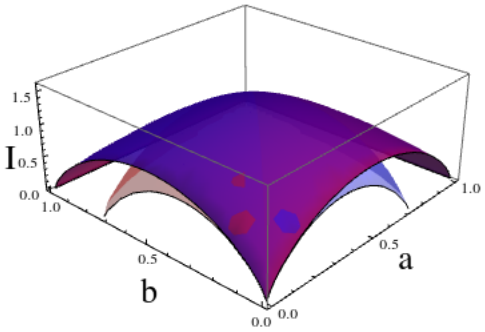

Figure 2.4: The combination of both extensions gives the actual information complexity.

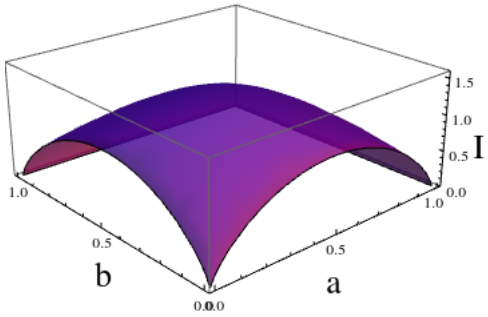

Figure 2.5: Better view of the actual information complexity.

2.6.2 System of PDEs for Information Complexity of General Functions

The main purpose of this section is to generalize the analysis performed in the previous section and state a system of PDEs together with boundary conditions that, we conjecture, characterize the information complexity of general functions. At present, nothing is known about solutions to this system of PDEs, so we limit our discussion to the motivation behind the statement of the system and the statement of the system.

Notation:

- For convenience we shall consider functions f of the form $f : [N] \times [N] \to \{0, 1\}$.

- For probability distribution μ on $[N] \times [N]$ we define $\mu_{\mathcal{A}}(i) := \sum_{j=1}^{N} \mu(i,j)$ and $\mu_{\mathcal{B}}(j) := \sum_{i=1}^{N} \mu(i,j)$.
- We shall write μ to refer to the distribution as well as its $N \times N$ matrix.
- We shall use $[\mu]$ to denote the linearization of the matrix μ, i.e., $[\mu]$ is the $N^2 \times 1$ column vector indexed by (i,j). We shall omit the square brackets when there is little chance of ambiguity.
- We use $E_{i,j}$ to denote the $N \times N$ matrix that is 0 everywhere except entry (i,j), which is 1.
- We define $\mathcal{I}(\mu) := \mathrm{IC}_{\mu}(f, 0)$, and we consider it as a function of N^2 (in reality only $N^2 - 1$) variables, where we denote an individual variable by its index (i,j).
- $\mathrm{Hes}\,\mathcal{I}(\mu)$ refers to the Hessian of the function $\mathcal{I}$ at μ, i.e., it is the $N^2 \times N^2$ matrix such that the entry at the intersection of row (i_1, j_1) and column (i_2, j_2) is $\frac{\partial^2 \mathcal{I}}{\partial(i_1,j_1)\partial(i_2,j_2)}\Big|_{\mu}$.

Definition 2.6.5. For $\epsilon \in [-1,1], i,j \in [N]$ define $B_{\mathcal{A}}(\epsilon, i, j)$ to be the random variable such that

- $P(B_{\mathcal{A}}(\epsilon,i,j) = 0 | X = i) = \frac{1}{2} + \frac{\epsilon}{2}$,
- $P(B_{\mathcal{A}}(\epsilon,i,j) = 0 | X = j) = \frac{1}{2} - \frac{\epsilon \mu_{\mathcal{A}}(i)}{2\mu_{\mathcal{A}}(j)}$,
- $P(B_{\mathcal{A}}(\epsilon,i,j) = 0 | X = k) = \frac{1}{2}$ for $k \neq i, j$,
- $P(B_{\mathcal{A}}(\epsilon,i,j) = 0) = P(B_{\mathcal{A}}(\epsilon,i,j) = 1) = \frac{1}{2}$.

$B_{\mathcal{B}}(\epsilon,i,j)$ is defined analogously.

Definition 2.6.6. We say $B_{\mathcal{A}}(\epsilon,i,j)$ is optimal at μ if the following equality holds

$$\mathcal{I}(\mu) = \frac{1}{2}\mathcal{I}(\mu_0) + \frac{1}{2}\mathcal{I}(\mu_1) + I(B_{\mathcal{A}}(\epsilon,i,j); X|Y),$$

where $\mu_z(x,y) = P(X = x, Y = y | B_{\mathcal{A}}(\epsilon,i,j) = z)$ for $z \in \{0,1\}$. We define optimality of $B_{\mathcal{B}}(\epsilon,i,j)$ at μ analogously.

A quick calculation reveals:

$$\begin{aligned}
\mu_0(x,y) &= \frac{P(B_{\mathcal{A}}(\epsilon,i,j)=0|X=x,Y=y)}{P(B_{\mathcal{A}}(\epsilon,i,j)=0)}\mu(x,y) && \text{Bayes' thm}\\
&= \frac{P(B_{\mathcal{A}}(\epsilon,i,j)=0|X=x)}{P(B_{\mathcal{A}}(\epsilon,i,j)=0)}\mu(x,y) && \text{independence}\\
&= \begin{cases} (1+\epsilon)\mu(x,y) & x=i \\ \left(1-\epsilon\frac{\mu_{\mathcal{A}}(i)}{\mu_{\mathcal{A}}(j)}\right)\mu(x,y) & x=j \\ \mu(x,y) & x\neq i,j \end{cases}
\end{aligned}$$

Similarly, we have

$$\mu_1(x,y) = \begin{cases} (1-\epsilon)\mu(x,y) & x=i \\ \left(1+\epsilon\frac{\mu_{\mathcal{A}}(i)}{\mu_{\mathcal{A}}(j)}\right)\mu(x,y) & x=j \\ \mu(x,y) & x\neq i,j \end{cases}$$

Using the notation introduced at the beginning of this section we can write

$$\mu_z = \mu + (-1)^z\epsilon\left(E_{i,i} - \frac{\mu_{\mathcal{A}}(i)}{\mu_{\mathcal{A}}(j)}E_{j,j}\right)\mu$$

Lemma 2.6.6. *If the second derivative of $\mathcal{I}$ exists at μ and $\exists\epsilon>0$ such that $B_{\mathcal{A}}(\epsilon,i,j)$ is optimal at μ then*

$$\lim_{\epsilon\to 0}\frac{-2I(B_{\mathcal{A}}(\epsilon,i,j);X|Y)}{\epsilon^2} = \left[\left(E_{i,i} - \frac{\mu_{\mathcal{A}}(i)}{\mu_{\mathcal{A}}(j)}E_{j,j}\right)\mu\right]^t \operatorname{Hes}\mathcal{I}(\mu)\left[\left(E_{i,i} - \frac{\mu_{\mathcal{A}}(i)}{\mu_{\mathcal{A}}(j)}E_{j,j}\right)\mu\right]$$

Analogous statement holds for Bob's optimal bits.

Proof. Applying Taylor's theorem for $\mathcal{I}$ at μ we have

$$\mathcal{I}(\nu) = \mathcal{I}(\mu) + \nabla\mathcal{I}(\mu)\cdot[(\nu-\mu)] + \frac{1}{2}[(\nu-\mu)]^t\operatorname{Hes}\mathcal{I}(\mu)[(\nu-\mu)] + o(||\nu-\mu||_2^2)$$

Using the above statement with $\nu=\mu_z$, plugging the resulting formulas into Definition 2.6.6, rearranging, taking limits, and simplifying gives the desired conclusion. □

We need a few last definitions before we can state the system of PDEs:

Definition 2.6.7. We define Alice's boundary conditions $\text{Bound}_{\mathcal{A}}$ to be the following set

$$\text{Bound}_{\mathcal{A}} = \{\mu \mid \forall i, j_1, j_2 \in [N] \quad (i, j_1), (i, j_2) \in \text{supp}(\mu) \Rightarrow f(i, j_1) = f(i, j_2)\}$$

We define Bob's boundary conditions $\text{Bound}_{\mathcal{B}}$ analogously.

Definition 2.6.8. We define the internal boundary of regions of Alice and Bob as

$$\begin{aligned}\mathcal{R}_{\mathcal{AB}} = \{\mu \mid \text{supp}(\mu) = [N] \times [N] \text{ and} \\ \forall \epsilon \in (0, 1], i, j \in [N] \text{ neither } B_{\mathcal{A}}(\epsilon, i, j) \text{ nor } B_{\mathcal{B}}(\epsilon, i, j) \text{ is optimal at } \mu\}\end{aligned}$$

Finally, we are ready to state the system of PDEs for the information complexity of general functions. The two unknown functions are going to be $\mathcal{I}_{\mathcal{A}}$ and $\mathcal{I}_{\mathcal{B}}$, which should be thought of as Alice's and Bob's extensions of $\mathcal{I}$, respectively.

System of PDEs for information complexity of general functions:

$$\begin{aligned}&\left[\left(E_{i,i} - \frac{\mu_{\mathcal{A}}(i)}{\mu_{\mathcal{A}}(j)} E_{j,j}\right)\mu\right]^t \text{Hes}\, \mathcal{I}_{\mathcal{A}}(\mu) \left[\left(E_{i,i} - \frac{\mu_{\mathcal{A}}(i)}{\mu_{\mathcal{A}}(j)} E_{j,j}\right)\mu\right] = \\ &= \lim_{\epsilon \to 0} \frac{-2I(B_{\mathcal{A}}(\epsilon, i, j); X|Y)}{\epsilon^2} \quad \forall i, j \in [N] \\ &\left[\left(E_{i,i} - \frac{\mu_{\mathcal{B}}(i)}{\mu_{\mathcal{B}}(j)} E_{j,j}\right)\mu\right]^t \text{Hes}\, \mathcal{I}_{\mathcal{B}}(\mu) \left[\left(E_{i,i} - \frac{\mu_{\mathcal{B}}(i)}{\mu_{\mathcal{B}}(j)} E_{j,j}\right)\mu\right] = \\ &= \lim_{\epsilon \to 0} \frac{-2I(B_{\mathcal{B}}(\epsilon, i, j); X|Y)}{\epsilon^2} \quad \forall i, j \in [N] \\ &\mathcal{I}_{\mathcal{A}}(\mu) = H_{(i,j)\sim\mu}(f(i, j)) \quad \forall \mu \in \text{Bound}_{\mathcal{A}} \\ &\mathcal{I}_{\mathcal{B}}(\mu) = H_{(i,j)\sim\mu}(f(i, j)) \quad \forall \mu \in \text{Bound}_{\mathcal{B}} \\ &D^z \mathcal{I}_{\mathcal{A}}(\mu) = D^z \mathcal{I}_{\mathcal{B}}(\mu) \quad \forall z \in \{0, 1, 2\} \ \ \forall \mu \in \mathcal{R}_{\mathcal{AB}}\end{aligned}$$

2.7 Rate of Convergence of $\text{IC}^r_\mu(\text{AND}, 0)$ to $\text{IC}_\mu(\text{AND}, 0)$

In this section we prove that for most distributions μ the rate at which $\text{IC}^r_\mu(\text{AND}, 0)$ converges to $\text{IC}_\mu(\text{AND}, 0)$ is $\Theta(1/r^2)$. The empirical evidence that the rate of convergence is $\Theta(1/r^2)$ has appeared in the information theory literature prior to our

work. In [37], Ma and Ishwar consider the task f of computing AND when only Bob is required to learn the answer. They derive an explicit formula for $\mathrm{IC}_\mu(f)$ for product distributions μ and design an algorithm that computes $\mathrm{IC}^r_\mu(f)$ to within a desired accuracy. Ishwar and Ma generously provided their scripts, which we used to generate Figure 2.6 (it is a variant of Figure 4(a) from [37]). Figure 2.6 demonstrates that $\max_{\mu \text{ - product}} \mathrm{IC}^r_\mu(f) - \mathrm{IC}_\mu(f)$ asymptotically behaves like $\Theta(1/r^2)$.

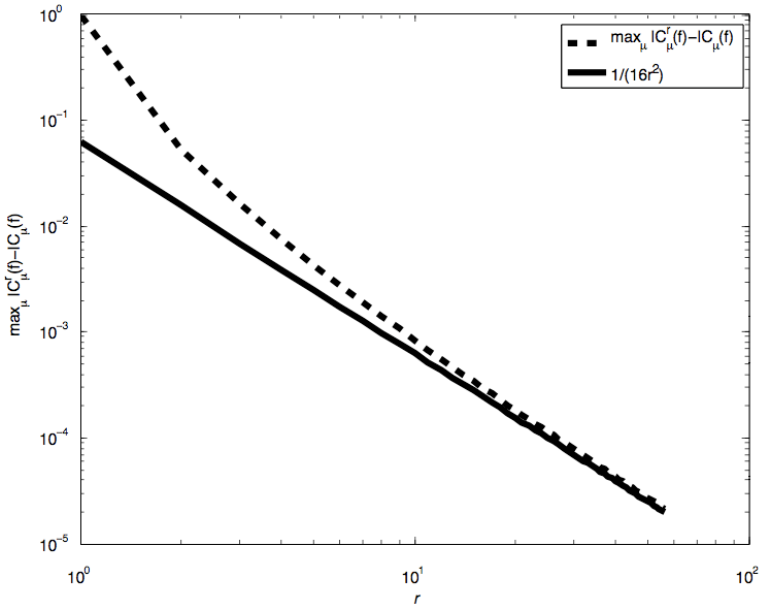

Figure 2.6: Empirical evidence that rate of convergence is $\Theta(1/r^2)$. The log-log scale figure shows the graph of $\max_{\mu \text{ - product}} \mathrm{IC}^r_\mu(f) - \mathrm{IC}_\mu(f)$ for a range of values r together with the line $1/(16r^2)$. The x-axis is the number of rounds r. The y-axis is the change in the information cost $\max_{\mu \text{ - product}} \mathrm{IC}^r_\mu(f) - \mathrm{IC}_\mu(f)$.

Our proof consists of two main parts: (1) the lower bound $\Omega(1/r^2)$ on the rate of convergence and (2) the matching upper bound of $O(1/r^2)$. The high-level idea for the lower bound is to show that any r-round protocol, when viewed as a random walk on $\Delta(\mathcal{X} \times \mathcal{Y})$ (see Section 1.5), has to travel a large distance in the wrong region. In other words, Alice often speaks in Bob's region, and Bob often speaks in Alice's region. Then we can use the formulas from Section 2.5.6 to conclude that each such step wastes a lot of information as compared to the optimal protocol. Aggregating this wastage over all rounds, $\Omega(1/r^2)$ information has to be wasted overall. The

upper bound is obtained by carefully analyzing a discretized version of our infeasible protocol for AND from Section 2.5.3. Both upper and lower bounds require a number of technical lemmas, which we also include in the text.

The rest of this section is organized as follows. In Subsection 2.7.1 we prove the lower bound on the rate of convergence modulo two technical lemmas. Subsection 2.7.2 contains the proof of the first lemma, which quantifies how much information is wasted by a feasible protocol versus an optimal infeasible one in terms of the distance traveled in the wrong region. Subsection 2.7.3 proves the second technical lemma from the lower bound on the rate of convergence. The second lemma gives a lower bound on the distance traveled in the wrong region by a protocol that solves the AND function. Finally, in Subsection 2.7.4 we prove the upper bound on the rate of convergence.

In this section it will be easier for us to work with general protocols and forgo the normal-form assumption.

2.7.1 Lower Bound on the Rate of Convergence

We say that a message M *crosses the diagonal* if this message starts at prior μ, has distribution on distributions $(\{\mu_m\}, \{p_m\})$, and there exists m such that the interval $[\mu, \mu_m]$ intersects the diagonal region, i.e., the interval $[\mu, \mu_m]$ contains a symmetric distribution.

We begin by showing that we can split a message that crosses the diagonal into two that do not cross the diagonal.

Lemma 2.7.1 ([9])**.** *Let M be a message sent by one of the players such that M crosses the diagonal. There exists two messages M_1 and M_2 such that neither M_1, nor M_2 crosses the diagonal, and (M_1, M_2) has the same distribution on distributions as M.*

Proof. The idea of the proof is that each message M is simply a sequence of bits, so the player can generate M bit by bit until there is a danger of the next bit crossing the diagonal. If the player is about to generate a crossing bit, the player will instead split that bit into two using the Splitting Lemma (Lemma 2.5.6). The split happens

in such a way that after the first bit is sent the player either ends up on the diagonal, or moves away from the diagonal. If the player does not jump to the diagonal, then the process continues in the same way. If the player happens to jump to the diagonal that signifies the end of message M_1 and beginning of M_2.

All that is left to show is that a crossing bit may be split into two non-crossing bits while preserving the distribution on distributions. Suppose that the player sends a bit B starting at prior μ and splitting μ into μ_0 and μ_1, such that $[\mu, \mu_1]$ contains a symmetric distribution μ_D. Since $\mu \in [\mu_0, \mu_D]$ there is a signal B_1 that splits μ into μ_0 and μ_D (by the Splitting Lemma). Also, since $\mu_D \in [\mu_0, \mu_1]$ there is a signal B_2 that splits μ_D into μ_0 and μ_1. Now instead of sending bit B, the player first sends B_1. If $B_1 = 0$ the message is terminated, otherwise the player sends B_2. This new message induces the same distribution on distributions as B, because (B_1, B_2) and B express μ as a convex combination of μ_0, μ_1, which is unique. Note that we allow B, B_1 and B_2 be biased. □

Theorem 2.7.2 ([9]). *For all* $\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \gamma & \delta \\ \hline \end{array}$ *with* $\{\alpha, \beta, \gamma\} \subseteq \mathrm{supp}(\mu)$ *we have*

$$\mathrm{IC}^r_\mu(\mathrm{AND}, 0) = \mathrm{IC}_\mu(\mathrm{AND}, 0) + \Omega_\mu\left(\frac{1}{r^2}\right).$$

Proof. Fix an arbitrary r-round protocol π that solves AND with 0-error and distribution $\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \gamma & \delta \\ \hline \end{array}$ with $\alpha, \beta, \gamma \neq 0$. Using Lemma 2.7.1 we obtain a protocol π' with $m \leq 2r$ messages, such that no message crosses the diagonal and $\mathrm{IC}_\mu(\pi') = \mathrm{IC}_\mu(\pi)$. We shall view π' as a random walk on the set of distributions $\Delta(\{0,1\} \times \{0,1\})$. For technical reasons we shall restrict this random walk to the subset S of $\Delta(\{0,1\}^2)$ defined as follows

$$S := \{\mu' \mid \alpha' \geq 0.01\alpha \text{ and } \min(\beta', \gamma') \geq 0.01 \min(\beta, \gamma)\}.$$

Using ideas from the proof of Lemma 2.7.1 we can always impose a constraint that π' does not make steps that cross from S into $\overline{S}$ without stopping at the boundary

of S. We let π'' denote such a modification of π'. Clearly, $\mathrm{IC}_\mu(\pi'') = \mathrm{IC}_\mu(\pi')$ and the number of messages in the first part of π'' that proceeds only until the boundary of S is at most m.

We shall show that $\mathrm{IC}_\mu(\pi'') = \mathrm{IC}_\mu(\mathrm{AND}, 0) + \Omega_\mu\left(\frac{1}{r^2}\right)$ by showing that the part of π'' until the boundary of S already wastes $\Omega_\mu\left(\frac{1}{r^2}\right)$ amount of information as compared to the optimal protocol.

Let T_i denote the ith message of π'' for $i \leq m$. Then the whole transcript T until the boundary of S is $T_1 T_2 \cdots T_m$.

A transcript t gives rise to $m+1$ distributions $\mu_0^{t_{\leq 0}}, \mu_1^{t_{\leq 1}}, \ldots, \mu_m^{t_{\leq m}}$ traced out by the protocol π when viewed as a random walk on $\Delta(\mathcal{X} \times \mathcal{Y})$ (see Section 1.5). We write $\mu_0^{t_{\leq 0}} = \mu$. We define the central object of this proof: $\delta_i^{t_{\leq i}}$ - the distance traveled by a player in the wrong region during the ith round. More formally

$$\delta_i^{t_{\leq i}} = \begin{cases} \|[\mu_{i-1}^{t_{\leq i-1}}, \mu_i^{t_{\leq i}}] \cap \Delta_{\mathcal{A}}\| & \text{if the } i\text{th message is} \\ & \text{transmitted by Bob,} \\ \|[\mu_{i-1}^{t_{\leq i-1}}, \mu_i^{t_{\leq i}}] \cap \Delta_{\mathcal{B}}\| & \text{if the } i\text{th message is} \\ & \text{transmitted by Alice.} \end{cases}$$

In the above $\Delta_{\mathcal{A}}$ denotes Alice's region, and $\Delta_{\mathcal{B}}$ denotes Bob's region (see Section 2.5.4). The lower bound $\Omega_\mu(1/r^2)$ on the overall wastage of protocol π'' follows from two crucial observations:

Lemma 2.7.3 ([9])**.**

$$\mathrm{IC}_\mu(\pi'') - \mathrm{IC}_\mu(\mathrm{AND}, 0) = \Omega_\mu\left(\sum_{i=1}^{m} (\mathbb{E}_t \delta_i^{t_{\leq i}})^3\right).$$

Lemma 2.7.4 ([9])**.**

$$\mathbb{E}_t\left(\sum_{i=1}^{m} \delta_i^{t_{\leq i}}\right) = \Omega_\mu(1).$$

We prove the above lemmas later in Subsections 2.7.2 and 2.7.3. Now, by Hölder's

inequality we have

$$\sum_{i=1}^{m} \left(\mathbb{E}_t \delta_i^{t\leq i}\right)^3 \geq \left(\mathbb{E}_t \left(\sum_{i=1}^{m} \delta_i^{t\leq i}\right)\right)^3 /m^2 = \Omega_\mu(1/r^2),$$

where the last step follows from Lemma 2.7.4 and the fact that $m \leq 2r$. This finishes the proof by Lemma 2.7.3. □

2.7.2 Informational Wastage

The goal of the current subsection is to prove Lemma 2.7.3 that appears in the proof of the lower bound on the rate of convergence. For definitions of relevant mathematical objects see Subsection 2.7.1. Recall that Lemma 2.7.3 asserts that the information wasted by an m-message protocol as compared to the optimal infeasible protocol is roughly the sum of the cubed distances traveled in the wrong region. The proof of this lemma consists of a sequence of reductions. We start with analyzing how much information is wasted by a single bit and gradually build up the result to the entire protocol.

We start by formally defining what it means for a particular step in a protocol, which consists of one of the players sending a message, to waste information.

Definition 2.7.1 ([9])**.** Suppose that Bob sends message M with distribution on distributions $(\{\mu_m\}, \{p_m\})$ from prior μ. *The informational wastage* of M is defined as

$$\mathrm{IW}(\mu, M) := \left(\sum_{m \in \mathrm{range}(M)} p_m \, \mathrm{IC}_{\mu_m}(\mathrm{AND}, 0)\right) + I(M; Y|X) - \mathrm{IC}_\mu(\mathrm{AND}, 0).$$

The informational wastage of Alice's messages is defined analogously.

The information wasted is how much extra information is revealed by a protocol that sends message M and then proceeds optimally versus the protocol that proceeds optimally from the start.

When the message is a single bit B sent by Bob from a symmetric prior μ, the above definition simplifies to

$$\mathrm{IW}(\mu, B) = p\,\mathrm{IC}_{\mu_0}(\mathrm{AND}, 0) + (1-p)\,\mathrm{IC}_{\mu_1}(\mathrm{AND}, 0) + I(B; Y|X) - \mathrm{IC}_{\mu}(\mathrm{AND}, 0),$$

where $\mu_0 := P(X = x, Y = y|B = 0)$ belongs to Bob's region and $\mu_1 := P(X = x, Y = y|B = 1)$ belongs to Alice's region.

Observe that the formulas from Section 2.5.6 imply that for a *uniform bit* B and symmetric prior μ we have

$$\mathrm{IC}(\mu, B) \geq C(\mu)||\mu_1 - \mu||^3 = \Omega(||\mu_1 - \mu||^3), \tag{2.11}$$

where $C(\mu) = \frac{\alpha\beta}{12(\alpha+\beta)(1-\alpha-\beta)^3}$ is a continuous positive function of μ. In other words, the information wasted is roughly the cube of the distance traveled in the wrong region.

Remark 2.7.1. In what follows we only consider the information wasted from a symmetric prior, because the information wasted when a player speaks starting in the wrong region is strictly larger (see formulas at the end of Section 2.5.6).

Now we extend this result to *nonuniform* bits. As expected, for a nonuniform bit the cube of the distance traveled in the wrong region gets scaled by the probability of jumping into the wrong region.

Lemma 2.7.5 ([9])**.** *Suppose that Bob sends bit B from symmetric prior μ with distribution on distributions $(\{\mu_0, \mu_1\}, \{p, 1-p\})$. If $\mu_1 + 2p(\mu_0 - \mu_1) \in \Delta(\{0,1\} \times \{0,1\})$ then*

$$\mathrm{IW}(\mu, B) \geq C(\mu)(1-p)||\mu_1 - \mu||^3 = \Omega((1-p)||\mu_1 - \mu||^3),$$

where $C(\mu) = \frac{\alpha\beta}{12(\alpha+\beta)(1-\alpha-\beta)^3}$. Similarly for Alice.

Proof. **Case $p \leq 1/2$.** Let $\mu_0' := \mu_1 + 2p(\mu_0 - \mu_1) \in [\mu_0, \mu_1]$. Then we have $\mu = (1/2)\mu_1 + (1/2)\mu_0'$, so there exists signal B' that Bob can send with distribution on

distributions $(\{\mu_0', \mu_1\}, \{1/2, 1/2\})$. Clearly, we have $\mathrm{IW}(\mu, B) \geq \mathrm{IW}(\mu, B')$. Finally, from Equation (2.11) we obtain $\mathrm{IW}(\mu, B') \geq C(\mu)||\mu_1 - \mu||^3 \geq C(\mu)(1-p)||\mu_1 - \mu||^3$.

Case $p > 1/2$**.** Let $\mu_0' := \mu_1 + 2p(\mu_0 - \mu_1)$. By conditions of the lemma, μ_0' is a valid distribution. Then we have $\mu_0 := ((1-p)/p)\mu_0' + ((2p-1)/p)\mu$, so there exists bit B' that Bob can send from prior μ_0 with distribution on distributions $(\{\mu_0', \mu\}, \{(1-p)/p, (2p-1)/p\})$. By Claim 2.5.9 we have $\mathrm{IW}(\mu_0, B') = 0$ thus

$$\mathrm{IW}(\mu, B) = \mathrm{IW}(\mu, B) + p\,\mathrm{IW}(\mu_0, B') = \mathrm{IW}(\mu, M),$$

where M is message (B, B') that has distribution on distributions $(\{\mu_0', \mu, \mu_1\}, \{1-p, 2p-1, 1-p\})$. Since $\mu = (1/2)\mu_0' + (1/2)\mu_1$ there exists the signal B'' with distribution on distributions $(\{\mu_0', \mu_1\}, \{1/2, 1/2\})$.
Define $I(\nu) := \mathrm{IC}_\nu(\mathrm{AND}, 0)$. Then we have

$$\begin{aligned}
\mathrm{IW}(\mu, M) &= (1-p)I(\mu_0') + (1-p)I(\mu_1) + (2p-1)I(\mu) + I(M;Y|X) - I(\mu) \\
&= 2(1-p)[(1/2)I(\mu_0') + (1/2)I(\mu_1) + I(M;Y|X)/(2(1-p)) - I(\mu)] \\
&= 2(1-p)[(1/2)I(\mu_0') + (1/2)I(\mu_1) + I(B'';Y|X) - I(\mu)] \\
&= 2(1-p)\,\mathrm{IW}(\mu, B'') \\
&\geq 2(1-p)C(\mu)||\mu_1 - \mu||^3,
\end{aligned}$$

$I(M;Y|X)/(2(1-p)) = I(B'';Y|X)$, since sending M and *staying at prior* μ *with probability* $2(1-p)$ *and otherwise sending* B'' induce the same distribution on distributions.The last step follows from Equation (2.11). □

The next step is to extend our lower bound on the information wasted in a single step of a protocol to messages. Suppose that Bob sends a message M with distribution on distributions $(\{\mu_m\}, \{p_m\})$ from symmetric prior μ. Define sets $\mathcal{M}_1 := \{m \mid \mu_m(0,1) > \mu_m(1,0)\}$ and $\mathcal{M}_2 := \{m \mid \mu_m(0,1) \leq \mu_m(1,0)\}$. The set $\mathcal{M}_1$ contains all the messages that lead to Alice's region and $\mathcal{M}_2$ contains all the messages that lead to Bob's region. Let μ_1 be the average of $\mu_m \in \mathcal{M}_1$ and μ_0 to be the average of μ_m in $\mathcal{M}_2$. If $\mu_1 + 2p(\mu_0 - \mu_1) \in \Delta(\{0,1\}^2)$ then the information wasted by sending M

is at least $\Omega(P(m \in \mathcal{M}_1)||\mu_1 - \mu||^3) = \Omega((\mathbb{E}_m||[\mu_m, \mu] \cap \Delta_A||)^3)$.

Lemma 2.7.6 ([9]). *Suppose that the conditions specified in the above paragraph hold for a message M sent by Bob, then we have*

$$\mathrm{IW}(\mu, M) \geq C(\mu)(\mathbb{E}_m||[\mu_m, \mu] \cap \Delta_A||)^3,$$

where $C(\mu) = \frac{\alpha\beta}{12(\alpha+\beta)(1-\alpha-\beta)^3}$. Similar inequality holds for Alice.

Proof. Define the indicator random variable Z as follows

$$Z = \begin{cases} 1 & \text{if } \mu_M(0,1) > \mu_M(1,0) \\ 0 & \text{otherwise} \end{cases}$$

In other words, Z indicates if after sending M the players end up in Alice's region.

Consider the two protocols:

1. π_1 - Bob first sends M and then players proceed optimally.
2. π_2 - Bob first sends Z, then $M|Z$ and then players proceed optimally.

Clearly, sending Z followed by $M|Z$ produces the same distribution on distributions as simply sending M, thus π_1 and π_2 have the same information cost. Therefore they have the same informational wastage. Observe that if Bob sends $Z = 1$ then the players update their distribution μ to distribution $\mu_1 = \mathbb{E}_{m \sim M|Z=1}(\mu_m)$. It is easy to see that $||\mu_1 - \mu|| = \mathbb{E}_{m \sim M|Z=1}(||\mu_m - \mu||)$ (note that this matches the definition of μ_1 we gave in a paragraph prior to the statement of this lemma). Now we are in a position to apply Lemma 2.7.5 to the bit Z. All in all, we have $\mathrm{IW}(\mu, (Z, M|Z)) \geq \mathrm{IW}(\mu, Z) \geq P(Z = 1)C(\mu)||\mu_1 - \mu||^3 \geq C(\mu)(P(Z = 1)\mathbb{E}_{m \sim M|Z=1}(||\mu_m - \mu||))^3 \geq C(\mu)(\mathbb{E}_m||[\mu_m, \mu] \cap \Delta_A||)^3$. □

Now we are in a position to prove Lemma 2.7.3.

Lemma (2.7.3 restated,[9]).

$$\mathrm{IC}_\mu(\pi'') - \mathrm{IC}_\mu(\mathrm{AND}, 0) = \Omega_\mu\left(\sum_{i=1}^{m}(\mathbb{E}_t\delta_i^{t \leq i})^3\right).$$

Proof of Lemma 2.7.3. By Lemma 2.7.6 the informational waste of the ith message T_i given a *fixed partial transcript* $t_{\leq i-1}$ is at least

$$C(\mu_{i-1}^{t_{\leq i-1}})(\mathbb{E}_{t_i \sim T_i | t_{\leq i-1}}(\delta_i^{t_{\leq i}}))^3 \geq \frac{(0.01\alpha)(0.01\min(\beta,\gamma))}{12}(\mathbb{E}_{t_i \sim T_i | t_{\leq i-1}}(\delta_i^{t_{\leq i}}))^3,$$

where the last step follows from the fact that $\mu_{i-1}^{t_{\leq i-1}} \in S$ and $C(\mu') \geq \frac{\alpha'\beta'}{12}$ (see proof of Theorem 2.7.2 for the relevant definitions). Aggregating this over all messages T_i we finish the proof of the lemma

$$\mathrm{IC}_\mu(\pi'') - \mathrm{IC}_\mu(\mathrm{AND}, 0) \geq \frac{(0.01\alpha)(0.01\min(\beta,\gamma))}{12}\left(\sum_{i=1}^{m}(\mathbb{E}_t \delta_i^{t_{\leq i}})^3\right).$$

□

2.7.3 Distance Traveled in the Wrong Region

The goal of the current subsection is to prove Lemma 2.7.4 appearing in the proof of the lower bound on the rate of convergence. See Subsection 2.7.1 for the relevant definitions. Lemma 2.7.4 asserts that when a protocol is viewed as a random walk on the space of distributions, the protocol has to spend non-trivial amount of time in the wrong region if it solves the AND function.

The proof relies on the following observation. Consider a protocol that solves AND correctly on all inputs. We view it as a random walk on the space of distributions (see Section 1.5). Recall that a single move multiplies rows or columns of the current distribution. Thus if the random walk starts from a non-trivial distribution (i.e., the players cannot derive the answer to AND from it immediately), the protocol would have to multiply some row or column by 0. This immediately implies that a protocol solving AND correctly has to travel statistical distance at least $\min(\beta,\gamma)$ overall. A more careful analysis reveals that in fact such a protocol has to travel $\min(\beta,\gamma)$ in the "wrong region". This is proved in this section via an invariant argument (see Lemma 2.7.8). We start by proving the following lemma, which shows that a certain process defined by a random walk of the protocol is a supermartingale.

Lemma 2.7.7 ([9]). *Let π be a protocol that starts at prior μ. For a (partial) transcript t, let $\mu_t = \begin{array}{|c|c|}\hline \alpha_t & \beta_t \\ \hline \gamma_t & \delta_t \\ \hline \end{array}$ denote the resulting distribution arising from t. Then $\beta_T\gamma_T$ is a supermartingale.*

Proof. Let B be a bit sent by Bob from μ. Then $\mu_i(x,y) = P(X = x, Y = y|B = i)$ for $i \in \{0,1\}$. We need to show that

$$\mathbb{E}_{b\sim B}(\beta_b\gamma_b) \leq \beta\gamma.$$

Let $p := P(B = 0)$. Recall that the jth column of μ_i is simply a multiple of the jth column of μ. We can write $\mu_0 = \begin{array}{|c|c|}\hline C_0\alpha & C_1\beta \\ \hline C_0\gamma & C_1\delta \\ \hline \end{array}$ and $\mu_1 = \begin{array}{|c|c|}\hline D_0\alpha & D_1\beta \\ \hline D_0\gamma & D_1\delta \\ \hline \end{array}$, where $C_i = P(B = 0|Y = i)/P(B = 0)$ and $D_i = P(B = 1|Y = i)/P(B = 1)$. Observe that $D_i = (1 - C_ip)/(1 - p)$. Therefore we have

$$\begin{aligned}
\mathbb{E}_{b\sim B}(\beta_b\gamma_b) &= pC_0C_1\beta\gamma + (1-p)D_0D_1\beta\gamma \\
&= \beta\gamma(pC_0C_1 + (1 - C_0p)(1 - C_1p)/(1 - p)) \\
&= \beta\gamma((1-p)pC_0C_1 + (1 - C_0p)(1 - C_1p))/(1 - p) \\
&= \beta\gamma(1 - p + (C_1 - 1)(C_0 - 1)p)/(1 - p) \\
&= \beta\gamma(1 + (C_1 - 1)(C_0 - 1)p/(1 - p)) \\
&\leq \beta\gamma,
\end{aligned}$$

where the last step follows from the fact that $C_i \leq 1 \iff C_{1-i} \geq 1$, so $(C_1 - 1)(C_0 - 1) \leq 0$. □

The next lemma proves an invariant of a protocol solving the AND function. The lemma says that in order for a protocol to decrease the value of $\min(\beta,\gamma)$ by a certain amount, the protocol has to spend an equivalent amount of time in the wrong region.

Lemma 2.7.8 ([9]).

$$\mathbb{E}_t\left(\min(\beta_m^t, \gamma_m^t) - \min(\beta,\gamma) + \sum_{i=1}^{m} \delta_i^{t\leq i}\right) \geq 0.$$

Proof. We prove the claim for all m-message protocols π and for all distributions μ by induction on m.

Base case is obvious, because it happens when $m = 0$ and we have $\min(\beta_0^t, \gamma_0^t) = \min(\beta, \gamma)$.

Now, consider the inductive step. We have

$$\begin{aligned}
&\mathbb{E}_t(\min(\beta_m^t, \gamma_m^t) - \min(\beta, \gamma) + \sum_{i=1}^{m} \delta_i^{t_{\leq i}}) \\
&= \mathbb{E}_t(\min(\beta_m^t, \gamma_m^t) - \min(\beta_1^{t_1}, \gamma_1^{t_1}) + \sum_{i=2}^{m} \delta_i^{t_{\leq i}} + \min(\beta_1^{t_1}, \gamma_1^{t_1}) - \min(\beta, \gamma) + \delta_1^{t_1}) \\
&\geq \mathbb{E}_{t_1}(\min(\beta_1^{t_1}, \gamma_1^{t_1}) - \min(\beta, \gamma) + \delta_1^{t_1}),
\end{aligned}$$

where the last step follows by induction. To complete the inductive step it is left to show that $\mathbb{E}_{t_1}(\min(\beta_1^{t_1}, \gamma_1^{t_1}) - \min(\beta, \gamma) + \delta_1^{t_1}) \geq 0$. We shall assume that the first message is sent by Bob. The case when Alice sends the first message is similar.

There are two possibilities, which we analyze separately.

First possibility is that μ is not a symmetric prior. So μ either belongs to Bob's region, or Alice's region. Consider the case when μ belongs to Bob's region ($\gamma > \beta$). Then $\min(\beta, \gamma) = \gamma$. Moreover, since the messages in our protocol do not cross the diagonal, we have that $\gamma_1^{t_1} \geq \beta_1^{t_1}$ for all $t_1 \in T_1$. Consequently $\min(\beta_1^{t_1}, \gamma_1^{t_1}) = \gamma_1^{t_1}$. Since $\gamma_i^{T_i}$ is a martingale, we have

$$\mathbb{E}_{t_1}(\min(\beta_1^{t_1}, \gamma_1^{t_1}) - \min(\beta, \gamma)) = 0.$$

Adding $\mathbb{E}_{t_1}(\delta_1^{t_1})$ to the above only increases the right-hand side. Similar calculation works for the case when μ belongs to Alice's region.

Second possibility is that μ is a symmetric prior, i.e., $\gamma = \beta$. Recall that the prior gets modified by multiplying the columns:

$$\begin{aligned}
\mu_1^{t_1}(x, y) &= P(X = x, Y = y | T_1 = t_1) \\
&= \tfrac{P(T_1 = t_1 | X = x, Y = y)}{P(T_1 = t_1)} P(X = x, Y = y) \\
&= \tfrac{P(T_1 = t_1 | Y = y)}{P(T_1 = t_1)} \mu(x, y).
\end{aligned}$$

Thus on the first message t_1 the first column of μ gets multiplied by $C_0^{t_1} := P(T_1 = t_1|Y = 0)/P(T_1 = t_1)$ and the second column gets multiplied by $C_1^{t_1} := P(T_1 = t_1|Y = 1)/P(T_1 = t_1)$. Next we define two sets of messages $\mathcal{S} := \{t_1|C_0^{t_1} < C_1^{t_1}\}$ and $\mathcal{R} := \{t_1|C_0^{t_1} \geq C_1^{t_1}\}$. Observe that $C_0^{t_1}P(Y = 0) + C_1^{t_1}P(Y = 1) = 1$. Hence if $C_0^{t_1} < C_1^{t_1}$ then $C_0^{t_1} < 1$ and $C_1^{t_1} > 1$; similarly, if $C_0^{t_1} > C_1^{t_1}$ then $C_0^{t_1} > 1$ and $C_1^{t_1} < 1$,

Observe that

- $(\forall t_1 \in \mathcal{R})(\delta_1^{t_1} = 0)$,
- $(\forall t_1 \in \mathcal{S})(\delta_1^{t_1} = (1 - C_0^{t_1})(\alpha + \beta) + (C_1^{t_1} - 1)(\beta + \delta))$,
- $(\forall t_1 \in \mathcal{R})(\min(\beta_1^{t_1}, \gamma_1^{t_1}) = C_1^{t_1}\beta)$, and
- $(\forall t_1 \in \mathcal{S})(\min(\beta_1^{t_1}, \gamma_1^{t_1}) = C_0^{t_1}\beta)$.

Introduce notation $p_{t_1} := P(T_1 = t_1)$. Then we have

$$\begin{aligned}
&\mathbb{E}_{t_1}\left(\min(\beta_1^{t_1}, \gamma_1^{t_1}) + \delta_1^{t_1}\right) \\
&= \sum_{t_1 \in \mathcal{S}} p_{t_1} C_0^{t_1}\beta + \sum_{t_1 \in \mathcal{R}} p_{t_1} C_1^{t_1}\beta + \sum_{t_1 \in \mathcal{S}} p_{t_1}((1 - C_0^{t_1})(\alpha + \beta) + (C_1^{t_1} - 1)(\beta + \delta)) \\
&\geq \sum_{t_1 \in \mathcal{S}} p_{t_1} C_0^{t_1}\beta + \sum_{t_1 \in \mathcal{R}} p_{t_1} C_1^{t_1}\beta + \sum_{t_1 \in \mathcal{S}} p_{t_1}(1 - C_0^{t_1})\beta + \sum_{t_1 \in \mathcal{S}} p_{t_1}(C_1^{t_1} - 1)\beta \\
&= \sum_{t_1 \in \mathcal{S}} p_{t_1} C_1^{t_1}\beta + \sum_{t_1 \in \mathcal{R}} p_{t_1} C_1^{t_1}\beta \\
&= \beta.
\end{aligned}$$

□

Finally we are in a position to prove Lemma 2.7.4.

Lemma (2.7.4 restated,[9])**.**

$$\mathbb{E}_t\left(\sum_{i=1}^{m} \delta_i^{t_{\leq i}}\right) = \Omega_\mu(1).$$

Proof of Lemma 2.7.4. By Lemma 2.7.7 $\beta_i^{T_i}\gamma_i^{T_i}$ is a supermartingale. Therefore $-2\beta_i^{T_i}\gamma_i^{T_i} = (\beta_i^{T_i} - \gamma_i^{T_i})^2 - (\beta_i^{T_i})^2 - (\gamma_i^{T_i})^2$ is a submartingale. By optional stopping theorem we have

$$\mathbb{E}_t\left((\beta_m^T - \gamma_m^T)^2 - (\beta_m^T)^2 - (\gamma_m^T)^2\right) \geq (\beta-\gamma)^2 - \beta^2 - \gamma^2.$$

Rearranging we get

$$\mathbb{E}_t\left((\beta_m^T - \gamma_m^T)^2 - (\beta-\gamma)^2\right) \geq \mathrm{Var}(\beta_m^T) + \mathrm{Var}(\gamma_m^T).$$

By definition of S, when transcript t is observed exactly one of the following three cases happens:

1. $\beta_m^t = 0.01\min(\beta,\gamma)$

   This transcript contributes at least $(0.99\min(\beta,\gamma))^2$ to $\mathrm{Var}(\beta_m^T)$.

2. $\gamma_m^t = 0.01\min(\beta,\gamma)$

   This contributes at least $(0.99\min(\beta,\gamma))^2$ to $\mathrm{Var}(\gamma_m^T)$.

3. $\alpha_m^t = 0.01\alpha$

   We do not have a guarantee on the contribution to $\mathrm{Var}(\beta_m^T)$ or $\mathrm{Var}(\gamma_m^T)$, but since $\alpha_m^{T_i}$ is a martingale we have $\mathbb{E}_t(\alpha_m^t) = \alpha$. In addition, $\alpha_m^t \leq 1$. Thus $P(\alpha_m^T > 0.01\alpha) \geq 0.99\alpha$.

From the above it follows that

$$\mathrm{Var}(\beta_m^T) + \mathrm{Var}(\gamma_m^T) \geq (0.99\alpha)(0.99\min(\beta,\gamma))^2 =: c_\mu.$$

Consequently

$$\mathbb{E}_t\left((\beta_m^t - \gamma_m^t)^2 - (\beta-\gamma)^2\right) \geq c_\mu. \tag{2.12}$$

Observe that $|\beta_m^t - \gamma_m^t| + |\beta - \gamma| \leq 2$. Thus

$$|\beta_m^t - \gamma_m^t| - |\beta-\gamma| \geq ((\beta_m^t - \gamma_m^t)^2 - (\beta-\gamma)^2)/2.$$

Taking expectation of both sides and using inequality (2.12) we obtain

$$\mathbb{E}_t(|\beta_m^t - \gamma_m^t| - |\beta - \gamma|) \geq c_\mu/2. \tag{2.13}$$

By Lemma 2.7.8 we have

$$\mathbb{E}_t\left(\min(\beta_m^t, \gamma_m^t) - \min(\beta, \gamma) + \sum_{i=1}^{m} \delta_i^{t \leq i}\right) \geq 0.$$

Using $\min(a, b) = (a + b)/2 - |a - b|/2$ we derive

$$\begin{aligned}
&\mathbb{E}_t\left(\frac{\beta_m^t + \gamma_m^t}{2} - \frac{|\beta_m^t - \gamma_m^t|}{2} - \frac{\beta + \gamma}{2} + \frac{|\beta - \gamma|}{2} + \sum_{i=1}^{m} \delta_i^{t \leq i}\right) \\
&= \mathbb{E}_t\left(-\frac{|\beta_m^t - \gamma_m^t|}{2} + \frac{|\beta - \gamma|}{2} + \sum_{i=1}^{m} \delta_i^{t \leq i}\right) \geq 0,
\end{aligned}$$

where the first step follows from the fact that $\beta_i^{T_i}$ and $\gamma_i^{T_i}$ are martingales. Rearranging and applying inequality (2.13) we finally arrive at the conclusion of the statement.

$$\mathbb{E}_t\left(\sum_{i=1}^{m} \delta_i^{t \leq i}\right) \geq (1/2)\mathbb{E}_t(|\beta_m^t - \gamma_m^t| - |\beta - \gamma|) \geq c_\mu/4.$$

□

2.7.4 Upper Bound on the Rate of Convergence

In this subsection we present an r-round discretization (see Protocol 6) of our optimal protocol (see Protocol 4) for AND. We shall prove that the discretized AND protocol achieves $O(1/r^2)$ upper bound on the rate of convergence. This matches the lower bound on the rate of convergence proven in Subsection 2.7.1.

Recall that the informational wastage is how much extra information a particular bit, or message, or a protocol reveals when compared to the optimal protocol.

The most natural way to discretize our continuous AND protocol would be to sample numbers N^A and N^B uniformly at random from the set $\{0, \ldots, r-1\}$ when the

Protocol 6 Discretized r-round protocol π_r for the AND function

Require:

$x \in \{0,1\}$ - known to Alice

$y \in \{0,1\}$ - known to Bob

$\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \gamma & \delta \\ \hline \end{array}, r \in \mathbb{N}$ - known to Alice and Bob

if $\beta < \gamma$ **then**

  Bob sends bit B as follows $B = \begin{cases} 1 & \text{if } y = 1 \\ 0 & \text{with probability } 1 - \beta/\gamma \text{ if } y = 0 \\ 1 & \text{with probability } \beta/\gamma \text{ if } y = 0 \end{cases}$

  if $B = 0$ **then**

    The protocol terminates, the players output 0

if $\beta > \gamma$ **then**

  Alice sends bit B as follows $B = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{with probability } 1 - \gamma/\beta \text{ if } x = 0 \\ 1 & \text{with probability } \gamma/\beta \text{ if } x = 0 \end{cases}$

  if $B = 0$ **then**

    The protocol terminates, the players output 0.

if $x = 0$ **then**

  Alice samples $N^A \in \{0, 1, \ldots, r-1\}$ with $P(N^A = i) = \frac{2r-2i-1}{r^2}$

else

  Alice sets $N^A = r$

if $y = 0$ **then**

  Bob samples $N^B \in \{0, 1, \ldots, r-1\}$ with $P(N^B = i) = \frac{2r-2i-1}{r^2}$

else

  Bob sets $N^B = r$

Both players set C to 0

while $C \leq r - 1$ **do**

  if $C = N^A$ **then**

    Alice sends 1 to Bob, protocol terminates, players output 0

  else

    Alice sends 0 to Bob

  if $C = N^B$ **then**

    Bob sends 1 to Alice, protocol terminates, players output 0

  else

    Bob sends 0 to Alice

  Both players update C to $C + 1$

Protocol terminates, players output 1

corresponding player(s) have 0 as input. While analyzing this option, we discovered that this discretization wastes increasing amounts information in later rounds as the counter C approaches r. This leads to a total information wasted $\approx \frac{1}{r^2}\sum_{i=1}^{r}\frac{1}{i} = \Theta\left(\frac{\log r}{r^2}\right)$. A natural remedy is to select numbers N^A and N^B non-uniformly, assigning less probability mass to the later rounds. Indeed, Protocol 6 assigns probability $\frac{2r-2i-1}{r^2}$ to the ith value of N^A and N^B leading to the correct $O(\frac{1}{r^2})$ bound on the total information wasted. In the rest of this section we prove this claim formally.

We start with two technical lemmas.

Lemma 2.7.9 ([9])**.** *Suppose that Alice sends bit B distributed as follows*

$$B = \begin{cases} 1 & \text{if } X = 1 \\ 0 & \text{with probability } \psi \text{ if } X = 0 \\ 1 & \text{with probability } 1-\psi \text{ if } X = 0 \end{cases}$$

from prior $\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \beta & \delta \\ \hline \end{array}$*. Then the informational wastage of B is*

$$O\left(\frac{\alpha\beta}{\alpha+\beta}\psi^3 + \frac{2\alpha\beta(\beta^2+3\alpha\beta+2\alpha^2)}{(1-\psi)^3\beta^3}\psi^4\right).$$

Proof. The informational wastage of bit B is

$$\mathrm{IW}(\alpha,\beta,\psi) = I(B;X|Y) + P(B=1)\,\mathrm{IC}_{\mu'}(\mathrm{AND},0) - \mathrm{IC}_{\mu}(\mathrm{AND},0),$$

where

$$\mu' = \begin{array}{|c|c|}\hline (1-\psi)\alpha/t & (1-\psi)\beta/t \\ \hline \beta/t & \delta/t \\ \hline \end{array},$$

and $t = (1-\psi)\alpha + (2-\psi)\beta + \delta$. Furthermore, we have

$$\begin{aligned} I(B;X|Y) &= H(B|Y) - H(B|XY) \\ &= (\alpha+\beta)H\left(\frac{\alpha}{\alpha+\beta}\psi\right) + (\beta+\delta)H\left(\frac{\beta}{\beta+\delta}\psi\right) - (\alpha+\beta)H(\psi). \end{aligned}$$

Writing Taylor series for $\mathrm{IW}(\alpha,\beta,\psi)$ for ψ around 0 we obtain

$$\exists \zeta \in [0,\psi] \text{ s.th. } \mathrm{IW}(\alpha,\beta,\psi) = \frac{\alpha\beta}{(\alpha+\beta)\ln 64}\psi^3 + R(\zeta)\frac{\psi^4}{24},$$

where $R(\zeta) = \frac{2\alpha\beta(\beta^2+3\alpha\beta(1-\zeta)+\alpha^2(2-3\zeta+\zeta^3))}{(1-\zeta)^3(\alpha+\beta-\alpha\zeta)^3\ln 2}$.

The above expressions were obtained with help from Wolfram Mathematica. The statement of the lemma follows immediately. □

Suppose that players start with a symmetric prior $\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \beta & \delta \\ \hline \end{array}$. Observe that the counter C in Protocol 6 can be viewed as a discrete implementation of a continuous clock from Protocol 4. The hand of our clock now moves in discrete steps from position z to position $z+\phi$ where $z = 1-\left(\frac{r-i}{r}\right)^2$ and $\phi = \frac{2r-2i-1}{r^2}$ for $i \in \{0,\ldots,r-1\}$. We now analyze how a single such move is accomplished by Alice and Bob in our protocol and how much information is wasted during this move.

At time z the prior μ becomes $\mu_z = \begin{array}{|c|c|}\hline (1-z)^2\alpha & (1-z)\beta \\ \hline (1-z)\beta & \delta \\ \hline \end{array}$ normalized. Thus, when the players move from time z to time $z+\phi$ it is equivalent to first Alice sending bit B as in Lemma 2.7.9 with $\psi = \frac{\phi}{1-z}$ followed by a similar bit B' sent by Bob. Note that after Alice sends bit B, the prior moves into Bob's region. In the optimal protocol, Bob would send *exactly* bit B'. Hence Bob's bit wastes no information. Therefore the total informational wastage incurred while moving clock hand from time z to time $z+\phi$ in 2 rounds of communication comes from bit B.

Lemma 2.7.10 ([9]). *Let* $\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \beta & \delta \\ \hline \end{array}$ *be a distribution with full support. When Alice and Bob in 2 rounds of communication advance the clock from position* z *to* $z+\phi$ *with* $\frac{\phi}{1-z} \leq \frac{2}{3}$ *they waste a total of* $O_\mu\left(\frac{\phi^3}{1-z}\right)$ *information.*

Proof. As discussed in the paragraph before the statement of the lemma, we simply have to apply Lemma 2.7.9 to Alice's signal B with $\psi = \frac{\phi}{1-z}$ and distribution

$$\mu_z = \begin{array}{|c|c|}\hline (1-z)^2\alpha/n & (1-z)\beta/n \\ \hline (1-z)\beta/n & \delta/n \\ \hline \end{array},$$

where $n = (1-z)^2\alpha + 2(1-z)\beta + \delta$. Note that by assumptions of the lemma we have $\psi \leq 2/3$, therefore we have $1/(1-\psi)^3 \leq 27$. Furthermore we have $n \geq \delta$ and $\phi \leq 1 - z$. Plugging all this in Lemma 2.7.9 and simplifying we obtain that the total information wasted is big-Oh of the following expression:

$$\begin{aligned}
&\frac{(1-z)^3\alpha\beta}{n((1-z)^2\alpha + (1-z)\beta)}\frac{\phi^3}{(1-z)^3} + \\
&+ 27\frac{2(1-z)^3\alpha\beta n^3}{(1-z)^3\beta^3 n^2}\frac{(1-z)^2\beta^2 + 3(1-z)^3\alpha\beta + 2(1-z)^4\alpha}{n^2}\frac{\phi^4}{(1-z)^4} \\
&= O\left(\frac{\alpha}{\delta}\frac{\phi^3}{1-z} + \frac{\alpha}{\delta\beta^2}\frac{\phi^4}{(1-z)^2}\right) \\
&= O_\mu\left(\frac{\phi^3}{1-z}\right).
\end{aligned}$$

□

Now we are in a position to prove the main result of this subsection.

Theorem 2.7.11 ([9]). *For distributions* $\mu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \gamma & \delta \\ \hline \end{array}$ *with full support we have*

$$\mathrm{IC}^r_\mu(\mathrm{AND}, 0) - \mathrm{IC}_\mu(\mathrm{AND}, 0) = O_\mu\left(\frac{1}{r^2}\right).$$

Proof. Let π_r denote Protocol 6 and π denote Protocol 4. In the first stage of protocol π_r the player who is more likely to have 0 sends a bit that either terminates the protocol or moves the prior to the diagonal. This stage is exactly the same in protocol π. Thus the difference in the information cost of the two protocols arises only from the second (which we previously called symmetric) stage of π and π_r. Thus for the rest of the proof we shall assume that μ is symmetric, i.e., $\beta = \gamma$.

Observe that for the ith jump of the clock we have $\phi_i = \frac{2r-2i-1}{r^2}$ and $z_i = 1 - \left(\frac{r-i}{r}\right)^2$. Therefore $\frac{\phi_i}{1-z_i} = \frac{2r-2i-1}{(r-i)^2} \leq \frac{2}{r-i}$. Hence $\frac{\phi_i}{1-z_i} \leq 2/3$ for all i except $i \in \{r-2, r-1\}$. The later event happens with probability $O(1/r^2)$ conditioned on Alice having 0 as input. In addition if $X = 1$, Alice learns the entire Bob's bit. Hence the difference in the information cost of π_r and π arises from the events when Alice or Bob have 0 as input. Consequently we may ignore the last two movements of the clock as they

contribute at most $O(1/r^2)$ to the informational wastage. For the rest of the clock movements we may apply Lemma 2.7.10 which says that the information wasted in the ith movement is $O_\mu\left(\frac{\phi_i^3}{1-z_i}\right)$. Aggregating it over the first $r-2$ movements of the clock we get the total amount of information wasted is

$$O_\mu\left(\sum_{i=0}^{r-3}\frac{(2r-2i-1)^3r^2}{r^6(r-i)^2}\right)=O_\mu\left(\sum_{i=0}^{r-3}\frac{(r-i)}{r^4}\right)=O_\mu\left(\frac{1}{r^2}\right).$$

□

2.8 Communication Complexity of $\bigvee$-type Functions

The main result of this section is a characterization of the randomized communication complexity of $\bigvee$-type functions in terms of a certain version of information complexity.

Definition 2.8.1. A function g is called $\bigvee$-type if it can be written as $g=\vee_{i=1}^n f(x_i,y_i)$ for some function f.

Definition 2.8.2 ([9])**.** We shall call a protocol π *good for* f if π solves f correctly on *all inputs*. Let $\mathcal{U}_0$ and $\mathcal{U}_1$ be the sets of distributions supported on $f^{-1}(0)$ and $f^{-1}(1)$ respectively. Define

$$\mathrm{IC}^{\mathrm{zero}}(f,0)=\inf_{\pi\text{ good for }f}\max_{\mu\in\mathcal{U}_0}\mathrm{IC}_\mu(\pi)$$
$$\mathrm{IC}^{\mathrm{one}}(f,0)=\inf_{\pi\text{ good for }f}\max_{\mu\in\mathcal{U}_1}\mathrm{IC}_\mu(\pi)$$

Remark 2.8.1.

- $\mathrm{IC}^{\mathrm{zero}}(f,0)$ and $\mathrm{IC}^{\mathrm{one}}(f,0)$ measure the zero-error information complexity of f with respect to restricted families of distributions.
- The maximum in the above definitions is allowed since the corresponding sets of distributions are compact and information cost is continuous in the input distribution.

- In spite of measuring information cost of the protocol with respect to restricted sets of distributions, we still require that the protocol is correct on *all inputs*.
- $\mathrm{IC}^{\mathrm{one}}(f, 0) = \mathrm{IC}^{\mathrm{zero}}(\neg f, 0)$.

We start by establishing the following theorem via a minimax argument similar to the one in [7].

Theorem 2.8.1 ([9])**.** *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function. Then*

$$\mathrm{IC}^{zero}(f, 0) = \inf_{\pi \text{ good for } f} \max_{\mu \in \mathcal{U}_0} \mathrm{IC}_\mu(\pi) = \max_{\mu \in \mathcal{U}_0} \inf_{\pi \text{ good for } f} \mathrm{IC}_\mu(\pi)$$

Similarly, we have

$$\mathrm{IC}^{one}(f, 0) = \inf_{\pi \text{ good for } f} \max_{\mu \in \mathcal{U}_1} \mathrm{IC}_\mu(\pi) = \max_{\mu \in \mathcal{U}_1} \inf_{\pi \text{ good for } f} \mathrm{IC}_\mu(\pi)$$

Proof. Clearly, we have $\mathrm{IC}^{\mathrm{zero}}(f, 0) \geq \max_{\mu \in \mathcal{U}_0} \inf_{\pi \text{ good for } f} \mathrm{IC}_\mu(\pi)$. We shall establish the reverse inequality. Let G be the set of protocols that are good for f.

Lemma 2.8.2. *Let H be any finite subset of G. Then for any α such that*

$$\alpha \geq \max_{\mu \in \mathcal{U}_0} \min_{\pi \in H} \mathrm{IC}_\mu(\pi)$$

there exists a protocol $\tau \in G$ such that $\forall \mu \in \mathcal{U}_0$ we have $\mathrm{IC}_\mu(\tau) \leq \alpha$.

We define the following zero-sum two-player game $\mathcal{G}_0$. Player A will come up with a randomized two-party protocol $\pi \in H$. Player B will come up with a distribution $\mu \in \mathcal{U}_0$. Player B's payoff is given by:

$$P_B(\pi, \mu) = \mathrm{IC}_\mu(\pi).$$

Then we have:

Claim 2.8.3. *The value $V_B(\mathcal{G}_0) \leq \alpha$.*

Proof. Let ν_B be a probability distribution representing a mixed strategy for player B. Thus, ν_B is a distribution on probability distributions μ over $\mathcal{X} \times \mathcal{Y}$. We need

to show that there exists a zero-error protocol $\tau \in H$ such that $\mathbb{E}_{\mu\sim\nu_B}(\mathrm{IC}_\mu(\tau)) \leq \alpha$. Let $\bar{\mu}$ be a distribution on $\mathcal{X} \times \mathcal{Y}$ that is obtained by taking the average of $\mu \sim \nu_B$. Formally,

$$\bar{\mu}(x, y) := \mathbb{E}_{\mu\sim\nu_B}(\mu(x, y)).$$

Since all distributions μ are in $\mathcal{U}_0$, we have $\bar{\mu} \in \mathcal{U}_0$. Since $\alpha \geq \max_{\mu\in\mathcal{U}_0} \min_{\pi\in H} \mathrm{IC}_\mu(\pi)$, there exists a protocol $\tau \in H$ such that $\mathrm{IC}_{\bar{\mu}}(\tau) \leq \alpha$. By the concavity of information cost, it holds that

$$\mathbb{E}_{\mu\sim\nu_B} (\mathrm{IC}_\mu(\tau)) \leq \mathrm{IC}_{\bar{\mu}}(\tau).$$

Thus, the value of the game is bounded by α. □

The minimax theorem holds for our game by an ϵ-net argument and continuity of $\mathrm{IC}_\mu(\pi)$. By the minimax theorem, there is a mixed strategy for player A such that for each response by player B the value of the game for player B is at most α. A mixed strategy for player A is a distribution ν_A on protocols. In other words,

$$\forall \mu \in \mathcal{U}_0 \quad \mathbb{E}_{\pi\sim\nu_A}(P_B(\pi, \mu)) \leq \alpha. \tag{2.14}$$

Let $\bar{\pi}$ be the randomized protocol obtained by publicly sampling $\pi \sim \nu_A$ and then applying π to the inputs. We claim that $\bar{\pi}$ is the protocol we are looking for. In other words, the randomized protocol $\bar{\pi}$ has the desired payoff properties. Clearly $\bar{\pi} \in G$.

Claim 2.8.4. *For all $\mu \in \mathcal{U}_0$ we have $\mathrm{IC}_\mu(\bar{\pi}) \leq \alpha$.*

Proof. We shall prove that

$$I_{(X,Y)\sim\mu}(\bar{\Pi}(X, Y); X|Y) \leq \mathbb{E}_{\Pi\sim\nu_A} \left(I_{(X,Y)\sim\mu}(\Pi(X, Y); X|Y) \right). \tag{2.15}$$

In other words, the amount of information revealed by $\bar{\pi}$ is bounded by the average amount of information revealed by Π that is drawn according to ν_A.

To establish (2.15), consider the following four random variables. Let S be a "selector" random variable, that picks the protocol Π to run according to the distribution ν_A. Let (X, Y) be inputs distributed according to μ independently of S. Finally, let

$\Pi = \pi(X, Y)$ be the transcript of the selected protocol executed on X and Y. We have:

$$\mathbb{E}_{\Pi \sim \nu_A} \left(I_{(X,Y)\sim\mu}(\Pi(X, Y); X|Y) \right) = I(\Pi; X|YS),$$

and

$$I_{(X,Y)\sim\mu}(\bar{\Pi}(X, Y); X|Y) = I(\Pi; X|Y).$$

Since the protocol Π is selected independently of the inputs, we have $I(X; S|\Pi Y) = 0$. By substituting $A = \Pi$, $B = X$, $C = Y$, and $D = S$ into Proposition 1.4.3 we get

$$I(\Pi; X|Y) \leq I(\Pi; X|YS), \tag{2.16}$$

establishing (2.15). Similarly to (2.15), the following inequality can be established:

$$I_{(X,Y)\sim\mu}(\bar{\Pi}(X, Y); Y|X) \leq \mathbb{E}_{\Pi \sim \nu_A} \left(I_{(X,Y)\sim\mu}(\Pi(X, Y); Y|X) \right). \tag{2.17}$$

Together, (2.15) and (2.17) imply

$$\mathrm{IC}_\mu(\bar{\pi}) \leq \mathbb{E}_{\Pi \sim \nu_A} \left(\mathrm{IC}_\mu(\pi) \right). \tag{2.18}$$

□

We use a compactness argument to complete the proof. Fix any α such that

$$\alpha > \max_{\mu \in \mathcal{U}_0} \inf_{\pi \in G} \mathrm{IC}_\mu(\pi).$$

Define

$$A(\pi) := \{\mu \in \mathcal{U}_0 : \mathrm{IC}_\mu(\pi) \geq \alpha\}$$

Then $\cap_{\pi \in G} A(\pi) = \emptyset$. Since $\mathcal{U}_0$ is compact and the sets $A(\pi)$ are closed because of the continuity of $\mathrm{IC}(\pi)$, we get that there is a finite set of protocols $H \subset G$ such that $\cap_{\pi \in H} A(\pi) = \emptyset$. Thus, $\forall \mu \in \mathcal{U}_0$ we have that $\min_{\pi \in H} \mathrm{IC}_\mu(\pi) < \alpha$. Then by Lemma 2.8.2, there exists a protocol $\tau \in G$ such that $\forall \mu \in \mathcal{U}_0$ we have $\mathrm{IC}_\mu(\tau) \leq \alpha$.

Thus

$$\inf_{\pi \in G} \max_{\mu \in \mathcal{U}_0} \mathrm{IC}_\mu(\pi) \leq \max_{\mu \in \mathcal{U}_0} \inf_{\pi \in G} \mathrm{IC}(\pi)$$

which completes the proof. □

Next, we show that in the definition of $\mathrm{IC}^{\mathrm{zero}}(f, 0)$ we could replace distributions with 0 mass on $f^{-1}(1)$ and protocols that are correct on all inputs with distributions that place ϵ mass on $f^{-1}(1)$ and protocols that are correct on the support of μ.

Theorem 2.8.5 ([9]). *For all $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ we have*

$$(1) \lim_{\epsilon \to 0} \max_{\mu : \mu(f^{-1}(1)) \leq \epsilon} \mathrm{IC}_\mu(f, 0) = \lim_{\epsilon \to 0} \max_{\mu : \mu(f^{-1}(1)) \leq \epsilon} \inf_{\pi \textit{ solves } f \textit{ on all inputs}} \mathrm{IC}_\mu(\pi)$$

$$(2) \mathrm{IC}^{zero}(f, 0) = \lim_{\epsilon \to 0} \max_{\mu : \mu(f^{-1}(1)) \leq \epsilon} \inf_{\pi \textit{ solves } f \textit{ on all inputs}} \mathrm{IC}_\mu(\pi)$$

Proof. (1) We have the following

$$\begin{aligned} \max_{\mu : \mu(f^{-1}(1)) \leq \epsilon} \mathrm{IC}_\mu(f, 0) &= \max_{\mu : \mu(f^{-1}(1)) \leq \epsilon} \inf_{\pi \text{ solves } f \text{ on } \mathrm{supp}(\mu)} \mathrm{IC}_\mu(\pi) \\ &\leq \max_{\mu : \mu(f^{-1}(1)) \leq \epsilon} \inf_{\pi \text{ solves } f \text{ on all inputs}} \mathrm{IC}_\mu(\pi) \end{aligned}$$

The above inequality holds since the second infimum is over a smaller set of protocols. Taking the limits, we obtain:

$$\lim_{\epsilon \to 0} \max_{\mu : \mu(f^{-1}(1)) \leq \epsilon} \mathrm{IC}_\mu(f, 0) \leq \lim_{\epsilon \to 0} \max_{\mu : \mu(f^{-1}(1)) \leq \epsilon} \inf_{\pi \text{ solves } f \text{ on all inputs}} \mathrm{IC}_\mu(\pi)$$

Next, we show the reverse inequality. Fix $\epsilon > 0$ and $\delta > 0$, and let ν be the distribution that maximizes $\max_{\mu : \mu(f^{-1}(1)) \leq \epsilon} \inf_{\pi \text{ solves } f \text{ on all inputs}} \mathrm{IC}_\mu(\pi)$. Let $\mathcal{U}$ denote the uniform distribution on $\{0,1\}^k \times \{0,1\}^k$ and define $\nu_\epsilon = \epsilon \mathcal{U} + (1 - \epsilon)\nu$. A protocol solves f on all inputs if and only if that protocol solves f on the support of ν_ϵ. Therefore, we have

$$\mathrm{IC}_{\nu_\epsilon}(f, 0) = \inf_{\pi \text{ solves } f \text{ on all inputs}} \mathrm{IC}_{\nu_\epsilon}(\pi).$$

Putting it all together, we have

$$\begin{aligned}
\max_{\mu:\mu(f^{-1}(1))\leq 2\epsilon} \mathrm{IC}_\mu(f,0) &\geq \mathrm{IC}_{\nu_\epsilon}(f,0) \quad (\text{since } \nu_\epsilon(f^{-1}(1)) \leq 2\epsilon) \\
&= \inf_{\pi \text{ solves } f \text{ on all inputs}} \mathrm{IC}_{\nu_\epsilon}(\pi) \quad (\text{by above}) \\
&= \inf_{\pi \text{ solves } f \text{ on all inputs}} \mathrm{IC}_{\nu}(\pi) + O(\sqrt{\epsilon}) \quad (\text{by Lemma 2.4.6}) \\
&= \max_{\mu:\mu(f^{-1}(1))\leq \epsilon} \inf_{\pi \text{ solves } f \text{ on all inputs}} \mathrm{IC}_{\mu}(\pi) + O(\sqrt{\epsilon}) \\
&\quad (\text{by choice of } \nu)
\end{aligned}$$

Taking the limit as $\epsilon \to 0$ on both sides finishes the proof of the other inequality and the first part of the theorem.

(2) Fix $\epsilon > 0$, then

$$\max_{\mu:\mu\in\mathcal{U}_0} \inf_{\pi \text{ solves } f \text{ on all inputs}} \mathrm{IC}_\mu(\pi) \leq \max_{\mu:\mu(f^{-1}(1))\leq \epsilon} \inf_{\pi \text{ solves } f \text{ on all inputs}} \mathrm{IC}_\mu(\pi),$$

since the maximum on the right hand side is taken over a larger set. Taking the limit as $\epsilon \to 0$ shows that

$$\mathrm{IC}^{\mathrm{zero}}(f,0) \leq \lim_{\epsilon\to 0} \max_{\mu:\mu(f^{-1}(1))\leq \epsilon} \inf_{\pi \text{ solves } f \text{ on all inputs}} \mathrm{IC}_\mu(\pi).$$

Next, we prove the reverse inequality. For a distribution μ we define

$$\mu^{\mathrm{zero}}(x,y) = \frac{\chi((x,y)\in f^{-1}(0))\mu(x,y)}{\mu(f^{-1}(0))},$$

where χ is the indicator function of the event specified in brackets. If $\mu(f^{-1}(1)) \leq \epsilon$ then we have

$$|\mu^{\mathrm{zero}}(x,y) - \mu(x,y)| \leq \begin{cases} \frac{\epsilon}{1-\epsilon}\mu(x,y) & (x,y)\in f^{-1}(0) \\ \mu(x,y) & (x,y)\in f^{-1}(1) \end{cases}$$

Therefore, $|\mu^{\text{zero}} - \mu| \leq \epsilon + \frac{\epsilon}{1-\epsilon} \leq 3\epsilon$ assuming $\epsilon \leq 1/2$. Then

$$\begin{aligned}\max_{\mu:\mu(f^{-1}(1))\leq\epsilon} \inf_{\pi \text{ solves } f \text{ on all inputs}} \mathrm{IC}_\mu(\pi) &\leq \max_{\mu:\mu(f^{-1}(1))\leq\epsilon} \inf_{\pi \text{ solves } f \text{ on all inputs}} \mathrm{IC}_{\mu^{\text{zero}}}(\pi) + O(\sqrt{\epsilon}) \\ &\leq \max_{\mu\in\mathcal{U}_0} \inf_{\pi \text{ solves } f \text{ on all inputs}} \mathrm{IC}_\mu(\pi) + O(\sqrt{\epsilon}) \\ &= \mathrm{IC}^{\text{zero}}(f,0) + O(\sqrt{\epsilon})\end{aligned}$$

In the above, the first inequality is due to Lemma 2.4.6 and the second inequality is due to the fact that for all μ we have $\mu^{\text{zero}} \in \mathcal{U}_0$. Taking the limit as $\epsilon \to 0$ on both sides finishes the proof of the reverse inequality and the second part of the theorem. □

In the rest of this section we prove Theorem 2.2.4 and apply it to the DISJ_n function. For convenience, we restate Theorem 2.2.4 below.

Theorem (Theorem 2.2.4 restated, [9])**.** *Let $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ and $g_n : \{0,1\}^{nk} \times \{0,1\}^{nk} \to \{0,1\}$ be functions, such that $g_n(x,y) = \vee_{i=1}^n f(x_i, y_i)$, where $x = \{x_i\}_{i=1}^n, y = \{y_i\}_{i=1}^n$ and $x_i, y_i \in \{0,1\}^k$. Then for all $\epsilon > 0$, there exists $\delta = \delta(f,\epsilon) > 0$ such that $\delta \to 0$ as $\epsilon \to 0$ and*

$$(\mathrm{IC}^{zero}(f,0) - \delta)n \leq \mathrm{R}(g_n,\epsilon) \leq \mathrm{IC}^{zero}(f,0)n + o(n)k.$$

Remark 2.8.2. Theorem 2.2.4 is stated for $\bigvee$-type functions, but an analogous result immediately follows (via De Morgan's laws) for $\bigwedge$-type functions with $\mathrm{IC}^{\text{zero}}(f,0)$ replaced by $\mathrm{IC}^{\text{one}}(f,0)$.

2.8.1 Lower Bound on Communication Complexity of $\bigvee$-type Functions

In this subsection we prove the '$\geq$' direction of Theorem 2.2.4.

Lemma 2.8.6 ([9])**.** *For all $\epsilon > 0$, there exists $\delta = \delta(f,\epsilon) > 0$ such that $\delta \to 0$ as $\epsilon \to 0$ and $\mathrm{R}(g_n,\epsilon) \geq (\mathrm{IC}^{zero}(f,0) - \delta)n$.*

The idea of the proof. We reduce f to g_n with a factor of n less information complexity with respect to a restricted family of distributions. This kind of a reduction was first introduced in [3]. It has since been used in a number of works (for example, see [4, 12, 7]) in the context of direct sums for information complexity. Since we start with a protocol that solves g_n with some non-zero error, our reduction produces a protocol for f with non-zero error also. We then use the continuity of information complexity at error tolerance 0 (see Section 2.4) to reach the desired conclusion.

Proof. Let π be a protocol for computing g_n that is provided by the definition of $\mathrm{R}(g_n, \epsilon)$. That is π computes g_n with error probability at most ϵ on all inputs and has communication cost $\mathrm{R}(g_n, \epsilon)$. Fix $\nu \in \mathcal{U}_0$. Let τ denote the protocol described in Protocol 7.

For all $x, y \in \{0,1\}^k$ we have

$$\begin{aligned}
&P(\mathrm{T}(x,y) \neq f(x,y)) = \\
&= P_{(X_i,Y_i)\sim\nu, J\in[n]}\left(\Pi(X_{<J}xX_{>j}, Y_{<J}yY_{>J}) \neq f(x,y)\right) \\
&= P_{(X_i,Y_i)\sim\nu, J\in[n]}\left(\Pi(X_{<J}xX_{>j}, Y_{<J}yY_{>J}) \neq \bigvee_{i\neq J} f(X_i,Y_i) \vee f(x,y)\right) \\
&\quad (\text{since } (X_i,Y_i)\sim\nu \text{ and } \mathrm{supp}(\nu) \in f^{-1}(0)) \\
&= P_{(X_i,Y_i)\sim\nu, J\in[n]}\left(\Pi(X_{<J}xX_{>j}, Y_{<J}yY_{>J}) \neq g(X_{<J}xX_{>j}, Y_{<J}yY_{>J})\right) \\
&\leq \epsilon \quad (\text{by the choice of } \pi)
\end{aligned}$$

Therefore τ solves f with probability of error at most ϵ on *all inputs*. Moreover, measuring information cost of τ with respect to ν we obtain:

Lemma 2.8.7 (Theorem 3.17 in [12])**.**

$$\mathrm{IC}_\nu(\tau) \leq \frac{\mathrm{IC}_{\nu^n}(\pi)}{n} \leq \frac{\mathrm{R}(g_n,\epsilon)}{n}$$

By the continuity of information complexity at error 0 (see Theorem 2.2.8) we have $\mathrm{IC}_\nu(\tau) \geq \mathrm{IC}_\nu(f,0) - \delta(f,\epsilon)$, where $\delta(f,\epsilon) \to 0$ as $\epsilon \to 0$. Putting this all together we

have:

$$\mathrm{R}(g_n, \epsilon) \geq n \, \mathrm{IC}_\nu(\tau) \geq n(\mathrm{IC}_\nu(f, 0) - \delta(f, \epsilon)).$$

Since $\nu \in \mathcal{U}_0$ is arbitrary, we can take the maximum over $\nu \in \mathcal{U}_0$ and apply Theorem 2.8.1 to reach the desired conclusion:

$$\mathrm{R}(g_n, \epsilon) \geq n(\mathrm{IC}^{\mathrm{zero}}(f, 0) - \delta(f, \epsilon)).$$

□

Protocol 7 Protocol τ for f that is constructed out of the protocol π for g_n.

Require:

$x \in \{0,1\}^k$ - known to Alice
$y \in \{0,1\}^k$ - known to Bob
$\nu \in \mathcal{U}_0, \pi$ - known to Alice and Bob

1: Both players use public randomness to sample $J \in [n]$ uniformly at random.
2: Both players use public randomness to sample $X_1, \ldots, X_{J-1}, Y_{J+1}, \ldots, Y_n$ according to the marginals of ν. Each coordinate is sampled independently.
3: For $k \in \{J+1, \ldots, n\}$ Alice privately samples X_k from ν conditioned on Y_k.
4: For $k \in \{1, \ldots, J-1\}$ Bob privately samples Y_k from ν conditioned on X_k.
5: Both players execute

$$\pi((X_1, \ldots, X_{J-1}, x, X_{J+1}, \ldots, X_n), (Y_1, \ldots, Y_{J-1}, y, Y_{J+1}, \ldots, Y_n))$$

6: Output of π is declared to be the output of the current protocol.

2.8.2 Upper Bound on Communication Complexity of $\bigvee$-type Functions

In this section we prove an upper bound on the communication complexity of $\bigvee$-type functions. We start by upper bounding the information complexity of g_n.

Lemma 2.8.8 ([9])**.** *Let $f : \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ and $g_n : \{0,1\}^{nk} \times \{0,1\}^{nk} \to \{0,1\}$ be functions such that $g_n(x_1, x_2, \ldots x_n, y_1, y_2, \ldots y_n) = \vee_{i=1}^n f(x_i, y_i)$, where $x_i, y_i \in$*

$\{0,1\}^k$. *Then we have*

$$\mathrm{IC}(g_n, 0) \leq n\,\mathrm{IC}^{zero}(f, 0) + O(n^{2/3}\log(n)k).$$

The idea of the proof. For $i \in [n]$ let $\nu_i(x_i, y_i) = \sum_{x_{<i}, y_{<i}, x_{>i}, y_{>i}} \mu(x, y)$ denote the marginal distribution of μ on the ith input. If many of the ν_i place sufficient mass on $f^{-1}(1)$ then Alice and Bob can compute the value of g_n with little communication (and hence information) by sampling and exchanging inputs on a few of the i. We can make the communication cost (and hence the information cost) of this step as low as $O(n^{2/3}\log(n)k)$. Thus, for hard distributions μ most of the ν_i have little mass on $f^{-1}(1)$. To deal with such distributions we run n copies of a low information protocol provided by the definition of $\mathrm{IC}^{\mathrm{zero}}(f, 0)$. The information cost of this step is $n\,\mathrm{IC}^{\mathrm{zero}}(f, 0)$. Combining these two steps we can handle arbitrary distributions μ.

Proof. Fix $\delta > 0$. Let π be a protocol that solves f with 0 error on all inputs and such that $\max_{\mu \in \mathcal{U}_0} \mathrm{IC}_\mu(\pi) \leq \mathrm{IC}^{\mathrm{zero}}(f, 0) + \delta$. Consider the protocol π_n described in Protocol 8

Protocol 8 Protocol π_n for computing g_n that is based on protocol π for f.

Require:
$x \in \{0,1\}^{nk}$ - known to Alice
$y \in \{0,1\}^{nk}$ - known to Bob
π - known to Alice and Bob

Both players use public randomness to sample $n^{2/3}$ random coordinates $i \in [n]$ independently and uniformly at random.
Let J denote the multiset of coordinates sampled in the above step.
For all $i \in J$ Alice sends x_i to Bob.
For all $i \in J$ Bob sends y_i to Alice.
if $\exists i \in J \;\; f(x_i, y_i) = 1$ **then**
    Both players output 1, protocol terminates.
else
    for $i \in [n]$ **do**
        Both players run $\pi(x_i, y_i)$
        if $\pi(x_i, y_i)$ outputs 1 **then**
            Both players output 1, protocol terminates.
    Both players output 0, protocol terminates.

It is clear that π_n solves g_n correctly on all inputs, since π solves f correctly on all inputs. To finish the proof we shall show that for an arbitrary μ – distribution on $\{0,1\}^{nk} \times \{0,1\}^{nk}$ – we have $\mathrm{IC}_\mu(\pi_n) \leq n\,\mathrm{IC}^{zero}(f,0) + O(n^{2/3}\log(n)k)$.

Let μ be a distribution on $\{0,1\}^{nk} \times \{0,1\}^{nk}$. Let $(X,Y) \sim \mu$ denote the random inputs to π_n. Then the public randomness concatenated with the transcript of π_n is $\Pi_n = J\Pi_n^1\Pi_n^2$, where J denotes the publicly sampled random coordinates, Π_n^1 denotes the inputs corresponding to J that are exchanged by the players in π_n, and Π_n^2 denotes the rest of the transcript of π_n. Let E denote the indicator random variable for the event "$\exists i \in J \;\; f(X_i, Y_i) = 1$." We have

$$\begin{aligned}
I(\Pi_n; X|Y) &= I(J\Pi_n^1\Pi_n^2; X|Y) \\
&= I(J; X|Y) + I(\Pi_n^1\Pi_n^2; X|YJ) \\
&= I(\Pi_n^1\Pi_n^2; X|YJ) \quad (\text{since } J \text{ is indep. of } X|Y) \\
&= I(\Pi_n^1; X|YJ) + I(\Pi_n^2; X|YJ\Pi_n^1) \\
&\leq 2kn^{2/3} + I(\Pi_n^2; X|YJ\Pi_n^1) \quad (|\Pi_n^1| \leq 2kn^{2/3}) \\
&\leq 2kn^{2/3} + I(\Pi_n^2; X|YJ\Pi_n^1E) \quad (E \text{ is a function of } J, \Pi_n^1) \\
&= 2kn^{2/3} + H(\Pi_n^2|YJ\Pi_n^1E) - H(\Pi_n^2|XYJ\Pi_n^1E) \\
&\leq 2kn^{2/3} + H(\Pi_n^2|YE) - H(\Pi_n^2|XYJ\Pi_n^1E) \quad (\text{conditioning}) \\
&= 2kn^{2/3} + H(\Pi_n^2|YE) - H(\Pi_n^2|XYE) \quad (\Pi_n^2 \text{ is indep. of } J, \Pi_n^1|E) \\
&= 2kn^{2/3} + I(\Pi_n^2; X|YE) \\
&= 2kn^{2/3} + P(E=0)I(\Pi_n^2; X|Y, E=0) \quad |\Pi_n^2| = 0 \text{ if } E = 1
\end{aligned}$$

If $P(E=0) \leq 1/n^{1/3}$, then $P(E=0)I(\Pi_n^2; X|Y, E=0) \leq n^{2/3}k$ and the claim follows. Therefore, we assume that $P(E=0) \geq 1/n^{1/3}$.

For $x, y \in \{0,1\}^{nk}$ we define $N(x,y) = |\{(x_i, y_i) \mid f(x_i, y_i) = 1\}|$. Abusing the notation, we write

$$\mu(d) = P_{(X,Y)\sim\mu}(N(X,Y) = d).$$

Let $\mu' = \mu|(E=0)$. For (x,y) such that $N(x,y) = d$ the probability of sampling $n^{2/3}$ coordinates and not hitting any (x_j, y_j) such that $f(x_j, y_j) = 1$ is bounded by

$\exp(-2d^2/n^{4/3})$ by the Chernoff bound. Thus

$$\begin{aligned}\mu'(d) &\leq \frac{\mu(d)\exp(-2d^2/n^{4/3})}{P(E=0)} \\ &\leq \mu(d)\exp(-2d^2/n^{4/3})n^{1/3}\end{aligned}$$

Thus for $d \geq n^{2/3}\log(n)$, $\mu'(d)$ is small. Hence

$$\mathbb{E}_{(X,Y)\sim\mu'}(N(X,Y)) \leq O(n^{2/3}\log(n))$$

Let μ'_i denote the marginal distribution of μ' on the ith block. Define

$$\epsilon_i = P_{(X_i,Y_i)\sim\mu'_i}(f(X_i,Y_i)=1).$$

The following lemma states that the information cost of a protocol π, that runs a protocol τ independently on many copies, is less than the sum of the information costs of the protocol τ on different copies w.r.t the marginal distributions.

Lemma 2.8.9 (Theorem 4.2) in [7]). *Let μ be a distribution on $\{0,1\}^{nk}\times\{0,1\}^{nk}$. Divide the input into n blocks of size k each and let μ_i denote the marginal distribution on the ith block. Let τ be a protocol on the input space $\{0,1\}^k\times\{0,1\}^k$. Then* $\mathrm{IC}_\mu(\tau^n) \leq \sum_{i=1}^n \mathrm{IC}_{\mu_i}(\tau)$.

Let $\nu = (\sum_{i=1}^n \mu'_i)/n$. We have

$$\begin{aligned}I(\Pi_n^2; X|Y, E=0) + I(\Pi_n^2; Y|X, E=0) &\leq \sum_{i=1}^n \mathrm{IC}_{\mu'_i}(\pi) \quad \text{(by Lemma 2.8.9)} \\ &\leq n\,\mathrm{IC}_\nu(\pi) \quad \text{(by the concavity of IC)}\end{aligned}$$

By the linearity of expectation we have

$$\sum_{i=1}^n \epsilon_i = \mathbb{E}_{(X,Y)\sim\mu'}(N(X,Y)) \leq O(n^{2/3}\log(n))$$

Thus $\nu = (\sum_{i=1}^n \mu'_i)/n$ has $O(\log(n)/n^{1/3})$ mass on $f^{-1}(1)$ and hence is $O(\log(n)/n^{1/3})$

close to distribution ν' in $\mathcal{U}_0$. Using Lemma 2.4.6 along with the fact that $\mathrm{IC}_{\nu'}(\pi) \leq \mathrm{IC}^{\mathrm{zero}}(f,0) + \delta$ gives us that

$$\mathrm{IC}_{\nu}(\pi) \leq \mathrm{IC}^{\mathrm{zero}}(f,0) + \delta + O(\log(n)/n^{1/3}k) + H(O(\log(n)/n^{1/3})).$$

Hence

$$\begin{aligned} &I(\Pi_n^2; X|Y, E=0) + I(\Pi_n^2; Y|X, E=0) \\ &\leq n(\mathrm{IC}^{\mathrm{zero}}(f,0) + \delta) + O(n^{2/3}\log(n)k) \end{aligned}$$

Thus we get that $\mathrm{IC}(g_n, 0) \leq n(\mathrm{IC}^{\mathrm{zero}}(f,0) + \delta) + O(n^{2/3}\log(n)k)$. Since $\delta > 0$ is arbitrary, the claim follows. □

The next theorem proves the upper bound on the communication complexity of g_n.

Theorem 2.8.10 ([9]). *For any $\epsilon > 0$ we have $\mathrm{R}(g_M, \epsilon) \leq M\,\mathrm{IC}^{zero}(f,0) + o(M)k$.*

We will need the following non-distributional version of "information equals amortized communication" from [7].

Theorem 2.8.11 ([7]). *Let $g : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a function, and let $\mathrm{IC}(g,0) = I$. Then for each $\delta_1, \delta_2 > 0$, there exists $C = C(g, \delta_1, \delta_2)$ such that for all $N \geq C$, there exists a protocol $\pi_N = \pi_N((x_1, x_2, \ldots, x_N), (y_1, y_2, \ldots, y_N))$ for computing N instances of g. The protocol has communication complexity $< NI(1+\delta_1)$ and answers on all coordinates correctly except with probability δ_2.*

Proof. (of Theorem 2.8.10) We shall prove the statement for all sufficiently large M. We do not specified what "sufficiently large" means explicitly. Fix such an integer M. Choose n to be the largest integer such that $nC(g_n, 1/n, \epsilon) \leq M$, where the function C is guaranteed by Theorem 2.8.11. Without loss of generality, assume that $nC(g_n, 1/n, \epsilon) = M$, and let $N = C(g_n, 1/n, \epsilon)$. By Theorem 2.8.11, there exists a protocol π_N for solving N instances of g_n with communication $< N\,\mathrm{IC}(g_n, 0)(1 + 1/n)$ and solving all instances correctly except with probability ϵ. Now consider the protocol τ for solving g_M described in Protocol 9.

Protocol 9 Protocol τ for computing g_M that is based on protocol π_N for g_N.

Require:

$x \in \{0,1\}^M$ - known to Alice
$y \in \{0,1\}^M$ - known to Bob
π_N - known to Alice and Bob

1: Players divide the input into N blocks of size n each.
2: Players run π_N to solve these N instances of g_n.
3: Players output 1 if π_N outputs 1 on some instance.

Clearly the protocol has error $\leq \epsilon$. The communication cost of the protocol is at most :

$$\begin{aligned} N \operatorname{IC}(g_n, 0)(1 + 1/n) &\leq N(n \operatorname{IC}^{\text{zero}}(f, 0) + O(n^{2/3} \log(n)k))(1 + 1/n) \\ &\leq M \operatorname{IC}^{\text{zero}}(f, 0) + o(M)k \end{aligned}$$

□

2.8.3 Application: Exact Communication Complexity of DISJ_n

In this section we show how our previous results of this chapter imply Theorem 2.2.5. The negation of disjointness is an $\bigvee$-type function with f being the AND function. The tools obtained in Section 2.5 enable us to compute $\operatorname{IC}^{\text{zero}}(\text{AND}, 0)$. Therefore, we can use Theorem 2.2.4 to obtain the exact randomized communication complexity of DISJ_n with error tending to 0.

Theorem (Theorem 2.2.5 restated, [9]). *For all $\epsilon > 0$, there exists $\delta = \delta(\epsilon) > 0$ such that $\delta \to 0$ as $\epsilon \to 0$ and*

$$(C_{\mathrm{DISJ}} - \delta)n \leq \mathrm{R}(\mathrm{DISJ}_n, \epsilon) \leq C_{\mathrm{DISJ}} n + o(n).$$

where $C_{\mathrm{DISJ}} \approx 0.4827$ bits.

Note that the reductions in the proof of the upper bound and lower bound of Theorem 2.2.4 preserve the number of rounds. Hence, by Theorem 2.2.3, an r-round

protocol for DISJ_n is suboptimal, when compared to an unbounded round protocol, by at least $\Omega(n/r^2)$ communication.

Proof. Theorem 2.5.3 says that

$$\lim_{\epsilon \to 0} \max_{\mu : \mu(1,1) \leq \epsilon} \mathrm{IC}_\mu(\mathrm{AND}, 0) = C_{\mathrm{DISJ}} \approx 0.4827\ldots.$$

By Theorem 2.8.5 it follows that $\mathrm{IC}^{\mathrm{zero}}(\mathrm{AND}, 0) = C_{\mathrm{DISJ}}$. Now, Theorem 2.2.5 follows immediately from Theorem 2.2.4. □

2.9 Exact Communication Complexity of DISJ_n^k

We also study the DISJ_n problem with the promise that both Alice and Bob have sets of size $\leq k$. Recall from Section 2.1 that this partial function is denoted by DISJ_n^k. This problem was studied in [26]. It is also one of the problems that gives a separation between deterministic communication complexity and average-case 0-error communication complexity (e.g. see [33]). In [26] they proved the following theorem:

Theorem 2.9.1 ([26])**.** *For $\epsilon > 0$ we have*

$$\mathrm{R}(\mathrm{DISJ}_n^k, \epsilon) \leq O(k).$$

A lower bound of $\Omega(k)$ is immediate from the $\Omega(n)$ lower bound on the communication complexity of DISJ_n. We are able to determine the exact communication complexity of this problem except for some regimes.

Theorem 2.9.2 ([9])**.** *Let n, k be such that $k = \omega(1)$ and $n/k = \omega(1)$. Then for all $\epsilon > 0$, we have*

$$\left(\frac{2}{\ln 2} - O(\sqrt{\epsilon})\right) k - o(k) \leq \mathrm{R}(\mathrm{DISJ}_n^k, \epsilon) \leq \frac{2}{\ln 2} k + o(k).$$

The following subsections are devoted to proving this theorem. In the next subsection we shall start with the lower bound.

2.9.1 Lower Bound

Lemma 2.9.3 ([9]). *Let n, k be such that $k = \omega(1)$ and $n/k = \omega(1)$. Then*

$$\mathrm{R}(\mathrm{DISJ}_n^k, \epsilon) \geq \left(\frac{2}{\ln 2} - O(\sqrt{\epsilon})\right) k - o(k).$$

Proof. Similar to the proof of the lower bound for $\bigvee$-type functions, we will reduce the AND function to DISJ_n^k. Let π be a protocol such that π solves DISJ_n^k with error probability at most ϵ on all inputs and communication cost of π is $\mathrm{R}(\mathrm{DISJ}_n^k, \epsilon)$. Consider the following distribution μ on the input space $\{0,1\} \times \{0,1\}$:

$$\mu = \begin{array}{|c|c|} \hline 1 - 2(k - k^{2/3})/(n-1) & (k - k^{2/3})/(n-1) \\ \hline (k - k^{2/3})/(n-1) & 0 \\ \hline \end{array}$$

Let τ be the protocol obtained by applying Protocol 7 to π with distribution ν equal to μ as above. As before, we have

$$\mathrm{IC}_\mu(\tau) \leq \frac{\mathrm{IC}_{\mu^n}(\pi)}{n} \leq \frac{\mathrm{CC}(\pi)}{n} = \frac{\mathrm{R}(\mathrm{DISJ}_n^k, \epsilon)}{n}.$$

Recall that in τ the players sample $((X_i, Y_i) \sim \mu)_{i \neq J}$ and apply the protocol π to $(X_{<J} x X_{>J}, Y_{<J} y Y_{>J})$. Since μ puts 0 mass on $(1,1)$ entry for AND, the value of DISJ_n^k is determined by the value of $x \wedge y$ provided that the number of ones appearing in $X_{<J} X_{>J}$ is at most k, as well as the number of ones appearing in $Y_{<J} Y_{>J}$. Note that $P(X_i = 1) = \frac{k - k^{2/3}}{n-1}$. Let $S = \sum_{i \neq J} X_i$. By the multiplicative Chernoff bound we have

$$P(S > k-1) \leq \exp\left(\left(\frac{k^{2/3} - 1}{k - k^{2/3}}\right)^2 (k - k^{2/3})/3\right) \leq \exp\left(-\frac{k^{1/3}}{3}\right).$$

By the union bound, we have that the probability of either Alice or Bob sampling over $k-1$ ones in the coordinates other than J is at most $2\exp(-k^{1/3}/3)$. Provided that Alice and Bob sample at most $k-1$ ones in the coordinates other than J, protocol π outputs the correct value of DISJ_n^k, and therefore $x \wedge y$, with probability $\leq \epsilon$. Thus,

for all inputs (x, y) protocol τ computes the value of $\mathrm{AND}(x, y)$ with probability of error at most $\epsilon + \exp\left(-k^{\Omega(1)}\right)$. Overall, we have:

$$\mathrm{R}(\mathrm{DISJ}_n^k, \epsilon) \geq n \, \mathrm{IC}_\mu(\tau) \geq n \, \mathrm{IC}_\mu^{\mathrm{all}}\left(\mathrm{AND}, \epsilon + \exp\left(-k^{\Omega(1)}\right)\right).$$

To finish the proof we would like to apply a continuity argument to relate $\mathrm{IC}_\mu^{\mathrm{all}}(\mathrm{AND}, 0)$ to $\mathrm{IC}_\mu^{\mathrm{all}}(\mathrm{AND}, \epsilon + \exp\left(-k^{\Omega(1)}\right)$ and use Claim 2.5.1. Unfortunately, we cannot simply apply Theorem 2.2.8, since the convergence rate guaranteed by that theorem depends on the distribution μ and in our case μ depends on n. Thus Theorem 2.2.8 is not fine enough to let us control the convergence rate times n. Thus, we need to obtain a stronger convergence rate for this particular case:

Lemma 2.9.4 ([9])**.** *Let* $\nu = \begin{array}{|c|c|}\hline 1-2k/n & k/n \\ \hline k/n & 0 \\ \hline \end{array}$*, then we have*

$$\mathrm{IC}_\nu^{all}(\mathrm{AND}, 0) = \mathrm{IC}_\nu^{all}(\mathrm{AND}, \epsilon) + O\left(\frac{k}{n}\sqrt{\epsilon}\right).$$

We shall prove this lemma at the end of this subsection. Now, we finish the proof of the current lemma assuming Lemma 2.9.4. By Claim 2.5.1 we have

$$\begin{aligned}\mathrm{IC}_\mu^{\mathrm{all}}(\mathrm{AND}, 0) &= \frac{\beta}{\ln 2} + \frac{\beta^2}{\alpha}\log\frac{\beta}{\beta+\alpha} + \alpha\log\frac{\alpha+\beta}{\alpha} \\ &= \frac{2}{\ln 2}\frac{k}{n-1} \pm o\left(\frac{k}{n-1}\right)\end{aligned}$$

Thus by Lemma 2.9.4

$$\begin{aligned}\mathrm{R}(\mathrm{DISJ}_n^k, \epsilon) &\geq n \, \mathrm{IC}_\mu^{\mathrm{all}}\left(\mathrm{AND}, \epsilon + \exp\left(-k^{\Omega(1)}\right)\right) \\ &= n\left(\mathrm{IC}_\mu^{\mathrm{all}}(\mathrm{AND}, 0) - O\left(\frac{(k-k^{2/3})}{n-1}\sqrt{\epsilon + \exp\left(-k^{\Omega(1)}\right)}\right)\right) \\ &= n\left(\frac{2}{\ln 2}\frac{k}{n-1} + o\left(\frac{k}{n-1}\right) - O\left(\frac{(k-k^{2/3})}{n-1}\sqrt{\epsilon + \exp\left(-k^{\Omega(1)}\right)}\right)\right) \\ &\geq \left(\frac{2}{\ln 2} - O\left(\sqrt{\epsilon}\right)\right)k - o(k)\end{aligned}$$

□

In the rest of this subsection we prove Lemma 2.9.4. We shall need the following proposition:

Proposition 2.9.5 ([9])**.** *Let* $\chi = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \gamma & 0 \\ \hline \end{array}$ *be a distribution on* $\{0,1\} \times \{0,1\}$ *such that* $\beta \leq \gamma$*. Then we have*

$$\mathrm{IC}_{\chi}^{all}(\mathrm{AND}, 0) \leq O(\beta \log(2\gamma/\beta)).$$

Proof. Let us first consider the information complexity of AND with respect to a symmetric distribution. Consider the distribution $\nu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \beta & 0 \\ \hline \end{array}$
By Claim 2.5.1 we have

$$\begin{aligned} \mathrm{IC}_{\nu}^{\mathrm{all}}(\mathrm{AND}, 0) &= \frac{\beta}{\ln 2} + \frac{\beta^2}{\alpha} \log \frac{\beta}{\beta + \alpha} + \alpha \log \frac{\alpha + \beta}{\alpha} \\ &\leq \frac{2\beta}{\ln 2} \end{aligned}$$

By Claim 2.5.2 $\mathrm{IC}_{\mu}^{\mathrm{all}}(\mathrm{AND}, 0)$ consists of the information cost of the symmetrization step plus t times the information complexity of AND with respect to the resulting symmetric distribution, which is $\leq tO(\beta/t) \leq O(\beta)$. As for the symmetrization step, its information cost is

$$\begin{aligned} &(\alpha + \beta) H\left(\frac{\beta}{\alpha + \beta} + \frac{\beta}{\gamma} \cdot \frac{\alpha}{\alpha + \beta}\right) - \alpha H\left(\frac{\beta}{\gamma}\right) \\ &= (\alpha + \beta) H\left(\frac{\beta}{\gamma} + \frac{\beta}{\alpha + \beta}\left(1 - \frac{\beta}{\gamma}\right)\right) - \alpha H\left(\frac{\beta}{\gamma}\right) \\ &\leq (\alpha + \beta)\left(H\left(\frac{\beta}{\gamma}\right) + \frac{\beta}{\alpha + \beta}\left(1 - \frac{\beta}{\gamma}\right) H'\left(\frac{\beta}{\gamma}\right)\right) - \alpha H\left(\frac{\beta}{\gamma}\right) \\ &= \beta H\left(\frac{\beta}{\gamma}\right) + \beta\left(1 - \frac{\beta}{\gamma}\right)\left(\log\left(1 - \frac{\beta}{\gamma}\right) - \log\left(\frac{\beta}{\gamma}\right)\right) \\ &= \beta \log\left(\frac{\gamma}{\beta}\right) \end{aligned}$$

The inequality in the above follows from the concavity of the entropy function. This finishes the proof of this proposition. □

Now we are ready to prove Lemma 2.9.4. Fix $\delta > 0$ and let π be a protocol such that on all inputs π solves AND correctly except with probability of error at most ϵ and $\mathrm{IC}_\nu(\pi) \leq \mathrm{IC}^{\mathrm{all}}_\nu(\mathrm{AND}, \epsilon) + \delta$. Let σ be the protocol obtained from π and ν as described in Protocol 10. It is clear that σ solves AND correctly on all inputs. It is left to analyze the information cost of σ with respect to ν.

Protocol 10 Protocol σ for computing AND with 0 error that is based on ϵ-error protocol π for AND.

Require:

$x \in \{0, 1\}$ - known to Alice
$y \in \{0, 1\}$ - known to Bob
π, ν - known to Alice and Bob

1: Players tun the protocol π on x and y.
2: Upon reaching a leaf ℓ, players run a protocol τ such that $\mathrm{IC}_{\nu_\ell}(\tau) \leq \mathrm{IC}^{\mathrm{all}}_{\nu_\ell}(\mathrm{AND}, 0) + \delta$ and τ solves AND on all inputs with 0 error. Here, ν_ℓ is ν conditioned on reaching ℓ in π.
3: Players output according to the output of τ.

Alice and Bob start with the distribution $\nu = \begin{array}{|c|c|}\hline 1-2k/n & k/n \\ \hline k/n & 0 \\ \hline \end{array}$. Let ℓ be a leaf in the communication tree of π. Let p_ℓ denote the probability of reaching ℓ in π, where the probability is taken over the inputs $(X, Y) \sim \nu$ and the randomness of the protocol. Let ν_ℓ be the distribution on the players' inputs conditioned on reaching the leaf ℓ. Since the initial distribution ν places 0 mass on $(1, 1)$ entry, the distribution ν_ℓ also places 0 mass on $(1, 1)$ entry (by Corollary 2.4.2). In particular, ν_ℓ has the following form:

$$\nu_\ell = \begin{array}{|c|c|}\hline \alpha_\ell & \beta_\ell \\ \hline \gamma_\ell & 0 \\ \hline \end{array}.$$

By Proposition 2.9.5, the amount of extra information leaked by σ as compared to π is

$$O\left(\sum_{\ell:\beta_\ell \leq \gamma_\ell} p_\ell \beta_\ell \log \frac{2\gamma_\ell}{\beta_\ell} + \sum_{\ell:\beta_\ell > \gamma_\ell} p_\ell \gamma_\ell \log \frac{2\beta_\ell}{\gamma_\ell}\right).$$

Thus, we need to show that the above expression is $O\left(\frac{k}{n}\sqrt{\epsilon}\right)$. We shall upper bound the terms with $\beta_\ell \leq \gamma_\ell$, since the other terms can be upper bounded similarly. Therefore, for the rest of the argument we shall assume that $\beta_\ell \leq \gamma_\ell$. First, we establish the probability of π reaching a certain leaf.

Claim 2.9.6 ([9])**.**

1. $P_{(X,Y)\sim\nu}(\Pi(X,Y)$ *reaches leaf* $\ell \mid X = 1, Y = 1) = p_\ell \frac{\beta_\ell}{k/n} \frac{\gamma_\ell}{k/n} \frac{1-2k/n}{\alpha_\ell}$.
2. $P_{(X,Y)\sim\nu}(\Pi(X,Y)$ *reaches leaf* $\ell \mid X = 1, Y = 0) = p_\ell \frac{\gamma_\ell}{k/n}$.

Note: the above probabilities are also taken over the randomness of π*, but for brevity purposes we suppress that notation.*

Proof. By the Bayes' rule we have

$$\begin{aligned}
&P_{(X,Y)\sim\nu}(\Pi(X,Y) \text{ reaches leaf } \ell \mid X = a, Y = b) \\
&= \frac{P_{(X,Y)\sim\nu}(X = a, Y = b \mid \Pi(X,Y) \text{ reaches leaf } \ell) P_{(X,Y)\sim\nu}(\Pi(X,Y) \text{ reaches leaf } \ell)}{P_{(X,Y)\sim\nu}(X = a, Y = b)} \\
&= \frac{\nu_\ell(a,b) p_\ell}{\nu(a,b)}
\end{aligned}$$

In particular, $P_{(X,Y)\sim\nu}(\Pi(X,Y) \text{ reaches leaf } \ell \mid X = 1, Y = 0) = p_\ell \frac{\gamma_\ell}{k/n}$, which establishes the second part of the claim. Also, we have $P_{(X,Y)\sim\nu}(\Pi(X,Y) \text{ reaches leaf } \ell \mid X = 0, Y = 1) = p_\ell \frac{\beta_\ell}{k/n}$ and $P_{(X,Y)\sim\nu}(\Pi(X,Y) \text{ reaches leaf } \ell \mid X = 0, Y = 0) = p_\ell \frac{\alpha_\ell}{1-2k/n}$.

Let $\mathcal{L}$ denote the set of all leaves of π. Due to the rectangular nature of communication protocols (see [3]), there exists functions $p_{\mathcal{A}} : \{0,1\} \times \mathcal{L} \to [0,1]$ and $p_{\mathcal{B}} : \{0,1\} \times \mathcal{L} \to [0,1]$ such that

$$P_{(X,Y)\sim\nu}(\Pi(X,Y) \text{ reaches leaf } \ell \mid X = a, Y = b) = p_{\mathcal{A}}(a,\ell) p_{\mathcal{B}}(b,\ell).$$

Using this notation, we restate the results from the beginning of the proof as follows:

$$p_{\mathcal{A}}(1,\ell)p_{\mathcal{B}}(0,\ell) = p_\ell \frac{\gamma_\ell}{k/n}$$
$$p_{\mathcal{A}}(0,\ell)p_{\mathcal{B}}(1,\ell) = p_\ell \frac{\beta_\ell}{k/n}$$
$$p_{\mathcal{A}}(0,\ell)p_{\mathcal{B}}(0,\ell) = p_\ell \frac{\alpha_\ell}{1-2k/n}$$

Finally, we have

$$\begin{aligned} p_{\mathcal{A}}(1,\ell)p_{\mathcal{B}}(1,\ell) &= \frac{p_{\mathcal{A}}(1,\ell)p_{\mathcal{B}}(0,\ell)p_{\mathcal{A}}(0,\ell)p_{\mathcal{B}}(1,\ell)}{p_{\mathcal{A}}(0,\ell)p_{\mathcal{B}}(0,\ell)} \\ &= \left(p_\ell \frac{\gamma_\ell}{k/n}\right)\left(p_\ell \frac{\beta_\ell}{k/n}\right) / \left(p_\ell \frac{\alpha_\ell}{1-2k/n}\right) \\ &= p_\ell \frac{\beta_\ell}{k/n}\frac{\gamma_\ell}{k/n}\frac{1-2k/n}{\alpha_\ell}, \end{aligned}$$

which establishes the first part of the claim. □

We shall require one other claim:

Claim 2.9.7 ([9])**.** *For $\beta \le \gamma$ we have either $\beta \log(2\gamma/\beta) < 4\sqrt{2\epsilon}$ or $\beta \log(2\gamma/\beta) < \frac{\beta\gamma}{\sqrt{2\epsilon}}$.*

Proof. We shall prove this statement by contradiction. Assume that

1. $\beta \log(2\gamma/\beta) \ge 4\sqrt{2\epsilon}$, and
2. $\beta \log(2\gamma/\beta) \ge \frac{\beta\gamma}{\sqrt{2\epsilon}}$.

Since $\beta \le \gamma$ there exists $q \ge 1$ such that $\frac{2\gamma}{\beta} = 2^q$. Then we can rewrite the above inequalities as follows:

1. $\beta q \ge 4\sqrt{2\epsilon}$, and
2. $\sqrt{2\epsilon} \ge \frac{\gamma}{q}$.

Combining the above inequalities we obtain $q^2 \ge 2^{q+1}$. Contradiction. □

For $i \in \{0,1\}$ define $\mathcal{L}_i = \{\ell \mid \beta_\ell \leq \gamma_\ell \text{ and } \pi \text{ outputs } i \text{ on } \ell\}$. By Claim 2.9.6 the probability of π making a mistake on $(1,1)$ entry is

$$\sum_{\ell \in \mathcal{L}_0} p_\ell \frac{\beta_\ell}{k/n} \frac{\gamma_\ell}{k/n} \frac{1-2k/n}{\alpha_\ell} \leq \epsilon \quad (\pi \text{ is } \epsilon\text{-error})$$
$$\implies \sum_{\ell \in \mathcal{L}_0} p_\ell \frac{\beta_\ell}{k/n} \frac{\gamma_\ell}{k/n} \leq 2\epsilon \quad (k/n = o(1))$$

Define $\beta'_\ell = \frac{\beta}{k/n}$ and $\gamma'_\ell = \frac{\gamma}{k/n}$.The contribution of extra information cost from leaves from $\mathcal{L}$ is

$$\begin{aligned}
\sum_{\ell \in \mathcal{L}_0} p_\ell \beta_\ell \log \frac{2\gamma_\ell}{\beta_\ell} &= \frac{k}{n} \sum_{\ell \in \mathcal{L}_0} p_\ell \beta'_\ell \log \frac{2\gamma'_\ell}{\beta'_\ell} \\
&\leq \frac{k}{n} \left(4\sqrt{2\epsilon} + \sum_{\ell \in \mathcal{L}_0} p_\ell \frac{\beta'_\ell \gamma'_\ell}{\sqrt{2\epsilon}} \right) \quad \text{(by Claim 2.9.7)} \\
&\leq \frac{k}{n} \left(4\sqrt{2\epsilon} + 2\epsilon/\sqrt{2\epsilon} \right) \quad \text{(by above)} \\
&= \frac{k}{n} (5\sqrt{2\epsilon})
\end{aligned}$$

This proves that the extra information contributed by all leaves in $\mathcal{L}_0$ is $O(k\sqrt{\epsilon}/n)$, as claimed.

Now, we consider leaves in $\mathcal{L}_1$. If a leaf in $\mathcal{L}_1$ is reached on input $X = 1$ and $Y = 0$ then the protocol π makes a mistake. Since π has at most ϵ error on all inputs, it follows by Claim 2.9.6 that

$$\sum_{\ell \in \mathcal{L}_1} p_\ell \frac{\gamma_\ell}{k/n} \leq \epsilon$$

The contribution of the leaves from $\mathcal{L}_1$ to the extra information cost is

$$\begin{aligned}\sum_{\ell\in\mathcal{L}_1} p_\ell\beta_\ell \log(2\gamma_\ell/\beta_\ell)) &= \frac{k}{n}\sum_{\ell\in\mathcal{L}_1} p_\ell\beta'_\ell \log(2\gamma'_\ell/\beta'_\ell)) \\ &\le \frac{k}{n}\sum_{\ell\in\mathcal{L}_1} p_\ell 2\gamma'_\ell \quad (\text{since } \beta'_\ell \log(2\gamma'_\ell/\beta'_\ell) \le 2\gamma'_\ell) \\ &\le \frac{k}{n}(2\epsilon)\end{aligned}$$

This completes the proof of Lemma 2.9.4.

2.9.2 Upper Bound

Now we prove the upper bound on the communication complexity of DISJ^k_n. We start by proving an upper bound on the 0-error information complexity of the problem. Notation: we let $S^k_n = \{(x,y) \mid x,y \in \{0,1\}^n, |x|,|y| \le k\} \subseteq \{0,1\}^n \times \{0,1\}^n$.

Lemma 2.9.8 ([9])**.** *Let n and k be such that $k = \omega(1)$ and $n/k = \omega(1)$. Then there exists a protocol π_n such that*

1. *for all $(x,y) \in S^k_n$ protocol π_n correctly outputs $\mathrm{DISJ}^k_n(x,y)$, and*
2. *for all distributions μ over $\{0,1\}^n \times \{0,1\}^n$ if $\mathrm{supp}(\mu) \subseteq S^k_n$ then $\mathrm{IC}_\mu(\pi_n) \le \frac{2}{\ln 2}k + o(k)$.*

Remark 2.9.1. If there exists a protocol π_n that solves DISJ^k_n on inputs of size exactly k with information cost $\frac{2}{\ln 2}k + o(k)$ then there exists a protocol π'_n that solves DISJ^k_n on inputs of size $\le k$ with the same information cost. Let x and y be inputs to Alice and Bob in π'_n such that $|x|,|y| \le k$. Alice appends $k + |x|$ zeroes followed by $k - |x|$ ones to x and calls it x', and Bob appends $k - |y|$ ones followed by $k + |y|$ zeroes and calls it y'. Then players run π_{n+2k} on (x',y') and output whatever π_{n+2k} outputs. Since $|x|,|y| \le k$ we have $\mathrm{DISJ}^k_n(x,y) = \mathrm{DISJ}^k_{n+2k}(x',y')$. Moreover, since the information cost of π_n does not depend on the universe size, the information cost of π'_n is the same as that of π_n. Therefore, without loss of generality we may assume that players receive inputs of size exactly k. The same argument holds for communication complexity.

Proof. By the above remark, assume that Alice and Bob both have sets of size exactly k.

Let τ denote the protocol solving AND correctly on all inputs and having information cost $\mathrm{IC}_\nu^{\mathrm{all}}(\mathrm{AND}, 0)$, where

$$\nu = \begin{array}{|c|c|}\hline 1-2k/n & k/n \\ \hline k/n & 0 \\ \hline \end{array}$$

For the clarity of exposition, we shall ignore the issue of information complexity not being achievable by a single protocol. If we wanted to, we could deal with this issue in a standard way – if the optimal information complexity is I then for all $\delta > 0$ there exists a protocol of information cost $I + \delta$. Fix such a protocol, and carry out the whole argument with this δ additive term. At the end of the argument the δ term would disappear, since $\delta > 0$ is arbitrary.

Let σ be the 0-error protocol for the equality function EQ ($\mathrm{EQ}(x, y) = 1$ if and only if $x = y$) described in [7]. In [7] it is proved that σ has constant information cost independent of the size of the universe. Let π_n be the protocol described in Protocol 11.

It is clear that for all $(x, y) \in S_n^k$ we have $\pi_n(x, y) = \mathrm{DISJ}_n^k(x, y)$, which establishes the first part of the lemma. Next, we analyze the information cost of this protocol.

Let μ be a distribution on $\{0,1\}^n \times \{0,1\}^n$ such that $\mathrm{supp}(\mu) \subseteq S_n^k \setminus S_n^{k-1}$. Let $(X, Y) \sim \mu$ denote the input random variables. We can write $\Pi_n = J\Pi_n^1\Pi_n^2$, where J denotes the random coordinates sampled at the beginning of π_n, Π_n^1 denotes the transcript corresponding to Step 1 (lines 3-7) of π_n and Π_n^2 denotes the transcript for Step 2 (lines 8-12) of the protocol. Then

$$\begin{aligned}
\mathrm{IC}_\mu(\pi_n) &= I(\Pi_n; X|YJ) + I(\Pi_n; Y|XJ) \\
&= I(\Pi_n^1; X|YJ) + I(\Pi_n^1; Y|XJ) \\
&+ I(\Pi_n^2; X|YJ\Pi_n^1) + I(\Pi_n^2; Y|XJ\Pi_n^1)
\end{aligned}$$

Protocol 11 Protocol π for computing DISJ_n^k that is based on protocol τ for AND and σ for EQ.

Require:
$x \in \{0,1\}^n, |x| = k$ - known to Alice
$y \in \{0,1\}^n, |y| = k$ - known to Bob
τ, σ - known to Alice and Bob

Alice samples a set $S_{\mathcal{A}} \subseteq [n]$ of $n/k^{2/3}$ coordinates using public randomness.
Bob samples a set $S_{\mathcal{B}} \subseteq [n]$ of $n/k^{2/3}$ coordinates using public randomness.
for $a \in S_{\mathcal{A}} \cap x$ **do**
    for $b \in S_{\mathcal{B}} \cap y$ **do**
        Alice and Bob run σ on (a, b)
        if $\sigma(a, b) = 1$, i.e, $a = b$ **then**
            Players output 0, protocol terminates.
for $i \in [n]$ **do**
    Players run τ on (x_i, y_i)
    if $\tau(x_i, y_i) = 1$, i.e., $x_i \wedge y_i = 1$ **then**
        Players output 0, protocol terminates.
Players output 1, protocol terminates

Claim 2.9.9 ([9])**.**

$$I(\Pi_n^1; X|YJ) + I(\Pi_n^1; Y|XJ) = O(k^{2/3})$$

Proof. We have

$$\begin{aligned}
I(\Pi_n^1; X|YJ) &= I(\Pi_n^1; X|Y(S_{\mathcal{B}} \cap Y)J) \quad (J \text{ and } Y \text{ determine } S_{\mathcal{B}} \cap Y) \\
&= I(\Pi_n^1; X(S_{\mathcal{A}} \cap X)|Y(S_{\mathcal{B}} \cap Y)J) \quad (J \text{ and } X \text{ determine } S_{\mathcal{A}} \cap X) \\
&= I(\Pi_n^1; (S_{\mathcal{A}} \cap X)|Y(S_{\mathcal{B}} \cap Y)J) \\
&\leq I(\Pi_n^1; (S_{\mathcal{A}} \cap X)|(S_{\mathcal{B}} \cap Y)J) \\
&\quad \text{(last two steps are due to } \Pi_n^1 \text{ being determined by } S_{\mathcal{B}} \cap Y, S_{\mathcal{A}} \cap X)
\end{aligned}$$

Similarly, we have $I(\Pi_n^1; Y|XJ) \leq I(\Pi_n^1; (S_{\mathcal{A}} \cap Y)|(S_{\mathcal{B}} \cap X)J)$. Now, we have

$$I(\Pi_n^1; A|BJ) + I(\Pi_n^1; B|AJ) = \mathbb{E}_J(I(\Pi_1; A|BJ) + I(\Pi_1; B|AJ))$$

Since σ is a 0-error protocol for EQ that has $O(1)$ information cost with respect to every distribution (Proposition 3.21 in [7]), we have

$$I(\Pi_n^1; A|BJ) + I(\Pi_n^1; B|AJ) = \mathbb{E}_J\left(O(|S_{\mathcal{A}} \cap X||S_{\mathcal{B}} \cap Y|)\right)$$

Let A_i denote the indicator random variable for the event that ith sampled element is in Alice's set. Define B_i, analogously, for Bob.

$$\begin{aligned}
\mathbb{E}_J(|S_{\mathcal{A}} \cap X||S_{\mathcal{B}} \cap Y|) &= \mathbb{E}_J(|S_{\mathcal{A}} \cap X|)\mathbb{E}_J(|S_{\mathcal{B}} \cap Y|) \quad \text{(indep.)} \\
&= \mathbb{E}\left(\left(\sum_{i=1}^{n/k^{2/3}} A_i\right)\left(\sum_{i=1}^{n/k^{2/3}} B_i\right)\right) \\
&= \left(\frac{k}{n}\frac{n}{k^{2/3}}\right)^2 \\
&= k^{2/3}
\end{aligned}$$

This completes the proof of the claim. □

Next, we analyze the information cost of Step 2 of π_n. Let E denote the indicator random variable for the event that the players find a common element in Step 1. Since E is determined by J and Π_n^1, we have

$$\begin{aligned}
I(\Pi_n^2; X|YJ\Pi_n^1) &= I(\Pi_n^2; X|YJ\Pi_n^1 E) \\
&= P(E=0)I(\Pi_n^2; X|YJ\Pi_n^1, E=0) \\
&\leq P(E=0)I(\Pi_n^2; X|Y, E=0) \qquad (2.19)
\end{aligned}$$

The last inequality follows from the fact that J, Π_n^1 are independent of Π_n^2 conditioned on $E = 0$.

We shall need the following proposition :

Proposition 2.9.10 ([9]). *Let* $\nu = \begin{array}{|c|c|}\hline \alpha & \beta \\ \hline \beta & \gamma \\ \hline \end{array}$ *Then* $\mathrm{IC}_\nu(\mathrm{AND}, 0) = O(\beta + \gamma)$

Proof. Follows from Claim 2.5.4 and the fact that $\ln(1+x) \leq x$. □

Now let $\mu' = \mu|(E=0)$ and let μ'_i denote the marginal distribution of μ' on the ith coordinate. Then using Lemma 2.8.9, we get that

$$\begin{aligned} &I(\Pi_n^2; X|Y, E=0) + I(\Pi_n^2; Y|X, E=0) \\ &\leq \sum_{i=1}^{n} \mathrm{IC}_{\mu'_i}(\tau) \leq n\,\mathrm{IC}_{\sum_{i=1}^n \mu'_i/n}(\tau) \end{aligned} \tag{2.20}$$

where the last inequality follows from the concavity of information cost. Define $N_{(a,b)}(x,y) = |\{\ i \text{ s.t. } x_i = a, y_i = b\ \}|$. Let $\nu = (\sum_{i=1}^n \mu'_i)/n$. Then $\nu(a,b) = \mathbb{E}_{(X,Y)\sim\mu'}(N_{(a,b)}(X,Y)/n)$. Since we assumed that both Alice and Bob have sets of size k, we get that $\nu(1,0) + \nu(1,1) = k/n$ and $\nu(0,1) + \nu(1,1) = k/n$. Thus by Proposition 2.9.10, $\mathrm{IC}_\nu(\tau) \leq O(k/n)$ (note that the information optimal protocol τ for AND is the same with respect to all symmetric distributions). Thus

$$I(\Pi_2; X|Y, E=0) + I(\Pi_2; Y|X, E=0) \leq O(k)$$

Now, if $P(E=0) \leq 1/k^{1/3}$, then from equation (2.19) and Claim 2.9.9 it follows that information cost of the entire protocol is $O(k^{2/3})$ and we are done. Therefore, we assume that $P(E=0) \geq 1/k^{1/3}$ for the rest of the argument.

Abusing the notation, we let $\mu(d)$ denote the mass on strings such that $N_{(1,1)}(x,y) = d$. Observe that $\mu(> k) = 0$. Now

$$\mu'(d) = \frac{\mu(d) P(E=0|d \text{ common elements})}{P(E=0)}$$

By the multiplicative Chernoff bound, we have

$$P(E=0|d \text{ common elements}) \leq e^{-d/2k^{2/3}}$$

Thus for $d \geq k^{3/4}$, we have

$$\mu'(d) \leq \mu(d) e^{-k^{\Omega(1)}} k^{1/3}$$

This implies that $\nu(1,1) = \mathbb{E}_{(X,Y)\sim\mu'}(N_{(1,1)}(X,Y)/n) \leq k^{3/4}/n$. Now, consider the information complexity of AND with respect to ν. By Claim 2.5.4, the information complexity is at most

$$\frac{\beta}{\ln 2} + 2\gamma \log \frac{\beta+\gamma}{\gamma} + 2\beta \log \frac{\beta+\gamma}{\beta} + \alpha \log \frac{\alpha+\beta}{\alpha}$$

where $k/n - k^{3/4} \leq \beta \leq k/n$, $\gamma \leq k^{3/4}/n$ and $\alpha = 1 - 2\beta - \gamma$. It is easy to observe that this is $\leq \frac{2}{\ln 2}\frac{k}{n} + o\left(\frac{k}{n}\right)$. Thus, the contribution from Step 2 of π_n is $\leq \frac{2}{\ln 2}k + o(k)$ by (2.19) and (2.20). This finishes the proof of the lemma. □

In fact, using the above arguments, we can get the following stronger lemma :

Lemma 2.9.11 ([9])**.** *Let n and k be such that $k = \omega(1)$ and $n/k = \omega(1)$. Then there exists a protocol π_n such that*

1. *π_n solves SETINT_n correctly on all inputs $(x,y) \in S_n^k$ with $|x \wedge y| = o(k)$, and*
2. *for all distributions μ over $\{0,1\}^n \times \{0,1\}^n$ if $\mathrm{supp}(\mu) \subseteq S_n^k$ then $\mathrm{IC}_\mu(\pi_n) \leq \frac{2}{\ln 2}k + o(k)$.*

We shall need this lemma later. Next, we prove the upper bound on the communication complexity of DISJ_n^k.

Theorem 2.9.12 ([9])**.** *Let M and k be such that $k = \omega(1)$ and $M/k = \omega(1)$. Then for all $\epsilon > 0$ we have $\mathrm{R}(\mathrm{DISJ}_M^k, \epsilon) \leq \frac{2}{\ln 2}k + o(k)$.*

We prove this theorem via self-reducibility. Since DISJ_n^k is a partial function (= promise problem), we start by generalizing the information complexity to promise problems.

Definition 2.9.1. Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a function. Let $S \subseteq \mathcal{X} \times \mathcal{Y}$. The promise problem associated with f and S, denoted by f_S, is to compute f via a communication protocol that is correct only on inputs from S and outputs whatever it wants on inputs from $\mathcal{X} \times \mathcal{Y} \setminus S$. Let us call a protocol good for f_S if it answers correctly on all inputs in S. Then *the 0-error information complexity of the promise problem f_S* s defined as follows: $\mathrm{IC}(f_S, 0) = \inf_{\pi \text{ good}} \max_{\mu \in \Delta(S)} \mathrm{IC}_\mu(\pi)$

The following non-distributional version of "information equals amortized communication" for promise problems follows from the same techniques as in Theorem 2.8.11 from [7].

Theorem 2.9.13 ([7]). *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a function, let $S \subseteq \mathcal{X} \times \mathcal{Y}$. let f_S be the corresponding promise problem. Let $\mathrm{IC}(f_S, 0) = I$. Let $P \subseteq \mathcal{X}^N \times \mathcal{Y}^N$ be the subset of inputs such that promise holds at each coordinate i.e. $(x_i, y_i) \in S$ for all i. Then for each $\delta_1, \delta_2 > 0$, there exists $C = C(f_S, \delta_1, \delta_2)$ such that for each $N \geq C$, there exists a protocol $\pi_N = \pi_N((x_1, x_2, \ldots, x_N), (y_1, y_2, \ldots, y_N))$ for computing N instances of f_S. The protocol has communication complexity $< NI(1+\delta_1)$ and for all inputs in P the protocol π_N answers on all coordinates correctly except with probability δ_2.*

We will prove prove the upper bound in several steps. Initially, we will prove it for the values of k that are close to the universe size. After that, we shall generalize this upper bound to all values of k such that $k = \omega(1)$. Let $\epsilon > 0$. Let $t(M) =$ the largest $n \in \mathbb{N}$ such that $nC(\mathrm{DISJ}_n^{n^{3/4}}, 1/n, \epsilon - 1/n) \leq M$. Note that $t(M)$ is an extremely slowly growing function of M but still it goes to infinity with M.

Lemma 2.9.14 ([9]). *For all $\epsilon > 0, k$, and M such that $\omega(1) \leq M/k \leq t(M)^{3/4}$ we have $\mathrm{R}(\mathrm{DISJ}_M^k, \epsilon) \leq \frac{2}{\ln 2}k + o(k)$.*

Proof. Let $n = t(M)^{3/4}$ and $N \geq C(\mathrm{DISJ}_n^{n^{3/4}}, 1/n, \epsilon - 1/n)$ be the largest integer such that $nN \leq M$. We can assume that $nN = M$. We shall prove the Lemma for $M/k = t(M)^{3/4}$. The Lemma for $M/k \leq t(M)^{3/4}$ can be proved by the same argument by taking a smaller yet super-constant n and a larger N. Let $\ell = n^{3/4} + n^{2/3}$. Using Theorem 2.9.13, there exists a protocol π_N for solving N copies of DISJ_n^ℓ with communication $< N\,\mathrm{IC}(\mathrm{DISJ}_n^\ell, 0)(1 + 1/n)$ such that if all copies have sets of size $\leq \ell$, then π_N answers correctly on all copies, except with probability $\epsilon - 1/n$. Now consider Protocol 12 for solving DISJ_M^k.

First, lets see what the error of the protocol is. Note that Alice and Bob both have sets of size $M/n^{1/4}$. Consider a particular block B. Define a super-martingale sequence $X_0, X_1, \ldots, X_n$ as follows : $X_0 = 0$, $X_i =$ (number of elements in the first

Protocol 12 Protocol π for computing DISJ_M^k that is based on protocol π_N for N copies of DISJ_n^ℓ.

Require:
$x \in \{0,1\}^M, |x| = k$ - known to Alice
$y \in \{0,1\}^M, |y| = k$ - known to Bob
$\pi_N, M = nN$ - known to Alice and Bob

1: Alice and Bob use public randomness to divide their inputs randomly into N blocks of size n.
2: Alice computes $c_{\mathcal{A}}$ – the number of blocks that contain more than ℓ elements of her input. Such blocks are called bad for Alice.
3: Bob computes $c_{\mathcal{B}}$ – the number of blocks that contain more than ℓ elements of his input. Such blocks are called bad for Bob.
4: **if** $c_{\mathcal{A}} \geq Nn^2 \exp(-2n^{1/3}) \vee c_{\mathcal{B}} \geq Nn^2 \exp(-2n^{1/3})$ **then**
5: Players outputs a random bit as the answer of the protocol. Protocol terminates.
6: Alice creates a string $q_{\mathcal{A}} \in \{0,1\}^N$ where $(q_{\mathcal{A}})_i = 1$ if and only if the ith block is bad for Alice.
7: Bob creates a string $q_{\mathcal{B}} \in \{0,1\}^N$ where $(q_{\mathcal{B}})_i = 1$ if and only if the ith block is bad for Bob.
8: Alice sends $q_{\mathcal{A}}$ to Bob.
9: Bob sends $q_{\mathcal{B}}$ to Alice.
10: Alice and Bob use brute force protocol to solve DISJ_n on blocks that are bad for at least one of them.
11: Alice and Bob use π_N to solve DISJ_n^ℓ on the rest of the blocks.
12: The players output 0 if either brute force or π_N outputs 0 on any single block; otherwise they output 1. The protocol terminates.

i coordinates of B) $- i\frac{M}{n^{1/4}(M-n)}$. It is clear that this is a super-martingale since conditioned on $X_0, \ldots, X_{i-1}$, the probability of the ith coordinate of B being an element of Alice is at most $\frac{M}{n^{1/4}(M-n)}$. Also we have $|X_i - X_{i-1}| \leq 1$. Therefore, by the Azuma's inequality we have:

$$P\left(X_n \geq n^{2/3}\right) \leq \exp(-2n^{1/3})$$

Thus, except with probability $\exp(-2n^{1/3})$, the number of elements of Alice in block B is $\leq n^{3/4} + n^{2/3}$ (we are ignoring the $\frac{M}{M-n} = \frac{N}{N-1}$ factor, since it is almost 1). By Markov's inequality, the probability that the number of bad blocks for Alice

$\geq Nn^2 \exp(-2n^{1/3})$ is $\leq 1/n^2$. Hence the probability that either Alice or Bob have a large number of bad blocks $\leq 2/n^2 \leq 1/n$. Now π_N answers correctly except with probability $\epsilon - 1/n$, thus the total error $\leq \epsilon$. The total communication cost is

$$\begin{aligned}
&< N \operatorname{IC}(\mathrm{DISJ}_n^\ell, 0)(1 + 1/n) + Nn^3 \exp(-2n^{1/3}) + O(N) \\
&= N\left(\frac{2}{\ln 2} n^{3/4} + o(n^{3/4})\right) + No(n^{3/4}) \\
&= \frac{2}{\ln 2} k + o(k)
\end{aligned}$$

□

Using the stronger lemma about the information complexity of the set intersection problem, Lemma 2.9.11, we get the following stronger lemma:

Lemma 2.9.15 ([9])**.** *For all $\epsilon > 0, k$, and M such that $\omega(1) \leq M/k \leq t(M)^{3/4}$, there exists a protocol π with communication $\leq \frac{2}{\ln 2} k + o(k)$ such that if Alice and Bob both have sets of size $\leq k$ and the size of their intersection is $o(k)$, then π outputs the intersection, otherwise it outputs that the intersection is large.*

Next, we generalize the upper bound for all values of k and M such that $k = \omega(1)$ and $M/k = \omega(1)$.

Proof of Theorem 2.9.12. The central idea of the proof is to reduce the size of the universe by hashing and then apply Lemma 2.9.15. Let ℓ be such that $\frac{\ell}{k} = t(\ell)^{3/4}$. Consider the following protocol :

It is clear that the protocol is correct with high probability. If $|X \cap Y| > k^{3/4}$ then the probability of each element in the intersection to survive selection in the first sampling step is $k^{-1/2}$, and thus the parties output 0 except with an exponentially small probability plus the probability that the Håstad-Wigderson protocol fails a majority of times, which is bounded by $1/k$. In all other cases the protocol will only fail if π returns that the intersection is too large or if the number of elements in the intersection bins exceeds $\frac{k\sqrt{k}}{\sqrt{\ell}}$. This happens if the number of collisions between elements of $h(X)$ and $h(Y)$ is very large ($\Omega(\frac{k\sqrt{k}}{\sqrt{\ell}})$). The expected number of collisions

Protocol 13 Protocol π for computing DISJ_n^k.

Require:
$x \in \{0,1\}^n, |x| = k$ - known to Alice
$y \in \{0,1\}^n, |y| = k$ - known to Bob

1: Alice and Bob each sample $k^{3/4}$ elements from their sets and then they run the Håstad-Wigderson protocol (Theorem 2.9.1) $\Omega(\log k)$ times in $O(k^{3/4} \log k)$ communication. If the majority output is 0, the parties output 0. The protocol terminates.
2: Alice and Bob choose a uniformly random hash function $h : [M] \to [\ell]$ and hash the universe into ℓ bins. If Alice and Bob have sets X and Y, they run the protocol π from Lemma 2.9.15 on $h(X)$ and $h(Y)$. If π returns that the sets are disjoint, then they output 1.
3: If π says that the intersection is too large, or if the number of elements in the intersection bins exceeds $\frac{k\sqrt{k}}{\sqrt{\ell}}$, the parties output a random answer and the protocol terminates. Otherwise the players continue.
4: Alice and Bob run the Håstad-Wigderson protocol on each of ℓ bins that have both Alice's and Bob's elements. Players output 1 if they don't find an intersection among any of the bins.

is bounded by $|X \cap Y| + k^2/\ell$. Thus, by the Markov's inequality, the probability of an abort in this step is $O(\sqrt{k/\ell})$ (in fact, due to concentration, it is much lower). Therefore, the protocol succeeds except with probability $o(1)$.

It remains to analyze the communication complexity of this protocol. First step has communication complexity $O(k^{3/4}) = o(k)$. The last step uses at most $O(k\sqrt{k/\ell}) = o(k)$ communication. Finally the bulk of the communication occurs in the middle of the protocol, where by Lemma 2.9.15, the communication is bounded by $\frac{2}{\ln 2}k + o(k)$. □

CHAPTER 3
INFORMATION COMPLEXITY BOUNDS VIA COMMUNICATION COMPLEXITY

The results of this chapter are based on the joint work of the author with Braverman, Garg, and Weinstein and have appeared in [10].

In this chapter we shall study communication complexity and information complexity of several explicit functions. For easy reference, we list these functions here.

Definition 3.0.2. Let $x, y \in \{0,1\}^n$. *The Hamming distance* between x and y, denoted by $\mathrm{HAM}(x,y)$, is defined by:

$$\mathrm{HAM}(x,y) = |\{i \mid x_i \neq y_i\}|.$$

Definition 3.0.3. Let $n, t, g \in \mathbb{N}$. *The Gap Hamming Distance* partial function with respect to parameters n, t, and g, denoted by $\mathrm{GHD}_{n,t,g} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$, is defined by:

$$\mathrm{GHD}_{n,t,g}(x,y) = \begin{cases} 1 & \text{if } \mathrm{HAM}(x,y) \geq t+g \\ 0 & \text{if } \mathrm{HAM}(x,y) \leq t-g \end{cases}$$

If the parameters t and g are omitted, then they are assumed to be $n/2$ and $\sqrt{n}$, respectively. In other words, we write

$$\mathrm{GHD}_n(x,y) = \mathrm{GHD}_{n,n/2,\sqrt{n}}(x,y).$$

Note that the Gap Hamming Distance is a *partial function* (promise problem).

Definition 3.0.4. *The Inner Product* function, denoted by $\mathrm{IP}_n : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$, is defined by:

$$\mathrm{IP}_n(x,y) = \sum_{i=1}^{n} x_i y_i \mod 2.$$

3.1 Introduction

In this chapter we develop a new self-reducibility technique for deriving information complexity lower bounds from communication complexity lower bounds. The technique works for functions that have a "self-reducible structure". Informally speaking f has a self-reducible structure, if for large enough inputs, solving f_{nk} essentially amounts to solving f_n^k (f_{nk} denotes the function f under inputs of length nk, while f_n^k denotes k independent copies of f under inputs of size n). Our departing point is a communication complexity lower bound for f_{nk} (that may be obtained by any means). Assuming self-reducibility, the same bound applies to f_n^k, which through the connection between information complexity and amortized communication complexity (see [12]), implies a lower bound on the information complexity of f_n. In this chapter we show how to make this reasoning go through for two functions: the Gap Hamming Distance, and the Inner Product.

Ideas of self-reducibility are central in applications of information complexity to communication complexity lower bounds, starting with the work of Bar-Yossef et al. [3]. These argument start with an information complexity lower bound for a (usually very simple) problem, and derive a communication complexity bound on many copies of the problem. We saw such an application in Chapter 2. The logic of this chapter is reversed: we start with a communication complexity lower bound, which we use as a black-box, and use self-reducibility to derive an amortized communication complexity bound, which translates into an information complexity lower bound.

3.2 Main Results of this Chapter

We prove that the information complexity of the Gap Hamming Distance problem with respect to the uniform distribution is linear. This was explicitly stated as an open problem by Chakrabarti et al. [18]. We prove

Theorem 3.2.1 ([10])**.** *There exists $\epsilon > 0$ such that*

$$\mathrm{IC}_{\mathcal{U}}(\mathrm{GHD}_n, \epsilon) = \Omega(n),$$

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

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

Theorem 3.2.2 ([10]). *For every $\delta > 0$, there exists $\epsilon > 0$, and n_0 such that $\forall n \geq n_0$ we have*

$$\mathrm{IC}_{\mathcal{U}}(\mathrm{IP}_n, \epsilon) \geq (1 - \delta)n,$$

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

Note that $\mathrm{IC}_{\mathcal{U}}(\mathrm{IP}_n, \epsilon) \leq (1 - 2\epsilon)(n + 1)$, since the parties can always give a random output with probability 2ϵ (this corresponds to error ϵ), and use the brute force protocol to compute IP_n otherwise. It is known that for all $\epsilon \in [0, 1/2)$ we have $\mathrm{IC}_{\mathcal{U}}(\mathrm{IP}_n, \epsilon) = \Omega(n)$ (see [13]). We prove that the information complexity of IP_n can be made arbitrarily close to the trivial upper bound n by decreasing the error (but keeping the error constant).

3.3 Information Complexity of Gap Hamming Distance

In a technical tour-de-force, Chakrabarti and Regev [19] proved that the randomized communication complexity of the Gap Hamming Distance problem is linear. Formally, they showed that

Theorem 3.3.1 ([19]). *For all $\gamma > 0$ and $\epsilon \in [0, 1/2)$ we have*

$$\mathrm{R}(\mathrm{GHD}_{n,n/2,\gamma\sqrt{n}}, \epsilon) \geq \Omega(n).$$

Chakrabarti and Regev [19] also proved the linear lower bound on the distributional communication complexity with respect to the *uniform distribution* $\mathcal{U}$. Specifically, they proved

Theorem 3.3.2 ([19]). *There exists $\epsilon > 0$ such that*

$$\mathrm{D}_{\mathcal{U}}(\mathrm{GHD}_{n,n/2,\sqrt{n}}, \epsilon) = \Omega(n),$$

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

Kerenidis et al. [32] proved that the information complexity of the Gap Hamming Distance is also linear with respect to some implicitly defined distribution. The proof of Kerenidis et al. relies on a reduction that shows that many of the communication complexity lower bound techniques translate to information complexity lower bounds – including the lower bound for the Gap Hamming Distance:

Theorem 3.3.3 ([32])**.** *There exists a distribution μ on $\{0,1\}^n \times \{0,1\}^n$ and $\epsilon > 0$ such that*

$$\mathrm{IC}_\mu(\mathrm{GHD}_{n,n/2,\sqrt{n}}, \epsilon) = \Omega(n).$$

Interestingly, while this approach yields an analogue of Theorem 3.3.1 for information complexity, it does not seem to yield an analogue of the stronger Theorem 3.3.2. In other words this approach does not immediately yield a lower bound on the information complexity of the Gap Hamming Distance with respect to the uniform distribution.

We present an alternative proof of the linear lower bound on the information complexity of GHD using the self-reducibility technique. Unlike the proof in [32], we do not need to dive into the details of the proof of the communication complexity lower bound for GHD. Rather, our starting point is Theorem 3.3.2, which we use as a black-box.

In fact, we shall prove a slightly weaker lemma, from which Theorem 3.2.1 follows via a reduction.

Lemma 3.3.4 ([10])**.** *There exists $\epsilon > 0$ and $\gamma > 0$ such that*

$$\mathrm{IC}_{\mathcal{U}}(\mathrm{GHD}_{n,n/2,\gamma\sqrt{n}}, \epsilon) = \Omega(n),$$

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

The idea of the proof. We use the self-reducibility argument described at the beginning of this chapter. Assume that for some $\epsilon > 0$ we have $\mathrm{IC}_{\mathcal{U}}(\mathrm{GHD}_n, \epsilon) = o(n)$. Using "information = amortized communication" there exists a protocol τ that solves N

copies of GHD_n with $o(nN)$ communication. Using τ we construct a protocol that solves GHD_{nN} with $o(nN)$ communication, which contradicts Theorem 3.3.1. Next, we describe the protocol. Alice and Bob are given $x, y \in \{0,1\}^{nN}$, respectively. They sample cnN random coordinates (for some constant c). Alice and Bob divide the sampled coordinates into cN blocks and run GHD_n on each block using $o(nN)$ communication in total. If $\mathrm{HAM}(x,y) = nN/2 + \sqrt{nN}$, then the expected Hamming distance of the inputs restricted to each block is $n/2 + \sqrt{n/N}$. Although the gain over $n/2$ is small, the Hamming distance is still biased towards being $> n/2$. We shall see that on each block, the protocol for GHD_n must gain an advantage of $\Omega(1/\sqrt{N})$ over random guessing. This in turn implies that cN copies suffice to get the correct answer with high probability.

3.3.1 Information Complexity of Small-Gap Instances

Assume that for some sufficiently small $\rho > 0$ (to be specified later) we have

$$\mathrm{IC}_{\mathcal{U}}(\mathrm{GHD}_{n,n/2,\sqrt{n}}, \rho) = o(n).$$

Thus $\forall \alpha > 0$ and for sufficiently large n we have

$$\mathrm{IC}_{\mathcal{U}}(\mathrm{GHD}_{n,n/2,\sqrt{n}}, \rho) \leq \alpha n.$$

We shall need the following theorem from [12]:

Theorem 3.3.5. *[12] Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a (possibly partial) function, let μ be a distribution on $\mathcal{X} \times \mathcal{Y}$, let $\rho > 0$, and let $I = \mathrm{IC}_\mu(f, \rho)$. Then for each $\delta_1, \delta_2 > 0$ there exists an $N = N(f, \rho, \mu, \delta_1, \delta_2)$ such that for each $n \geq N$, there is a protocol π_n for computing n instances of f over μ^n such that the error probability for each copy is $\leq \rho$. The communication cost of the protocol is $< n(1+\delta_1)I$. Moreover, if we let π be any protocol for computing f with information cost $\leq (1+\delta_1/3)I$ with respect to μ then we can design π_n so that for each set of inputs, the statistical distance between the output of π_n and π^n is $< \delta_2$, where π^n denotes n independent executions of π.*

In other words, Theorem 3.3.5 allows us to take a low-information protocol for f and turn it into a low-communication protocol for sufficiently many copies of f.

Step 1: From GHD to a tiny advantage.

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

Lemma 3.3.6 ([10])**.** *There exist $\tau > 0$, $\gamma > 0$ and $\rho > 0$ with the following property. Suppose that for all large enough n there exists a protocol π_n such that π_n solves $\mathrm{GHD}_{n,n/2,\gamma\sqrt{n}}$ with error ρ with respect to the uniform distribution. Then for all large enough n and for all $\epsilon < 1/n^2$ we have*

$$P_{(X,Y)\sim\mu_\epsilon}(\pi_n(X,Y)=1) - P_{(X,Y)\sim\mu_0}(\pi_n(X,Y)=1) > \tau\epsilon\sqrt{n}, \tag{3.1}$$

and

$$P_{(X,Y)\sim\mu_{-\epsilon}}(\pi_n(X,Y)=0) - P_{(X,Y)\sim\mu_0}(\pi_n(X,Y)=0) > \tau\epsilon\sqrt{n}. \tag{3.2}$$

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

We will establish (3.1), with (3.2) established similarly. We first focus on the region where $\mathrm{HAM}(x,y) \geq n/2$ and show that its contribution to (3.1) is at least $\Omega(\epsilon\sqrt{n})$. We break the region into two further subregions: (I) (x,y) with $n/2 < \mathrm{HAM}(x,y) < n/2 + \gamma\sqrt{n}$; (II) (x,y) with $n/2 + \gamma\sqrt{n} \leq \mathrm{HAM}(x,y)$ for appropriately chosen γ. We show that the contribution of region (II) is $\Omega(\epsilon\sqrt{n})$, while the fact that the contribution of region (I) being positive is easy to see.

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

$$\begin{aligned} &p_i(P_{\mu_\epsilon}(\mathrm{HAM}(X,Y) = n/2+i) - P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2+i)) \\ &= p_i P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2+i)\left((1-4\epsilon^2)^{n/2-i}(1+2\epsilon)^{2i} - 1\right). \end{aligned}$$

Now $(1-4\epsilon^2)^{n/2-i} \geq 1 - 2\epsilon/n$ and $(1+2\epsilon)^{2i} \leq e^2$ (since $\epsilon < 1/n^2$). Thus, we have

$$\sum_{i=0}^{n/2} p_i P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2+i)(2\epsilon/n)(1+2\epsilon)^{2i} \leq O(\epsilon/n).$$

Therefore, we can ignore the term $(1-4\epsilon^2)^{n/2-i}$ since then

$$\begin{aligned} &\sum_{i=0}^{n/2} p_i P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2+i)\left((1-4\epsilon^2)^{n/2-i}(1+2\epsilon)^{2i} - 1\right) \\ &- \sum_{i=0}^{n/2} p_i P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2+i)\left((1+2\epsilon)^{2i} - 1\right) \leq O(\epsilon/n) \end{aligned}$$

After ignoring the term, the contribution in region (I) is positive .

This leaves us with region (II), where we need to show that we actually get a non-negligible advantage. Let T be an appropriately chosen constant, so that $P_{\mu_0}(\sqrt{n} \leq$

$\mathrm{HAM}(X,Y) - n/2 \leq T\sqrt{n}) = \Omega(1)$. The advantage

$$\begin{aligned}
&\sum_{i=\gamma\sqrt{n}}^{n/2} p_i P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2+i)\left((1+2\epsilon)^{2i}-1\right) \\
&\geq \sum_{i=\gamma\sqrt{n}}^{T\sqrt{n}} p_i P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2+i) 4i\epsilon \\
&= \sum_{i=\gamma\sqrt{n}}^{T\sqrt{n}} P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2+i) 4i\epsilon \\
&- \sum_{i=\gamma\sqrt{n}}^{T\sqrt{n}} (1-p_i) P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2+i) 4i\epsilon \\
&\geq \Theta(\epsilon\sqrt{n}) - 4T\rho\epsilon\sqrt{n}
\end{aligned}$$

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

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

$$\begin{aligned}
&q_i(P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2-i) - P_{\mu_\epsilon}(\mathrm{HAM}(X,Y) = n/2-i)) \\
&= q_i P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2-i)(1-(1-4\epsilon^2)^{n/2-i}(1-2\epsilon)^{2i})
\end{aligned}$$

As before, we can ignore the term $(1-4\epsilon^2)^{n/2-i}$. In region (I) the negative contribution

is bounded in absolute terms by:

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

In region (III) the contribution is again bounded by

$$\begin{aligned}
&\sum_{i=T\sqrt{n}}^{n/2} P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2 - i)(1 - (1-2\epsilon)^{2i}) \\
&< \sum_{i=T\sqrt{n}}^{n/2} P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2 - i)4i\epsilon.
\end{aligned}$$

By the Chernoff bound, the probability $P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2 - i)$ is dominated by $e^{-\Omega(i^2/n)}$, and thus the sum can be made into an arbitrarily small multiple of $\epsilon\sqrt{n}$ by choosing T large enough. For region (II) the advantage

$$\begin{aligned}
&\sum_{i=\gamma\sqrt{n}}^{T\sqrt{n}} q_i P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2 - i)(1 - (1-2\epsilon)^{2i}) \\
&\leq \sum_{i=\gamma\sqrt{n}}^{T\sqrt{n}} q_i P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2 - i)4i\epsilon \\
&\leq 4T\epsilon\sqrt{n} \sum_{i=\gamma\sqrt{n}}^{T\sqrt{n}} q_i P_{\mu_0}(\mathrm{HAM}(X,Y) = n/2 - i) \\
&\leq 4T\rho\epsilon\sqrt{n}.
\end{aligned}$$

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

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

We can now apply Lemma 3.3.6 together with Theorem 3.3.5 to show that existence of a low-information protocol for $\mathrm{GHD}_{n,n/2,\gamma\sqrt{n}}$ with respect to the uniform distribution contradicts the communication complexity lower bound of Theorem 3.3.2.

Proof of Lemma 3.3.4. Assume for the sake of contradiction that for each α there

exists n and a protocol π_n with $\mathrm{IC}_{\mathcal{U}}(\pi_n) < \alpha n$ and which solves $\mathrm{GHD}_{n,n/2,\gamma\sqrt{n}}$ with error ρ, where the parameters γ and ρ are from Lemma 3.3.6. Let

$$N > \max(n^7, N(\mathrm{GHD}_{n,n/2,\gamma\sqrt{n}}, \rho, \mathcal{U}, \delta_1, \delta_2)),$$

where $\delta_1 = 1$ and $\delta_2 = \epsilon/2$, where ϵ is the error parameter in Theorem 3.3.2. Then using Theorem 3.3.5, for each $c > 1$, cN copies of π_n can be executed with communication $< 2\alpha cnN$ (as long as the inputs to each π_n are distributed according to $\mathcal{U}$) such that on each copy the error is at most ρ with respect to $\mathcal{U}$. Also for each set of inputs, the statistical distance between the output of the execution and $\pi_n^{cN} \leq \epsilon/2$.

Let $t = P_{(X,Y)\sim\mathcal{U}}(\pi_n(X,Y) = 1)$. Without loss of generality, we assume $t = 1/2$ (otherwise we can use a $\mathrm{threshold}_{tcN}$ instead of majority in the protocol). We solve $\mathrm{GHD}_{nN,nN/2,\sqrt{nN}}$ over the uniform distribution with a small constant error ϵ using the protocol depicted in Protocol 14.

Protocol 14 The protocol $\pi_{nN}(x, y)$

Require:

$x \in \{0,1\}^{nN}$ - known to Alice
$y \in \{0,1\}^{nN}$ - known to Bob

1: Players create cN instances of GHD_n by sampling n random coordinates each time (with replacement) using public randomness: $(x_1, y_1), \ldots, (x_{cN}, y_{cN}) \in \{0,1\}^n \times \{0,1\}^n$.
2: Players use compression (Theorem 3.3.5) to run $\pi_n(x_1, y_1), \ldots, \pi_n(x_{cN}, y_{cN})$ in communication $2\alpha cNn$.
3: Players return $\mathrm{MAJORITY}(\pi_n(x_1, y_1), \ldots, \pi_n(x_{cN}, y_{cN}))$.

The communication cost upper bound follows from the way the protocol $\pi_{nN}(x, y)$ is constructed. To finish the proof we need to analyze its success probability. Suppose that the Hamming distance between x and y is $nN/2 + \ell\sqrt{nN}$, where $\ell > 1$. Note that $\ell < n$ except with probability $e^{-\Omega(n^2)}$. The samples (x_i, y_i) are drawn iid according to the distribution $\mu_{\ell\sqrt{1/(nN)}}$. Since $N > n^7$ we have $\ell\sqrt{1/nN} < 1/n^2$. By Lemma 3.3.6, the output of π_n on each copy is thus $\tau\ell/\sqrt{N}$-biased towards 1. An application of the Chernoff bound along with the fact that, for each set of inputs, the statistical distance between the output of the execution and $\pi_n^{cN} \leq \epsilon/2$, implies that the probability that

the protocol π_{nN} outputs 1 is at least $1 - e^{-2\tau^2\ell^2 c} - \epsilon/2$. For constant τ, we can make this expression as close to $1 - \epsilon/2$ as we like by letting c be a sufficiently large constant. But this means that for an arbitrarily small constant $\alpha > 0$, $\pi_{nN}(x, y)$ will solve $\mathrm{GHD}_{nN,nN/2,\sqrt{nN}}$ with error $\leq \epsilon$ (the case when the Hamming distance between x and y is $nN/2 - \ell\sqrt{nN}$ is symmetric) in communication $O(\alpha cNn)$, which can be made arbitrarily small relatively to Nn, leading to a contradiction. Note that we got a randomized protocol for solving $\mathrm{GHD}_{nN,nN/2,\sqrt{nN}}$ but we can fix the randomness to get a deterministic algorithm. □

3.3.2 The Reduction from a Small-Gap instance to a Large-Gap instance

Now we complete the proof of Theorem 3.2.1 by providing the details of the reduction. We will start by proving a few technical lemmas.

Lemma 3.3.7 ([10]). *Let $\alpha > 1$ be an integer. Let $\mathcal{U}_n$ be the uniform distribution over $\{0,1\}^n \times \{0,1\}^n$. Let $(X, Y) \sim \mathcal{U}_n$. Define a distribution μ over $\{0,1\}^{\alpha n} \times \{0,1\}^{\alpha n}$ by picking αn random coordinates of X, Y (with replacement) and then taking* XOR *with a random string $r \in_R \{0,1\}^{\alpha n}$ (let U', V' be the strings obtained by sampling αn random coordinates of X, Y. Then $U = U' \oplus r, V = V' \oplus r$ are the final strings sampled). Then for all $\epsilon > 0$ and n large enough, there exists a constant M_ϵ and a distribution μ_ϵ such that*

1. $|\mu - \mu_\epsilon| \leq \epsilon$
2. $\mu_\epsilon \leq M_\epsilon \mathcal{U}_{\alpha n}$

Proof. It is easy to see that the distribution μ is symmetric with respect to the Hamming distance i.e., if $(x, y) \in \{0,1\}^{\alpha n} \times \{0,1\}^{\alpha n}$, and $(x', y') \in \{0,1\}^{\alpha n} \times \{0,1\}^{\alpha n}$ such that $\mathrm{HAM}(x, y) = \mathrm{HAM}(x', y')$, then $\mu(x, y) = \mu(x', y')$. This is because μ is invariant under the application of a random permutation and a random XOR i.e., if $\sigma \in_R S_n$ and $r' \in_R \{0,1\}^n$, then $\mu(x, y) = \mu(\sigma(x \oplus r'), \sigma(y \oplus r'))$. With a slight abuse of notation let $\mu(d)$ denote the probability mass on strings of Hamming distance d,

and let $\mathcal{U}_{\alpha n}(d)$ denote the probability mass with respect to the uniform distribution. Let $N = \alpha n$.

The distribution μ_ϵ will be the restriction of the distribution μ to the interval $[N/2 - C\sqrt{N}, N/2 + C\sqrt{N}]$. Let T be the random variable that equals to $\mathrm{HAM}(X, Y)$. Then by the Chernoff bound we have

$$P(T \notin [n/2 - \beta\sqrt{n}, n/2 + \beta\sqrt{n}]) \leq 2e^{-2\beta^2}.$$

If we pick N random coordinates distributed according to $B_{\frac{1}{2}+p}$, where $|p| \leq \beta/\sqrt{n}$, then the expected number of 1's $\in [N/2 - \beta\sqrt{\alpha}\sqrt{N}, N/2 + \beta\sqrt{\alpha}\sqrt{N}]$. Thus by another application of the Chernoff bound and taking C large enough, we can make the statistical distance between μ_ϵ and μ small enough.

Let μ' be the distribution μ restricted to the interval $[N/2 - C\sqrt{N}, N/2 + C\sqrt{N}]$ for some constant C with a slight scaling (it is easy to see that the scaling will be at most 2 for large enough C), which we can ignore. We will show that there exists a constant M such that $\mu' \leq M\mathcal{U}_{\alpha n}$. By the symmetry properties of μ, it suffices to prove that for all d, $\mu'(d) \leq M\mathcal{U}_{\alpha n}(d)$. We have

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

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

$k = 3n/4$ to $k = n$ is small. So we consider

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

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

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

Now $\sum_{k=n/2+T}^{3n/4} \binom{n}{k} \left(1 + \frac{8T}{n}\right)^{k} \left(1 - \frac{8T}{n}\right)^{n-k} \leq 2^n$ by the binomial theorem. Since $T \leq C\sqrt{N}$ for some constant C, the following quantity is a constant, too:

$$\left(1 + \frac{8T}{n}\right)^{-n/2} \left(1 - \frac{8T}{n}\right)^{n/2} = \left(1 - \frac{64T^2}{n^2}\right)^{-n/2}.$$

This completes the proof. □

We need a lemma that relates the information cost of a protocol with respect to one distribution to the information cost of the same protocol with respect to a distribution subsuming the first distribution. Formally, the statement of the lemma

is as follows:

Lemma 3.3.8 ([10]). *Let μ_1 and μ_2 be distributions over $\{0,1\}^N \times \{0,1\}^N$ such that $\mu_1 \leq M\mu_2$ for some constant M. Let f be a function (possibly partial) with domain $\{0,1\}^N \times \{0,1\}^N$ and let π be a protocol for solving it. Then $\mathrm{IC}_{\mu_1}(\pi) \leq M\,\mathrm{IC}_{\mu_2}(\pi)$.*

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

$$I(\Pi_1; X_1|Y_1) = \mathbb{E}_{(X,Y)\sim\mu_1}(\mathbb{D}(\Pi_1|_{(X,Y)}||\Pi_1|_Y)) = \mathbb{E}_Y(\mathbb{E}_X(\mathbb{D}(\Pi_1|_{(X,Y)}||\Pi_1|_Y)))$$

By Fact 1.4.6, $\mathbb{E}_X(\mathbb{D}(\Pi_1|_{X,Y}||\Pi_1|_Y)) \leq \mathbb{E}_X(\mathbb{D}(\Pi_1|_{X,Y}||\Pi_2|_Y))$. Also $\Pi_1|_{X,Y} = \Pi_2|_{X,Y}$. Thus

$$\begin{aligned} I(\Pi_1; X_1|Y_1) \leq \mathbb{E}_{(X,Y)\sim\mu_1}(\mathbb{D}(\Pi_2|_{X,Y}||\Pi_2|_Y)) &\leq M\mathbb{E}_{(X,Y)\sim\mu_2}(\mathbb{D}(\Pi_2|_{X,Y}||\Pi_2|_Y) \\ &= MI(\Pi_2; X_2|Y_2) \end{aligned}$$

Hence $\mathrm{IC}_{\mu_1}(\pi) \leq M\,\mathrm{IC}_{\mu_2}(\pi)$. □

The next lemma says that if the information complexity of a function with respect to the distribution μ from Lemma 3.3.7 is high, then the information complexity with respect to the uniform distribution is high as well.

Lemma 3.3.9 ([10]). *Let $f : \{0,1\}^N \times \{0,1\}^N \to \{0,1\}$ be a function (possibly partial). Let μ be a distribution over $\{0,1\}^N \times \{0,1\}^N$, as defined in Lemma 3.3.7. If $\mathrm{IC}_\mu(f, \delta) = \Omega(N)$, for some $\delta > 0$, then $\mathrm{IC}_{\mathcal{U}_N}(f, \eta) = \Omega(N)$, for some $\eta > 0$.*

Proof. Let π be a protocol for computing f with error η with respect to the distribution $\mathcal{U}_N$, and let $I = \mathrm{IC}_{\mathcal{U}_N}(\pi)$. Let $\epsilon > 0$. Then by Lemma 3.3.7, for N large enough there exists a distribution μ_ϵ over $\{0,1\}^N \times \{0,1\}^N$ such that $|\mu - \mu_\epsilon| \leq \epsilon$ and $\mu_\epsilon \leq M_\epsilon \mathcal{U}_N$ for some constant M_ϵ. Then the error probability of the protocol π with respect to μ is $\leq M_\epsilon\eta + \epsilon$. Also the information cost of π with respect to μ is $\leq M_\epsilon I + 5N\epsilon$ (by Lemmas 2.4.6 and 3.3.8). If $M_\epsilon\eta + \epsilon \leq \delta$ then $M_\epsilon I + 5N\epsilon \geq cN$ for some constant c. Take $\epsilon = \min(\delta/2, c/10)$ and $\eta = (\delta - \epsilon)/M_\epsilon$. Then $I \geq cN/2M_\epsilon$. Thus, we have $\mathrm{IC}_{\mathcal{U}_N}(f, \eta) = \Omega(N)$. □

Proof of Theorem 3.2.1. Note that because of Lemma 3.3.9, we just need to prove that $\mathrm{IC}_\mu(\mathrm{GHD}_N, \epsilon) = \Omega(N)$ for some $\epsilon > 0$ for the distribution μ in Lemma 3.3.7. Assume that for all $\epsilon > 0$ we have $\mathrm{IC}_\mu(\mathrm{GHD}_N, \epsilon) = o(N)$. That is for all β, ϵ, and for N sufficiently large, $\mathrm{IC}_\mu(\mathrm{GHD}_N, \epsilon) \leq \beta N$. By Lemma 3.3.4, there exist constants $\epsilon' > 0$, $\gamma > 0$ and $c > 0$ such that $\mathrm{IC}_{\mathcal{U}_n}(\mathrm{GHD}_{n,n/2,\gamma\sqrt{n}}, \epsilon') \geq cn$.

Let α be a large enough integer to be determined later. Set $N = \alpha n$. Let π_N be a protocol that solves GHD_N with error $\leq \epsilon$ with respect to μ, and let the information cost of π_N with respect to μ be $\leq \beta N$. Consider the following protocol $\pi_n(x, y)$ for $\mathrm{GHD}_{n,n/2,\gamma\sqrt{n}}$: players pick N random coordinates of x, y, call them u', v'. Players pick a random string $r \in_R \{0,1\}^N$ and set $u = u' \oplus r$ and $v = v' \oplus r$. Players run π_N on u, v. Let $(X, Y) \sim \mathcal{U}_n$ be the inputs for π_n. Let U, V denote the random variables denoting the sampled coordinates. Note that $(U, V) \sim \mu$. Let Π denote the random variable for the transcript of running π_N on U, V. Then the transcript of running π_n on X, Y is ΠR, where R denotes the public randomness involved in sampling u, v from x, y. Now

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

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

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

and $I(\Pi; YR|UV) = 0$, $I(\Pi; X|UVYR) = 0$. Applying Fact 1.4.4, with $A = \Pi$, $B = U$, $C = X$ and $D = VYR$, we get that $I(\Pi; X|VYR) \leq I(\Pi; U|VYR)$. Also $I(\Pi; YR|UV) = 0$. Applying Fact 1.4.3 with $A = U$, $B = \Pi$, $C = V$ and $D = YR$, we get $I(\Pi; U|V) \geq I(\Pi; U|VYR)$. This implies that $I(\Pi R; X|Y) \leq I(\Pi; U|V)$. A similar argument shows that $I(\Pi R; Y|X) \leq I(\Pi; V|U)$ and hence $\mathrm{IC}_{\mathcal{U}_n}(\pi_n) \leq \mathrm{IC}_\mu(\pi_N)$.

Let us calculate the error probability of the protocol π_n. If $\mathrm{HAM}(x, y) \geq n/2 +$

$\gamma\sqrt{n}$, then for a random coordinate ℓ we have $P(x_\ell \oplus y_\ell = 1) \geq 1/2 + \gamma/\sqrt{n}$. Then the expected Hamming distance of N random coordinates is $N/2 + \gamma\sqrt{\alpha}\sqrt{N}$. Probability that the Hamming distance is $\leq N/2 + \frac{\gamma\sqrt{\alpha}}{2}\sqrt{N}$ is bounded by $e^{-\frac{\alpha\gamma^2}{2}}$. Similarly for the lower case. Choose α so that $\gamma\sqrt{\alpha} \geq 2$ and $e^{-\frac{\alpha\gamma^2}{2}} \leq \epsilon'/2$. Then

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

We have

$$P(\pi_n \text{ outputs } 0 \text{ on input } x,y) = \sum_{u,v} \mu(u,v|x,y) P(\pi_N \text{ outputs } 0 \text{ on input } u,v),$$

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

$$\begin{aligned}&\sum_{u,v} \mu(u,v|x,y) P(\pi_N \text{ outputs } 0 \text{ on input } u,v) \\ &\leq \sum_{u,v:\text{HAM}(u,v)\geq N/2+\sqrt{N}} \mu(u,v|x,y) P(\pi_N \text{ outputs } 0 \text{ on input } u,v) + \epsilon'/2\end{aligned}$$

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

$$\begin{aligned}\text{error}(\pi_n) \leq &\sum_{u,v:\text{HAM}(u,v)\geq N/2+\sqrt{N}} \mu(u,v) P(\pi_N \text{ outputs } 0 \text{ on input } u,v) \\ &+ \sum_{u,v:\text{HAM}(u,v)\leq N/2-\sqrt{N}} \mu(u,v) P(\pi_N \text{ outputs } 1 \text{ on input } u,v) + \epsilon'/2 \\ &= \text{error}(\pi_N) + \epsilon'/2\end{aligned}$$

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

3.4 Information Complexity of Inner Product

The proof of Theorem 3.2.2 exploits the self-reducible structure of the Inner Product function. Since IP_n is a highly sensitive function, we shall first prove a lower bound on its 0-error information complexity. We shall use the continuity of information complexity at 0 error to to finish the argument regarding the ϵ-error information complexity of IP_n.

We shall need the following lemma from [12]. It is similar to Theorem 3.3.5. The difference is that when dealing with 0 error we cannot ensure that the probability of error on each copy is 0. We just control the overall error, which is the error incurred if the compression fails.

Lemma 3.4.1 ([12])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a function, and let μ be a distribution over the input space $\mathcal{X} \times \mathcal{Y}$. Let π be a protocol that computes f with 0 error with respect to μ. Let $I = \mathrm{IC}_\mu(\pi)$. Then for all $\delta, \epsilon > 0$ there exists a protocol π_n for computing f^n with error ϵ with respect to μ^n such that the worst case communication cost of π is*

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

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

Lemma 3.4.2 ([4])**.** *Let $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ be a function, and let μ be a distribution over the input space $\mathcal{X} \times \mathcal{Y}$. Then we have*

$$\mathrm{IC}_{\mu^n}(\oplus_n f, \epsilon) \geq n(\mathrm{IC}_\mu(f, \epsilon) - 2)$$

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

Lemma 3.4.3 ([10])**.** *For all n we have*

$$\mathrm{IC}_{\mathcal{U}_n}(\mathrm{IP}_n, 0) \geq n,$$

where $\mathcal{U}_n$ is the uniform distribution over $\{0,1\}^n \times \{0,1\}^n$

Proof. It is known that $\mathrm{D}_{\mathcal{U}_n}(\mathrm{IP}_n, \epsilon) \geq n - c_\epsilon$, for all constant $\epsilon \in (0, 1/2)$, where c_ϵ is a constant depending just on ϵ [33, 22]. Assume that for some n we have $\mathrm{IC}_{\mathcal{U}_n}(\mathrm{IP}_n, 0) \leq n - c$. Then using Lemma 3.4.1 with $\delta = c/2$ and $\epsilon = 1/3$, we can get a protocol π for solving N copies of IP_n with overall error $1/3$ with respect to $\mathcal{U}_n^N$, and $\mathrm{CC}(\pi) \leq N(n - c + c/2)$. This gives us a protocol π' for solving IP_{Nn} with error $1/3$ with respect to the uniform distribution and $\mathrm{CC}(\pi') \leq Nn - Nc/2$ (divide the inputs into N chunks, solve the N chunks using π and XOR the answers). But $\mathrm{CC}(\pi') \geq Nn - c_{1/3}$, a contradiction. □

Proof of Theorem 3.2.2. Given $\delta > 0$, let $\ell = \lceil \frac{3}{\delta} \rceil$. Then

$$\mathrm{IC}_{\mathcal{U}_\ell}(\mathrm{IP}_\ell, 0) \geq \ell \geq (1 - \delta)\ell + 3$$

By Theorem 2.2.8, we have $\lim_{\epsilon \to 0} \mathrm{IC}_{\mathcal{U}_\ell}(\mathrm{IP}_\ell, \epsilon) = \mathrm{IC}_{\mathcal{U}_\ell}(\mathrm{IP}_\ell, 0)$. Thus, there exists $\epsilon(\ell, \delta) = \epsilon(\delta)$ such that

$$\mathrm{IC}_{\mathcal{U}_\ell}(\mathrm{IP}_\ell, \epsilon) \geq (1 - \delta)\ell + 2$$

Now using Lemma 3.4.2, we get that $\mathrm{IC}_{\mathcal{U}_\ell^N}(\oplus_N \mathrm{IP}_\ell, \epsilon) \geq (1 - \delta)N\ell$. Thus we have

$$\mathrm{IC}_{\mathcal{U}_{N\ell}}(\mathrm{IP}_{N\ell}, \epsilon) \geq (1 - \delta)N\ell.$$

Thus for sufficiently large n we have $\mathrm{IC}_{\mathcal{U}_n}(\mathrm{IP}_n, \epsilon) \geq (1 - \delta)n$. □

CHAPTER 4
PUBLIC VS. PRIVATE RANDOMNESS IN INFORMATION COMPLEXITY

4.1 Introduction

Based on the definitions given in Section 1.3 it is easy to see that for every $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ and for every $\epsilon \geq 0$ we have

$$\mathrm{R}(f,\epsilon) \leq \mathrm{R}^{\mathrm{priv}}(f,\epsilon).$$

Ilan Newman [40] showed that the reverse inequality holds up to constant multiplicative factors and a logarithmic additive term. More precisely, he showed

Theorem 4.1.1 (Newman [40])**.** $\forall f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}, \forall \epsilon \in [0,1/2), \forall \delta \in (0,1]$, *we have*

$$\mathrm{R}^{priv}(f,(1+\delta)\epsilon) = O\left(\mathrm{R}(f,\epsilon) + \log \frac{n}{\epsilon\delta}\right).$$

In other words, for the purpose of minimizing communication complexity it is always beneficial to have public randomness instead of private randomness; however, asymptotic improvements in communication are only possible for functions of sublogarithmic complexity. For the rest of the discussion, we shall need the following definition.

Definition 4.1.1. Let π be a protocol with public and private randomness. Define π_r to be the protocol obtained from π by fixing public randomness to r. Note that π_r is a protocol that uses only private randomness.

The value of public vs. private randomness for the purpose of minimizing information complexity is reversed. More specifically, for the purpose of minimizing information complexity it is always beneficial to have private randomness. Formally, we have the following folklore fact.

Fact 4.1.2. *For all functions* $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$*, distributions* μ *on* $\{0,1\}^n \times \{0,1\}^n$*, and* $\epsilon \geq 0$*, we have*

$$\mathrm{IC}_\mu^{priv}(f,\epsilon) = \mathrm{IC}_\mu(f,\epsilon).$$

Proof. The inequality $\mathrm{IC}_\mu^{\mathrm{priv}}(f,\epsilon) \geq \mathrm{IC}_\mu(f,\epsilon)$ is clear from the definition, since the infimum on the right hand side is over a larger set of protocols. Now, let π be a protocol nearly achieving $\mathrm{IC}_\mu(f,\epsilon)$. The protocol π might use publicly-known random string R. Now, let π' be the following protocol that uses only private randomness: Alice uses her private randomness to sample R, then Alice sends R to Bob, and both players run π_R in such a way that Alice never reuses her private random bits that were used to generate R. Clearly, the first message of Alice carries 0 information about her input in π'. Therefore $\mathrm{IC}_\mu(\pi') = \mathrm{IC}_\mu(\pi)$. This establishes $\mathrm{IC}_\mu^{\mathrm{priv}}(f,\epsilon) \leq \mathrm{IC}_\mu(f,\epsilon)$, completing the proof. □

This establishes that $\mathrm{IC}_\mu^{\mathrm{priv}}(f,\epsilon)$ is the same as $\mathrm{IC}_\mu(f,\epsilon)$. A natural question is how close $\mathrm{IC}_\mu^{\mathrm{pub}}(f,\epsilon)$ is to $\mathrm{IC}_\mu(f,\epsilon)$? In general, the answer to this question is not known. The special case of 1-round protocols was answered by Braverman and Garg [8]. They proved that $\mathrm{IC}_\mu^{1,\mathrm{pub}}(f,\epsilon) \leq \mathrm{IC}_\mu^1(f,\epsilon) + \log \mathrm{IC}_\mu^1(f,\epsilon) + O(1)$. The general form of the question was considered in [44]. We showed that if $\mathrm{IC}_\mu^{\mathrm{pub}}(f,\epsilon)$ is close to $\mathrm{IC}_\mu(f,\epsilon)$ then a strong compression of communication to information follows. This result was independently discovered in [15]. In the rest of this chapter we prove this result. We shall need the following definition and fact.

Definition 4.1.2. Let π be a protocol with public and private randomness. Define π_r to be the protocol obtained from π by fixing public randomness to r.

Fact 4.1.3 (Barack et al. [4])**.** *Let* π *be a protocol with public and private randomness on inputs* $\{0,1\}^n \times \{0,1\}^n$*. Let* $\mu \in \Delta(\{0,1\}^n \times \{0,1\}^n)$*. Then*

$$\mathrm{IC}_\mu(\pi) = \mathbb{E}_r(\mathrm{IC}_\mu(\pi_r)),$$

where the expectation is over public randomness r*.*

Proof. For the purpose of this proof we shall write out public randomness R explicitly in the definition of information cost. Consider the first term in the definition of $\mathrm{IC}_\mu(\pi)$.

$$\begin{aligned} I(X; R\Pi_R(X,Y)|Y) &= H(R\Pi_R(X,Y)|Y) - H(R\Pi_R(X,Y)|YX) \\ &= H(R|Y) + H(\Pi_R(X,Y)|YR) - \\ &\quad -H(R|YX) - H(\Pi_R(X,Y)|YXR) \\ &= I(R; X|Y) + I(\Pi_R(X,Y); X|YR) \\ &= 0 + I(\Pi_R(X,Y); X|YR). \end{aligned}$$

Where $I(R; X|Y) = 0$ because R and X are independent. Similarly, we obtain $I(Y; R\Pi_R(X,Y)|X) = I(\Pi_R(X,Y); Y|XR)$.
Finally, we have $\mathrm{IC}_\mu(\pi) = I(\Pi_R(X,Y); X|YR) + I(\Pi_R(X,Y); Y|XR) = \mathbb{E}_r(\mathrm{IC}_\mu(\pi_r))$.
□

4.2 Efficient Simulation of Private Randomness with Public Randomness Leads to Strong Compression

In this section we prove the following lemma and investigate its consequences.

Lemma 4.2.1. *Letl π be a protocol on inputs from $\{0,1\}^n \times \{0,1\}^n$ such that π uses public randomness (R) only (no private randomness). Let $\mu \in \Delta(\{0,1\}^n \times \{0,1\}^n)$, $\gamma \in (0, 1/2)$ and $k \in \mathbb{N}$. Then there exists a protocol $\tau_{k,\gamma}$ with public randomness only and an event E such that*

- $\mathrm{CC}(\tau_{k,\gamma}) = O\left(k \log \frac{\mathrm{CC}(\pi)}{\gamma}\right)$
- $P_{(X,Y)\sim\mu, R\sim\mathcal{R}}(\mathcal{E}) \leq \frac{\mathrm{IC}_\mu(\pi)}{k} + k\gamma$
- *conditioned on $\neg\mathcal{E}$ and (x,y) Alice and Bob both output a transcript $\sim \Pi(x,y)$*

Overview of the proof. The proof of Lemma 4.2.1 proceeds by applying the compression scheme of [4] to a protocol π with public randomness only. In the compressed protocol, the players try to guess a transcript of π without communication, and then communicate to verify their guess. In most cases, their guess is likely to have a mistake. Fortunately, the players can reuse the part of their previous guess that did not

contain a mistake to construct a new guess. The crucial observation is that the expected number of times that the players need to correct their guesses in order to create a correct transcript of π is $\mathrm{IC}_\mu(\pi)$. This bound is only known to hold for protocols without private randomness. Before we describe the formal proof of Lemma 4.2.1, we mention a useful definition and a result that will be used for fixing players' guesses.

Definition 4.2.1. *The first difference* function $\mathrm{FDIFF}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1,\ldots,n\}$ is defined as follows:

$$\mathrm{FDIFF}_n(x,y) = \begin{cases} \min\{i \mid x_i \neq y_i\} & \text{if } x \neq y, \\ 0 & \text{otherwise.} \end{cases}$$

Feige et al [24] showed that the first difference function can be computed with logarithmic communication complexity. Feige et al [24] do not state this result in terms of communication complexity, so we say that it appears there implicitly.

Lemma 4.2.2 (Feige et al [24] (implicit))**.**

$$\mathrm{R}(\mathrm{FDIFF}_n, \epsilon) = O\left(\log(n/\epsilon)\right).$$

Proof of Lemma 4.2.1. For each fixing of public randomness r the protocol π_r is deterministic. Recall that π_r can be viewed as a complete binary tree T_r with additional structure (see Section 1.3 and Algorithm 1). In particular, we associate the following objects with each node u of T_r:

1. Owner of u, i.e., either Alice or Bob.
2. $p_u : \{0,1\}^n \to \{0,1\}$ - the function of owner's input specifying the bit to be transmitted by the owner of u upon reaching u.

The above tree with the entire additional structure is known to both players; however, the players do not know the entire input. Therefore, in general only the owner of u knows the true value of p_u. In general, the non owner of u only has a guess about the value of p_u. This guess is based on the joint distribution μ from which the inputs to the players were sampled. In particular, if μ is such that X and Y are highly

correlated, the non owner of u can simply evaluate p_u on their input and get a good estimate on the true value of p_u. We shall use $\tilde{p}_u$ to denote the non owner's belief that p_u evaluates to 1. More precisely,

$$\begin{aligned}\tilde{p}_u(x) &= P_{Y\sim\mu_x}(p_u(Y)=1|\text{reached } u, X=x) \quad \text{if Bob is owner of } u,\\ \tilde{p}_u(y) &= P_{X\sim\mu_y}(p_u(X)=1|\text{reached } u, Y=y) \quad \text{if Alice is owner of } u.\end{aligned}$$

With each path $v = v_0, v_1, \ldots, v_k$ in T_r we associate a binary string $\langle v\rangle \in \{0,1\}^k$ in a natural way: for all $i \geq 1$ $\langle v\rangle_i$ is 0 if v_i is a left child of v_{i-1} and 1 otherwise. Algorithm 15 shows how to construct a protocol τ_γ (that depends on the parameter $\gamma > 0$) with communication cost similar to the information cost of π.

Define $\tau_{k,\gamma}$ to be the protocol obtained from τ_γ by restricting the number of iterations of the while loop on line 4 to at most k times. By Lemma 4.2.2 we have

$$\mathrm{CC}(\tau_{k,\gamma}) = O\left(k\log\frac{\mathrm{CC}(\pi)}{\gamma}+k\right) = O\left(k\log\frac{\mathrm{CC}(\pi)}{\gamma}\right).$$

It is left to analyze the simulation properties of Algorithm 15. Define $\mathcal{E}_1$ to be the event that there exists at least one iteration in which players learned incorrect value of j on line 24. Define $\mathcal{E}_2$ to be the event that $\tau_{k,\gamma}$ terminates without u being a leaf of T_r. Finally, define $\mathcal{E} = \mathcal{E}_1 \cup \mathcal{E}_2$. To analyze $P(\mathcal{E})$ we shall need the following lemma.

Lemma 4.2.3. *Conditioned on all executions of* FDIFF *protocol being correct on line 24, the expected number of times the while loop on line 4 is executed in τ_γ is at most* $\mathrm{IC}_\mu(\pi)$.

We defer the proof of Lemma 4.2.3 to finish the rest of the current proof. By Markov's inequality we have $P(\mathcal{E}_2|\neg\mathcal{E}_1) \leq \frac{\mathrm{IC}_\mu(\pi)}{k}$. By union bound, we have $P(\mathcal{E}_1) \leq k\gamma$. Overall, we get that

$$\begin{aligned}P(\mathcal{E}) = P(\mathcal{E}_1\cup\mathcal{E}_2) &= P((\mathcal{E}_1\cup\mathcal{E}_2)\cap\mathcal{E}_1) + P((\mathcal{E}_1\cup\mathcal{E}_2)\cap\neg\mathcal{E}_1) \leq P(\mathcal{E}_1) + P(\mathcal{E}_2|\neg\mathcal{E}_1)\\ &\leq \frac{\mathrm{IC}_\mu(\pi)}{k} + k\gamma.\end{aligned}$$

Finally, note that conditioned on $\neg\mathcal{E}$ the players end up with u being the correct leaf

Algorithm 15 Constructing τ_γ out of π

Require:

$x \in \{0,1\}^n$ - known to Alice
$y \in \{0,1\}^n$ - known to Bob
μ, R - known to Alice and Bob

Using public randomness, players jointly sample r and construct T_r
Using public randomness, players jointly sample $t_u \in [0,1]$ for each node $u \in T_r$
Both players set the current node u to the root of T_r
while u is not a leaf **do**
    Alice builds a path $a = a_0, \ldots, a_k$ without communicating with Bob as follows
    $a_0 \leftarrow u, i \leftarrow 0$
    do
        if owner of a_i is Alice **then**
            $a_{i+1} \leftarrow$ left child of a_i if $p_{a_i}(x) = 0$ and right child of a_i otherwise
        else
            $a_{i+1} \leftarrow$ left child of a_i if $t_{a_i} > \tilde{p}_{a_i}(x)$ and right child of a_i otherwise
        $i \leftarrow i + 1$
    while a_i is not a leaf
    Bob builds a path $b = b_0, \ldots, b_\ell$ without communicating with Alice as follows
    $b_0 \leftarrow u, i \leftarrow 0$
    do
        if owner of b_i is Bob **then**
            $b_{i+1} \leftarrow$ left child of b_i if $p_{b_i}(y) = 0$ and right child of b_i otherwise
        else
            $b_{i+1} \leftarrow$ left child of b_i if $t_{b_i} > \tilde{p}_{b_i}(y)$ and right child of b_i otherwise
        $i \leftarrow i + 1$
    while b_i is not a leaf
    Players run protocol from Lemma 4.2.2 with error tolerance γ
    Players learn j, such that $j = \mathrm{FDIFF}(\langle a \rangle, \langle b \rangle)$ with probability $1 - \gamma$
    if $j = 0$ **then**
        Protocol terminates
    else
        Owner of $a_{j-1} (= b_{j-1})$ sends the true value of $p_{a_{j-1}}$
        if the true value is 0 **then**
            Players update u to the left child of a_{j-1}
        else
            Players update u to the right child of a_{j-1}

of T_r reached from the given inputs. Since leaves are in one-to-one correspondence with transcripts of π we are done. □

Proof of Lemma 4.2.3. Let $\tau_{k,\gamma,r}$ denote the protocol obtained from $\tau_{k,\gamma}$ after line 1 has been executed, i.e., after public randomness r has been sampled. Let $S_{i,r}$ be the indicator random variable indicating whether a mistake occurred at depth i of T_r, where $i \in [\mathrm{CC}(\pi)]$.

Suppose that the current node in the execution of $\tau_{k,\gamma,r}(x,y)$ is u and that depth of u is i. Without loss of generality, assume that u is owned by Alice and $p_u(x) = 1$. The probability that Bob guesses the incorrect child of node u is

$$1 - \tilde{p}_u(y) \leq \ln\left(\frac{1}{\tilde{p}_u(y)}\right) < \log\left(\frac{1}{\tilde{p}_u(y)}\right).$$

Let $\pi_r^i(x,y)$ denote the ith bit sent during the execution of protocol π_r on input (x,y). Note that $\pi_r^i(x,y)$ is not random. When (parts of) the input (X,Y) is a random variable, $\Pi_r^i(X,Y)$ becomes a random variable, so we denote it by a capital letter. Let $\pi_r^{<i}(x,y)$ denote the concatenation of the first $i-1$ bits sent during the execution of protcol π_r on input (x,y). Then we have

$$\mathbb{D}(\pi_r^i(x,y)||(\Pi_r^i(X,y)|\Pi_r^{<i}(X,y) = \pi_r^{<i}(x,y))) = \log\left(\frac{1}{\tilde{p}_u(y)}\right).$$

Combining the above equations, we get that Bob guesses the incorrect child of u with probability at most $\mathbb{D}(\pi_r^i(x,y)||(\Pi_r^i(X,y)|\Pi_r^{<i}(X,y) = \pi_r^{<i}(x,y)))$. The same conclusion holds in case $p_u(x) = 0$. Similarly, if Bob owns node u, then Alice makes the incorrect guess at node u with probability at most $\mathbb{D}(\pi_r^i(x,y)||(\Pi_r^i(x,Y)|\Pi_r^{<i}(x,Y) = \pi_r^{<i}(x,y)))$. Therefore, we get the following:

$$\begin{aligned}
\mathbb{E}_{x,y,\{t_u\}}(S_{i,r}) &\leq \mathbb{E}_{x,y}\left(\mathbb{D}(\pi_r^i(x,y)||(\Pi_r^i(X,y)|\Pi_r^{<i}(X,y) = \pi_r^{<i}(x,y)))\right) + \\
&\quad + \mathbb{E}_{x,y}\left(\mathbb{D}(\pi_r^i(x,y)||(\Pi_r^i(x,Y)|\Pi_r^{<i}(x,Y) = \pi_r^{<i}(x,y)))\right) \\
&= I(X;\Pi_r^i(X,Y)|Y\Pi_r^{<i}(X,Y)) + I(Y;\Pi_r^i(X,Y)|X\Pi_r^{<i}(X,Y)).
\end{aligned}$$

Applying the chain rule for mutual information, we obtain

$$\begin{aligned}
\mathbb{E}_{x,y,\{t_u\}}\left(\sum_{i=1}^{\mathrm{CC}(\pi)} S_{i,r}\right) &= \sum_{i=1}^{\mathrm{CC}(\pi)} \mathbb{E}_{x,y,\vec{t_u}}(S_{i,r}) \\
&\leq \sum_{i=1}^{\mathrm{CC}(\pi)} I(X;\Pi_r^i(X,Y)|Y\Pi_r^{<i}(X,Y))+ \\
&+ \sum_{i=1}^{\mathrm{CC}(\pi)} I(Y;\Pi_r^i(X,Y)|X\Pi_r^{<i}(X,Y)) \\
&= I(X;\Pi_r(X,Y)|Y) + I(Y;\Pi_r(X,Y)|X) \\
&= \mathrm{IC}_\mu(\pi_r).
\end{aligned}$$

Let S_i be the indicator random variable indicating that a mistake occurs at step i during the execution of τ_γ. Using Fact 4.1.3 we obtain

$$\mathbb{E}_{x,y,\{t_u\},r}\left(\sum_{i=1}^{\mathrm{CC}(\pi)} S_i\right) = \mathbb{E}_{x,y,\{t_u\},r}\left(\sum_{i=1}^{\mathrm{CC}(\pi)} S_{i,r}\right) \leq \mathbb{E}_r(\mathrm{IC}_\mu(\pi_r)) = \mathrm{IC}_\mu(\pi).$$

□

The following theorem follows easily from Lemma 4.2.1.

Theorem 4.2.4. *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function, $\mu \in \Delta(\{0,1\}^n \times \{0,1\}^n)$, and $\epsilon \in (0,1/2)$ be a parameter. Let π be a communication protocol with public randomness only that solves f with probability of error $\leq \epsilon$ when inputs are sampled from μ. Then we have*

$$\mathrm{D}_\mu(f,3\epsilon) = O\left(\frac{\mathrm{IC}_\mu(\pi)}{\epsilon}\log\frac{\mathrm{CC}(\pi)\,\mathrm{IC}_\mu(\pi)}{\epsilon^2}\right).$$

Consequently, there exists μ such that $\mathrm{R}(f,3\epsilon) = O\left(\inf_\pi \frac{\mathrm{IC}_\mu(\pi)}{\epsilon}\log\frac{\mathrm{CC}(\pi)\,\mathrm{IC}_\mu(\pi)}{\epsilon^2}\right)$.

Proof. Let $k = \frac{\mathrm{IC}_\mu(\pi)}{\epsilon}$ and $\gamma = \frac{\epsilon^2}{\mathrm{IC}_\mu(\pi)}$. Apply Lemma 4.2.1 to π to get $\tau_{k,\gamma}$ with k and

γ as specified. Thus, we have that

$$\mathrm{CC}(\tau_{k,\gamma}) = O\left(\frac{\mathrm{IC}_\mu(\pi)}{\epsilon}\log\frac{\mathrm{CC}(\pi)\,\mathrm{IC}_\mu(\pi)}{\epsilon^2}\right) \text{ and}$$

$$P_{x,y,\{t_u\},r}(\tau_{k,\gamma,r}(x,y) \neq \pi_r(x,y)) \leq \frac{\mathrm{IC}_\mu(\pi)}{\mathrm{IC}_\mu(\pi)/\epsilon} + \frac{\mathrm{IC}_\mu(\pi)}{\epsilon}\frac{\epsilon^2}{\mathrm{IC}_\mu(\pi)} \leq 2\epsilon.$$

Given that $P_{x,y,r}(\pi_r(x,y) \neq f(x,y)) \leq \epsilon$ we get that $P_{x,y,\{t_u\},r}(\tau_{k,\gamma,r}(x,y) \neq f(x,y)) \leq 3\epsilon$. In particular, there exists a fixing of $\{t_u\}$ and r such that $P_{x,y}(\tau_{k,\gamma,r}(x,y) \neq f(x,y)) \leq 3\epsilon$. Thus $\mathrm{D}_\mu(f,3\epsilon) = O\left(\frac{\mathrm{IC}_\mu(\pi)}{\epsilon}\log\frac{\mathrm{CC}(\pi)\,\mathrm{IC}_\mu(\pi)}{\epsilon^2}\right)$. The consequence part of the statement of the theorem follows from Yao's minimax principle. □

CHAPTER 5
CONCLUSIONS

5.1 Open Problems

In this section we list several open problems inspired by the results described in previous chapters. We list the problems in a perceived order of increasing abstractness.

We derived exact and closed-form formulas for the information complexity of the AND function with error tolerance 0. A natural question is whether this method can be adapted to the information complexity of the AND function with some fixed error tolerance $\epsilon > 0$. This task will likely require new techniques. While the approach from Chapter 2 should work in theory, it becomes intractable in practice with an increased number of parameters of the information complexity function.

Open Problem 5.1.1. Compute $\mathrm{IC}_\mu(\mathrm{AND}, \epsilon)$ for $\epsilon > 0$ exactly.

Another open problem is to extend the techniques of Chapter 2 to handle functions with alternating levels of $\bigvee$ and $\bigwedge$.

Open Problem 5.1.2. Compute the exact communication complexity of functions with alternating levels of $\bigvee$ and $\bigwedge$, e.g., $\bigvee_{i=1}^{\sqrt{n}} \bigwedge_{j=1}^{\sqrt{n}} x_{ij} \vee y_{ij}$.

With regards to Chapter 3, resolving the following problem would likely lead to new insights in connection to the information complexity of XOR of n copies of a function.

Open Problem 5.1.3. What is the best value of $\alpha > 0$ such that $\mathrm{IC}_{\mathcal{U}}(\mathrm{IP}_n, \epsilon) \geq (1 - \alpha\epsilon)n$?

For the Gap Hamming Distance problem, the question of interest is whether its information complexity approaches the trivial bound as the error tolerance goes to 0 under a natural setting of parameters.

Open Problem 5.1.4. Is it true that for all $\epsilon > 0$ there exist $\delta > 0$ and measure μ such that $\mathrm{IC}_\mu(\mathrm{GHD}_{n,n/2,\delta\sqrt{n}}, \epsilon) \geq (1-\epsilon)n$?

With regards to Chapter 4 the main open problem is to find the exact relationship between $\mathrm{IC}^{\mathrm{pub}}_\mu(f, \epsilon)$ and $\mathrm{IC}_\mu(f, \epsilon)$.

Open Problem 5.1.5. Quantify the relationship between $\mathrm{IC}^{\mathrm{pub}}_\mu(f, \epsilon)$ and $\mathrm{IC}_\mu(f, \epsilon)$.

In Chapter 2 we described a partial differential equation formulation of information complexity. We believe that this observation deserves further attention. The following questions need to be addressed first:

Open Problem 5.1.6. Does the system of PDEs describing information complexity always have a solution? Is the solution unique? If the system of PDEs is, in fact, a reasonable characterization of information complexity, then what properties of information complexity can be inferred from the rich area of elliptic PDEs?

As we briefly mentioned in the introduction, it is rather difficult to obtain strong lower bounds in the number-on-the-forehead (NOF) multiparty communication complexity model. The only known methods for such lower bounds are based on combinatorial and analytic techniques. In particular, at present there are no viable candidates for the concept of information complexity in the NOF setting. There are some known obstacles towards this goal. If one tries to preserve the direct sum property when extending the definition of information complexity from 2 players to k playes, one is led to the notion of randomness-on-the-forehead. However, it is known that with randomness-on-the-forehead information complexity trivializes – it becomes constant *for all functions*. Thus, the following open problem is one of the central ones in the area of information complexity.

Open Problem 5.1.7. Define a notion of information complexity for the NOF multiparty communication complexity. This notion should obey the direct sum property, or some reasonable weakening of the direct sum property, and it should provide a reasonable lower bound for the k-party AND function – for example, $\Omega(1/2^k)$.

We finish with the most abstract and ambitious open problem. A posteriori, one can view the development of information complexity as the following program:

consider a complexity measure – communication complexity, define a lower bound in terms of informational quantities such that this lower bound has nice properties (e.g., direct sum), use techniques from information theory (e.g., chain rule) to reprove and improve the known results regarding the original complexity measure, prove new results regarding the complexity measure. The ambitious open problem is to implement this program for other complexity measures in other areas of complexity theory.

Open Problem 5.1.8. Reprove (improve?) known lower bounds in other areas of complexity theory via information-theoretic arguments. Find suitable information-theoretic measures for computational models other than communication.

5.2 Conclusions

We described three contributions that establish new connections between communication complexity and information complexity.

In the first contribution, we computed the information complexity of the smallest nontrivial two-party function: the AND function. This led to new techniques in communication complexity and information complexity and ultimately allowed us to compute the *exact communication complexity* of several functions on n-bit inputs. Among other results, we showed that the exact communication complexity of the set disjointness function on n-bit inputs is $C_{\text{DISJ}}n + o(n)$, where $C_{\text{DISJ}} \approx 0.4827$. Such precise statements are quite common in the area of information theory, but were not known in the area of communication complexity prior to this work. The significance of this result is that information complexity exactly captures the minimum amount of communication that is necessary and sufficient to solve a whole class of functions – the $\bigvee$-type functions.

In the second contribution, we show that lower bounds on communication complexity of self-reducible functions imply lower bounds on the information complexity in a black-box manner. Numerous previous works used information complexity as a lower bound method for communication complexity, and we demonstrate that the connection between information complexity and communication complexity is a two-

way connection for the self-reducible functions. We highlighted this connection by proving strong lower bounds on the information complexity of GHD and IP functions.

In the third contribution, we proved that protocols using only public randomness and having small information cost can be efficiently compressed in terms of communication. This implies that if one hopes to prove strong separations between information and communication complexities, then one has to crucially rely on private randomness, as in [25]. If one wishes to disprove strong separations between information and communication complexities (to the extent not ruled out by [25]) then it suffices to show how to simulate private randomness with public randomness without increasing information complexity too much.

The first and second contributions are based on the joint work with Braverman, Garg, and Weinstein. The third contribution is my work alone.

REFERENCES

[1] Farid Ablayev. Lower bounds for one-way probabilistic communication complexity and their application to space complexity. *Theoret. Comp. Sci.*, 157(2):139–159, 1996.

[2] Noga Alon, Yossi Matias, and Mario Szegedy. The space complexity of approximating the frequency moments. In *Proc. 28th STOC*, pages 20–29, 1996.

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

[4] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proc. 42nd STOC*, pages 67–76, 2010.

[5] Richard Beigel and Jun Tarui. On ACC. In *Proc. 32nd FOCS*, pages 783–792, 1991.

[6] Eric Blais, Joshua Brody, and Kevin Matulef. Property testing lower bounds via communication complexity. *Comput. Compl.*, 21(2):311–358, 2012.

[7] Mark Braverman. Interactive information complexity. In *Proc. 44th STOC*, pages 505–524, 2012.

[8] Mark Braverman and Ankit Garg. Public vs private coin in bounded-round information. In *Automata, Languages, and Programming*, volume 8572, pages 502–513. 2014.

[9] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. From information to exact communication. In *Proc. of 45th STOC*, pages 151–160, 2013.

[10] Mark Braverman, Ankit Garg, Denis Pankratov, and Omri Weinstein. Information lower bounds via self-reducibility. In *Comput. Sci. Theory and Applications*, volume 7913, pages 183–194. 2013.

[11] Mark Braverman and Ankur Moitra. An information complexity approach to extended formulations. In *Proc. 45th STOC*, pages 161–170, 2013.

[12] Mark Braverman and Anup Rao. Information equals amortized communication. In *Proc. 52nd FOCS*, pages 748–757, 2011.

[13] Mark Braverman and Omri Weinstein. A discrepancy lower bound for information complexity. *Approx., Random., and Combinat. Optim. Algorithms and Techniques*, 7408:459–470, 2012.

[14] Mark Braverman and Omri Weinstein. An interactive information odometer with applications. *ECCC*, 2014.

[15] Joshua Brody, Harry Buhrman, Michal Koucky, Bruno Loff, Florian Speelman, and Nikolay Vereshchagin. Towards a reverse Newman's theorem in interactive information complexity. In *Proc. of CCC*, pages 24–33, 2013.

[16] Joshua Brody, Amit Chakrabarti, and Ranganath Kondapally. Certifying equality with limited interaction. *ECCC*, 2012.

[17] Amit Chakrabarti, Subhash Khot, and Xiaodong Sun. Near-optimal lower bounds on the multi-party communication complexity of set disjointness. In *Proc. CCC*, pages 107–117, 2003.

[18] Amit Chakrabarti, Ranganath Kondapally, and Zhenghui Wang. Information complexity versus corruption and applications to orthogonality and gap-hamming. *Approx., Random., and Combinat. Optim. Algorithms and Techniques*, 7408:483–494, 2012.

[19] Amit Chakrabarti and Oded Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. In *Proc. 43rd STOC*, pages 51–60, 2011.

[20] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Proc. 42nd FOCS*, pages 270–278, 2001.

[21] Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. *arXiv preprint arXiv:0801.3624*, 2008.

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

[23] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley series in telecommunications. J. Wiley and Sons, New York, 1991.

[24] Uriel Feige, David Peleg, Prabhakar Raghavan, and Eli Upfal. Computing with noisy information. *SIAM J. of Comput.*, 23(5):1001–1018, 1994.

[25] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication for boolean functions. *ECCC*, 2014.

[26] Johan Håstad and Avi Wigderson. The randomized communication complexity of set disjointness. *Theory Of Comput.*, 3:211–219, 2007.

[27] David A. Huffman. A method for the construction of minimum-redundancy codes. In *Proc. of the IRE*, volume 40, pages 1098–1101, 1952.

[28] Prakash Ishwar and Nan Ma. Personal communication.

[29] Thathachar Jayram. Hellinger strikes back: A note on the multi-party information complexity of AND. *Approx., Random., and Combinat. Optim.. Algorithms and Techniques*, pages 562–573, 2009.

[30] Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. on Disc. Mathem.*, 5(4):545–557, 1992.

[31] Mauricio Karchmer and Avi Wigderson. Monotone circuits for connectivity require super-logarithmic depth. *SIAM J. on Disc. Mathem.*, 3(2):255–265, 1990.

[32] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. In *Proc. 53rd FOCS*, pages 500–509, 2012.

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

[34] Troy Lee and Adi Shraibman. Lower bounds in communication complexity. *Found. and Trends in Theoret. Comp. Sci.*, 3(4):263–399, 2007.

[35] Iordanis Kerenidis Sophie Laplante Mathieu Lauriere Jérémie Roland Lila Fontes, Rahul Jain. Relative discrepancy does not separate information and communication complexity. *ECCC*, 2015.

[36] Nan Ma and Prakash Ishwar. Two-terminal distributed source coding with alternating messages for function computation. In *Proc. ISIT*, pages 51–55, 2008.

[37] Nan Ma and Prakash Ishwar. Infinite-message distributed source coding for two-terminal interactive computing. In *Proc. 47th Allerton Conf. on Comm., Control, and Comp.*, pages 1510–1517, 2009.

[38] Nan Ma and Prakash Ishwar. Some results on distributed source coding for interactive function computation. *IEEE Trans. on Inform. Theory*, 57(9):6180–6195, 2011.

[39] Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. On data structures and asymmetric communication complexity. In *Proc. 27th STOC*, pages 103–111, 1995.

[40] Ilan Newman. Private vs. common random bits in communication complexity. *Inform. Proc. Letters*, 39(2):67–71, 1991.

[41] Noam Nisan and Avi Wigderson. Rounds in communication complexity revisited. *SIAM J. on Comput.*, 22(1):211–219, 1993.

[42] Alon Orlitsky. Worst-case interactive communication. I. Two messages are almost optimal. *IEEE Transact. on Inform. Theory*, 36(5):1111–1126, 1990.

[43] Alon Orlitsky. Worst-case interactive communication. II. Two messages are not optimal. *IEEE Transact. on Inform. Theory*, 37(4):995–1005, 1991.

[44] Denis Pankratov. Direct sum questions in classical communication complexity, 2012. Available at http://people.cs.uchicago.edu/~pankratov/.

[45] Alexander A. Razborov. On the distributed complexity of disjointness. *Theoret. Comput. Sci.*, 106, 1992.

[46] Alexander A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67:145–159, 2003.

[47] Mert Saglam and Gábor Tardos. On the communication complexity of sparse set disjointness and exists-equal problems. In *Proc. 54th FOCS*, pages 678–687, 2013.

[48] Claude E. Shannon. A mathematical theory of communication. *Bell Syst. Techn. J.*, 27, 1948. Monograph B-1598.

[49] Alexander A. Sherstov. The pattern matrix method. *SIAM J. on Comput.*, 40:1969–2000, 2008.

[50] Alexander A. Sherstov. The multiparty communication complexity of set disjointness. In *Proc. 44th STOC*, pages 525–548, 2012.

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

[52] Andrew C.C. Yao. Some complexity questions related to distributive computing (preliminary report). In *Proc. 11th STOC*, pages 209–213, 1979.