Electronic Colloquium on Computational Complexity, Report No. 194 (2023)

XOR Lemmas for Communication via Marginal Information

Siddharth Iyer*
siyer@cs.washington.edu

Anup Rao*
anuprao@cs.washington.edu

December 4, 2023

Abstract

We define the *marginal information* of a communication protocol, and use it to prove XOR lemmas for communication complexity. We show that if every C-bit protocol has bounded advantage for computing a Boolean function f, then every $\widetilde{\Omega}(C\sqrt{n})$-bit protocol has advantage $\exp(-\Omega(n))$ for computing the n-fold xor $f^{\oplus n}$. We prove exponentially small bounds in the average case setting, and near optimal bounds for product distributions and for bounded-round protocols.

Contents

*Supported by NSF award 2131899.

ISSN 1433-8092

1. Introduction

If a function is hard to compute, is it even harder to compute it many times? This old question is often challenging, and new answers are usually accompanied by foundational ideas. We give new answers in the framework of communication complexity, accompanied by a new measure of complexity called *marginal information*. This definition provides a new tool for proving lower bounds in theoretical computer science.

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 [KRW95], lower bounds on data structures [Păt11] and lower bounds on the extension complexity of polytopes [BP16, JSY23, Rot17, Sin18], to name a few nice examples. We refer the reader to the textbook [RY20] for an introduction to the basic definitions and concepts in communication complexity, the role played by the questions we address here, and the connections to other areas.

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

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

So, f^n computes f on n different pairs of inputs, and $f^{\oplus n}$ computes the parity of the outputs of f^n. If f is hard to compute, are f^n and $f^{\oplus n}$ even harder to compute? For deterministic communication complexity, Feder, Kushilevitz, Naor and Nisan [FKNN95] 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. In this work, we study randomized communication complexity. Let $\|\pi\|$ denote the communication complexity of a randomized communication protocol π and define the advantage:

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

This quantity measures the best worst-case advantage achievable by a C-bit protocol over random guessing. We can now state our main result:

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

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

The constant $1/2$ is not important, it can be replaced by any constant less than 1. Some assumption of the type $C > 1/\kappa$ is necessary, because if $x, y \in \{0,1\}$ and $f(xy) = x \oplus y$, then $\mathsf{adv}(2, f^{\oplus n}) = 1$. Prior to our work, the best known upper bound was proved by the second author with Barak, Braverman and Chen [BBCR10]. They showed that under a similar bound on the communication, the advantage is bounded by $1/2$. Our work builds on the work of Yu [Yu22], who proved exponentially small bounds in the setting of bounded-round protocols. Our ideas lead to many results similar to Theorem 1. Next, we review the history that led us to the notion of marginal information, and explain the intuitions behind the choices made in the definition before describing all of our results in Section 1.2.

[1] Throughout, we drop the delimiters between variables. $f(xy)$ is to be read as $f(x, y)$.

1.1. The evolution of information complexity

Marginal information is the most recent advance in an evolution of definitions about 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 5.

Step 2 Marginal information is subadditive, so the marginal information for computing f is smaller by a factor of n; see Theorem 6.

Step 3 Small marginal information can be compressed to give protocols with small communication; see Theorems 7 to 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 [SK87]. The proof was subsequently simplified by Razborov [Raz92], who gave a beautiful short argument that used Shannon's notion of entropy [Sha48] 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 two 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 subadditivity of entropy. Step 3 is also not too difficult.

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 proved the first exponentially small bounds [Raz95] in this context, using the Kullback-Liebler divergence as a measure of information. Given a distribution $p(xy)$, and a carefully chosen event W, Raz measured the divergence

$$\begin{aligned} &\mathop{\mathbb{E}}_{p(xy|W)}\Big[\mathsf{D}(p(x|yW)||p(x|y)) + \mathsf{D}(p(y|xW)||p(y|x))\Big] \\ &= \mathop{\mathbb{E}}_{p(xy|W)}\Big[\log\Big(\frac{p(x|yW)}{p(x|y)}\cdot\frac{p(y|xW)}{p(y|x)}\Big)\Big]. \end{aligned} \tag{1}$$

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 subadditivity of divergence and a similar set of clever random variables as in [Raz92] to split the dependence and accomplish Step 2. Later, Holenstein [Hol07] introduced a method called *correlated sampling* to simplify the analogue of Step 3 in Raz's proof, and obtained better bounds. The second author used these tools to prove optimal bounds for parallel repetition in the setting relevant to probabilistically checkable proofs [Rao11].

Chakrabarti, Shi, Wirth and Yao [CSWY01] were the first to propose using general measures of information complexity to address the questions we consider in this paper. Let xy denote the inputs, m denote the public randomness and transcript of a communication protocol and $p(xym)$ denote the joint distribution induced by the protocol[2]. [CSWY01] proposed to measure

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

the mutual information

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

Years later, this measure was renamed *external information* by [BBCR10]. 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 subadditivity 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 [JRS03], and Harsha, Jain, McAllester and Radhakrishnan [HJMR10] 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 [BYJKS02] showed how 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 the second author together with Barak, Braverman and Chen [BBCR10] 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

$$\mathsf{I}(M : X|Y) + \mathsf{I}(M : Y|X) = \mathop{\mathbb{E}}_{p(xym)} \Big[\log \Big(\frac{p(x|ym)}{p(x|y)} \cdot \frac{p(y|xm)}{p(y|x)} \Big) \Big]. \tag{2}$$

The internal information measures what is learned by each party about the other's input. Equation (1) was the inspiration for Equation (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 [BBCR10] proved that subadditivity 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 $\mathsf{adv}(C, f) \leq 1/2$, then

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

which was the first result along the lines of Theorem 1. Later, the second author and Braverman [BR11] argued that this is the *right* definition of information, because the internal information cost of a function is equal to the amortized communication complexity of that function. This suggested that the internal information might well be the last word in this evolution of definitions, because it could be defined purely using the concept of communication complexity. It seemed like the only path to better results was through better methods to compress internal information. This is a belief we no longer hold.

Nevertheless, a flurry of ideas about compressing protocols with internal information I and communication C followed. Braverman [Bra15] showed how to obtain protocols with communication $\approx 2^{O(I)}$. The second author and Ramamoorthy [RR15] 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 $\approx I_A + \sqrt[4]{I_B \cdot C^3}$. Two excellent papers, the first by Kol [Kol16] and the second by Sherstov [She18], showed that $\approx I \log^2 I$ communication can be achieved when the inputs come from a product distribution. Ganor, Kol and Raz [GKR16] (see also [RS18]) 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 $\approx 2^I$.

The next definition to evolve was proposed by the second author together with Braverman, Weinstein and Yehudayoff [BRWY13b, BRWY13a], inspired by the work of Jain, Pereszlényi and Yao [JPY12]. Rather than bounding the information under the distribution $p(xym)$ induced by the protocol, 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

$$\inf_q \mathsf{I}_q(M : X|Y) + \mathsf{I}_q(M : Y|X) = \inf_q \mathop{\mathbb{E}}_{q(xym)} \Big[\log \Big(\frac{q(x|ym)}{q(x|y)} \cdot \frac{q(y|xm)}{q(y|x)} \Big) \Big], \tag{3}$$

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) 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 huge amount of information. Still, [BRWY13b] 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). [BRWY13b] 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) may not seem very different from Equation (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.

In a paper full of new ideas, Yu [Yu22] 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 [Yu22]:

Definition 2. *Given a set Q consisting of triples (xym), we say that Q is* rectangular *if its indicator function can be expressed as*

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

for some Boolean functions $\mathbb{1}_A, \mathbb{1}_B$. 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) 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. *For $I \geq 1$ and[3] $\delta = 1/15$, define the* marginal information *of a protocol p for computing f:*

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

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 with Equations (2) and (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 $\mathbb{E}_{q(xy|m)}[(-1)^{f(xy)}]$, 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 4. *For $I \geq 1$ and $\delta = 1/15$, define the* external marginal information *of a protocol p for computing f:*

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

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

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

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

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

are working with have a bounded number of rounds, define the worst-case and average case quantities:

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

where throughout, the supremums 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 3, we show that a protocol with small communication and large advantage has small marginal information, to handle Step 1:

   Theorem 5. *For every Boolean function $f(xy)$ and every protocol p of communication complexity C,*

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

   If $\mathsf{adv}_\mu(C, f^{\oplus n}) \ge \exp(-m)$ via a protocol corresponding to the distribution p, then the above theorem implies that $\mathsf{M}_I(p, f^{\oplus n}) \le O(C + Im)$. 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, we prove that marginal information is sub-additive with respect to the n-fold xor of f. If the transcript $m = m_0, m_1, m_2, \ldots$ where m_j denotes the j'th message of the protocol, we show

   Theorem 6. *There is a universal constant Δ such that if $I \ge 1$ and p is a protocol distribution with $p(xy) = \prod_{i=1}^n p(x_i y_i)$, then there is a protocol p_i 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*

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

   If $\mathsf{M}_I(p, f^{\oplus n}) \le O(In)$, this theorem proves that $\mathsf{M}_I(p_i, f) \le O(I)$. This might well be the most technically novel part of our proof; it is certainly where we spent the most time. The heart of the matter is the case $n = 2$. If $\mathsf{M}_I(p, f)$ is small and $n = 2$, then there is 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 $\mathsf{M}_I(p, f)$. We show how to use q, p to generate a new pair

   $$q(x_1 y_1 m^{(1)}), p(x_1 y_1 m^{(1)})$$

   or a new pair

   $$q(x_2 y_2 m^{(2)}), p(x_2 y_2 m^{(2)})$$

   so that the appropriate marginal information is more or less reduced by a factor of 2. This is accomplished by Theorem 27, and Theorem 6 is a straightforward consequence. The proof of Theorem 27 is delicate. A significant first step is the construction of two pairs of rectangular/protocol distributions with the properties described in Equations (13) to (16).

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.

In Theorem 6 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, because in our proof of Theorem 27 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 major obstacle for the high-level strategy.

3. In Sections 7 and 9 to 11, we show how to compress marginal information to handle Step 3. We have been able to match many of the prior results [BBCR10, BR11, Bra15] 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 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 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 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 [Kol16, She18, BK18], but we have not managed to obtain such results. We also obtain results that are independent of the communication complexity:

   Theorem 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 $\mathsf{adv}_\mu(\Delta I, f) \geq \exp(-I\Delta)$.*

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

   Theorem 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 1, as well as several other results of that flavor.

1.2. 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:

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

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

$$\begin{aligned}
\mathsf{adv}(C,f) &= \inf_\mu \mathsf{adv}_\mu(C,f),\\
\mathsf{suc}(C,f) &= \inf_\mu \mathsf{suc}_\mu(C,f),\\
\mathsf{adv}^r(C,f) &= \inf_\mu \mathsf{adv}^r_\mu(C,f),\\
\mathsf{suc}^r(C,f) &= \inf_\mu \mathsf{suc}^r_\mu(C,f).
\end{aligned} \tag{4}$$

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 1 is proved by proving this stronger bound:

Theorem 11. *There is a universal constant $\kappa > 0$ such that if $C > 1/\kappa$ and $\mathsf{adv}_\mu(C,f) \le \kappa$, then $\mathsf{adv}_{\mu^n}(\kappa C\sqrt{n}/\log(Cn), f^{\oplus n}) \le \exp(-\kappa n)$.*

To prove Theorem 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 $T/n \ge 1$, we set $I = T/n$ and apply Theorem 5 to show that $\mathsf{M}_I(p, f^{\oplus n}) \le O(T + \kappa I n) \le O(In)$. Next, apply Theorem 6 to find a protocol p' with $\mathsf{M}_I(p', f) \le O(I)$. Finally, apply Theorem 7 to obtain a protocol computing f with advantage $\Omega(1)$ and communication proportional to

$$\begin{aligned}
\frac{T}{n} + 2\sqrt{IT}\log(T) &\le \frac{T}{n} + 2\frac{T\log T}{\sqrt{n}}\\
&\lesssim \frac{\kappa C}{\log nC}\cdot\log T \lesssim \kappa C.
\end{aligned}$$

If $T/n < 1$, set $I = 1$ and apply Theorem 5 to show that $\mathsf{M}_I(p, f^{\oplus n}) \le O(In)$. Next, apply Theorem 6 to find a protocol p' with $\mathsf{M}_I(p', f) \le O(I) = O(1)$. Finally, we apply Theorem 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 1 can be obtained from Theorem 11 using Equation (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 1 and 11 yield bounds on the success probability for computing f^n as well:

Corollary 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 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 [BRWY13b] 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

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

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 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 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 5 to show that $\mathsf{M}_I(p, f^{\oplus n}) \leq O(nI)$. Next, apply Theorem 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 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 5 to show that $\mathsf{M}_I(p, f^{\oplus n}) \leq O(n)$. Then, apply Theorem 6 to find a protocol p' with $\mathsf{M}_I(p', f) \leq O(I) = O(1)$. Lastly, we apply Theorem 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 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 [BRWY13b] using a different approach. For the bounded-round setting, we prove:

Theorem 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 [Yu22] 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 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 5. Applying Theorem 6 gives an r-round protocol with M_I bounded by $O(I)$, and applying Theorem 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 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 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 $\mathsf{suc}^r((\kappa C/r - \log r)n, f^n) < \exp(-\kappa n)$.*

In the rest of the paper, we prove Theorems 5 to 10. We prove Theorem 5 in Section 3, and Theorem 6 in Section 4. In Section 5 we gather several results related to the *trimming* technique borrowed from [Yu22] that are used in these first two steps. In Section 6 we gather several consequences of small marginal information that are used to analyze our compression schemes. We prove the general simulation for marginal information, Theorem 7, in Section 7. In Section 8 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 [BBCR10]. We then show how to compress smooth protocols to prove Theorem 8 in Section 9. We prove Theorem 10 in Section 10 and finally, in Section 11 we prove Theorem 9.

Acknowledgements

Thanks to Paul Beame, Makrand Sinha, Oscar Sprumont, Michael Whitmeyer and Amir Yehudayoff for helpful conversations.

2. Preliminaries

Throughout, we assume that $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $m \in \mathcal{M}$ for some finite sets $\mathcal{X}, \mathcal{Y}, \mathcal{M}$. Let $\mu(xy)$ be a distribution on pairs of inputs. To ease the notation, we often write ab instead of

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

Definition 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 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 2.

Let $x = x_1 x_2$ and $y = y_1 y_2$. Let $\mu(xy) = \mu(x_1 y_1) \cdot \mu(x_2 y_2)$ be a product distribution. It will be helpful to define $w = (x_1 y_2 m)$. Given $m = (m_0, \ldots, m_r)$ and y_2, we denote

$$\begin{aligned} m^{(1)} &= (m_0 y_2, m_1, m_2, \ldots, m_r), \\ m^{(2)} &= (m_0, x_1 m_1, m_2, \ldots, m_r). \end{aligned} \tag{5}$$

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

Proposition 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. $v(x_1 | y_1 m^{(1)}) \cdot v(x_2 | y_2 m^{(2)}) = v(x|ym)$, *and*
4. $v(y_1 | x_1 m^{(1)}) \cdot v(y_2 | x_2 m^{(2)}) = v(y|xm)$.

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

$$\begin{aligned} v(xy|w) = \frac{v(xyw)}{v(w)} &= \frac{\mu(x_1 y_1) \cdot \mu(x_2 y_2) \cdot A(xm) \cdot B(ym)}{\sum_{x_2' y_1'} \mu(x_1 y_1') \cdot \mu(x_2' y_2) \cdot A(x_1 x_2' m) \cdot B(y_1' y_2 m)} \\ &= \frac{\mu(x_1 y_1) \cdot B(y_1 y_2 m)}{\sum_{y_1'} \mu(x_1 y_1') \cdot B(y_1' y_2 m)} \cdot \frac{\mu(x_2 y_2) \cdot A(x_1 x_2 m)}{\sum_{x_2'} \mu(x_2' y_2) \cdot A(x_1 x_2' m)} \\ &= v(y_1|w) \cdot v(x_2|w). \end{aligned}$$

For the second identity,

$$\begin{aligned} v(xw) \cdot v(yw) &= v(w) \cdot v(yw) \cdot v(x|w) \\ &= v(w) \cdot v(yw) \cdot v(x_2 | x_1 ym) && \text{(by the first identity)} \\ &= v(w) \cdot v(xym). \end{aligned}$$

For the third identity,

$$\begin{aligned} v(x_1 | y_1 m^{(1)}) \cdot v(x_2 | y_2 m^{(2)}) &= v(x_1 | ym) \cdot v(x_2 | x_1 y_2 m) \\ &= v(x_1 | ym) \cdot v(x_2 | x_1 ym) && \text{(by the first identity)} \\ &= v(x|ym). \end{aligned}$$

A similar calculation yields the fourth identity. □

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

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

We recall the definition of divergence and some relevant facts from information theory.

Definition 22. *Given two distributions $a(u)$ and $b(u)$, the divergence of a from b is defined as*

$$\mathop{\mathbb{E}}_{a(u)} \left[\log \frac{a(u)}{b(u)} \right].$$

Proposition 23. *For two distributions $a(u)$ and $b(u)$, we have that*

$$\mathop{\mathbb{E}}_{a(u)} \left[\log \frac{a(u)}{b(u)} \right] \geq 0, \tag{6}$$

$$\sqrt{\mathop{\mathbb{E}}_{a(u)} \left[\log \frac{a(u)}{b(u)} \right]} \geq \frac{\|a(u) - b(u)\|_1}{2}. \tag{7}$$

The proof of this proposition can be found in [CT91].

The following version of a protocol due to [BR11] appears as Lemma 43 in [Yu22], the parameters we cite here are clear from the proof.

Lemma 24. *Let u, v denote two distributions on some finite set $\mathcal{M}$. For every $\varepsilon > 0$, there is a 1-round protocol distribution $\pi(uvs)$ (here uv correspond to the inputs of the protocol, and s corresponds to the transcript), and functions $a(us) \in \mathcal{M}, b(vs) \in \mathcal{M} \cup \{\perp\}$ (here $\perp \notin \mathcal{M}$), such that $\pi(uv)$ is supported on all pairs uv and for every uv and $z \in \mathcal{M}$,*

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

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

Additionally, we need the following Lemma, which appears as Lemma 4.14 in [BBCR10].

Lemma 25. *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, it 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.*

3. Marginal information of efficient protocols

In this section, we prove Theorem 5. Let R be a rectangular set that maximizes the quantity

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

We shall use trimming to prove the following claim:

Claim 26. *There exists a rectangular set $T \subseteq R$ with $p(T|R) \geq 1/4$ such that for any xym in the support of T, we have*

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

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

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

Observe that

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

and so by the choice of R,

$$p(Q|T) \geq \frac{1}{2} \cdot \mathop{\mathbb{E}}_{p(m|T)} \left| \mathop{\mathbb{E}}_{p(xy|mT)} \left[(-1)^f \right] \right| \geq \frac{\Omega(1)}{p(R)^{\delta}} \cdot \mathop{\mathbb{E}}_{p(m)} \left| \mathop{\mathbb{E}}_{p(xy|m)} \left[(-1)^f \right] \right|. \tag{8}$$

Define the rectangular distribution $q(xym) = p(xym|Q)$. By the definition of Q and Claim 26, we have that for all m in the support of q:

$$\left| \mathop{\mathbb{E}}_{p(xy|mQ)} \left[(-1)^f \right] \right| \geq \frac{1}{2} \cdot \mathop{\mathbb{E}}_{p(m|T)} \left| \mathop{\mathbb{E}}_{p(xy|mT)} \left[(-1)^f \right] \right| \geq \frac{\Omega(1)}{p(R)^{\delta}} \cdot \mathop{\mathbb{E}}_{p(m)} \left| \mathop{\mathbb{E}}_{p(xy|m)} \left[(-1)^f \right] \right|. \tag{9}$$

Using Claim 26 and Eqs. (8) and (9) and the definition of Q, we can bound the marginal information cost by

$$\begin{aligned} 2^{\mathsf{M}(p,f)} &\leq \sup_{xym} \left(\frac{p(x|ymQ)}{p(x|y)} \cdot \frac{p(y|xmQ)}{p(y|x)} \right) \cdot \left(\frac{p(xym|Q)}{p(xym)} \right)^I \cdot \left(\left| \mathop{\mathbb{E}}_{p(xy|mQ)} \left[(-1)^f \right] \right| \right)^{-12I/\delta} \\ &\leq \sup_{xym} \left(\frac{p(x|ymT)}{p(x|y)} \cdot \frac{p(y|xmT)}{p(y|x)} \right) \cdot \left(\frac{p(xym|T)}{p(xym) \cdot p(Q|T)} \right)^I \cdot \left(\left| \mathop{\mathbb{E}}_{p(xy|mQ)} \left[(-1)^f \right] \right| \right)^{-12I/\delta} \\ &\leq O(1) \cdot 2^{2C} \cdot p(R)^{-2} \cdot 2^{O(I)} \cdot p(R)^{-I(1-\delta)+12I} \cdot \left(\mathop{\mathbb{E}}_{p(m)} \left| \mathop{\mathbb{E}}_{p(xy|m)} \left[(-1)^f \right] \right| \right)^{-I(1+12/\delta)} \\ &\leq O(1) \cdot 2^{2C} \cdot 2^{O(I)} \cdot \left(\mathop{\mathbb{E}}_{p(m)} \left| \mathop{\mathbb{E}}_{p(xy|m)} \left[(-1)^f \right] \right| \right)^{-I(1+12/\delta)}, \end{aligned}$$

where in the last inequality we used the fact that $I(12+\delta-1)-2>0$ since $I \geq 1$. This completes the proof of the theorem.

It only remains to prove Claim 26. We have

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

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

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

Define the rectangular set

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

Markov's inequality implies that $p(G|R) \geq 1/2$. We apply Lemma 29 with $a(xym) = p(xym|G)$, $b(xym) = p(xym)$, and $\kappa = 1/6$ to obtain a rectangular set $T \subseteq G$ with $p(T|G) \geq 1/2$ and

$$\frac{p(xm|T)}{p(xm)}, \frac{p(ym|T)}{p(ym)} \geq \frac{1}{6}, \tag{10}$$

for all points in the support of T. We compute

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

Let us verify that T satisfies the remaining conditions promised by Claim 26. We have

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

To prove the second identity, use the first identity, the definition of G and Equation (10):

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

A similar calculation yields

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

Finally, applying Lemma 30 with $v(xym) = p(xym)$, $Z = T$, and noting that $p(Z|R) \geq 1/4$, we get

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

This completes the proof of Claim 26.

4. Marginal information is subadditive

In this section we prove Theorem 6. Recall the definitions of $m^{(1)}, m^{(2)}$, which are given in Equation (5). The core of the proof is the following statement.

Theorem 27. *Let $f(x_1y_1)$ and $g(x_2y_2)$ be two Boolean functions and let $p(xym)$ be a protocol distribution. Then, for every $1/3 \le \gamma \le 2/3$, 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*

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

We shall prove Theorem 6 assuming Theorem 27, whose proof we supply right after.

Proof of Theorem 6

Let $k_0 > 1$ be a large constant, to be determined. Define $f_i(x_iy_i) = f(x_iy_i)$. For $\ell = 0, 1, 2, \ldots, n$, define

$$k(\ell) = \max\left\{\inf_{\substack{S\subset[n],|S|=\ell\\ p'}}\mathsf{M}(p', \oplus_{i\in S}f_i), k_0I\right\},$$

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

$$T = \max\{k(n), k_0nI\}.$$

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

$$k(\ell) \le \frac{\ell T}{n} + 12I\cdot\log\frac{\ell T}{In}, \tag{11}$$

then set $\gamma = \lceil\ell/2\rceil/\ell$. Since $1/3 \le \gamma \le 2/3$, for k_0 chosen large enough, Theorem 27 shows that for some $\ell' \in \{\lceil\ell/2\rceil, \lfloor\ell/2\rfloor\}$, we have

$$\begin{aligned}k(\ell') &\le \max\{\frac{\ell'}{\ell}\cdot k(\ell) + 3I\log\frac{k(\ell)}{I}, k_0I\}\\ &\le \frac{\ell'}{\ell}\cdot\frac{\ell T}{n} + \frac{\ell'}{\ell}\cdot 12I\cdot\log\frac{\ell T}{In} + 3I\log\left(\frac{\ell T}{In} + 12\log\frac{\ell T}{In}\right)\\ &\qquad\qquad\text{(by the choice of } T \text{ and Equation (11))}\\ &\le \frac{\ell' T}{n} + 8I\cdot\log\frac{\ell T}{In} + 3I\log\left(13\cdot\frac{\ell T}{In}\right)\\ &= \frac{\ell' T}{n} + 11I\cdot\log\frac{\ell' T}{In} + 11I\cdot\log\frac{\ell}{\ell'} + 3I\log 13\\ &\le \frac{\ell' T}{n} + 11I\cdot\log\frac{\ell' T}{In} + 11I\cdot\log\frac{3}{2} + 3I\log 13\\ &\le \frac{\ell' T}{n} + 12I\cdot\log\frac{\ell' T}{In},\end{aligned}$$

for k_0 chosen large enough.

So, starting with $\ell = n$, we obtain a smaller and smaller ℓ satisfying Equation (11), until $\ell = 1$, which completes the proof.

Proof of Theorem 27

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

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

so $\mathsf{M}(p,h) = \inf_q \mathsf{M}(q,p,h)$. We shall prove that there are 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

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

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 preliminaries. 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 (13) to (16). 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

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

but the advantage terms do not add nicely in the marginal information cost: in Equation (16), 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

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

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 (18). This allows us to convert the advantage term in $\mathsf{M}(q,p,f\oplus g)$ into the kind of term where Equation (16) can be applied, and we use it to get subadditivity as described in Equation (19). 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 (20). 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

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

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 $2^{-\Omega(\mathsf{M}(q,p,f\oplus g)}$ (Claim 28). We then cluster the w and pass to a subset B of density $\Omega(\mathsf{M}(q,p,f\oplus g)^{-1})$ 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 $\log\mathsf{M}$ 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 (23)). All of these steps leave us with a subset of the inputs xym where the proof gives good control on the quantity $\mathsf{M}(q_1,p_1,f)$, but we now need to define a distribution r_1 supported on these points where $\mathsf{M}(r_1,p_1,f)$ 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

$$
\begin{aligned}
q_1(x_1y_1m^{(1)}) &= q(x_1y_1m^{(1)}) = q(yw)\\
q_2(x_2y_2m^{(2)}) &= q(x_2y_2m^{(2)}) = q(xw)\\
p_1(x_1y_1m^{(1)}) &= p(x_1y_1)\cdot q(m_0y_2)\cdot \prod_{i=1,3,5,\ldots} q(m_i|x_1y_2,m_{<i})\cdot p(m_{i+1}|ym_{\le i})\\
p_2(x_2y_2m^{(2)}) &= p(x_2y_2)\cdot p(m_0)\cdot q(x_1|m_0y_2)\cdot \prod_{i=1,3,5,\ldots} p(m_i|xm_{<i})\cdot q(m_{i+1}|x_1y_2m_{\le i})
\end{aligned}
$$

These distributions have been carefully chosen to have many nice properties. First, observe that $q_1(x_1y_1m^{(1)}), q_2(x_2y_2m^{(2)})$ are rectangular and $p_1(x_1y_1m^{(1)}), p_2(x_2y_2m^{(2)})$ are protocol distributions. Using the fact that p is a protocol and x_1y_1 and x_2y_2 are independent under p, we can compute:

$$
\begin{aligned}
p_1(x_1y_1m^{(1)})\cdot p_2(x_2y_2m^{(2)}) &= p(xy)\cdot p(m_0)\cdot q(x_1y_2m_0)\cdot \prod_{i>0} p(m_i|xym_{<i})\cdot q(m_i|x_1y_2m_{<i})\\
&= p(xym)\cdot q(x_1y_2m)\\
&= p(xym)\cdot q(w). \qquad (12)
\end{aligned}
$$

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:

$$
\frac{q_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})}\cdot\frac{q_2(x_2y_2m^{(2)})}{p_2(x_2y_2m^{(2)})} = \frac{q(xym)}{p(xym)}, \qquad (13)
$$

$$
\frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\cdot\frac{q_2(x_2|y_2m^{(2)})}{p_2(x_2|y_2)} = \frac{q(x|ym)}{p(x|y)}, \qquad (14)
$$

$$
\frac{q_1(y_1|x_1m^{(1)})}{p_1(y_1|x_1)}\cdot\frac{q_2(y_2|x_2m^{(2)})}{p_2(y_2|x_2)} = \frac{q(y|xm)}{p(y|x)}, \qquad (15)
$$

$$
\left|\mathop{\mathbb{E}}_{q_1(x_1y_1|w)}\left[(-1)^f\right]\right|\cdot\left|\mathop{\mathbb{E}}_{q_2(x_2y_2|w)}\left[(-1)^g\right]\right| = \left|\mathop{\mathbb{E}}_{q(xy|w)}\left[(-1)^{f\oplus g}\right]\right|. \qquad (16)
$$

To prove Equation (13), use Equation (12) and Proposition 20 to obtain

$$
\frac{q_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})}\cdot\frac{q_2(x_2y_2m^{(2)})}{p_2(x_2y_2m^{(2)})} = \frac{q(xym)}{p(xym)}\cdot\frac{q(w)}{q(w)} = \frac{q(xym)}{p(xym)}.
$$

Equations (14) and (15) follow directly from Proposition 20. We use the fact that $q(xy|w) = q(x_2|w)\cdot q(y_1|w)$ from Proposition 20 to prove Equation (16):

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

These identities suggest that the costs in the first and second coordinates should sum to $\mathsf{M}(q,p,f\oplus g)$. The main challenge in applying this intuition is that the advantage terms in Equation (16) are not the ones needed for $\mathsf{M}(q_1,p_1,f)$ and $\mathsf{M}(q_2,p_2,g)$. 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

$$q(G|m)^{\delta} \cdot \left| \mathop{\mathbb{E}}_{q(xy|mG)} \left[(-1)^{f \oplus g} \right] \right|, \tag{17}$$

among all such sets. In Lemma 31, we prove that for all w in the support of G:

$$\left| \mathop{\mathbb{E}}_{q(xy|w)} \left[(-1)^{f \oplus g} \right] \right| \geq (1-\delta) \cdot q(G|m)^{-\delta} \cdot \left| \mathop{\mathbb{E}}_{q(xy|m)} \left[(-1)^{f \oplus g} \right] \right|. \tag{18}$$

This gives us an effective way to split the costs for q, p. Using Equations (13) to (16), we obtain that for all xym in the support of G,

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

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:

$$\begin{aligned}
&\left| \mathop{\mathbb{E}}_{q_1(x_1y_1|w)} \left[(-1)^f \right] \right|^{-12I/\delta} \cdot \sup_{y_1} \left(\frac{q_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})} \right)^I \cdot \frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)} \frac{q_1(y_1|x_1m^{(1)})}{p_1(y_1|x_1)} \\
\times &\left| \mathop{\mathbb{E}}_{q_2(x_2y_2|w)} [(-1)^g] \right|^{-12I/\delta} \cdot \sup_{x_2} \left(\frac{q_2(x_2y_2m^{(2)})}{p_2(x_2y_2m^{(2)})} \right)^I \cdot \frac{q_2(x_2|y_2m^{(2)})}{p_2(x_2|y_2)} \frac{q_2(y_2|x_2m^{(2)})}{p_2(y_2|x_2)} \\
\leq &2^{\mathsf{M}(q,p,f \oplus g)} \cdot O(q(G|m))^{12I}.
\end{aligned} \tag{19}$$

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

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

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 (19), for all w in the support of $U \cap L$ we have

$$\left| \mathop{\mathbb{E}}_{q_1(x_1 y_1|w)} \left[(-1)^f\right] \right|^{-12I/\delta} \cdot \left(\sup_{y_1} \frac{q_1(x_1 y_1 m^{(1)})}{p_1(x_1 y_1 m^{(1)})} \right)^I \cdot \frac{q_1(x_1|y_1 m^{(1)})}{p_1(x_1|y_1)} \frac{q_1(y_1|x_1 m^{(1)})}{p_1(y_1|x_1)}$$
$$\leq 2^{\gamma \cdot \mathsf{M}(q,p,f\oplus g)} \cdot O(q(G|m))^{12\gamma \cdot I}. \tag{20}$$

Our next barrier is that in $\mathsf{M}(q_1, p_1, f)$ 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 $m^{(1)}$

$$\left| \mathop{\mathbb{E}}_{q_1(x_1 y_1|m^{(1)})} \left[(-1)^f\right] \right| \ll \left| \mathop{\mathbb{E}}_{q_1(x_1 y_1|w)} \left[(-1)^f\right] \right|. \tag{21}$$

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,

$$\left| \mathop{\mathbb{E}}_{q_1(x_1 y_1|w)} \left[(-1)^f\right] \right| \geq \frac{1}{2} \cdot \left| \mathop{\mathbb{E}}_{q_1(x_1 y_1|w')} \left[(-1)^f\right] \right|.$$

This will ensure that Equation (21) does not happen. To find this subset B, we first prune away some problematic points to ensure all advantage terms in Equation (20) are reasonably large. This is accomplished by Claim 28 below.

Claim 28. *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*

$$\left| \mathop{\mathbb{E}}_{q_1(x_1 y_1|w)} \left[(-1)^f\right] \right|^{-1} \leq \alpha,$$

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

We defer the proof of Claim 28 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 $\mathbb{E}_{q_1(x_1 y_1|w)}\left[(-1)^f\right]$, and the value of

$$\left\lceil \log \left| \mathop{\mathbb{E}}_{q_1(x_1 y_1|w)} \left[(-1)^f\right] \right| \right\rceil.$$

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 $1_B(x_1 y_1 m^{(1)})$ is determined by w, such that for every $m^{(1)}$,

$$q_1(B|m^{(1)} L') \geq \frac{1}{O(\log \alpha)}, \tag{22}$$

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

$$\left| \mathop{\mathbb{E}}_{q_1(x_1 y_1|m^{(1)} B)} \left[(-1)^f\right] \right| \geq \frac{1}{2} \cdot \left| \mathop{\mathbb{E}}_{q_1(x_1 y_1|w)} \left[(-1)^f\right] \right|. \tag{23}$$

Let $R \subseteq B \subseteq L'$ be the rectangular set in $x_1 y_1 m^{(1)}$ such that for every $m^{(1)}$, R maximizes

$$q_1(R|m^{(1)} B)^\delta \cdot \left| \mathop{\mathbb{E}}_{q_1(x_1 y_1|m^{(1)} R)} \left[(-1)^f\right] \right|. \tag{24}$$

Define the rectangular distribution

$$\begin{aligned} r(x_1y_1m^{(1)}) &= q_1(x_1y_1m^{(1)}) \cdot \frac{1_{U'}(m) \cdot 1_R(x_1y_1m^{(1)})}{q_1(U') \cdot q_1(L'|m) \cdot q_1(R|m^{(1)}L')} \\ &= \frac{q_1(m)1_{U'}(m)}{q_1(U')} \cdot \frac{q_1(y_2|m)q_1(L'|m^{(1)})}{q_1(L'|m)} \cdot \frac{q_1(x_1y_1|m^{(1)}) \cdot 1_R(x_1y_1m^{(1)})}{q_1(L'|m^{(1)}) \cdot q_1(R|m^{(1)}L')} \\ &= q_1(m|U') \cdot q_1(y_2|mL') \cdot \frac{q_1(x_1y_1|m^{(1)}) \cdot 1_R(x_1y_1m^{(1)})}{q_1(R|m^{(1)})} \\ &= q_1(m|U') \cdot q_1(y_2|mL') \cdot q_1(x_1y_1|m^{(1)}R). \end{aligned} \tag{25}$$

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 (25), it is clear that r is a distribution. We have the following bound:

$$\begin{aligned} \frac{r(x_1y_1m^{(1)})}{q_1(x_1y_1m^{(1)})} &= \frac{1_{U'}(m) \cdot 1_R(x_1y_1m^{(1)})}{q_1(U') \cdot q_1(L'|m) \cdot q_1(R|m^{(1)}L')} \\ &\leq \frac{O(1)}{q_1(L'|mG) \cdot q_1(G|m) \cdot q_1(R|m^{(1)}B) \cdot q_1(B|m^{(1)}L')} \\ &\leq \frac{O(\log \alpha)}{q_1(G|m) \cdot q_1(R|m^{(1)}B)}, \end{aligned} \tag{26}$$

where here we used Claim 28 and Equation (22). Apply Lemma 29 with $a = r$, $b = q_1$ and $\kappa = 1/6$ to obtain a rectangular set T with $r(T) \geq 1/2$ such that

$$\frac{r(x_1m^{(1)}|T)}{q_1(x_1m^{(1)})}, \frac{r(y_1m^{(1)}|T)}{q_1(y_1m^{(1)})}, \frac{r(m^{(1)}|T)}{r(m^{(1)})} \geq \frac{1}{6}. \tag{27}$$

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 $\mathsf{M}(r_1, p_1, f)$. For all $x_1y_1m^{(1)}$ in the support of r_1, we have

$$\begin{aligned} \frac{r_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})} &= \frac{q_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})} \cdot \frac{r(x_1y_1m^{(1)})}{q_1(x_1y_1m^{(1)})} \cdot \frac{1}{r(T)} \\ &\leq \frac{q_1(x_1y_1m^{(1)})}{p(x_1y_1m^{(1)})} \cdot \frac{O(\log \alpha)}{q_1(G|m) \cdot q_1(R|m^{(1)}B)}, \end{aligned} \tag{28}$$

using Equation (26) and the fact that $r(T) \geq 1/2$. For the next term,

$$\begin{aligned} \frac{r_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1m^{(1)})} &= \frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1m^{(1)})} \cdot \frac{r_1(x_1|y_1m^{(1)})}{q_1(x_1|y_1m^{(1)})} \\ &= \frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1m^{(1)})} \cdot \frac{r(x_1y_1m^{(1)})}{q_1(x_1y_1m^{(1)})} \cdot \frac{1}{r(T)} \cdot \frac{q_1(y_1m^{(1)})}{r(y_1m^{(1)}|T)} \\ &\leq \frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1m^{(1)})} \cdot \frac{O(\log \alpha)}{q_1(G|m) \cdot q_1(R|m^{(1)}B)}, \end{aligned} \tag{29}$$

using Equations (26) and (27), and the fact that $r(T) \geq 1/2$. The symmetric argument gives:

$$\frac{r_1(y_1|x_1m^{(1)})}{p_1(y_1|x_1m^{(1)})} \leq \frac{q_1(y_1|x_1m^{(1)})}{p_1(y_1|x_1m^{(1)})} \cdot \frac{O(\log \alpha)}{q_1(G|m) \cdot q_1(R|m^{(1)}B)}. \tag{30}$$

To bound the advantage, first note that

$$q_1(T|m^{(1)}R) = \frac{q_1(T|R)\cdot q_1(m^{(1)}|T)}{q_1(m^{(1)}|R)} = \frac{r(T)\cdot r(m^{(1)}|T)}{r(m^{(1)})} \geq \frac{1}{12}, \tag{31}$$

by Equation (27). For each $m^{(1)}$, we apply Lemma 30 with $v(x_1y_1m^{(1)}) = q_1(x_1y_1m^{(1)}|m^{(1)}B)$, R and $Z = T$. Note here that $v(m^{(1)}) = 1$. We obtain the bound:

$$\begin{aligned}
\left|\mathop{\mathbb{E}}_{r_1(x_1y_1|m^{(1)})}\left[(-1)^f\right]\right| &= \left|\mathop{\mathbb{E}}_{q_1(x_1y_1|m^{(1)}T)}\left[(-1)^f\right]\right| \\
&\geq \frac{1-\delta^2-\delta/q_1(T|m^{(1)}R)}{q_1(R|m^{(1)}B)^\delta}\cdot\left|\mathop{\mathbb{E}}_{q_1(x_1y_1|m^{(1)}B)}\left[(-1)^f\right]\right| \\
&\geq \frac{\Omega(1)}{q_1(R|m^{(1)}B)^\delta}\cdot\left|\mathop{\mathbb{E}}_{q_1(x_1y_1|m^{(1)}B)}\left[(-1)^f\right]\right|,
\end{aligned} \tag{32}$$

by the choice of δ and Equation (31).

Now, we are ready to put all these bounds together to complete the proof of the theorem. By Equation (23), Equation (28), Equation (29), Equation (30), Equation (32), we get that for every $x_1y_1m^{(1)}$ in the support of r_1,

$$\begin{aligned}
&\frac{r_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\cdot\frac{r_1(y_1|x_1m^{(1)})}{p(y_1|x_1)}\cdot\left(\frac{r_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})}\right)^I\cdot\left|\mathop{\mathbb{E}}_{r_1(x_1y_1|m^{(1)})}\left[(-1)^f\right]\right|^{-12I/\delta} \\
&\leq \frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\cdot\frac{q_1(y_1|x_1m^{(1)})}{p(y_1|x_1)}\cdot\left(\frac{q_1(x_1y_1m^{(1)})}{p_1(x_1y_1m^{(1)})}\right)^I\cdot\left|\mathop{\mathbb{E}}_{q_1(x_1y_1|w)}\left[(-1)^f\right]\right|^{-12I/\delta} \\
&\quad\times\frac{O(\log(\alpha))^{I+2}}{q_1(G|m)^{I+2}\cdot q_1(R|m^{(1)}B)^{I+2-12I}} \\
&\leq 2^{\gamma\cdot\mathsf{M}(q,p,f\oplus g)}\cdot O(\log(\alpha))^{3I}\cdot q_1(G|m)^{12\gamma\cdot I-I-2}\cdot 2^{O(I)}\cdot q_1(R|m^{(1)}B)^{11I-2} \\
&\leq 2^{\gamma\cdot\mathsf{M}(q,p,f\oplus g)}\cdot O(\log(\alpha))^{3I}\cdot q_1(G|m)^{3I-2}\cdot 2^{O(I)} \\
&\leq 2^{\gamma\cdot\mathsf{M}(q,p,f\oplus g)}\cdot O\left(\frac{\mathsf{M}(q,p,f\oplus g)}{I}\right)^{3I},
\end{aligned}$$

where in the last three inequalities we used Equation (20), the fact that $I \geq 1$ and Claim 28. This implies that

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

completing the proof of the theorem.

Proof of Claim 28: We have

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

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

$$\frac{q_1(m)}{p_1(m)} \leq 1/4 \tag{33}$$

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

$$q(U') \geq 1/2 - 1/4 = 1/4. \tag{34}$$

Next, we delete w from L if either

$$\frac{q_1(w|m)}{p_1(w|m)} < \frac{q_1(G|m)}{8}, \tag{35}$$

or

$$\mathop{\mathbb{E}}_{q_1(y_1|w)}\left[\log \frac{q_1(x_1|y_1m^{(1)})}{p_1(x_1|y_1)}\right] < \log \frac{q_1(G|m)}{8}. \tag{36}$$

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

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

so Markov's inequality implies that for each m the total mass of w for which Equation (35) is violated is at most $1/8$. By the concavity of the log function, the w deleted because of Equation (36) satisfy

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

On the other hand, because w determines 1_G,

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

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

$$q(L'|mG) \geq 1/2 - 1/8 - 1/8 = 1/4. \tag{37}$$

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

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

where here $\exp(z)$ denotes 2^z. Now we use the fact that all the m, w violating Equations (33),

(35) and (36) have been deleted and use Equation (6) to bound

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

Combining this bound with Equation (20), we get that for all w consistent with U', L',

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

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

$$\left|\mathop{\mathbb{E}}_{q_1(x_1y_1|w)}\left[(-1)^f\right]\right|^{-1}\leq O(2^{\mathsf{M}(q,p,f\oplus g)\cdot(\delta\gamma/12I)})=\alpha, \tag{38}$$

as required.

5. Trimming and advantage preserving sets

In this section, we gather a few lemmas about trimming rectangular sets to pass to sub-rectangles with nice features. The idea of trimming comes from the work of Yu [Yu22].

Lemma 29. *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

$$\frac{a(xm|T)}{b(xm)}<\kappa, \tag{39}$$

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 (39), 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 (39) 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. □

Next, we gather a couple of nice lemmas about finding subrectangles with nice properties.

Lemma 30. *For any distribution $v(xym)$ and a Boolean function $h(xy)$, suppose R is a rectangular set maximizing*

$$v(R)^{\delta} \cdot \mathop{\mathbb{E}}_{v(m|R)} \left| \mathop{\mathbb{E}}_{v(xy|mR)} \left[(-1)^h\right] \right|. \tag{40}$$

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

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

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

$$\mathop{\mathbb{E}}_{v(m|R)} \left| \mathop{\mathbb{E}}_{v(xy|Rm)} \left[(-1)^h\right] \right| \leq \sum_{i=0}^{2} v(Z_i|R) \cdot \mathop{\mathbb{E}}_{v(m|Z_i)} \left| \mathop{\mathbb{E}}_{v(xy|mZ_i)} \left[(-1)^h\right] \right| \tag{41}$$

Let us bound the contribution of Z_1, Z_2:

$$\begin{aligned}
&\sum_{i=1}^{2} v(Z_i|R) \cdot \mathop{\mathbb{E}}_{v(m|Z_i)} \left| \mathop{\mathbb{E}}_{v(xy|mZ_i)} \left[(-1)^h\right] \right| \\
&= \sum_{i=1}^{2} v(Z_i|R)^{1-\delta} \cdot \left(\frac{v(Z_i)}{v(R)}\right)^{\delta} \cdot \mathop{\mathbb{E}}_{v(m|Z_i)} \left| \mathop{\mathbb{E}}_{v(xy|mZ_i)} \left[(-1)^h\right] \right| \\
&\leq \sum_{i=1}^{2} v(Z_i|R)^{1-\delta} \cdot \mathop{\mathbb{E}}_{v(m|R)} \left| \mathop{\mathbb{E}}_{v(xy|mR)} \left[(-1)^h\right] \right| \quad \text{(because } R \text{ is the maximizer of Equation (40))} \\
&\leq 2^{\delta} \cdot \Big(\sum_{i=1}^{2} v(Z_i|R)\Big)^{1-\delta} \cdot \mathop{\mathbb{E}}_{v(m|R)} \left| \mathop{\mathbb{E}}_{v(xy|mR)} \left[(-1)^h\right] \right| \quad \text{(by Hölder's inequality)} \\
&= 2^{\delta} \cdot \Big(1 - v(Z|R)\Big)^{1-\delta} \cdot \mathop{\mathbb{E}}_{v(m|R)} \left| \mathop{\mathbb{E}}_{v(xy|mR)} \left[(-1)^h\right] \right|.
\end{aligned}$$

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

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

Putting this back into Equation (41) and reaarranging, we get:

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

where we again used the fact that R maximizes Equation (40). □

Lemma 31. *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*

$$q(G|m)^\delta \cdot \left| \mathop{\mathbb{E}}_{q(xy|mG)} \left[(-1)^{f\oplus g}\right]\right|, \tag{42}$$

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

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

Proof. Fix $w = x_1y_2m$ 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

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

where in the last line we used the fact that G is the maximizer of Equation (42).

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

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

Rearranging gives:

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

where in the second inequality we once again used the fact that G is the maximizer of Equation (42). □

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

$$\frac{q(xym)}{p(xym)} = \frac{\mu(xy)\cdot A(xm)\cdot B(ym)}{\mu(xy)\cdot p(m_0)\cdot \prod_{i=1,3,5,\dots} p(m_i|xm_{<i})\cdot p(m_{i+1}|ym_{\le i})} = g_1(xm)\cdot g_2(ym), \quad (43)$$

for appropriate functions g_1 and g_2.

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

$$S_K = \{xym : |\lceil \log g_1(xm)\rceil + \log g_2(ym)| \le 3(\mathsf{M}_I(p,f)+KI)/I\}, \quad (44)$$

$$R_K = \{xym : p(m_1|xm_0) \le 2^{6(\mathsf{M}_I(p,f)+KI)}\cdot p(m_1|ym_0)\}.. \quad (45)$$

Proposition 32. *For $xym \in S_K$,*

$$-\frac{3(\mathsf{M}_I(p,f)+KI)}{I} - 1 \le \log\frac{q(xym)}{p(xym)} \le \frac{3(\mathsf{M}_I(p,f)+KI)}{I}. \quad (46)$$

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

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

and

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

□

Claim 33. *If $K \ge 3$, $q(S_K^c), q(R_K^c|S_K) \le 5\cdot 2^{-(\mathsf{M}_I(p,f)+KI)/I}$.*

Proof. Define

$$G_1 = \{xym : q(x|ym) \ge 2^{-(\mathsf{M}_I(p,f)+KI)/I}\cdot p(x|y)\},$$
$$G_2 = \{xym : q(y|xm) \ge 2^{-(\mathsf{M}_I(p,f)+KI)/I}\cdot p(y|x)\},$$
$$G_3 = \{xym : q(xym) \ge 2^{-3(\mathsf{M}_I(p,f)+KI)/I}\cdot p(xym)\}, \text{ and}$$
$$G_4 = \{xym : q(x|ym) \ge 2^{-(\mathsf{M}_I(p,f)+KI)/I}\cdot p(x|m_0m_1y)\}.$$

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

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

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

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

Moreover, for $xym \in G_3$,

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

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 $xym \in G_2 \cap G_3 \cap G_4$,

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

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 34. *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 (43). For every K, define*

$$S_K = \{xym : |\lceil \log g_1(xm) \rceil + \log g_2(ym)| \leq 3(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I\} \text{ and} \tag{47}$$

$$R_K = \{xym : p(m_1|xm_0) \leq 2^{5(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)} \cdot p(m_1|m_0)\}. \tag{48}$$

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

Proof. Define

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

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

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

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

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

Moreover, for $xym \in G_2$

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

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 $q(S_K^c) \leq 2\cdot 2^{-(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I}$ as desired. To see the upper bound on $q(G_1^c)$, we may write

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

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 $q(R_K^c|S_K) \leq 5\cdot 2^{-(\mathsf{M}_I^{\mathsf{ext}}(p,f)+KI)/I}$. 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 $xym \in G_2 \cap G_3$,

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

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

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

since $K \geq 2$. □

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

Claim 35. *Let $K \geq 3$, and S_K be the set defined in Equation (44). Let $p(xym)$ be an r-round protocol and define*

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

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

Proof. Define the sets

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

We claim that

$$\bigcap_{j=1}^{3}(G_j \cap G_j') \cap G_4 \cap S_K \subseteq T_K. \tag{49}$$

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

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

becase $I \geq 1$, and by the definition of G_2' and Equation (46). Rearranging implies the first inequality below, and the second has a similar proof:

$$\frac{q(x|ym)}{p(x|y)}, \frac{q(y|xm)}{p(y|x)} \leq 2^{5(\mathsf{M}_I(p,f)+KI)/I} \cdot (r+1). \tag{50}$$

By Equation (50), for $xym \in \bigcap_{j=1}^{3}(G_j \cap G_j') \cap G_4 \cap S_K$ and all i,

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

Moreover,

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

where we used Equation (46) as well as the definitions of G_4, G_1' and G_2 in the last step. Thus,

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

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 (49) 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,

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

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

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

where we used Claim 33 and the fact that $K \ge 3$. □

Claim 36. *For any $K \ge 1$, let S_K be the set defined in Equation (44) and define*

$$T_K = \{xym : p(m|xy) \le 2^{6(\mathsf{M}_I(p,f)+KI)} \cdot \min\{p(m|x), p(m|y)\}\}. \tag{51}$$

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

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

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

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

$$\begin{aligned}
\mathsf{M}_I(p,f) &\geq \log\left(\frac{q(x|ym)}{p(x|y)}\cdot\frac{q(y|xm)}{p(y|x)}\cdot\left(\frac{q(xym)}{p(xym)}\right)^I\cdot\left|\mathop{\mathbb{E}}_{q(xy|m)}\left[(-1)^f\right]\right|^{-12I/\delta}\right)\\
&\geq \log\frac{q(x|ym)}{p(x|y)} - \frac{(\mathsf{M}_I(p,f)+KI)}{I} - 3(\mathsf{M}_I(p,f)+KI) - I\\
&\qquad\qquad\text{(by Equation (46) and definition of } G_2)\\
&\geq \log\frac{q(x|ym)}{p(x|ym)} + \log\frac{p(x|ym)}{p(x|y)} - 4(\mathsf{M}_I(p,f)+KI) - I \qquad\text{(since } I\geq 1)\\
&\geq -\frac{(\mathsf{M}_I(p,f)+KI)}{I} + \log\frac{p(m|xy)}{p(m|y)} - 4(\mathsf{M}_I(p,f)+KI) - I,
\end{aligned}$$

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 37.

$$\mathop{\mathbb{E}}_{q(xym)}\left[\sum_{i\geq 2}^{C}\|p(m_i|xm_{<i}) - p(m_i|ym_{<i})\|_1\right] \leq 8\sqrt{C\cdot\mathsf{M}_I(p,f)} \tag{52}$$

$$\mathop{\mathbb{E}}_{q(xym)}\left[\left|\mathop{\mathbb{E}}_{q(xy|m)}\left[(-1)^f\right]\right|\right] \geq 2^{-\delta\mathsf{M}_I(p,f)/12I}. \tag{53}$$

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

$$\begin{aligned}
\mathsf{M}_I(p,f) &= \max_{xym\in\mathsf{supp}(q)}\log\left(\frac{q(x|ym)}{p(x|y)}\cdot\frac{q(y|xm)}{p(y|x)}\cdot\left(\frac{q(xym)}{p(xym)}\right)^I\cdot\left|\mathop{\mathbb{E}}_{q(xy|m)}\left[(-1)^f\right]\right|^{-12I/\delta}\right)\\
&\geq \mathop{\mathbb{E}}_{q(xym)}\left[\log\frac{q(x|ym)}{p(x|y)}\right] + \mathop{\mathbb{E}}_{q(xym)}\left[\log\frac{q(y|xm)}{p(y|x)}\right]\\
&\quad + I\cdot\mathop{\mathbb{E}}_{q(xym)}\left[\log\frac{q(xym)}{p(xym)}\right] + \mathop{\mathbb{E}}_{q(xym)}\left[\log\left|\mathop{\mathbb{E}}_{q(xy|m)}\left[(-1)^f\right]\right|^{-12I/\delta}\right]
\end{aligned}\tag{54}$$

By Equation (6) 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

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

thus giving Equation (53). For Equation (52), we have

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{q(xym)}\left[\sum_{i\geq 2}^{C}\|p(m_i|xm_{<i})-p(m_i|ym_{<i})\|_1\right]\\
&\leq \mathop{\mathbb{E}}_{q(xym)}\left[\sum_i \|p(m_i|xm_{<i})-q(m_i|xym_{<i})\|_1+\|q(m_i|xym_{<i})-p(m_i|ym_{<i})\|_1\right]\\
&\leq 2\mathop{\mathbb{E}}_{q(xym)}\left[\sum_i\sqrt{\mathop{\mathbb{E}}_{q(m_i|xym_{<i})}\left[\log\frac{q(m_i|xym_{<i})}{p(m_i|xm_{<i})}\right]}+\sqrt{\mathop{\mathbb{E}}_{q(m_i|xym_{<i})}\left[\log\frac{q(m_i|xym_{<i})}{p(m_i|ym_{<i})}\right]}\right]\\
&\qquad\qquad\text{(by Equation (7))}\\
&\leq 2\sqrt{C\cdot\mathop{\mathbb{E}}_{q(xym)}\left[\sum_i\log\frac{q(m_i|xym_{<i})}{p(m_i|xm_{<i})}\right]}+2\sqrt{C\cdot\mathop{\mathbb{E}}_{q(xym)}\left[\sum_i\log\frac{q(m_i|xym_{<i})}{p(m_i|ym_{<i})}\right]}\\
&\qquad\qquad\text{(by concavity of }\sqrt{\cdot}\text{)}\\
&= 2\sqrt{C\cdot\mathop{\mathbb{E}}_{q(xym)}\left[\log\frac{q(m|xy)}{p(m|x)}\right]}+2\sqrt{C\cdot\mathop{\mathbb{E}}_{q(xym)}\left[\log\frac{q(m|xy)}{p(m|y)}\right]}.
\end{aligned}
$$

To complete the proof, we claim that

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

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,

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

where in the first inequality, we used the concavity of $\log(\cdot)$ and in the last one, we used Equation (54). □

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

Lemma 38. *Let q be a distribution achieving $\mathsf{M}^{\mathsf{ext}}_I(p,f)$. Then,*

$$\mathop{\mathbb{E}}_{q(xym)}\left[\log\frac{p(m|xy)}{p(m)}\right] \le \mathsf{M}^{\mathsf{ext}}_I(p,f), \tag{55}$$

$$\mathop{\mathbb{E}}_{q(xym)}\left[\left|\mathop{\mathbb{E}}_{q(xy|m)}\left[(-1)^f\right]\right|\right] \ge 2^{-\delta\mathsf{M}^{\mathsf{ext}}_I(p,f)/(12I)}. \tag{56}$$

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

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

By Equation (6) 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

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

thus giving Equation (56). Moreover,

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

and this implies Equation (55) since the first term in the sum is non-negative. □

7. Compressing marginal information

Here we prove Theorem 7. 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 $\mathsf{M}(p,f) \ge 0$, we have $\mathsf{M} \ge KI$. Let g_1, g_2 be as in Equation (43).

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 $h : \mathbb{Z} \to \{1,\dots,\lceil 1/\varepsilon\rceil\}$.
2. Run the protocol π from Lemma 24 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 $m_1^A = a(\pi(us))$, Bob sets $m_1^B = b(\pi(vs))$. 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

$$m_i^A = \begin{cases} 1 & \text{if } \rho_i \leq p(m_i = 1 | x m_{<i}^A), \\ 0 & \text{otherwise.} \end{cases} \tag{57}$$

$$m_i^B = \begin{cases} 1 & \text{if } \rho_i \leq p(m_i = 1 | y m_{<i}^B), \\ 0 & \text{otherwise.} \end{cases}, \tag{58}$$

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

4. Run τ from Lemma 25 to find the smallest j with $m_j^A \neq m_j^B$. 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 (57). 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 (58). The players repeat this process at most $2^{7\mathsf{M}/I}\sqrt{C\mathsf{M}}$ times. If by this point τ reports that $m^A \neq m^B$, 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 $\eta^A \leq g_1(x\langle m^A \rangle) \cdot 2^{-\lceil \log g_1(x\langle m^A \rangle) \rceil}$, 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

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

Bob sends $\mathsf{sign}\Big(\mathbb{E}_{q(xy|\langle m^B \rangle)}[(-1)^f] \Big) \in \{\pm 1\}$ 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

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

when $i > 1$ is even, and setting

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

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

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

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

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

Let S and R be the sets defined in Equations (44) and (45) for our choice of K. In addition to S and R, we need the following sets to analyze the simulating protocol:

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

$$\mathcal{E} = \{ \langle m^A \rangle \langle m^B \rangle m : \langle m^A \rangle = \langle m^B \rangle = m \},$$

$$\mathcal{Z} = \{ xymh : \exists \text{ a unique integer } z \text{ with } |z + \log g_2(ym)| \leq 3\mathsf{M}/I \text{ and } h(z) = h(\lceil \log g_1(xm) \rceil) \}.$$

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

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

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:

$$\Gamma(\mathcal{EZ}SQ) \cdot \mathop{\mathbb{E}}_{\Gamma(xym|\mathcal{EZ}SQ)} [\mathcal{A}(xym)] - \Gamma(\mathcal{G}(\mathcal{EZ}SQ)^c). \tag{59}$$

We shall prove:

$$\mathop{\mathbb{E}}_{\Gamma(xym|\mathcal{EZ}QS)} [\mathcal{A}(xym)] \geq \Omega(2^{-\delta \mathsf{M}/(12I)}), \tag{60}$$

$$\Gamma(\mathcal{EZ}QS) \geq \Omega(2^{-3\mathsf{M}/I}), \tag{61}$$

$$\Gamma(\mathcal{G}(\mathcal{EZ}SQ)^c) \leq O(2^{-4\mathsf{M}/I}). \tag{62}$$

By Equation (59), 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:

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

The definition of the protocol ensures that $\Gamma(\mathcal{Z}^c|\mathcal{G}\mathcal{E}) = 0$. Moreover, we claim that $\Gamma(Q^c|\mathcal{G}\mathcal{E}\mathcal{Z}S) = 0$, because if the event $\mathcal{E}\mathcal{Z}S$ 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 π 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 (62).

Now, we turn to proving Equation (61). 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 $xym \in S$,

$$\Gamma(Q|xym) = \frac{g_1(xm) \cdot 2^{-\lceil \log g_1(xm) \rceil} \cdot g_2(ym) \cdot 2^{\lceil \log g_1(xm) \rceil}}{2^{3\mathsf{M}/I}} = \frac{q(xym)}{p(xym)} \cdot \frac{1}{2^{3\mathsf{M}/I}}, \tag{63}$$

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,

$$\Gamma(QS) = \sum_{xym \in S} p(xym) \cdot \frac{q(xym)}{p(xym)} \cdot \frac{1}{2^{3\mathsf{M}/I}} = \frac{q(S)}{2^{3\mathsf{M}/I}}$$
$$\geq \frac{(1 - 5 \cdot 2^{-\mathsf{M}/I})}{2^{3\mathsf{M}/I}} = \Omega(2^{-3\mathsf{M}/I}), \tag{64}$$

where in the last line, we used Claim 33.

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

$$\Gamma(\mathcal{Z}|xymQS) = \Gamma(\mathcal{Z}|xym) \geq 1 - O(\varepsilon\mathsf{M}/I). \tag{65}$$

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

$$\Gamma(QS\mathcal{Z}) = \Gamma(QS) \cdot \Gamma(\mathcal{Z}|QS) \geq \Omega(2^{-3\mathsf{M}/I}), \tag{66}$$

For all $xym \in S$,

$$\begin{aligned}
\Gamma(xym|QS\mathcal{Z}) &= \frac{\Gamma(xym) \cdot \Gamma(QS\mathcal{Z}|xym)}{\Gamma(QS\mathcal{Z})} \\
&= \frac{p(xym)}{\Gamma(QS)} \cdot \Gamma(Q|xym) \cdot \frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} \\
&= \frac{p(xym)}{\Gamma(QS)} \cdot \frac{q(xym)}{p(xym) \cdot 2^{3\mathsf{M}/I}} \cdot \frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} && \text{(By Equation (63))} \\
&= \frac{q(xym)}{q(S)} \cdot \frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} && \text{(By Equation (64))} \\
&= q(xym|S) \cdot (1 \pm O(\varepsilon\mathsf{M}/I)), && (67)
\end{aligned}$$

where the last line follows by Equation (65).

Given Equation (66), to complete the proof of Equation (61), it will be enough to prove that $\Gamma(\mathcal{E}|QS\mathcal{Z}) \geq 1/2$. We shall prove that

$$\begin{aligned}
&\Gamma(\mathcal{E}^c|QS\mathcal{Z}) \\
&\leq \Gamma(R^c|QSZ) + \Gamma\left(\mathcal{E}^c \middle| QS\mathcal{Z}R, \sum_{i=2}^{C} E_i \leq 2^{7\mathsf{M}/I} \cdot \sqrt{C\mathsf{M}}\right) + \Gamma\left(\sum_{i=2}^{C} E_i > 2^{7\mathsf{M}/I} \cdot \sqrt{C\mathsf{M}} \middle| QS\mathcal{Z}\right) \\
&\leq O(2^{-\mathsf{M}/I}). && (68)
\end{aligned}$$

By Equation (67) and Claim 33,

$$\Gamma(R^c|QS\mathcal{Z}) \leq q(R^c|S)(1 + O(\varepsilon\mathsf{M}/I)) \leq 2^{-\mathsf{M}/I+3}. \tag{69}$$

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 π make an error that leaves Alice and Bob with inconsistent messages, or if π aborts. We claim that the probability that π 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 24. 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

$$\Gamma\left(\mathcal{E}^c \middle| QSZR, \sum_{i=2}^{C} E_i \le 2^{7\mathsf{M}/I}\cdot\sqrt{C\mathsf{M}}\right) \le O(\varepsilon 2^{7\mathsf{M}/I}\sqrt{C\mathsf{M}}). \tag{70}$$

We shall prove at the end of this section that

$$\Gamma\left(\sum_{i=2}^{C} E_i > 2^{7\mathsf{M}/I}\cdot\sqrt{C\mathsf{M}} \middle| QSZ\right) < O(2^{-\mathsf{M}/I}). \tag{71}$$

Equations (69) to (71) together prove Equation (68), and so conclude the proof of Equation (61). Next, we prove Equation (60). Since $|\mathcal{A}(xym)| \le 1$, we have

$$\mathop{\mathbb{E}}_{\Gamma(xym|QSZ)}[\mathcal{A}(xym)] \le \Gamma(\mathcal{E}|QSZ)\cdot \mathop{\mathbb{E}}_{\Gamma(xym|QSZ\mathcal{E})}[\mathcal{A}(xym)] + \Gamma(\mathcal{E}^c|QSZ),$$

and since $\Gamma(\mathcal{E}|QSZ) \le 1$, this gives

$$\begin{aligned}
\mathop{\mathbb{E}}_{\Gamma(xym|QSZ\mathcal{E})}[\mathcal{A}(xym)] &\ge \mathop{\mathbb{E}}_{\Gamma(xym|QSZ)}[\mathcal{A}(xym)] - \Gamma(\mathcal{E}^c|QSZ) \\
&\ge \mathop{\mathbb{E}}_{q(xym|S)}[\mathcal{A}(xym)] - O(\varepsilon\mathsf{M}/I) - O(2^{-\mathsf{M}/I}) \\
&\qquad\qquad \text{(using Equations (67) and (68))} \\
&\ge \mathop{\mathbb{E}}_{q(xym)}[\mathcal{A}(xym)] - O(2^{-\mathsf{M}/I}) \qquad \text{(by Claim 33)} \\
&= \mathop{\mathbb{E}}_{q(xym)}\left[\mathsf{sign}\left(\mathop{\mathbb{E}}_{q(x'y'|m)}\left[(-1)^f\right]\right)\cdot(-1)^{f(xy)}\right] - O(2^{-\mathsf{M}/I}) \\
&= \mathop{\mathbb{E}}_{q(m)}\left[\left|\mathop{\mathbb{E}}_{q(xy|m)}\left[(-1)^{f(xy)}\right]\right|\right] - O(2^{-\mathsf{M}/I}) \ge \Omega(2^{-\delta\mathsf{M}/(12I)}),
\end{aligned}$$

by Equation (53). This completes the proof of Equation (60).

It only remains to prove Equation (71). Define the function

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

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

We have

$$\Gamma\left(\sum_{i=2}^{C} E_i > 2^{7\mathsf{M}/I}\cdot\sqrt{C\mathsf{M}} \middle| QSZ\right) \le \Gamma(t \ne \perp|QSZ) + \Gamma\left(\sum_{i=2}^{C} E_i > 2^{7\mathsf{M}/I}\cdot\sqrt{C\mathsf{M}} \middle| QSZ, t = \perp\right),$$

so let us bound each of these terms.

$$\begin{aligned}
\Gamma(t \ne \perp|QSZ) &\le q(t \ne \perp|S)\cdot(1 + O(\varepsilon\mathsf{M}/I)) \qquad \text{(by Equation (67))} \\
&\le q(t \ne \perp)\cdot(1 + O(2^{-3\mathsf{M}/I}))\cdot(1 + O(\varepsilon\mathsf{M}/I)), \qquad \text{(by Claim 33)}
\end{aligned}$$

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

$$\Gamma(t \neq \perp | QS\mathcal{Z}) \leq O(2^{-3\mathsf{M}/I}). \tag{72}$$

Next, we show that

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

which would complete the proof of Equation (71). This follows from Markov's inequality and the bound

$$\mathop{\mathbb{E}}_{\Gamma}\Big[\sum_{i=2}^{C} E_i \Big| QS\mathcal{Z}, t=\perp \Big] \leq O(2^{6\mathsf{M}/I}\sqrt{C\mathsf{M}}), \tag{73}$$

which we prove next. We have:

$$\mathop{\mathbb{E}}_{\Gamma}\Big[\sum_{i=2}^{C} E_i \Big| QS\mathcal{Z}, t=\perp \Big] = \sum_{i=2}^{C} \frac{\Gamma(E_i=1, QS\mathcal{Z}, t=\perp)}{\Gamma(QS\mathcal{Z}, t=\perp)} \tag{74}$$
$$\leq O(2^{3\mathsf{M}/I}) \cdot \sum_{i=2}^{C} \Gamma(E_i=1, t=\perp). \quad \text{(by Equations (66) and (72))}$$

Moreover,

$$\Gamma(E_i=1, t=\perp) = \sum_{xym_{<i}} \Gamma(xym_{<i}) \cdot \Gamma(E_i=1|xym_{<i}) \cdot \Gamma(t=\perp|E_i=1, xym_{<i})$$
$$= \sum_{xym_{<i}} p(xym_{<i}) \cdot \|p(m_i|xm_{<i}) - p(m_i|ym_{<i})\|_1 \cdot \Gamma(t=\perp|E_i=1, xym_{<i})$$
$$\leq O(2^{3\mathsf{M}/I}) \cdot \sum_{xym_{<i}} q(xym_{<i}) \cdot \|p(m_i|xm_{<i}) - p(m_i|ym_{<i})\|_1 \tag{75}$$

Therefore,

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

by Equation (52), which completes the proof of Equation (73).

8. Smoothing protocols

A smooth protocol is a protocol where each message bit is close to being uniformly distributed:

Definition 39. *Given a protocol distribution $p(xym)$ with C messages satisfying $m_2, \ldots, m_C \in \{0,1\}$, 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 40. *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. 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 $\{0,1\}^L$:

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

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

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

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

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

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 $i > 1$,

$$p'(m_im'_i|xym_{<i}m'_{<i}) = p(m_i|xy, m_{<i} = D(m')_{<i}) \cdot s_{m_i}(m'_i)$$

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,

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

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

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

Thus, we get that

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

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 $j \ge 2$, define the j-th divergence costs:

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

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

$$\begin{aligned} d_{j+1}^A(xm) - d_j^A(xm) &\le \log \frac{1/2+\beta}{1/2-\beta} \le 5\beta, \\ d_{j+1}^B(ym) - d_j^B(ym) &\le \log \frac{1/2+\beta}{1/2-\beta} \le 5\beta. \end{aligned} \tag{76}$$

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 $r(xym) \ne r(xym')$, 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

$$\begin{aligned} F_{r,\alpha} &= \Big\{ xym : \Big| \sum_{i \ge 2}^{r(xym)} \log \frac{p(m_i|xym_{<i})}{p(m_i|m_{<i})} - d_{r(xym)}(xym) \Big| \ge \alpha \Big\}, \\ F_{r,\alpha}^A &= \Big\{ xym : \Big| \sum_{i \ge 2 \text{ odd}}^{r(xym)} \log \frac{p(m_i|xym_{<i})}{p(m_i|m_{\le i})} - d_{r(xym)}^A(xm) \Big| \ge \alpha \Big\}, \\ F_{r,\alpha}^B &= \Big\{ xym : \Big| \sum_{i \ge 2 \text{ even}}^{r(xym)} \log \frac{p(m_i|xym_{<i})}{p(m_i|m_{\le i})} - d_{r(xym)}^B(ym) \Big| \ge \alpha \Big\}. \end{aligned} \tag{77}$$

Lemma 41. *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 $t_i = z_i - \mathbb{E}_{p(m_i|xym_{<i})}[z_i]$. Then by definition $\mathbb{E}[t_i|t_{<i}] = 0$. 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

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

It is well known that if $\mathbb{E}[t_i] = 0$, then $\mathbb{E}[\exp(t_i)] \le \exp((\sup(t_i) - \inf(t_i))^2/8)$ (see Lemma 2.6 in [JHM+98]). We can use this inequality to bound:

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

So by Markov's inequality, we get:

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

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. □

9. Compressing external marginal information

Here we prove Theorem 8. Set $\mathsf{M}^{\mathsf{ext}} = \mathsf{M}_I^{\mathsf{ext}}(p,f) + KI$, for a large constant K to be chosen later. By Theorem 40, it is no loss of generality to assume that p is β-smooth, with $\beta = 1/(K\log(C2^{5\mathsf{M}^{\mathsf{ext}}/I}))$. Let g_1, g_2 be as in Equation (43).

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

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

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 π from Lemma 24 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 $m_1^A = a(\pi(us))$, Bob sets $m_1^B = b(\pi(vs))$. 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 $r_{\ell+1}^A(x\widetilde{m})$ and $r_{\ell+1}^B(y\widetilde{m})$ to determine
   $$k = \min\{r_{\ell+1}^A(x\widetilde{m}), r_{\ell+1}^B(y\widetilde{m})\}.$$
   (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 $m_{\le k} \leftarrow \widetilde{m}_{\le k}$.

4. If $\eta^A \le g_1(xm) \cdot 2^{-\lceil \log g_1(xm) \rceil}$, 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

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

Bob sends $\mathsf{sign}\Big(\mathbb{E}_{q(xy|m)}[(-1)^f]\Big) \in \{\pm 1\}$ 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

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

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

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 (47) and (48) for our choice of K. In addition to S and R, we need the following sets to analyze the simulating protocol:

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

$$\mathcal{Z} = \Big\{xymh : \exists \text{ unique integer } z \text{ with } |z + \log g_2(ym)| \le \frac{3\mathsf{M}^{\mathsf{ext}}}{I} \text{ and } h(z) = h(\lceil \log g_1(xm) \rceil)\Big\}.$$

Let $\mathcal{G}$ denote the event that the protocol reaches the final step without aborting and 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

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

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

$$\Gamma(\mathcal{Z}QS\mathcal{G}) \cdot \mathop{\mathbb{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QS\mathcal{G}] - \Gamma(\mathcal{G}(\mathcal{Z}QS)^c) \tag{78}$$

We shall prove:

$$\mathop{\mathbb{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QS\mathcal{G}] \ge \Omega(2^{-\delta\mathsf{M}^{\mathsf{ext}}/(12I)}), \tag{79}$$

$$\Gamma(\mathcal{Z}QS\mathcal{G}) \ge \Omega(2^{-3\mathsf{M}^{\mathsf{ext}}/I}), \tag{80}$$

$$\Gamma(\mathcal{G}(\mathcal{Z}QS)^c) \le O(2^{-4\mathsf{M}^{\mathsf{ext}}/I}). \tag{81}$$

By Equation (78), since $\delta \le 1$, we can choose K to be large enough to prove the theorem, since $\alpha + K \ge \mathsf{M}^{\mathsf{ext}}/I \ge K$.

We first prove Equation (81). By the union bound, we have:

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

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:

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

Additionally, $\Gamma(S^c\mathcal{G}\mathcal{Z}) \le \Gamma(S^c\mathcal{Z}) \le O(\varepsilon\mathsf{M}/I)$, 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 (79) and (80), 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 $m_{\le \ell}$ has been fixed. For m consistent with this prefix $m_{\le \ell}$, 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 $m_{\le r(xym)}$ on the frontier. By the definition of $r^A_{\ell+1}, r^B_{\ell+1}$, and by Equation (76), we have that for all m, $d_{r(xym)}(m) - d_\ell(m) \le 45\beta$. Setting $\tau = 45\beta$ and $\alpha = 1/4$, we apply Lemma 41 to conclude that if

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

then

$$p(F^A \cup F^B | xym_{\le \ell}) \le 4\exp(-\Omega(1/\beta)) \le C^{-1} \cdot 2^{-5\mathsf{M}^{\mathsf{ext}}/I}. \tag{82}$$

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 $xym_{\le \ell}$, the probability that W occurs is

$$\begin{aligned}\Gamma(W | xym_{\le \ell}) &\ge \sum_{m'_{r(xym')} : xym'_{r(xym)} \notin F^A \cup F^B} p(m'_{\le r(xym')} | xym_{\le \ell})/4 \\ &\ge 1/4 - p(F^A \cup F^B | xym_{\le \ell})/4 \ge 1/4 - C^{-1} \cdot 2^{-5\mathsf{M}^{\mathsf{ext}}/I} \ge 1/8,\end{aligned} \tag{83}$$

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

$$\frac{p(m_{\le r(xym)} | xym_{\le \ell})/4}{1/4 - p(F^A \cup F^B | xym_{\le \ell})/4} = p(m_{\le r(xym)} | xym_{\le \ell}) \cdot (1 \pm O(C^{-1} \cdot 2^{-5\mathsf{M}^{\mathsf{ext}}/I})). \tag{84}$$

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 (82),

$$p(B) \leq O(2^{-5\mathsf{M}^{\mathsf{ext}}/I}). \tag{85}$$

Moreover, by Equation (84), for $xym \notin B$,

$$\Gamma(xym) = p(xym) \cdot (1 \pm O(2^{-5\mathsf{M}^{\mathsf{ext}}/I})). \tag{86}$$

Equations (85) and (86) imply

$$\Gamma(B) = 1 - \Gamma(B^c) \leq 1 - p(B^c) \cdot (1 - O(2^{-5\mathsf{M}^{\mathsf{ext}}/I})) \leq O(2^{-5\mathsf{M}^{\mathsf{ext}}/I}). \tag{87}$$

Additionally, we have

$$q(SB^c) = q(S) - q(BS) \geq q(S) - 2^{3\mathsf{M}^{\mathsf{ext}}/I} \cdot p(B) \geq 1 - \Omega(2^{-\mathsf{M}/I}), \tag{88}$$

by the definition of S, Claim 33 and Eq. (85).

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

$$\Gamma(Q|xym) = \frac{g_1(xm) \cdot 2^{-\lceil \log g_1(xm) \rceil} \cdot g_2(ym) \cdot 2^{\lceil \log g_1(xm) \rceil}}{2^{3\mathsf{M}^{\mathsf{ext}}/I}} = \frac{q(xym)}{p(xym)} \cdot \frac{1}{2^{3\mathsf{M}^{\mathsf{ext}}/I}}, \tag{89}$$

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

$$\begin{aligned}
\Gamma(QSB^c) &= \sum_{xym \in S \cap B^c} \Gamma(xym) \cdot \Gamma(Q|xym) \\
&= \sum_{xym \in S \cap B^c} p(xym) \cdot \Gamma(Q|xym) \cdot (1 \pm O(2^{-5\mathsf{M}^{\mathsf{ext}}/I})). && \text{(by Equation (86))} \\
&= 2^{-3\mathsf{M}^{\mathsf{ext}}/I} \cdot q(SB^c) \cdot (1 \pm O(2^{-5\mathsf{M}^{\mathsf{ext}}/I})) && (90) \\
&= \Omega(2^{-3\mathsf{M}^{\mathsf{ext}}/I}), && (91)
\end{aligned}$$

by Equations (85) and (88). We claim that for all $xym \in SB^c$,

$$\Gamma(\mathcal{Z}|xymQSB^c) = \Gamma(\mathcal{Z}|xym) \geq 1 - O(\varepsilon \mathsf{M}^{\mathsf{ext}}/I). \tag{92}$$

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

$$\Gamma(\mathcal{Z}|QS) \geq 1 - O(\varepsilon \mathsf{M}^{\mathsf{ext}}/I). \tag{93}$$

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

$$\begin{aligned}
\Gamma(xym|\mathcal{Z}QSB^c) &= \frac{\Gamma(xym)\cdot\Gamma(\mathcal{Z}QSB^c|xym)}{\Gamma(\mathcal{Z}QSB^c)} \\
&= \frac{p(xym)\cdot\Gamma(Q|xym)\cdot\Gamma(\mathcal{Z}|xym)}{\Gamma(\mathcal{Z}QSB^c)}\cdot(1\pm O(2^{-5\mathsf{M}^{\mathsf{ext}}/I})) \\
&\qquad\qquad \text{(by Equation (86), and since } xym \text{ determine } S, B^c) \\
&= \frac{p(xym)}{\Gamma(QSB^c)}\cdot\frac{q(xym)}{p(xym)\cdot 2^{3\mathsf{M}^{\mathsf{ext}}/I}}\cdot\frac{\Gamma(\mathcal{Z}|xym)}{\Gamma(\mathcal{Z}|QSB^c)}\cdot(1\pm O(2^{-5\mathsf{M}^{\mathsf{ext}}/I})) \\
&\qquad\qquad \text{(By Equation (89))} \\
&= \frac{q(xym)}{q(SB^c)}\cdot(1\pm O(\varepsilon\mathsf{M}^{\mathsf{ext}}/I + 2^{-5\mathsf{M}^{\mathsf{ext}}/I})). \quad \text{(By Equations (90) and (92))} \\
&= q(xym|SB^c)\cdot(1\pm O(\varepsilon\mathsf{M}^{\mathsf{ext}}/I + 2^{-5\mathsf{M}^{\mathsf{ext}}/I})). \qquad (94)
\end{aligned}$$

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

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

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 (77). By Equation (76), and the the choice of r, $d_{r(xym)}(xym) \le 2\mathsf{M}^{\mathsf{ext}}\cdot 2^{10\mathsf{M}^{\mathsf{ext}}/I} + 5\beta$, so we can apply Lemma 41 to conclude that

$$p(FH) \le p(F_{r,\mathsf{M}^{\mathsf{ext}}\cdot 2^{10\mathsf{M}^{\mathsf{ext}}/I}}) \le 2\exp(-\Omega(\mathsf{M}^{\mathsf{ext}}\cdot 2^{10\mathsf{M}^{\mathsf{ext}}/I})). \qquad (95)$$

We have

$$\begin{aligned}
q(F|SB^c) &\le \frac{q(FS)}{q(SB^c)} \\
&\le \frac{q(FHS)+q(H^c)}{q(SB^c)} \\
&\le O(q(FHS)+q(H^c)) \qquad \text{(by Equation (88))} \\
&\le O(q(FHS)+2^{-10\mathsf{M}^{\mathsf{ext}}/I}) \qquad \text{(by Markov's inequality and Equation (55))} \\
&\le O(p(FHS)\cdot 2^{3\mathsf{M}^{\mathsf{ext}}/I} + 2^{-10\mathsf{M}^{\mathsf{ext}}/I}) \qquad \text{(using the definition of } S) \\
&\le O(\exp(-\Omega(\mathsf{M}^{\mathsf{ext}}\cdot 2^{10\mathsf{M}^{\mathsf{ext}}/I}))\cdot 2^{3\mathsf{M}^{\mathsf{ext}}/I} + 2^{-10\mathsf{M}^{\mathsf{ext}}/I}),
\end{aligned}$$

by Equation (95). Putting this bound back into Equation (94), we get

$$\Gamma(F|\mathcal{Z}QSB^c) \le O(2^{-10\mathsf{M}^{\mathsf{ext}}/I}) \qquad (96)$$

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 (83) and a standard calculation. Moreover, in each round, the players communicate at most $2+2\log C$ bits to exchange two indices in $\{1,\dots,C\}$. 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

$$\Gamma(\mathcal{G}|\mathcal{Z}QSB^cF^c) = 1 - O(2^{-5\mathsf{M}^{\mathsf{ext}}/I}). \tag{97}$$

So, we can conclude that

$$\begin{aligned}
\Gamma(\mathcal{Z}QS\mathcal{G}) &\geq \Gamma(\mathcal{Z}QSB^cF^c\mathcal{G})\\
&\geq \Gamma(QSB^c)\cdot\Gamma(\mathcal{Z}|QSB^c)\cdot\Gamma(F^c|\mathcal{Z}QSB^c)\cdot\Gamma(\mathcal{G}|\mathcal{Z}QSB^cF^c)\\
&\geq \Omega(2^{-3\mathsf{M}^{\mathsf{ext}}/I}), && \text{(by Equations (91), (92), (96) and (97))}
\end{aligned}$$

proving Equation (80). Observe that by Equations (80), (87) and (96),

$$\Gamma(B|\mathcal{Z}QS\mathcal{G}) \leq \frac{\Gamma(B)}{\Gamma(\mathcal{Z}QS\mathcal{G})} \leq O(2^{-2\mathsf{M}^{\mathsf{ext}}/I}), \tag{98}$$

$$\Gamma(F|\mathcal{Z}QS\mathcal{G}) \leq \frac{\Gamma(F\mathcal{Z}QSB^c)+\Gamma(B)}{\Gamma(\mathcal{Z}QS\mathcal{G})} \leq O(2^{-2\mathsf{M}^{\mathsf{ext}}/I}). \tag{99}$$

Moreover, we have

$$\begin{aligned}
\mathop{\mathbb{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QSB^cF^c] &\leq \Gamma(\mathcal{G}|\mathcal{Z}QSB^cF^c)\cdot\mathop{\mathbb{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QS\mathcal{G}B^cF^c] + \Gamma(\mathcal{G}^c|\mathcal{Z}QSB^cF^c)\\
&\leq \mathop{\mathbb{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QS\mathcal{G}B^cF^c] + O(2^{-5\mathsf{M}^{\mathsf{ext}}/I}), &(100)\\
\mathop{\mathbb{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QSB^c] &\leq \Gamma(F^c|\mathcal{Z}QSB^c)\cdot\mathop{\mathbb{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QSB^cF^c] + \Gamma(F|\mathcal{Z}QSB^c)\\
&\leq \mathop{\mathbb{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QSB^cF^c] + O(2^{-10\mathsf{M}^{\mathsf{ext}}/I}), &(101)
\end{aligned}$$

by Equation (96).

We are now ready to prove Equation (79). We have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QS\mathcal{G}]\\
&\geq \Gamma(B^cF^c|\mathcal{Z}QS\mathcal{G})\cdot\mathop{\mathbb{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QS\mathcal{G}B^cF^c] - \Gamma(B|\mathcal{Z}QS\mathcal{G}) - \Gamma(F|\mathcal{Z}QS\mathcal{G})\\
&\geq (1-O(2^{-2\mathsf{M}^{\mathsf{ext}}/I}))\cdot\mathop{\mathbb{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QS\mathcal{G}B^cF^c] - \Omega(2^{-2\mathsf{M}^{\mathsf{ext}}/I}) && \text{(by Equations (98) and (99))}\\
&\geq (1/2)\cdot\mathop{\mathbb{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QSB^cF^c] - \Omega(2^{-5\mathsf{M}^{\mathsf{ext}}/I}) - \Omega(2^{-2\mathsf{M}^{\mathsf{ext}}/I}) && \text{(by Equation (100))}\\
&\geq (1/2)\cdot\mathop{\mathbb{E}}_{\Gamma}[\mathcal{A}(xym)|\mathcal{Z}QSB^c] - \Omega(2^{-10\mathsf{M}^{\mathsf{ext}}/I}) - \Omega(2^{-2\mathsf{M}^{\mathsf{ext}}/I}) && \text{(by Equation (101))}\\
&\geq (1/4)\cdot\mathop{\mathbb{E}}_{q}[\mathcal{A}(xym)|SB^c] - \Omega(2^{-2\mathsf{M}^{\mathsf{ext}}/I}) && \text{(by Equation (94))}\\
&\geq (1/4)\cdot\mathop{\mathbb{E}}_{q(xym)}[\mathcal{A}(xym)] - \Omega(2^{-\mathsf{M}^{\mathsf{ext}}/I}) && \text{(by Equation (88))}\\
&= (1/4)\cdot\mathop{\mathbb{E}}_{q(xym)}\left[\mathsf{sign}\left(\mathop{\mathbb{E}}_{q(x'y'|m)}\left[(-1)^f\right]\right)\cdot(-1)^{f(xy)}\right] - \Omega(2^{-\mathsf{M}^{\mathsf{ext}}/I})\\
&= (1/4)\cdot\mathop{\mathbb{E}}_{q(m)}\left[\left|\mathop{\mathbb{E}}_{q(xy|m)}\left[(-1)^{f(xy)}\right]\right|\right] - \Omega(2^{-\mathsf{M}^{\mathsf{ext}}/I})\\
&\geq \Omega(2^{-\delta\mathsf{M}^{\mathsf{ext}}/(12I)}). && \text{(by Equation (56))}
\end{aligned}$$

This concludes the proof of the theorem.

10. Compressing bounded-round protocols

We prove the Theorem 10 in this section. Let $p(xym)$ be a protocol distribution such that $p(xy) = \mu(xy)$. We have $m = (m_0, \ldots, m_r)$, which is 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 (43). 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 π from Lemma 24 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 π from Lemma 24 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

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

   Bob sends $\mathsf{sign}\Big(\mathbb{E}_{q(xy|\langle m^B \rangle b)}[(-1)^f]\Big) \in \{\pm 1\}$ 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

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

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 π from Lemma 24 in step 2 of Γ. Given this assumption, define m as follows: $m_0 = m_0^A = m_0^B$, and for all $i \geq 1$, $m_i = a_i(p(m_i | m_{<i} xy)s)$, where s is a transcript of the protocol π from Lemma 24 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 24, it is clear that $\Gamma(xym) = p(xym)$.

Let S be the set defined in Equation (44) for our choice of K. In addition to S, we need the following sets to analyze the simulating protocol.

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

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

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

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:

$$\Gamma(\mathcal{EZSQ})\cdot \mathop{\mathbb{E}}_{\Gamma(xym|\mathcal{EZSQ})}[\mathcal{A}(xym)] - \Gamma(\mathcal{G}(\mathcal{EZSQ})^c). \tag{102}$$

We shall prove each of the following bounds:

$$\mathop{\mathbb{E}}_{\Gamma(xym|\mathcal{EZQS})}[\mathcal{A}(xym)] \ge \Omega(2^{-\delta\mathsf{M}/(12I)}), \tag{103}$$
$$\Gamma(\mathcal{EZQS}) \ge \Omega(2^{-3\mathsf{M}/I}), \tag{104}$$
$$\Gamma(\mathcal{G}(\mathcal{EZSQ})^c) \le O(2^{-4\mathsf{M}/I}). \tag{105}$$

Because $\delta \le 1$ and $(\alpha + K) \ge \mathsf{M}/I \ge K$, 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:

$$\Gamma(\mathcal{G}(\mathcal{EZSQ})^c) \le \Gamma(\mathcal{GE}^c) + \Gamma(\mathcal{Z}^c|\mathcal{GE}) + \Gamma(S^c\mathcal{GEZ}) + \Gamma(Q^c|\mathcal{GEZS}).$$

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

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

The event $\mathcal{GE}^c$ implies that π 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 $\varepsilon\cdot r \le 2^{-4\mathsf{M}/I}$, by our choice of ε. Finally, $\Gamma(S^c\mathcal{GEZ}) \le \Gamma(S^c\mathcal{EZ}) \le O(\varepsilon\mathsf{M}/I)$, 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 (105).

Now, we turn to proving Equation (104). 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 $xym \in S$,

$$\Gamma(Q|xym) = \frac{g_1(xm)\cdot 2^{-\lceil \log g_1(xm)\rceil}\cdot g_2(ym)\cdot 2^{\lceil \log g_1(xm)\rceil}}{2^{3\mathsf{M}/I}} = \frac{q(xym)}{p(xym)}\cdot\frac{1}{2^{3\mathsf{M}/I}}, \tag{106}$$

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,

$$\begin{aligned}\Gamma(QS) = \sum_{xym \in S} p(xym) \cdot \frac{q(xym)}{p(xym)} \cdot \frac{1}{2^{3\mathsf{M}/I}} &= \frac{q(S)}{2^{3\mathsf{M}/I}} \\ &\geq \frac{(1 - 5 \cdot 2^{-\mathsf{M}/I})}{2^{3\mathsf{M}/I}} = \Omega(2^{-3\mathsf{M}/I}), \end{aligned} \tag{107}$$

where in the last line, we used Claim 33.

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

$$\Gamma(\mathcal{Z}|xymQS) = \Gamma(\mathcal{Z}|xym) \geq 1 - O(\varepsilon \mathsf{M}/I). \tag{108}$$

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

$$\Gamma(QS\mathcal{Z}) = \Gamma(QS) \cdot \Gamma(\mathcal{Z}|QS) \geq \Omega(2^{-3\mathsf{M}/I}), \tag{109}$$

For all $xym \in S$,

$$\begin{aligned}\Gamma(xym|QS\mathcal{Z}) &= \frac{\Gamma(xym) \cdot \Gamma(QS\mathcal{Z}|xym)}{\Gamma(QS\mathcal{Z})} \\ &= \frac{p(xym)}{\Gamma(QS)} \cdot \Gamma(Q|xym) \cdot \frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} \\ &= \frac{p(xym)}{\Gamma(QS)} \cdot \frac{q(xym)}{p(xym) \cdot 2^{3\mathsf{M}/I}} \cdot \frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} && \text{(By Equation (106))} \\ &= \frac{q(xym)}{q(S)} \cdot \frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} && \text{(By Equation (107))} \\ &= q(xym|S) \cdot (1 \pm O(\varepsilon \mathsf{M}/I)), \end{aligned} \tag{110}$$

where the last line follows from Equation (108).

Given Equation (109), to complete the proof of Equation (104), 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 35 for our choice of K. We have

$$\begin{aligned}\Gamma(\mathcal{E}^c|QS\mathcal{Z}) &\leq \Gamma(T^c|QSZ) + \Gamma(\mathcal{E}^c|QS\mathcal{Z}T) \\ &\leq q(T^c|S) \cdot (1 + O(\varepsilon \mathsf{M}/I)) + \Gamma(\mathcal{E}^c|QS\mathcal{Z}T) && \text{(By Equation (110))} \\ &\leq O(2^{-\mathsf{M}/I}) + \Gamma(\mathcal{E}^c|QS\mathcal{Z}T) && \text{(By Claim 35)} \\ &\leq O(2^{-\mathsf{M}/I}) + 2\varepsilon \cdot r \leq O(2^{-\mathsf{M}/I}) \leq 1/2 \end{aligned} \tag{111}$$

where in the last line, we used the fact that given $QS\mathcal{Z}T$, item 2 and 3 of Lemma 24 guarantee that $\mathcal{E}^c$ can only happen with probability at most 2ε in each of the r rounds. Equations (109) and (111) together prove Equation (104).

Next, we prove Equation (103). Since $|\mathcal{A}(xym)| \leq 1$, we have

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

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

$$\begin{aligned}
\mathop{\mathbb{E}}_{\Gamma(xym|QS\mathcal{Z}\mathcal{E})}[\mathcal{A}(xym)] &\ge \mathop{\mathbb{E}}_{\Gamma(xym|QS\mathcal{Z})}[\mathcal{A}(xym)] - \Gamma(\mathcal{E}^c|QS\mathcal{Z}) \\
&\ge \mathop{\mathbb{E}}_{q(xym|S)}[\mathcal{A}(xym)] - \Omega(\varepsilon \mathsf{M}/I + 2^{-\mathsf{M}/I}) \\
&\qquad\qquad \text{(using Equations (110) and (111))} \\
&\ge \mathop{\mathbb{E}}_{q(xym)}[\mathcal{A}(xym)] - \Omega(2^{-\mathsf{M}/I}) \qquad \text{(by Claim 33)} \\
&= \mathop{\mathbb{E}}_{q(xym)}\left[\mathsf{sign}\left(\mathop{\mathbb{E}}_{q(x'y'|m)}\left[(-1)^{f}\right]\right)\cdot(-1)^{f(xy)}\right] - \Omega(2^{-\mathsf{M}/I}) \\
&= \mathop{\mathbb{E}}_{q(m)}\left[\left|\mathop{\mathbb{E}}_{q(xy|m)}\left[(-1)^{f(xy)}\right]\right|\right] - \Omega(2^{-\mathsf{M}/I}) \ge \Omega(2^{-\delta\mathsf{M}/(12I)}),
\end{aligned}$$

by Equation (53). This completes the proof of Equation (103).

11. Compression independent of communication

In this section, we prove Theorem 9. 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 $\mathsf{M}_I(p,f) \le \alpha I$. Let $q(xym)$ be a rectangular distribution that realizes $\mathsf{M}_I(p,f)$.

Define $\mathsf{M} = \mathsf{M}_I(p,f) + KI$. Since $\mathsf{M}(p,f) \ge 0$, we have $\mathsf{M} \ge KI$. Let g_1, g_2 be as in Equation (43). 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

$$2\log 1/\varepsilon \le O(6\mathsf{M}/I + 8\mathsf{M}) = \Delta I,$$

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 $h, t : \mathbb{Z} \to \{1, \dots, \lceil 1/\varepsilon \rceil\}$.
2. Jointly sample an infinite sequence of triples $(m^1, \rho_A^1, \rho_B^1), (m^2, \rho_A^2, \rho_B^2), \dots$, 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

$$\begin{aligned}
\rho_A^{i_A} &\le \prod_{j \text{ odd}} p(m_j^{i_A} | x m_{<j}^{i_A}), \\
\rho_B^{i_A} &\le 2^{6\mathsf{M}} \cdot \prod_{j \text{ even}} p(m_j^{i_A} | x m_{<j}^{i_A}).
\end{aligned}$$

   Alice checks if $\eta^A \le g_1(x m^{i_A}) \cdot 2^{\lceil \log g_1(x m^{i_A}) \rceil}$, 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

$$\begin{aligned}
\rho_A^{i_B} &\le 2^{6\mathsf{M}} \cdot \prod_{j \text{ odd}} p(m_j^{i_B} | y m_{<j}^{i_B}), \\
\rho_B^{i_B} &\le \prod_{j \text{ even}} p(m_j^{i_B} | y m_{<j}^{i_B}).
\end{aligned}$$

If $t(i_B) = t(i_A)$, he checks if there is a unique integer z such that

$$|z + \log g_2(ym^{i_B})| \le 3\mathsf{M}/I,$$
$$h(z) = h(\lceil \log g_1(xm^{i_A}) \rceil),$$
$$\eta^B \le g_2(ym^{i_B}) \cdot 2^{z - 3\mathsf{M}/I},$$

If all these conditions are satisfied, he sends $\mathsf{sign}\Big(\mathbb{E}_{q(xy|m^{i_B})}[(-1)^f]\Big) \in \{\pm 1\}$ 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

$$\prod_{j \text{ odd}} p(m_j^{i_*} | x m_{<j}^{i_*}) \ge \rho_A^{i_*} \text{ and } \prod_{j \text{ even}} p(m_j^{i_*} | x m_{<j}^{i_*}) \ge \rho_B^{i_*}.$$

Let $m = m^{i_*}$. We note that $\Gamma(xym) = p(xym)$.

Let S and T be the sets defined in Equation (44) and Equation (51) respectively for our choice of K. In addition to S, we need the following sets to analyze the simulating protocol:

$$Q = \Big\{xym\eta^A\eta^B : \eta^A \le g_1(xm) \cdot 2^{-\lceil \log g_1(xm) \rceil}, \eta^B \le g_2(ym) \cdot 2^{\lceil \log g_1(xm) \rceil - 3\mathsf{M}/I}\Big\},$$
$$\mathcal{E} = \Big\{i_A i_B i_* : i_A = i_B = i_*\Big\},$$
$$\mathcal{Z} = \Big\{xymh : \exists \text{ a unique integer } z \text{ with } |z + \log g_2(ym)| \le \frac{3\mathsf{M}}{I} \text{ and } h(z) = h(\lceil \log g_1(xm) \rceil)\Big\}.$$

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

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

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

$$\Gamma(\mathcal{EZSQ}) \cdot \mathop{\mathbb{E}}_{\Gamma(xym|\mathcal{EZSQ})}[\mathcal{A}(xym)] - \Gamma(\mathcal{G}(\mathcal{EZSQ})^c). \tag{112}$$

We shall prove:

$$\mathop{\mathbb{E}}_{\Gamma(xym|\mathcal{EZSQ})}[\mathcal{A}(xym)] \ge \Omega(2^{-\delta\mathsf{M}/(12I)}), \tag{113}$$
$$\Gamma(\mathcal{EZSQ}) \ge \Omega(2^{-6\mathsf{M}/I - 6\mathsf{M}}), \tag{114}$$
$$\Gamma(\mathcal{G}(\mathcal{EZSQ})^c) \le O(2^{-6\mathsf{M}/I - 7\mathsf{M}}). \tag{115}$$

By Equation (112), since $\delta \le 1$, we can choose K to be large enough to prove the theorem, since $\alpha + K \ge \mathsf{M}/I \ge K$.

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

$$\Gamma(\mathcal{G}(\mathcal{EZSQ})^c) \le \Gamma(\mathcal{GE}^c) + \Gamma(\mathcal{Z}^c|\mathcal{GE}) + \Gamma(S^c\mathcal{GEZ}) + \Gamma(Q^c|\mathcal{GEZS}).$$

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:

$$\eta^A \le g_1(xm^{i_A}) \cdot 2^{\lceil \log g_1(xm^{i_A}) \rceil} = g_1(xm) \cdot 2^{\lceil \log g_1(xm) \rceil},$$
$$\eta^B \le g_2(ym^{i_B}) \cdot 2^{z - 3\mathsf{M}/I} = g_2(ym) \cdot 2^{\lceil \log g_1(xm) \rceil - 3\mathsf{M}/I}.$$

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, $\Gamma(S^c\mathcal{GEZ}) \le \Gamma(S^c\mathcal{EZ}) \le O(\varepsilon \mathsf{M}/I)$, 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 $2^{-6\mathsf{M}/I-8\mathsf{M}}(2+\mathsf{M}/I) \le 2^{-6\mathsf{M}/I-7\mathsf{M}}$, for K sufficiently large. This implies Equation (115).

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 $xym \in S$,

$$\Gamma(Q|xym) = \frac{g_1(xm)\cdot 2^{-\lceil \log g_1(xm)\rceil}\cdot g_2(ym)\cdot 2^{\lceil \log g_1(xm)\rceil}}{2^{3\mathsf{M}/I}} = \frac{q(xym)}{p(xym)}\cdot\frac{1}{2^{3\mathsf{M}/I}}, \tag{116}$$

where the first equality follows from the fact that

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

by the definition of S. Therefore,

$$\begin{aligned}\Gamma(QS) = \sum_{xym\in S} p(xym)\cdot\frac{q(xym)}{p(xym)}\cdot\frac{1}{2^{3\mathsf{M}/I}} &= \frac{q(S)}{2^{3\mathsf{M}/I}} \\ &\ge \frac{(1-5\cdot 2^{-\mathsf{M}/I})}{2^{3\mathsf{M}/I}} = \Omega(2^{-3\mathsf{M}/I}),\end{aligned} \tag{117}$$

where in the last line, we used Claim 33.

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

$$\Gamma(\mathcal{Z}|xymQS) = \Gamma(\mathcal{Z}|xym) \ge 1 - O(\varepsilon\mathsf{M}/I). \tag{118}$$

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 $|z + \log g_2(ym)| \le 3\mathsf{M}/I$. The probability that this happens is at most $O(\varepsilon\cdot\mathsf{M}/I)$. Therefore, $\Gamma(\mathcal{Z}|xym) \ge 1 - O(\varepsilon\mathsf{M}/I) \ge 1/2$, by our choice of ε. We conclude that

$$\Gamma(QS\mathcal{Z}) = \Gamma(QS)\cdot\Gamma(\mathcal{Z}|QS) \ge \Omega(2^{-3\mathsf{M}/I}), \tag{119}$$

Let W be the event that $\min\{i_A, i_B, i_*\} = 1$ and let T be the set defined in Equation (51) for our choice of K. We claim that TW^c implies $i_* > 1$, since if $xym \in T$ then $p(m|xy) \le 2^{6\mathsf{M}}\cdot\min\{p(m|x), p(m|y)\}$, which implies

$$\begin{aligned}\prod_{j \text{ even}} p(m_j|ym_{<j}) &\le 2^{6\mathsf{M}}\cdot\prod_{j\text{ even}} p(m_j|xm_{<j}), \\ \prod_{j \text{ odd}} p(m_j|xm_{<j}) &\le 2^{6\mathsf{M}}\cdot\prod_{j\text{ even}} p(m_j|ym_{<j}),\end{aligned}$$

and hence if $i_* = 1$ then in fact $i_A = i_B = 1$.

Now, we compute $\Gamma(\mathcal{E}|QS\mathcal{Z})$.

$$\begin{aligned}\Gamma(\mathcal{E}|QS\mathcal{Z}) &= \Gamma(\mathcal{E}|QS\mathcal{Z}W) \\ &\ge \frac{\Gamma(i_A = i_B = i_* = 1|QS\mathcal{Z})}{\Gamma(i_A = 1|QS\mathcal{Z}) + \Gamma(i_B = 1|QS\mathcal{Z}) + \Gamma(i_* = 1|QS\mathcal{Z})} \\ &\ge \frac{\Gamma(i_A = i_B = i_* = 1, QS\mathcal{Z})}{\Gamma(i_A = 1) + \Gamma(i_B = 1) + \Gamma(i_* = 1)}.\end{aligned} \tag{120}$$

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,

$$\begin{aligned}
&\Gamma(i_A = i_B = i_* = 1, QS\mathcal{Z}) \\
&\geq \sum_{xym \in S \cap T} \Gamma(i_A = i_B = i_* = 1, xym, Q\mathcal{Z}) \\
&= \sum_{xym \in S \cap T} \Gamma(i_A = i_B = i_* = 1, xym) \cdot \Gamma(Q\mathcal{Z}|xym) \\
&\qquad\qquad\qquad\qquad\qquad\qquad (\text{given } xym,\ Q\mathcal{Z} \text{ is independent of } i_A, i_B, i_*) \\
&= \sum_{xym \in S \cap T} p(xym) \cdot \frac{1}{|\mathcal{M}|} \cdot \Gamma(Q\mathcal{Z}|xym) \qquad (\text{by the definition } \Gamma \text{ and } T) \\
&= \sum_{xym \in S \cap T} p(xym) \cdot \frac{1}{|\mathcal{M}|} \cdot \frac{q(xym)}{p(xym)2^{3\mathsf{M}/I}} \cdot \Gamma(\mathcal{Z}|xym) \qquad (\text{by Equation (116)}) \\
&\geq q(ST) \cdot \frac{1}{|\mathcal{M}| \cdot 2^{3\mathsf{M}/I}} \cdot (1 - \Omega(\varepsilon\mathsf{M}/I)). \qquad (\text{by Equation (118)})
\end{aligned}$$

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 (120) 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 33 and Claim 36. Using Equation (119) we get that $\Gamma(QS\mathcal{Z}\mathcal{E}) = \Omega(2^{-6\mathsf{M}-(6\mathsf{M}/I)})$ as claimed in Equation (114).

For all $xym \in S$,

$$\begin{aligned}
\Gamma(xym|QS\mathcal{Z}) &= \frac{\Gamma(xym) \cdot \Gamma(QS\mathcal{Z}|xym)}{\Gamma(QS\mathcal{Z})} \\
&= \frac{p(xym)}{\Gamma(QS)} \cdot \Gamma(Q|xym) \cdot \frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} \\
&= \frac{p(xym)}{\Gamma(QS)} \cdot \frac{q(xym)}{p(xym) \cdot 2^{3\mathsf{M}/I}} \cdot \frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} \qquad (\text{By Equation (116)}) \\
&= \frac{q(xym)}{q(S)} \cdot \frac{\Gamma(\mathcal{Z}|xymQS)}{\Gamma(\mathcal{Z}|QS)} \qquad (\text{By Equation (117)}) \\
&= q(xym|S) \cdot (1 \pm O(\varepsilon\mathsf{M}/I)), \qquad (121)
\end{aligned}$$

where the last line follows by Equation (118).

Next, we note that

$$\Gamma(\mathcal{E}|QS\mathcal{Z}, i_* = 1) \geq \Gamma(\mathcal{E}, T|QS\mathcal{Z}, i_* = 1) = \Gamma(T|QS\mathcal{Z}), \tag{122}$$

where we used the fact that the event $T, i_* = 1$ implies $\mathcal{E}$ and that xym is distributed independently of i_*. For any $xym \in S \cap T$

$$\begin{aligned}
\Gamma(xym|QS\mathcal{Z}\mathcal{E}) &= \Gamma(xym|QS\mathcal{Z}\mathcal{E}W) \quad (xym \text{ is independent of } W \text{ even conditioned on } QS\mathcal{Z}\mathcal{E}) \\
&= \Gamma(xym|QS\mathcal{Z}\mathcal{E}, i_* = 1) \quad (\text{the event } \mathcal{E}W \text{ is the same as the event } \mathcal{E}, i_* = 1) \\
&= \frac{\Gamma(xym\mathcal{E}|QS\mathcal{Z}, i_* = 1)}{\Gamma(\mathcal{E}|QS\mathcal{Z}i^* = 1)} \\
&= \Gamma(xym|QS\mathcal{Z}) \cdot \frac{\Gamma(\mathcal{E}|xym, i_* = 1)}{\Gamma(\mathcal{E}|QS\mathcal{Z}i^* = 1)} \\
&= \frac{\Gamma(xym|QS\mathcal{Z})}{\Gamma(\mathcal{E}|QS\mathcal{Z}i^* = 1)} \qquad (\text{because } xym \in S \cap T) \\
&= \Gamma(xym|QS\mathcal{Z}) \cdot (1 \pm O(\Gamma(T^c|QS\mathcal{Z})))
\end{aligned}$$

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 (122). Together with Equation (121) we get that for any $xym \in S \cap T$

$$\begin{aligned}
\Gamma(xym|QS\mathcal{Z}\mathcal{E}) &= q(xym|S) \cdot (1 \pm O(\Gamma(T^c|QS\mathcal{Z}) + \varepsilon\mathsf{M}/I)) \\
&= q(xym|S) \cdot (1 \pm O(q(T^c|S) + \varepsilon\mathsf{M}/I)) \\
&= q(xym|S) \cdot (1 \pm O(2^{-\mathsf{M}/I} + \varepsilon\mathsf{M}/I)),
\end{aligned} \tag{123}$$

where the last line follows by Claim 36.

Now, we complete the proof of Equation (113). We have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\Gamma(xym|QS\mathcal{Z}\mathcal{E})}[\mathcal{A}(xym)] \\
&\geq \sum_{xym \in S \cap T} \Gamma(xym|QS\mathcal{Z}\mathcal{E}) \cdot \mathcal{A}(xym) - \Gamma(T^c|QS\mathcal{Z}\mathcal{E}) \\
&\geq \sum_{xym \in S \cap T} q(xym|S) \cdot \mathcal{A}(xym) - \Omega(2^{-\mathsf{M}/I} + \varepsilon\mathsf{M}/I) - 1 + \Gamma(T|QS\mathcal{Z}\mathcal{E}) \quad (\text{by Equation (123)}) \\
&\geq \mathop{\mathbb{E}}_{q(xym|S)}[\mathcal{A}(xym)] - q(T^c|S) - \Omega(2^{-\mathsf{M}/I} + \varepsilon\mathsf{M}/I) - 1 + q(T|S) \cdot (1 - O(2^{-\mathsf{M}/I} + \varepsilon\mathsf{M}/I)) \\
&\qquad\qquad (\text{by Equation (123)}) \\
&\geq \mathop{\mathbb{E}}_{q(xym)}[\mathcal{A}(xym)] - q(S^c) - 2q(T^c|S) - \Omega(2^{-\mathsf{M}/I} + \varepsilon\mathsf{M}/I) \\
&= \mathop{\mathbb{E}}_{q(xym)}[\mathcal{A}(xym)] - \Omega(2^{-\mathsf{M}/I} + \varepsilon\mathsf{M}/I) \qquad (\text{by Claim 33 and Claim 36}) \\
&= \mathop{\mathbb{E}}_{q(xym)}\left[\mathsf{sign}\left(\mathop{\mathbb{E}}_{q(x'y'|m)}\left[(-1)^f\right]\right) \cdot (-1)^{f(xy)}\right] - \Omega(2^{-\mathsf{M}/I}) \\
&= \mathop{\mathbb{E}}_{q(m)}\left[\left|\mathop{\mathbb{E}}_{q(xy|m)}\left[(-1)^{f(xy)}\right]\right|\right] - \Omega(2^{-\mathsf{M}/I}) \geq \Omega(2^{-\delta\mathsf{M}/(12I)}),
\end{aligned}$$

by Equation (53).

References

[BBCR10] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, STOC '10, page 67–76, New York, NY, USA, 2010. Association for Computing Machinery.

[BK18] 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, page 964–977, New York, NY, USA, 2018. Association for Computing Machinery.

[BP16] Gábor Braun and Sebastian Pokutta. Common information and unique disjointness. *Algorithmica*, 76(3):597–629, 2016.

[BR11] Mark Braverman and Anup Rao. Information equals amortized communication. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 748–757, Los Alamitos, CA, USA, oct 2011. IEEE Computer Society.

[Bra15] Mark Braverman. Interactive information complexity. *SIAM Journal on Computing*, 44(6):1698–1739, 2015.

[BRWY13a] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct product via round-preserving compression. In Fedor V. Fomin, Rusins Freivalds, Marta Z. Kwiatkowska, and David Peleg, editors, *Automata, Languages, and Programming - 40th International Colloquium, ICALP 2013, Riga, Latvia, July 8-12, 2013, Proceedings, Part I*, volume 7965 of *Lecture Notes in Computer Science*, pages 232–243. Springer, 2013.

[BRWY13b] Mark Braverman, Anup Rao, Omri Weinstein, and Amir Yehudayoff. Direct products in communication complexity. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 746–755, 2013.

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

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

[CT91] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley series in telecommunications. J. Wiley and Sons, New York, 1991.

[FKNN95] Tomás Feder, Eyal Kushilevitz, Moni Naor, and Noam Nisan. Amortized communication complexity. *SIAM Journal on Computing*, 24(4):736–750, 1995.

[GKR16] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication for boolean functions. *J. ACM*, 63(5), nov 2016.

[HJMR10] Prahladh Harsha, Rahul Jain, David McAllester, and Jaikumar Radhakrishnan. The communication complexity of correlation. *IEEE Transactions on Information Theory*, 56(1):438–449, 2010.

[Hol07] 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, page 411–419, New York, NY, USA, 2007. Association for Computing Machinery.

[JHM+98] Mark Jerrum, Michel Habib, Colin McDiarmid, Jorge L. Ramirez-Alfonsin, and Bruce Reed. *Probabilistic Methods for Algorithmic Discrete Mathematics*, volume 16 of *Algorithms and Combinatorics*. Springer-Verlag, 1998.

[JPY12] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A direct product theorem for the two-party bounded-round public-coin communication complexity. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 167–176, 2012.

[JRS03] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. In Jos C. M. Baeten, Jan Karel Lenstra, Joachim Parrow, and Gerhard J. Woeginger, editors, *Automata, Languages and Programming*, pages 300–315, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

[JSY23] 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)*, pages 1635–1654. SIAM, 2023.

[Kol16] Gillat Kol. Interactive compression for product distributions. In *Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '16, page 987–998, New York, NY, USA, 2016. Association for Computing Machinery.

[KRW95] Mauricio Karchmer, Ran Raz, and Avi Wigderson. Super-logarithmic depth lower bounds via the direct sum in communication complexity. *Comput. Complex.*, 5(3/4):191–204, 1995.

[Păt11] Mihai Pătraşcu. Unifying the landscape of cell-probe lower bounds. *SIAM J. Comput.*, 40(3):827–847, jun 2011.

[Rao11] Anup Rao. Parallel repetition in projection games and a concentration bound. *SIAM Journal on Computing*, 40(6):1871–1891, 2011.

[Raz92] Alexander A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992.

[Raz95] Ran Raz. A parallel repetition theorem. In *Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '95, page 447–456, New York, NY, USA, 1995. Association for Computing Machinery.

[Rot17] Thomas Rothvoss. The matching polytope has exponential extension complexity. *J. ACM*, 64(6), sep 2017.

[RR15] Sivaramakrishnan Natarajan Ramamoorthy and Anup Rao. How to compress asymmetric communication. In *Proceedings of the 30th Conference on Computational Complexity*, CCC '15, page 102–123, Dagstuhl, DEU, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[RS18] Anup Rao and Makrand Sinha. Simplified separation of information and communication. *Theory of Computing*, 14(20):1–29, 2018.

[RY20] Anup Rao and Amir Yehudayoff. *Communication Complexity: and Applications*. Cambridge University Press, 2020.

[Sha48] Claude Elwood Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 1948.

[She18] Alexander A. Sherstov. Compressing interactive communication under product distributions. *SIAM Journal on Computing*, 47(2):367–419, 2018.

[Sin18] Makrand Sinha. Lower bounds for approximating the matching polytope. In Artur Czumaj, editor, *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 1585–1604. SIAM, 2018.

[SK87] 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.

[Yu22] Huacheng Yu. Strong xor lemma for communication with bounded rounds : (extended abstract). In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1186–1192, 2022.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il