Improved XOR Lemmas for Communication Complexity

Siddharth Iyer Vaidyanathan

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Anup Rao, Chair

Paul Beame

Thomas Rothvoß

Program Authorized to Offer Degree:
Computer Science & Engineering

University of Washington

Abstract

Improved XOR Lemmas for Communication Complexity

Siddharth Iyer Vaidyanathan

Chair of the Supervisory Committee:
Anup Rao
Department of Computer Science & Engineering

We give communication lower bounds for computing the n-fold XOR of a given Boolean function f, denoted $f^{\oplus n}(x, y) := f(x_1, y_1) \oplus \ldots \oplus f(x_n, y_n)$, in both the deterministic and the randomized setting. In addition, we also give deterministic communication lower bounds on computing the composition of 2 functions, $g \circ f(x, y) := g(f(x_1, y_1), \ldots, f(x_n, y_n))$. Below for some absolute constant $C_0 > 0$ and all $C > C_0$ we show the following:

1. **Randomized XOR Lemma.** If f requires C bits to be computed with some constant success probability then, computing $f^{\oplus n}$ with probability at least $1/2 + \exp(-\Omega(n))$ requires $\widetilde{\Omega}(C\sqrt{n})$ bits.

2. **Deterministic XOR Lemma.** If f requires C bits to be computed deterministically then, computing $f^{\oplus n}$ deterministically requires $\Omega(n\sqrt{C})$ bits.

3. **Lifting Theorem.** For any function g, having sensitivity s and degree d, and any f requiring C bits to be computed deterministically, computing $g \circ f$ deterministically requires $\Omega(\min\{s, d\} \cdot \sqrt{C})$ bits.

We prove the above results using information theory. In particular, the randomized XOR lemma is proved using a new notion of information that we call marginal information.

TABLE OF CONTENTS

Page

LIST OF FIGURES

ACKNOWLEDGMENTS

This journey would not have been possible without several people. A lot of the credit goes to my advisor, Anup Rao. In the course of my PhD, I have learnt a lot of mathematics and gained valuable experience working with him. I am extremely grateful for the overall direction and support he has given me and also for the time and freedom while working on the problems. A big reason this thesis exists is due to him, without his encouragement, there is a good chance I would not have worked on XOR lemmas.

The CS theory faculty at UW have played a large part in my graduate education, for which I am very grateful. Paul Beame and Thomas Rothvoß, served on my committee and also wrote letters for me. I am especially thankful to Paul, for his advice and feedback on several occasions.

Among others who have contributed significantly to my PhD journey, are my co-authors. I have had the privilege to have worked with and learned from Anup Rao, Victor Reis, Thomas Rothvoß, Amir Yehudayoff, and Michael Whitmeyer.

I am lucky to have had a wonderful community of students at the theory group, and more broadly at the CS department at UW. I am especially grateful to count among my friends, Ashrujit Ghoshal, Michael Whitmeyer, Oscar Sprumont and Siva Ramamoorthy. I have had many enjoyable moments and made a lot of great memories playing CSE-mafia and clocktower – thanks a lot to both these groups at CSE. Many thanks also go to Anup, Ashrujit, Avi, Michael, Oscar, Srini and Aditi for their company in all the moments spent outdoors.

My extended family has played a huge part in me having an enjoyable and comfortable time in the US. Thanks to Archana Chiti and Param Chitappa for all the support, especially in the early days and during COVID. Thanks to Priya Chiti and Vignesh Chitappa for all

their support and the great times in the Bay Area. I am also very fortunate to have had great roommates in Seattle, thanks especially to Srini and Aditi, you have been like family away from family. I am also grateful to Eshwar Ram, for his friendship, support and great conversations over the years.

I am quite lucky to have had the opportunity to have worked on CS theory research during my undergraduate years. I am grateful to Samir Datta, who introduced me to several topics in theoretical computer science, and mentored and worked with me on research projects. Thanks also to Ran Gelles for giving me the opportunity to work on interactive codes.

This thesis would not be if not for the support and sacrifices of my parents. Thanks to you I have had a fantastic education and great experiences. You continue to inspire me everyday.

DEDICATION

To Amma and Appa

Chapter 1

INTRODUCTION

In this thesis we aim to further our understanding of some natural questions in computational complexity. Computational complexity is the branch of computer science that studies the limits of computation. The questions we consider in thesis have the following theme.

> *How does the computational difficulty of computing several copies of a given task scale with the number of copies?*

Indeed, if a task can be computed by an algorithm using some C resources then computing n instances of the task can be done with $n \cdot C$ resources by simply repeating the algorithm on each instance. Is this the best that one can do? The *direct sum problem* asks exactly this: does computing n instances of a given task require n times the resources needed to compute a single one?

One can also consider a variant of this problem by relaxing the requirement of exact computation to approximate computation; that is, we allow randomized algorithms that make errors with some small (fixed) probability. The study of randomized algorithms is an important area within theoretical computer science – for several problems, we know simple randomized algorithms whose performance is comparable to the best known deterministic ones.

Suppose we have a function $f(x)$ such that the best randomized algorithm computing it with C resources succeeds with probability $2/3$. What is the relationship between the success probability of computing n instances of f with the amount of computational resources used. If one repeats the best randomized algorithm for f on each instance independently, the probability of succeeds on all n instances is $(2/3)^n$ while the amount of resources used is $n \cdot C$. In other words, naively repeating the best algorithm for f incurs a linear resource

blowup while the measure of approximation (the success probability) decreases exponentially with n. Similar to the direct sum problem, the *strong direct product problem* asks if this is necessary: if f requires C resources to be computed with $2/3$ success probability, is the success probability of computing n instances of f with $n \cdot C$ exponentially small in n?

A reason why computing n copies of f might be computationally expensive could simply be because one needs to express the output of f on n instances. Indeed, the process of verifying whether or not the output of a computational procedure on n copies of f is successful is itself quite expensive since one has to check the output for each copy. What if we replace the requirement of computing f on n copies with that of computing a single bit of information regarding the n outputs? Let us state this more precisely. Consider some Boolean f and let $f^{\oplus n}$ denote the parity[1] of f on n instances. Now, one can ask: how does the computational hardness of computing $f^{\oplus n}$ depend on n and the hardness of f? Is it as hard to compute (or approximate) $f^{\oplus n}$ as it is to compute (or approximate) f on n copies?

As in the case of direct sum/product questions, let us consider what happens if we naively repeat the best algorithm for f. First, suppose that one needs C resources to compute f with probability at least $2/3$. If we repeat the best procedure for f on each instance and take the parity of the resulting outputs, the success probability is roughly $1/2 + 3^{-n}/2$. One can see this by reasoning about a quantity very related to the success probability, known as *advantage*. Roughly speaking, the advantage of a randomized procedure for computing a Boolean function f is the expected value of the random variable which is $+1$ is the procedure is successful and -1 if not. It turns out that the advantage multiplies when we repeat the same procedure multiple times. In other words, the advantage of the repeated protocol to compute $f^{\oplus n}$ is exponentially small in n. Note that $f^{\oplus n}$ can be computed with probability $1/2$, equivalently, zero advantage; the naive protocol is barely better than random guessing even at the expense of a linear blowup in the resources. Is it possible to do better: does computing $f^{\oplus n}$ with advantage at least $\exp(-O(n))$ require a linear blow-up in the resources?

A statement that shows the hardness of computing the $f^{\oplus n}$ in terms of n as well as the

[1] In general, one can choose any sufficiently complex Boolean function g and ask about the hardness of computing $g \circ f$. The case when g is the parity function is a simple and a natural choice.

hardness of computing f is called an *XOR lemma*. The main contributions of this thesis are new XOR lemmas in the computational model of communication complexity. Before discussing our results we review some prior work on direct sums/products and XOR lemmas in theoretical computer science.

1.1 Direct Sums, Direct Products & XOR Lemmas

As one might imagine, the need to repeatedly compute a given task occurs frequently in computer science. Studying this problem from the perspective of lower bounds has led to important results in various areas in theoretical computer science, for example circuit complexity [35, 34], cryptography [66, 37], hardness of approximation [53, 49], etc. Below, we expand on some of the above examples to provide context for these questions.

Matrix Multiplication. Matrix multiplication provides a classic example of a task for which one can compute several copies significantly faster than computing each copy separately. Indeed, multiplying two $n \times n$ matrices can be thought of as computing the product of n different vectors with a fixed matrix. To compute a single matrix-vector product, one needs to at the very least read the matrix and the vector, which requires n^2 operations. Repeating this n times yields an algorithm that takes n^3 steps. Yet, in 1969, Strassen [63] demonstrated a surprising algorithm that multiplies two matrices in time roughly $n^{2.81}$ – a noticeable savings over the naive algorithm. The current best algorithm for matrix multiplication runs in time $O(n^{2.32})$ [2]! Can we always expect to obtain such savings or are there tasks for which computing n instances necessarily requires an n-fold increase in the resources?

Cryptography and Yao's XOR Lemma. A basic principle of cryptography is that we can use computational tasks that are hard to compute, yet whose solutions are easy to verify – like factoring discussed above – to give security guarantees. Given this, one might hope that there are tasks for which computing several instances really requires a significant increase in the computational cost. Indeed, a natural way to amplify the complexity of a task is by simply repeating it. Using computational tasks whose complexity increases upon

repetition we can hope to build more secure cryptographic systems. A textbook example of this is the amplification of *weak one-way functions* to *strong* one-way functions [66, 37], which was obtained by the famous *Yao's XOR lemma*. A one-way function is a function that is easy to compute but hard to invert. While we do not yet know whether one-way functions exist, we have candidates, a notable one being multiplication. Multiplying two numbers is easy, but factoring – the inverse of multiplication – seems computationally hard. The weak-to-strong amplification of one-way functions essentially says that using a Boolean function f that is "mildly" hard to invert, we can construct a function g that is significantly harder to invert. Moreover, the construction is very simple: g is simply the XOR of f on polynomially many copies.

We note perhaps a subtle point here. It might well be possible to invert g without actually inverting each copy of f – an algorithm that inverts several copies of f certainly yields an algorithm to invert g, but the converse need not hold. Hence, the fact that g is hard to invert is a stronger statement than saying that inverting f on several instances is hard. Our main results are XOR lemmas in the computational model of communication complexity, that we introduce shortly.

Small Circuits vs Shallow Circuits. Our last example comes from an attempt to answer a question raised earlier: can efficient algorithms be made even faster with access to more processors? It turns out that efficient algorithms correspond to circuits of small[2] size, and algorithms that lend themselves to be efficiently parallelized correspond to circuits that are shallow[3]. Hence, the question mentioned above boils down to whether or not small circuits can be simulated by shallow ones? We believe that this is not the case; Karchmer, Raz and Wigderson [34] gave an example of a function which can be computed by polynomial size circuits yet it seems hard to compute with circuits of logarithmic depth. We shall describe this next (see also Figure 1.1). For an appropriate Boolean function

[2] size that is polynomial in the length of the input

[3] where the depth of any leaf is some polynomial in the logarithm of the input length

$f : \{0,1\}^k \to \{0,1\}$, define the function $f^{(1)} : \{0,1\}^{k^2} \to \{0,1\}$ as

$$f^{(1)}(x_1, \ldots, x_k) = f \circ f(x_1, \ldots, x_k) := f(f(x_1), \ldots, f(x_k)).$$

We can recursively define $f^{(m)} : \{0,1\}^{k^m} \to \{0,1\}$ to be $f^{(m)} := f \circ f^{(m-1)}$, where the $\circ$ operator denotes composition, as written above. If we can show non-trivial lower bounds on the depth of circuits computing compositions then, for appropriate values of m, we can hope to show that $f^{(m)}$ can be computed by small circuits but not by shallow ones!

In the next section, we describe the model of communication complexity, which will be the setting where we formalize the questions stated at the beginning of this chapter.

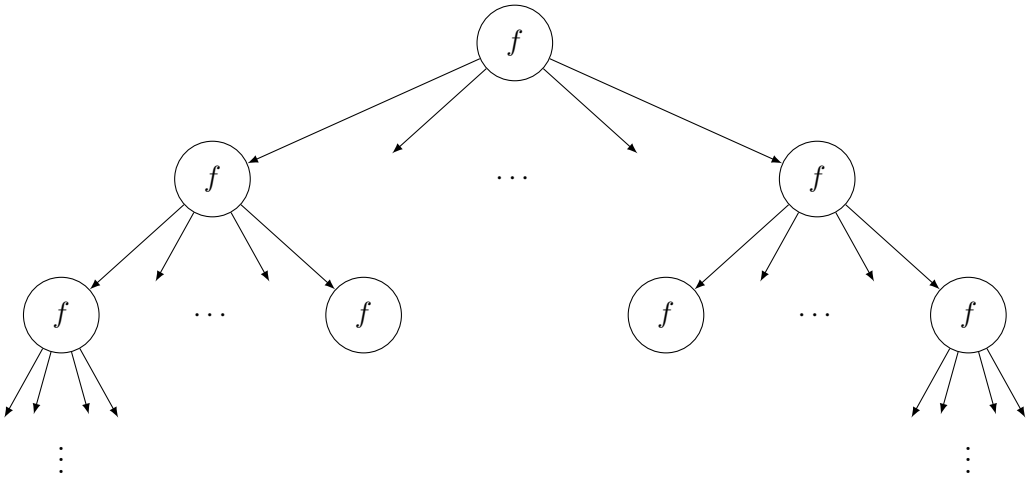

Figure 1.1: The bottommost layer (not pictured) would be the inputs, and the outputs of intermediate nodes are the inputs to the parent node. The topmost node is the output of the function.

1.2 Communication Complexity

Communication is inherent in computation. For example, in distributed computing one typically uses many communicating processors to compute functions; in the streaming setting, a user communicates several times with the cloud to carry out computations over the

stored data; similarly, a data structure algorithm may repeatedly query the data to compute various statistics, and so on. The model of communication complexity captures the communication involved in computing in meaningful manner.

In this model, there are two players Alice and Bob. Alice knows an input x, Bob knows an input y and the players wish to compute a function $f(x, y)$. They do so by executing a communication protocol π – an algorithm that specifies the message that each player conveys to the other in an alternating manner. Sometimes, we will allow the players to share a common (public) random tape and give them access to (private) random coins. Access to randomness dramatically changes the power of protocols, and so we will make it clear whether the protocol is deterministic or randomized. A detailed background on communication complexity can be found in [52]. The deterministic communication complexity of f, denoted by $D(f)$, is the length of the cheapest deterministic protocol computing f.

Since its introduction in 1979 by Yao [67], the study of the communication complexity of functions has proven extremely fruitful in theoretical computer science. Indeed, this model has several connections to other objects in computer science, such as, circuit complexity [35, 34], data structures [1, 43], streaming algorithms [3, 5], and proofs [54, 21] just to name a few. Moreover, researchers have had a lot of success proving lower bounds in this model, often times using ideas from several different fields of mathematics [55, 51, 39].

Let us now turn to the main contributions of this thesis: XOR lemmas in communication complexity.

1.3 Our Contributions

Direct-sum type statements have been studied in communication complexity for several years now. A major reason to understand these type of questions is due to the connection between circuits and communication protocols that was discovered by Karchmer and Wigderson [35]. They observed an equivalence between the depth of the best circuit computing a given function f and the length of the shortest protocol computing the so-called *Karchmer-Wigderson* game for f, denoted KW_f. In the communication game KW_f, Alice receives an input $x \in f^{-1}(0)$ and Bob receives $y \in f^{-1}(1)$. Their goal is to output an index i such that $x_i \neq y_i$. As they showed, the length of the shortest protocol achieving this is

exactly equal to the depth of the best circuit computing f. Hence, obtaining lower bounds on the communication complexity of $\mathrm{KW}_{f\circ g}$ in terms of that of KW_f and KW_g is an avenue to show the existence of functions computable by small circuits but not by shallow ones.

Understanding the communication complexity of Karchmer-Wigderson games has been a major challenge, mainly because KW_f is a relation: for a given pairs of inputs x, y, there could be several valid answers. However, over the last few decades researchers have developed several tools to understand the communication complexity of functions. Perhaps the earliest direct sum statement in communication complexity is due to Feder, Kushilevitz, Naor and Nisan [18]. In words, they showed that computing n copies of a function f requires communication roughly n times the square-root of the communication for a single copy.

Randomized Communication. Direct sums and its variants have been extensively studied in randomized communication complexity as well. A central work in this area is due to Barak, Braverman, Chen and Rao [6]. They showed that for any function f, computing

$$f^n(x, y) := (f(x_1, y_1), \ldots, f(x_n, y_n))$$

with probability 0.99 requires communication roughly $\sqrt{n}$ times the communication required to compute f with probability 0.99. They also showed a similar result to compute the n-fold XOR of f, denoted

$$f^{\oplus n}(x, y) = f(x_1, y_1) \oplus \ldots \oplus f(x_n, y_n).$$

They proved that for any function f with sufficiently large randomized communication complexity, computing $f^{\oplus n}$ with constant success probability requires $\widetilde{\Omega}(\sqrt{n})$ times the communication for a single copy. Building on a recent work of Yu [68], we strengthen the previous lower bound. In Chapter 3, we show that computing $f^{\oplus n}$ with probability that is barely larger, $1/2 + \exp(-\Omega(n))$, requires roughly $\sqrt{n}$ times the communication for a single copy. In Chapter 4, we supply the details of the proof of this XOR lemma.

Deterministic Communication. In Chapter 5, we prove two related results. First, we show a deterministic XOR lemma in communication complexity; i.e., we show a lower

bound on the communication required to compute $f^{\oplus n}(x, y)$ in terms of n, the communication complexity of f, and a related measure known as the rank. The rank of f, denoted $\mathsf{rk}(f)$, is simply the rank of the matrix encoding the function f. Rank and communication complexity are known to be be related: the communication complexity of a function is at least the logarithm of its rank. Moreover, the *log-rank conjecture*, a well-known conjecture in the area, asserts that the communication can be bounded by some polynomial in the logarithm of the rank. We show that

$$D(f^{\oplus}) \geq n \cdot \left(\frac{\Omega(D(f))}{\log \mathsf{rk}(f)} - \log \mathsf{rk}(f) \right).$$

Using the fact that the rank of n-fold XOR of f is $(\mathsf{rk}(f)-1)^n$, we can obtain as a corollary of the above result that if $D(f)$ is a sufficiently large constant, the communication complexity of the n-fold XOR is at least $\Omega(n \cdot \sqrt{D(f)})$.

Next, we generalize the preceding result to show a similar lower bound on the communication required to compute $g \circ f(x, y) := g(f(x_1, y_n), \ldots, f(x_n, y_n))$ for arbitrary g and f. There are several well-studied complexity measures of Boolean functions, such as sensitivity, block-sensitivity, degree, and decision-tree complexity, etc. The above measures are all related to each other up to polynomial factors, and our lower bound is in terms of one such measure – the sensitivity. We show that

$$D(g \circ f) \geq \mathsf{s}(g) \cdot \left(\frac{\Omega(D(f))}{\log \mathsf{rk}(f)} - \log \mathsf{rk}(f) \right),$$

where $\mathsf{s}(g)$ is the *sensitivity* of g. The *sensitivity of g at a point x* is the number of (sensitive) coordinates i such that if x_i is flipped while keeping the others unchanged, then the output of g also flips. The *sensitivity* of g is the largest sensitivity that it can have at any point. Similar to how we used the tensor property of rank in the case of XOR above, we can show that $\mathsf{rk}(g \circ f) \geq (\mathsf{rk}(f) - 1)^{\mathsf{deg}(g)}$, where denotes the degree of the unique, multilinear polynomial that computes g. Using this fact, we obtain as a corollary that if $D(f)$ is a sufficiently large constant then the communication complexity of $g \circ f$ is $\Omega\Big(\min\{\mathsf{s}(g), \mathsf{deg}(g)\} \cdot \sqrt{D(f)} \Big)$.

The last statement is called a *lifting theorem* since it "lifts" a lower bound from a weaker model to a stronger model. In our case, we lift sensitivity/degree, which are complexity measures of Boolean functions, to communication complexity. In prior work, several lift-

ing theorems have been proved with applications in communication complexity and proof complexity, and we review the relevant work in Chapter 5.

1.4 XOR Lemmas via Information Measures

The proofs of the 3 results mentioned above use information theory – a powerful mathematical tool with applications in several areas of mathematics, computer science, engineering, and statistics. We obtain the XOR lemma for deterministic communication as well as the lifting theorem using entropy – a basic definition from information theory. Perhaps more interestingly, the XOR lemma for randomized communication is obtained using a new definition of information, known as *marginal information*. Next, we give a brief background on information theory in the context of obtaining XOR lemma-type statements in communication complexity.

Information theory goes back to the work of Shannon [59], who wanted to know the minimum number of bits required to communicate a message sampled from some known distribution. For instance, communicating two uniformly random bit can be done by sending the bits themselves, however, if the distribution is skewed one can optimize the average the number of bits sent. For example, if "00" is sampled with probability 0.99 and the rest of them equally likely then, one could use the following strategy: use the bit 0 to encode "00", and send all other messages by sending 1 followed by the message itself. Now, the expected communication is $0.99 \times 1 + 0.01 \times 3 = 1.02$, which is much lesser than sending the 2 bits in the clear.

Shannon showed that in general, the communication required is given by a quantity known as the *entropy* of the message. At a high level, this quantity captures how much Bob learns about the message. In the example above, the most likely case is that Alice wants to send the message "00" – this happens with probability 0.99. Hence, when Bob receives a zero, he does not learn much. When he receives a longer message, such as "111", he learns that Alice actually wanted to send "11" which is surprising since it occurs less frequently.

The proofs of the deterministic XOR lemma and the lifting theorem both use the notion of entropy as well as one of its key properties, known as *sub-additivity*. Historically, information theoretic methods have been used extensively for randomized communication but

to a lesser extent for deterministic. Our results are one of the first instances of the application of information theory to obtain an XOR lemma and a lifting theorem for deterministic communication.

Before returning to the randomized XOR lemma, we briefly revisit the example at the start of this section. There, Alice and Bob were able to come up with a strategy that leveraged the skewed nature of the distribution, to effectively "compress" the contents of the message. Indeed, instead of sending 2 bits in the clear Alice just sends a single bit most of the time.

A natural question is whether or not we can do something similar for an entire conversation, instead of just a single message. This is of particular interest for communication complexity: we might hope to reduce the length of a protocol by compressing it in a similar manner as above.

It is worth noting that a natural way to compress a protocol is to compress each message one after the other; this approach has led to direct sum type statements in the *bounded-round* setting of communication complexity [31, 23, 68]. Such statements typically show that if r-round, C-bit protocols cannot succeed to compute f with good probability (or advantage) then r-round protocols with $O(nC/r)$ communication fail to compute f^n (of $f^{\oplus n}$) with good probability (or advantage). Such statements are trivial in the setting where $r = C$ – each message is a single bit.

As it turns out, like entropy, a similar quantity known as the *information complexity*, determines the extent to which general communication protocols can be compressed. The definition of information complexity is due to Barak, Braverman, Chen and Rao [6], and is inspired by variants of this quantity from prior works [53, 15, 4]. Barak et al. showed that protocols with communication C and information I can be compressed down to communication roughly $\sqrt{IC}\log C$. They used this to give a direct sum theorem for randomized communication complexity. They showed that to compute n instances of a function $f(x, y)$ with probability 0.99, the communication required is roughly $\sqrt{n}$ times the communication required for a single instance. Subsequently, Braverman, Weinstein, Rao and Yehudayoff [12] used this definition to strengthen the previous result and proved a direct product statement. They showed that to compute n instances of a function $f(x, y)$ with probability as

small as $\exp(-\Omega(n))$ requires $\widetilde{\Omega}(\sqrt{n} \cdot C)$.

Very recently, Yu [68] proved an strong XOR lemma for bounded-round protocols. Roughly speaking, he showed that if r-round, C-bit protocols fails to compute f with advantage more than $1/3$ then r-round protocols with $O(nC/r^{O(r)})$ bits of communication fail to compute $f^{\oplus n}$ with advantage $\exp(-O(n))$. A major challenge in proving an XOR lemma for randomized communication with exponentially small advantage is that the techniques developed for the direct sum/product problems show that information of protocols that compute the n-fold repetition must be large. However, there are protocols that simultaneously have both exponentially small advantage and exponentially small information. Hence, it seems unlikely that we can prove an XOR lemma for exponentially small advantage using information complexity.

Yu's proof had several new ideas, including that of a potential function which built on the concept of information cost. Our result generalizes Yu's work to all protocols and combines it with the compression methods similar to Barak et al. [6]. In the course of the proof of our XOR lemma, we give a new definition of information that we call *marginal information*. We provide a more detailed account of information cost, its variants and the related results in Chapter 3.

1.5 Organization

We review some preliminary mathematical definitions and facts in Chapter 2. In Chapter 3 we state our randomized XOR lemma, and in Chapter 4 we give the complete proof. Both these chapters are based on the following joint work with Anup Rao [29].

- Siddharth Iyer and Anup Rao. "XOR Lemmas for Communication via Marginal Information". In: STOC 2024. 2024, pp. 652–658. ISBN: 9798400703836. DOI: 10.1145/3618260.3649726. URL: https://doi.org/10.1145/3618260.3649726

Next, in Chapter 5 we prove the deterministic XOR lemma and the lifting result.

- The XOR lemma for deterministic communication is based on the following joint work with Anup Rao [28].

Siddharth Iyer and Anup Rao. *An XOR Lemma for Deterministic Communication Complexity*. 2024. arXiv: 2407.01802 [cs.CC]. URL: https://arxiv.org/abs/2407.01802

- The lifting result is based on [27].

  Siddharth Iyer. "Lifting for Arbitrary Gadgets". In: *Electron. Colloquium Comput. Complex.* TR25-036 (2025). ECCC: TR25-036. URL: https://eccc.weizmann.ac.il/report/2025/036

We end with some open problems in Chapter 6.

Chapter 2

PRELIMINARIES

In this chapter, we set up notation and review some facts from linear algebra, information theory, and communication complexity.

We use $[n]$ to denote the set $\{1, 2, \ldots, n\}$. Given a tuple $x = (x_1 \ldots x_n)$, we let $x_{<i}$ and $x_{\geq i}$ to denote the prefix $(x_1, \ldots, x_{i-1})$ and the suffix $(x_i, \ldots, x_n)$ respectively. For shorthand, we skip commas when referring to several variables; for instance, we write $f(xy)$ in place of $f(x, y)$ and ABC in place of A, B, C etc. When discussing random processes, we use capital letters to denote random variables and lower-case letters to denote values taken by them. If XY are jointly distributed random variables according to some law $p(XY)$ and y is an outcome for Y, we use $p(X|y)$ to denote the conditional distribution of X given that $Y = y$. For a random variable X, we use $\mathsf{supp}(X)$ to denote the set of points in the support of $p(X)$. Given a random variable X, distributed according to some law p, and a function $g(x)$, we denote the expected value of g as $\mathbf{E}_{p(x)}[g(x)] := \sum_x p(x) \cdot g(x)$. In the last expression, we slightly abuse notation by overloading $p(x)$ to mean the distribution of x according sampled according to p rather than the probability that $X = x$.

Given a Boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, define the functions $f^n : \mathcal{X}^n \times \mathcal{Y}^n \to \{0, 1\}^n$, $f^{\oplus n} : \mathcal{X}^n \times \mathcal{Y}^n \to \{0, 1\}$ as follows:

$$f^n(xy) = (f(x_1y_1), f(x_2y_2), \ldots, f(x_ny_n)),$$
$$f^{\oplus n}(xy) = f(x_1y_1) \oplus f(x_2y_2) \oplus \cdots \oplus f(x_ny_n).$$

So, f^n computes f on n different pairs of inputs, and $f^{\oplus n}$ computes the parity of the outputs of f^n. Additionally, given a Boolean function $g : \{0, 1\}^n \to \{0, 1\}$ and a f as above, denote the composed function $g \circ f : \mathcal{X}^n \times \mathcal{Y}^n \to \{0, 1\}$ as

$$g \circ f(xy) = g(f(x_1y_1), \ldots, f(x_ny_n)).$$

2.1 Information Theory

We record some basic information-theoretic facts in this section whose proofs, if omitted, can be found in [17].

Definition 2.1 (Entropy)**.** *Given a random variable A distributed according $p(A)$ the entropy of A is given by*

$$\mathbf{H}(A) := \mathop{\mathbf{E}}_{p(a)} \left[\log \frac{1}{p(a)} \right].$$

Fact 2.2. *If A has finite support, then $\mathbf{H}(A) \leq \log |\mathsf{supp}(p(a))|$, with equality if $p(A)$ is the uniform distribution.*

Given two jointly distributed random variables A and B distributed according to $p(AB)$, the conditional entropy of B given A is defined as

$$\mathbf{H}(B|A) := \mathop{\mathbf{E}}_{p(ab)} \left[\log \frac{1}{p(b|a)} \right].$$

It is well-known that $\mathbf{H}(B|A) \leq \mathbf{H}(B)$. We also recall the chain rule for entropy

$$\mathbf{H}(AB) = \mathbf{H}(A) + \mathbf{H}(B|A). \tag{2.1}$$

Next, we recall the notion of KL-divergence as well as ℓ_1-distance between distributions.

Definition 2.3. *Given two probability distributions $p(A)$ and $q(A)$, the* ℓ_1-distance *between p and q is defined as $\|p(A) - q(A)\|_1 := \sum_a |p(a) - q(a)|$.*

Similarly, the KL-divergence *between p and q is defined as*

$$\mathbf{D}(p(A)||q(A)) := \mathop{\mathbf{E}}_{p(a)} \left[\log \frac{p(a)}{q(a)} \right].$$

Fact 2.4. *For any two distributions $p(A)$ and $q(A)$, it holds that $\mathbf{D}(p(A)||q(A)) \geq 0$. Moreover, $\sqrt{\mathbf{D}(p(A)||q(A))} \geq (1/2) \cdot \|p(A) - q(A)\|_1$.*

Lastly, we define the *mutual information* between two random variables A and B that are jointly distributed according to $p(AB)$ to be

$$\mathbf{I}(A : B) := \mathop{\mathbf{E}}_{p(ab)} \left[\log \frac{p(ab)}{p(a) \cdot p(b)} \right].$$

Given three random variables A, B and C distributed according to $p(ABC)$, we define mutual information of A and B conditioned on C as

$$\mathbf{I}(A : B|C) := \mathop{\mathbf{E}}_{p(abc)} \left[\log \frac{p(ab|c)}{p(a|c) \cdot p(b|c)} \right].$$

2.2 Communication Complexity

As mentioned in the introduction, in the model of communication complexity, two players Alice and Bob receive inputs $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ respectively. They wish to compute a known function[1] $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ by executing a communication protocol, denoted π. The length of the protocol π is the maximum number of bits communicated in any execution of π, and is denoted by $\|\pi\|$. The deterministic communication complexity of f, denoted $D(f)$, is the length of shortest protocol computing f.

In the model of randomized communication complexity, the players have access to a shared random tape, known as public randomness, as well as private randomness that they can use when sending any message. We assume that the transcripts of π come from some space of messages $\mathcal{M}$ and for any $m \in \mathcal{M}$ let m_i be the i-th message sent. For ease of notation, we prepend the shared random string to the start of the protocol's transcript and denote it by m_0. Given a protocol π, we use $\pi(xy)$ to refer to the random variable for the output of π on input x and y; for a deterministic protocol, $\pi(xy)$ is fixed, otherwise, it depends on the randomness of π.

Let $\mathsf{suc}(f, C)$ denote the success probability of the best randomized protocol computing f with at most C bits of communication. Formally,

$$\mathsf{suc}(f, C) := \inf_{\pi : \|\pi\| \leq C} \sup_{xy} \mathbf{Pr}[f(xy) = \pi(xy)].$$

Closely related to the notion of success probability, is that of the advantage of protocols computing Boolean functions. For a Boolean function f, the advantage of f for protocols of length at most C is defined as

$$\mathsf{adv}(f, C) := \sup_{\pi : \|\pi\| \leq C} \inf_{xy} \mathbf{E}[(-1)^{f(xy) \oplus \pi(xy)}].$$

We recall the following two protocols from prior work, which we use during the compression step in Chapter 3. The first is based on a protocol in [10] and appears as Lemma 43 in [68].

[1] The function f need not be Boolean in general, but in this thesis, we restrict ourselves to Boolean functions.

Lemma 2.5. *[10, 68] Let u, v denote two distributions on some finite set $\mathcal{M}$. For every $\varepsilon > 0$, there is a 1-round protocol distribution $\psi(uvs)$ (here uv correspond to the inputs of the protocol, and s corresponds to the transcript), and functions $a(us) \in \mathcal{M}, b(vs) \in \mathcal{M} \cup \{\perp\}$ with $\perp \notin \mathcal{M}$ such that $\psi(uv)$ is supported on all pairs uv and, for every uv and $z \in \mathcal{M}$,*

1. $\psi(a(us) = z|uv) = u(z)$,
2. $\psi(a(us) \neq b(vs)|uv, a(us)) \leq \varepsilon + \max\{0, 1 - 2^L \cdot \frac{v(a(us))}{u(a(us))}\}$.
3. $\psi\left(b(vs) \notin \{a(us), \perp\}|uv\right) \leq \varepsilon$.

Moreover, the communication complexity of ψ is $L + \log\log 1/\varepsilon + \log 1/\varepsilon + O(1)$.

The next lemma appears as Lemma 4.14 in [6].

Lemma 2.6. *[6] There is a randomized protocol τ with communication complexity at most $O(\log(C/\varepsilon))$ such that on input two C-bit strings m^A, m^B, τ outputs the first index $i \in [C]$ such that $m_i^A \neq m_i^B$ with probability at least $1 - \varepsilon$, if such an i exists.*

2.3 Linear Algebra

In this section we gather some facts regarding communication complexity and its connections to the rank of matrices, whose proofs can be found in [52]. First, we recall that rank is sub-additive.

Fact 2.7. *For two matrices A_1 and A_2, we have $\mathsf{rk}(A_1 + A_2) \leq \mathsf{rk}(A_1) + \mathsf{rk}(A_2)$.*

For a function $f(xy)$, we denote the matrix corresponding to f as M_f whose xy-th entry is simply $(-1)^{f(xy)}$. The rank of f, denoted $\mathsf{rk}(f)$ is the rank of the matrix M_f. We note that the communication complexity of a function is at least the logarithm of rank of the corresponding matrix.

Fact 2.8. *For any function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$, we have $D(f) \geq \lceil \log \mathsf{rk}(f) \rceil$.*

Lastly, we need the fact that a protocol with a small number of leaves can be simulated by a short protocol (see [52] Theorem 1.7).

Fact 2.9. *Given a protocol π with ℓ leaves, there exists a protocol with communication at most $\lceil 2\log_{3/2}\ell\rceil$ that outputs $\pi(x,y)$ on inputs x and y.*

2.4 Boolean Function Complexity Measures

In this section, we recall some concrete complexity measures of Boolean Functions, such as sensitivity, degree and decision-tree complexity as well as the relationships between them. For a detailed background, we refer the reader to the survey of Buhrman and de Wolf [14].

The sensitivity of a Boolean function $g:\{0,1\}^n\to\{0,1\}$ at a point z is defined as

$$\mathsf{s}_z(g) := |\{i : g(z) \neq g(z_{<i}, 1-z_i, z_{>i})\}|,$$

and the sensitivity of g is $\mathsf{s}(g) = \max_z \mathsf{s}_z(g)$.

A decision tree of depth d is an adaptive (deterministic) query algorithm, making at most d queries to compute a given function. The algorithm queries variables $x_{i_1},\ldots,x_{i_d}$ adaptively and outputs a bit based on the values of the variables it has queried. We say that a decision tree computes a function g, if on every input x, the algorithm outputs $g(x)$. The decision tree complexity of g, denoted $\mathsf{DT}(g)$ is the least depth of a decision tree among those that compute g.

We also recall that for every function $g:\{0,1\}^n\to\{0,1\}$, there exists a unique real, multilinear polynomial

$$q(x) = \sum_{S\subseteq[n]} c_S \cdot \prod_{i\in S} x_i,$$

such that $q(x)=g(x)$ for all $x\in\{0,1\}^n$. The degree of g, denoted $\mathsf{deg}(g)$, is the degree of q.

The above complexity measures are known to be related to each other up to polynomial factors. In particular, we know that for any f,

$$\mathsf{deg}(f), \mathsf{s}(f) \leq \mathsf{DT}(f), \tag{2.2}$$

$$\sqrt{\mathsf{deg}(f)} \leq \mathsf{s}(f) \leq 2\cdot\mathsf{deg}(f)^2, \text{ and} \tag{2.3}$$

$$\mathsf{DT}(f) \leq 2\cdot\mathsf{deg}(f)^3. \tag{2.4}$$

In the preceding facts, Equation (2.2) is due to Nisan and Szegedy [45] (see also [44]). The lower bound in Equation (2.3) is due to Huang [25], and the upper bound is again due to Nisan and Szegedy [45]. Lastly, Equation (2.4) was shown by Midrijānis [42].

Chapter 3

MARGINAL INFORMATION AND A STRONG XOR LEMMA

In this chapter, we state a strong XOR lemma for randomized communication complexity, using a new notion of information, called *marginal information*. A wide variety of important lower bounds in computer science ultimately rely on information-theoretic lower bounds in communication complexity, including lower bounds on the depth of monotone circuits [34], lower bounds on data structures [47] and lower bounds on the extension complexity of polytopes [7, 57, 62, 33], to name a few nice examples.

For deterministic communication complexity, Feder, Kushilevitz, Naor and Nisan [18] proved that if $|\mathcal{X}|, |\mathcal{Y}| \leq 2^\ell$ and f requires C bits of communication, then f^n requires at least $n(\sqrt{C} - \log_2 \ell - 1)$ bits of communication.

For randomized communication, we can ask the following question: if f requires C bits to be computed with probability $2/3$ then how much communication is required to compute f^n with probability $2/3$? This is known as the *direct sum problem* for randomized communication complexity. We note that there is a natural protocol for f^n – simply repeat the best protocol for f on each instance. This communicates at most $n \cdot C$ bits and in the worst case, succeeds only if all instances succeed, which happens with probability $(1/3)^n$. The stronger version of the direct sum, known as the strong direct product problem, asks if the success probability decays exponentially even ?

More generally, we can ask about the randomized communication complexity of computing compositions: what is the communication required to compute $g \circ f$ with constant (or smaller) probability? In the special case where g is the parity function ($\oplus$), an assertion that computing $f^{\oplus n}$ is significantly harder than computing f is called an XOR Lemma. Note that $f^{\oplus n}$ can trivially be computed with probability $1/2$ – simply output a random bit. The naive protocol that repeatedly computes f on each instance and then computes the

parity, succeeds with probability $1/2 + 3^{-n}/2$. This can be seen by an inductive argument:

$$\begin{aligned}\mathbf{Pr}[\text{computing } f^{\oplus n}] &= \mathbf{Pr}[\text{computing } f^{\oplus n-1}] \cdot \mathbf{Pr}[\text{computing } f] \\ &\quad + (1 - \mathbf{Pr}[\text{computing } f^{\oplus n-1}]) \cdot (1 - \mathbf{Pr}[\text{computing } f]) \\ &\leq 2\,\mathbf{Pr}[\text{computing } f^{\oplus n-1}]/3 + (1 - \mathbf{Pr}[\text{computing } f^{\oplus n-1}])/3 \\ &\leq (1 + \mathbf{Pr}[\text{computing } f^{\oplus n-1}])/3 \leq 1/2 + 3^{-n}/2.\end{aligned}$$

We point the reader to Section 2.2 for the notation and definitions of different quantities in randomized communication complexity such as advantage and success probability. Our main result is as follows.

Theorem 3.1. *There is a universal constant $\kappa > 0$ such that if $C > 1/\kappa$ and $\mathsf{adv}(C, f) < 1/2$, then*

$$\mathsf{adv}\Big(\frac{\kappa C\sqrt{n}}{\log(Cn)}, f^{\oplus n}\Big) < \exp(-\kappa n).$$

The constant $1/2$ is not important, it can be replaced by any constant less than 1. We note that a condition of the form $C > 1/\kappa$ is necessary, because if $x, y \in \{0, 1\}$ and $f(xy) = x \oplus y$, then $\mathsf{adv}(1, f) = 0$, yet $\mathsf{adv}(2, f^{\oplus n}) = 1$.

Prior to this result, the best known upper bound was proved by Barak, Braverman, Chen and Rao [6], who showed that the advantage is at most $1/2$ for a similar choice of the other parameters. Our work builds on the work of Yu [68], who proved exponentially small bounds on the advantage in the setting of bounded-round communication protocols. Using similar ideas we can prove several other results similar to Theorem 3.1.

Next, we give an overview of the past work that led us to the notion of marginal information, explain the intuitions behind the choices made in the definition, and then describe all of our results in Section 3.2.

3.1 The evolution of information complexity

Our definition of marginal information is the most recent advance in an evolution of definitions about information. Using the new definition of marginal information we relate bounds on the communication and advantage for computing f to the corresponding parameters for $f^{\oplus n}$ via a scheme that has been applied many times before. We prove:

Step 1 Every protocol computing $f^{\oplus n}$ with significant advantage and small communication has small marginal information; see Theorem 3.5.

Step 2 Marginal information is subadditive, so the marginal information for computing f is smaller by a factor of n; see Theorem 3.6.

Step 3 Small marginal information can be compressed to give protocols with small communication; see Theorems 3.7 to 3.10.

Definitions of information are famously subtle. In order to make this strategy work, the marginal information needs to permit all 3 steps, and even minor changes to the definition can make one of the steps infeasible.

Our current definition builds on important insights and intuitions developed in theoretical computer science over a period of decades. An early precursor to the use of information theory in computer science is the work of Kalyanasundaram and Schnitger, who used Kolmogorov complexity to prove lower bounds on the randomized communication complexity of the disjointness function [58]. The proof was subsequently simplified by Razborov [55], who gave a beautiful short argument that used Shannon's notion of entropy [59] and implicitly followed the outline of the steps 1,2,3 described above. This is related to the questions we study here because the disjointness function can be thought of as a way to compute the AND of 2 bits n times. Step 1 is relatively easy for this problem. Step 2 involved a clever way to split the dependence between random variables, and was accomplished using the sub-additivity of entropy. Step 3 is also not too difficult.

3.1.1 Parallel Repetition

The next chapter of the story was written during the study of parallel repetition, a vital tool in the development of probabilistically checkable proofs. Raz [53] proved the first exponentially small bounds in this context using the KL-divergence as a measure of information.

Given a distribution $p(xy)$, and a carefully chosen event W, Raz measured the divergence

$$\begin{aligned}&\mathop{\mathbf{E}}_{p(xy|W)}\Big[\mathbf{D}(p(x|yW)||p(x|y))+\mathbf{D}(p(y|xW)||p(y|x))\Big]\\&=\mathop{\mathbf{E}}_{p(xy|W)}\Big[\log\Big(\frac{p(x|yW)}{p(x|y)}\cdot\frac{p(y|xW)}{p(y|x)}\Big)\Big].\end{aligned}\tag{3.1}$$

In the proof, it is crucial that the event W is *rectangular*, meaning that if x, y are independent, then they remain independent even after conditioning on W. Once again, Step 1 is not too difficult. Raz used the sub-additivity of divergence and a similar set of clever random variables as in [55] to split the dependence and accomplish Step 2. Later, Holenstein [24] introduced a method called *correlated sampling* to simplify the analogue of Step 3 in Raz's proof, and obtained better bounds. Rao [49] used these tools to prove optimal bounds for parallel repetition in the setting relevant to probabilistically checkable proofs.

3.1.2 Direct Sums via Internal and External Information

Chakrabarti, Shi, Wirth and Yao [15] were the first to propose using general measures of information complexity to address the questions we consider in this paper. For inputs xy and transcripts m, we denote by $p(xym)$ the joint distribution induced by the protocol[1] p. Chakrabarti et al. [15] proposed to measure the mutual information

$$\mathbf{I}(M:XY)=\mathop{\mathbf{E}}_{p(xym)}\Big[\log\frac{p(xy|m)}{p(xy)}\Big].$$

Years later, this measure was renamed *external information* by [6]. The external information measures the information learned by an external observer about the parties' inputs. Step 1 is easy for this measure of information. However, the sub-additivity of Step 2 does not hold in general; the proof only goes through when the input distribution $p(xy)$ is a product distribution. Jain, Radhakrishnan and Sen [31], and Harsha, Jain, McAllester and Radhakrishnan [23] gave ways to implement Step 3 that led to bounds on the success probability for computing f^n in the setting where the inputs are assumed to come from a product distribution and the communication protocols are restricted to having a bounded number of rounds. Meanwhile, Bar-yossef, Jayram, Kumar and Sivakumar [4] showed how

[1] We often say $p(xym)$ is a protocol when we mean that it is a distribution induced by a protocol.

to reframe Razborov's proof using mutual information instead of entropy, and proved other results using this formulation which contained hints of the definition of information that came next.

The first upper bounds on the success probability in the general setting came when Barak, Braverman, Chen and Rao [6] adapted the methods developed in the study of parallel repetition to these problems. In contrast with the external information, they defined the *internal information*, which is the sum of two mutual information terms

$$\mathbf{I}(M : X|Y) + \mathbf{I}(M : Y|X) = \mathop{\mathbf{E}}_{p(xym)} \Big[\log \Big(\frac{p(x|ym)}{p(x|y)} \cdot \frac{p(y|xm)}{p(y|x)} \Big) \Big]. \tag{3.2}$$

The internal information measures what is learned by each party about the other's input. Equation (3.1) was the inspiration for Equation (3.2); indeed, each setting of m corresponds to a rectangular event. When the inputs come from a product distribution, the internal and external information are the same, and [6] proved that sub-additivity holds for internal information using an argument similar to the one used in the context of parallel repetition. Moreover, they showed how to leverage the technique of correlated sampling developed by Holenstein to simulate protocols with information I and communication C using $\approx \sqrt{IC}/\log C$ communication. They gave near optimal simulations of $\approx I \log^2 C$ for protocols with small external information using rejection sampling and a variant of Azuma's concentration inequality. These results proved that there is a constant κ such that if $\mathsf{adv}(C, f) < 1/2$, then

$$\mathsf{adv}\Big(\frac{\kappa C \sqrt{n}}{\log(Cn)}, f^{\oplus n} \Big) < 1/2,$$

which was the first result along the lines of Theorem 3.1. Later, Braverman and Rao [10] showed that the internal information cost of a function is equal to the amortized communication complexity of that function, suggesting that this is the *right* definition of information in the interactive setting.

Since then, several researchers have studied the problem of compressing protocols with internal information I and communication C. Braverman [8] showed how to obtain protocols with communication $\approx 2^{O(I)}$. Ramamoorthy and Rao [48] showed that if I_A, I_B denote the internal information learned by each party, then you can achieve communication $\approx I_A \cdot 2^{O(I_B)}$

and can also achieve communication $\approx I_A + \sqrt[4]{I_B \cdot C^3}$. The work of Sherstov [60], building on the work of Kol [36], showed that $\approx I \log^2 I$ communication can be achieved when the inputs come from a product distribution. Ganor, Kol and Raz [19] (see also [50]) gave a nice counterexample: a function that can be computed with communication $\approx 2^{2^{O(I)}}$, and internal information $\approx I$, but cannot be computed with communication $\approx 2^I$.

3.1.3 Direct Products via New Information Measures and Proxy Distributions

The next definition to evolve was proposed by Braverman, Weinstein, Rao and Yehudayoff [12, 11], inspired by the work of Jain, Pereszlényi and Yao [30]. Rather than bounding the information under the distribution $p(xym)$, they bounded the infimum of information achieved in the ball of distributions that are close to the protocol. They defined the information to be the infimum

$$\inf_q \mathbf{I}_q(M : X|Y) + \mathbf{I}_q(M : Y|X) = \inf_q \mathop{\mathbf{E}}_{q(xym)} \left[\log \left(\frac{q(x|ym)}{q(x|y)} \cdot \frac{q(y|xm)}{q(y|x)} \right) \right], \tag{3.3}$$

where here the infimum is taken over all distributions $q(xym)$ that are close to $p(xym)$ in statistical distance. This quantity was ultimately bounded by setting $q(xym) = p(xym|W)$, where here W is a reasonably large event (not necessarily rectangular) that implies that the protocol correctly computes the function. The bound on Equation (3.3) does not lead to a bound on the information according to $p(xym)$, because it is quite possible that the points outside W reveal a lot of information. Still, [12] were able to follow all 3 steps of the high-level approach to prove their results. Step 1 remained easy, but Steps 2 and 3 became more difficult using Equation (3.3). [12] obtained exponentially small upper bounds for the success probability of computing f^n, but did not manage to prove new bounds on the advantage for $f^{\oplus n}$ using this approach. Equation (3.3) may not seem very different from Equation (3.2), but it does involve a proxy q, and we pursue the use of such proxies further in the definition of marginal information that we discuss next.

3.1.4 XOR Lemma via Marginal Information

In a paper full of new ideas, Yu [68] recently proved exponentially small bounds on the advantage of bounded-round protocols computing $f^{\oplus n}$. Although Yu's paper involves a

potential function that superficially looks like a definition of information, his proof does not involve a method to compress protocols whose potential is small, and we are unable to extract a definition of information from his work. Still, his ideas inspired many of the choices made in our definition. To define the marginal information, we need the concept of a rectangular distribution, which was defined in [68]:

Definition 3.2. *Given a set Q consisting of triples (xym), we say that Q is* rectangular *if its indicator function can be expressed as*

$$1_Q(xym) = 1_A(xm) \cdot 1_B(ym),$$

for some Boolean functions $1_A, 1_B$. Given a distribution $q(xym)$ and a distribution $\mu(xy)$, we say that q is rectangular with respect to μ *if it can be expressed as*

$$q(xym) = \mu(xy) \cdot A(xm) \cdot B(ym),$$

for some functions A, B.

For intuition, it is helpful to think of a rectangular distribution as the result of conditioning a protocol distribution $p(xym)$ on a rectangular event. That would produce a rectangular distribution, but the space of rectangular distributions actually contains other distributions that cannot be obtained in this way.

From our perspective, the most useful insight of Yu's work is that if q is restricted to being rectangular, then one can allow q to be quite far from p in Equation (3.3) and still carry out a meaningful compression of a protocol p to implement Step 3. That is because the rectangular nature of q allows the parties to use hashing and rejection sampling to convert a protocol that samples from p into a protocol that samples from q. If $q(xym) = p(xym|R)$ for a rectangular event R, this is easy to understand: the parties can communicate 2 bits to compute if $xym \in R$ and output the most likely value of f under q with $xym \in R$. If $xym \notin R$ they can output a random guess for the value of f. So, it is enough to bound the information terms for $xym \in R$, and enough to guarantee that the compression is efficient for such points. This observation is very powerful, because it allows us to throw away problematic points in the support of the distributions we are working with and pass to appropriate sub-rectangles throughout our proofs.

For all of this to work, it is crucial that the protocol retains some advantage within the support of q. For this reason, we need to keep track of the information in the support of q as well as the advantage within the support of q, and so, for the first time, the measure of information is going to depend on the function f that the protocol computes. We are ready to state the definition:

Definition 3.3 (Marginal Information)**.** *For $I \geq 1$ and[2] $\delta = 1/15$, the* marginal information *of a protocol p for computing f is defined as*

$$\mathsf{M}_I(p,f) = \inf_q \sup_{xym} \log\Big(\frac{q(x|ym)}{p(x|y)} \cdot \frac{q(y|xm)}{p(y|x)} \cdot \Big(\frac{q(xym)}{p(xym)}\Big)^I \cdot \Big|\mathop{\mathbf{E}}_{q(x'y'|m)}[(-1)^{f(x'y')}]\Big|^{-12I/\delta}\Big),$$

where the infimum is taken over all distributions q that are rectangular with respect to the input distribution $p(xy)$, and the supremum is taken over all xym in the support of q.

We use the letter I above because it turns out that protocols computing f can be efficiently compressed when $\mathsf{M}_I = O(I)$, and any compression must have communication $\Omega(I)$. Compare Definition 3.3 with Equations (3.2) and (3.3). The fact that q must be tethered to p is ensured by including the term $q(xym)/p(xym)$. If $q(xym) = p(xym|R)$ for a rectangular event R, $q(xym)/p(xym)$ will be equal to $1/p(R)$. The last term in the product computes the advantage of q for computing f, because under q and given m, the best guess for the value of f is determined by the sign of $\mathbf{E}_{q(xy|m)}[(-1)^{f(xy)}]$, and its advantage is the absolute value of this quantity. In words, the marginal information measures the supremum over all xym of the information per unit of advantage, of the best rectangular approximation q.

In analogy with the external information, we define the external marginal information:

Definition 3.4. *For $I \geq 1$ and $\delta = 1/15$, the* external marginal information *of a protocol p for computing f is defined as:*

$$\mathsf{M}_I^{\mathsf{ext}}(p,f) = \inf_q \sup_{xym} \log\Big(\frac{q(xy|m)}{p(xy)} \cdot \Big(\frac{q(xym)}{p(xym)}\Big)^I \cdot \Big|\mathop{\mathbf{E}}_{q(x'y'|m)}[(-1)^{f(x'y')}]\Big|^{-12I/\delta}\Big),$$

where the infimum is taken over all distributions q that are rectangular with respect to the input distribution $p(xy)$, and the supremum is taken over all xym in the support of q.

[2]Even though δ is a fixed constant, we choose to write it in the definition because it eases the notation throughout the paper.

We prove that the external marginal information is equal to the marginal information when the distribution on inputs is a product distribution in Lemma 3.21.

To state our results about marginal information, we first define the average-case measure of advantage. Given a distribution $\mu(xy)$ on inputs, define

$$\mathsf{adv}_\mu(C, f) = \sup_{\|\pi\| \le C} \mathbf{E}[(-1)^{\pi(xy)+f(xy)}],$$

where here the expectation is over the choice of inputs xy as well as the random coins of the communication protocol. To study the more restricted setting where the protocols we are working with have a bounded number of rounds, define the worst-case and average case quantities:

$$\mathsf{adv}^r(C, f) = \sup_{\|\pi\| \le C} \inf_{xy} \mathbf{E}[(-1)^{\pi(xy)+f(xy)}],$$

$$\mathsf{adv}^r_\mu(C, f) = \sup_{\|\pi\| \le C} \mathbf{E}[(-1)^{\pi(xy)+f(xy)}],$$

where throughout, the suprema are taken over r-round protocols.

Returning to our high-level approach, we prove the following results about marginal information, which allow us to carry out Steps 1,2,3:

1. In Section 4.1, we show that a protocol with small communication and large advantage has small marginal information, to handle Step 1:

   Theorem 3.5. *For every Boolean function $f(xy)$ and every protocol p of communication complexity C,*

   $$\mathsf{M}_I(p, f) \le 2C - (1 + 12/\delta) \cdot I \cdot \log \left(\mathop{\mathbf{E}}_{p(m)} \left| \mathop{\mathbf{E}}_{p(xy|m)} \left[(-1)^{f(xy)} \right] \right| \right) + O(I).$$

   For any fixed m, the quantity $|\mathbf{E}_{p(xy|m)}[(-1)^{f(xy)}]|$ measures the advantage of the protocol for computing f conditioned on that value of m. So, if $\mathsf{adv}_\mu(C, f^{\oplus n}) \ge \exp(-n)$ via a protocol corresponding to the distribution p, then the above theorem implies that $\mathsf{M}_I(p, f^{\oplus n}) \le O(C + In)$. Unlike all previous definitions, for marginal information Step 1 involves significant work. Our proof crucially uses the fact that the protocol has bounded communication complexity: for example it would not be enough to start with a bound on the internal information.

2. In Section 4.2, we prove that marginal information is sub-additive with respect to the n-fold XOR of f. If the transcript $m = (m_0, m_1, \ldots, m_C)$, where m_j denotes the j-th message of the protocol, we show

Theorem 3.6. *There is a universal constant Δ such that if $I \geq 1$ and p is a protocol distribution for computing $f^{\oplus n}$ with $p(xy) = \prod_{i=1}^{n} p(x_i y_i)$, then there is a protocol p_i for computing f such that $p_i(x_i y_i) = p(x_i y_i)$, p_i has the same number of messages as p, for $j > 1$ the support of m_j is identical in p_i and p, and moreover,*

$$\mathsf{M}_I(p_i, f) \leq \frac{\mathsf{M}_I(p, f^{\oplus n})}{n} + \Delta I \cdot \left(1 + \log \frac{\mathsf{M}_I(p, f^{\oplus n})}{n \cdot I}\right).$$

If $\mathsf{M}_I(p, f^{\oplus n}) \leq O(In)$, this theorem proves that $\mathsf{M}_I(p_i, f) \leq O(I)$. This might well be the most technically novel part of our proof; it is certainly where we spent the most time. The main challenge is proving the result for $n = 2$, which is very delicate. This case is captured by Theorem 4.2, and Theorem 3.6 is a straightforward consequence. If $n = 2$ and $\mathsf{M}_I(p, f^{\oplus 2})$ is small, then there is a rectangular distribution q such that the pair

$$q(x_1 x_2 y_1 y_2 m), p(x_1 x_2 y_1 y_2 m)$$

leads to a small value of $\mathsf{M}_I(p, f^{\oplus 2})$. We show how to use q, p to generate a new pair

$$q_1(x_1 y_1 m^{(1)}), p_1(x_1 y_1 m^{(1)})$$

or a new pair

$$q_2(x_2 y_2 m^{(2)}), p_2(x_2 y_2 m^{(2)})$$

proving that either $\mathsf{M}_I(p_1, f)$ or $\mathsf{M}_I(p_2, f)$ is more or less bounded by $\mathsf{M}_I(p, f^{\oplus 2})/2$. A significant first step is the construction of two pairs of rectangular/protocol distributions with the properties described in Equations (4.6) to (4.9). Given this step, we need to eliminate various problematic points from the support of the distributions while preserving the rectangular nature of the distribution to ultimately construct the promised pair of distributions.

We are unable to bound the length of the first message of p_i in terms of the length of the corresponding message of p in Theorem 3.6, because in our proof of Theorem 4.2

the first message $m_1^{(1)}$ or $m_1^{(2)}$ needs to encode one of the inputs of the original protocol. Fortunately, this is not a significant obstacle for the high-level strategy.

3. In Sections 4.3 and 4.5 to 4.7, we show how to compress marginal information to handle Step 3. We have been able to match many of the prior results [6, 10, 8] about compressing information and external information with corresponding results about compressing marginal information and external marginal information, though our proofs are much more technical. Our most general simulation is captured by the following theorem:

Theorem 3.7. *For every $\alpha > 0$ there is a $\Delta > 0$ such that if $\mathsf{M}_I(p, f) \leq \alpha I$, $\mu(xy) = p(xy)$ and moreover the messages $m = (m_0, \ldots, m_C)$ are such that $m_2, \ldots, m_C \in \{0, 1\}$, then $\mathsf{adv}_\mu(\Delta(I + \sqrt{CI}\log(CI)), f) \geq 1/\Delta$.*

Theorem 3.7 shows that if the marginal information is $O(I)$, then one can obtain a protocol with communication $\widetilde{O}(\sqrt{CI})$ that has $\Omega(1)$ advantage for computing f. For the external marginal information, we prove:

Theorem 3.8. *For every $\alpha > 0$, there is a $\Delta > 0$ such that if $\mathsf{M}^{\mathsf{ext}}_I(p, f) \leq \alpha I$, $\mu(xy) = p(xy)$, and moreover the messages $m = (m_0, \ldots, m_C)$ are such that $m_2, \ldots, m_C \in \{0, 1\}$, then $\mathsf{adv}_\mu(\Delta I \log^2 C, f) \geq 1/\Delta$.*

This theorem gives improved results when the inputs come from a product distribution. It is quite possible that even better simulations can be obtained using the ideas of [36, 60, 9], but we have not managed to obtain such results. We also obtain results that are independent of the communication complexity:

Theorem 3.9. *For every $\alpha > 0$, there is a $\Delta > 0$ such that if $\mathsf{M}_I(p, f) \leq \alpha I$ and $\mu(xy) = p(xy)$, then $\mathsf{adv}_\mu(\Delta I, f) \geq \exp(-\Delta I)$.*

When the number of rounds of the protocol is bounded, we prove:

Theorem 3.10. *For every $\alpha > 0$, there is a $\Delta > 0$ such that if $\mathsf{M}_I(p, f) \leq \alpha I$, $\mu(xy) = p(xy)$, p has r-rounds and $m_r \in \{0, 1\}$, then $\mathsf{adv}^r_\mu(\Delta r(I + \log r), f) \geq 1/\Delta$.*

These results about the marginal information cost allow us to prove Theorem 3.1, as well as several other results of that flavor.

3.2 Using marginal information to prove XOR lemmas

To state all of our results, let us define the average-case and worst-case measures of success. Similar to the definition of $\mathsf{suc}(C, f)$ in Section 2.2, we denote

$$\begin{aligned}
\mathsf{suc}^r(C, f) &= \sup_{\|\pi\| \leq C} \inf_{xy} \mathbf{Pr}[\pi(xy) = f(xy)] \\
\mathsf{suc}_\mu(C, f) &= \sup_{\|\pi\| \leq C} \mathbf{Pr}[\pi(xy) = f(xy)] \\
\mathsf{suc}^r_\mu(C, f) &= \sup_{\|\pi\| \leq C} \mathbf{Pr}[\pi(xy) = f(xy)],
\end{aligned}$$

where in $\mathsf{suc}^r, \mathsf{suc}^r_\mu$ the supremum is taken over r-round protocols, and in $\mathsf{suc}_\mu, \mathsf{suc}^r_\mu$ the probability is over inputs sampled from $\mu(xy)$. Yao's min-max theorem yields

$$\begin{aligned}
\mathsf{adv}(C, f) &= \inf_\mu \mathsf{adv}_\mu(C, f), \\
\mathsf{suc}(C, f) &= \inf_\mu \mathsf{suc}_\mu(C, f), \\
\mathsf{adv}^r(C, f) &= \inf_\mu \mathsf{adv}^r_\mu(C, f), \\
\mathsf{suc}^r(C, f) &= \inf_\mu \mathsf{suc}^r_\mu(C, f). \qquad (3.4)
\end{aligned}$$

Given any distribution μ on $\mathcal{X} \times \mathcal{Y}$, define the n-fold product distribution μ^n on $\mathcal{X}^n \times \mathcal{Y}^n$ by $\mu^n(xy) = \prod_{j=1}^n \mu(x_j y_j)$. Theorem 3.1 is proved by proving this stronger bound:

Theorem 3.11. *There is a universal constant $\kappa > 0$ such that if $C > 1/\kappa$ and $\mathsf{adv}_\mu(C, f) \leq \kappa$, then $\mathsf{adv}_{\mu^n}(\kappa C\sqrt{n}/\log(Cn), f^{\oplus n}) \leq \exp(-\kappa n)$.*

To prove Theorem 3.11, suppose that there is a protocol p computing $f^{\oplus n}$ with advantage $\exp(-\kappa n)$ and communication $T = \kappa C \cdot \sqrt{n}/\log(Cn)$. If $T/n \geq 1$, we set $I = T/n$ and apply Theorem 3.5 to show that $\mathsf{M}_I(p, f^{\oplus n}) \leq O(T + \kappa I n) \leq O(In)$. Next, apply Theorem 3.6 to

find a protocol p' with $\mathsf{M}_I(p', f) \leq O(I)$. Finally, apply Theorem 3.7 to obtain a protocol computing f with advantage $\Omega(1)$ and communication proportional to

$$\frac{T}{n} + 2\sqrt{IT}\log(T) \leq \frac{T}{n} + 2\frac{T\log T}{\sqrt{n}}$$
$$\lesssim \frac{\kappa C}{\log nC} \cdot \log T \lesssim \kappa C.$$

If $T/n < 1$, set $I = 1$ and apply Theorem 3.5 to show that $\mathsf{M}_I(p, f^{\oplus n}) \leq O(In)$. Next, apply Theorem 3.6 to find a protocol p' with $\mathsf{M}_I(p', f) \leq O(I) = O(1)$. Finally, we apply Theorem 3.9 to obtain a protocol computing f with advantage $\Omega(1)$ and communication $O(1)$. Setting κ sufficiently small, we obtain a contradiction in either case, which proves that there is no protocol p as above. Theorem 3.1 can be obtained from Theorem 3.11 using Equation (3.4) and the fact that the worst-case success probability of a communication protocol can be increased by taking the majority outcome of several runs of the protocol. We leave these details to the reader.

Theorems 3.1 and 3.11 yield bounds on the success probability for computing f^n as well:

Corollary 3.12. *There is a universal constant $\kappa > 0$ such that if $C > 1/\kappa$ and $\mathsf{adv}(C, f) < \kappa$, then $\mathsf{suc}(\kappa C\sqrt{n}/\log(Cn)), f^n) < \exp(-\kappa n)$.*

Corollary 3.13. *There is a universal constant $\kappa > 0$ such that if $C > 1/\kappa$ and $\mathsf{adv}_\mu(C, f) < \kappa$, then $\mathsf{suc}_{\mu^n}(\kappa C\sqrt{n}/\log(Cn)), f^n) < \exp(-\kappa n)$.*

This matches the result proved by [12] mentioned earlier. These corollaries are obtained by observing that if $S \subseteq \{1, 2, \ldots, n\}$ is chosen uniformly at random, and xy are sampled according to μ^n, then

$$\mathbf{E}\left[(-1)^{\sum_{j\in S}\pi(xy)_j + f(x_j y_j)}\right] = \mathbf{Pr}[\pi(xy) = f^n(xy)],$$

so a protocol computing f^n with success probability $\exp(-n/2)$ yields a set of $n' = \Omega(n)$ coordinates where the protocol computes $f^{\oplus n'}$ with advantage $\exp(-\Omega(n))$. Again, we leave the details to the reader. When the distribution $\mu(xy) = \mu(x)\cdot\mu(y)$ is a product distribution, we obtain stronger bounds:

Theorem 3.14. *There is a universal constant $\kappa > 0$ such that for every product distribution μ, if $C > 1/\kappa$ and $\mathsf{adv}_{\mu}(C, f) < \kappa$, then $\mathsf{adv}_{\mu^n}(\kappa Cn/\log^2(Cn), f^{\oplus n}) < \exp(-\kappa n)$.*

To prove Theorem 3.14, suppose we are given a protocol p computing $f^{\oplus n}$ with advantage $\exp(-\kappa n)$ and communication $T = \kappa Cn/\log^2(Cn)$. If $T/n \geq 1$, we set $I = T/n$ and apply Theorem 3.5 to show that $\mathsf{M}_I(p, f^{\oplus n}) \leq O(nI)$. Next, apply Theorem 3.6 to find a protocol p' with $\mathsf{M}_I(p', f) \leq O(I)$. Finally, using the fact that for product distributions, $\mathsf{M}_I^{\mathsf{ext}}(p, f) = \mathsf{M}_I(p, f)$, we can apply Theorem 3.8 to obtain a protocol computing f with advantage $\Omega(1)$ and communication $O(I \log^2(Cn)) \leq O(\kappa C)$. Otherwise, if $T/n < 1$, set $I = 1$ and apply Theorem 3.5 to show that $\mathsf{M}_I(p, f^{\oplus n}) \leq O(n)$. Then, apply Theorem 3.6 to find a protocol p' with $\mathsf{M}_I(p', f) \leq O(I) = O(1)$. Lastly, we apply Theorem 3.9 to obtain a protocol computing f with advantage $\Omega(1)$ and communication $O(1)$. Setting κ to be small enough gives a contradiction in either case.

As before, this yields a corollary for computing f^n:

Corollary 3.15. *There is a universal constant $\kappa > 0$ such that for every product distribution μ, if $C > 1/\kappa$ and $\mathsf{adv}_{\mu}(C, f) < \kappa$, then $\mathsf{suc}_{\mu^n}(\kappa Cn/\log^2(Cn), f^n) < \exp(-\kappa n)$.*

Again, this is identical to a bound proved by [12] using a different approach. For the bounded-round setting, we prove:

Theorem 3.16. *There is a universal constant $\kappa > 0$ such that if $C > (r(\log r) + 1)/\kappa$, and $\mathsf{adv}^r_{\mu}(C, f) < \kappa$, then $\mathsf{adv}^r_{\mu^n}((\kappa C/r - \log r)n, f^{\oplus n}) < \exp(-\kappa n)$.*

Yu [68] proves the same bound on the advantage with a communication budget that grows like $\Omega((C/r^r - O(1))n)$. Our bound eliminates the exponential dependence on r. To prove Theorem 3.14, set $T = (\kappa C/r - \log r)n$, and suppose there is a protocol computing f with r rounds, communication T and advantage $\exp(-\kappa n)$. Set $I = T/n \geq 1$. Then, M_I can be bounded by $O(T + \kappa In)$ by Theorem 3.5. Applying Theorem 3.6 gives an r-round protocol with M_I bounded by $O(I)$, and applying Theorem 3.10 gives an r-round protocol with communication complexity $O(r(I + \log r)) = O(\kappa C)$ computing f with advantage $\Omega(1)$. Setting κ to be small enough proves the result. As usual, we obtain the following corollaries:

Corollary 3.17. *There is a universal constant $\kappa > 0$ such that if $C > 7(r \log r)/\kappa$ and $\mathsf{adv}^r_\mu(C, f) < \kappa$, then $\mathsf{suc}^r_{\mu^n}((\kappa C/r - \log r)n, f^n) < \exp(-\kappa n)$.*

Corollary 3.18. *There is a universal constant $\kappa > 0$ such that if $C > 7(r \log r)/\kappa$, and $\mathsf{adv}^r(C, f) < \kappa$, then $\mathsf{suc}^r((\kappa C/r - \log r)n, f^n) < \exp(-\kappa n)$.*

3.3 Organization

In the next section, we setup some notation and record some basic properties of marginal information Definition 3.3. In Section 3.5 we gather several results related to the *trimming* technique borrowed from [68] that will be repeatedly used in the proofs of Sections 4.1 and 4.2. In Section 3.6 we gather several consequences of small marginal information that are used to analyze our compression schemes.

In the next chapter, we prove Theorems 3.5 to 3.10. We prove Theorem 3.5 in Section 4.1, Theorem 3.6 in Section 4.2, and prove the general simulation theorem for marginal information Theorem 3.7, in Section 4.3. Afterwards, in Section 4.4 we prove that if the external marginal information is small, then there is a *smooth* protocol with small external marginal information, mirroring a similar result in [6]. We then show how to compress smooth protocols to prove Theorem 3.8 in Section 4.5. We prove Theorem 3.10 in Section 4.6 and finally, in Section 4.7 we prove Theorem 3.9.

After the next section, we encourage the reader to skip to the next chapter and refer to the facts proved in Section 3.5 and Section 3.6 as and when necessary.

3.4 Definitions and Basic Properties of Marginal Information

Everywhere in the chapter, we assume that $\delta > 0$ is a sufficiently small constant; $\delta = 1/15$ will suffice.

Definition 3.19. *We say that $p(xym)$ is a protocol distribution if it can be expressed as*

$$p(xym) = p(xy) \cdot p(m_0) \cdot \prod_{i=1,3,5,\ldots} p(m_i | xm_{<i}) \cdot p(m_{i+1} | ym_{\leq i}).$$

Every randomized worst-case protocol corresponds to some protocol distribution $p(xym)$, where $p(xy)$ can be taken to be the uniform distribution on all possible inputs. Given

a distribution $\mu(xy)$ on inputs, and any protocol generating the messages m, the joint distribution of xym corresponds again to a protocol distribution $p(xym)$, with $p(xy) = \mu(xy)$.

Recall Definition 3.2. Note that if q is rectangular with respect to $\mu(xy)$ and p is a protocol with $p(xy) = \mu(xy)$, it is not necessary that $q(xy) = \mu(xy)$. For the purpose of intuition, it may be helpful to think of a rectangular distribution as the result of conditioning $\mu(xy)$ on the event that it lies in a disjoint union of rectangles indexed by m, though this statement is not without loss of generality, and we do use the full generality of Definition 3.2.

Let $x = x_1x_2$ and $y = y_1y_2$. Let $\mu(xy) = \mu(x_1y_1) \cdot \mu(x_2y_2)$ be a product distribution. It will be helpful to define $w = (x_1y_2m)$. Given $m = (m_0, \ldots, m_r)$ and y_2, we denote

$$\begin{aligned} m^{(1)} &= (m_0, y_2m_1, m_2, \ldots, m_r), \\ m^{(2)} &= (m_0x_1, m_1, m_2, \ldots, m_r). \end{aligned} \tag{3.5}$$

Let us gather some basic facts about rectangular distributions in this setting:

Proposition 3.20. *If v is rectangular, then*

1. $v(xy|w) = v(y_1|w) \cdot v(x_2|w)$,
2. $v(xw) \cdot v(yw) = v(xym) \cdot v(w)$,
3. $v(x_1|y_1m^{(1)}) \cdot v(x_2|y_2m^{(2)}) = v(x|ym)$, *and*
4. $v(y_1|x_1m^{(1)}) \cdot v(y_2|x_2m^{(2)}) = v(y|xm)$.

Proof. For the first identity, let A, B be such that $v(xym) = \mu(xy) \cdot A(xm) \cdot B(ym)$. Then

$$\begin{aligned} v(xy|w) = \frac{v(xyw)}{v(w)} &= \frac{\mu(x_1y_1) \cdot \mu(x_2y_2) \cdot A(xm) \cdot B(ym)}{\sum_{x_2'y_1'} \mu(x_1y_1') \cdot \mu(x_2'y_2) \cdot A(x_1x_2'm) \cdot B(y_1'y_2m)} \\ &= \frac{\mu(x_1y_1) \cdot B(y_1y_2m)}{\sum_{y_1'} \mu(x_1y_1') \cdot B(y_1'y_2m)} \cdot \frac{\mu(x_2y_2) \cdot A(x_1x_2m)}{\sum_{x_2'} \mu(x_2'y_2) \cdot A(x_1x_2'm)} \\ &= v(y_1|w) \cdot v(x_2|w). \end{aligned}$$

For the second identity,

$$\begin{aligned} v(xw) \cdot v(yw) &= v(w) \cdot v(yw) \cdot v(x|w) \\ &= v(w) \cdot v(yw) \cdot v(x_2|x_1ym) && \text{(by the first identity)} \\ &= v(w) \cdot v(xym). \end{aligned}$$

For the third identity,

$$\begin{aligned} v(x_1|y_1m^{(1)}) \cdot v(x_2|y_2m^{(2)}) &= v(x_1|ym) \cdot v(x_2|x_1y_2m) \\ &= v(x_1|ym) \cdot v(x_2|x_1ym) && \text{(by the first identity)} \\ &= v(x|ym). \end{aligned}$$

A similar calculation yields the fourth identity. □

It is easy to check that the external marginal information and marginal information are the same when the distribution on inputs is product:

Lemma 3.21. *If $p(xy) = \mu(xy)$ is a product distribution then we have $\mathsf{M}_I^{\mathsf{ext}}(p, f) = \mathsf{M}_I(p, f)$.*

Proof. For all rectangular q, we have

$$\begin{aligned} q(xy|m) &= \frac{q(xym)}{q(m)} \\ &= \frac{\mu(xy) \cdot A(xm) \cdot B(ym)}{\sum_{x'y'} \mu(x'y') \cdot A(x'm) \cdot B(y'm)} \\ &= \frac{\mu(x)\mu(y) \cdot A(xm) \cdot B(ym)}{\sum_{x'y'} \mu(x')\mu(y') \cdot A(x'm) \cdot B(y'm)} \\ &= \frac{\mu(x) \cdot A(xm)}{\sum_{x'} \mu(x') \cdot A(x'm)} \cdot \frac{\mu(y) \cdot B(ym)}{\sum_{y'} \mu(y') \cdot B(y'm)}, \end{aligned}$$

proving that $q(xy|m)$ is a product distribution.

Thus:

$$\frac{q(x|ym)}{p(x|y)} \cdot \frac{q(y|xm)}{p(y|x)} = \frac{q(xy|m)}{p(xy)},$$

and so $\mathsf{M}^{\mathsf{ext}}(p, f) = \mathsf{M}_I(p, f)$. □

3.5 Trimming and advantage preserving sets

In this section, we gather a few lemmas about trimming rectangular sets to pass to subrectangles with nice features. The idea of trimming comes from the work of Yu [68].

Lemma 3.22. *For every $1 > \kappa > 0$, if $a(xym), b(xym)$ are two distributions, there exists a rectangular set T such that $a(T) \geq 1 - 3\kappa$ and for all $xym \in T$, we have*

$$\frac{a(xm|T)}{b(xm)}, \frac{a(ym|T)}{b(ym)}, \frac{a(m|T)}{a(m)} \geq \kappa.$$

Proof. The set T is constructed by an iterative process. Initially, T is the set of all triples xym. In each iteration, if there is xm such that

$$\frac{a(xm|T)}{b(xm)} < \kappa, \tag{3.6}$$

then delete xm from the support of T, if there is ym such that

$$\frac{a(ym|T)}{b(ym)} < \kappa,$$

then delete ym from the support of T, and if there is m such that

$$\frac{a(m|T)}{a(m)} < \kappa,$$

then delete m from the support of T. The process halts when there are no more elements to delete. Because the distributions we are working with have finite support, this process must eventually terminate. Initially, T is rectangular, and each deletion step leaves us with another rectangular set T, so the final T is also rectangular.

Let us bound $a(T)$. For each pair xm that was deleted from the support of T because of Equation (3.6), let T_{xm} denote the set T right before xm was deleted. If xm was not deleted, let T_{xm} denote the empty set.

The total mass deleted using Equation (3.6) is exactly

$$\sum_{xm} a(xmT_{xm}) = \sum_{xm} a(T_{xm}) \cdot a(xm|T_{xm}) < \sum_{xm} \kappa \cdot b(xm) = \kappa.$$

Similarly, the total mass deleted using each of the other rules is also at most κ. By the union bound, this proves that $a(T) \geq 1 - 3\kappa$ when the process terminates. □

As alluded to in the high-level overview, in the course of the proof, we repeatedly pass through rectangular sets whose points have some nice properties[3]. While doing so, it is important for us to ensure that we do not move to a rectangular set where the advantage vanishes. As in Yu's proof, we enforce this constraint by working only with "advantage preserving sets".

Lemma 3.23. *For any distribution $v(xym)$ and a Boolean function $h(xy)$, suppose R is a rectangular set maximizing*

$$v(R)^{\delta} \cdot \mathop{\mathbf{E}}_{v(m|R)} \left| \mathop{\mathbf{E}}_{v(xy|mR)} \left[(-1)^{h(xy)}\right] \right|. \tag{3.7}$$

Then, for any rectangular $Z \subseteq R$, we have

$$\mathop{\mathbf{E}}_{v(m|Z)} \left| \mathop{\mathbf{E}}_{v(xy|mZ)} \left[(-1)^{h(xy)}\right] \right| \geq \frac{1 - \delta^2 - \delta/v(Z|R)}{v(R)^{\delta}} \cdot \mathop{\mathbf{E}}_{v(m)} \left| \mathop{\mathbf{E}}_{v(xy|m)} \left[(-1)^{h(xy)}\right] \right|.$$

Proof. Since R and Z are rectangular, we have

$$1_R(xym) = 1_A(xm) \cdot 1_B(ym),$$

and

$$1_Z(xym) = 1_{A'}(xm) \cdot 1_{B'}(ym),$$

for appropriate sets A, A' and B, B'. R can be partitioned into three rectangular sets, $Z_0 = Z, Z_1$ and Z_2, where

$$1_{Z_1}(xym) = 1_{A \setminus A'}(xm) \cdot 1_B(ym)$$

and

$$1_{Z_2}(xym) = 1_{A'}(xm) \cdot 1_{B \setminus B'}(ym).$$

By the triangle inequality, we get

$$\mathop{\mathbf{E}}_{v(m|R)} \left| \mathop{\mathbf{E}}_{v(xy|Rm)} \left[(-1)^{h(xy)}\right] \right| \leq \sum_{i=0}^{2} v(Z_i|R) \cdot \mathop{\mathbf{E}}_{v(m|Z_i)} \left| \mathop{\mathbf{E}}_{v(xy|mZ_i)} \left[(-1)^{h(xy)}\right] \right| \tag{3.8}$$

[3]For example, we would like to only consider those points where the information ratios are bounded.

Let us bound the contribution of Z_1, Z_2:

$$\begin{aligned}
&\sum_{i=1}^{2} v(Z_i|R) \cdot \mathop{\mathbf{E}}_{v(m|Z_i)} \Big| \mathop{\mathbf{E}}_{v(xy|mZ_i)} \Big[(-1)^{h(xy)}\Big]\Big| \\
&= \sum_{i=1}^{2} v(Z_i|R)^{1-\delta} \cdot \Big(\frac{v(Z_i)}{v(R)}\Big)^{\delta} \cdot \mathop{\mathbf{E}}_{v(m|Z_i)} \Big| \mathop{\mathbf{E}}_{v(xy|mZ_i)} \Big[(-1)^{h(xy)}\Big]\Big| \\
&\le \sum_{i=1}^{2} v(Z_i|R)^{1-\delta} \cdot \mathop{\mathbf{E}}_{v(m|R)} \Big| \mathop{\mathbf{E}}_{v(xy|mR)} \Big[(-1)^{h(xy)}\Big]\Big| \\
&\qquad\qquad \text{(because } R \text{ is the maximizer of Equation (3.7))} \\
&\le 2^{\delta} \cdot \Big(\sum_{i=1}^{2} v(Z_i|R)\Big)^{1-\delta} \cdot \mathop{\mathbf{E}}_{v(m|R)} \Big| \mathop{\mathbf{E}}_{v(xy|mR)} \Big[(-1)^{h(xy)}\Big]\Big| \qquad \text{(by Hölder's inequality)} \\
&= 2^{\delta} \cdot \Big(1 - v(Z|R)\Big)^{1-\delta} \cdot \mathop{\mathbf{E}}_{v(m|R)} \Big| \mathop{\mathbf{E}}_{v(xy|mR)} \Big[(-1)^{h(xy)}\Big]\Big|.
\end{aligned}$$

Using the inequalities $(1-t)^{1-\delta} \le 1 - t(1-\delta))$ and $2^{\delta} \le 1 + \delta$ which hold for $t, \delta \in [0,1]$:

$$\le (1+\delta) \cdot (1 - (1-\delta)v(Z|R)) \cdot \mathop{\mathbf{E}}_{v(m|R)} \Big| \mathop{\mathbf{E}}_{v(xy|mR)} \Big[(-1)^{h(xy)}\Big]\Big|.$$

Putting this back into Equation (3.8) and rearranging, we get:

$$\begin{aligned}
\mathop{\mathbf{E}}_{v(m|Z)} \Big| \mathop{\mathbf{E}}_{v(xy|mZ)} \Big[(-1)^{h(xy)}\Big]\Big| &\ge \frac{-\delta + (1-\delta^2)v(Z|R)}{v(Z|R)} \mathop{\mathbf{E}}_{v(m|R)} \Big| \mathop{\mathbf{E}}_{v(xy|mR)} \Big[(-1)^{h(xy)}\Big]\Big| \\
&\ge (1 - \delta^2 - \delta/v(Z|R)) \cdot v(R)^{-\delta} \cdot \mathop{\mathbf{E}}_{v(m)} \Big| \mathop{\mathbf{E}}_{v(xy|m)} \Big[(-1)^{h(xy)}\Big]\Big|,
\end{aligned}$$

where we again used the fact that R maximizes Equation (3.7). □

Lemma 3.24. *Let $q(xym)$ be a rectangular distribution, with $x = x_1x_2$ and $y = y_1y_2$. Let $f(x_1y_1), g(x_2y_2)$ be Boolean functions. Let G be a subset of triples xym such that the indicator function $1_G(xym)$ depends only on $w = x_1y_2m$, and for each m, G maximizes*

$$q(G|m)^{\delta} \cdot \Big| \mathop{\mathbf{E}}_{q(xy|mG)} \Big[(-1)^{f \oplus g(xy)}\Big]\Big|, \tag{3.9}$$

among all such sets. Then for any w in the support of G, we have

$$\Big| \mathop{\mathbf{E}}_{q(xy|w)} \Big[(-1)^{f \oplus g(xy)}\Big]\Big| \ge (1-\delta) \cdot q(G|m)^{-\delta} \cdot \Big| \mathop{\mathbf{E}}_{q(xy|m)} \Big[(-1)^{f \oplus g(xy)}\Big]\Big|.$$

Proof. Fix $w = x_1 y_2 m$ and define $G' \subseteq G$ to be the subset of G obtained by deleting all triples xym consistent with w. Using the triangle inequality, we can write

$$\begin{aligned}
&\left| \mathop{\mathbf{E}}_{q(xy|mG)} \left[(-1)^{f \oplus g(xy)} \right] \right| \\
&\leq q(w|mG) \cdot \left| \mathop{\mathbf{E}}_{q(xy|w)} \left[(-1)^{f \oplus g(xy)} \right] \right| + q(G'|mG) \cdot \left| \mathop{\mathbf{E}}_{q(xy|mG')} \left[(-1)^{f \oplus g(xy)} \right] \right| \\
&= q(w|mG) \cdot \left| \mathop{\mathbf{E}}_{q(xy|w)} \left[(-1)^{f \oplus g(xy)} \right] \right| + q(G'|mG)^{1-\delta} \cdot \frac{q(G'|m)^{\delta}}{q(G|m)^{\delta}} \cdot \left| \mathop{\mathbf{E}}_{q(xy|mG')} \left[(-1)^{f \oplus g(xy)} \right] \right| \\
&\leq q(w|mG) \cdot \left| \mathop{\mathbf{E}}_{q(xy|w)} \left[(-1)^{f \oplus g(xy)} \right] \right| + q(G'|mG)^{1-\delta} \cdot \left| \mathop{\mathbf{E}}_{q(xy|mG)} \left[(-1)^{f \oplus g(xy)} \right] \right|,
\end{aligned}$$

where in the last line we used the fact that G is the maximizer of Equation (3.9).

Because $q(G'|mG) = 1 - q(w|mG)$, and using the inequality $(1-t)^{\gamma} \leq 1 - t\gamma$, which holds for $t, \gamma \in [0,1]$, we obtain

$$\begin{aligned}
\left| \mathop{\mathbf{E}}_{q(xy|mG)} \left[(-1)^{f \oplus g(xy)} \right] \right| \leq q(w|mG) \cdot \left| \mathop{\mathbf{E}}_{q(xy|w)} \left[(-1)^{f \oplus g(xy)} \right] \right| \\
+ (1 - (1-\delta) \cdot q(w|mG)) \cdot \left| \mathop{\mathbf{E}}_{q(xy|mG)} \left[(-1)^{f \oplus g(xy)} \right] \right|.
\end{aligned}$$

Rearranging gives:

$$\begin{aligned}
\left| \mathop{\mathbf{E}}_{q(xy|w)} \left[(-1)^{f \oplus g(xy)} \right] \right| &\geq (1-\delta) \cdot \left| \mathop{\mathbf{E}}_{q(xy|mG)} \left[(-1)^{f \oplus g(xy)} \right] \right| \\
&\geq (1-\delta) \cdot q(G|m)^{-\delta} \cdot \left| \mathop{\mathbf{E}}_{q(xy|m)} \left[(-1)^{f \oplus g(xy)} \right] \right|,
\end{aligned}$$

where in the second inequality we once again used the fact that G is the maximizer of Equation (3.9). □

3.6 Consequences of small marginal information

Let q be a rectangular distribution achieving $\mathsf{M}_I(p, f)$. Since q is rectangular, we can write

$$\frac{q(xym)}{p(xym)} = \frac{\mu(xy) \cdot A(xm) \cdot B(ym)}{\mu(xy) \cdot p(m_0) \cdot \prod_{i=1,3,5,\ldots} p(m_i | xm_{<i}) \cdot p(m_{i+1} | ym_{\leq i})} = g_1(xm) \cdot g_2(ym), \tag{3.10}$$

for appropriate functions g_1 and g_2.

For every $K \geq 1$, define the sets

$$S_K = \{xym : |\lceil \log g_1(xm) \rceil + \log g_2(ym)| \leq 3(\mathsf{M}_I(p,f) + KI)/I\}, \tag{3.11}$$

$$R_K = \{xym : p(m_1|xm_0) \leq 2^{6(\mathsf{M}_I(p,f)+KI)} \cdot p(m_1|ym_0)\}. \tag{3.12}$$

Proposition 3.25. *For $xym \in S_K$,*

$$-\frac{3(\mathsf{M}_I(p,f)+KI)}{I} - 1 \leq \log \frac{q(xym)}{p(xym)} \leq \frac{3(\mathsf{M}_I(p,f)+KI)}{I}. \tag{3.13}$$

Proof. Because $\log(q(xym)/p(xym)) = \log g_1(xm) + \log g_2(ym)$,

$$\log \frac{q(xym)}{p(xym)} \geq \lceil \log g_1(xm) \rceil + \log g_2(ym) - 1 \geq -\frac{3(\mathsf{M}_I(p,f)+KI)}{I} - 1,$$

and

$$\log \frac{q(xym)}{p(xym)} \leq \lceil \log g_1(xm) \rceil + \log g_2(ym) \leq \frac{3(\mathsf{M}_I(p,f)+KI)}{I}.$$

□

Claim 3.26. *If $K \geq 3$, $q(S_K^c), q(R_K^c|S_K) \leq 5 \cdot 2^{-(\mathsf{M}_I(p,f)+KI)/I}$.*

Proof. Define

$$G_1 = \{xym : q(x|ym) \geq 2^{-(\mathsf{M}_I(p,f)+KI)/I} \cdot p(x|y)\},$$

$$G_2 = \{xym : q(y|xm) \geq 2^{-(\mathsf{M}_I(p,f)+KI)/I} \cdot p(y|x)\},$$

$$G_3 = \{xym : q(xym) \geq 2^{-3(\mathsf{M}_I(p,f)+KI)/I} \cdot p(xym)\}, \text{ and}$$

$$G_4 = \{xym : q(x|ym) \geq 2^{-(\mathsf{M}_I(p,f)+KI)/I} \cdot p(x|m_0m_1y)\}.$$

If $xym \in G_1 \cap G_2$, then

$$\begin{aligned}
\mathsf{M}_I(p,f) &\geq \log \left(\frac{q(x|ym)}{p(x|y)} \cdot \frac{q(y|xm)}{p(y|x)} \cdot \left(\frac{q(xym)}{p(xym)} \right)^I \cdot \left| \mathop{\mathbf{E}}_{q(xy|m)} \left[(-1)^f \right] \right|^{-12I/\delta} \right) \\
&\geq -\frac{2(\mathsf{M}_I(p,f)+KI)}{I} + I \cdot \log \frac{q(xym)}{p(xym)} && \text{(because } xym \in G_1 \cap G_2\text{)} \\
&\geq -2\mathsf{M}_I(p,f) - 2KI + I \cdot (\lceil \log g_1(xm) \rceil + \log g_2(ym) - 1) && \text{(using } I \geq 1\text{)}
\end{aligned}$$

and rearranging this and using the fact that $KI \geq 1$ gives

$$\lceil \log g_1(xm) \rceil + \log g_2(ym) \leq 3(\mathsf{M}_I(p,f) + KI)/I.$$

Moreover, for $xym \in G_3$,

$$\lceil \log g_1(xm) \rceil + \log g_2(ym) \geq \log \frac{q(xym)}{p(xym)} \geq -3(\mathsf{M}_I(p,f) + KI)/I,$$

so we have $G_1 \cap G_2 \cap G_3 \subseteq S_K$. We shall prove that $q(S_K^c) \leq 3 \cdot 2^{-(\mathsf{M}_I(p,f)+KI)/I}$ by proving that $q(G_1^c), q(G_2^c), q(G_3^c)$ and $q(G_4^c)$ are all less than $2^{-(\mathsf{M}_I(p,f)+KI)/I}$. We have

$$\begin{aligned} q(G_1^c) = \sum_{xym \notin G_1} q(xym) &< \sum_{xym \notin G_1} q(ym) \cdot p(x|y) \cdot 2^{-(\mathsf{M}_I(p,f)+KI)/I} \\ &\leq 2^{-(\mathsf{M}_I(p,f)+KI)/I} \cdot \sum_{xym} q(ym) \cdot p(x|y) \\ &\leq 2^{-(\mathsf{M}_I(p,f)+KI)/I}, \end{aligned}$$

and similar calculations show that $q(G_2^c), q(G_3^c), q(G_4^c) < 2^{-(\mathsf{M}_I(p,f)+KI)/I}$.

It only remains to bound $q(R_K^c|S_K)$. We have

$$\frac{p(m_1|xm_0)}{p(m_1|ym_0)} = \frac{p(m_1|xym_0)}{p(m_1|ym_0)} = \frac{p(x|ym_0m_1)}{p(x|ym_0)} = \frac{p(x|ym_0m_1)}{q(x|ym)} \cdot \frac{q(x|ym)}{p(x|y)},$$

so, for every $xym \in G_2 \cap G_3 \cap G_4$,

$$\begin{aligned} \mathsf{M}_I(p,f) &\geq \log \left(\frac{q(x|ym)}{p(x|y)} \cdot \frac{q(y|xm)}{p(y|x)} \cdot \left(\frac{q(xym)}{p(xym)} \right)^I \cdot \left| \mathop{\mathbf{E}}_{q(x'y'|m)} \left[(-1)^{f(x'y')} \right] \right|^{-12I/\delta} \right) \\ &\geq \log \left(\frac{p(m_1|xm_0)}{p(m_1|ym_0)} \cdot \frac{q(x|ym)}{p(x|ym_0m_1)} \cdot \frac{q(y|xm)}{p(y|x)} \cdot \left(\frac{q(xym)}{p(xym)} \right)^I \right) \\ &\geq \log \frac{p(m_1|xm_0)}{p(m_1|ym_0)} - \frac{2(\mathsf{M}_I(p,f) + KI)}{I} - 3(\mathsf{M}_I(p,f) + KI) \\ &\geq \log \frac{p(m_1|xm_0)}{p(m_1|ym_0)} - 5(\mathsf{M}_I(p,f) + KI), \end{aligned}$$

since $I \geq 1$. Rearranging, we get $p(m_1|xm_0) \leq 2^{6(\mathsf{M}_I(p,f)+KI)} \cdot p(m_1|ym_0)$, so $G_2 \cap G_3 \cap G_4 \subseteq R_K$. The union bound then gives:

$$q(R_K^c|S_K) \leq \frac{q(G_2^c) + q(G_3^c) + q(G_4^c)}{q(S_K)} < \frac{3 \cdot 2^{-(\mathsf{M}_I(p,f)+KI)/I}}{1 - 3 \cdot 2^{-(\mathsf{M}_I(p,f)+KI)/I}} \leq 5 \cdot 2^{-3(\mathsf{M}_I(p,f)+KI)/I},$$

since $K \geq 3$. □

An argument analogous to the one in the previous claim allows us to obtain similar bounds if marginal information cost is replaced by external marginal information cost:

Claim 3.27. *Let q be a rectangular distribution achieving $\mathsf{M}_I^{\mathsf{ext}}(p,f)$ and let g_1, g_2 be as defined in Equation (3.10). For every K, define*

$$S_K = \{xym : |\lceil \log g_1(xm) \rceil + \log g_2(ym)| \le 3(\mathsf{M}_I^{\mathsf{ext}}(p,f) + KI)/I\} \text{ and} \tag{3.14}$$

$$R_K = \{xym : p(m_1|xm_0) \le 2^{5(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)} \cdot p(m_1|m_0)\}. \tag{3.15}$$

Then, for all $K \ge 2$, it holds that $q(S_K^c), q(R_K^c|S_K) \le 4 \cdot 2^{-(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I}$.

Proof. Define

$$\begin{aligned}
G_1 &= \{xym : q(xy|m) \ge 2^{-(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I} \cdot p(xy)\},\\
G_2 &= \{xym : q(xym) \ge 2^{-3(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I} \cdot p(xym)\},\\
G_3 &= \{xym : q(xy|m) \ge 2^{-(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I} \cdot p(xy|m_0m_1)\}.
\end{aligned}$$

For $xym \in G_1 \cap G_2$, we have

$$\begin{aligned}
\mathsf{M}_I^{\mathsf{ext}}(p,f) &\ge \log\left(\frac{q(xy|m)}{p(xy)} \cdot \left(\frac{q(xym)}{p(xym)}\right)^I \cdot \left|\mathop{\mathbf{E}}_{q(x'y'|m)}\left[(-1)^{f(x'y')}\right]\right|^{-12I/\delta}\right)\\
&\ge -\frac{(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)}{I} + I \cdot \log\frac{q(xym)}{p(xym)}\\
&\ge -(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI) + I \cdot (\lceil \log g_1(xm)\rceil + \log g_2(ym) - 1),
\end{aligned}$$

since $K, I \ge 1$. Rearranging gives

$$\lceil \log g_1(xm)\rceil + \log g_2(ym) \le \frac{3(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)}{I}.$$

Moreover, for $xym \in G_2$

$$\lceil \log g_1(xm)\rceil + \log g_2(ym) \ge \log\frac{q(xym)}{p(xym)} \ge -\frac{3(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)}{I},$$

proving that $G_1 \cap G_2 \subseteq S_K$.

We show that $q(G_1^c), q(G_2^c)$ and $q(G_3^c)$ are all less than $2^{-(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I}$, which implies that $q(S_K^c) \le 2 \cdot 2^{-(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I}$ as desired. To see the upper bound on $q(G_1^c)$, we may write

$$q(G_1^c) = \sum_{xym \notin G_1} q(xym) < \sum_{xym \notin G_1} q(m) \cdot p(xy) \cdot 2^{-3(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I} \le 2^{-(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I}.$$

A similar calculation shows that $q(G_2^c)$ and $q(G_3^c) < 2^{-(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I}$.

Now, we prove that $q(R_K^c|S_K) \le 5 \cdot 2^{-(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I}$. We have

$$\frac{p(m_1|xm_0)}{p(m_1|m_0)} = \frac{p(m_1|xym_0)}{p(m_1|m_0)} = \frac{p(xy|m_0m_1)}{p(xy)} = \frac{p(xy|m_0m_1)}{q(xy|m)} \cdot \frac{q(xy|m)}{p(xy)},$$

so for every $xym \in G_2 \cap G_3$,

$$\begin{aligned}
\mathsf{M}_I^{\mathsf{ext}}(p,f) &\ge \log\left(\frac{q(xy|y)}{p(xy)} \cdot \left(\frac{q(xym)}{p(xym)}\right)^I \cdot \left|\mathop{\mathbf{E}}_{q(x'y'|m)}\left[(-1)^{f(x'y')}\right]\right|^{-12I/\delta}\right) \\
&\ge \log\left(\frac{p(m_1|xm_0)}{p(m_1|m_0)} \cdot \frac{q(xy|m)}{p(xy|m_0m_1)} \cdot \left(\frac{q(xym)}{p(xym)}\right)^I\right) \\
&\ge \log\frac{p(m_1|xym_0)}{p(m_1|m_0)} - \frac{(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)}{I} - (\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI) \\
&\ge \log\frac{p(m_1|xym_0)}{p(m_1|m_0)} - 4(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI),
\end{aligned}$$

since $I \ge 1$. Rearranging, we get $p(m_1|xym_0) \le 2^{5(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)} \cdot p(m_1|m_0)$ for all $xym \in G_2 \cap G_3$, and so $G_2 \cap G_3 \subseteq R_K$. The union bound then gives:

$$q(R_K^c|S_K) \le \frac{q(G_2^c)+q(G_3^c)}{q(S_K)} < \frac{2 \cdot 2^{-(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I}}{1-2\cdot 2^{-(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I}} \le 4 \cdot 2^{-(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I},$$

since $K \ge 2$. □

For the bounded-round simulation protocol, we need the following claim.

Claim 3.28. *Let $K \ge 3$, and S_K be the set defined in Equation (3.11). Let $p(xym)$ be an r-round protocol and define*

$$T_K = \left\{xym : \forall i, \frac{p(m_i|xym_{<i})}{p(m_i|ym_{<i})}, \frac{p(m_i|xym_{<i})}{p(m_i|xm_{<i})} \le 2^{14(\mathsf{M}_I(p,f)+KI)} \cdot (r+1)^5\right\}.$$

Then $q(T_K^c|S_K) \le 22 \cdot 2^{-(\mathsf{M}_I(p,f)+KI)/I}$.

Proof. Define the sets

$$G_1 = \{xm : \forall i, q(xm_{\le i}) \ge 2^{-(\mathsf{M}_I(p,f)+KI)/I} \cdot (r+1)^{-1} \cdot p(xm_{\le i})\},$$
$$G_1' = \{ym : \forall i : q(ym_{\le i}) \ge 2^{-(\mathsf{M}_I(p,f)+KI)/I} \cdot (r+1)^{-1} \cdot p(ym_{<i})\},$$
$$G_2 = \{xym : \forall i, q(x|ym_{\le i}) \ge 2^{-(\mathsf{M}_I(p,f)+KI)/I} \cdot (r+1)^{-1} \cdot p(x|y)\},$$
$$G_2' = \{xym : \forall i, q(y|xm_{\le i}) \ge 2^{-(\mathsf{M}_I(p,f)+KI)/I} \cdot (r+1)^{-1} \cdot p(y|x)\},$$
$$G_3 = \{xym : \forall i, q(x|ym) \ge 2^{-(\mathsf{M}_I(p,f)+KI)/I} \cdot (r+1)^{-1} \cdot p(x|ym_{\le i})\},$$
$$G_3' = \{xym : \forall i, q(y|xm) \ge 2^{-(\mathsf{M}_I(p,f)+KI)/I} \cdot (r+1)^{-1} \cdot p(y|xm_{\le i})\},$$
$$G_4 = \{xym : \forall i, q(m_{\ge i}|xym_{<i}) \ge 2^{-(\mathsf{M}_I(p,f)+KI)/I} \cdot (r+1)^{-1} \cdot p(m_{\ge i}|xym_{<i})\}.$$

We claim that

$$\bigcap_{j=1}^{3} (G_j \cap G_j') \cap G_4 \cap S_K \subseteq T_K. \tag{3.16}$$

For $xym \in G_2' \cap G_2 \cap S_K$,

$$\begin{aligned}
\mathsf{M}_I(p,f) &\ge \log \left(\frac{q(x|ym)}{p(x|y)} \cdot \frac{q(y|xm)}{p(y|x)} \cdot \left(\frac{q(xym)}{p(xym)} \right)^I \cdot \left| \mathop{\mathbf{E}}_{q(x'y'|m)} \left[(-1)^{f(x'y')} \right] \right|^{-12I/\delta} \right) \\
&\ge \log \frac{q(x|ym)}{p(x|y)} - \frac{(\mathsf{M}_I(p,f)+KI)}{I} - \log(r+1) - 3(\mathsf{M}_I(p,f)+KI) - I \\
&\ge \log \frac{q(x|ym)}{p(x|y)} - 4\mathsf{M}_I(p,f) - 5KI - \log(r+1),
\end{aligned}$$

becase $I \ge 1$, and by the definition of G_2' and Equation (3.13). Rearranging implies the first inequality below, and the second has a similar proof:

$$\frac{q(x|ym)}{p(x|y)}, \frac{q(y|xm)}{p(y|x)} \le 2^{5(\mathsf{M}_I(p,f)+KI)/I} \cdot (r+1). \tag{3.17}$$

By Equation (3.17), for $xym \in \bigcap_{j=1}^{3} (G_j \cap G_j') \cap G_4 \cap S_K$ and all i,

$$\frac{p(m_{\le i}|xy)}{p(m_{\le i}|y)} = \frac{p(x|ym_{\le i})}{p(x|y)} = \frac{p(x|ym_{\le i})}{q(x|ym)} \cdot \frac{q(x|ym)}{p(x|y)} \le 2^{(\mathsf{M}_I(p,f)+KI)/I} \cdot 2^{5(\mathsf{M}_I(p,f)+KI)} \cdot (r+1)^2.$$

Moreover,

$$\begin{aligned}\frac{p(m_{\le i}|xy)}{p(m_{\le i}|y)} &= \frac{p(x|ym_{\le i})}{p(x|y)} \\ &= \frac{p(x|ym_{\le i})}{q(x|ym_{\le i})} \cdot \frac{q(x|ym_{\le i})}{p(x|y)} \\ &= \frac{p(xym_{\le i})}{q(xym_{\le i})} \cdot \frac{q(ym_{\le i})}{p(ym_{\le i})} \cdot \frac{q(x|ym_{\le i})}{p(x|y)} \\ &= \frac{p(xym)}{q(xym)} \cdot \frac{q(m_{>i}|xym_{\le i})}{p(m_{>i}|xym_{\le i})} \cdot \frac{q(ym_{\le i})}{p(ym_{\le i})} \cdot \frac{q(x|ym_{\le i})}{p(x|y)} \ge \frac{2^{-6(\mathsf{M}_I(p,f)+KI)/I}}{(r+1)^3},\end{aligned}$$

where we used Equation (3.13) as well as the definitions of G_4, G_1' and G_2 in the last step. Thus,

$$\begin{aligned}\frac{p(m_i|xym_{<i})}{p(m_i|ym_{<i})} = \frac{p(m_{\le i}|xy)}{p(m_{\le i}|y)} \cdot \frac{p(m_{<i}|y)}{p(m_{<i}|xy)} &\le 2^{9(\mathsf{M}_I(p,f)+KI)/I} \cdot 2^{5(\mathsf{M}_I(p,f)+KI)} \cdot (r+1)^5 \\ &\le 2^{14(\mathsf{M}_I(p,f)+KI)} \cdot (r+1)^5,\end{aligned}$$

since $I \ge 1$. A similar calculation shows that $\frac{p(m_i|xym_{<i})}{p(m_i|xm_{<i})} < 2^{14(\mathsf{M}_I(p,f)+KI)} \cdot (r+1)^5$. We conclude that Equation (3.16) holds.

Next, we show that $q(G_4^c) < 2^{-(\mathsf{M}_I(p,f)+KI)/I}$. Define

$$t(xym) = \begin{cases} \min\{i : q(m_{\ge i}|xym_{<i}) < \frac{2^{-(\mathsf{M}_I(p,f)+KI)/I} \cdot p(m_{\ge i}|xym_{<i})}{r+1}\} & \text{if such } i \text{ exists,} \\ \perp & \text{otherwise.} \end{cases}$$

We have,

$$\begin{aligned}q(G_4^c) = q(t \ne \perp) &= \sum_{i=0}^{r} \sum_{\substack{xym, \\ t(xym)=i}} q(xym) \\ &< \frac{2^{-(\mathsf{M}_I(p,f)+KI)/I}}{r+1} \cdot \sum_{i=0}^{r} \sum_{\substack{xym \\ t(xym)=i}} q(xym_{<i}) \cdot p(m_{\ge i}|xym_{<i}) \\ &\le 2^{-(\mathsf{M}_I(p,f)+KI)/I}.\end{aligned}$$

A similar argument shows that $q(G_j^c), q(G_j'^c) < 2^{-(\mathsf{M}_I(p,f)+KI)/I}$, for all $j \in \{1,2,3\}$. Thus, we can bound

$$\begin{aligned}q(T_K^c|S_K) &\le \sum_{j=1}^{3} \frac{q(G_j^c) + q(G_j'^c)}{q(S_K)} + \frac{q(G_4^c) + q(S_K^c)}{q(S_K)} \\ &\le 11 \cdot \frac{2^{-(\mathsf{M}_I(p,f)+KI)/I}}{1 - 2^{-(\mathsf{M}_I(p,f)+KI)/I+2}} \le 22 \cdot 2^{-(\mathsf{M}_I(p,f)+KI)/I},\end{aligned}$$

where we used Claim 3.26 and the fact that $K \geq 3$. □

Claim 3.29. *For any $K \geq 1$, let S_K be the set defined in Equation (3.11) and define*

$$T_K = \{xym : p(m|xy) \leq 2^{6(\mathsf{M}_I(p,f)+KI)} \cdot \min\{p(m|x), p(m|y)\}\}. \tag{3.18}$$

Then, for all $K \geq 3$, $q(T_K^c|S_K) \leq 6 \cdot 2^{-(\mathsf{M}_I(p,f)+KI)/I}$.

Proof. Define the sets

$$G_1 = \{xym : q(m|xy) < 2^{-(\mathsf{M}_I(p,f)+KI)/I} p(m|xy)\}$$
$$G_2 = \{xym : q(x|ym) < 2^{-(\mathsf{M}_I(p,f)+KI)/I} p(x|y)\}$$
$$G_3 = \{xym : q(y|xm) < 2^{-(\mathsf{M}_I(p,f)+KI)/I} p(y|x)\}.$$

We claim that $q(G_1^c), q(G_2^c)$ and $q(G_3^c)$ are all smaller than $2^{-(\mathsf{M}_I(p,f)+KI)/I}$. Indeed, to bound $q(G_1^c)$, we see that

$$q(G_1^c) = \sum_{xym \in G_1} q(xym) < 2^{-(\mathsf{M}_I(p,f)+KI)/I} \cdot \sum_{xym \in G_1} q(xy) \cdot p(m|xy) \leq 2^{-(\mathsf{M}_I(p,f)+KI)/I}.$$

The proof for the bounds on $q(G_2^c)$ and $q(G_3^c)$ are similar. For any $xym \in G_1 \cap G_2 \cap S_K$, we have

$$\begin{aligned}
\mathsf{M}_I(p,f) &\geq \log \left(\frac{q(x|ym)}{p(x|y)} \cdot \frac{q(y|xm)}{p(y|x)} \cdot \left(\frac{q(xym)}{p(xym)} \right)^I \cdot \left| \mathop{\mathbf{E}}_{q(x'y'|m)} \left[(-1)^{f(x'y')} \right] \right|^{-12I/\delta} \right) \\
&\geq \log \frac{q(x|ym)}{p(x|y)} - \frac{(\mathsf{M}_I(p,f)+KI)}{I} - 3(\mathsf{M}_I(p,f)+KI) - I \\
&\qquad \text{(by Equation (3.13) and definition of } G_2) \\
&\geq \log \frac{q(x|ym)}{p(x|ym)} + \log \frac{p(x|ym)}{p(x|y)} - 4(\mathsf{M}_I(p,f)+KI) - I \qquad \text{(since } I \geq 1) \\
&\geq -\frac{(\mathsf{M}_I(p,f)+KI)}{I} + \log \frac{p(m|xy)}{p(m|y)} - 4(\mathsf{M}_I(p,f)+KI) - I,
\end{aligned}$$

where in the last step we used the fact $p(m|xy)/p(m|y) = p(x|ym)/p(x|y)$. Rearranging, we get that for every $xym \in G_1 \cap G_2 \cap S_K$

$$\log \frac{p(m|xy)}{p(m|y)} \leq 6(\mathsf{M}_I(p,f)+KI).$$

A similar calculation shows that for every $xym \in G_1 \cap G_3 \cap S_K$ it holds that

$$\log \frac{p(m|xy)}{p(m|x)} \leq 6(\mathsf{M}_I(p,f) + KI).$$

Therefore,

$$q(T_K^c|S_K) \leq \frac{q(G_1^c) + q(G_2^c) + q(G_3^c)}{q(S_K)} \leq 6 \cdot 2^{-(\mathsf{M}_I(p,f)+KI)/I}.$$

□

Additionally, the bound on the marginal information cost implies the following lemma which will be useful in our simulation.

Lemma 3.30.

$$\mathop{\mathbf{E}}_{q(xym)} \left[\sum_{i \geq 2}^{C} \|p(m_i|xm_{<i}) - p(m_i|ym_{<i})\|_1 \right] \leq 8\sqrt{C \cdot \mathsf{M}_I(p,f)} \tag{3.19}$$

$$\mathop{\mathbf{E}}_{q(xym)} \left[\left| \mathop{\mathbf{E}}_{q(xy|m)} \left[(-1)^{f(xy)} \right] \right| \right] \geq 2^{-\delta \mathsf{M}_I(p,f)/12I}. \tag{3.20}$$

Proof. By our bound on the marginal information cost, we get

$$\begin{aligned}
\mathsf{M}_I(p,f) &= \max_{xym \in \mathsf{supp}(q)} \log \left(\frac{q(x|ym)}{p(x|y)} \cdot \frac{q(y|xm)}{p(y|x)} \cdot \left(\frac{q(xym)}{p(xym)} \right)^I \cdot \left| \mathop{\mathbf{E}}_{q(x'y'|m)} \left[(-1)^{f(x'y')} \right] \right|^{-12I/\delta} \right) \\
&\geq \mathop{\mathbf{E}}_{q(xym)} \left[\log \frac{q(x|ym)}{p(x|y)} \right] + \mathop{\mathbf{E}}_{q(xym)} \left[\log \frac{q(y|xm)}{p(y|x)} \right] \\
&\quad + I \cdot \mathop{\mathbf{E}}_{q(xym)} \left[\log \frac{q(xym)}{p(xym)} \right] + \mathop{\mathbf{E}}_{q(xym)} \left[\log \left| \mathop{\mathbf{E}}_{q(xy|m)} \left[(-1)^{f} \right] \right|^{-12I/\delta} \right]
\end{aligned} \tag{3.21}$$

By Fact 2.4 and the fact that the advantage is always at most 1, each of the expectations appearing above is non-negative, and so each term is bounded by $\mathsf{M}_I(p,f)$. This implies

$$\log \mathop{\mathbf{E}}_{q(xym)} \left[\left| \mathop{\mathbf{E}}_{q(xy|m)} \left[(-1)^{f(xy)} \right] \right| \right] \geq \mathop{\mathbf{E}}_{q(xym)} \left[\log \left| \mathop{\mathbf{E}}_{q(xy|m)} \left[(-1)^{f(xy)} \right] \right| \right] \geq -\frac{\delta \mathsf{M}_I(p,f)}{12I},$$

thus giving Equation (3.20). For Equation (3.19), we have

$$\begin{aligned}
&\mathop{\mathbf{E}}_{q(xym)}\left[\sum_{i\geq 2}^{C}\|p(m_i|xm_{<i})-p(m_i|ym_{<i})\|_1\right]\\
&\leq \mathop{\mathbf{E}}_{q(xym)}\left[\sum_i \|p(m_i|xm_{<i})-q(m_i|xym_{<i})\|_1+\|q(m_i|xym_{<i})-p(m_i|ym_{<i})\|_1\right]\\
&\leq 2\mathop{\mathbf{E}}_{q(xym)}\left[\sum_i \sqrt{\mathop{\mathbf{E}}_{q(m_i|xym_{<i})}\left[\log\frac{q(m_i|xym_{<i})}{p(m_i|xm_{<i})}\right]}+\sqrt{\mathop{\mathbf{E}}_{q(m_i|xym_{<i})}\left[\log\frac{q(m_i|xym_{<i})}{p(m_i|ym_{<i})}\right]}\right]\\
&\qquad\qquad\text{(by Fact 2.4)}\\
&\leq 2\sqrt{C\cdot\mathop{\mathbf{E}}_{q(xym)}\left[\sum_i\log\frac{q(m_i|xym_{<i})}{p(m_i|xm_{<i})}\right]}+2\sqrt{C\cdot\mathop{\mathbf{E}}_{q(xym)}\left[\sum_i\log\frac{q(m_i|xym_{<i})}{p(m_i|ym_{<i})}\right]}\\
&\qquad\qquad\text{(by concavity of }\sqrt{\cdot}\text{)}\\
&= 2\sqrt{C\cdot\mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{q(m|xy)}{p(m|x)}\right]}+2\sqrt{C\cdot\mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{q(m|xy)}{p(m|y)}\right]}.
\end{aligned}$$

To complete the proof, we claim that

$$\mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{q(m|xy)}{p(m|x)}\right],\mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{q(m|xy)}{p(m|y)}\right]\leq \mathsf{M}_I(p,f)\cdot(1+1/I).$$

We show this for the first term; the proof for the second term is identical. First, we have

$$\begin{aligned}
\frac{q(m|xy)}{p(m|x)}&=\frac{q(m|xy)}{p(m|xy)}\cdot\frac{p(m|xy)}{p(m|x)}\\
&=\frac{q(m|xy)}{p(m|xy)}\cdot\frac{p(x|ym)}{p(x|y)}\\
&=\frac{q(m|xy)}{p(m|xy)}\cdot\frac{p(x|ym)}{q(x|ym)}\cdot\frac{q(x|ym)}{p(x|y)}\\
&=\frac{q(xym)}{p(xym)}\cdot\frac{p(xy)}{q(xy)}\cdot\frac{p(x|ym)}{q(x|ym)}\cdot\frac{q(x|ym)}{p(x|y)}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{q(m|xy)}{p(m|x)}\right]\\
&=\mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{q(xym)}{p(xym)}\right]+\mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{p(xy)}{q(xy)}\right]+\mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{p(x|ym)}{q(x|ym)}\right]+\mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{q(x|ym)}{p(x|y)}\right]\\
&\leq\mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{q(xym)}{p(xym)}\right]+\log\mathop{\mathbf{E}}_{q(xym)}\left[\frac{p(xy)}{q(xy)}\right]+\log\mathop{\mathbf{E}}_{q(xym)}\left[\frac{p(x|ym)}{q(x|ym)}\right]+\mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{q(x|ym)}{p(x|y)}\right]\\
&\leq\mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{q(xym)}{p(xym)}\right]+\mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{q(x|ym)}{p(x|y)}\right]\leq\frac{\mathsf{M}_I(p,f)}{I}+\mathsf{M}_I(p,f),
\end{aligned}$$

where in the first inequality, we used the concavity of $\log(\cdot)$ and in the last one, we used Equation (3.21). □

For the simulation of external marginal information, we need a claim analogous to the previous one.

Lemma 3.31. *Let q be a distribution achieving $\mathsf{M}_I^{\mathsf{ext}}(p,f)$. Then,*

$$\mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{p(m|xy)}{p(m)}\right] \le \mathsf{M}_I^{\mathsf{ext}}(p,f), \tag{3.22}$$

$$\mathop{\mathbf{E}}_{q(xym)}\left[\left|\mathop{\mathbf{E}}_{q(xy|m)}\left[(-1)^{f(xy)}\right]\right|\right] \ge 2^{-\delta\mathsf{M}_I^{\mathsf{ext}}(p,f)/(12I)}. \tag{3.23}$$

Proof. By our bound on the marginal information cost, we get

$$\begin{aligned}\mathsf{M}_I^{\mathsf{ext}}(p,f) &= \max_{xym\in\mathsf{supp}(q)} \log\left(\frac{q(xy|m)}{p(xy)}\cdot\left(\frac{q(xym)}{p(xym)}\right)^I\cdot\left|\mathop{\mathbf{E}}_{q(x'y'|m)}\left[(-1)^{f(x'y')}\right]\right|^{-12I/\delta}\right)\\ &\ge \mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{q(xy|m)}{p(xy)}\right] + I\cdot\mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{q(xym)}{p(xym)}\right]\\ &\quad - \frac{12I}{\delta}\cdot\mathop{\mathbf{E}}_{q(xym)}\left[\log\left|\mathop{\mathbf{E}}_{q(xy|m)}\left[(-1)^{f(xy)}\right]\right|\right].\end{aligned}$$

By Fact 2.4 and the fact that the advantage is always at most 1, each of the expectations appearing above is non-negative, and so each term is bounded by $\mathsf{M}_I(p,f)$. This implies

$$\log\mathop{\mathbf{E}}_{q(xym)}\left[\left|\mathop{\mathbf{E}}_{q(xy|m)}\left[(-1)^{f(xy)}\right]\right|\right] \ge \mathop{\mathbf{E}}_{q(xym)}\left[\log\left|\mathop{\mathbf{E}}_{q(xy|m)}\left[(-1)^{f(xy)}\right]\right|\right] \ge -\frac{\delta\mathsf{M}_I(p,f)}{12I},$$

thus giving Equation (3.23). Moreover,

$$\mathsf{M}_I^{\mathsf{ext}}(p,f) \ge \mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{q(xy|m)}{p(xy)}\right] = \mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{q(xy|m)}{p(xy|m)}\right] + \mathop{\mathbf{E}}_{q(xym)}\left[\log\frac{p(xy|m)}{p(xy|m)}\right],$$

and this implies Equation (3.22) since the first term in the sum is non-negative. □

Chapter 4

PROOF OF THE XOR LEMMA

In this chapter we give the details of the statements that were used to prove Theorems 3.1, 3.11 and 3.14.

4.1 Marginal information of efficient protocols

In this section, we prove Theorem 3.5. For convenience, we restate it below.

Theorem 3.5 Restated. *For every Boolean function $f(xy)$ and every protocol p of communication complexity C,*

$$\mathsf{M}_I(p,f) \le 2C - (1+12/\delta)\cdot I \cdot \log\Big(\mathop{\mathbf{E}}_{p(m)}\Big|\mathop{\mathbf{E}}_{p(xy|m)}\Big[(-1)^{f(x,y)}\Big]\Big|\Big) + O(I).$$

Let R be a rectangular set that maximizes the quantity

$$p(R)^{\delta}\cdot \mathop{\mathbf{E}}_{p(m|R)}\Big|\mathop{\mathbf{E}}_{p(xy|mR)}\Big[(-1)^{f(xy)}\Big]\Big|.$$

We shall use trimming to prove the following claim:

Claim 4.1. *There exists a rectangular set $T \subseteq R$ with $p(T|R) \ge 1/4$ such that for any xym in the support of T, we have*

$$\begin{aligned}
\frac{p(xym|T)}{p(xym)} &\le \frac{4}{p(R)},\\
\log\frac{p(x|ymT)}{p(x|y)}, \log\frac{p(y|xmT)}{p(y|x)} &\le \frac{96\cdot 2^C}{p(R)^2},\\
\mathop{\mathbf{E}}_{p(m|T)}\Big|\mathop{\mathbf{E}}_{p(xy|mT)}\Big[(-1)^{f(xy)}\Big]\Big| &\ge \frac{\Omega(1)}{p(R)^{\delta}}\cdot \mathop{\mathbf{E}}_{p(m)}\Big|\mathop{\mathbf{E}}_{p(xy|m)}\Big[(-1)^{f(xy)}\Big]\Big|.
\end{aligned}$$

We defer the proof of the above claim to the end of this section. Let $Q \subseteq T$ be the sub-rectangle obtained by keeping only the messages m' for which the advantage is at least

half of the average advantage:

$$Q = \left\{ x'y'm' \in T : \left| \mathop{\mathbf{E}}_{p(xy|m'T)} \left[(-1)^{f(xy)} \right] \right| \geq \frac{1}{2} \cdot \mathop{\mathbf{E}}_{p(m|T)} \left| \mathop{\mathbf{E}}_{p(xy|mT)} \left[(-1)^{f(xy)} \right] \right| \right\}.$$

Observe that

$$\begin{aligned} \mathop{\mathbf{E}}_{p(m|T)} \left| \mathop{\mathbf{E}}_{p(xy|mT)} \left[(-1)^{f(xy)} \right] \right| &< p(Q|T) + \sum_{\substack{m': \\ p(m'|Q)=0}} p(m'|T) \cdot \frac{1}{2} \cdot \mathop{\mathbf{E}}_{p(m|T)} \left| \mathop{\mathbf{E}}_{p(xy|mT)} \left[(-1)^{f(xy)} \right] \right| \\ &\leq p(Q|T) + \frac{1}{2} \cdot \mathop{\mathbf{E}}_{p(m|T)} \left| \mathop{\mathbf{E}}_{p(xy|mT)} \left[(-1)^{f(xy)} \right] \right|, \end{aligned}$$

and so by the choice of R,

$$p(Q|T) \geq \frac{1}{2} \cdot \mathop{\mathbf{E}}_{p(m|T)} \left| \mathop{\mathbf{E}}_{p(xy|mT)} \left[(-1)^{f(xy)} \right] \right| \geq \frac{\Omega(1)}{p(R)^{\delta}} \cdot \mathop{\mathbf{E}}_{p(m)} \left| \mathop{\mathbf{E}}_{p(xy|m)} \left[(-1)^{f(xy)} \right] \right|. \tag{4.1}$$

Define the rectangular distribution $q(xym) = p(xym|Q)$. By the definition of Q and Claim 4.1, we have that for all m in the support of q:

$$\left| \mathop{\mathbf{E}}_{p(xy|mQ)} \left[(-1)^{f(xy)} \right] \right| \geq \frac{1}{2} \cdot \mathop{\mathbf{E}}_{p(m|T)} \left| \mathop{\mathbf{E}}_{p(xy|mT)} \left[(-1)^{f(xy)} \right] \right| \geq \frac{\Omega(1)}{p(R)^{\delta}} \cdot \mathop{\mathbf{E}}_{p(m)} \left| \mathop{\mathbf{E}}_{p(xy|m)} \left[(-1)^{f(xy)} \right] \right|. \tag{4.2}$$

Using Claim 4.1 and Eqs. (4.1) and (4.2) and the definition of Q, we can bound the marginal information cost by

$$\begin{aligned} & 2^{\mathsf{M}_I(p,f)} \\ & \leq \sup_{xym} \left(\frac{p(x|ymQ)}{p(x|y)} \cdot \frac{p(y|xmQ)}{p(y|x)} \right) \cdot \left(\frac{p(xym|Q)}{p(xym)} \right)^I \cdot \left(\left| \mathop{\mathbf{E}}_{p(x'y'|mQ)} \left[(-1)^{f(x'y')} \right] \right| \right)^{-12I/\delta} \\ & \leq \sup_{xym} \left(\frac{p(x|ymT)}{p(x|y)} \cdot \frac{p(y|xmT)}{p(y|x)} \right) \cdot \left(\frac{p(xym|T)}{p(xym) \cdot p(Q|T)} \right)^I \cdot \left(\left| \mathop{\mathbf{E}}_{p(x'y'|mQ)} \left[(-1)^{f(x'y')} \right] \right| \right)^{-12I/\delta} \\ & \leq O(1) \cdot 2^{2C} \cdot p(R)^{-4} \cdot 2^{O(I)} \cdot p(R)^{-I(1-\delta)+12I} \cdot \left(\mathop{\mathbf{E}}_{p(m)} \left| \mathop{\mathbf{E}}_{p(xy|m)} \left[(-1)^{f(xy)} \right] \right| \right)^{-I(1+12/\delta)} \\ & \leq O(1) \cdot 2^{2C} \cdot 2^{O(I)} \cdot \left(\mathop{\mathbf{E}}_{p(m)} \left| \mathop{\mathbf{E}}_{p(xy|m)} \left[(-1)^{f(xy)} \right] \right| \right)^{-I(1+12/\delta)}, \end{aligned}$$

where in the last inequality we used the fact that $I(12 + \delta - 1) - 4 > 0$ since $I \geq 1$. This completes the proof of the theorem.

4.1.1 Proof of Claim 4.1

It only remains to prove Claim 4.1. We have

$$\mathop{\mathbf{E}}_{p(ym|R)}\left[\frac{1}{p(m|y)}\right] = \sum_{ym}\frac{p(ym|R)}{p(m|y)} \le \frac{1}{p(R)}\sum_{ym}\frac{p(ym)}{p(m|y)} = \frac{\sum_{ym}p(y)}{p(R)} \le \frac{2^C}{p(R)},$$

since the communication complexity of p is bounded by C. A similar argument proves

$$\mathop{\mathbf{E}}_{p(xm|R)}\left[\frac{1}{p(m|x)}\right] \le \frac{2^C}{p(R)}.$$

Define the rectangular set

$$G = \left\{xym \in R : \frac{1}{p(m|y)}, \frac{1}{p(m|x)} \le 4\cdot\frac{2^C}{p(R)}\right\}.$$

Markov's inequality implies that $p(G|R) \ge 1/2$. We apply Lemma 3.22 with $a(xym) = p(xym|G)$, $b(xym) = p(xym)$, and $\kappa = 1/6$ to obtain a rectangular set $T \subseteq G$ with $p(T|G) \ge 1/2$ and

$$\frac{p(xm|T)}{p(xm)}, \frac{p(ym|T)}{p(ym)} \ge \frac{1}{6}, \tag{4.3}$$

for all points in the support of T. We compute

$$p(T|R) = p(G|R)\cdot p(T|G) \ge \frac{1}{4}.$$

Let us verify that T satisfies the remaining conditions promised by Claim 4.1. We have

$$\frac{p(xym|T)}{p(xym)} = \frac{1}{p(T)} = \frac{1}{p(T|R)\cdot p(R)} \le \frac{4}{p(R)}.$$

To prove the second identity, use the first identity, the definition of G and Equation (4.3):

$$\begin{aligned}\frac{p(x|ymT)}{p(x|y)} &= \frac{1}{p(ym|T)}\cdot\frac{p(xym|T)}{p(x|y)}\\ &\le \frac{6}{p(ym)}\cdot\frac{4\cdot p(xym)}{p(x|y)\cdot p(R)}\\ &= \frac{24\cdot p(m|xy)}{p(m|y)\cdot p(R)}\\ &\le \frac{96\cdot 2^C}{p(R)^2}.\end{aligned}$$

A similar calculation yields

$$\frac{p(y|xmT)}{p(y|x)} \leq \frac{96 \cdot 2^C}{p(R)^2}.$$

Finally, applying Lemma 3.23 with $v(xym) = p(xym)$, $Z = T$, and noting that $p(Z|R) \geq 1/4$, we get

$$\begin{aligned}\mathop{\mathbf{E}}_{p(m|T)} \left| \mathop{\mathbf{E}}_{p(xy|mT)} \left[(-1)^{f(xy)}\right]\right| &\geq \frac{1-\delta^2-4\delta}{p(R)^\delta} \cdot \mathop{\mathbf{E}}_{p(m)} \left| \mathop{\mathbf{E}}_{p(xy|m)} \left[(-1)^{f(xy)}\right]\right| \\ &\geq \frac{\Omega(1)}{p(R)^\delta} \cdot \mathop{\mathbf{E}}_{p(m)} \left| \mathop{\mathbf{E}}_{p(xy|m)} \left[(-1)^{f(xy)}\right]\right|.\end{aligned}$$

This completes the proof of Claim 4.1.

4.2 Marginal information is subadditive

In this section we prove Theorem 3.6.

Theorem 3.6 Restated. *There is a universal constant Δ such that if $I \geq 1$ and p is a protocol distribution for computing $f^{\oplus n}$ with $p(xy) = \prod_{i=1}^n p(x_i y_i)$, then there is a protocol p_i for computing f such that $p_i(x_i y_i) = p(x_i y_i)$, p_i has the same number of messages as p, for $j > 1$ the support of m_j is identical in p_i and p, and moreover,*

$$\mathsf{M}_I(p_i, f) \leq \frac{\mathsf{M}_I(p, f^{\oplus n})}{n} + \Delta I \cdot \left(1 + \log \frac{\mathsf{M}_I(p, f^{\oplus n})}{n \cdot I}\right).$$

Recall the definitions of $m^{(1)}, m^{(2)}$, which are given in Equation (3.5). The core of the proof is the following statement.

Theorem 4.2. *Let $f(x_1y_1)$ and $g(x_2y_2)$ be two Boolean functions and let $p(xym)$ be a protocol distribution such that $p(x_1y_1x_2y_2) = p(x_1y_1) \cdot p(x_2y_2)$. Then, for every $1/3 \leq \gamma \leq 2/3$, there are protocol distributions $p_1(x_1y_1m^{(1)})$, $p_2(x_2y_2m^{(2)})$ such that $p_1(x_1y_1) = p(x_1y_1)$, $p_2(x_2y_2) = p(x_2y_2)$, and*

$$\begin{aligned}&\min\left\{\mathsf{M}_I(p_1, f) - \gamma \cdot \mathsf{M}_I(p, f \oplus g), \mathsf{M}_I(p_2, g) - (1-\gamma) \cdot \mathsf{M}_I(p, f \oplus g)\right\} \\ &\quad \leq 3I \cdot \log \frac{\mathsf{M}_I(p, f \oplus g)}{I} + O(I).\end{aligned}$$

We shall prove Theorem 3.6 assuming Theorem 4.2, whose proof we supply right after.

4.2.1 Proof of Theorem 3.6

Let $k_0 > 1$ be a large constant, to be determined. Define $f_i(x_i y_i) = f(x_i y_i)$. For $\ell = 1, 2, \ldots, n$, define

$$k(\ell) = \max\Big\{ \inf_{\substack{S\subset[n],|S|=\ell \\ p'}} \mathsf{M}_I(p', \oplus_{i\in S} f_i), k_0 I \Big\},$$

where the infimum is taken over all protocols p' with C messages such that the support of $m_2, \ldots, m_C$ is the same as in p. Define

$$T = \max\{k(n), k_0 n I\}.$$

Note that

$$k(n) \le T \le \frac{nT}{n} + 12I \cdot \log \frac{nT}{I \cdot n}.$$

For any $\ell > 1$, suppose we have

$$k(\ell) \le \frac{\ell T}{n} + 12I \cdot \log \frac{\ell T}{In}, \tag{4.4}$$

then set $\gamma = \lceil \ell/2 \rceil / \ell$. Since $1/3 \le \gamma \le 2/3$, for k_0 chosen large enough, Theorem 4.2 shows that for some $\ell' \in \{\lceil \ell/2 \rceil, \lfloor \ell/2 \rfloor\}$, we have

$$\begin{aligned}
k(\ell') &\le \max\{\frac{\ell'}{\ell} \cdot k(\ell) + 3I \log \frac{k(\ell)}{I}, k_0 I\} \\
&\le \frac{\ell'}{\ell} \cdot \frac{\ell T}{n} + \frac{\ell'}{\ell} \cdot 12I \cdot \log \frac{\ell T}{In} + 3I \log\Big(\frac{\ell T}{In} + 12 \log \frac{\ell T}{In}\Big) \\
&\qquad\qquad \text{(by the choice of } T \text{ and Equation (4.4))} \\
&\le \frac{\ell' T}{n} + 8I \cdot \log \frac{\ell T}{In} + 3I \log\Big(13 \cdot \frac{\ell T}{In}\Big) \\
&= \frac{\ell' T}{n} + 11I \cdot \log \frac{\ell' T}{In} + 11I \cdot \log \frac{\ell}{\ell'} + 3I \log 13 \\
&\le \frac{\ell' T}{n} + 11I \cdot \log \frac{\ell' T}{In} + 11I \cdot \log \frac{3}{2} + 3I \log 13 \\
&\le \frac{\ell' T}{n} + 12I \cdot \log \frac{\ell' T}{In},
\end{aligned}$$

for k_0 chosen large enough.

So, starting with $\ell = n$, we obtain a smaller and smaller ℓ satisfying Equation (4.4), until $\ell = 1$, which completes the proof.

4.2.2 Proof of Theorem 4.2

Given a Boolean function $h(xy)$, a protocol distribution $p(xym)$ and $q(xym)$ that is rectangular with respect to $p(xy)$, it will be convenient to define

$$\mathsf{M}_I(q,p,h) = \sup_{\substack{xym\in\\ \mathsf{supp}(q)}} \log\left(\frac{q(x|ym)}{p(x|y)}\cdot\frac{q(y|xm)}{p(y|x)}\cdot\left(\frac{q(xym)}{p(xym)}\right)^I\cdot\left|\mathop{\mathbf{E}}_{q(x'y'|m)}\left[(-1)^{h(x'y')}\right]\right|^{-12I/\delta}\right),$$

so $\mathsf{M}_I(p,h) = \inf_q \mathsf{M}_I(q,p,h)$. We exhibit protocol distributions $p_1(x_1y_1m^{(1)})$, $p_2(x_2y_2m^{(2)})$ with $p_1(x_1y_1) = p(x_1y_1)$ and $p_2(x_2y_2) = p(x_2y_2)$ and rectangular distributions r_1, r_2 such that

$$\begin{aligned}&\min\left\{\mathsf{M}_I(r_1,p_1,f)-\gamma\cdot\mathsf{M}_I(q,p,f\oplus g), \mathsf{M}_I(r_2,p_2,g)-(1-\gamma)\cdot\mathsf{M}_I(q,p,f\oplus g)\right\}\\ &\quad\le 3I\cdot\log\frac{\mathsf{M}_I(q,p,f\oplus g)}{I}+O(I),\end{aligned}$$

from which the theorem follows.

Before we give the actual proof, let us give a high level overview of all the steps. Recall the definitions of $m^{(1)}, m^{(2)}$ and w, given in the paragraph preceding Equation (3.5). We start by defining rectangular distributions $q_1(x_1y_1m^{(1)}), q_2(x_2y_2m^{(2)})$ and protocol distributions $p_1(x_1y_1m^{(1)}), p_2(x_2y_2m^{(2)})$ that satisfy the identities described in Equations (4.6) to (4.9). The distributions q_1, q_2 are not be the same as our final rectangular distributions r_1, r_2, but they are closely related. We would like to prove that

$$\mathsf{M}_I(q_1,p_1,f)+\mathsf{M}_I(q_2,p_2,g)\le\mathsf{M}_I(q,p,f\oplus g),$$

but the advantage terms do not add nicely in the marginal information cost: in Equation (4.9), the advantage is computed with respect to w, and not $m^{(1)}, m^{(2)}$ or m. For example, there may be some mw in the support for which

$$\left|\mathop{\mathbf{E}}_{q(xy|w)}\left[(-1)^{f\oplus g(xy)}\right]\right| \ll \left|\mathop{\mathbf{E}}_{q(xy|m)}\left[(-1)^{f\oplus g(xy)}\right]\right|.$$

To resolve this issue, we define a subset G whose indicator function $1_G(xym)$ depends only on w, and yet for all mw in the support of G, the advantage is preserved in the sense of Equation (4.11). This allows us to convert the advantage term in $\mathsf{M}_I(q,p,f\oplus g)$ into the kind of term where Equation (4.9) can be applied, and we use it to get sub-additivity as

described in Equation (4.12). This equation shows that the costs add up pointwise, and so we can pass to a large subset $U \cap L$ where the costs in, say, the first coordinate are a γ-fraction of the total, see Equation (4.13). We are left with our final obstacle: once again the advantage term that we have control over is not exactly the one we want, it may well be that

$$\left| \mathop{\mathbf{E}}_{q_1(x_1y_1|m^{(1)})} \left[(-1)^{f(x_1y_1)} \right] \right| \ll \left| \mathop{\mathbf{E}}_{q_1(x_1y_1|w)} \left[(-1)^{f(x_1y_1)} \right] \right|.$$

To address this, we show that after passing to a suitable set $U' \cap L'$ (whose indicator function depends only on w), the advantage for each fixed w is at least $2^{-\Omega(\mathsf{M}_I(q,p,f\oplus g))}$ (Claim 4.3). We then cluster the w and pass to a subset B of density $\Omega(\mathsf{M}_I(q,p,f\oplus g)^{-1})$ where the advantage terms for each w are within a factor of 2 of each other. The low density of this set is what leads to the $\log \mathsf{M}_I$ factor in the statement of the theorem. This allows us to show that the advantage with respect to $m^{(1)}$ is comparable to the advantage with respect to w (Equation (4.16)). All of these steps leave us with a subset of the inputs xym where the proof gives good control on the quantity $\mathsf{M}_I(q_1,p_1,f)$, but we now need to define a distribution r_1 supported on these points where $\mathsf{M}_I(r_1,p_1,f)$ can be bounded. To do this we need to carefully control the sizes of all the sets we encounter during the proof, and define the distribution of r_1 carefully.

Now we begin the actual proof. Define

$$\begin{aligned}
q_1(x_1y_1m^{(1)}) &= q(x_1y_1m^{(1)}) = q(yw) \\
q_2(x_2y_2m^{(2)}) &= q(x_2y_2m^{(2)}) = q(xw) \\
p_1(x_1y_1m^{(1)}) &= p(x_1y_1) \cdot p(m_0) \cdot q(y_2|x_1m_0) \cdot \prod_{i=1,3,5,\ldots} q(m_i|x_1y_2m_{<i}) \cdot p(m_{i+1}|ym_{\leq i}) \\
p_2(x_2y_2m^{(2)}) &= p(x_2y_2) \cdot q(m_0x_1) \cdot \prod_{i=1,3,5,\ldots} p(m_i|xm_{<i}) \cdot q(m_{i+1}|x_1y_2m_{\leq i})
\end{aligned}$$

These distributions have been carefully chosen to have many nice properties. First, observe that $p_1(x_1y_1m^{(1)}), p_2(x_2y_2m^{(2)})$ are protocol distributions, with the same number of rounds of communication as p, though the length of the $m_1^{(1)}$ is longer than m_1, and the length of $m_0^{(2)}$ is longer than m_0. Since q is rectangular with respect to $p(xy)$, we have

$q(xym) = p(xy) \cdot A(xm) \cdot B(ym)$ for some functions A, B. So, we get

$$\begin{aligned}
&q_1(x_1 y_1 m^{(1)}) \\
&= \sum_{x_2} q(xym) \\
&= \sum_{x_2} p(x_1 y_1) \cdot p(x_2 y_2) \cdot A(xm) \cdot B(ym) \\
&= p(x_1 y_1) \cdot \left(\sum_{x_2} p(x_2 y_2) \cdot A(xm) \right) \cdot B(ym) = p(x_1 y_1) \cdot A'(x_1 m^{(1)}) \cdot B'(y_1 m^{(1)}),
\end{aligned}$$

proving that q_1 is rectangular with respect to $p(x_1 y_1)$. A similar calculation calculation shows that $q_2(x_2 y_2 m^{(2)})$ is rectangular with respect to $p(x_2 y_2)$. Using the fact that p is a protocol and $x_1 y_1$ and $x_2 y_2$ are independent under p, we can compute:

$$\begin{aligned}
p_1(x_1 y_1 m^{(1)}) \cdot p_2(x_2 y_2 m^{(2)}) &= p(xy) \cdot p(m_0) \cdot q(x_1 y_2 m_0) \cdot \prod_{i>0} p(m_i | xym_{<i}) \cdot q(m_i | x_1 y_2 m_{<i}) \\
&= p(xym) \cdot q(x_1 y_2 m) \\
&= p(xym) \cdot q(w). \qquad (4.5)
\end{aligned}$$

The pairs p_1, p_2, q_1, q_2 have been engineered so that the various terms in the marginal cost add up nicely across the pairs. We have:

$$\frac{q_1(x_1 y_1 m^{(1)})}{p_1(x_1 y_1 m^{(1)})} \cdot \frac{q_2(x_2 y_2 m^{(2)})}{p_2(x_2 y_2 m^{(2)})} = \frac{q(xym)}{p(xym)}, \qquad (4.6)$$

$$\frac{q_1(x_1 | y_1 m^{(1)})}{p_1(x_1 | y_1)} \cdot \frac{q_2(x_2 | y_2 m^{(2)})}{p_2(x_2 | y_2)} = \frac{q(x|ym)}{p(x|y)}, \qquad (4.7)$$

$$\frac{q_1(y_1 | x_1 m^{(1)})}{p_1(y_1 | x_1)} \cdot \frac{q_2(y_2 | x_2 m^{(2)})}{p_2(y_2 | x_2)} = \frac{q(y|xm)}{p(y|x)}, \qquad (4.8)$$

$$\left| \mathop{\mathbf{E}}_{q_1(x_1 y_1 | w)} \left[(-1)^{f(x_1 y_1)} \right] \right| \cdot \left| \mathop{\mathbf{E}}_{q_2(x_2 y_2 | w)} \left[(-1)^{g(x_2 y_2)} \right] \right| = \left| \mathop{\mathbf{E}}_{q(xy|w)} \left[(-1)^{f \oplus g(xy)} \right] \right|. \qquad (4.9)$$

To prove Equation (4.6), use Equation (4.5) and Proposition 3.20 to obtain

$$\frac{q_1(x_1 y_1 m^{(1)})}{p_1(x_1 y_1 m^{(1)})} \cdot \frac{q_2(x_2 y_2 m^{(2)})}{p_2(x_2 y_2 m^{(2)})} = \frac{q(xym)}{p(xym)} \cdot \frac{q(w)}{q(w)} = \frac{q(xym)}{p(xym)}.$$

Equations (4.7) and (4.8) follow directly from Proposition 3.20. We use the fact that

$q(xy|w) = q(x_2|w) \cdot q(y_1|w)$ from Proposition 3.20 to prove Equation (4.9):

$$\begin{aligned}
&\left|\mathop{\mathbf{E}}_{q_1(x_1y_1|w)}\left[(-1)^{f(x_1y_1)}\right]\right| \cdot \left|\mathop{\mathbf{E}}_{q_2(x_2y_2|w)}\left[(-1)^{g(x_2y_2)}\right]\right| \\
&= \left|\mathop{\mathbf{E}}_{q(y_1|w)}\left[(-1)^{f(x_1y_1)}\right] \cdot \mathop{\mathbf{E}}_{q(x_2|w)}\left[(-1)^{g(x_2y_2)}\right]\right| \\
&= \left|\mathop{\mathbf{E}}_{q(xy|w)}\left[(-1)^{f\oplus g(xy)}\right]\right|.
\end{aligned}$$

These identities suggest that the costs in the first and second coordinates should sum to $\mathsf{M}_I(q, p, f \oplus g)$. The main challenge in applying this intuition is that the advantage terms in Equation (4.9) are not the ones needed for $\mathsf{M}_I(q_1, p_1, f)$ and $\mathsf{M}_I(q_2, p_2, g)$. To resolve this, we need to remove some problematic points in the support of q. We need to do this while retaining the rectangular structure of q_1, q_2 and preserving the sub-additivity of the other terms in the marginal cost.

Let G be a subset of triples xym such that the indicator function $1_G(xyw)$ depends only on w, and for each fixed m, the set G maximizes

$$q(G|m)^{\delta} \cdot \left|\mathop{\mathbf{E}}_{q(xy|mG)}\left[(-1)^{f\oplus g(xy)}\right]\right|, \tag{4.10}$$

among all such sets. In Lemma 3.24, we prove that for all w in the support of G:

$$\left|\mathop{\mathbf{E}}_{q(xy|w)}\left[(-1)^{f\oplus g(xy)}\right]\right| \geq (1-\delta) \cdot q(G|m)^{-\delta} \cdot \left|\mathop{\mathbf{E}}_{q(xy|m)}\left[(-1)^{f\oplus g(xy)}\right]\right|. \tag{4.11}$$

This gives us an effective way to split the costs for q, p. Using Equations (4.6) to (4.9), we obtain that for all xym in the support of G,

$$\begin{aligned}
&\frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\frac{q_1(y_1|x_1m^{(1)})}{p_1(y_1|x_1)} && \cdot\left(\frac{q_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})}\right)^I && \cdot\left|\mathop{\mathbf{E}}_{q_1(x_1y_1'|w)}\left[(-1)^{f(x_1'y_1')}\right]\right|^{-12I/\delta} \\
\times&\frac{q_2(x_2|y_2m^{(2)})}{p_2(x_2|y_2)}\frac{q_2(y_2|x_2m^{(2)})}{p_2(y_2|x_2)} && \cdot\left(\frac{q_2(x_2y_2m^{(2)})}{p_2(x_2y_2m^{(2)})}\right)^I && \cdot\left|\mathop{\mathbf{E}}_{q_2(x_2'y_2'|w)}\left[(-1)^{g(x_2'y_2')}\right]\right|^{-12I/\delta} \\
=&\frac{q(x|ym)}{p(x|y)}\frac{q(y|xm)}{p(y|x)} && \cdot\left(\frac{q(xym)}{p(xym)}\right)^I && \cdot\left|\mathop{\mathbf{E}}_{q(x'y'|w)}\left[(-1)^{f\oplus g(x'y')}\right]\right|^{-12I/\delta} \\
\leq&\frac{q(x|ym)}{p(x|y)}\frac{q(y|xm)}{p(y|x)} && \cdot\left(\frac{q(xym)}{p(xym)}\right)^I && \cdot\left|\mathop{\mathbf{E}}_{q(x'y'|m)}\left[(-1)^{f\oplus g(x'y')}\right]\right|^{-12I/\delta} \cdot O(q(G|m))^{12I} \\
\leq&2^{\mathsf{M}_I(q,p,f\oplus g)} && \cdot O(q(G|m))^{12I}
\end{aligned}$$

In this product, the quantity in the first line does not depend on the choice of x_2', and the quantity in the second line does not depend on y_1'. Thus, for every fixed value of w, we obtain:

$$\begin{aligned}
&\left|\mathop{\mathbf{E}}_{q_1(x_1y_1|w)}\left[(-1)^{f(x_1y_1)}\right]\right|^{-12I/\delta}\cdot\sup_{y_1}\left(\frac{q_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})}\right)^I\cdot\frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\frac{q_1(y_1|x_1m^{(1)})}{p_1(y_1|x_1)}\\
\times&\left|\mathop{\mathbf{E}}_{q_2(x_2y_2|w)}\left[(-1)^{g(x_2y_2)}\right]\right|^{-12I/\delta}\cdot\sup_{x_2}\left(\frac{q_2(x_2y_2m^{(2)})}{p_2(x_2y_2m^{(2)})}\right)^I\cdot\frac{q_2(x_2|y_2m^{(2)})}{p_2(x_2|y_2)}\frac{q_2(y_2|x_2m^{(2)})}{p_2(y_2|x_2)}\\
\leq&2^{\mathsf{M}_I(q,p,f\oplus g)}\cdot O(q(G|m))^{12I}. \qquad (4.12)
\end{aligned}$$

Let $L\subseteq G$ be a subset whose indicator function $1_L(xym)$ depends only on w, such that $1_L(w)=1$ if and only if

$$\begin{aligned}
&\left(\left|\mathop{\mathbf{E}}_{q_1(x_1y_1|w)}\left[(-1)^{f(x_1y_1)}\right]\right|^{-12I/\delta}\cdot\sup_{y_1}\left(\frac{q_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})}\right)^I\cdot\frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\frac{q_1(y_1|x_1m^{(1)})}{p_1(y_1|x_1)}\right)^{1/\gamma}\\
\leq&\left(\left|\mathop{\mathbf{E}}_{q_2(x_2y_2|w)}\left[(-1)^{g(x_2y_2)}\right]\right|^{-12I/\delta}\cdot\sup_{x_2}\left(\frac{q_2(x_2y_2m^{(2)})}{p_2(x_2y_2m^{(2)})}\right)^I\cdot\frac{q_2(x_2|y_2m^{(2)})}{p_2(x_2|y_2)}\frac{q_2(y_2|x_2m^{(2)})}{p_2(y_2|x_2)}\right)^{1/(1-\gamma)}.
\end{aligned}$$

Let U denote the set whose indicator function depends only on m, such that $1_U(m)=1$ if and only if $q(L|mG)\geq 1/2$. If $q(U)\geq 1/2$, we carry out the reduction in the first coordinate. Otherwise, we carry out the reduction in the second coordinate, using the complements of U, L instead. Without loss of generality, we assume that $q(U)\geq 1/2$. By the definition of U, L, and by Equation (4.12), for all w in the support of $U\cap L$ we have

$$\begin{aligned}
&\left|\mathop{\mathbf{E}}_{q_1(x_1y_1|w)}\left[(-1)^{f(x_1y_1)}\right]\right|^{-12I/\delta}\cdot\left(\sup_{y_1}\frac{q_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})}\right)^I\cdot\frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\frac{q_1(y_1|x_1m^{(1)})}{p_1(y_1|x_1)}\\
&\leq 2^{\gamma\cdot\mathsf{M}_I(q,p,f\oplus g)}\cdot O(q(G|m))^{12\gamma\cdot I}. \qquad (4.13)
\end{aligned}$$

Our next barrier is that in $\mathsf{M}_I(q_1,p_1,f)$ the advantage term is not exactly the same as what we have bounded in the above expressions; it might well be that for most w consistent with $m^{(1)}$

$$\left|\mathop{\mathbf{E}}_{q_1(x_1y_1|m^{(1)})}\left[(-1)^{f(x_1y_1)}\right]\right|\ll\left|\mathop{\mathbf{E}}_{q_1(x_1y_1|w)}\left[(-1)^{f(x_1y_1)}\right]\right|. \qquad (4.14)$$

To resolve this issue, we condition on a dense subset B of the w's such that given any two $w, w' \in B$ that are consistent with the same m,

$$\left|\mathop{\mathbf{E}}_{q_1(x_1y_1|w)}\left[(-1)^{f(x_1y_1)}\right]\right| \geq \frac{1}{2}\cdot\left|\mathop{\mathbf{E}}_{q_1(x_1y_1|w')}\left[(-1)^{f(x_1y_1)}\right]\right|.$$

This will ensure that Equation (4.14) does not happen. To find this subset B, we first prune away some problematic points to ensure all advantage terms in Equation (4.13) are reasonably large. This is accomplished by Claim 4.3 below.

Claim 4.3. *There are subsets $U' \subseteq U, L' \subseteq L$ such that $1_{U'}(xym)$ only depends on m, $1_{L'}(xym)$ only depends on w, $q(U') \geq 1/4$, and for all mw in the support of $U' \cap L'$, we have $q(L'|mG) \geq 1/4$ and*

$$\left|\mathop{\mathbf{E}}_{q_1(x_1y_1|w)}\left[(-1)^{f(x_1y_1)}\right]\right|^{-1} \leq \alpha,$$

for some $\alpha \leq 2^{O(\mathsf{M}_I(q,p,f\oplus g)/I)} \cdot O(1)$.

We defer the proof of Claim 4.3 to the end of this section. Assuming the claim, we can now bucket the w according to their advantage. For each fixing of $m^{(1)}$, partition the space of w consistent with $m^{(1)}$ into disjoint buckets according to the sign of $\mathbf{E}_{q_1(x_1y_1|w)}\left[(-1)^{f(x_1y_1)}\right]$, and the value of

$$\left\lceil \log\left|\mathop{\mathbf{E}}_{q_1(x_1y_1|w)}\left[(-1)^{f(x_1y_1)}\right]\right| \right\rceil.$$

There can be at most $O(\log \alpha)$ such buckets, and so by picking the heaviest bucket for each $m^{(1)}$, we obtain a set $B \subseteq L'$ whose indicator function $1_B(x_1y_1m^{(1)})$ is determined by w, such that for every $m^{(1)}$,

$$q_1(B|m^{(1)}L') \geq \frac{1}{O(\log \alpha)}, \tag{4.15}$$

and moreover, for every $wm^{(1)}$ in the support of B,

$$\left|\mathop{\mathbf{E}}_{q_1(x_1y_1|m^{(1)}B)}\left[(-1)^{f(x_1y_1)}\right]\right| \geq \frac{1}{2}\cdot\left|\mathop{\mathbf{E}}_{q_1(x_1y_1|w)}\left[(-1)^{f(x_1y_1)}\right]\right|. \tag{4.16}$$

Let $R \subseteq B \subseteq L'$ be the rectangular set consisting of $x_1y_1m^{(1)}$'s such that for every $m^{(1)}$, R maximizes

$$q_1(R|m^{(1)}B)^{\delta} \cdot \left|\mathop{\mathbf{E}}_{q_1(x_1y_1|m^{(1)}R)}\left[(-1)^{f(x_1y_1)}\right]\right|. \tag{4.17}$$

Define the rectangular distribution

$$\begin{aligned} r(x_1y_1m^{(1)}) &= q_1(x_1y_1m^{(1)}) \cdot \frac{1_{U'}(m) \cdot 1_R(x_1y_1m^{(1)})}{q_1(U') \cdot q_1(L'|m) \cdot q_1(R|m^{(1)}L')} \\ &= \frac{q_1(m)1_{U'}(m)}{q_1(U')} \cdot \frac{q_1(y_2|m)q_1(L'|m^{(1)})}{q_1(L'|m)} \cdot \frac{q_1(x_1y_1|m^{(1)}) \cdot 1_R(x_1y_1m^{(1)})}{q_1(L'|m^{(1)}) \cdot q_1(R|m^{(1)}L')} \\ &= q_1(m|U') \cdot q_1(y_2|mL') \cdot \frac{q_1(x_1y_1|m^{(1)}) \cdot 1_R(x_1y_1m^{(1)})}{q_1(R|m^{(1)})} \\ &= q_1(m|U') \cdot q_1(y_2|mL') \cdot q_1(x_1y_1|m^{(1)}R). \end{aligned} \tag{4.18}$$

Because r is defined as the product of a rectangular distribution with a function that is also rectangular, r is rectangular. From the last line in Equation (4.18), it is clear that r is a distribution. We have the following bound:

$$\begin{aligned} \frac{r(x_1y_1m^{(1)})}{q_1(x_1y_1m^{(1)})} &= \frac{1_{U'}(m) \cdot 1_R(x_1y_1m^{(1)})}{q_1(U') \cdot q_1(L'|m) \cdot q_1(R|m^{(1)}L')} \\ &\leq \frac{O(1)}{q_1(L'|mG) \cdot q_1(G|m) \cdot q_1(R|m^{(1)}B) \cdot q_1(B|m^{(1)}L')} \\ &\leq \frac{O(\log \alpha)}{q_1(G|m) \cdot q_1(R|m^{(1)}B)}, \end{aligned} \tag{4.19}$$

where here we used Claim 4.3 and Equation (4.15). Apply Lemma 3.22 with $a = r$, $b = q_1$ and $\kappa = 1/6$ to obtain a rectangular set T with $r(T) \geq 1/2$ such that

$$\frac{r(x_1m^{(1)}|T)}{q_1(x_1m^{(1)})}, \frac{r(y_1m^{(1)}|T)}{q_1(y_1m^{(1)})}, \frac{r(m^{(1)}|T)}{r(m^{(1)})} \geq \frac{1}{6}. \tag{4.20}$$

Finally, we define $r_1(x_1y_1m^{(1)}) = r(x_1y_1m^{(1)}|T)$. Because r is a rectangular distribution and T is a rectangular set, r_1 is a rectangular distribution. It only remains to bound $\mathsf{M}_I(r_1, p_1, f)$. For all $x_1y_1m^{(1)}$ in the support of r_1, we have

$$\begin{aligned} \frac{r_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})} &= \frac{q_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})} \cdot \frac{r(x_1y_1m^{(1)})}{q_1(x_1y_1m^{(1)})} \cdot \frac{1}{r(T)} \\ &\leq \frac{q_1(x_1y_1m^{(1)})}{p(x_1y_1m^{(1)})} \cdot \frac{O(\log \alpha)}{q_1(G|m) \cdot q_1(R|m^{(1)}B)}, \end{aligned} \tag{4.21}$$

using Equation (4.19) and the fact that $r(T) \geq 1/2$. For the next term,

$$\begin{aligned}\frac{r_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1m^{(1)})} &= \frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1m^{(1)})} \cdot \frac{r_1(x_1|y_1m^{(1)})}{q_1(x_1|y_1m^{(1)})} \\ &= \frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1m^{(1)})} \cdot \frac{r(x_1y_1m^{(1)})}{q_1(x_1y_1m^{(1)})} \cdot \frac{1}{r(T)} \cdot \frac{q_1(y_1m^{(1)})}{r(y_1m^{(1)}|T)} \\ &\leq \frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1m^{(1)})} \cdot \frac{O(\log \alpha)}{q_1(G|m) \cdot q_1(R|m^{(1)}B)}, \end{aligned} \tag{4.22}$$

using Equations (4.19) and (4.20), and the fact that $r(T) \geq 1/2$. The symmetric argument gives:

$$\frac{r_1(y_1|x_1m^{(1)})}{p_1(y_1|x_1m^{(1)})} \leq \frac{q_1(y_1|x_1m^{(1)})}{p_1(y_1|x_1m^{(1)})} \cdot \frac{O(\log \alpha)}{q_1(G|m) \cdot q_1(R|m^{(1)}B)}. \tag{4.23}$$

To bound the advantage, first note that

$$q_1(T|m^{(1)}R) = \frac{q_1(T|R) \cdot q_1(m^{(1)}|T)}{q_1(m^{(1)}|R)} = \frac{r(T) \cdot r(m^{(1)}|T)}{r(m^{(1)})} \geq \frac{1}{12}, \tag{4.24}$$

by Equation (4.20). For each $m^{(1)}$, we apply Lemma 3.23 by setting $v(x_1y_1m^{(1)}) = q_1(x_1y_1m^{(1)}|m^{(1)}B)$, $Z = T \subseteq R$. Note here that $v(m^{(1)}) = 1$. We obtain the bound:

$$\begin{aligned}\left| \mathop{\mathbf{E}}_{r_1(x_1y_1|m^{(1)})} \left[(-1)^{f(x_1y_1)}\right] \right| &= \left| \mathop{\mathbf{E}}_{q_1(x_1y_1|m^{(1)}T)} \left[(-1)^{f(x_1y_1)}\right] \right| \\ &\geq \frac{1 - \delta^2 - \delta/q_1(T|m^{(1)}R)}{q_1(R|m^{(1)}B)^{\delta}} \cdot \left| \mathop{\mathbf{E}}_{q_1(x_1y_1|m^{(1)}B)} \left[(-1)^{f(x_1y_1)}\right] \right| \\ &\geq \frac{\Omega(1)}{q_1(R|m^{(1)}B)^{\delta}} \cdot \left| \mathop{\mathbf{E}}_{q_1(x_1y_1|m^{(1)}B)} \left[(-1)^{f(x_1y_1)}\right] \right|, \end{aligned} \tag{4.25}$$

by the choice of δ and Equation (4.24).

Now, we are ready to put all these bounds together to complete the proof of the theorem. By Equations (4.16), (4.21) to (4.23) and (4.25), we get that for every $x_1y_1m^{(1)}$ in the support of r_1,

$$\begin{aligned}
&\frac{r_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\cdot\frac{r_1(y_1|x_1m^{(1)})}{p(y_1|x_1)}\cdot\left(\frac{r_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})}\right)^I\cdot\left|\mathop{\mathbf{E}}_{r_1(x_1'y_1'|m^{(1)})}\left[(-1)^{f(x_1'y_1')}\right]\right|^{-12I/\delta}\\
&\leq\frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\cdot\frac{q_1(y_1|x_1m^{(1)})}{p(y_1|x_1)}\cdot\left(\frac{q_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})}\right)^I\cdot\left|\mathop{\mathbf{E}}_{q_1(x_1'y_1'|w)}\left[(-1)^{f(x_1'y_1')}\right]\right|^{-12I/\delta}\\
&\quad\times\frac{O(\log(\alpha))^{I+2}}{q_1(G|m)^{I+2}\cdot q_1(R|m^{(1)}B)^{I+2-12I}}\\
&\leq 2^{\gamma\cdot\mathsf{M}_I(q,p,f\oplus g)}\cdot O(\log(\alpha))^{3I}\cdot q_1(G|m)^{12\gamma\cdot I-I-2}\cdot 2^{O(I)}\cdot q_1(R|m^{(1)}B)^{11I-2}\\
&\leq 2^{\gamma\cdot\mathsf{M}_I(q,p,f\oplus g)}\cdot O(\log(\alpha))^{3I}\cdot q_1(G|m)^{3I-2}\cdot 2^{O(I)}\\
&\leq 2^{\gamma\cdot\mathsf{M}_I(q,p,f\oplus g)}\cdot O\left(\frac{\mathsf{M}_I(q,p,f\oplus g)}{I}\right)^{3I},
\end{aligned}$$

where in the last three inequalities we used Equation (4.13), the fact that $I\geq 1$ and Claim 4.3. This implies that

$$\mathsf{M}_I(r_1,p_1,f)\leq\gamma\cdot\mathsf{M}_I(q,p,f\oplus g)+3I\log\frac{\mathsf{M}_I(q,p,f\oplus g)}{I}+O(I),$$

completing the proof of the theorem.

Proof of Claim 4.3

We have

$$\mathop{\mathbf{E}}_{q_1(m)}\left[\frac{p_1(m)}{q_1(m)}\right]\leq 1,$$

and so by Markov's inequality, the total mass of $m\in\mathsf{supp}(q)$ for which

$$\frac{q_1(m)}{p_1(m)}\leq 1/4 \tag{4.26}$$

is at most $1/4$. We delete all such m from the support of U. We are left with a set U' with

$$q(U')\geq 1/2-1/4=1/4. \tag{4.27}$$

Next, we delete w from L if either

$$\frac{q_1(w|m)}{p_1(w|m)}<\frac{q_1(G|m)}{8}, \tag{4.28}$$

or

$$\mathop{\mathbf{E}}_{q_1(y_1|w)}\left[\log\frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\right] < \log\frac{q_1(G|m)}{8}. \tag{4.29}$$

We claim that for all m in the support of U', $q(L'|mG) \geq 1/4$. To see this, observe that

$$q_1(G|m)\cdot\mathop{\mathbf{E}}_{q_1(w|mG)}\left[\frac{p_1(w|m)}{q_1(w|m)}\right] \leq \mathop{\mathbf{E}}_{q_1(w|m)}\left[\frac{p_1(w|m)}{q_1(w|m)}\right] \leq 1,$$

so Markov's inequality implies that for each m the total mass of w for which Equation (4.28) is violated is at most $1/8$. By the concavity of the log function, the w deleted because of Equation (4.29) satisfy

$$\log\mathop{\mathbf{E}}_{q_1(y_1|w)}\left[\frac{p_1(x_1|y_1m^{(1)})}{q_1(x_1|y_1)}\right] \geq \mathop{\mathbf{E}}_{q_1(y_1|w)}\left[\log\frac{p_1(x_1|y_1m^{(1)})}{q_1(x_1|y_1)}\right] > \log\frac{8}{q_1(G|m)}.$$

On the other hand, because w determines 1_G,

$$\begin{aligned} q_1(G|m)\cdot\mathop{\mathbf{E}}_{q_1(w|mG)}\left[\mathop{\mathbf{E}}_{q_1(y_1|w)}\left[\frac{p_1(x_1|y_1)}{q_1(x_1|y_1m^{(1)})}\right]\right] &\leq \mathop{\mathbf{E}}_{q_1(y_1w|m)}\left[\frac{p_1(x_1|y_1)}{q_1(x_1|y_1m^{(1)})}\right] \\ &= \mathop{\mathbf{E}}_{q_1(y|m)}\left[\mathop{\mathbf{E}}_{q_1(x_1|ym)}\left[\frac{p_1(x_1|y_1)}{q_1(x_1|y_1m^{(1)})}\right]\right] \leq 1, \end{aligned}$$

so once again, Markov's inequality implies that the total mass of w deleted using this rule is at most $1/8$. This gives

$$q(L'|mG) \geq 1/2 - 1/8 - 1/8 = 1/4. \tag{4.30}$$

The result of these pruning steps is that we are left with large sets $U', L' \subseteq G$ such that for all m, w that are consistent with U', L', we have

$$\begin{aligned} &\sup_{y_1}\left(\frac{q_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})}\right)^I\cdot\frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\cdot\frac{q_1(y_1|x_1m^{(1)})}{p_1(y_1|x_1)} \\ &= \left(\frac{q_1(m)\cdot q_1(w|m)}{p_1(m)\cdot p_1(w|m)}\right)^I\cdot\sup_{y_1}\left(\frac{q_1(y_1|w)}{p_1(y_1|w)}\right)^I\cdot\frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\cdot\frac{q_1(y_1|x_1m^{(1)})}{p_1(y_1|x_1)} \\ &= \left(\frac{q_1(m)\cdot q_1(w|m)}{p_1(m)\cdot p_1(w|m)}\right)^I\cdot\exp\left(\sup_{y_1}\log\left(\left(\frac{q_1(y_1|w)}{p_1(y_1|w)}\right)^I\cdot\frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\cdot\frac{q_1(y_1|x_1m^{(1)})}{p_1(y_1|x_1)}\right)\right) \\ &\geq \left(\frac{q_1(m)\cdot q_1(w|m)}{p_1(m)\cdot p_1(w|m)}\right)^I\cdot\exp\left(\mathop{\mathbf{E}}_{q(y_1|w)}\log\left(\left(\frac{q_1(y_1|w)}{p_1(y_1|w)}\right)^I\cdot\frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\cdot\frac{q_1(y_1|x_1m^{(1)})}{p_1(y_1|x_1)}\right)\right), \end{aligned}$$

where here $\exp(z)$ denotes 2^z. Now we use the fact that all the m, w violating Equations (4.26), (4.28) and (4.29) have been deleted and use Fact 2.4 to bound

$$\begin{aligned}
&\geq \left(\frac{1}{4}\cdot\frac{q_1(G|m)}{8}\right)^I \\
&\quad\cdot\exp\left(\mathop{\mathbf{E}}_{q_1(y_1|w)}\left[I\log\frac{q_1(y_1|w)}{p_1(y_1|w)}+\log\left(\frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\right)+\log\left(\frac{q_1(y_1|x_1m^{(1)})}{p_1(y_1|x_1)}\right)\right]\right)\\
&\geq \left(\frac{1}{4}\cdot\frac{q_1(G|m)}{8}\right)^I\cdot\exp\left(\mathop{\mathbf{E}}_{q_1(y_1|w)}\left[\log\left(\frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\right)\right]\right)\\
&\geq \Omega(q_1(G|m))^{1+I}.
\end{aligned}$$

Combining this bound with Equation (4.13), we get that for all w consistent with U', L',

$$\left|\mathop{\mathbf{E}}_{q_1(x_1y_1|w)}\left[(-1)^{f(x_1y_1)}\right]\right|^{-12I/\delta}\leq 2^{\gamma\cdot\mathsf{M}_I(q,p,f\oplus g)}\cdot O(q(G|m))^{12\cdot\gamma I-I-1},$$

so since $I\geq 1$ and $\gamma\geq 1/3$, this implies

$$\left|\mathop{\mathbf{E}}_{q_1(x_1y_1|w)}\left[(-1)^{f(x_1y_1)}\right]\right|^{-1}\leq O(2^{\mathsf{M}_I(q,p,f\oplus g)\cdot(\delta\gamma/12I)})=\alpha, \tag{4.31}$$

as required.

4.3 Compressing marginal information

Here we prove Theorem 3.7.

Theorem 3.7 Restated. *For every $\alpha>0$ there is a $\Delta>0$ such that if $\mathsf{M}_I(p,f)\leq\alpha I$, $\mu(xy)=p(xy)$ and moreover the messages $m=(m_0,\ldots,m_C)$ are such that $m_2,\ldots,m_C\in\{0,1\}$, then $\mathsf{adv}_\mu(\Delta(I+\sqrt{CI}\log(CI)),f)\geq 1/\Delta$.*

Let $p(xym)$ be a protocol distribution such that $p(xy)=\mu(xy)$, and $\mathsf{M}_I(p,f)=\alpha\cdot I$. Let $q(xym)$ be a rectangular distribution that realizes $\mathsf{M}_I(p,f)$. For a large constant K, let $\mathsf{M}=\mathsf{M}_I(p,f)+KI$. Since $\mathsf{M}(p,f)\geq 0$, we have $\mathsf{M}\geq KI$. Let g_1,g_2 be as in Equation (3.10).

Let ε be a parameter such that $\varepsilon\gg(2^{11\mathsf{M}/I}\sqrt{C\cdot\mathsf{M}})^{-1}$. We define a protocol Γ whose communication complexity is bounded by

$$O(\mathsf{M}+\log 1/\varepsilon+2^{7\mathsf{M}/I}\cdot\sqrt{C\mathsf{M}}\cdot\log(C/\varepsilon)).$$

Using the assumption that $\mathsf{M}_I(p,f) \le \alpha I$ we see that $\mathsf{M} \le \Delta_1 \cdot I$ and $\log 1/\varepsilon \le \Delta_2 \cdot \log(CI)$, where Δ_1, Δ_2 only depend on α. This implies the bound on the communication in the theorem. Here is a description of Γ:

1. Jointly sample $p(m_0)$. Alice sets $m_0^A = m_0$ and Bob sets $m_0^B = m_0$. Jointly sample $\eta^A, \eta^B \in [0,1]$ uniformly. Jointly sample uniformly random $\rho \in [0,1]^C$. Jointly sample a uniformly random function $h : \mathbb{Z} \to \{1, \ldots, \lceil 1/\varepsilon \rceil\}$.

2. Run the protocol ψ from Lemma 2.5 with $u = p(m_1|m_0^A x), v = p(m_1|m_0^B y), L = 6\mathsf{M}$, error parameter ε, to obtain functions a, b and transcript s. Alice sets $m_1^A = a(\psi(us))$, Bob sets $m_1^B = b(\psi(vs))$. If $m_1^B = \perp$, the protocol terminates. Bob sends a bit to Alice to indicate whether or not this occurs. The communication complexity of this step is $L + O(\log(1/\varepsilon))$.

3. Alice and Bob compute m^B, m^A by setting

$$m_i^A = \begin{cases} 1 & \text{if } \rho_i \le p(m_i = 1|x m_{<i}^A), \\ 0 & \text{otherwise.} \end{cases} \tag{4.32}$$

$$m_i^B = \begin{cases} 1 & \text{if } \rho_i \le p(m_i = 1|y m_{<i}^B), \\ 0 & \text{otherwise.} \end{cases}, \tag{4.33}$$

   for $i = 2, \ldots, C$.

4. Run τ from Lemma 2.6 to find the smallest j with $m_j^A \ne m_j^B$. If j is even, Alice flips the value of m_j^A to $1 - m_j^A$ and recomputes m_i^A for $i = j+1, \ldots, C$ using Equation (4.32). If j is odd, Bob flips the value of m_j^B to $1 - m_j^B$ and recomputes m_i^B for $i = j+1, \ldots, C$ using Equation (4.33). The players repeat this process at most $2^{7\mathsf{M}/I}\sqrt{C\mathsf{M}}$ times. If by this point τ reports that $m^A \ne m^B$, the players abort. Otherwise, they continue. Let $\langle m^A \rangle$, $\langle m^B \rangle$ denote the final values of m^A, m^B after this step. The communication complexity of this step is at most $O(2^{7\mathsf{M}/I} \cdot \sqrt{C\mathsf{M}} \cdot \log(C/\varepsilon))$.

5. If $\eta^A \le g_1(x\langle m^A \rangle) \cdot 2^{-\lceil \log g_1(x\langle m^A \rangle) \rceil}$, Alice sends $h(\lceil \log g_1(x\langle m^A \rangle) \rceil)$ to Bob, and otherwise she sends $\perp$ to indicate that the protocol should be aborted.

6. If there is a unique integer z such that

$$|z + \log g_2(y\langle m^B\rangle)| \leq 3\mathsf{M}/I,$$
$$h(z) = h(\lceil \log g_1(x\langle m^A\rangle)\rceil),$$
$$\eta^B \leq g_2(y\langle m^B\rangle) \cdot 2^{z-3\mathsf{M}/I},$$

Bob sends $\mathsf{sign}\Big(\mathbf{E}_{q(x'y'|\langle m^B\rangle)}[(-1)^{f(x'y')}]\Big) \in \{\pm 1\}$ to Alice. Otherwise, he sends $\perp$ to abort the protocol.

Let Γ denote the joint distribution of the inputs and transcript of the above protocol. In order to analyze the protocol, define m by setting $m_0 = m_0^A = m_0^B$, $m_1 = m_1^A$, and setting

$$m_i = \begin{cases} 1 & \text{if } \rho_i \leq p(m_i = 1|xm_{<i}), \\ 0 & \text{otherwise} \end{cases},$$

when $i > 1$ is even, and setting

$$m_i = \begin{cases} 1 & \text{if } \rho_i \leq p(m_i = 1|ym_{<i}), \\ 0 & \text{otherwise} \end{cases},$$

when $i > 1$ is odd. This definition ensures that

$$\Gamma(xym) = p(xym).$$

For $i = 2, 3, \ldots, C$ define

$$E_i = \begin{cases} 1 & \text{if } \rho_i \text{ is in between the numbers } p(m_i = 1|xm_{<i}) \text{ and } p(m_i = 1|ym_{<i}), \\ 0 & \text{otherwise.} \end{cases}$$

Let S and R be the sets defined in Equations (3.11) and (3.12) for our choice of K. In addition to S and R, we need the following sets to analyze the simulating protocol:

$$Q = \Big\{xym\eta^A\eta^B : \eta^A \leq g_1(xm) \cdot 2^{-\lceil \log g_1(xm)\rceil}, \eta^B \leq g_2(ym) \cdot 2^{\lceil \log g_1(xm)\rceil - 3\mathsf{M}/I}\Big\},$$
$$\mathcal{E} = \{\langle m^A\rangle\langle m^B\rangle m : \langle m^A\rangle = \langle m^B\rangle = m\},$$
$$\mathcal{Z} = \{xymh : \exists \text{ unique } z \in \mathbb{Z} \text{ s.t. } |z + \log g_2(ym)| \leq 3\mathsf{M}/I \text{ and } h(z) = h(\lceil \log g_1(xm)\rceil)\}.$$

Let $\mathcal{G}$ denote the event that the protocol reaches the final step without aborting, and define $\mathcal{A}(xym) \in \{\pm 1\}$ by

$$\mathcal{A}(xym) = \mathsf{sign}\Big(\mathop{\mathbf{E}}_{q(x'y'|m)}[(-1)^{f(x'y')}]\Big) \cdot (-1)^{f(xy)}.$$

Our protocol computes $f(xy)$ correctly when: $\mathcal{G}$ happens, $\mathcal{A}(xym) = 1$ and $m = m^B$. Since $\mathcal{EZ}SQ \subseteq \mathcal{G}$, and $\mathcal{E}$ implies $m = m^B$, the advantage of our protocol is at least:

$$\Gamma(\mathcal{EZ}SQ) \cdot \mathop{\mathbf{E}}_{\Gamma(xym|\mathcal{EZ}SQ)}[\mathcal{A}(xym)] - \Gamma(\mathcal{G}(\mathcal{EZ}SQ)^c). \tag{4.34}$$

We shall prove:

$$\mathop{\mathbf{E}}_{\Gamma(xym|\mathcal{EZ}QS)}[\mathcal{A}(xym)] \geq \Omega(2^{-\delta \mathsf{M}/(12I)}), \tag{4.35}$$

$$\Gamma(\mathcal{EZ}SQ) \geq \Omega(2^{-3\mathsf{M}/I}), \tag{4.36}$$

$$\Gamma(\mathcal{G}(\mathcal{EZ}SQ)^c) \leq O(2^{-4\mathsf{M}/I}). \tag{4.37}$$

By Equation (4.34), since $\delta \leq 1$, we can choose K to be large enough to prove the theorem, since $(\alpha + K) \geq \mathsf{M}/I \geq K$.

We first upper bound $\Gamma(\mathcal{G}(\mathcal{EZ}SQ)^c)$. By the union bound, we have:

$$\Gamma(\mathcal{G}(\mathcal{EZ}SQ)^c) \leq \Gamma(\mathcal{G}\mathcal{E}^c) + \Gamma(\mathcal{Z}^c|\mathcal{G}\mathcal{E}) + \Gamma(S^c\mathcal{G}\mathcal{E}\mathcal{Z}) + \Gamma(Q^c|\mathcal{G}\mathcal{E}\mathcal{Z}\mathcal{Z}S).$$

The definition of the protocol ensures that $\Gamma(\mathcal{Z}^c|\mathcal{G}\mathcal{E}) = 0$. Moreover, we claim that $\Gamma(Q^c|\mathcal{G}\mathcal{E}\mathcal{Z}\mathcal{Z}S) = 0$, because if the event $\mathcal{EZ}S$ happens and the parties do not abort, then:

$$\eta^A \leq g_1(x\langle m^A\rangle) \cdot 2^{\lceil \log g_1(x\langle m^A\rangle)\rceil} = g_1(xm) \cdot 2^{\lceil \log g_1(xm)\rceil},$$
$$\eta^B \leq g_2(y\langle m^B\rangle) \cdot 2^{z-3\mathsf{M}/I} = g_2(ym) \cdot 2^{\lceil \log g_1(xm)\rceil - 3\mathsf{M}/I}.$$

The event $\mathcal{G}\mathcal{E}^c$ implies that ψ or τ made an error, leaving Alice and Bob with strings that were not equal in some step. The probability that this happens is at most

$$O(\varepsilon \cdot (1 + 2^{7\mathsf{M}/I}\sqrt{C \cdot \mathsf{M}})) \leq 2^{-4\mathsf{M}/I},$$

by our choice of ε. Finally, $\Gamma(S^c\mathcal{G}\mathcal{E}\mathcal{Z}) \leq \Gamma(S^c\mathcal{E}\mathcal{Z}) \leq O(\varepsilon\mathsf{M}/I)$, since if $S^c\mathcal{G}\mathcal{E}\mathcal{Z}$ happens then there must have been a hash collision, which happens with probability at most $O(\varepsilon\mathsf{M}/I)$. This implies Equation (4.37).

Now, we turn to proving Equation (4.36). Let us first estimate $\Gamma(QS)$. We have,

$$\Gamma(QS) = \sum_{xym \in S} \Gamma(xym) \cdot \Gamma(Q|xym) = \sum_{xym \in S} p(xym) \cdot \Gamma(Q|xym).$$

For $xym \in S$,

$$\Gamma(Q|xym) = \frac{g_1(xm) \cdot 2^{-\lceil \log g_1(xm) \rceil} \cdot g_2(ym) \cdot 2^{\lceil \log g_1(xm) \rceil}}{2^{3\mathsf{M}/I}} = \frac{q(xym)}{p(xym)} \cdot \frac{1}{2^{3\mathsf{M}/I}}, \tag{4.38}$$

where the first equality follows from the fact that

$$g_2(ym) \cdot 2^{\lceil \log g_1(xm) \rceil} = 2^{\lceil \log g_1(xm) \rceil + \log g_2(ym)} \leq 2^{3\mathsf{M}/I},$$

by the definition of S. Therefore,

$$\begin{aligned} \Gamma(QS) = \sum_{xym \in S} p(xym) \cdot \frac{q(xym)}{p(xym)} \cdot \frac{1}{2^{3\mathsf{M}/I}} &= \frac{q(S)}{2^{3\mathsf{M}/I}} \\ &\geq \frac{(1 - 5 \cdot 2^{-\mathsf{M}/I})}{2^{3\mathsf{M}/I}} = \Omega(2^{-3\mathsf{M}/I}), \end{aligned} \tag{4.39}$$

where in the last line, we used Claim 3.26.

We claim that for all $xym \in S$,

$$\Gamma(\mathcal{Z}|xymQS) = \Gamma(\mathcal{Z}|xym) \geq 1 - O(\varepsilon \mathsf{M}/I). \tag{4.40}$$

The equality follows by observing that xym determine S and given xym, $\mathcal{Z}$ just depends on the choice of h, which is independent of Q. The event $\mathcal{Z}^c$ can happen only if there exists an integer z distinct from $\lceil \log g_1(xm) \rceil$ such that $h(\lceil g_1(xm) \rceil) = h(z)$ and $|z + \log g_2(ym)| \leq 3\mathsf{M}/I$. The probability that this happens is at most $O(\varepsilon \cdot \mathsf{M}/I)$. Therefore, $\Gamma(\mathcal{Z}|xym) \geq 1 - O(\varepsilon \mathsf{M}/I) \geq 1/2$, by our choice of ε. We conclude that

$$\Gamma(QS\mathcal{Z}) = \Gamma(QS) \cdot \Gamma(\mathcal{Z}|QS) \geq \Omega(2^{-3\mathsf{M}/I}), \tag{4.41}$$

For all $xym \in S$,

$$\begin{aligned}
\Gamma(xym|QS\mathcal{Z}) &= \frac{\Gamma(xym)\cdot\Gamma(QS\mathcal{Z}|xym)}{\Gamma(QS\mathcal{Z})} \\
&= \frac{p(xym)}{\Gamma(QS)}\cdot\Gamma(Q|xym)\cdot\frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} \\
&= \frac{p(xym)}{\Gamma(QS)}\cdot\frac{q(xym)}{p(xym)\cdot 2^{3\mathsf{M}/I}}\cdot\frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} && \text{(By Equation (4.38))} \\
&= \frac{q(xym)}{q(S)}\cdot\frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} && \text{(By Equation (4.39))} \\
&= q(xym|S)\cdot(1\pm O(\varepsilon\mathsf{M}/I)), && (4.42)
\end{aligned}$$

where the last line follows by Equation (4.40).

Given Equation (4.41), to complete the proof of Equation (4.36), it will be enough to prove that $\Gamma(\mathcal{E}|QS\mathcal{Z}) \geq 1/2$. We shall prove that

$$\begin{aligned}
&\Gamma(\mathcal{E}^c|QS\mathcal{Z}) \\
&\leq \Gamma(R^c|QSZ) + \Gamma\left(\mathcal{E}^c\middle|QS\mathcal{Z}R, \sum_{i=2}^{C} E_i \leq 2^{7\mathsf{M}/I}\cdot\sqrt{C\mathsf{M}}\right) + \Gamma\left(\sum_{i=2}^{C} E_i > 2^{7\mathsf{M}/I}\cdot\sqrt{C\mathsf{M}}\middle|QS\mathcal{Z}\right) \\
&\leq O(2^{-\mathsf{M}/I}). && (4.43)
\end{aligned}$$

By Equation (4.42) and Claim 3.26,

$$\Gamma(R^c|QS\mathcal{Z}) \leq q(R^c|S)(1+O(\varepsilon\mathsf{M}/I)) \leq 2^{-\mathsf{M}/I+3}. \tag{4.44}$$

Given $QS\mathcal{Z}R$ and the event $\sum_{i=2}^{C} E_i \leq 2^{7\mathsf{M}/I}\cdot\sqrt{C\mathsf{M}}$, the event $\mathcal{E}^c$ can happen only if τ or ψ make an error that leaves Alice and Bob with inconsistent messages, or if ψ aborts. We claim that the probability that ψ makes an error or aborts is at most 2ε. This is because every $xym \in R$ satisfies $p(m_1|xm_0) \leq 2^{6\mathsf{M}}\cdot p(m_1|m_0y)$, so we can apply Lemma 2.5. Moreover, the probability that τ ever makes an error is at most $O(\varepsilon 2^{7\mathsf{M}/I}\sqrt{C\mathsf{M}})$ by a union bound. So, we conclude that

$$\Gamma\left(\mathcal{E}^c\middle|QS\mathcal{Z}R, \sum_{i=2}^{C} E_i \leq 2^{7\mathsf{M}/I}\cdot\sqrt{C\mathsf{M}}\right) \leq O(\varepsilon 2^{7\mathsf{M}/I}\sqrt{C\mathsf{M}}). \tag{4.45}$$

We shall prove at the end of this section that

$$\Gamma\left(\sum_{i=2}^{C} E_i > 2^{7\mathsf{M}/I}\cdot\sqrt{C\mathsf{M}}\middle|QS\mathcal{Z}\right) < O(2^{-\mathsf{M}/I}). \tag{4.46}$$

Equations (4.44) to (4.46) together prove Equation (4.43), and so conclude the proof of Equation (4.36). Next, we prove Equation (4.35). Since $|\mathcal{A}(xym)| \leq 1$, we have

$$\mathop{\mathbf{E}}_{\Gamma(xym|QS\mathcal{Z})}[\mathcal{A}(xym)] \leq \Gamma(\mathcal{E}|QS\mathcal{Z}) \cdot \mathop{\mathbf{E}}_{\Gamma(xym|QS\mathcal{Z}\mathcal{E})}[\mathcal{A}(xym)] + \Gamma(\mathcal{E}^c|QS\mathcal{Z}),$$

and since $\Gamma(\mathcal{E}|QS\mathcal{Z}) \leq 1$, this gives

$$\begin{aligned}
\mathop{\mathbf{E}}_{\Gamma(xym|QS\mathcal{Z}\mathcal{E})}[\mathcal{A}(xym)] &\geq \mathop{\mathbf{E}}_{\Gamma(xym|QS\mathcal{Z})}[\mathcal{A}(xym)] - \Gamma(\mathcal{E}^c|QS\mathcal{Z}) \\
&\geq \mathop{\mathbf{E}}_{q(xym|S)}[\mathcal{A}(xym)] - O(\varepsilon \mathsf{M}/I) - O(2^{-\mathsf{M}/I}) \\
&\qquad \text{(using Equations (4.42) and (4.43))} \\
&\geq \mathop{\mathbf{E}}_{q(xym)}[\mathcal{A}(xym)] - O(2^{-\mathsf{M}/I}) \qquad \text{(by Claim 3.26)} \\
&= \mathop{\mathbf{E}}_{q(xym)}\left[\mathsf{sign}\left(\mathop{\mathbf{E}}_{q(x'y'|m)}\left[(-1)^{f(x'y')}\right]\right) \cdot (-1)^{f(xy)}\right] - O(2^{-\mathsf{M}/I}) \\
&= \mathop{\mathbf{E}}_{q(m)}\left[\left|\mathop{\mathbf{E}}_{q(xy|m)}\left[(-1)^{f(xy)}\right]\right|\right] - O(2^{-\mathsf{M}/I}) \geq \Omega(2^{-\delta \mathsf{M}/(12I)}),
\end{aligned}$$

by Equation (3.20). This completes the proof of Equation (4.35).

It only remains to prove Equation (4.46). Define the function

$$t(xym) = \begin{cases} \min\{j : q(xym_{\leq j}) < 2^{-3\mathsf{M}/I} \cdot p(xym_{\leq j})\} & \text{if such } j \text{ exists,} \\ \perp & \text{otherwise.} \end{cases}$$

Note that the function $t(xym)$ is determined by $xym_{\leq t(xym)}$.

We have

$$\begin{aligned}
&\Gamma\left(\sum_{i=2}^{C} E_i > 2^{7\mathsf{M}/I} \cdot \sqrt{C\mathsf{M}} \middle| QS\mathcal{Z}\right) \\
&\leq \Gamma(t \neq \perp | QSZ) + \Gamma\left(\sum_{i=2}^{C} E_i > 2^{7\mathsf{M}/I} \cdot \sqrt{C\mathsf{M}} \middle| QS\mathcal{Z}, t = \perp\right),
\end{aligned}$$

so let us bound each of these terms.

$$\begin{aligned}
\Gamma(t \neq \perp | QSZ) &\leq q(t \neq \perp | S) \cdot (1 + O(\varepsilon \mathsf{M}/I)) \qquad \text{(by Equation (4.42))} \\
&\leq q(t \neq \perp) \cdot (1 + O(2^{-3\mathsf{M}/I})) \cdot (1 + O(\varepsilon \mathsf{M}/I)), \qquad \text{(by Claim 3.26)}
\end{aligned}$$

and

$$q(t \neq \perp) = \sum_{j=0}^{C} q(t=j) = \sum_{j=0}^{C} \sum_{\substack{xym_{\leq j} \\ t(xym)=j}} q(xym_{\leq j})$$
$$< 2^{-3\mathsf{M}/I} \cdot \sum_{j=0}^{C} \sum_{\substack{xym_{\leq j} \\ t(xym)=j}} p(xym_{\leq j}) = 2^{-3\mathsf{M}/I} \cdot p(t \neq \perp) \leq 2^{-3\mathsf{M}/I},$$

so we conclude that

$$\Gamma(t \neq \perp | QS\mathcal{Z}) \leq O(2^{-3\mathsf{M}/I}). \tag{4.47}$$

Next, we show that

$$\Gamma\left(\sum_{i=2}^{C} E_i > 2^{7\mathsf{M}/I} \cdot \sqrt{C\mathsf{M}} \middle| QS\mathcal{Z}, t = \perp \right) < O(2^{-\mathsf{M}/I}),$$

which would complete the proof of Equation (4.46). This follows from Markov's inequality and the bound

$$\mathop{\mathbf{E}}_{\Gamma}\left[\sum_{i=2}^{C} E_i \middle| QS\mathcal{Z}, t = \perp \right] \leq O(2^{6\mathsf{M}/I}\sqrt{C\mathsf{M}}), \tag{4.48}$$

which we prove next. We have:

$$\mathop{\mathbf{E}}_{\Gamma}\left[\sum_{i=2}^{C} E_i \middle| QS\mathcal{Z}, t = \perp \right] = \sum_{i=2}^{C} \frac{\Gamma(E_i = 1, QS\mathcal{Z}, t = \perp)}{\Gamma(QS\mathcal{Z}, t = \perp)} \tag{4.49}$$
$$\leq O(2^{3\mathsf{M}/I}) \cdot \sum_{i=2}^{C} \Gamma(E_i = 1, t = \perp).$$
(by Equations (4.41) and (4.47))

Moreover,

$$\Gamma(E_i = 1, t = \perp) = \sum_{xym_{<i}} \Gamma(xym_{<i}) \cdot \Gamma(E_i = 1 | xym_{<i}) \cdot \Gamma(t = \perp | E_i = 1, xym_{<i})$$
$$= \sum_{xym_{<i}} p(xym_{<i}) \cdot \|p(m_i | xm_{<i}) - p(m_i | ym_{<i})\|_1 \cdot \Gamma(t = \perp | E_i = 1, xym_{<i})$$
$$\leq O(2^{3\mathsf{M}/I}) \cdot \sum_{xym_{<i}} q(xym_{<i}) \cdot \|p(m_i | xm_{<i}) - p(m_i | ym_{<i})\|_1 \tag{4.50}$$

Therefore,

$$\mathop{\mathbf{E}}_{\Gamma}\left[\sum_{i=2}^{C} E_i \Big| QSZ, t=\perp\right] \le O(2^{3\mathsf{M}/I}) \cdot \mathop{\mathbf{E}}_{q(xym)}\left[\sum_{i=2}^{C} \|p(m_i|xm_{<i}) - p(m_i|ym_{<i})\|_1\right]$$
$$\le O(2^{3\mathsf{M}/I} \cdot \sqrt{C\mathsf{M}}),$$

by Equation (3.19), which completes the proof of Equation (4.48).

4.4 Smoothing protocols

A smooth protocol is a protocol where the message bit is close to being uniformly distributed even conditioned on the transcript until then.

Definition 4.4. *Given a protocol distribution $p(xym)$ with C messages satisfying $m_i \in \{0,1\}$ for each $i \ge 2$, we say that the distribution is β-smooth if for all $i > 1$, $|p(m_i|xym_{<i}) - 1/2| \le \beta$.*

Here we prove the following theorem:

Theorem 4.5. *For every Boolean function f, every protocol distribution $p(xym)$ with C messages satisfying $m_2, \ldots, m_C \in \{0,1\}$, and every $\beta > 0$, assuming that $\mathsf{M}_I^{\mathsf{ext}}(p,f)$ is finite, there is a β-smooth protocol $p'(xym')$ with $C' \le O(C \cdot \log(IC)/\beta^2)$ messages such that $\mathsf{M}_I^{\mathsf{ext}}(p',f) \le \mathsf{M}_I^{\mathsf{ext}}(p,f) + 1$, and $m'_2, \ldots, m'_{C'} \in \{0,1\}$.*

Proof. We start by ensuring that if $p(m_i|xym_{<i}) \in [1/2-\beta, 1/2+\beta]$ for each $i \in \{2,\ldots,C\}$. Let $q(xym)$ be a rectangular distribution realizing $\mathsf{M}_I^{\mathsf{ext}}(p,f)$. Let $L > 1$ be a large odd number to be determined. Define the pair of distributions $q'(xym'), p'(xym')$ as follows. Let m'_0, m'_1 have the same support as m_0, m_1, and let $m'_2, \ldots, m'_C \in \{0,1\}^L$. In p', q', the i'th message will correspond to m'_i.

For $a \in \{0,1\}$, define the following distributions supported on $\{0,1\}^L$:

$$s_a(r) = \prod_{j=1}^{L} \frac{1}{2} + (-1)^{a+r_j} \cdot \beta$$

$$t_a(r) = s_a\Big(r \Big| (-1)^a \cdot \sum_{j=1}^{L} (-1)^{r_j} \geq 0\Big)$$

$$t'_a(r) = s_a\Big(r \Big| (-1)^a \cdot \sum_{j=1}^{L} (-1)^{r_j} < 0\Big).$$

In words, $s_a(r)$ is the distribution of L independent bits that are biased towards being equal to a, $t_a(r)$ is this distribution conditioned on the event that the majority of the bits is equal to a and $t'_a(r)$ is the distribution conditioned on the event that the majority is not a.

Now we define a protocol distribution $p'(xym')$ and a rectangular distribution $q'(xym')$. Given m', let $D(m'_i)$ denote the unique string satisfying $D(m'_0, m'_1) = (m'_0, m'_1)$, and

$$(-1)^{D(m'_i)} \cdot \sum_{j=1}^{L} (-1)^{m'_{i,j}} \geq 0.$$

In other words, D decodes each block of L bits by taking the majority. Below we abuse notation and write $D(m) = D(m_0), D(m_1), \ldots, D(m_C)$.

Define

$$q'(xym') = q(xyD(m')) \cdot \prod_{i=2}^{C} t_{D(m')_i}(m'_i).$$

The definition ensures that $q'(xym')$ is rectangular, and that conditioned on $D(m')$, xy is independent of m'. Define the distribution $p'(xymm')$ as follows:

$$p'(xym'_0m'_1) = p(xym'_0m'_1),$$

and for $i > 1$,

$$p'(m_im'_i|xym_{<i}m'_{<i}) = \begin{cases} p(m_i|xym_{<i} = D(m')_{<i}) \cdot s_{m_i}(m'_i), & \text{if } p(xym_{<i} = D(m')_{<i}) > 0, \\ 1/2 \cdot s_{1/2}(m'_i), & \text{otherwise.} \end{cases}$$

In words, in the protocol $p'(xym')$, the parties privately sample each message bit m_i according to the protocol distribution p. However, instead of sending this sampled bit, they

send m'_i sampled according to $s_{m_i}(m'_i)$. After this transmission, they continue the protocol using $D(m')_{<i}$. Strictly speaking, in order to ensure that the new protocol is a protocol distribution, we require that all the odd bits are transmitted by Alice and all even bits are sent by Bob. This can be easily achieved by inserting random bits into the transcript, but we leave out the details here.

There is some small chance that for $i > 1$, $D(m')_i \neq m_i$, but by the Chernoff bound,

$$p'(D(m')_i \neq m_i) \leq \exp(-\Omega(\beta^2 L)).$$

We have that for all xym' in the support of q',

$$\frac{q'(xym')}{p'(xym')} = \frac{q(xyD(m'))}{p'(xyD(m'))} \cdot \prod_{i=2}^{C} \frac{t_{D(m')_i}(m'_i)}{p'(m'_i|xyD(m')_{\leq i})}.$$

For any $xyD(m')$ such that $q(xyD(m')) > 0$, we can bound

$$\begin{aligned}
\frac{q(xyD(m'))}{p'(xyD(m'))} &= \frac{q(xyD(m'))}{p(xyD(m'))} \cdot \prod_{i=2}^{C} \frac{p(D(m')_i|xyD(m')_{<i})}{p'(D(m')_i|xyD(m')_{<i})} \\
&\leq \frac{q(xyD(m'))}{p(xyD(m'))} \cdot \prod_{i=2}^{C} \frac{1}{1-\exp(-\Omega(\beta^2 L))},
\end{aligned}$$

where we assumed that $p(xyD(m')) > 0$; if $p(xyD(m')) = 0$ then $q(xyD(m')) = 0$ for otherwise the marginal information cost would be unbounded. Next,

$$\begin{aligned}
\prod_{i=2}^{C} \frac{t_{D(m')_i}(m'_i)}{p'(m'_i|xyD(m')_{\leq i})} &= \prod_{i=2}^{C} \frac{t_{D(m')_i}(m'_i)}{\mathbf{E}_{p'(m_i|xyD(m')_{\leq i})}[p'(m'_i|xyD(m')_{\leq i}, m_i)]} \\
&= \prod_{i=2}^{C} \frac{t_{D(m')_i}(m'_i)}{\mathbf{E}_{p'(m_i|xyD(m')_{\leq i})}[s_{m_i}(m'_i|D(m'_i))]} \\
&= \prod_{i=2}^{C} \frac{t_{D(m')_i}(m'_i)}{p'(m_i = D(m')_i) \cdot t_{D(m')_i}(m'_i) + p'(m_i \neq D(m')_i) \cdot t'_{D(m')_i}(m'_i)} \\
&= \prod_{i=2}^{C} \frac{1}{p'(m_i = D(m')_i) + p'(m_i \neq D(m')_i) \cdot \frac{t'_{D(m')_i}(m'_i)}{t_{D(m')_i}(m'_i)}} \\
&\leq \prod_{i=2}^{C} \frac{1}{p'(m_i = D(m')_i)} \\
&\leq \prod_{i=2}^{C} \frac{1}{1-\exp(-\Omega(\beta^2 L))}.
\end{aligned}$$

So, we obtain the bound:

$$\frac{q'(xym')}{p'(xym')} = \frac{q(xyD(m'))}{p(xyD(m'))} \cdot (1 + 2C\exp(-\Omega(\beta^2 L))).$$

Moreover, for all xym' in the support of q', we have

$$\begin{aligned} q'(xy|m') &= \frac{q'(xym')}{q'(m')} \\ &= \frac{q(xyD(m')) \cdot \prod_{i=1}^{C} t_{D(m')_i}(m'_i)}{q(D(m')) \cdot \prod_{i=1}^{C} t_{D(m')_i}(m'_i)} \\ &= q(xy|D(m')). \end{aligned}$$

Finally, since $q'(xy|m') = q(xy|D(m'))$, we have

$$\Big| \mathop{\mathbf{E}}_{q'(xy|m')}[(-1)^f] \Big| = | \mathop{\mathbf{E}}_{q'(xy|D(m'))}[(-1)^f] \Big|.$$

Thus, we get that

$$\mathsf{M}^{\mathsf{ext}}_I(p', f) \leq \mathsf{M}^{\mathsf{ext}}_I(p, f) + IC \cdot \exp(-\Omega(\beta^2 L)).$$

Setting $L = O(\log(IC)/\beta^2)$ proves the theorem. □

Smooth protocols have the feature that the log-ratios of the information terms are tightly concentrated. To explain this phenomenon, we need to introduce a few definitions. For every xym in the support of p, and $j \geq 2$, define the j-th divergence costs:

$$\begin{aligned} d^A_j(xm) &= \sum_{\substack{2 \leq i \leq j \\ i \text{ odd}}} \mathop{\mathbf{E}}_{p(m_i|xm_{<i})} \Big[\log \frac{p(m_i|xm_{<i})}{p(m_i|m_{<i})} \Big], \\ d^B_j(ym) &= \sum_{\substack{2 \leq i \leq j \\ i \text{ even}}} \mathop{\mathbf{E}}_{p(m_i|ym_{<i})} \Big[\log \frac{p(m_i|ym_{<i})}{p(m_i|m_{<i})} \Big], \\ d_j(xym) &= d^A_j(xm) + d^B_j(ym). \end{aligned}$$

By the non-negativity of divergence, the divergence costs are monotone i.e. $d_j(xym) \leq d_{j+1}(xym)$. Since the protocol is β-smooth, we have

$$\begin{aligned} d^A_{j+1}(xm) - d^A_j(xm) &\leq \log \frac{1/2 + \beta}{1/2 - \beta} \leq 5\beta, \\ d^B_{j+1}(ym) - d^B_j(ym) &\leq \log \frac{1/2 + \beta}{1/2 - \beta} \leq 5\beta. \end{aligned} \tag{4.51}$$

We say a function $r(xym)$ taking values in $\{1,\dots,C\}$ is a *frontier* if every m contains exactly one prefix of the type $m'_{\le r(xym')}$, and that is the prefix $m_{\le r(xym)}$. Alternatively, for every m, m' such that $r(xym) \neq r(xym')$, it holds that both $r(xym)$ and $r(xym')$ are larger than the length of the longest common prefix of m and m'. Given a frontier $r(xym)$, define

$$\begin{aligned}
F_{r,\alpha} &= \Big\{xym : \Big| \sum_{i\ge 2}^{r(xym)} \log \frac{p(m_i|xym_{<i})}{p(m_i|m_{<i})} - d_{r(xym)}(xym)\Big| \ge \alpha\Big\},\\
F^A_{r,\alpha} &= \Big\{xym : \Big| \sum_{i\ge 2 \text{ odd}}^{r(xym)} \log \frac{p(m_i|xym_{<i})}{p(m_i|m_{\le i})} - d^A_{r(xym)}(xm)\Big| \ge \alpha\Big\},\\
F^B_{r,\alpha} &= \Big\{xym : \Big| \sum_{i\ge 2 \text{ even}}^{r(xym)} \log \frac{p(m_i|xym_{<i})}{p(m_i|m_{\le i})} - d^B_{r(xym)}(ym)\Big| \ge \alpha\Big\}. \qquad (4.52)
\end{aligned}$$

Lemma 4.6. *Let $r(xym)$ be a frontier such that for every xym, it holds that $d_{r(xym)}(xym) \le \tau$. Then $p(F_{r,\alpha}), p(F^A_{r,\alpha})$ and $p(F^B_{r,\alpha})$ are all at most $2\exp(-\Omega(\alpha^2/\tau))$.*

Proof. We prove the inequality for $p(F_{r,\alpha})$; the proofs for the other two terms are similar. Define the random variable $z_0, z_1 \dots$ where $z_0 = z_1 = 0$ and for every $i \ge 2$,

$$z_i = \begin{cases} \log \frac{p(m_i|xym_{<i})}{p(m_i|m_{<i})} & \text{if } i \le r(xym)\\ 0 & \text{otherwise.}\end{cases}$$

and let $t_i = z_i - \mathbf{E}_{p(m_i|xym_{<i})}[z_i]$. Then by definition $\mathbf{E}[t_i|t_{<i}] = 0$. Moreover, we have

$$\begin{aligned}
\sup(z_i|xym_{<i}) &\le \max_{m_i}\Big\{\log\frac{p(m_i|xym_{<i})}{p(m_i|m_{<i})}\Big\}\\
&\le \log\frac{1/2-\beta+\sqrt{d_i(xym)-d_{i-1}(xym)}}{1/2-\beta}\\
&\le O(\sqrt{d_i(xym)-d_{i-1}(xym)}).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\inf(z_i|xym_{<i}) &\ge \log\frac{1/2-\beta}{1/2-\beta-\sqrt{d_i(xym)-d_{i-1}(xym)}}\\
&\ge -O(\sqrt{d_i(xym)-d_{i-1}(xym)}).
\end{aligned}$$

So, if we define L as below, we have

$$\begin{aligned} L &= \sup_{xym} \sum_{i=2}^{C} (\sup(t_i|xym_{<i}) - \inf(t_i|xym_{<i}))^2 \\ &= \sup_{xym} \sum_{i=2}^{C} (\sup(z_i|xym_{<i}) - \inf(z_i|xym_{<i}))^2 \\ &\leq O(\tau). \end{aligned}$$

It is well known that if $\mathbf{E}[t_i] = 0$, then $\mathbf{E}[\exp(t_i)] \leq \exp((\sup(t_i) - \inf(t_i))^2/8)$ (see Lemma 2.6 in [32]). We can use this inequality to bound:

$$\begin{aligned} \mathop{\mathbf{E}}_{p(m|xy)} \Big[\exp(\frac{4\alpha}{L} \cdot \sum_{i=2}^{C} t_i) \Big] &\leq \mathop{\mathbf{E}}_{p(m_{\leq 2}|xy)} \Big[\exp(\frac{4\alpha}{L} t_2) \cdot \mathop{\mathbf{E}}_{p(m|xym_{<3})} \Big[\exp(\frac{4\alpha}{L} \cdot \sum_{i=3}^{C} t_i) \Big] \Big] \\ &\leq \dots \\ &\leq \exp\Big(\frac{(4\alpha/L)^2 \sup_{xym} \sum_{i=2}^{C} (\sup(t_i|xym_{<i}) - \inf(t_i|xym_{<i}))^2}{8} \Big) \\ &\leq \exp\Big(2\alpha^2/L\Big). \end{aligned}$$

So by Markov's inequality, we get:

$$p(\sum_{i=2}^{C} t_i > \alpha) \leq \mathbf{E}\Big[\exp(\frac{4\alpha}{L} \cdot \sum_{i=2}^{C} t_i) \Big] \cdot \exp(-4\alpha^2/L) \leq \exp(-\Omega(\alpha^2/\tau)).$$

Applying the same argument with $t_i = -t_i$ proves the other inequality. Defining z_i, t_i appropriately proves the other inequalities using the same proof. □

4.5 Compressing external marginal information

Here we prove Theorem 3.8.

Theorem 3.8 Restated. *For every $\alpha > 0$, there is a $\Delta > 0$ such that if $\mathsf{M}^{\mathsf{ext}}_I(p, f) \leq \alpha I$, $\mu(xy) = p(xy)$, and moreover the messages $m = (m_0, \dots, m_C)$ are such that $m_2, \dots, m_C \in \{0, 1\}$, then $\mathsf{adv}_\mu(\Delta I \log^2 C, f) \geq 1/\Delta$.*

Set $\mathsf{M}^{\mathsf{ext}} = \mathsf{M}^{\mathsf{ext}}_I(p, f) + KI$, for a large constant K to be chosen later. By Theorem 4.5, it is no loss of generality to assume that p is β-smooth, with $\beta = 1/(K \log(C 2^{5\mathsf{M}^{\mathsf{ext}}/I}))$. Let g_1, g_2 be as in Equation (3.10).

Define:

$$r_i^A(xm) = \begin{cases} \min\{j : d_j^A(xm) > 20\beta + d_{i-1}^A(xm)\} & \text{if such } j \text{ exists,} \\ C & \text{otherwise.} \end{cases}$$

$$r_i^B(ym) = \begin{cases} \min\{j : d_j^B(ym) > 20\beta + d_{i-1}^B(ym)\} & \text{if such } j \text{ exists,} \\ C & \text{otherwise.} \end{cases}$$

Note that $r_i^A(xm)$ is always odd, and $r_i^B(ym)$ is always even.

Because p is a protocol, we have

$$\mathop{\mathbf{E}}_{p(xy|m)}[d_j^A(xm) + d_j^B(ym)] = d_j(m).$$

Let ε be a parameter such that $\varepsilon \ll 2^{-5\mathsf{M}^{\mathsf{ext}}/I}$. Now, we describe a protocol Γ for computing $f(xy)$. Throughout this protocol, the parties will maintain a partial transcript $m_{<i}$. These partial transcripts may be inconsistent with each other, but we describe the protocol assuming that they are consistent with each other. In the analysis we shall show that the probability that the parties end up with inconsistent transcripts is negligible.

1. The parties sample m_0 using the distribution $p(m_0)$. The parties also sample a uniformly random function $h : \mathbb{Z} \to \{1, 2, \ldots, \lceil 1/\varepsilon \rceil\}$.

2. Run the protocol ψ from Lemma 2.5 with $u = p(m_1|m_0x), v = p(m_1|m_0), L = 5\mathsf{M}^{\mathsf{ext}}$, error parameter ε, to obtain functions a, b and transcript s. Alice sets $m_1^A = a(\psi(us))$, Bob sets $m_1^B = b(\psi(vs))$. If $m_1^B = \perp$, the protocol terminates. Bob sends a bit to Alice to indicate whether or not this occurs. The communication complexity of this step is $L + O(\log(1/\varepsilon))$.

3. Let $m_{\leq \ell}$ denote the part of the transcript sampled so far. Alice and Bob repeat the following steps until m corresponds to an entire transcript.

(a) Both parties use shared randomness to sample a full transcript $\widetilde{m}$ according to $p(m|m_{\leq \ell})$. They exchange the values of $r^A_{\ell+1}(x\widetilde{m})$ and $r^B_{\ell+1}(y\widetilde{m})$ to determine

$$k = \min\{r^A_{\ell+1}(x\widetilde{m}), r^B_{\ell+1}(y\widetilde{m})\}.$$

(b) Alice privately samples a number $\zeta^A \in [0,1]$ and sends 1 to Bob if

$$\zeta^A \leq \frac{1}{2} \cdot \prod_{\substack{i=\ell+2 \\ i \text{ odd}}}^{k} \frac{p(\widetilde{m}_i|x\widetilde{m}_{<i})}{p(\widetilde{m}_i|\widetilde{m}_{<i})},$$

and otherwise sends 0.

(c) Bob privately samples a number $\zeta^B \in [0,1]$ and sends 1 to Alice if

$$\zeta^B \leq \frac{1}{2} \cdot \prod_{\substack{i=\ell+2 \\ i \text{ even}}}^{k} \frac{p(\widetilde{m}_i|y\widetilde{m}_{<i})}{p(\widetilde{m}_i|\widetilde{m}_{<i})},$$

and otherwise sends 0.

(d) If both players receive 1 then, set $m_{\leq k} \leftarrow \widetilde{m}_{\leq k}$.

4. If $\eta^A \leq g_1(xm) \cdot 2^{-\lceil \log g_1(xm) \rceil}$, Alice sends $h(\lceil \log g_1(xm) \rceil)$ to Bob, and otherwise she sends $\perp$ to indicate that the protocol should be aborted.

5. If there is a unique integer z such that

$$\begin{aligned} |z + \log g_2(ym)| &\leq 3\mathsf{M}^{\mathsf{ext}}/I, \\ h(z) &= h(\lceil \log g_1(xm) \rceil), \\ \eta^B &\leq g_2(ym) \cdot 2^{z - 3\mathsf{M}^{\mathsf{ext}}/I}, \end{aligned}$$

Bob sends $\mathsf{sign}\Big(\mathbf{E}_{q(x'y'|m)}[(-1)^{f(x'y')}]\Big) \in \{\pm 1\}$ to Alice. Otherwise, he sends $\perp$ to abort the protocol.

To ensure the communication of the protocol is small, in our final protocol the parties abort and output a random bit if the communication in step 3 exceeds $(\mathsf{M}^{\mathsf{ext}} \cdot 2^{15\mathsf{M}^{\mathsf{ext}}/I} \cdot \log C)/\beta$. Then, the total communication is at most

$$5\mathsf{M}^{\mathsf{ext}} + \frac{\mathsf{M}^{\mathsf{ext}} \cdot 2^{15\mathsf{M}^{\mathsf{ext}}/I} \cdot \log C}{\beta} + O(\log 1/\varepsilon) = O(\mathsf{M}^{\mathsf{ext}} \cdot 2^{15\mathsf{M}^{\mathsf{ext}}/I} \cdot \log^2(C \cdot 2^{5\mathsf{M}^{\mathsf{ext}}/I})) \leq \Delta \cdot I \log^2 C,$$

for some Δ that depends only on α since $\mathsf{M}^{\mathsf{ext}} \leq (\alpha + K)I$.

Throughout the analysis below, we assume that in step 2, Alice always samples a message according to u, and Bob either accepts this sample or aborts, but never samples an inconsistent message. We can afford to make this assumption, because the probability of Bob sampling an inconsistent message without aborting is bounded by ε, which will be much smaller than our final advantage. Moreover, if Alice and Bob sample consistently in step 2 then the transcript they end up with after step 3 must be the same.

Let S and R be the sets defined in Equations (3.14) and (3.15) for our choice of K. In addition to S and R, we need the following sets to analyze the simulating protocol:

$$Q = \left\{ xym\eta^A\eta^B : \eta^A \leq g_1(xm) \cdot 2^{-\lceil \log g_1(xm) \rceil}, \eta^B \leq g_2(ym) \cdot 2^{\lceil \log g_1(xm) \rceil - 3\mathsf{M}^{\mathsf{ext}}/I} \right\},$$

$$\mathcal{Z} = \left\{ xymh : \exists \text{ unique } z \in \mathbb{Z} \text{ s.t. } |z + \log g_2(ym)| \leq \frac{3\mathsf{M}^{\mathsf{ext}}}{I} \text{ and } h(z) = h(\lceil \log g_1(xm) \rceil) \right\}.$$

Let $\mathcal{G}$ denote the event that the protocol reaches the final step without aborting and having communicated at most $(\mathsf{M}^{\mathsf{ext}} \cdot 2^{15\mathsf{M}^{\mathsf{ext}}/I} \cdot \log C)/\beta$ bits in step 3. Define $\mathcal{A}(xym) \in \{\pm 1\}$ by

$$\mathcal{A}(xym) = \mathsf{sign}\Big(\mathop{\mathbf{E}}_{q(x'y'|m)} [(-1)^{f(x'y')}] \Big) \cdot (-1)^{f(xy)}.$$

Our protocol computes $f(xy)$ correctly when $\mathcal{G}$ happens and $\mathcal{A}(xym) = 1$. The advantage of the protocol is at least

$$\Gamma(\mathcal{Z}QS\mathcal{G}) \cdot \mathop{\mathbf{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QS\mathcal{G}] - \Gamma(\mathcal{G}(\mathcal{Z}QS)^c) \tag{4.53}$$

We shall prove:

$$\mathop{\mathbf{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QS\mathcal{G}] \geq \Omega(2^{-\delta\mathsf{M}^{\mathsf{ext}}/(12I)}), \tag{4.54}$$

$$\Gamma(\mathcal{Z}QS\mathcal{G}) \geq \Omega(2^{-3\mathsf{M}^{\mathsf{ext}}/I}), \tag{4.55}$$

$$\Gamma(\mathcal{G}(\mathcal{Z}QS)^c) \leq O(2^{-4\mathsf{M}^{\mathsf{ext}}/I}). \tag{4.56}$$

By Equation (4.53), since $\delta \leq 1$, we can choose K to be large enough to prove the theorem, since $\alpha + K \geq \mathsf{M}^{\mathsf{ext}}/I \geq K$.

We first prove Equation (4.56). By the union bound, we have:

$$\Gamma(\mathcal{G}(\mathcal{Z}QS)^c) \leq \Gamma(\mathcal{Z}^c\mathcal{G}) + \Gamma(S^c\mathcal{G}\mathcal{Z}) + \Gamma(Q^c\mathcal{G}\mathcal{Z}S).$$

The definition of the protocol ensures that $\Gamma(\mathcal{Z}^c\mathcal{G}) = 0$. Moreover, $\Gamma(Q^c\mathcal{G}\mathcal{Z}S) = 0$, because if the event $\mathcal{Z}S$ happens and the parties do not abort, then:

$$\eta^A \leq g_1(xm) \cdot 2^{\lceil \log g_1(xm) \rceil} \text{ and}$$
$$\eta^B \leq g_2(ym) \cdot 2^{z - 3\mathsf{M}^{\mathsf{ext}}/I} = g_2(ym) \cdot 2^{\lceil \log g_1(xm) \rceil - 3\mathsf{M}^{\mathsf{ext}}/I}.$$

Additionally, $\Gamma(S^c\mathcal{G}\mathcal{Z}) \leq \Gamma(S^c\mathcal{Z}) \leq O(\varepsilon\mathsf{M}/I)$, since if $S^c\mathcal{Z}$ happens then there must have been a hash collision, which happens with probability at most $O(\varepsilon\mathsf{M}/I)$.

In order to prove Equations (4.54) and (4.55), we need to first establish that $\Gamma(xym)$ is typically quite close to $p(xym)$. Indeed, consider a particular execution of step 3 in the protocol. At this point, some prefix $m_{\leq \ell}$ has been fixed. For m consistent with this prefix $m_{\leq \ell}$, define the frontier

$$r(xym) = \min\{r^A_{\ell+1}(xm), r^B_{\ell+1}(ym)\}.$$

When the parties finally accept a sample, it will be a string $m_{\leq r(xym)}$ on the frontier. By the definition of $r^A_{\ell+1}, r^B_{\ell+1}$, and by Equation (4.51), we have that for all m, $d_{r(xym)}(m) - d_\ell(m) \leq 45\beta$. Setting $\tau = 45\beta$ and $\alpha = 1/4$, we apply Lemma 4.6 to conclude that if

$$F^A = \Big\{xym : \prod_{j=\ell+1 \text{ odd}}^{r(xym)} p(m_j | xym_{\leq \ell}) \geq 2 \cdot \prod_{j=\ell+1 \text{ odd}}^{r(xym)} p(m_{\leq r(xym)} | m_{\leq \ell})\Big\},$$
$$F^B = \Big\{xym : \prod_{j=\ell+1 \text{ even}}^{r(xym)} p(m_j | xym_{\leq \ell}) \geq 2 \cdot \prod_{j=\ell+1 \text{ even}}^{r(xym)} p(m_{\leq r(xym)} | m_{\leq \ell})\Big\}.$$

then

$$p(F^A \cup F^B | xym_{\leq \ell}) \leq 4\exp(-\Omega(1/\beta)) \leq C^{-1} \cdot 2^{-5\mathsf{M}^{\mathsf{ext}}/I}. \tag{4.57}$$

Now, we perform a standard analysis of rejection sampling. Let W denote the event that the first sample of $m_{r(xym)}$ is accepted in the protocol. Given $xym_{\leq \ell}$, the probability that W occurs is

$$\begin{aligned}\Gamma(W | xym_{\leq \ell}) &\geq \sum_{m'_{r(xym')} : xym'_{r(xym)} \notin F^A \cup F^B} p(m'_{\leq r(xym')} | xym_{\leq \ell})/4 \\ &\geq 1/4 - p(F^A \cup F^B | xym_{\leq \ell})/4 \geq 1/4 - C^{-1} \cdot 2^{-5\mathsf{M}^{\mathsf{ext}}/I} \geq 1/8, \end{aligned} \tag{4.58}$$

where here we abused notation to write $xym'_{r(xym)} \notin F^A \cup F^B$ to mean that the prefix is not consistent with any m in $F^A \cup F^B$.

It is clear that the sampled point is independent of the event $\neg W$, so it is also independent of W. So, the probability that a particular prefix $m_{r(xym)}$ is sampled is the same as the probability that it is sampled conditioned on W. When $m_{r(xym)}$ is not consistent with $F^A \cup F^B$, the probability of such a point is

$$\frac{p(m_{\leq r(xym)}|xym_{\leq \ell})/4}{1/4 - p(F^A \cup F^B|xym_{\leq \ell})/4} = p(m_{\leq r(xym)}|xym_{\leq \ell}) \cdot (1 \pm O(C^{-1} \cdot 2^{-5\mathsf{M}^{\mathsf{ext}}/I})). \tag{4.59}$$

Let B denote the event that the final sample xym is such that at some point a prefix was sampled in $F^A \cup F^B$ during step 3. Whenever step 3 accepts a sample, the length of the transcript increases by at least 1, so the number of times step 3 accepts a sample is at most C. Thus, by the union bound and Equation (4.57),

$$p(B) \leq O(2^{-5\mathsf{M}^{\mathsf{ext}}/I}). \tag{4.60}$$

Moreover, by Equation (4.59), for $xym \notin B$,

$$\Gamma(xym) = p(xym) \cdot (1 \pm O(2^{-5\mathsf{M}^{\mathsf{ext}}/I})). \tag{4.61}$$

Equations (4.60) and (4.61) imply

$$\Gamma(B) = 1 - \Gamma(B^c) \leq 1 - p(B^c) \cdot (1 - O(2^{-5\mathsf{M}^{\mathsf{ext}}/I})) \leq O(2^{-5\mathsf{M}^{\mathsf{ext}}/I}). \tag{4.62}$$

Additionally, we have

$$q(SB^c) = q(S) - q(BS) \geq q(S) - 2^{3\mathsf{M}^{\mathsf{ext}}/I} \cdot p(B) \geq 1 - \Omega(2^{-\mathsf{M}/I}), \tag{4.63}$$

by the definition of S, Claim 3.26 and Eq. (4.60).

Now we can begin to understand $\Gamma(\mathcal{Z}QS\mathcal{G})$. For $xym \in S$,

$$\Gamma(Q|xym) = \frac{g_1(xm) \cdot 2^{-\lceil \log g_1(xm) \rceil} \cdot g_2(ym) \cdot 2^{\lceil \log g_1(xm) \rceil}}{2^{3\mathsf{M}^{\mathsf{ext}}/I}} = \frac{q(xym)}{p(xym)} \cdot \frac{1}{2^{3\mathsf{M}^{\mathsf{ext}}/I}}, \tag{4.64}$$

where the first equality follows from the fact that

$$g_2(ym) \cdot 2^{\lceil \log g_1(xm) \rceil} = 2^{\lceil \log g_1(xm) \rceil + \log g_2(ym)} \leq 2^{3\mathsf{M}/I},$$

by the definition of S.

We can bound

$$\begin{aligned}
\Gamma(QSB^c) &= \sum_{xym \in S \cap B^c} \Gamma(xym) \cdot \Gamma(Q|xym) \\
&= \sum_{xym \in S \cap B^c} p(xym) \cdot \Gamma(Q|xym) \cdot (1 \pm O(2^{-5\mathsf{M}^{\mathsf{ext}}/I})). \qquad \text{(by Equation (4.61))} \\
&= 2^{-3\mathsf{M}^{\mathsf{ext}}/I} \cdot q(SB^c) \cdot (1 \pm O(2^{-5\mathsf{M}^{\mathsf{ext}}/I})) \qquad (4.65) \\
&= \Omega(2^{-3\mathsf{M}^{\mathsf{ext}}/I}), \qquad (4.66)
\end{aligned}$$

by Equations (4.60) and (4.63). We claim that for all $xym \in SB^c$,

$$\Gamma(\mathcal{Z}|xymQSB^c) = \Gamma(\mathcal{Z}|xym) \geq 1 - O(\varepsilon \mathsf{M}^{\mathsf{ext}}/I). \qquad (4.67)$$

The equality follows by observing that xym determine SB^c, and given xym, $\mathcal{Z}$ just depends on the choice of h, which is independent of Q. The inequality follows from the fact that for each xym in S, the event $\mathcal{Z}^c$ can happen only if there exists an integer z distinct from $\lceil \log g_1(xm) \rceil$ such that $h(\lceil \log g_1(xm) \rceil) = h(z)$ and $|z + \log g_2(ym)| \leq 3\mathsf{M}^{\mathsf{ext}}/I$. The probability that this happens is at most $O(\varepsilon \cdot \mathsf{M}^{\mathsf{ext}}/I)$. In particular, this implies

$$\Gamma(\mathcal{Z}|QS) \geq 1 - O(\varepsilon \mathsf{M}^{\mathsf{ext}}/I). \qquad (4.68)$$

For $xym \in S \cap B^c$, we have

$$\begin{aligned}
\Gamma(xym|\mathcal{Z}QSB^c) &= \frac{\Gamma(xym) \cdot \Gamma(\mathcal{Z}QSB^c|xym)}{\Gamma(\mathcal{Z}QSB^c)} \\
&= \frac{p(xym) \cdot \Gamma(Q|xym) \cdot \Gamma(\mathcal{Z}|xym)}{\Gamma(\mathcal{Z}QSB^c)} \cdot (1 \pm O(2^{-5\mathsf{M}^{\mathsf{ext}}/I})) \\
&\qquad \text{(by Equation (4.61), and since } xym \text{ determine } S, B^c) \\
&= \frac{p(xym)}{\Gamma(QSB^c)} \cdot \frac{q(xym)}{p(xym) \cdot 2^{3\mathsf{M}^{\mathsf{ext}}/I}} \cdot \frac{\Gamma(\mathcal{Z}|xym)}{\Gamma(\mathcal{Z}|QSB^c)} \cdot (1 \pm O(2^{-5\mathsf{M}^{\mathsf{ext}}/I})) \\
&\qquad \text{(By Equation (4.64))} \\
&= \frac{q(xym)}{q(SB^c)} \cdot (1 \pm O(\varepsilon \mathsf{M}^{\mathsf{ext}}/I + 2^{-5\mathsf{M}^{\mathsf{ext}}/I})). \\
&\qquad \text{(By Equations (4.65) and (4.67))} \\
&= q(xym|SB^c) \cdot (1 \pm O(\varepsilon \mathsf{M}^{\mathsf{ext}}/I + 2^{-5\mathsf{M}^{\mathsf{ext}}/I})). \qquad (4.69)
\end{aligned}$$

To argue that the protocol does not have too much communication, we show that typically the divergence costs of the accepted transcripts are small. Define the sets

$$H = \{xym : \log \frac{p(m|xy)}{p(m)} \leq \mathsf{M}^{\mathsf{ext}} \cdot 2^{10\mathsf{M}^{\mathsf{ext}}/I}\},$$
$$F = \{xym : d_C(xym) > 2\mathsf{M}^{\mathsf{ext}} \cdot 2^{10\mathsf{M}^{\mathsf{ext}}/I}\}$$

and the frontier

$$r(xym) = \begin{cases} \min\{i : d_i(xym) > 2\mathsf{M}^{\mathsf{ext}} \cdot 2^{10\mathsf{M}^{\mathsf{ext}}/I}\} & \text{if such } i \text{ exists,} \\ C & \text{otherwise.} \end{cases}$$

We have $F \cap H \subseteq F_{r,\mathsf{M}^{\mathsf{ext}} \cdot 2^{10\mathsf{M}^{\mathsf{ext}}/I}}$, where $F_{r,\mathsf{M}^{\mathsf{ext}} \cdot 2^{10\mathsf{M}^{\mathsf{ext}}/I}}$ is the set from Equation (4.52). By Equation (4.51), and the the choice of r, $d_{r(xym)}(xym) \leq 2\mathsf{M}^{\mathsf{ext}} \cdot 2^{10\mathsf{M}^{\mathsf{ext}}/I} + 5\beta$, so we can apply Lemma 4.6 to conclude that

$$p(FH) \leq p(F_{r,\mathsf{M}^{\mathsf{ext}} \cdot 2^{10\mathsf{M}^{\mathsf{ext}}/I}}) \leq 2\exp(-\Omega(\mathsf{M}^{\mathsf{ext}} \cdot 2^{10\mathsf{M}^{\mathsf{ext}}/I})). \tag{4.70}$$

We have

$$\begin{aligned} q(F|SB^c) &\leq \frac{q(FS)}{q(SB^c)} \\ &\leq \frac{q(FHS) + q(H^c)}{q(SB^c)} \\ &\leq O(q(FHS) + q(H^c)) && \text{(by Equation (4.63))} \\ &\leq O(q(FHS) + 2^{-10\mathsf{M}^{\mathsf{ext}}/I}) && \text{(by Markov's inequality and Equation (3.22))} \\ &\leq O(p(FHS) \cdot 2^{3\mathsf{M}^{\mathsf{ext}}/I} + 2^{-10\mathsf{M}^{\mathsf{ext}}/I}) && \text{(using the definition of } S) \\ &\leq O(\exp(-\Omega(\mathsf{M}^{\mathsf{ext}} \cdot 2^{10\mathsf{M}^{\mathsf{ext}}/I})) \cdot 2^{3\mathsf{M}^{\mathsf{ext}}/I} + 2^{-10\mathsf{M}^{\mathsf{ext}}/I}), \end{aligned}$$

by Equation (4.70). Putting this bound back into Equation (4.69), we get

$$\Gamma(F|\mathcal{Z}QSB^c) \leq O(2^{-10\mathsf{M}^{\mathsf{ext}}/I}) \tag{4.71}$$

We note that every time step 3 accepts a sample, the divergence cost of the transcript increases by 20β, and in expectation, the number of rounds of rejection sampling involved to accept a sample is at most 8 by Equation (4.58) and a standard calculation. Moreover,

in each round, the players communicate at most $2 + 2\log C$ bits to exchange two indices in $\{1, \ldots, C\}$. Hence, given xy and a transcript m the expected communication to sample m is at most $16 \cdot (1 + \log C) \cdot d_C(xym)/(20\beta)$. Recall that $\mathcal{G}$ occurs when the protocol reaches the final step having communicated at most $(\mathsf{M}^{\mathsf{ext}} \cdot 2^{15\mathsf{M}^{\mathsf{ext}}/I} \cdot \log C)/\beta$. Thus, Markov's inequality implies that

$$\Gamma(\mathcal{G}|\mathcal{Z}QSB^cF^c) = 1 - O(2^{-5\mathsf{M}^{\mathsf{ext}}/I}). \tag{4.72}$$

So, we can conclude that

$$\begin{aligned}
\Gamma(\mathcal{Z}QS\mathcal{G}) &\geq \Gamma(\mathcal{Z}QSB^cF^c\mathcal{G}) \\
&\geq \Gamma(QSB^c) \cdot \Gamma(\mathcal{Z}|QSB^c) \cdot \Gamma(F^c|\mathcal{Z}QSB^c) \cdot \Gamma(\mathcal{G}|\mathcal{Z}QSB^cF^c) \\
&\geq \Omega(2^{-3\mathsf{M}^{\mathsf{ext}}/I}), \qquad \text{(by Equations (4.66), (4.67), (4.71) and (4.72))}
\end{aligned}$$

proving Equation (4.55). Observe that by Equations (4.55), (4.62) and (4.71),

$$\Gamma(B|\mathcal{Z}QS\mathcal{G}) \leq \frac{\Gamma(B)}{\Gamma(\mathcal{Z}QS\mathcal{G})} \leq O(2^{-2\mathsf{M}^{\mathsf{ext}}/I}), \tag{4.73}$$

$$\Gamma(F|\mathcal{Z}QS\mathcal{G}) \leq \frac{\Gamma(F\mathcal{Z}QSB^c) + \Gamma(B)}{\Gamma(\mathcal{Z}QS\mathcal{G})} \leq O(2^{-2\mathsf{M}^{\mathsf{ext}}/I}). \tag{4.74}$$

Moreover, we have

$$\begin{aligned}
\mathop{\mathbf{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QSB^cF^c] &\leq \Gamma(\mathcal{G}|\mathcal{Z}QSB^cF^c) \cdot \mathop{\mathbf{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QS\mathcal{G}B^cF^c] + \Gamma(\mathcal{G}^c|\mathcal{Z}QSB^cF^c) \\
&\leq \mathop{\mathbf{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QS\mathcal{G}B^cF^c] + O(2^{-5\mathsf{M}^{\mathsf{ext}}/I}), \qquad (4.75) \\
\mathop{\mathbf{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QSB^c] &\leq \Gamma(F^c|\mathcal{Z}QSB^c) \cdot \mathop{\mathbf{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QSB^cF^c] + \Gamma(F|\mathcal{Z}QSB^c) \\
&\leq \mathop{\mathbf{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QSB^cF^c] + O(2^{-10\mathsf{M}^{\mathsf{ext}}/I}), \qquad (4.76)
\end{aligned}$$

by Equation (4.71).

We are now ready to prove Equation (4.54). We have

$$
\begin{aligned}
&\mathop{\mathbf{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QS\mathcal{G}] \\
&\geq \Gamma(B^cF^c|\mathcal{Z}QS\mathcal{G}) \cdot \mathop{\mathbf{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QS\mathcal{G}B^cF^c] - \Gamma(B|\mathcal{Z}QS\mathcal{G}) - \Gamma(F|\mathcal{Z}QS\mathcal{G}) \\
&\geq (1 - O(2^{-2\mathsf{M}^{\mathrm{ext}}/I})) \cdot \mathop{\mathbf{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QS\mathcal{G}B^cF^c] - \Omega(2^{-2\mathsf{M}^{\mathrm{ext}}/I}) \\
&\qquad\qquad \text{(by Equations (4.73) and (4.74))} \\
&\geq (1/2) \cdot \mathop{\mathbf{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QSB^cF^c] - \Omega(2^{-5\mathsf{M}^{\mathrm{ext}}/I}) - \Omega(2^{-2\mathsf{M}^{\mathrm{ext}}/I}) \qquad \text{(by Equation (4.75))} \\
&\geq (1/2) \cdot \mathop{\mathbf{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QSB^c] - \Omega(2^{-10\mathsf{M}^{\mathrm{ext}}/I}) - \Omega(2^{-2\mathsf{M}^{\mathrm{ext}}/I}) \qquad \text{(by Equation (4.76))} \\
&\geq (1/4) \cdot \mathop{\mathbf{E}}_{q}[\mathcal{A}(xym)|SB^c] - \Omega(2^{-2\mathsf{M}^{\mathrm{ext}}/I}) \qquad \text{(by Equation (4.69))} \\
&\geq (1/4) \cdot \mathop{\mathbf{E}}_{q(xym)}[\mathcal{A}(xym)] - \Omega(2^{-\mathsf{M}^{\mathrm{ext}}/I}) \qquad \text{(by Equation (4.63))} \\
&= (1/4) \cdot \mathop{\mathbf{E}}_{q(xym)}\left[\mathsf{sign}\left(\mathop{\mathbf{E}}_{q(x'y'|m)}\left[(-1)^{f(x'y')}\right]\right) \cdot (-1)^{f(xy)}\right] - \Omega(2^{-\mathsf{M}^{\mathrm{ext}}/I}) \\
&= (1/4) \cdot \mathop{\mathbf{E}}_{q(m)}\left[\left|\mathop{\mathbf{E}}_{q(xy|m)}\left[(-1)^{f(xy)}\right]\right|\right] - \Omega(2^{-\mathsf{M}^{\mathrm{ext}}/I}) \\
&\geq \Omega(2^{-\delta\mathsf{M}^{\mathrm{ext}}/(12I)}). \qquad \text{(by Equation (3.23))}
\end{aligned}
$$

This concludes the proof of the theorem.

4.6 Compressing bounded-round protocols

We prove Theorem 3.10 in this section.

Theorem 3.10 Restated. *For every $\alpha > 0$, there is a $\Delta > 0$ such that if $\mathsf{M}_I(p, f) \leq \alpha I$, $\mu(xy) = p(xy)$, p has r-rounds and $m_r \in \{0, 1\}$, then $\mathsf{adv}^r_\mu(\Delta r(I + \log r), f) \geq 1/\Delta$.*

Let $p(xym)$ be a protocol distribution such that $p(xy) = \mu(xy)$. In the bounded-round setting, we have $m = (m_0, \ldots, m_r)$ – the transcript consisting of r messages along with the shared randomness. By assumption, $\mathsf{M}_I(p, f) = \alpha I$, and $m_r \in \{0, 1\}$. We assume without loss of generality that r is even. Let $q(xym)$ be a rectangular distribution that realizes $\mathsf{M}_I(p, f)$. For a large constant K, let $\mathsf{M} = \mathsf{M}_I(p, f) + KI$. Since $\mathsf{M}(p, f) \geq 0$, we have $\mathsf{M} \geq KI$. Let g_1, g_2 be as in Equation (3.10). Let ε be a parameter such that $\varepsilon = (2^{4\mathsf{M}/I} \cdot (r + 1))^{-1}$.

We define a protocol Γ whose communication complexity is bounded by

$$O(r \cdot (\mathsf{M} + \log(r/\varepsilon))).$$

Since $\mathsf{M}_I(p, f) = \alpha I$ we get that $\mathsf{M} \leq (\alpha + K) \cdot I$, and it follows that the communication is bounded by $\Delta r(I + \log r)$ for some Δ that only depends on α. Now, we describe Γ.

1. Jointly sample $p(m_0)$. Alice sets $m_0^A = m_0$ and Bob sets $m_0^B = m_0$. Jointly sample $\eta_0^A, \eta_1^A, \eta^B \in [0, 1]$ uniformly and independently. Jointly sample a uniformly random function $h : \mathbb{Z} \to \{1, \ldots, \lceil 1/\varepsilon \rceil\}$.

2. For each $i \in \{1, \ldots, r-1\}$:

   (a) If i is odd, run the protocol ψ from Lemma 2.5 with $u = p(m_i | m_{<i}^A x)$, $v = p(m_i | m_{<i}^B y)$, $L = 14\mathsf{M} + 5\log(r+1)$ and error parameter ε, to obtain functions a_i, b_i and transcript s. Alice sets $m_i^A = a_i(us)$, Bob sets $m_i^B = b_i(vs)$. If $m_i^B = \perp$, Bob signals to abort in the next round and sends a random bit to Alice, which they both output.

   (b) If i is even, run the protocol ψ from Lemma 2.5 with $u = p(m_i | m_{<i}^B y)$, $v = p(m_i | m_{<i}^A x)$, $L = 14\mathsf{M} + 5\log(r+1)$ and error parameter ε, to obtain functions a_i, b_i and transcript s. Bob sets $m_i^B = a_i(us)$, Alice sets $m_i^A = b_i(vs)$. If $m_i^A = \perp$, Alice signals to abort in the next round and sends a random bit to Bob, which they both output.

   Let $\langle m^A \rangle, \langle m^B \rangle$ denote the values of m^A and m^B after the first $r-1$ rounds.

3. For each $b \in \{0, 1\}$, Alice sends a message to Bob. If

$$\eta_b^A \leq \log g_1(x\langle m^A \rangle b) \cdot 2^{-\lceil \log g_1(x\langle m^A \rangle b) \rceil},$$

   Alice sends $h(\lceil \log g_1(x\langle m^A \rangle b) \rceil)$ to Bob, otherwise she sends 0.

4. Bob samples a bit b according to $p(m_r|\langle m^B\rangle y)$. If there is a unique integer z such that

$$|z + \log g_2(y\langle m^B\rangle b)| \le 3\mathsf{M}/I,$$
$$h(z) = h(\lceil \log g_1(x\langle m^A\rangle b)\rceil),$$
$$\eta^B \le g_2(y\langle m^B\rangle b)\cdot 2^{z-3\mathsf{M}/I},$$

Bob sends $\mathsf{sign}\Big(\mathbf{E}_{q(x'y'|\langle m^B\rangle b)}[(-1)^{f(x'y')}]\Big) \in \{\pm 1\}$ to Alice. Otherwise, he sends $\perp$ to abort the protocol.

We note that the above protocol involves at most r rounds of communication, and in each of the first $r-1$ rounds, the communication from step 2 is at most

$$14\mathsf{M} + 5\log(r+1) + O(\log 1/\varepsilon) \le O(\mathsf{M} + \log(r/\varepsilon)).$$

In step 3, Alice additionally sends $O(\log 1/\varepsilon)$ bits for the hashes. Hence, the total communication is at most $O(r\cdot(\mathsf{M} + \log(r/\varepsilon)))$.

We may assume that at the beginning of Γ, the players sample r independent random tapes, where the i-th random tape is used for the i-th execution of the protocol ψ from Lemma 2.5 in step 2 of Γ. Given this assumption, define m as follows: $m_0 = m_0^A = m_0^B$, and for all $i \ge 1$, $m_i = a_i(p(m_i|m_{<i}xy)s)$, where s is a transcript of the protocol ψ from Lemma 2.5 that is determined given $x, y, m_{<i}$ and the i-th random tape, and a_i is the function promised by the lemma. From item 1 of Lemma 2.5, it is clear that $\Gamma(xym) = p(xym)$.

Let S be the set defined in Equation (3.11) for our choice of K. In addition to S, we need the following sets to analyze the simulating protocol.

$$Q = \Big\{xym\eta^A_{m_r}\eta^B : \eta^A_{m_r} \le g_1(xm)\cdot 2^{-\lceil\log g_1(xm)\rceil}, \eta^B \le g_2(ym)\cdot 2^{\lceil\log g_1(xm)\rceil - 3\mathsf{M}/I}\Big\},$$
$$\mathcal{E} = \{\langle m^A\rangle\langle m^B\rangle m_{<r} : \langle m^A\rangle = \langle m^B\rangle = m_{<r}\},$$
$$\mathcal{Z} = \{xymh : \exists \text{ unique } z\in\mathbb{Z} \text{ s.t. } |z + \log g_2(ym)| \le 3\mathsf{M}/I \text{ and } h(z) = h(\lceil\log g_1(xm)\rceil)\}.$$

Let $\mathcal{G}$ denote the event that the protocol reaches the final step without aborting, and define $\mathcal{A}(xym)\in\{\pm 1\}$ by

$$\mathcal{A}(xym) = \mathsf{sign}\Big(\mathop{\mathbf{E}}_{q(x'y'|m)}[(-1)^{f(x'y')}]\Big)\cdot(-1)^{f(xy)}.$$

Our protocol computes $f(xy)$ correctly when: $\mathcal{G}$ happens, $\mathcal{A}(xym) = 1$ and $m_{<r} = \langle m^B \rangle$. Since $\mathcal{EZSQ} \subseteq \mathcal{G}$, and $\mathcal{E}$ implies $m_{<r} = \langle m^B \rangle$, the advantage of our protocol is at least:

$$\Gamma(\mathcal{EZSQ}) \cdot \mathop{\mathbf{E}}_{\Gamma(xym|\mathcal{EZSQ})}[\mathcal{A}(xym)] - \Gamma(\mathcal{G}(\mathcal{EZSQ})^c). \tag{4.77}$$

We shall prove each of the following bounds:

$$\mathop{\mathbf{E}}_{\Gamma(xym|\mathcal{EZQS})}[\mathcal{A}(xym)] \geq \Omega(2^{-\delta \mathsf{M}/(12I)}), \tag{4.78}$$

$$\Gamma(\mathcal{EZQS}) \geq \Omega(2^{-3\mathsf{M}/I}), \tag{4.79}$$

$$\Gamma(\mathcal{G}(\mathcal{EZSQ})^c) \leq O(2^{-4\mathsf{M}/I}). \tag{4.80}$$

Because $\delta \leq 1$ and $(\alpha + K) \geq \mathsf{M}/I \geq K$, we can choose K to be large enough to prove the theorem.

We first upper bound $\Gamma(\mathcal{G}(\mathcal{EZSQ})^c)$. By the union bound, we have:

$$\Gamma(\mathcal{G}(\mathcal{EZSQ})^c) \leq \Gamma(\mathcal{GE}^c) + \Gamma(\mathcal{Z}^c|\mathcal{GE}) + \Gamma(S^c\mathcal{GEZ}) + \Gamma(Q^c|\mathcal{GEZS}).$$

The definition of the protocol ensures that $\Gamma(\mathcal{Z}^c|\mathcal{GE}) = 0$. Moreover, we claim that $\Gamma(Q^c|\mathcal{GEZS}) = 0$, because if the event $\mathcal{EZS}$ happens and the parties do not abort, then

$$\eta^A_{m_r} \leq g_1(x\langle m^A \rangle m_r) \cdot 2^{\lceil \log g_1(x\langle m^A \rangle m_r) \rceil} = g_1(xm) \cdot 2^{\lceil \log g_1(xm) \rceil},$$
$$\eta^B \leq g_2(y\langle m^B \rangle m_r) \cdot 2^{z - 3\mathsf{M}/I} = g_2(ym) \cdot 2^{\lceil \log g_1(xm) \rceil - 3\mathsf{M}/I}.$$

The event $\mathcal{GE}^c$ implies that ψ made an error in one of the r rounds, leaving Alice and Bob with strings that were not equal. The probability that this happens is at most $\varepsilon \cdot r \leq 2^{-4\mathsf{M}/I}$, by our choice of ε. Finally, $\Gamma(S^c\mathcal{GEZ}) \leq \Gamma(S^c\mathcal{EZ}) \leq O(\varepsilon \mathsf{M}/I)$, since if $S^c\mathcal{EZ}$ happens then there must have been a hash collision, which happens with probability at most $O(\varepsilon \mathsf{M}/I)$. This implies Equation (4.80).

Now, we turn to proving Equation (4.79). Let us first estimate $\Gamma(QS)$. We have,

$$\Gamma(QS) = \sum_{xym \in S} \Gamma(xym) \cdot \Gamma(Q|xym) = \sum_{xym \in S} p(xym) \cdot \Gamma(Q|xym).$$

For $xym \in S$,

$$\Gamma(Q|xym) = \frac{g_1(xm) \cdot 2^{-\lceil \log g_1(xm) \rceil} \cdot g_2(ym) \cdot 2^{\lceil \log g_1(xm) \rceil}}{2^{3\mathsf{M}/I}} = \frac{q(xym)}{p(xym)} \cdot \frac{1}{2^{3\mathsf{M}/I}}, \tag{4.81}$$

where the first equality follows from the fact that

$$g_2(ym) \cdot 2^{\lceil \log g_1(xm) \rceil} = 2^{\lceil \log g_1(xm) \rceil + \log g_2(ym)} \leq 2^{3\mathsf{M}/I},$$

by the definition of S. Therefore,

$$\begin{aligned}\Gamma(QS) = \sum_{xym \in S} p(xym) \cdot \frac{q(xym)}{p(xym)} \cdot \frac{1}{2^{3\mathsf{M}/I}} = \frac{q(S)}{2^{3\mathsf{M}/I}} \\ \geq \frac{(1 - 5 \cdot 2^{-\mathsf{M}/I})}{2^{3\mathsf{M}/I}} = \Omega(2^{-3\mathsf{M}/I}),\end{aligned} \tag{4.82}$$

where in the last line, we used Claim 3.26.

We claim that for all $xym \in S$,

$$\Gamma(\mathcal{Z}|xymQS) = \Gamma(\mathcal{Z}|xym) \geq 1 - O(\varepsilon \mathsf{M}/I). \tag{4.83}$$

The equality follows by noting that xym determine S and given xym, $\mathcal{Z}$ just depends on the choice of h, which is independent of Q. The event $\mathcal{Z}^c$ can happen only if there exists an integer z distinct from $\lceil \log g_1(xm) \rceil$ such that $h(\lceil g_1(xm) \rceil) = h(z)$ and $|z + \log g_2(ym)| \leq 3\mathsf{M}/I$. The probability that this happens is at most $O(\varepsilon \cdot \mathsf{M}/I)$. Therefore, $\Gamma(\mathcal{Z}|xym) \geq 1 - O(\varepsilon \mathsf{M}/I) \geq 1/2$, by our choice of ε. We conclude that

$$\Gamma(QS\mathcal{Z}) = \Gamma(QS) \cdot \Gamma(\mathcal{Z}|QS) \geq \Omega(2^{-3\mathsf{M}/I}), \tag{4.84}$$

For all $xym \in S$,

$$\begin{aligned}\Gamma(xym|QS\mathcal{Z}) &= \frac{\Gamma(xym) \cdot \Gamma(QS\mathcal{Z}|xym)}{\Gamma(QS\mathcal{Z})} \\ &= \frac{p(xym)}{\Gamma(QS)} \cdot \Gamma(Q|xym) \cdot \frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} \\ &= \frac{p(xym)}{\Gamma(QS)} \cdot \frac{q(xym)}{p(xym) \cdot 2^{3\mathsf{M}/I}} \cdot \frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} && \text{(By Equation (4.81))} \\ &= \frac{q(xym)}{q(S)} \cdot \frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} && \text{(By Equation (4.82))} \\ &= q(xym|S) \cdot (1 \pm O(\varepsilon \mathsf{M}/I)), && (4.85)\end{aligned}$$

where the last line follows from Equation (4.83).

Given Equation (4.84), to complete the proof of Equation (4.79), it will be enough to prove that $\Gamma(\mathcal{E}|QS\mathcal{Z}) \geq 1/2$. Let T be the set T_K defined in Claim 3.28 for our choice of K. We have

$$\begin{aligned}
\Gamma(\mathcal{E}^c|QS\mathcal{Z}) &\leq \Gamma(T^c|QSZ) + \Gamma(\mathcal{E}^c|QS\mathcal{Z}T) \\
&\leq q(T^c|S) \cdot (1 + O(\varepsilon \mathsf{M}/I)) + \Gamma(\mathcal{E}^c|QS\mathcal{Z}T) && \text{(By Equation (4.85))} \\
&\leq O(2^{-\mathsf{M}/I}) + \Gamma(\mathcal{E}^c|QS\mathcal{Z}T) && \text{(By Claim 3.28)} \\
&\leq O(2^{-\mathsf{M}/I}) + 2\varepsilon \cdot r \leq O(2^{-\mathsf{M}/I}) \leq 1/2 && (4.86)
\end{aligned}$$

where in the last line, we used the fact that given $QS\mathcal{Z}T$, item 2 and 3 of Lemma 2.5 guarantee that $\mathcal{E}^c$ can only happen with probability at most 2ε in each of the r rounds. Equations (4.84) and (4.86) together prove Equation (4.79).

Next, we prove Equation (4.78). Since $|\mathcal{A}(xym)| \leq 1$, we have

$$\mathop{\mathbf{E}}_{\Gamma(xym|QS\mathcal{Z})}[\mathcal{A}(xym)] \leq \Gamma(\mathcal{E}|QS\mathcal{Z}) \cdot \mathop{\mathbf{E}}_{\Gamma(xym|QS\mathcal{Z}\mathcal{E})}[\mathcal{A}(xym)] + \Gamma(\mathcal{E}^c|QS\mathcal{Z}),$$

and since $\Gamma(\mathcal{E}|QS\mathcal{Z}) \leq 1$, this gives

$$\begin{aligned}
\mathop{\mathbf{E}}_{\Gamma(xym|QS\mathcal{Z}\mathcal{E})}[\mathcal{A}(xym)] &\geq \mathop{\mathbf{E}}_{\Gamma(xym|QS\mathcal{Z})}[\mathcal{A}(xym)] - \Gamma(\mathcal{E}^c|QS\mathcal{Z}) \\
&\geq \mathop{\mathbf{E}}_{q(xym|S)}[\mathcal{A}(xym)] - \Omega(\varepsilon\mathsf{M}/I + 2^{-\mathsf{M}/I}) \\
&\qquad \text{(using Equations (4.85) and (4.86))} \\
&\geq \mathop{\mathbf{E}}_{q(xym)}[\mathcal{A}(xym)] - \Omega(2^{-\mathsf{M}/I}) && \text{(by Claim 3.26)} \\
&= \mathop{\mathbf{E}}_{q(xym)}\left[\mathsf{sign}\left(\mathop{\mathbf{E}}_{q(x'y'|m)}\left[(-1)^{f(x'y')}\right]\right) \cdot (-1)^{f(xy)}\right] - \Omega(2^{-\mathsf{M}/I}) \\
&= \mathop{\mathbf{E}}_{q(m)}\left[\left|\mathop{\mathbf{E}}_{q(xy|m)}\left[(-1)^{f(xy)}\right]\right|\right] - \Omega(2^{-\mathsf{M}/I}) \geq \Omega(2^{-\delta\mathsf{M}/(12I)}),
\end{aligned}$$

by Equation (3.20). This completes the proof of Equation (4.78).

4.7 Compression independent of communication

In this section, we prove Theorem 3.9.

Theorem 3.9 Restated. *For every $\alpha > 0$, there is a $\Delta > 0$ such that if $\mathsf{M}_I(p, f) \leq \alpha I$ and $\mu(xy) = p(xy)$, then $\mathsf{adv}_\mu(\Delta I, f) \geq \exp(-\Delta I)$.*

Let K be a sufficiently large constant to be determined later. Let $p(xym)$ be a protocol distribution such that $p(xy) = \mu(xy)$ and $\mathsf{M}_I(p, f) \leq \alpha I$. Let $q(xym)$ be a rectangular distribution that realizes $\mathsf{M}_I(p, f)$.

Define $\mathsf{M} = \mathsf{M}_I(p, f) + KI$. Since $\mathsf{M}(p, f) \geq 0$, we have $\mathsf{M} \geq KI$. Let g_1, g_2 be as in Equation (3.10). Let ε be a parameter such that $\varepsilon = 2^{-6\mathsf{M}/I - 8\mathsf{M}}$. We define a protocol Γ whose communication complexity is bounded by

$$2 \log 1/\varepsilon \leq O(6\mathsf{M}/I + 8\mathsf{M}) = \Delta I,$$

for some Δ that depends only on α.

We describe the protocol Γ.

1. Jointly sample $\eta^A, \eta^B \in [0, 1]$ uniformly. Jointly sample two uniformly random functions $h, t : \mathbb{Z} \to \{1, \ldots, \lceil 1/\varepsilon \rceil\}$.

2. Jointly sample an infinite sequence of triples $(m^1, \rho_A^1, \rho_B^1), (m^2, \rho_A^2, \rho_B^2), \ldots$, where m^i is sampled uniformly at random from the set of all transcripts and ρ_A^i, ρ_B^i are sampled uniformly at random in $[0, 1]$.

3. Alice finds the first index i_A such that

$$\rho_A^{i_A} \leq \prod_{j \text{ odd}} p(m_j^{i_A} | x m_{<j}^{i_A}),$$
$$\rho_B^{i_A} \leq 2^{6\mathsf{M}} \cdot \prod_{j \text{ even}} p(m_j^{i_A} | x m_{<j}^{i_A}).$$

   Alice checks if $\eta^A \leq g_1(x m^{i_A}) \cdot 2^{\lceil \log g_1(x m^{i_A}) \rceil}$, in which case she sends $t(i_A)$ and $h(\lceil \log g_1(x m^{i_A}) \rceil)$ to Bob. Otherwise, she sends $\perp$ signaling to abort.

4. Bob finds the first index i_B such that

$$\rho_A^{i_B} \leq 2^{6\mathsf{M}} \cdot \prod_{j \text{ odd}} p(m_j^{i_B} | y m_{<j}^{i_B}),$$
$$\rho_B^{i_B} \leq \prod_{j \text{ even}} p(m_j^{i_B} | y m_{<j}^{i_B}).$$

If $t(i_B) = t(i_A)$, he checks if there is a unique integer z such that

$$|z + \log g_2(ym^{i_B})| \leq 3\mathsf{M}/I,$$
$$h(z) = h(\lceil \log g_1(xm^{i_A}) \rceil),$$
$$\eta^B \leq g_2(ym^{i_B}) \cdot 2^{z - 3\mathsf{M}/I},$$

If all these conditions are satisfied, he sends $\mathsf{sign}\Big(\mathbf{E}_{q(x'y'|m^{i_B})}[(-1)^{f(x'y')}] \Big) \in \{\pm 1\}$ to Alice. Otherwise, he sends $\perp$ to abort the protocol.

The protocol has the feature that Alice sends at most $2 \log 1/\varepsilon$ bits to Bob. Let i_* be the smallest index such that

$$\prod_{j \text{ odd}} p(m_j^{i_*} | x m_{<j}^{i_*}) \geq \rho_A^{i_*} \text{ and } \prod_{j \text{ even}} p(m_j^{i_*} | x m_{<j}^{i_*}) \geq \rho_B^{i_*}.$$

Let $m = m^{i_*}$. We note that $\Gamma(xym) = p(xym)$.

Let S and T be the sets defined in Equation (3.11) and Equation (3.18) respectively for our choice of K. In addition to S, we need the following sets to analyze the simulating protocol:

$$Q = \Big\{ xym\eta^A\eta^B : \eta^A \leq g_1(xm) \cdot 2^{-\lceil \log g_1(xm) \rceil}, \eta^B \leq g_2(ym) \cdot 2^{\lceil \log g_1(xm) \rceil - 3\mathsf{M}/I} \Big\},$$
$$\mathcal{E} = \Big\{ i_A i_B i_* : i_A = i_B = i_* \Big\},$$
$$\mathcal{Z} = \Big\{ xymh : \exists \text{ unique } z \in \mathbb{Z} \text{ s.t. } |z + \log g_2(ym)| \leq \frac{3\mathsf{M}}{I} \text{ and } h(z) = h(\lceil \log g_1(xm) \rceil) \Big\}.$$

Let $\mathcal{G}$ denote the event that the protocol reaches the final step without aborting, and define $\mathcal{A}(xym) \in \{\pm 1\}$ by

$$\mathcal{A}(xym) = \mathsf{sign}\Big(\mathop{\mathbf{E}}_{q(x'y'|m)}[(-1)^{f(x'y')}] \Big) \cdot (-1)^{f(xy)}.$$

Our protocol computes $f(xy)$ correctly when: $\mathcal{G}$ happens, $\mathcal{A}(xym) = 1$ and $m = m^{i_B}$. Since $\mathcal{E}\mathcal{Z}SQ \subseteq \mathcal{G}$, and $\mathcal{E}$ implies $m = m^{i_B}$, the advantage of our protocol is at least:

$$\Gamma(\mathcal{E}\mathcal{Z}SQ) \cdot \mathop{\mathbf{E}}_{\Gamma(xym|\mathcal{E}\mathcal{Z}SQ)}[\mathcal{A}(xym)] - \Gamma(\mathcal{G}(\mathcal{E}\mathcal{Z}SQ)^c). \tag{4.87}$$

We shall prove:

$$\mathop{\mathbf{E}}_{\Gamma(xym|\mathcal{EZ}SQ)}[\mathcal{A}(xym)] \geq \Omega(2^{-\delta \mathsf{M}/(12I)}), \tag{4.88}$$

$$\Gamma(\mathcal{EZ}SQ) \geq \Omega(2^{-6\mathsf{M}/I-6\mathsf{M}}), \tag{4.89}$$

$$\Gamma(\mathcal{G}(\mathcal{EZ}SQ)^c) \leq O(2^{-6\mathsf{M}/I-7\mathsf{M}}). \tag{4.90}$$

By Equation (4.87), since $\delta \leq 1$, we can choose K to be large enough to prove the theorem, since $\alpha + K \geq \mathsf{M}/I \geq K$.

We first upper bound $\Gamma(\mathcal{G}(\mathcal{EZ}SQ)^c)$. By the union bound, we have:

$$\Gamma(\mathcal{G}(\mathcal{EZ}SQ)^c) \leq \Gamma(\mathcal{GE}^c) + \Gamma(\mathcal{Z}^c|\mathcal{GE}) + \Gamma(S^c\mathcal{GEZ}) + \Gamma(Q^c|\mathcal{GEZ}S).$$

The definition of the protocol ensures that $\Gamma(\mathcal{Z}^c|\mathcal{GE}) = 0$. Moreover, we claim that $\Gamma(Q^c|\mathcal{GEZ}S) = 0$, because if the event $\mathcal{EZ}S$ happens and the parties do not abort, then:

$$\eta^A \leq g_1(xm^{i_A}) \cdot 2^{\lceil \log g_1(xm^{i_A})\rceil} = g_1(xm) \cdot 2^{\lceil \log g_1(xm)\rceil},$$
$$\eta^B \leq g_2(ym^{i_B}) \cdot 2^{z-3\mathsf{M}/I} = g_2(ym) \cdot 2^{\lceil \log g_1(xm)\rceil - 3\mathsf{M}/I}.$$

The event $\mathcal{GE}^c$ implies that there was a hash error for the triples accepted by Alice and Bob. The probability of this happening is at most ε. Finally, $\Gamma(S^c\mathcal{GEZ}) \leq \Gamma(S^c\mathcal{EZ}) \leq O(\varepsilon \mathsf{M}/I)$, since if $S^c\mathcal{EZ}$ happens then there must have been a hash collision, which happens with also occurs with probability at most 2ε. By our choice of ε, the total error is bounded by $2^{-6\mathsf{M}/I-8\mathsf{M}}(2 + \mathsf{M}/I) \leq 2^{-6\mathsf{M}/I-7\mathsf{M}}$, for K sufficiently large. This implies Equation (4.90).

Let us estimate $\Gamma(QS)$. We have,

$$\Gamma(QS) = \sum_{xym \in S} \Gamma(xym) \cdot \Gamma(Q|xym) = \sum_{xym \in S} p(xym) \cdot \Gamma(Q|xym).$$

For $xym \in S$,

$$\Gamma(Q|xym) = \frac{g_1(xm) \cdot 2^{-\lceil \log g_1(xm)\rceil} \cdot g_2(ym) \cdot 2^{\lceil \log g_1(xm)\rceil}}{2^{3\mathsf{M}/I}} = \frac{q(xym)}{p(xym)} \cdot \frac{1}{2^{3\mathsf{M}/I}}, \tag{4.91}$$

where the first equality follows from the fact that

$$g_2(ym) \cdot 2^{\lceil \log g_1(xm)\rceil} = 2^{\lceil \log g_1(xm)\rceil + \log g_2(ym)} \leq 2^{3\mathsf{M}/I},$$

by the definition of S. Therefore,

$$\Gamma(QS) = \sum_{xym \in S} p(xym) \cdot \frac{q(xym)}{p(xym)} \cdot \frac{1}{2^{3\mathsf{M}/I}} = \frac{q(S)}{2^{3\mathsf{M}/I}}$$
$$\geq \frac{(1 - 5 \cdot 2^{-\mathsf{M}/I})}{2^{3\mathsf{M}/I}} = \Omega(2^{-3\mathsf{M}/I}), \tag{4.92}$$

where in the last line, we used Claim 3.26.

We claim that for all $xym \in S$,

$$\Gamma(\mathcal{Z}|xymQS) = \Gamma(\mathcal{Z}|xym) \geq 1 - O(\varepsilon \mathsf{M}/I). \tag{4.93}$$

The equality follows by observing that xym determine S and given xym, $\mathcal{Z}$ just depends on the choice of h, which is independent of Q. The event $\mathcal{Z}^c$ can happen only if there exists an integer z distinct from $\lceil \log g_1(xm) \rceil$ such that $h(\lceil g_1(xm) \rceil) = h(z)$ and $|z + \log g_2(ym)| \leq 3\mathsf{M}/I$. The probability that this happens is at most $O(\varepsilon \cdot \mathsf{M}/I)$. Therefore, $\Gamma(\mathcal{Z}|xym) \geq 1 - O(\varepsilon \mathsf{M}/I) \geq 1/2$, by our choice of ε. We conclude that

$$\Gamma(QS\mathcal{Z}) = \Gamma(QS) \cdot \Gamma(\mathcal{Z}|QS) \geq \Omega(2^{-3\mathsf{M}/I}), \tag{4.94}$$

Let W be the event that $\min\{i_A, i_B, i_*\} = 1$ and let T be the set defined in Equation (3.18) for our choice of K. We claim that TW^c implies $i_* > 1$, since if $xym \in T$ then $p(m|xy) \leq 2^{6\mathsf{M}} \cdot \min\{p(m|x), p(m|y)\}$, which implies

$$\prod_{j \text{ even}} p(m_j|ym_{<j}) \leq 2^{6\mathsf{M}} \cdot \prod_{j \text{ even}} p(m_j|xm_{<j}),$$
$$\prod_{j \text{ odd}} p(m_j|xm_{<j}) \leq 2^{6\mathsf{M}} \cdot \prod_{j \text{ even}} p(m_j|ym_{<j}),$$

and hence if $i_* = 1$ then in fact $i_A = i_B = 1$.

Now, we compute $\Gamma(\mathcal{E}|QS\mathcal{Z})$.

$$\begin{aligned}\Gamma(\mathcal{E}|QS\mathcal{Z}) &= \Gamma(\mathcal{E}|QS\mathcal{Z}W) \\ &\geq \frac{\Gamma(i_A = i_B = i_* = 1|QS\mathcal{Z})}{\Gamma(i_A = 1|QS\mathcal{Z}) + \Gamma(i_B = 1|QS\mathcal{Z}) + \Gamma(i_* = 1|QS\mathcal{Z})} \\ &\geq \frac{\Gamma(i_A = i_B = i_* = 1, QS\mathcal{Z})}{\Gamma(i_A = 1) + \Gamma(i_B = 1) + \Gamma(i_* = 1)}.\end{aligned} \tag{4.95}$$

Now, we estimate the numerator and denominator in the last expression. Let $\mathcal{M}$ be the set of all transcripts in the support of p. We have,

$$\begin{aligned}
&\Gamma(i_A = i_B = i_* = 1, QS\mathcal{Z}) \\
&\geq \sum_{xym \in S \cap T} \Gamma(i_A = i_B = i_* = 1, xym, Q\mathcal{Z}) \\
&= \sum_{xym \in S \cap T} \Gamma(i_A = i_B = i_* = 1, xym) \cdot \Gamma(Q\mathcal{Z}|xym) \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{(given } xym\text{, } Q\mathcal{Z} \text{ is independent of } i_A, i_B, i_*) \\
&= \sum_{xym \in S \cap T} p(xym) \cdot \frac{1}{|\mathcal{M}|} \cdot \Gamma(Q\mathcal{Z}|xym) \qquad \text{(by the definition } \Gamma \text{ and } T) \\
&= \sum_{xym \in S \cap T} p(xym) \cdot \frac{1}{|\mathcal{M}|} \cdot \frac{q(xym)}{p(xym)2^{3\mathsf{M}/I}} \cdot \Gamma(\mathcal{Z}|xym) \qquad \text{(by Equation (4.91))} \\
&\geq q(ST) \cdot \frac{1}{|\mathcal{M}| \cdot 2^{3\mathsf{M}/I}} \cdot (1 - \Omega(\varepsilon \mathsf{M}/I)). \qquad \text{(by Equation (4.93))}
\end{aligned}$$

Next,

$$\begin{aligned}
\Gamma(i_A = 1) = \sum_{xm'} \Gamma(i_A = 1, xm') &\leq \sum_{xm'} p(x) \cdot \frac{1}{|\mathcal{M}|} \cdot \prod_{j \text{ odd}} p(m'_j|xm'_{<j}) \cdot 2^{6\mathsf{M}} \cdot \prod_{j \text{ even}} p(m'_j|xm'_{<j}) \\
&\leq \sum_{xm'} p(xm') \cdot \frac{2^{6\mathsf{M}}}{|\mathcal{M}|} \leq \frac{2^{6\mathsf{M}}}{|\mathcal{M}|}.
\end{aligned}$$

An identical calculation shows that $\Gamma(i_B = 1) \leq 2^{6\mathsf{M}}/|\mathcal{M}|$. Furthermore,

$$\begin{aligned}
\Gamma(i_* = 1) = \sum_{xym} \Gamma(i_* = 1, xym) &= \sum_{xym} p(x) \cdot \frac{1}{|\mathcal{M}|} \cdot \prod_{j \text{ odd}} p(m_j|xm_{<j}) \cdot \prod_{j \text{ even}} p(m_j|ym_{<j}) \\
&\leq \sum_{xym} p(xym) \cdot \frac{1}{|\mathcal{M}|} \leq \frac{1}{|\mathcal{M}|}.
\end{aligned}$$

Plugging this into Equation (4.95) we get

$$\Gamma(\mathcal{E}|QS\mathcal{Z}) \geq \frac{q(ST) \cdot (1 - \Omega(\varepsilon \mathsf{M}/I))}{2^{6\mathsf{M}+(3\mathsf{M}/I)+2}} = \Omega(2^{-6\mathsf{M}-(3\mathsf{M}/I)}),$$

by Claim 3.26 and Claim 3.29. Using Equation (4.94) we get that $\Gamma(QS\mathcal{Z}\mathcal{E}) = \Omega(2^{-6\mathsf{M}-(6\mathsf{M}/I)})$ as claimed in Equation (4.89).

For all $xym \in S$,

$$\begin{aligned}
\Gamma(xym|QS\mathcal{Z}) &= \frac{\Gamma(xym)\cdot\Gamma(QS\mathcal{Z}|xym)}{\Gamma(QS\mathcal{Z})} \\
&= \frac{p(xym)}{\Gamma(QS)}\cdot\Gamma(Q|xym)\cdot\frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} \\
&= \frac{p(xym)}{\Gamma(QS)}\cdot\frac{q(xym)}{p(xym)\cdot 2^{3\mathsf{M}/I}}\cdot\frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} && \text{(By Equation (4.91))} \\
&= \frac{q(xym)}{q(S)}\cdot\frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} && \text{(By Equation (4.92))} \\
&= q(xym|S)\cdot(1\pm O(\varepsilon\mathsf{M}/I)), && (4.96)
\end{aligned}$$

where the last line follows by Equation (4.93).

Next, we note that

$$\Gamma(\mathcal{E}|QS\mathcal{Z}, i_*=1) \geq \Gamma(\mathcal{E},T|QS\mathcal{Z}, i_*=1) = \Gamma(T|QS\mathcal{Z}), \tag{4.97}$$

where we used the fact that the event $T, i_* = 1$ implies $\mathcal{E}$ and that xym is distributed independently of i_*. For any $xym \in S \cap T$

$$\begin{aligned}
\Gamma(xym|QS\mathcal{Z}\mathcal{E}) &= \Gamma(xym|QS\mathcal{Z}\mathcal{E}W) \\
&\qquad (xym \text{ is independent of } W \text{ even conditioned on } QS\mathcal{Z}\mathcal{E}) \\
&= \Gamma(xym|QS\mathcal{Z}\mathcal{E}, i_*=1) \\
&\qquad (\text{the event } \mathcal{E}W \text{ is the same as the event } \mathcal{E}, i_*=1) \\
&= \frac{\Gamma(xym\mathcal{E}|QS\mathcal{Z}, i_*=1)}{\Gamma(\mathcal{E}|QS\mathcal{Z}i^*=1)} \\
&= \Gamma(xym|QS\mathcal{Z})\cdot\frac{\Gamma(\mathcal{E}|xym, i_*=1)}{\Gamma(\mathcal{E}|QS\mathcal{Z}i^*=1)} \\
&= \frac{\Gamma(xym|QS\mathcal{Z})}{\Gamma(\mathcal{E}|QS\mathcal{Z}i^*=1)} && (\text{because } xym \in S\cap T) \\
&= \Gamma(xym|QS\mathcal{Z})\cdot(1\pm O(\Gamma(T^c|QS\mathcal{Z})))
\end{aligned}$$

where the last inequality used the fact that $1 \geq \Gamma(\mathcal{E}|QS\mathcal{Z}i^*=1) \geq 1-\Gamma(T^c|QS\mathcal{Z})$ by

Equation (4.97). Together with Equation (4.96) we get that for any $xym \in S \cap T$

$$\begin{aligned}\Gamma(xym|QS\mathcal{Z}\mathcal{E}) &= q(xym|S) \cdot (1 \pm O(\Gamma(T^c|QS\mathcal{Z}) + \varepsilon \mathsf{M}/I)) \\ &= q(xym|S) \cdot (1 \pm O(q(T^c|S) + \varepsilon \mathsf{M}/I)) \\ &= q(xym|S) \cdot (1 \pm O(2^{-\mathsf{M}/I} + \varepsilon \mathsf{M}/I)), \end{aligned} \tag{4.98}$$

where the last line follows by Claim 3.29.

Now, we complete the proof of Equation (4.88). We have

$$\begin{aligned}&\mathop{\mathbf{E}}_{\Gamma(xym|QS\mathcal{Z}\mathcal{E})}[\mathcal{A}(xym)] \\ &\geq \sum_{xym \in S \cap T} \Gamma(xym|QS\mathcal{Z}\mathcal{E}) \cdot \mathcal{A}(xym) - \Gamma(T^c|QS\mathcal{Z}\mathcal{E}) \\ &\geq \sum_{xym \in S \cap T} q(xym|S) \cdot \mathcal{A}(xym) - \Omega(2^{-\mathsf{M}/I} + \varepsilon \mathsf{M}/I) - 1 + \Gamma(T|QS\mathcal{Z}\mathcal{E}) \\ &\qquad\qquad \text{(by Equation (4.98))} \\ &\geq \mathop{\mathbf{E}}_{q(xym|S)}[\mathcal{A}(xym)] - q(T^c|S) - \Omega(2^{-\mathsf{M}/I} + \varepsilon \mathsf{M}/I) - 1 + q(T|S) \cdot (1 - O(2^{-\mathsf{M}/I} + \varepsilon \mathsf{M}/I)) \\ &\qquad\qquad \text{(by Equation (4.98))} \\ &\geq \mathop{\mathbf{E}}_{q(xym)}[\mathcal{A}(xym)] - q(S^c) - 2q(T^c|S) - \Omega(2^{-\mathsf{M}/I} + \varepsilon \mathsf{M}/I) \\ &= \mathop{\mathbf{E}}_{q(xym)}[\mathcal{A}(xym)] - \Omega(2^{-\mathsf{M}/I} + \varepsilon \mathsf{M}/I) \qquad \text{(by Claim 3.26 and Claim 3.29)} \\ &= \mathop{\mathbf{E}}_{q(xym)}\left[\mathsf{sign}\left(\mathop{\mathbf{E}}_{q(x'y'|m)}\left[(-1)^{f(x'y')}\right]\right) \cdot (-1)^{f(xy)}\right] - \Omega(2^{-\mathsf{M}/I}) \\ &= \mathop{\mathbf{E}}_{q(m)}\left[\left|\mathop{\mathbf{E}}_{q(xy|m)}\left[(-1)^{f(xy)}\right]\right|\right] - \Omega(2^{-\mathsf{M}/I}) \geq \Omega(2^{-\delta \mathsf{M}/(12I)}),\end{aligned}$$

by Equation (3.20).

Chapter 5

XOR LEMMA FOR DETERMINISTIC COMMUNICATION & LIFTING

Given two functions f and g, how much harder is it to compute their composition than it is to compute each of the functions? In this chapter, we give some answers to this question in the model of deterministic communication complexity. We recall that for a function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$, $D(f)$ denotes the deterministic communication complexity of f. Moreover, we denote by $C(f)$, the *cover number* of f, which is the minimum of number of rectangles needed to cover $\mathcal{X} \times \mathcal{Y}$ such that each rectangle is constant for f.

We recall the notation of function composition from Chapter 2: given functions $f(x, y)$ and a function $g : \{0,1\}^n \to \{0,1\}$ we denote the composed function by $g \circ f(xy) := g(f(x_1, y_1), \ldots, f(x_n, y_n))$. In the special case when the outer function is parity, the n-fold XOR of f is denoted by $f^{\oplus n}(xy) := f(x_1y_1) \oplus \ldots \oplus f(x_ny_n)$. In this chapter, we study the communication complexity of $g \circ f$ and $f^{\oplus n}$ in terms of the communication complexity of f and various complexity measures of f. Before proceeding, we remark that in the lifting literature, it is common to flip the notation for inner and outer functions in the composition – the outer function is f, and the inner function, known as a gadget, is g. We prefer to use $g \circ f$ as the other results of this paper, which are XOR lemmas, are typically written with the inner function being f.

5.1 Background

We start with a simple and natural upper bound for computing $g \circ f$, which was observed by Buhrman, Cleve and Wigderson [13]. Given a decision tree T for g, simulate T by running the communication protocol to compute $z_i = f(x_iy_i)$ whenever it queries z_i. Using the best decision tree for g, we get

$$D(g \circ f) \leq \mathsf{DT}(g) \cdot D(f). \tag{5.1}$$

A natural question is whether or not the above bound is optimal. From the perspective of lower bounds, a related and perhaps simpler question is whether or not computing n copies of f requires n times the communication of a single copy. This latter problem was studied by Feder, Kushilevitz, Naor and Nisan [18] who gave a direct sum theorem for deterministic communication complexity.

Theorem 5.1 ([18]). $D(f^n) \geq \log C(f^n) \geq n \cdot (\sqrt{D(f)} - \log\log(|\mathcal{X}| \cdot |\mathcal{Y}|) - 1)$.

The preceding theorem gives a lower bound on the communication needed to compute n copies of f, however, as observed in Equation (5.1), computing $g \circ f$ may not require computing f on all the coordinates.

There is a large body of work showing that Equation (5.1) is indeed optimal for several (fixed) functions f. The earliest such result is due to Raz and McKenzie [54] who considered the index function gadget, $\mathsf{Ind}_m : [m] \times \{0,1\}^m \to \{0,1\}$ given by $\mathsf{Ind}_m(x,y) = y_x$. They showed that for $m = n^{O(1)}$, and any function $g : \{0,1\}^n \to \{0,1\}$, $D(g \circ \mathsf{Ind}_m) = \Theta(\mathsf{DT}(g) \cdot D(\mathsf{Ind}_m))$.

Their proof was simplified by Göös, Pitassi, and Watson [22], and was recently improved by [40] who showed the same result as that of [54] for $m = O(n^{1+\varepsilon})$ (for any fixed $\varepsilon > 0$). Chattopadhyay, Koucký, Loff, and Mukhopadhyay [16] showed a similar result for the inner product gadget (as well as for any gadget with a certain pseudorandom property) $\mathsf{IP}_m : \{0,1\}^m \times \{0,1\}^m \to \{0,1\}$ given by $\mathsf{IP}_m(x,y) = x_1y_1 \oplus \ldots \oplus x_my_m$. They showed that for $m = \Omega(\log n)$ and any function g, $D(g \circ \mathsf{IP}_m) = \Theta(\mathsf{DT}(g) \cdot D(\mathsf{IP}_m))$. Manor and Meir [41] proved that $D(g \circ f) = \Omega(\mathsf{DT}(g) \cdot D(f))$ for all functions f with discrepancy at most $n^{-O(1)}$.

Several works have also considered lifting complexity measures other than decision tree complexity to communication. Zhang [69] showed that for any $f_1, \ldots, f_n \in \{\vee, \wedge\}$ and any function g, computing $g(f_1(x_1y_1), \ldots, f_n(x_n, y_n))$ with constant success probability requires communication $\Omega(\mathsf{DT}(g)^{1/3})$. Huynh and Nordström [26] lifted critical block-sensitivity to the randomized communication complexity of certain search problems. This was simplified by Göos and Pitassi [21] and applied to obtain depth lower bounds for monotone circuits. Sherstov [61] lifted approximate degree to the randomized communication complexity of functions of the form $g \circ \mathsf{Ind}_m$.

5.2 Discussion of Main Results

In this chapter we prove two related results. First, we prove an XOR lemma for communication complexity.

Theorem 5.2 (XOR lemma for Deterministic Communication)**.** *There exists $c_0 \geq 1$ such that for any function f with $D(f) \geq c_0$, $D(f^{\oplus n}) \geq \log C(f^{\oplus n}) \geq n \cdot \left(\frac{\Omega(D(f))}{\log \mathsf{rk}(f)} - \log \mathsf{rk}(f)\right)$.*

Then, we generalize the above to obtain a lifting theorem from sensitivity to deterministic communication complexity for arbitrary gadgets.

Theorem 5.3 (Lifting Theorem)**.** *There exists $c_0 \geq 1$ such that for any function f with $D(f) \geq c_0$ and any function $g : \{0,1\}^n \to \{0,1\}$,*

$$D(g \circ f) \geq \log C(g \circ f) \geq \mathsf{s}(g) \cdot \left(\frac{\Omega(D(f))}{\log \mathsf{rk}(f)} - \log \mathsf{rk}(f)\right).$$

We provide a few remarks on the above results. First, we note that the sensitivity of the XOR function is n since, on any input, changing any coordinate results in the output changing. It follows that Theorem 5.3 implies Theorem 5.2. Next, we address the relationship between rank and communication and its implications for the above results. We recall that $D(f) \geq \log \mathsf{rk}(f)$ and moreover, the log-rank conjecture due to Lovász and Saks [38], asserts a partial converse.

Conjecture 5.4 (Log-rank Conjecture [38])**.** *There exists a constant k such that for any function $f(x, y)$, $D(f) \leq \log^k \mathsf{rk}(f)$.*

The best upper bound in this direction is due to Sudakov and Tomon [64] which improves a bound of Lovett [39] to get $D(f) = O(\sqrt{\mathsf{rk}(f)})$. Additionally, Göös, Pitassi, and Watson [22], gave an example of a function with rank r and communication complexity $\widetilde{\Omega}(\log^2 r)$.

Using the fact that the rank "tensorizes" i.e. $\mathsf{rk}(g \circ f) \geq (\mathsf{rk}(f) - 1)^{\mathsf{deg}(g)}$ [56, 28], the above conjecture implies that $D(g \circ f) = \Omega(\mathsf{deg}(g) \cdot (D(f))^{1/k})$, where k is the constant in Conjecture 5.4. Yang [65] used this property of rank and observed that when g is the XOR function, one can use Theorem 5.2 to conclude that $D(f^{\oplus n}) = \Omega(n \cdot \sqrt{D(f)})$, for any f whose communication complexity is a sufficiently large constant. This is a strengthening

of Theorem 5.1, and the same ideas yield the following corollary, which is an unconditional version of the consequence of the log-rank conjecture.

Corollary 5.5. *There exists $c_0 \geq 1$ such that for any function f with $D(f) \geq c_0$ and any function $g : \{0,1\}^n \to \{0,1\}$,*

$$D(g \circ f) \geq \frac{\mathsf{s}(g) \cdot \mathsf{deg}(g)}{2\mathsf{s}(g) + \mathsf{deg}(g)} \cdot \Omega\left(\frac{D(f)}{\log \mathsf{rk}(f)} + \log \mathsf{rk}(f)\right) = \Omega\big(\min\{\mathsf{s}(g), \mathsf{deg}(g)\} \cdot \sqrt{D(f)}\big).$$

We note that $D(f)$ needs to be a sufficiently large constant for such a lower bound since there are functions f and g for which the communication complexity of computing $g \circ f$ does not grow. For example, the deterministic communication complexity of computing the parity function on 2 bits is 2, and moreover, the n-fold XOR of the parity function can also be computed with 2 bits. The same holds if f and g are both the AND/OR functions.

Combining Corollary 5.5 with Equations (2.3) and (2.4) we get that for any f whose communication complexity is a sufficiently large constant, and any function g,

$$D(g \circ f) \geq \Omega(\mathsf{DT}(g)^{1/6} \cdot \sqrt{D(f)}),$$

which is partial progress towards establishing the tightness of Equation (5.1).

The main ideas proof of Theorem 5.2 and Theorem 5.3 are similar and use information theory. In the XOR case, we first show that if $\oplus \circ f$ admits a small cover via monochromatic rectangles, f itself must contain a large monochromatic rectangle.

Lemma 5.6. *If $\oplus \circ f$ can be covered with 2^T monochromatic rectangles, then f contains a monochromatic rectangle of density $2^{-T/n-2}$.*

For the general case, we strengthen the previous lemma to obtain the following similar one.

Lemma 5.7. *If $g \circ f$ can be covered with 2^T monochromatic rectangles, then f contains a monochromatic rectangle of density $2^{-2T/\mathsf{s}(g)} \cdot (4 \cdot \mathsf{rk}(f))^{-2}$.*

Nisan and Wigderson [46] had observed that a large monochromatic rectangle can be used to partition the inputs to f into 2 parts – in one part, the rank decreases and in the other, the size decreases. Using this they showed that if one can repeatedly obtain dense

monochromatic rectangles, then there is an efficient communication protocol for f. We can repeatedly apply Lemmas 5.6 and 5.7 to obtain such a protocol, and as a consequence we get that if $g \circ f$ admits a small cover then f admits a protocol much shorter than $D(f)$, a contradiction.

Organization. This chapter is organized as follows. First, we prove Lemmas 5.6 and 5.7 in Section 5.3. Then, in Section 5.4 we explain how the lemmas together with the Nisan and Wigderson protocol yield Theorems 5.2 and 5.3. In Section 5.5, we prove Corollary 5.5 from Theorem 5.3. Section 5.6 closes this chapter with some additional observations and future questions.

5.3 Dense Monochromatic Rectangles from Small Covers

We first give a short and simple proof of Lemma 5.6 before discussing a slightly more complicated argument for Lemma 5.7.

5.3.1 Proof of Lemma 5.6

Since $\oplus \circ f$ can be covered with 2^T monochromatic rectangles, there exists one of density at least 2^{-T}, say R. Let X and Y be a uniformly random row and column respectively in R. Since R is a rectangle, X and Y are independent. Using the chain rule, we get

$$\begin{aligned}
H(XY) &= H(X) + H(Y) && \text{(because } X, Y \text{ are independent)} \\
&= \sum_{i=1}^{n} H(X_i|X_{<i}) + H(Y_i|Y_{>i}) && \text{(by the chain rule)} \\
&= \sum_{i=1}^{n} H(X_i|X_{<i}Y_{>i}) + H(Y_i|X_{<i}Y_{>i}X_i) && \text{(because } X, Y \text{ are independent)} \\
&= \sum_{i=1}^{n} H(X_iY_i|X_{<i}Y_{>i}). && \text{(by the chain rule)}
\end{aligned}$$

This implies there exist $i, x_{<i}, y_{>i}$ such that

$$H(X_iY_i|x_{<i}y_{>i}) = \frac{H(XY)}{n} \geq \log(|\mathcal{X}| \cdot |\mathcal{Y}|) - \frac{T}{n}.$$

Define the random variables $U = f(x_1Y_1) \oplus \ldots \oplus f(x_{i-1}Y_{i-1})$ and $V = f(X_{i+1}y_{i+1}) \oplus \ldots \oplus f(X_n y_n)$. By the chain rule, and since U, V are bits, we get

$$\begin{aligned} H(X_iY_i|x_{<i}y_{>i}UV) + 2 &\geq H(X_iY_i|x_{<i}y_{>i}UV) + H(UV|x_{<i}y_{>i}) \\ &= H(X_iY_iUV|x_{<i}y_{>i}) \\ &\geq H(X_iY_i|x_{<i}y_{>i}), \end{aligned}$$

so there is some fixed value of u, v such that

$$H(X_iY_i|x_{<i}y_{>i}uv) \geq \log(|\mathcal{X}| \cdot |\mathcal{Y}|) - \frac{T}{n} - 2.$$

The desired rectangle is the set given by the support of $\mathsf{supp}(X_iY_i|x_{<i}y_{>i}uv)$. Because (X, Y) is distributed uniformly in R, the distribution of (X_i, Y_i) conditioned on $(x_{<i}, y_{>i}, u, v)$ is a product distribution, and hence this set is a rectangle. Moreover, since R is monochromatic for $\oplus \circ f$ and $\oplus_{j \neq i} f(x_j y_j) = u \oplus v$ holds for all $(x, y) \in \mathsf{supp}(XY|uv)$, it follows that $\mathsf{supp}(X_iY_i|x_{<i}y_{>i}uv)$ is monochromatic. Using Fact 2.2, the density of this rectangle is given by

$$\frac{|\mathsf{supp}(X_iY_i|x_{<i}y_{>i}uv)|}{|\mathcal{X}| \cdot |\mathcal{Y}|} \geq 2^{-T/n-2}.$$

5.3.2 Proof of Lemma 5.7

Before giving the proof, we briefly explain why the previous argument fails for general function compositions, and how we resolve the issues that arise.

A key step in the preceding argument was the fixing of the parities of the outputs of f among first $i-1$ and the last $n-i$ coordinates. This step strongly used the fact that the outer function is the parity function, which is no longer true for us. To prove lower bounds for general function composition, we need an appropriate generalization of this parity constraint. We enforce such a constraint by switching from the uniform distribution to a correlated distribution.

For simplicity, assume $\mathsf{s}(g) = n$ and that g is balanced. By definition, there exists $z \in \{0,1\}^n$ with $g(z) \neq g(z_{<i}, 1 - z_i, z_{>i})$, for all i. Consider the distribution $p(xy)$ on $\mathcal{X}^n \times \mathcal{Y}^n$ obtained by sampling each $(x_i, y_i) \in g^{-1}(z_i)$ uniformly at random; let XY be

random variables jointly distributed according to $p(xy)$. Now, using the sub-additivity of entropy, we obtain a slightly weaker statement than in the previous argument. We show that there exists a rectangle R that is constant for $g \circ f$, a coordinate i and inputs $x_{<i}, y_{>i}$ such that the sets

$$A = \mathsf{supp}(X_i | x_{<i} y_{>i} R) \quad \text{and} \quad B = \mathsf{supp}(Y_i | x_{<i} y_{>i} R),$$

satisfy $|A| \geq \Omega(|\mathcal{X}| \cdot 2^{-T/n})$ and $|B| \geq \Omega(|\mathcal{Y}| \cdot 2^{-T/n})$.

We finish the proof by showing that $A \times B$ is a monochromatic rectangle for f. Note that

$$\mathsf{supp}(p(x_i, y_i | x_1, \ldots, x_{i-1}, y_{i+1}, \ldots, y_n, R)) \subseteq A \times B,$$

and although f is constant on the former set (by the definition of p), it is not obvious that the same holds for the $A \times B$.

The main difference between the above high-level description and the proof of Lemma 5.7 is that g need not be balanced. To address this, we consider two cases. First, we assume that g is extremely biased, say $\mathbf{Pr}[g(x, y) = 1] > 1 - 1/(4 \cdot \mathsf{rk}(g))$. In this case, we obtain a monochromatic rectangle for g using an observation of Gavinsky and Lovett [20], ignoring the cover for $f \circ g$. Otherwise, g is not too biased and we can apply the above discussion, albeit with a loss of $1/(4 \cdot \mathsf{rk}(g))$ in the final bound. Let us now turn to the actual proof.

First, assume that $|\mathbf{E}_{xy}[f(xy)] - 1/2| > 1/2 - 1/(4 \cdot \mathsf{rk}(f))$. In this case, we use Lemma 3.3 from [20] to infer that f contains a monochromatic rectangle of constant density. For completeness, we supply the proof. Indeed,

$$\left| \mathop{\mathbf{E}}_{xy}[f(xy)] - \frac{1}{2} \right| = \max \left\{ \mathop{\mathbf{Pr}}_{xy}[f(xy) = 1] - \frac{1}{2}, \mathop{\mathbf{Pr}}_{xy}[f(xy) = 0] - \frac{1}{2} \right\},$$

and we can assume without loss of generality that

$$\mathop{\mathbf{Pr}}_{xy}[f(xy) = 1] > 1 - \frac{1}{4 \cdot \mathsf{rk}(f)}.$$

Let E be the set of all x such that $\mathbf{Pr}_y[f(xy) = 1] \leq 1 - 1/(2 \cdot \mathsf{rk}(f))$. We have,

$$\begin{aligned} \mathop{\mathbf{Pr}}_{xy}[f(xy) = 1] &\leq \mathop{\mathbf{Pr}}_{x}[x \in E] \cdot \left(1 - \frac{1}{2 \cdot \mathsf{rk}(f)}\right) + \mathop{\mathbf{Pr}}_{x}[x \notin E] \\ &= 1 - \frac{\mathbf{Pr}_x[x \in E]}{2 \cdot \mathsf{rk}(f)}, \end{aligned}$$

which implies that $\mathbf{Pr}_x[x \in E] \leq 1/2$. Let $x_1, \ldots, x_r \in E^c$ be such that the corresponding rows are maximally linearly independent in M_g. Moreover, let $G = \{y : f(x_i y) = 1, \forall i \in [r]\}$. By a union bound, we have

$$\mathop{\mathbf{Pr}}_y[y \notin G] \leq r \cdot \frac{1}{2\mathsf{rk}(f)} \leq \frac{1}{2}.$$

We observe that any row in $E^c \times G$ is all 1s or all 0s since it can be expressed as a linear combination of $x_1, \ldots, x_r$, each of which is all 1s. Hence, we have a monochromatic rectangle of density at least $1/8$. Since

$$2^{-2T/\mathsf{s}(g)} \cdot (4 \cdot \mathsf{rk}(f))^{-2} \leq 1/16 < 1/8,$$

we have found a monochromatic rectangle of the desired density.

Next, suppose that $|\mathbf{E}_{xy}[f(xy)] - 1/2| < 1/2 - 1/(4 \cdot \mathsf{rk}(f))$. We have

$$\frac{1}{4 \cdot \mathsf{rk}(f)} \leq \mathop{\mathbf{Pr}}_{xy}[f(xy) = 0], \mathop{\mathbf{Pr}}_{xy}[f(xy) = 1] < 1 - \frac{1}{4 \cdot \mathsf{rk}(f)}. \tag{5.2}$$

For shorthand, let s be the sensitivity of g. By definition, there exists an input $z \in \{0,1\}^n$ and a set $S \subseteq [n]$ such that for all $i \in S$

$$g(z) \neq g(z_{<i}, 1 - z_i, z_{>i}).$$

We may assume without loss of generality that $S \supseteq [s]$, for otherwise, this can be ensured by renaming the coordinates. Let $u(xy)$ denote the uniform distribution over all inputs $(x, y) \in \mathcal{X}^n \times \mathcal{Y}^n$, and let $p(xy)$ be a distribution obtained by sampling each (x_i, y_i) randomly and independently subject to $g(x_i y_i) = z_i$. Additionally, let XY be random variables jointly distributed according to $p(xy)$. By Equation (5.2) we have the following inequality relating the two distributions:

$$\begin{aligned}\max_{x,y} \frac{p(x_{\leq s} y_{\leq s})}{u(x_{\leq s} y_{\leq s})} &= \max_{xy} \prod_{i \leq s} \frac{u(x_i y_i | g(x_i y_i) = z_i)}{u(x_i y_i)} \\ &= \prod_{i \leq s} \frac{1}{\mathbf{Pr}_{u(x_i, y_i)}[g(x_i, y_i) = z_i]} \leq (4 \cdot \mathsf{rk}(g))^s. \end{aligned} \tag{5.3}$$

Since $g \circ f$ can be covered with at most 2^T monochromatic rectangles, say $R_1, \ldots, R_{2^T}$, there exists a rectangle R in the cover with $p(R) \geq 2^{-T}$. Using the sub-additivity of entropy we prove the following claim.

Claim 5.8.

$$\sum_{i\in[s]} H(X_i|X_{<i}X_{>s}Y_{>i}) + H(Y_i|X_{<i}X_{>s}Y_{>i}) \geq s\log\frac{|\mathcal{X}|\cdot|\mathcal{Y}|}{(4\cdot\mathsf{rk}(f))^2} - 2T.$$

Proof. Applying the chain rule for entropy we get

$$\begin{aligned}&\sum_{i\in[s]} H(X_i|X_{<i}X_{>s}Y_{>i}) + H(Y_i|X_{<i}X_{>s}Y_{>i})\\ &\geq \sum_{i\in[s]} H(X_i|X_{<i}X_{>s}Y) + H(Y_i|XY_{>i}) = H(X|X_{>s}Y) + H(Y|XY_{>s}).\end{aligned}$$

Let $p'(xy)$ be the distribution obtained by sampling (x_j, y_j) uniformly at random, for each $j \in [s]$, and according to $p(x_jy_j)$ for each $j \notin S$. Using this notation, we bound the term $H(X|X_{>s}Y)$ above as follows

$$\begin{aligned}&H(X|YX_{>s})\\ &= \mathop{\mathbf{E}}_{p(xy|R)}\left[\log\frac{1}{p(x|yx_{>s}R)}\right]\\ &= \mathop{\mathbf{E}}_{p(x,y|R)}\left[\log\frac{p(R)\cdot p(x_{>s}y|R)}{p(xy)}\right]\\ &\geq \mathop{\mathbf{E}}_{p(xy|R)}\left[\log\frac{p(x_{>s}y|R)}{(4\cdot\mathsf{rk}(f))^s\cdot u(x_{\leq s}y_{\leq s})\cdot p(x_{>s}y_{>s})}\right] - T \quad \text{(Equation (5.3) and } p(R)\geq 2^{-T})\\ &= \mathop{\mathbf{E}}_{p(xy|R)}\left[\log\frac{|\mathcal{X}|^s\cdot p(x_{>s}y|R)}{p'(x_{>s}y)}\right] - T - s\log(4\cdot\mathsf{rk}(f))\\ &= s\log\frac{|\mathcal{X}|}{4\cdot\mathsf{rk}(f)} + \mathbf{D}(p(x_{>s}y|R)||p'(x_{>s}y)) - T \geq s\log\frac{|\mathcal{X}|}{4\cdot\mathsf{rk}(f)} - T,\end{aligned}$$

which follows by Fact 2.4. A similar calculation shows that $H(Y|XY_{>s}) \geq s\log\frac{|\mathcal{Y}|}{4\cdot\mathsf{rk}(f)} - T$, yielding the desired bound. □

By an averaging argument, we obtain an index $i \in [s]$ and $x_{<i}, x_{>s}, y_{>i}$ such that

$$H(X_i|x_{<i}x_{>s}y_{>i}) + H(Y_i|x_{<i}x_{>s}y_{>i}) \geq \log\frac{|\mathcal{X}|\cdot|\mathcal{Y}|}{(4\cdot\mathsf{rk}(f))^2} - \frac{2T}{s}.$$

For shorthand, let

$$A := \mathsf{supp}(X_i|x_{<i}x_{>s}y_{>i}R) \text{ and } B := \mathsf{supp}(Y_i|x_{<i}x_{>s}y_{>i}R).$$

Using Fact 2.2 we conclude that the rectangle given by $A \times B$ has size at least

$$\frac{|\mathcal{X}| \cdot |\mathcal{Y}|}{16 \cdot \mathsf{rk}(f)^2 \cdot 2^{2T/s}}.$$

Moreover, we claim that $A \times B$ is monochromatic for f. Indeed, for any $x_i \in A$, there exists a row $x' \in \mathsf{supp}(X|x_{<i}x_{>s}y_{>i}R)$ such that $x'_i = x_i$. Similarly, for any $y_i \in B$, there exists a column $y' \in \mathsf{supp}(Y|x_{<i}x_{>s}y_{>i}R)$ such that $y'_i = y_i$. In particular, $(x', y') \in R$ and in addition, $x'_j = x_j$ for all $j < i$ and $y'_j = y_j$ for all $j > i$.

Since $y' \in \mathrm{supp}(Y|x_{<i}x_{>s}y_{>i}R)$, we get $f(x'_t, y'_t) = f(x_t, y'_t) = z_t$ for all $t < i$. Similarly, we have $f(x'_t, y'_t) = f(x'_t, y_t) = z_t$ for all $t > i$. Since $i \in [s]$, if $f(x'_i y'_i) \neq z_i$, then $g \circ f(x'y') = g(z_{<i}, 1 - z_i, z_{>i}) \neq g(z)$. However, this contradicts the fact that R is monochromatic for $g \circ f$ because for every $(x, y) \in \mathsf{supp}(XY|R)$, we know that $g \circ f(xy) = g(z)$.

5.4 Proofs of the XOR Lemma and Lifting Theorem

Below, we prove Theorem 5.3; the proof of Theorem 5.2 is identical. At a high level, we apply Lemma 5.7 repeatedly to find dense monochromatic rectangles and combine this with the arguments of [46] to obtain a protocol for f.

Fix some two functions f and g, and consider any cover of $g \circ f$ with some 2^T monochromatic rectangles. For shorthand, let s denote the sensitivity of g. Applying Lemma 5.7, we obtain a monochromatic rectangle R in M_f with density $2^{-2T/s} \cdot (4 \cdot \mathsf{rk}(f))^{-2}$.

By renaming the rows and columns of M_f appropriately, we can rewrite it as

$$\begin{bmatrix} R & A \\ B & Z \end{bmatrix},$$

for some matrices A, B and Z. Now, we observe that

$$\min\left\{\mathsf{rk}\left(\begin{bmatrix} R & A \end{bmatrix}\right), \mathsf{rk}\left(\begin{bmatrix} R \\ B \end{bmatrix}\right)\right\} \leq \frac{\mathsf{rk}(f) + 3}{2}. \tag{5.4}$$

Since R has rank one we get

$$\begin{aligned}
\mathsf{rk}\left(\begin{bmatrix} R & A \end{bmatrix}\right) + \mathsf{rk}\left(\begin{bmatrix} R \\ B \end{bmatrix}\right) &\leq \mathsf{rk}(A) + \mathsf{rk}(B) + 2 && \text{(by Fact 2.7)} \\
&\leq \mathsf{rk}\left(\begin{bmatrix} 0 & A \\ B & Z \end{bmatrix}\right) + 2 && \text{(Gaussian Elimination)} \\
&\leq \mathsf{rk}\left(\begin{bmatrix} R & A \\ B & Z \end{bmatrix}\right) + 3 && \text{(by Fact 2.7)} \\
&= \mathsf{rk}(g) + 3,
\end{aligned}$$

and Equation (5.4) follows.

If $\mathsf{rk}\left(\begin{bmatrix} R & A \end{bmatrix}\right) \leq (\mathsf{rk}(f)+3)/2$, Alice sends a bit to Bob indicating whether or not her input is consistent with the rows of R. Otherwise, Bob sends a bit to Alice indicating whether or not his input is consistent with the columns of R. We can assume without loss of generality that $\mathsf{rk}\left(\begin{bmatrix} R & A \end{bmatrix}\right) \leq (\mathsf{rk}(f)+3)/2$ as the proof is symmetric.

Let f' and f'' denote the functions encoded by the matrices $\begin{bmatrix} R & A \end{bmatrix}$ and $\begin{bmatrix} B & Z \end{bmatrix}$ respectively. We note that a cover of $M_{g\circ f}$ also gives a cover of both $M_{g\circ f'}$ and $M_{g\circ f''}$. If Alice's input is consistent with the rows of R, the players repeat the above argument using the rectangle cover for $M_{g\circ f'}$. Otherwise, they repeat the argument using the rectangle cover for $M_{g\circ f''}$. In the former case, we have $\mathsf{rk}(f') \leq (\mathsf{rk}(f)+3)/2$ and in the latter case, the size of $\mathcal{X}\times\mathcal{Y}$ shrinks by a factor of $1 - 2^{-2T/s}\cdot(4\cdot\mathsf{rk}(f))^{-2}$.

We claim that after $(4\cdot\mathsf{rk}(f))^3\cdot 2^{2T/s} + O(\log \mathsf{rk}(f))$ recursive steps either the rank is at most 5 or the size of the matrix is at most 1. Indeed, as long as the $\mathsf{rk}(f) \geq 5$, we have $\mathsf{rk}(f') \leq (\mathsf{rk}(f)+3)/2 \leq 4\cdot\mathsf{rk}(f)/5$. Hence, there can only be after $\log_{5/4}\mathsf{rk}(f)$ many steps where the rank reduces by a factor of $4/5$. Similarly, there can be only $k = (4\cdot\mathsf{rk}(f))^3\cdot 2^{2T/s}$ many steps where the size of the matrix reduces, since

$$\left(1 - \frac{1}{2^{2T/s}(4\cdot\mathsf{rk}(f))^2}\right)^k \leq \exp\left(-\frac{k}{2^{2T/s}(4\cdot\mathsf{rk}(f))^2}\right) = e^{-4\cdot\mathsf{rk}(f)} \leq \frac{1}{|\mathcal{X}|\cdot|\mathcal{Y}|},$$

where we used the fact that $|\mathcal{X}|$ and $|\mathcal{Y}|$ are both at most $2^{\mathsf{rk}(f)}$ in the last step.

Every leaf of this protocol either corresponds to a size 1 matrix or a matrix of rank at most 5. Thus, with constantly more bits of communication, we get a protocol for f with

the following upper bound on the number of leaves:

$$\binom{(4\cdot \mathsf{rk}(f))^3\cdot 2^{2T/s}+O(\log \mathsf{rk}(f))}{O(\log \mathsf{rk}(f))}\cdot O(1) \le O(\mathsf{rk}(f)^3\cdot 2^{2T/s})^{O(\log \mathsf{rk}(f))}$$
$$\le 2^{O((T/s+\log \mathsf{rk}(f))\cdot \log \mathsf{rk}(f))},$$

where all the inequalities hold for c_0 large enough.

By Fact 2.9 the above protocol can be rebalanced to have communication at most

$$O\left(\left(\frac{T}{s}+\log \mathsf{rk}(f)\right)\cdot \log \mathsf{rk}(f)\right).$$

Since f requires communication at least $D(f)$,we have

$$\left(\frac{T}{s}+\log \mathsf{rk}(f)\right)\cdot \log \mathsf{rk}(f) \ge \Omega(D(f)),$$

and the theorem follows by rearranging.

Remark 5.9. The main difference between the above analysis and the proof of Theorem 5.2 is the number of steps needed to obtain either a matrix of size 1 or one of rank at most 5. In the case of XOR, if the rank does not reduce by a factor of 2, the size shrinks by a factor of $2^{-T/n-2}$. The number of times this can happen is at most $O(2^{T/n}\cdot \mathsf{rk}(f))$. Hence, the number of leaves is now

$$\binom{O(\mathsf{rk}(f)\cdot 2^{T/n}+\log \mathsf{rk}(f))}{O(\log \mathsf{rk}(f))}\cdot O(1) \le O(\mathsf{rk}(f)\cdot 2^{T/n})^{O(\log \mathsf{rk}(f))}$$
$$\le 2^{O((T/n+\log \mathsf{rk}(f))\cdot \log \mathsf{rk}(f))},$$

which is asymptotically the same as in the general case. The remainder of the proof is identical.

5.5 A Lifting Theorem without Rank

In this section, we proof Corollary 5.5. The main idea is based on the following claim; a similar statement can be found in Rezende et al [56]. Below, we include a proof for the sake of completeness.

Lemma 5.10. *For any two functions $g:\{0,1\}^n\to\{0,1\}$ and $f:\mathcal{X}\times\mathcal{Y}\to\{0,1\}$, it holds that $\mathsf{rk}(g\circ f)\ge(\mathsf{rk}(f)-1)^{\mathsf{deg}(g)}$.*

Proof. For shorthand, denote by d, the degree of f. By definition, there exists a subset of size d whose corresponding coefficient in the polynomial expansion of g is non-zero. We can assume without loss of generality that this set is $[d]$, otherwise, we can rename the variables to ensure this. Let $u_1, \ldots, u_r$ be a maximal set of linearly independent rows of M_f, and let $x_1, \ldots, x_r$ be the corresponding inputs. Further, define the vectors $\widetilde{u}_1, \ldots, \widetilde{u}_r$, where $\widetilde{u}_i$ is the projection of u_i onto the space orthogonal to the all-ones vector, $\mathbf{1}$. We note that the dimension of $\mathsf{span}(\widetilde{u}_1, \ldots, \widetilde{u}_r)$ is at least $r - 1$.

In what follows, we adopt the following notation for the tensor product of 2 (or more vectors). Given two vectors $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^k$, we denote the tensor product of u with v by $u \otimes v \in \mathbb{R}^{mk}$ where $u \otimes v[i, j] = u(i) \cdot v(j)$.

The key observation is that the projection of the rows of $M_{g \circ f}$ to the space

$$\mathcal{V} := \mathsf{span}(\{v_1 \otimes \ldots \otimes v_d \otimes \underbrace{\mathbf{1} \otimes \ldots \otimes \mathbf{1}}_{n-d \text{ times}} : v_i \in \{\widetilde{u}_1, \ldots, \widetilde{u}_r\}\})$$

has full rank. Indeed, consider any function $h : [n] \to [r]$ and let u_h be the row corresponding to the inputs $x_{h(1)}, \ldots, x_{h(n)}$. For any $y_1, \ldots, y_n$, using the multilinear polynomial for g we can write

$$\begin{aligned} u_h(y_1, \ldots, y_n) &= g(f(x_1, y_1), \ldots, f(x_n, y_n)) \\ &= \sum_{S \subseteq [n]} \alpha_S \cdot \prod_{i \in S} f(x_i, y_i) \\ &= \sum_{S \subseteq [n]} \alpha_S \cdot \prod_{i \in S} u_{h(i)}(y_i). \end{aligned}$$

For any set S, the last quantity above can be written as tensor product. For example, if we let $S = [t]$ then

$$\begin{aligned} \prod_{i \in [t]} u_{h(i)}(y_i) &= u_{h(1)} \otimes \ldots u_{h(t)}[y_1, \ldots, y_t] \\ &= u_{h(1)} \otimes \ldots u_{h(t)} \otimes \underbrace{\mathbf{1} \otimes \ldots \otimes \mathbf{1}}_{n-t \text{ times}}[y_1, \ldots, y_n]. \end{aligned}$$

Applying this to a general set S, we can write

$$\prod_{i \in S} u_{h(i)}(y_i) = \otimes_{i \in S} u_{h(i)} \otimes_{i \notin S} \mathbf{1}[y_1, \ldots, y_n],$$

where the subscript is used to denote the vector in the i-th coordinate of the tensor product depending on whether or not $i \in S$.

For any set $S \neq [d]$ of size at most d, the projection of $\otimes_{i\in S} u_{h(i)} \otimes_{i\notin S} \mathbf{1}$ onto $\mathcal{V}$ is zero, since there exists $i \in [d] \setminus S$ such that the vector in the i-th coordinate of the tensor product is $\mathbf{1}$. Moreover, by the definition of degree, $\alpha_S = 0$ for sets S of size larger than d. Lastly, the projection of $\alpha_{[d]} \cdot \otimes_{i\in[d]} u_{h(i)} \otimes_{i=d+1}^{n} \mathbf{1}$ is exactly $\alpha_{[d]} \cdot \otimes_{i\in[d]} \widetilde{u}_{h(i)} \otimes_{i=d+1}^{n} \mathbf{1}$.

This establishes that the projection of the rows of $M_{g\circ f}$ to $\mathcal{V}$ has full rank. It follows the rank of $M_{g\circ f}$ is at least the dimension of $\mathcal{V}$, which is at least $(r-1)^d = (\mathsf{rk}(f)-1)^{\mathsf{deg}(g)}$. □

We are now ready to prove Corollary 5.5. The proof uses an observation pointed out to us by Yang [65] after the publication of [28].

Proof of Corollary 5.5. Recalling Fact 2.8, we have $D(g \circ f) \geq \lceil \log \mathsf{rk}(g \circ f) \rceil$. Therefore, we can combine Theorem 5.3 and Lemma 5.10 to conclude

$$D(g \circ f) \geq \max\left\{\mathsf{s}(g) \cdot \left(\frac{\Omega(D(f))}{\log \mathsf{rk}(f)} - \log \mathsf{rk}(f)\right), \mathsf{deg}(g) \cdot \log(\mathsf{rk}(g) - 1)\right\}.$$

Using the fact that $\max\{a, b\} \geq \lambda \cdot a + (1-\lambda) \cdot b$, for any $\lambda \in [0, 1]$, we can set $\lambda = \mathsf{deg}(g)/(2\mathsf{s}(g) + \mathsf{deg}(g))$ to get

$$\begin{aligned} D(g \circ f) &\geq \frac{\mathsf{s}(g) \cdot \mathsf{deg}(g)}{2\mathsf{s}(g) + \mathsf{deg}(g)} \cdot \left(\frac{\Omega(D(f))}{\log \mathsf{rk}(f)} - \log \mathsf{rk}(f) + 2\log(\mathsf{rk}(f) - 1)\right) \\ &\geq \frac{\mathsf{s}(g) \cdot \mathsf{deg}(g)}{2\mathsf{s}(g) + \mathsf{deg}(g)} \cdot \Omega\left(\frac{D(f)}{\log \mathsf{rk}(f)} + \log \mathsf{rk}(f)\right), \end{aligned}$$

where we used the fact that for c_0 large enough, $(\mathsf{rk}(f) - 1)^2 \geq \mathsf{rk}(f)^{3/2}$.

Lastly, we can write $\mathsf{s}(g) \cdot \mathsf{deg}(g)/(2\mathsf{s}(g) + \mathsf{deg}(g)) \geq \min\{\mathsf{s}(g), \mathsf{deg}(g)\}/3$ and by the AM-GM inequality, we have $D(f)/\log \mathsf{rk}(f) + \log \mathsf{rk}(f) \geq 2\sqrt{D(f)}$. It follows that $D(g \circ f) = \Omega(\min\{\mathsf{s}(g), \mathsf{deg}(g)\} \cdot \sqrt{D(f)})$. □

5.6 Conclusions

In this chapter, we have tried to reason about the communication complexity of computing $g \circ f$ for arbitrary f and g. We expect that this is $\Omega(\mathsf{DT}(g) \cdot D(f))$, for any sufficiently complex gadget g. Corollary 5.5 gives the lower bound $\Omega(\min\{\mathsf{s}(g), \mathsf{deg}(g)\} \cdot \sqrt{D(f)}) \geq$

$\Omega(\mathsf{DT}(g)^{1/6} \cdot \sqrt{D(f)})$ which can be seen as progress towards this. Below, we show that for some gadgets this can be further improved.

1. For certain gadgets g, we can obtain $D(g \circ f) \geq \Omega(\mathsf{DT}(g)^{1/3} \cdot \sqrt{D(f)})$ using the notion of *block-sensitivity*, another well-studied [45] Boolean function complexity measure. The block-sensitivity of g at $z \in \{0,1\}^n$ is the maximum number of disjoint sets $S_1, \ldots, S_t$ such that for all $i \in [t]$,

$$g(z) \neq g(z^{\oplus S_i}), \text{ where } z_j^{\oplus S_i} = \begin{cases} 1 - z_j, & \text{if } j \in S_i \text{ and} \\ z_j, & \text{otherwise.} \end{cases}$$

   The block-sensitivity of g, denoted $\mathsf{bs}(g)$, is the maximum across all z, of the block-sensitivity of f at z. By definition, $\mathsf{s}(g) \leq \mathsf{bs}(g)$. Moreover, the block-sensitivity is known to give a better upper bound for the decision-tree complexity than the sensitivity. In particular, Midrijānis [42] showed that $\mathsf{DT}(g) \leq \mathsf{bs}(g) \cdot \mathsf{deg}(g)$. By the lower bound in Equation (2.3) we know that $\mathsf{deg}(g) \leq \mathsf{s}(g)^2 \leq \mathsf{bs}(g)^2$. Hence, $\mathsf{DT}(g) \leq \mathsf{bs}(g)^3$.

   We say a function $f(x, y)$ is *row-symmetric*[1] if for any x, there exists $\overline{x} \neq x$ such that for all y, we have $f(x, y) = 1 - f(\overline{x}, y)$. An example of such a function is the index function, Ind_m.

   Theorem 5.11. *There exists an absolute constant $c_0 > 0$ such that for any row-symmetric function f with $D(f) \geq c_0$ and any function $g : \{0,1\}^n \to \{0,1\}$ we have $D(g \circ f) = \Omega(\min\{\mathsf{bs}(g), \mathsf{deg}(g)\} \cdot \sqrt{D(f)}) = \Omega(\mathsf{DT}(g)^{1/3} \cdot \sqrt{D(f)})$.*

   Proof. Suppose g has block-sensitivity b, achieved at a point $\widetilde{z}$ by sets $S_1, \ldots, S_b$. Consider the function $g' : \{0,1\}^b \to \{0,1\}$ given by

$$g'(z) = g(z'), \text{ where } z'_j = \begin{cases} |\widetilde{z}_j - z_i|, & \text{if } \exists i \text{ such that } j \in S_i \text{ and} \\ \widetilde{z}_j, & \text{otherwise.} \end{cases}$$

[1] A related notion is that of *flippability*, defined in [21].

We note that g' has sensitivity b, since $g'(0) = g(\widetilde{z})$, and $g'(0^{\oplus\{i\}}) = g(\widetilde{z}^{\oplus S_i})$. Moreover, any protocol that computes $g \circ f$ can also be used to compute $g' \circ f$ in the following way. Suppose Alice and Bob gets inputs $x_1, \ldots, x_b$ and $y_1, \ldots, y_b$. For each set S_i and coordinate $j \in S_i$, Alice sets $x_j = x_i$ if $\widetilde{z}_j = 0$, and otherwise, sets $x_j = \overline{x}_i$ (from the row-symmetry property). Bob sets $y_j = y_i$ for each $j \in S_i$. For every coordinate $j \notin S_1 \cup \ldots \cup S_b$, the players arbitrarily fix inputs such that $f(x_j, y_j) = \widetilde{z}_j$. They can now run the protocol for $g \circ f$ to compute $g' \circ f$ and it follows that $D(g' \circ f) \leq D(g \circ f)$. Moreover, Theorem 5.3 shows that

$$D(g' \circ f) \geq \mathsf{s}(g') \cdot \left(\frac{\Omega(D(f))}{\log \mathsf{rk}(f)} - \log \mathsf{rk}(f) \right) = \mathsf{bs}(g) \cdot \left(\frac{\Omega(D(f))}{\log \mathsf{rk}(f)} - \log \mathsf{rk}(f) \right).$$

The theorem follows by applying the same arguments as in the proof of Corollary 5.5.

□

2. Anup Rao observed that for certain gadgets g, such as inner product IP_m, one can improve Theorem 5.3 to obtain $D(g \circ f) = \Omega(\mathsf{s}(g) \cdot D(f)/\log \mathsf{rk}_2(f))$, where $\mathsf{rk}_2(f)$ is the rank of M_f over $\mathbb{F}_2$. This can be seen by modifying the proof of Lemma 5.7 and Theorem 5.3 to keep track of $\mathsf{rk}_2(f)$ instead of $\mathsf{rk}(f)$.

Theorem 5.12. *There exists an absolute constant $c_0 > 0$ such that for any function $f(xy)$ with $D(f) \geq c_0$ and any $g : \{0,1\}^n \to \{0,1\}$ we have*

$$D(g \circ f) = \mathsf{s}(g) \cdot \left(\frac{\Omega(D(f))}{\log \mathsf{rk}_2(f)} - \log \mathsf{rk}_2(f) \right).$$

We sketch a proof of this theorem. First, the statement of Lemma 5.7 can be strengthened to show that there is a monochromatic rectangle of density at least $2^{-2T/\mathsf{s}(g)} \cdot (4 \cdot \mathsf{rk}_2(f))^{-2}$. To see this, we make the following slight modification to the proof. At the start, we suppose that f is biased – the appropriate threshold for the bias must be updated to $|\mathbf{E}_{xy}[f(xy)] - 1/2| > 1/2 - 1/(4 \cdot \mathsf{rk}_2(f))$. Next, as in the proof, we conclude that at least half the rows are all biased towards the same bit with probability at least $1 - 1/(2 \cdot \mathsf{rk}_2(f))$. Furthermore, there exist some r rows, $x_1, \ldots, x_r$

that are maximally linearly independent[2] over $\mathbb{F}_2$. Again, like in the proof, we can show that for half the columns the entries corresponding to $x_1, \ldots, x_r$ are all the same bit. Using the linear dependence of any other row on $x_1, \ldots, x_r$ over $\mathbb{F}_2$, we can infer the existence of a rectangle of density at least $1/8$. The remainder of the proof of Lemma 5.7 proceeds as before and does not use any algebraic property of rank.

Lastly, we need to change the proof of Theorem 5.3 in the following way. Each time we find a monochromatic rectangle R for g using Lemma 5.7, we can recurse on a sub-matrix where either $\mathsf{rk}_2(g)$ goes down by a factor of $4/5$ or the size of the matrix shrinks by the appropriate amount. If the rank over $\mathbb{F}_2$ is at most 5, one can just use 6 bits of communication to compute the function since $D(f) \leq \mathsf{rk}_2(f) + 1$. This calculation yields

$$D(g \circ f) = \mathsf{s}(g) \cdot \left(\frac{\Omega(D(f))}{\log \mathsf{rk}_2(f)} - \log \mathsf{rk}_2(f) \right).$$

Furthermore, for any gadget satisfying $D(f) = \Omega(\log^2 \mathsf{rk}(f))$, we get that $D(g \circ f) = \Omega(\mathsf{s}(g) \cdot D(f)/\log \mathsf{rk}_2(f))$. In particular, for the inner product gadget, we know that $\mathsf{rk}_2(\mathrm{IP}_m) \leq m$ and $D(\mathrm{IP}_m) = \Omega(m)$.

[2] this step is slightly different from the proof: the proof assumes that $x_1, \ldots, x_r$ are maximally linearly independent over $\mathbb{R}$.

Chapter 6

OPEN PROBLEMS AND CONCLUDING REMARKS

6.1 Randomized Communication

In Chapter 3 we gave an XOR lemma for randomized communication showing that if f requires C bits to be computed with constant success probability then computing $f^{\oplus n}$ with probability at least $1/2 + \exp(-\Omega(n))$ requires communication $\widetilde{\Omega}(\sqrt{n} \cdot C)$. We built on information-theoretic techniques developed over a long line of works that addressed problems in parallel repetition and direct sum theorems in randomized communication. The new XOR lemma makes use of some new definitions, such as rectangular distributions and marginal information. Armed with these new definitions, we wonder if we can go back to obtain improved parallel repetition theorems and optimal direct-sum statements? One way to obtain an improved direct-sum result for randomized communication is to prove a better compression result for marginal information.

Open Question 6.1. *Given a Boolean function f and a protocol p with communication C and marginal information $\mathsf{M}_I(p, f) \leq O(I)$, does there exist a protocol τ simulating p with communication $O(I^{O(1)} \log C)$ and constant advantage?*

Another direction that can potentially benefit from the definition of marginal information is understanding the randomized communication complexity of computing $g \circ f$ for arbitrary Boolean functions g. A simpler version of this is to show that computing $f^{\vee n}(x, y) := f(x_1, y_1) \vee \ldots \vee f(x_n, y_n)$ requires large communication.

Open Question 6.2. *Does there exist $C > 0$ and $\varepsilon \in (0, 1)$ such that for any Boolean f with $\mathsf{suc}(C, f) \leq \varepsilon$, it holds that $\mathsf{suc}(\widetilde{O}(\sqrt{n}C), f^{\vee n}) \leq \varepsilon$?*

The challenge in proving the above statement is the sub-additivity step; if μ is the hard distribution for f then the hard distribution for $f^{\vee n}$ is different from μ^n. Indeed, if we were to use μ^n then $f^{\vee n}$ can be computed with probability $1 - \exp(-\Omega(n))$ by simply outputting

1. The candidate hard distribution for $f^{\vee n}$ is the following one: choose a random coordinate $i \in [n]$ and sample the inputs in i-th coordinate from μ, and everywhere else according to $\mu(x, y|f^{-1}(0))$. The distribution above unlike μ^n, has correlations across the coordinates. In our proof of Theorem 4.2, we crucially used the independence of the input distribution across the coordinates. We note that such a distribution was used by Razborov [55] to prove a lower bound on the randomized communication complexity of disjointness, which can be thought of as the n-fold OR of the AND function on 2 bits.

6.2 Deterministic Communication

A natural follow-up to Theorem 5.3 and Corollary 5.5 is whether $D(g \circ f) = \Omega(\mathsf{DT}(g) \cdot D(f))$ for any f with communication complexity that is a sufficiently large constant. This is open even for specific gadgets f. For example, the best known bound for f being the index gadget Ind_m is due to Lovett et al. [40] who showed that for $m \geq \Omega(n^{1+\varepsilon})$, it holds that $D(g \circ \mathsf{Ind}_m) = \Omega(\mathsf{DT}(g) \cdot D(\mathsf{Ind}_m))$. We outline a direction to prove a near-optimal lifting theorem for a different gadget, IP_m (inner-product). Indeed, as discussed in item 2 in Section 5.6, $\mathsf{rk}_2(\mathsf{IP}_m) = \log m$ while $D(\mathsf{IP}_m) = m$. Hence, if sensitivity in Theorem 5.3 can be replaced with decision-tree complexity, specifically for the inner-product function then, one could show that for any m that is a sufficiently large constant, $D(g \circ \mathsf{IP}_m) = \Omega(\mathsf{DT}(g) \cdot D(\mathsf{IP}_m)/\log D(\mathsf{IP}_m))$.

Open Question 6.3. *Does there exists $m_0 > 0$ such that for any $m \geq m_0$ and any function $g : \{0,1\}^n \to \{0,1\}$,*

$$D(g \circ \mathsf{IP}_m) \geq \mathsf{DT}(g) \cdot \left(\frac{\Omega(m)}{\log m} - \log m\right).$$

We reiterate a basic question posed by Karchmer, Raz and Wigderson [34], which asks about the existence of function f whose corresponding KW-game has large amortized communication complexity. We wonder if the information-theoretic techniques in this thesis can be used to attack this problem.

Open Question 6.4 ([34])**.** *Is there a Boolean $f : \{0,1\}^n \to \{0,1\}$ with*

$$\lim_{k \to \infty} \frac{D(\mathrm{KW}_f^k)}{k} = \omega(\log n)?$$

BIBLIOGRAPHY

[1] Miklós Ajtai. "A lower bound for finding predecessors in Yao's call probe model". In: *Comb.* 8.3 (1988), pp. 235–247. DOI: 10.1007/BF02126797. URL: https://doi.org/10.1007/BF02126797.

[2] Josh Alman et al. *More Asymmetry Yields Faster Matrix Multiplication*. 2024. arXiv: 2404.16349 [cs.DS]. URL: https://arxiv.org/abs/2404.16349.

[3] Noga Alon, Yossi Matias, and Mario Szegedy. "The space complexity of approximating the frequency moments". In: *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*. STOC '96. Philadelphia, Pennsylvania, USA: Association for Computing Machinery, 1996, pp. 20–29. ISBN: 0897917855. DOI: 10.1145/237814.237823. URL: https://doi.org/10.1145/237814.237823.

[4] Ziv Bar-Yossef et al. "An Information Statistics Approach to Data Stream and Communication Complexity". In: *43rd Symposium on Foundations of Computer Science (FOCS 2002), 16-19 November 2002, Vancouver, BC, Canada, Proceedings*. IEEE Computer Society, 2002, pp. 209–218. DOI: 10.1109/SFCS.2002.1181944. URL: https://doi.org/10.1109/SFCS.2002.1181944.

[5] Ziv Bar-Yossef et al. "Counting Distinct Elements in a Data Stream". In: *Proceedings of the 6th International Workshop on Randomization and Approximation Techniques*. RANDOM '02. Berlin, Heidelberg: Springer-Verlag, 2002, pp. 1–10. ISBN: 3540441476.

[6] Boaz Barak et al. "How to Compress Interactive Communication". In: *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*. STOC '10. Cambridge, Massachusetts, USA: Association for Computing Machinery, 2010, pp. 67–76. ISBN: 9781450300506. DOI: 10.1145/1806689.1806701. URL: https://doi.org/10.1145/1806689.1806701.

[7] Gábor Braun and Sebastian Pokutta. "Common Information and Unique Disjointness". In: *Algorithmica* 76.3 (2016), pp. 597–629. DOI: 10.1007/S00453-016-0132-0. URL: https://doi.org/10.1007/s00453-016-0132-0.

[8] Mark Braverman. "Interactive Information Complexity". In: *SIAM Journal on Computing* 44.6 (2015), pp. 1698–1739. DOI: 10.1137/130938517. eprint: https://doi.org/10.1137/130938517. URL: https://doi.org/10.1137/130938517.

[9] Mark Braverman and Gillat Kol. "Interactive Compression to External Information". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2018. Los Angeles, CA, USA: Association for Computing Machinery, 2018, pp. 964–977. ISBN: 9781450355599. DOI: 10.1145/3188745.3188956. URL: https://doi.org/10.1145/3188745.3188956.

[10] Mark Braverman and Anup Rao. "Information Equals Amortized Communication". In: *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. Los Alamitos, CA, USA: IEEE Computer Society, Oct. 2011, pp. 748–757. DOI: 10.1109/FOCS.2011.86. URL: https://doi.ieeecomputersociety.org/10.1109/FOCS.2011.86.

[11] Mark Braverman et al. "Direct Product via Round-Preserving Compression". In: *Automata, Languages, and Programming - 40th International Colloquium, ICALP 2013, Riga, Latvia, July 8-12, 2013, Proceedings, Part I*. Ed. by Fedor V. Fomin et al. Vol. 7965. Lecture Notes in Computer Science. Springer, 2013, pp. 232–243. DOI: 10.1007/978-3-642-39206-1_20. URL: https://doi.org/10.1007/978-3-642-39206-1%5C_20.

[12] Mark Braverman et al. "Direct Products in Communication Complexity". In: *54th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2013, 26-29 October, 2013, Berkeley, CA, USA*. IEEE Computer Society, 2013, pp. 746–755. DOI: 10.1109/FOCS.2013.85. URL: https://doi.org/10.1109/FOCS.2013.85.

[13] Harry Buhrman, Richard Cleve, and Avi Wigderson. "Quantum vs. classical communication and computation". In: *Proceedings of the Thirtieth Annual ACM Symposium*

on Theory of Computing. STOC '98. Dallas, Texas, USA: Association for Computing Machinery, 1998, pp. 63–68. ISBN: 0897919629. DOI: 10.1145/276698.276713. URL: https://doi.org/10.1145/276698.276713.

[14] Harry Buhrman and Ronald de Wolf. "Complexity measures and decision tree complexity: a survey". In: *Theor. Comput. Sci.* (2002).

[15] Amit Chakrabarti et al. "Informational Complexity and the Direct Sum Problem for Simultaneous Message Complexity". In: *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*. IEEE Computer Society, 2001, pp. 270–278. DOI: 10.1109/SFCS.2001.959901. URL: https://doi.org/10.1109/SFCS.2001.959901.

[16] Arkadev Chattopadhyay et al. "Simulation Theorems via Pseudo-random Properties". In: *Comput. Complex.* (2019).

[17] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley series in telecommunications. J. Wiley and Sons, New York, 1991.

[18] Tomás Feder et al. "Amortized Communication Complexity". In: *SIAM Journal on Computing* 24.4 (1995), pp. 736–750. DOI: 10.1137/S0097539792235864.

[19] Anat Ganor, Gillat Kol, and Ran Raz. "Exponential Separation of Information and Communication for Boolean Functions". In: *J. ACM* 63.5 (Nov. 2016). ISSN: 0004-5411. DOI: 10.1145/2907939. URL: https://doi.org/10.1145/2907939.

[20] Dmitry Gavinsky and Shachar Lovett. "En Route to the Log-Rank Conjecture: New Reductions and Equivalent Formulations". In: *Automata, Languages, and Programming*. Springer Berlin Heidelberg, 2014.

[21] Mika Göös and Toniann Pitassi. "Communication lower bounds via critical block sensitivity". In: *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*. STOC '14. New York, New York: Association for Computing Machinery, 2014, pp. 847–856. ISBN: 9781450327107. DOI: 10.1145/2591796.2591838. URL: https://doi.org/10.1145/2591796.2591838.

[22] Mika Göös, Toniann Pitassi, and Thomas Watson. "Deterministic Communication vs. Partition Number". In: *SIAM Journal on Computing* (2018).

[23] Prahladh Harsha et al. "The Communication Complexity of Correlation". In: *IEEE Transactions on Information Theory* 56.1 (2010), pp. 438–449. DOI: 10.1109/TIT.2009.2034824.

[24] Thomas Holenstein. "Parallel Repetition: Simplifications and the No-Signaling Case". In: *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*. STOC '07. San Diego, California, USA: Association for Computing Machinery, 2007, pp. 411–419. ISBN: 9781595936318. DOI: 10.1145/1250790.1250852. URL: https://doi.org/10.1145/1250790.1250852.

[25] Hao Huang. "Induced subgraphs of hypercubes and a proof of the Sensitivity Conjecture". In: *Annals of Mathematics* (2019).

[26] Trinh Huynh and Jakob Nordstrom. "On the virtue of succinct proofs: amplifying communication complexity hardness to time-space trade-offs in proof complexity". In: *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*. STOC '12. New York, New York, USA: Association for Computing Machinery, 2012, pp. 233–248. ISBN: 9781450312455. DOI: 10.1145/2213977.2214000. URL: https://doi.org/10.1145/2213977.2214000.

[27] Siddharth Iyer. "Lifting for Arbitrary Gadgets". In: *Electron. Colloquium Comput. Complex.* TR25-036 (2025). ECCC: TR25-036. URL: https://eccc.weizmann.ac.il/report/2025/036.

[28] Siddharth Iyer and Anup Rao. *An XOR Lemma for Deterministic Communication Complexity*. 2024. arXiv: 2407.01802 [cs.CC]. URL: https://arxiv.org/abs/2407.01802.

[29] Siddharth Iyer and Anup Rao. "XOR Lemmas for Communication via Marginal Information". In: STOC 2024. 2024, pp. 652–658. ISBN: 9798400703836. DOI: 10.1145/3618260.3649726. URL: https://doi.org/10.1145/3618260.3649726.

[30] Rahul Jain, Attila Pereszlényi, and Penghui Yao. "A Direct Product Theorem for the Two-Party Bounded-Round Public-Coin Communication Complexity". In: *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*. IEEE Computer Society, 2012, pp. 167–176. DOI: 10.1109/FOCS.2012.42. URL: https://doi.org/10.1109/FOCS.2012.42.

[31] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. "A Direct Sum Theorem in Communication Complexity via Message Compression". In: *Automata, Languages and Programming*. Ed. by Jos C. M. Baeten et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 300–315. ISBN: 978-3-540-45061-0.

[32] Mark Jerrum et al. *Probabilistic Methods for Algorithmic Discrete Mathematics*. Vol. 16. Algorithms and Combinatorics. Springer-Verlag, 1998.

[33] Xinrui Jia, Ola Svensson, and Weiqiang Yuan. "The Exact Bipartite Matching Polytope Has Exponential Extension Complexity". In: *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 2023, pp. 1635–1654. DOI: 10.1137/1.9781611977554.ch61. URL: https://epubs.siam.org/doi/abs/10.1137/1.9781611977554.ch61.

[34] Mauricio Karchmer, Ran Raz, and Avi Wigderson. "Super-logarithmic depth lower bounds via direct sum in communication complexity". In: *[1991] Proceedings of the Sixth Annual Structure in Complexity Theory Conference*. 1991, pp. 299–304. DOI: 10.1109/SCT.1991.160273.

[35] Mauricio Karchmer and Avi Wigderson. "Monotone circuits for connectivity require super-logarithmic depth". In: *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*. STOC '88. Chicago, Illinois, USA: Association for Computing Machinery, 1988, pp. 539–550. ISBN: 0897912640. DOI: 10.1145/62212.62265. URL: https://doi.org/10.1145/62212.62265.

[36] Gillat Kol. "Interactive Compression for Product Distributions". In: *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*. STOC '16. Cambridge, MA, USA: Association for Computing Machinery, 2016, pp. 987–998. ISBN:

9781450341325. DOI: 10.1145/2897518.2897537. URL: https://doi.org/10.1145/2897518.2897537.

[37] L A Levin. "One-way functions and pseudorandom generators". In: *Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing*. STOC '85. Providence, Rhode Island, USA: Association for Computing Machinery, 1985, pp. 363–365. ISBN: 0897911512. DOI: 10.1145/22145.22185. URL: https://doi.org/10.1145/22145.22185.

[38] László Miklós Lovász and Michael E. Saks. "Lattices, mobius functions and communications complexity". In: *[Proceedings 1988] 29th Annual Symposium on Foundations of Computer Science* (1988).

[39] Shachar Lovett. "Communication is bounded by root of rank". In: *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*. STOC '14. New York, New York: Association for Computing Machinery, 2014, pp. 842–846. ISBN: 9781450327107. DOI: 10.1145/2591796.2591799.

[40] Shachar Lovett et al. "Lifting with Sunflowers". In: *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2022.

[41] Yahel Manor and Or Meir. "Lifting with Inner Functions of Polynomial Discrepancy". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2022*. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022.

[42] Gatis Midrijanis. *Exact quantum query complexity for total Boolean functions*. 2004. arXiv: quant-ph/0403168 [quant-ph]. URL: https://arxiv.org/abs/quant-ph/0403168.

[43] Peter Bro Miltersen et al. "On data structures and asymmetric communication complexity". In: *Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing*. STOC '95. Las Vegas, Nevada, USA: Association for Computing Ma-

chinery, 1995, pp. 103–111. ISBN: 0897917189. DOI: 10.1145/225058.225093. URL: https://doi.org/10.1145/225058.225093.

[44] Noam Nisan. "CREW PRAMS and decision trees". In: *Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing*. Association for Computing Machinery, 1989.

[45] Noam Nisan and Mario Szegedy. "On the degree of Boolean functions as real polynomials". In: *Proceedings of the Twenty-Fourth Annual ACM Symposium on Theory of Computing*. Association for Computing Machinery, 1992.

[46] Noam Nisan and Avi Wigderson. "On Rank vs. Communication Complexity". In: *Comb.* (1995).

[47] Mihai Pătraşcu. "Unifying the Landscape of Cell-Probe Lower Bounds". In: *SIAM J. Comput.* 40.3 (June 2011), pp. 827–847. ISSN: 0097-5397. DOI: 10.1137/09075336X. URL: https://doi.org/10.1137/09075336X.

[48] Sivaramakrishnan Natarajan Ramamoorthy and Anup Rao. "How to Compress Asymmetric Communication". In: *Proceedings of the 30th Conference on Computational Complexity*. CCC '15. Portland, Oregon: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015, pp. 102–123. ISBN: 9783939897811.

[49] Anup Rao. "Parallel Repetition in Projection Games and a Concentration Bound". In: *SIAM Journal on Computing* 40.6 (2011), pp. 1871–1891. DOI: 10.1137/080734042. eprint: https://doi.org/10.1137/080734042. URL: https://doi.org/10.1137/080734042.

[50] Anup Rao and Makrand Sinha. "Simplified Separation of Information and Communication". In: *Theory of Computing* 14.20 (2018), pp. 1–29. DOI: 10.4086/toc.2018.v014a020. URL: https://theoryofcomputing.org/articles/v014a020.

[51] Anup Rao and Amir Yehudayoff. "Anticoncentration and the Exact Gap-Hamming Problem". In: *SIAM J. Discret. Math.* 36.2 (Jan. 2022), pp. 1071–1092. ISSN: 0895-4801. DOI: 10.1137/21M1435288. URL: https://doi.org/10.1137/21M1435288.

[52] Anup Rao and Amir Yehudayoff. *Communication Complexity: and Applications*. Cambridge University Press, 2020. DOI: 10.1017/9781108671644.

[53] Ran Raz. "A Parallel Repetition Theorem". In: *Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing*. STOC '95. Las Vegas, Nevada, USA: Association for Computing Machinery, 1995, pp. 447–456. ISBN: 0897917189. DOI: 10.1145/225058.225181. URL: https://doi.org/10.1145/225058.225181.

[54] Ran Raz and Pierre McKenzie. "Separation of the Monotone NC Hierarchy". In: *Comb.* (1999).

[55] Alexander A. Razborov. "On the Distributional Complexity of Disjointness". In: *Theor. Comput. Sci.* 106.2 (1992), pp. 385–390. DOI: 10.1016/0304-3975(92)90260-M. URL: https://doi.org/10.1016/0304-3975(92)90260-M.

[56] Susanna F. de Rezende et al. "Lifting with Simple Gadgets and Applications to Circuit and Proof Complexity". In: *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*. Ed. by Sandy Irani. IEEE, 2020, pp. 24–30. DOI: 10.1109/FOCS46700.2020.00011. URL: https://doi.org/10.1109/FOCS46700.2020.00011.

[57] Thomas Rothvoss. "The Matching Polytope Has Exponential Extension Complexity". In: *J. ACM* 64.6 (Sept. 2017). ISSN: 0004-5411. DOI: 10.1145/3127497. URL: https://doi.org/10.1145/3127497.

[58] Georg Schnitger and Bala Kalyanasundaram. "The probabilistic communication complexity of set intersection". In: *Proceedings of the Second Annual Conference on Structure in Complexity Theory, Cornell University, Ithaca, New York, USA, June 16-19, 1987*. IEEE Computer Society, 1987, pp. 41–47. URL: https://ieeexplore.ieee.org/document/10319253.

[59] Claude. E. Shannon. "A mathematical theory of communication". In: *The Bell System Technical Journal* 27.3 (1948), pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[60] Alexander A. Sherstov. "Compressing Interactive Communication Under Product Distributions". In: *SIAM Journal on Computing* 47.2 (2018), pp. 367–419. DOI: 10.1137/16M109380X. eprint: https://doi.org/10.1137/16M109380X. URL: https://doi.org/10.1137/16M109380X.

[61] Alexander A. Sherstov. "The Pattern Matrix Method". In: *SIAM J. Comput.* (2011).

[62] Makrand Sinha. "Lower Bounds for Approximating the Matching Polytope". In: *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*. Ed. by Artur Czumaj. SIAM, 2018, pp. 1585–1604. DOI: 10.1137/1.9781611975031.104. URL: https://doi.org/10.1137/1.9781611975031.104.

[63] Volker Strassen. "Gaussian elimination is not optimal". In: *Numerische Mathematik* (1969).

[64] Benny Sudakov and István Tomon. *Matrix discrepancy and the log-rank conjecture*. 2023. arXiv: 2311.18524 [math.CO].

[65] Guangxu Yang. Private Communication. Nov. 2024.

[66] Andrew C. Yao. "Theory and application of trapdoor functions". In: *23rd Annual Symposium on Foundations of Computer Science (sfcs 1982)*. 1982, pp. 80–91. DOI: 10.1109/SFCS.1982.45.

[67] Andrew Chi-Chih Yao. "Some complexity questions related to distributive computing(Preliminary Report)". In: *Proceedings of the Eleventh Annual ACM Symposium on Theory of Computing*. STOC '79. Atlanta, Georgia, USA: Association for Computing Machinery, 1979, pp. 209–213. ISBN: 9781450374385. DOI: 10.1145/800135.804414. URL: https://doi.org/10.1145/800135.804414.

[68] Huacheng Yu. "Strong XOR Lemma for Communication with Bounded Rounds : (extended abstract)". In: *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*. IEEE, 2022, pp. 1186–1192. DOI: 10.1109/FOCS54457.2022.00114. URL: https://doi.org/10.1109/FOCS54457.2022.00114.

[69] Shengyu Zhang. "On the Tightness of the Buhrman-Cleve-Wigderson Simulation". In: *Proceedings of the 20th International Symposium on Algorithms and Computation*. ISAAC '09. Honolulu, Hawaii: Springer-Verlag, 2009, pp. 434–440. ISBN: 9783642106309. DOI: 10.1007/978-3-642-10631-6_45. URL: https://doi.org/10.1007/978-3-642-10631-6_45.