Fiftieth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
October 1 - 5, 2012

Interactive Hypothesis Testing with Communication Constraints

Yu Xiang and Young-Han Kim
Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA 92093, USA
Email: {yxiang,yhk}@ucsd.edu

***Abstract*—This paper studies the problem of interactive hypothesis testing with communication constraints, in which two communication nodes separately observe one of two correlated sources and interact with each other to decide between two hypotheses on the joint distribution of the sources. When testing against independence, that is, the joint distribution of the sources under the alternative hypothesis is the product of the marginal distributions under the null hypothesis, a computable characterization is provided for the optimal tradeoff between the communication rates in two-round interaction and the testing performance measured by the type II error exponent such that the type I error probability asymptotically vanishes. An example is provided to show that interaction is strictly helpful.**

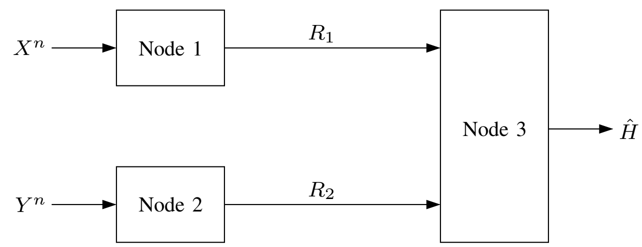

Fig. 1. Multiterminal hypothesis testing with communication constraints.

I. Introduction

Berger [1], in an inspiring attempt at combining information theory and statistical inference, formulated the problem of hypothesis testing with communication constraints as depicted in Fig. 1. Let $(X^n, Y^n) \sim \prod_{i=1}^n p_{X,Y}(x_i, y_i)$ be a pair of independent and identically distributed (i.i.d.) n-sequences generated by a two-component discrete memoryless source (2-DMS) (X, Y). Suppose that there are two hypotheses on the joint distribution of (X, Y), namely,

$$H_0 : (X, Y) \sim p_0(x, y),$$
$$H_1 : (X, Y) \sim p_1(x, y).$$

In order to decide which hypothesis is true, nodes 1 and 2 observe X^n and Y^n, respectively, compress their observed sequences into indices of rates R_1 and R_2, and communicate them over noiseless links to node 3, which then makes a decision $\hat{H} \in \{H_0, H_1\}$ based on the received compression indices. What is the impact of communication constraints on the performance of hypothesis testing? To answer this question, Berger [1] studied the optimal tradeoff between the communication rates and the testing performance that is measured by the exponent of the type II error probability such that the type I error probability is upper bounded by a given $\epsilon < 1$. Despite many natural applications, however, theoretical understanding of this problem is far from complete and a simple characterization of this rate–exponent tradeoff remains open in general.

In their celebrated work [2], Ahlswede and Csiszár studied the special case in which the sequence Y^n is fully available at the destination node, i.e., $R_2 = \infty$. They established single-letter inner and outer bounds on the optimal tradeoff between the communication rate R_1 and the type II error exponent and showed that these bounds are tight for testing against independence, i.e., the alternative hypothesis H_1 is $p_1(x, y) = p_0(x)p_0(y)$. Later, Han [3] and Shimokawa, Han, and Amari [4] provided a new coding scheme that improves upon the Ahlswede and Csiszar inner bound for the general hypothesis testing problem. The Shimokawa–Han–Amari scheme is similar to the Berger–Tung scheme [5], [6] for the distributed lossy source coding problem, where node 1 and node 2 perform joint typicality encoding followed by binning. A more comprehensive survey on the earlier literature can be found in [7]. Recently, several variations of this setup have been studied, including successive refinement hypothesis testing [8] and testing against conditional independence [9].

This paper studies an interactive version of hypothesis testing with communication constraints. Two nodes communicate with each other through noiseless links and one of the nodes is to perform hypothesis testing at the end of interactive communication. To be concrete, we focus on *two rounds* of interaction for hypothesis testing *against independence*. For this special case, we establish a single-letter characterization of the optimal tradeoff between the communication rates and the type II error probability when the type I error probability is arbitrarily small.

The rest of the paper is organized as follows. In Section II, we review the problem of one-way hypothesis testing with communication constraints. In Section III, we formulate the problem of interactive hypothesis testing with communication constraints and present our main theorem. In Section IV, we compare the interactive hypothesis testing problem with the interactive lossy source coding problem by Kaspi [10].

Throughout the paper, we closely follow the notation in [11]. In particular, for $X \sim p(x)$ and $\epsilon \in (0,1)$, we define the set of ϵ-typical n-sequences x^n (or the typical set in short) [12] as $\mathcal{T}_\epsilon^{(n)}(X) = \{x^n : |\#\{i : x_i = x\}/n - p(x)| \le \epsilon p(x) \text{ for all } x \in \mathcal{X}\}$. We say that $X \to Y \to Z$ form a Markov chain if $p(x,y,z) = p(x)p(y|x)p(z|y)$, that is, X and Z are conditionally independent of each other given Y.

II. One-way Hypothesis Testing with Communication Constraints

As before, let $(X^n, Y^n) \sim \prod_{i=1}^n p_{X,Y}(x_i, y_i)$ be a pair of i.i.d. sequences generated by a 2-DMS (X,Y) and consider two hypotheses

$$H_0 : (X,Y) \sim p_0(x,y),$$
$$H_1 : (X,Y) \sim p_1(x,y).$$

We consider the special case of the hypothesis testing problem depicted in Fig. 1, in which $R_2 = \infty$; see Fig. 2. Here, node 2 is equivalent to node 3 and is required to make a decision $\hat{H} \in \{H_0, H_1\}$.

A $(2^{nR_1}, n)$ hypothesis test consists of

- an encoder that assigns an index $m_1(x^n) \in [1 : 2^{nR_1}]$ to each sequence $x^n \in \mathcal{X}^n$, and
- a tester that assigns $\hat{h}(m_1, y^n) \in \{H_0, H_1\}$ to each $(m_1, y^n) \in [1 : 2^{nR_1}] \times \mathcal{Y}^n$.

The acceptance region is defined as

$$\mathcal{A}_n := \{(m_1, y^n) \in [1 : 2^{nR_1}] \times \mathcal{Y}^n : \hat{h}(m_1, y^n) = H_0\}.$$

Then the type I error probability is

$$\mathsf{P}_0(\mathcal{A}_n^c) = \sum_{(x^n,y^n):(m_1(x^n),y^n)\in\mathcal{A}_n^c} p_0(x^n, y^n)$$

and the type II error probability is

$$\mathsf{P}_1(\mathcal{A}_n) = \sum_{(x^n,y^n):(m_1(x^n),y^n)\in\mathcal{A}_n} p_1(x^n, y^n).$$

Fix $\epsilon \in (0,1)$ and define the optimal type II error probability as

$$\beta_n^*(R_1, \epsilon) := \min \mathsf{P}_1(\mathcal{A}_n),$$

where the minimum is over all $(2^{nR_1}, n)$ tests such that $\mathsf{P}_0(\mathcal{A}_n^c) \le \epsilon$. Further define the optimal type II error exponent as

$$\theta_1(R_1, \epsilon) := \lim_{n\to\infty} -\frac{1}{n} \log \beta_n^*(R_1, \epsilon).$$

Now suppose that the two hypotheses are

$$H_0 : (X,Y) \sim p_0(x,y),$$
$$H_1 : (X,Y) \sim p_1(x,y) = p_0(x)p_0(y).$$

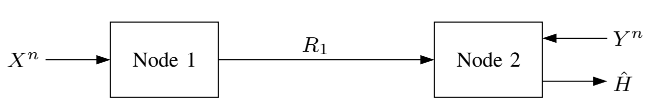

Fig. 2. One-way hypothesis testing with communication constraint.

Here $p_0(x)$ and $p_0(y)$ are marginal distributions of $p_0(x,y)$. For this special case of hypothesis testing *against independence*, Ahlswede and Csiszár established the following.

Theorem 1 (Ahlswede and Csiszár [2]): For every $\epsilon \in (0,1)$,

$$\theta_1(R_1, \epsilon) = \max_{p(u_1|x): R_1 \ge I(U_1;X)} I(U_1; Y), \tag{1}$$

where the cardinality bound for U_1 is $|\mathcal{U}_1| \le |\mathcal{X}| + 1$.

We illustrate the theorem with the following.

Example 1: Consider the following Z binary sources (X,Y) depicted in Fig. 3, where Y is the output of X through a Z channel (equivalently, X is the output of Y through an inverted Z channel) and

$$p_{X,Y}(0,0) = 1/2,$$
$$p_{X,Y}(0,1) = 0,$$
$$p_{X,Y}(1,0) = 1/4,$$
$$p_{X,Y}(1,1) = 1/4.$$

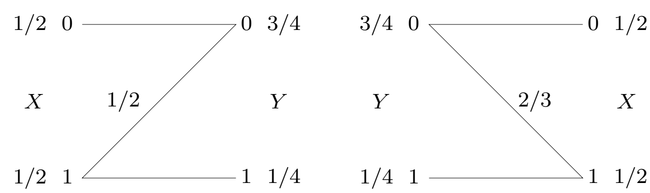

Fig. 3. Two equivalent representations of (X,Y).

We now apply Theorem 1 and evaluate the optimal type II error exponent in (1). Since $|\mathcal{U}_1| \le |\mathcal{X}_1| + 1 = 3$, we can optimize over all conditional pmfs $p(u_1|x)$ of the form in Fig. 4. Then we have

$$\theta_1(R_1, \epsilon) = \max\left(H_1 - \frac{1}{6}H_2 - \frac{3}{4}H_3\right),$$

where the maximum is over all (a,b,c,d) such that

$$R_1 \ge H_1 - \frac{1}{2}H_4 - \frac{1}{2}H_2$$

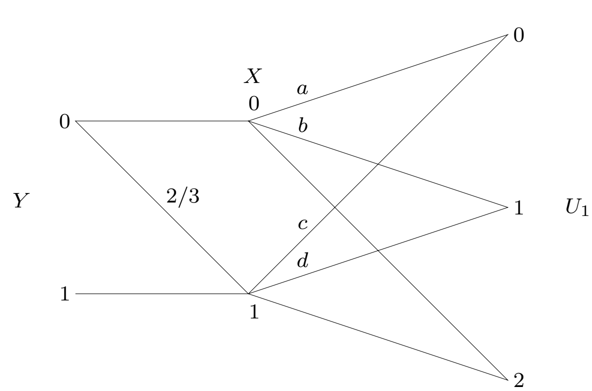

Fig. 4. Conditional pmf $p(u_1|x)$.

and H_1 through H_4 are defined as

$$
\begin{aligned}
H_1 &:= H\left(\frac{a+c}{2}, \frac{b+d}{2}, \frac{2-a-b-c-d}{2}\right),\\
H_2 &:= H(c, d, 1-c-d),\\
H_3 &:= H\left(\frac{a+2c}{3}, \frac{b+2d}{3}, \frac{3-a-b-2c-2d}{3}\right),\\
H_4 &:= H(a, b, 1-a-b).
\end{aligned}
$$

For example, when $R_1 = 1/2$, we have $\theta_1(R_1, \epsilon) \approx 0.1878$. The entire curve of $\theta_1(R_1, \epsilon)$ is plotted in Fig. 7 in Section III.

III. Interactive Hypothesis Testing with Communication Constraints

Suppose that instead of making an immediate decision based on one round of communication, the two nodes can interactively communicate over a noiseless bidirectional link before one of the nodes performs hypothesis testing. We wish to characterize the optimal tradeoff between the communication rates and the performance of hypothesis testing. For simplicity of discussion, we focus on the 2-round case depicted in Fig. 5.

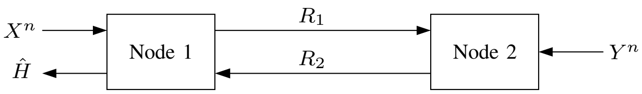

Fig. 5. Interactive hypothesis testing with communication constraints.

As before, we consider testing against independent, i.e.,

$$
\begin{aligned}
H_0 &: (X, Y) \sim p_0(x, y),\\
H_1 &: (X, Y) \sim p_1(x, y) = p_0(x)p_0(y).
\end{aligned}
$$

A $(2^{nR_1}, 2^{nR_2}, n)$ hypothesis test consists of

- a round 1 encoder that assigns an index $m_1(x^n) \in [1 : 2^{nR_1}]$ to each sequence $x^n \in \mathcal{X}^n$,
- a round 2 encoder that assigns an index $m_2(m_1, y^n) \in [1 : 2^{nR_2}]$ to each $(m_1, y^n) \in [1 : 2^{nR_1}] \times \mathcal{Y}^n$, and
- a tester that assigns $\hat{h}(m_2, x^n) \in \{H_0, H_1\}$ to each $(m_1, x^n) \in [1 : 2^{nR_1}] \times \mathcal{X}^n$.

The acceptance region is defined as

$$\mathcal{A}_n := \{(m_2, x^n) \in [1 : 2^{nR_2}] \times \mathcal{X}^n : \hat{h}(m_2, x^n) = H_0\}.$$

The type I error probability is $\mathsf{P}_0(\mathcal{A}_n^c)$ and the type II error probability is $\mathsf{P}_1(\mathcal{A}_n)$. Fix $\epsilon \in (0, 1)$ and define the optimal type II error probability as

$$\beta_n^*(R_1, R_2, \epsilon) := \min \mathsf{P}_1(\mathcal{A}_n)$$

where the minimum is over all $(2^{nR_1}, 2^{nR_2}, n)$ tests with $\mathsf{P}_0(\mathcal{A}_n^c) \le \epsilon$. Further define the optimal type II error exponent as

$$\theta_2(R_1, R_2, \epsilon) := \lim_{n\to\infty} -\frac{1}{n} \log \beta_n^*(R_1, R_2, \epsilon).$$

Remark 1: The optimal type II error exponent is lower bounded as

$$
\begin{aligned}
\theta_2(R_1, R_2, \epsilon) &\ge \max\{\theta_2(R_1 + R_2, 0, \epsilon),\, \theta_2(0, R_1 + R_2, \epsilon)\}\\
&\ge \theta_1(R_1 + R_2, \epsilon).
\end{aligned}
$$

We establish the optimal tradeoff between the rate constraints and the testing performance by characterizing $\theta_2(R_1, R_2, \epsilon)$ as $\epsilon \to 0$.

We are ready to state the main result of the paper.

Theorem 2:

$$\lim_{\epsilon\to 0} \theta_2(R_1, R_2, \epsilon) = \max_{\substack{p(u_1|x)p(u_2|u_1,y):\\ R_1 \ge I(U_1;X),\\ R_2 \ge I(U_2;Y|U_1)}} \big(I(U_1; Y) + I(U_2; X|U_1)\big), \tag{2}$$

where $|\mathcal{U}_1| \le |\mathcal{X}| + 1$ and $|\mathcal{U}_2| \le |\mathcal{Y}|\cdot|\mathcal{U}_1| + 1$.

Remark 2: By setting $U_2 = \emptyset$ and $R_2 = 0$, Theorem 2 recovers the optimal one-way type II error exponent in Theorem 1.

Remark 3: We can express the optimal tradeoff between communication constraints and the type II error exponent by the *rate–exponent region* that consists of all rate–exponent triples (R_1, R_2, θ) such that

$$
\begin{aligned}
R_1 &\ge I(U_1; X),\\
R_2 &\ge I(U_2; Y|U_1),\\
\theta &\le I(U_1; Y) + I(U_2; X|U_1)
\end{aligned}
$$

for some conditional pmfs $p(u_1|x)p(u_2|u_1, y)$.

Example 2 (Interaction helps): We revist the Z binary sources in Fig. 3. We show that the two-round interaction can strictly outperform the one-way case. While the optimal type II exponent in Theorem 2 can be directly evaluated, we instead use the simple lower bound on $\theta_2(R_1, R_2, \epsilon)$ discussed in Remark 1. Consider

$$
\begin{aligned}
\lim_{\epsilon\to 0} \theta_2(R_1, R_2, \epsilon) &\ge \lim_{\epsilon\to 0} \theta_2(0, R_1 + R_2, \epsilon)\\
&\overset{(a)}{\ge} \max_{p(u_2|y): R_1+R_2 \ge I(U_2;Y)} I(U_2; X),
\end{aligned}
$$

where (a) follows by Theorem 1. Since $|\mathcal{U}_2| \le 3$, we can again optimize over all conditional pmfs $p(u_2|y)$ of the form in Fig. 6, which yields

$$\theta_2(0, R_1 + R_2, \epsilon) = \max\left(H_1 - \frac{1}{2}H_2 - \frac{1}{2}H_3\right),$$

where the maximum is over all (a, b, c, d) such that

$$R_1 + R_2 \ge H_1 - \frac{3}{4}H_2 - \frac{1}{4}H_4$$

and

$$
\begin{aligned}
H_1 &:= H\left(\frac{3a+c}{4}, \frac{3b+d}{4}, \frac{4-3a-3b-c-d}{4}\right),\\
H_2 &:= H(a, b, 1-a-b),\\
H_3 &:= H\left(\frac{a+c}{2}, \frac{b+d}{2}, \frac{2-a-b-c-d}{2}\right),\\
H_4 &:= H(c, d, 1-c-d).
\end{aligned}
$$

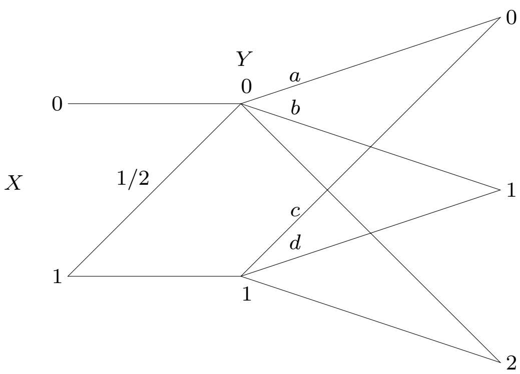

Fig. 6. Conditional pmf $p(u_2|y)$.

Fig. 7 numerically compares the one-way exponent $\theta_1(R_1 + R_2, \epsilon)$ with the lower bound $\theta_2(0, R_1 + R_2, \epsilon)$ on the two-round exponent $\theta_2(R_1, R_2, \epsilon)$ as $\epsilon \to 0$. For every value of the sum rate $R_1 + R_2 \in (0, 1)$,

$$\theta_2(0, R_1 + R_2, \epsilon) > \theta_1(R_1 + R_2, \epsilon)$$

and thus there is strict improvement by using interaction. In fact, it can be shown that the gap increases as the cross probability of the Z channel increases, that is, as the channel becomes more and more skewed.

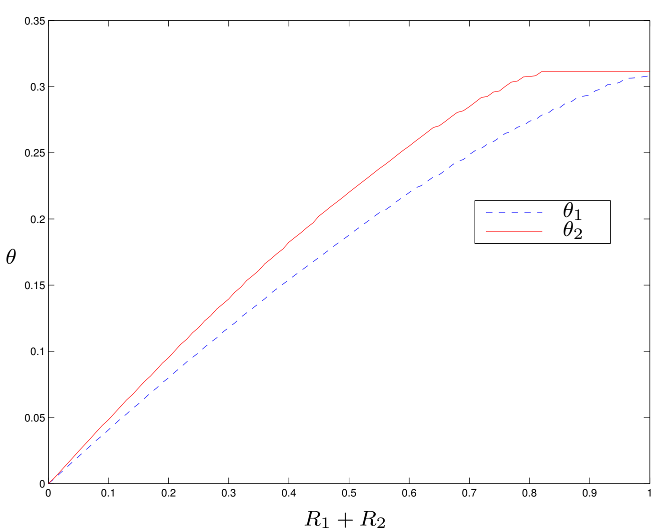

Fig. 7. Comparison of the one-way case with the two-round case. The solid red curve corresponds to the lower bound $\theta_2(0, R_1 + R_2, \epsilon)$ for the two-round case and the dotted blue curve corresponds to $\theta_1(R_1 + R_2, \epsilon)$ for the one-way case.

In the following two subsections, we prove Theorem 2 by establishing achievability and the weak converse.

A. Proof of Achievability

Codebook generation. Fix a conditional pmf $p(u_1, u_2|x, y) = p(u_1|x)p(u_2|u_1, y)$ that attains the maximum in (2). Let $p_0(u_1) = \sum_x p_0(x)p(u_1|x)$ and $p_0(u_2|u_1) = \sum_y p_0(y)p(u_2|u_1, y)$. Randomly and independently generate 2^{nR_1} sequences $u_1^n(m_1)$, $m_1 \in [1 : 2^{nR_1}]$, each according to $\prod_{i=1}^n p_0(u_{1i})$. Randomly and independently generate 2^{nR_2} sequences $u_2^n(m_2|m_1)$, $m_2 \in [1 : 2^{nR_2}]$, each according to $\prod_{i=1}^n p_0(u_{2i}|u_{1i})$. These sequences constitute the codebook $\mathcal{C}$, which is revealed to both nodes.

Encoding for round 1. Given a sequence x^n, node 1 finds an index m_1 such that $(x^n, u_1^n(m_1)) \in \mathcal{T}_\epsilon^{(n)}$. If there is more than one such index, it sends the smallest one among them. If there is no such index, it selects an index from $[1 : 2^{nR_1}]$ uniformly at random.

Encoding for round 2. Given y^n and m_1, node 2 finds an index m_2 such that $(y^n, u_1^n(m_1), u_2^n(m_2)) \in \mathcal{T}_\epsilon^{(n)}$. If there is more than one such index, it selects one of them uniformly at random. If there is no such index, it selects an index from $[1 : 2^{nR_2}]$ uniformly at random.

Testing. Upon receiving m_2, node 1 sets the acceptance region $\mathcal{A}_n$ for H_0 to

$$\mathcal{A}_n = \{(m_2, x^n) : (u_1^n(m_1), u_2^n(m_2), x^n) \in \mathcal{T}_\epsilon^{(n)}\},$$

where the jointly typical set $\mathcal{T}_\epsilon^{(n)} = \mathcal{T}_\epsilon^{(n)}(U_1, U_2, X)$ is defined with respect to $p_0(x, y)$, $p(u_1|x)$ and $p(u_2|u_1, y)$.

Analysis of two types of error. Let (M_1, M_2) denote the chosen indices at node 1 and 2 respectively. Node 1 chooses $\hat{H} \neq H_0$ iff one or more of the following events occur:

$$\begin{aligned}
\mathcal{E}_1 &= \{(U_1^n(m_1), X^n) \notin \mathcal{T}_{\epsilon''}^{(n)} \text{ for all } m_1 \in [1 : 2^{nR_1})\},\\
\mathcal{E}_2 &= \{(U_2^n(m_2), U_1^n(M_1), Y^n) \notin \mathcal{T}_{\epsilon'}^{(n)}\\
&\qquad\qquad \text{for all } m_2 \in [1 : 2^{nR_2})\},\\
\mathcal{E}_3 &= \{(U_1^n(M_1), U_2^n(M_2), X^n) \notin \mathcal{T}_\epsilon^{(n)}\}.
\end{aligned}$$

For the type I error probability, assume that H_0 is true. Then

$$\alpha_n = \mathsf{P}(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_3) \leq \mathsf{P}(\mathcal{E}_1) + \mathsf{P}(\mathcal{E}_1^c \cap \mathcal{E}_2) + \mathsf{P}(\mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_3).$$

We now bound each term. By the covering lemma [11, Section 3.7], $\mathsf{P}(\mathcal{E}_1)$ tends to zero as $n \to \infty$ if $R_1 \geq I(U_1; X) + \delta(\epsilon')$. Now we bound the second term. Since $\epsilon > \epsilon' > \epsilon''$, $\mathcal{E}_1^c = \{(U_1^n(M_1), X^n) \in \mathcal{T}_{\epsilon''}^{(n)}\}$ and $Y^n|\{U_1^n(M_1) = u_1^n, X^n = x^n\} \sim \prod_{i=1}^n p_0(y_i|u_{1i}, x_i) = \prod_{i=1}^n p_0(y_i|x_i)$, by the conditional typicality lemma [11, Section 2.5], then $\mathsf{P}\{(U_1^n(M_1), X^n, Y^n) \in \mathcal{T}_{\epsilon'}^{(n)}\}$ tends to zero as $n \to \infty$ and thus $\mathsf{P}\{(U_1^n(M_1), Y^n) \in \mathcal{T}_{\epsilon'}^{(n)}\}$ tends to zero as $n \to \infty$. Therefore, again by the covering lemma, $\mathsf{P}(\mathcal{E}_1^c \cap \mathcal{E}_2)$ tends to zero as $n \to \infty$ if $R_2 \geq I(U_2; Y|U_1) + \delta(\epsilon')$.

To bound the last term, we use a version of the Markov lemma [5] in [11, Section 12.1]. Let $(x^n, u_1^n, y^n) \in \mathcal{T}_{\epsilon'}^{(n)}$ and consider $\mathsf{P}\{U_2^n(M_2) = u_2^n|X^n = x^n, U_1^n(M_1) = u_1^n, Y^n = y^n\} = \mathsf{P}\{U_2^n(M_2) = u_2^n|U_1^n(M_1) = u_1^n, Y^n = y^n\} = p(u_2^n|u_1^n, y^n)$. First note that by the covering lemma, $\mathsf{P}\{U_2^n(M_2) \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n)|U_1^n(M_1) = u_1^n, Y^n = y^n\}$ tends to one as $n \to \infty$, that is, $p(u_2^n|u_1^n, y^n)$ satisfies the first condition in the Markov lemma. For the second condition, the following is proved in the Appendix.

Lemma 1: For every $u_2^n \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n)$ and n sufficiently large,

$$p(u_2^n|u_1^n, y^n) \doteq 2^{-nH(U_2|U_1,Y)}.$$

Hence, by the Markov lemma, $\mathsf{P}\{(x^n, u_1^n, y^n, U_2^n(M_2)) \in \mathcal{T}_{\epsilon}^{(n)} \mid X^n = x^n, U_1^n(M_1) = u_1^n, Y^n = y^n\}$ tends to one as $n \to \infty$, if $(u_1^n, x^n, y^n) \in \mathcal{T}_{\epsilon'}^{(n)}(U_1, X_n, Y^n)$ and $\epsilon' < \epsilon$ is sufficiently small. Therefore, $\mathsf{P}(\mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_3) \to 0$ as $n \to \infty$.

For the type II error probability, assume in this case that H_1 is true. Then

$$\beta_n = \mathsf{P}(\mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_3^c) = \mathsf{P}(\mathcal{E}_1^c)\, \mathsf{P}(\mathcal{E}_2^c|\mathcal{E}_1^c)\, \mathsf{P}(\mathcal{E}_3^c|\mathcal{E}_1^c \cap \mathcal{E}_2^c).$$

We now bound each factor. By the covering lemma, $\mathsf{P}(\mathcal{E}_1^c)$ tends to one as $n \to \infty$ if $R_1 \geq I(U_1; X) + \delta(\epsilon')$. Define the event

$$\mathcal{E}_0 = \{(U_1^n(M), Y^n) \notin \mathcal{T}_{\epsilon'}^{(n)}\}.$$

Then, given that $(U_1^n(M), X^n) \in \mathcal{T}_{\epsilon'}^{(n)}$,

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_2^c|\mathcal{E}_1^c) &= \mathsf{P}(\mathcal{E}_2^c \cap \mathcal{E}_0|\mathcal{E}_1^c) + \mathsf{P}(\mathcal{E}_2^c \cap \mathcal{E}_0^c|\mathcal{E}_1^c) \\
&= \mathsf{P}(\mathcal{E}_2^c \cap \mathcal{E}_0^c|\mathcal{E}_1^c) \\
&= \mathsf{P}(\mathcal{E}_2^c|\mathcal{E}_0^c)\, \mathsf{P}(\mathcal{E}_0^c|\mathcal{E}_1^c).
\end{aligned}$$

By the covering lemma, $\mathsf{P}(\mathcal{E}_2^c|\mathcal{E}_0^c)$ tends to one as $n \to \infty$ if $R_2 \geq I(U_2; Y|U_1) + \delta(\epsilon')$.

$$\mathsf{P}(\mathcal{E}_0^c|\mathcal{E}_1^c) = \sum_{(u^n, y^n) \in \mathcal{T}_{\epsilon'}^{(n)}} p_1(u_1^n) p_1(y^n) \leq 2^{-n(I(U_1;Y) - \delta(\epsilon'))}.$$

Thus we have

$$\mathsf{P}(\mathcal{E}_2^c|\mathcal{E}_1^c) \leq 2^{-n(I(U_1;Y) - \delta(\epsilon'))}.$$

For the third factor $\mathsf{P}(\mathcal{E}_3^c|\mathcal{E}_1^c \cap \mathcal{E}_2^c)$, we need the following.

Lemma 2: If H_1 is true, we have

$$p_1(u_2, x|u_1) = p_1(u_2|u_1) p_1(x|u_1).$$

Proof: We have

$$\begin{aligned}
p_1(u_2, x|u_1) &= \sum_y p_1(u_2, y, x|u_1) \\
&= \sum_y p_1(y, x|u_1) p_1(u_2|y, u_1) \\
&= \sum_y p_1(x|y, u_1) p_1(y|u_1) p_1(u_2|y, u_1) \\
&= \sum_y p_1(x|u_1) p_1(u_2, y|u_1) \\
&= p_1(u_2|u_1) p_1(x|u_1).
\end{aligned}$$

■

Now we bound $\mathsf{P}(\mathcal{E}_3^c|\mathcal{E}_1^c \cap \mathcal{E}_2^c)$. Given that $(U_1^n(M), X^n) \in \mathcal{T}_{\epsilon'}^{(n)}$ and $(U_2^n(K), U_1^n(M), Y^n) \in \mathcal{T}_{\epsilon'}^{(n)}$, we have

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}_3^c|\mathcal{E}_2^c \cap \mathcal{E}_1^c) \\
&= \sum_{(u_1^n, u_2^n, x^n) \in \mathcal{T}_{\epsilon}^{(n)}} \mathsf{P}\{U_1^n(M) = u_1^n, U_2^n(K) = u_2^n, \\
&\qquad\qquad X^n = x^n | \mathcal{E}_2^c \cap \mathcal{E}_1^c\} \\
&= \sum_{(u_1^n, u_2^n, x^n) \in \mathcal{T}_{\epsilon}^{(n)}} p_1(u_2^n|u_1^n) p_1(x^n, u_1^n) \\
&\leq 2^{n(H(U_1,U_2,X) + \delta(\epsilon))} 2^{-n(H(U_2|U_1) - \delta(\epsilon'))} 2^{-n(H(U_1,X) - \delta(\epsilon'))} \\
&= 2^{n(H(U_2|U_1,X) - H(U_2|U_1) - \delta(\epsilon))} \\
&= 2^{-n(I(U_2;X|U_1) - \delta(\epsilon))}.
\end{aligned}$$

Combining the bounds on the three factors, we have

$$\beta_n \leq 2^{-n(I(U_1;Y) + I(U_2;X|U_1) - \delta(\epsilon))}.$$

In summary, the type I error probability averaged over all codebooks is upper bounded by ϵ if $R_1 \geq I(U_1; X)$ and $R_2 \geq I(U_2; Y|U_1)$, while the type II error probability averaged over all codebooks is upper bounded (in exponent) by $2^{-n(I(U_1;Y)+I(U_2;X|U_1)-\delta(\epsilon))}$. Therefore, there exists a codebook such that

$$\theta_2(R_1, R_2, \epsilon) \geq I(U_1; Y) + I(U_2; X|U_1).$$

This completes the achievability proof.

B. Proof of the Converse

Given a $(2^{nR_1}, 2^{nR_2}, n)$ test characterized by the encoding functions m_1 and m_2, and the acceptance region $\mathcal{A}_n$, we have by the data processing inequality for relative entropy that

$$\begin{aligned}
&D\Bigg(\sum_{y^n} p_0(x^n, y^n) p(m_1|x^n) p(m_2|m_1, y^n) || \\
&\qquad \sum_{y^n} p_0(x^n) p_0(y^n) p(m_1|x^n) p(m_2|m_1, y^n)\Bigg) \\
&\geq (1 - \alpha) \log \frac{1 - \alpha}{\beta} + \alpha \log \frac{\alpha}{1 - \beta},
\end{aligned}$$

where $\alpha := \mathsf{P}_0(\mathcal{A}_n^c)$ and $\beta := \mathsf{P}_1(\mathcal{A}_n)$. Let $M_1 = m_1(X^n)$ and $M_2 = m_2(M_1, Y^n)$. Then by the definition of $\beta_n^*(R_1, R_2, \epsilon)$, we must have

$$\begin{aligned}
H(M_1) &\leq nR_1, \\
H(M_2) &\leq nR_2, \\
\alpha &\leq \epsilon, \\
\beta &\leq \beta_n^*(R_1, R_2, \epsilon).
\end{aligned}$$

Then,

$$\begin{aligned}
&(1 - \alpha) \log \frac{1 - \alpha}{\beta} + \alpha \log \frac{\alpha}{1 - \beta} \\
&\quad = (1 - \alpha) \log \frac{1}{\beta} + \alpha \log \frac{1}{1 - \beta} - H(\alpha) \\
&\quad \geq (1 - \alpha) \log \frac{1}{\beta} - H(\alpha) \\
&\quad \geq (1 - \epsilon) \log \frac{1}{\beta} - H(\alpha).
\end{aligned}$$

Thus we have the following multiletter expression upper bound as

$$\lim_{\epsilon\to 0}\theta(R_1,R_2,\epsilon)$$
$$\le \lim_{n\to\infty}\frac{1}{n}D\big(p_0(m_1,m_2,x^n)||p_1(m_1,m_2,x^n)\big),$$

where

$$p_0(m_1,m_2,x^n)=\sum_{y^n}p_0(x^n,y^n)p(m_1|x^n)p(m_2|m_1,y^n),$$
$$p_1(m_1,m_2,x^n)=\sum_{y^n}p_0(x^n)p_0(y^n)p(m_1|x^n)p(m_2|m_1,y^n).$$

The relative entropy term is upper bounded as

$$\begin{aligned}
&D\big(p_0(m_1,m_2,x^n)||p_1(m_1,m_2,x^n)\big)\\
&=D\big(p_0(x^n,m_2|m_1)p_0(m_1)||p_0(x^n|m_1)p_1(m_2|m_1)p_0(m_1)\big)\\
&=\sum_{x^n,m_1,m_2}p_0(x^n,m_1,m_2)\log\frac{p_0(m_2|m_1,x^n)}{p_1(m_2|m_1)}\\
&=\sum_{x^n,m_1,m_2}p_0(x^n,m_1,m_2)\log\bigg(\frac{p_0(m_2|m_1,x^n)}{p_0(m_2|m_1)}\cdot\frac{p_0(m_2|m_1)}{p_1(m_2|m_1)}\bigg)\\
&=I(X^n;M_2|M_1)+\sum_{m_1,m_2}p_0(m_1,m_2)\log\frac{p_0(m_2|m_1)}{p_1(m_2|m_1)}
\end{aligned} \tag{3}$$

where $p_0(m_2|m_1,x^n)$, $p_1(m_2|m_1)$, and $p_0(m_2|m_1)$ are defined as

$$p_0(m_2|m_1,x^n):=\sum_{y^n}p_0(y^n|x^n)p(m_2|m_1,y^n),$$
$$p_1(m_2|m_1):=\sum_{y^n}p_0(y^n)p(m_2|m_1,y^n),$$
$$p_0(m_2|m_1):=\sum_{y^n}p_0(y^n|m_1)p(m_2|m_1,y^n).$$

The second term in (3) is upper bounded as

$$\begin{aligned}
&\sum_{m_1,m_2}p_0(m_1,m_2)\log\frac{p_0(m_2|m_1)}{p_1(m_2|m_1)}\\
&=D(p_0(m_1,m_2)||p_1(m_1,m_2))\\
&=D\bigg(p_0(m_1)\sum_{y^n}p_0(y|m_1)p(m_2|m_1,y^n)||\\
&\qquad p_0(m_1)\sum_{y^n}p_0(y)p(m_2|m_1,y^n)\bigg)\\
&\overset{(a)}{\le} D(p_0(m_1)p_0(y|m_1)p(m_2|m_1,y^n)||\\
&\qquad p_0(m_1)p_0(y)p(m_2|m_1,y^n))\\
&=I(M_1;Y^n),
\end{aligned}$$

where (a) follows by the data processing inequality for relative entropy. Thus we have

$$\theta:=\lim_{\epsilon\to 0}\theta_2(R_1,R_2,\epsilon)$$
$$\le \lim_{n\to\infty}\frac{1}{n}\big(I(X^n;M_2|M_1)+I(M_1;Y^n)\big). \tag{4}$$

To complete the converse proof, we single-letterize the upper bound in (4) in the following steps. First consider

$$\begin{aligned}
nR_1&\ge H(M_1)\\
&\ge I(M_1;X^n)\\
&=\sum_{i=1}^{n}I(M_1;X_i|X^{i-1})\\
&\overset{(a)}{=}\sum_{i=1}^{n}I(M_1,X^{i-1},Y^{i-1};X_i),
\end{aligned}$$

where (a) follows from the fact that X^n and Y^n are i.i.d. Next consider

$$\begin{aligned}
nR_2&\ge H(M_2)\\
&\ge I(M_2;X^n,Y^n|M_1)\\
&=\sum_{i=1}^{n}I(M_2;X_i,Y_i|M_1,X^{i-1},Y^{i-1})\\
&\ge\sum_{i=1}^{n}I(M_2;Y_i|M_1,X^{i-1},Y^{i-1}).
\end{aligned}$$

Now the mutual information term $I(M_1;Y^n)$ is upper bounded as

$$\begin{aligned}
I(M_1;Y^n)&=\sum_{i=1}^{n}I(M_1;Y_i|Y^{i-1})\\
&=\sum_{i=1}^{n}I(M_1,Y^{i-1};Y_i)\\
&\le\sum_{i=1}^{n}I(M_1,X^{i-1},Y^{i-1};Y_i).
\end{aligned}$$

Finally $I(M_2;X^n|M_1)$ is upper bounded as

$$\begin{aligned}
&I(M_2;X^n|M_1)\\
&=\sum_{i=1}^{n}I(M_2;X_i|M_1,X^{i-1})\\
&\overset{(a)}{=}\sum_{i=1}^{n}I(M_2;X_i|M_1,X^{i-1},Y^{i-1})\\
&+I(M_2;Y^{i-1}|M_1,X^{i-1})-I(M_2;Y^{i-1}|M_1,X^{i-1},X_i)\\
&\overset{(b)}{\le}\sum_{i=1}^{n}I(M_2;X_i|M_1,X^{i-1},Y^{i-1}),
\end{aligned}$$

where (a) follows since

$$\begin{aligned}
&I(M_2;X_i,Y^{i-1}|M_1,X^{i-1})\\
&=I(M_2;X_i|M_1,X^{i-1})+I(M_2;Y^{i-1}|M_1,X^{i-1},X_i)\\
&=I(M_2;Y^{i-1}|M_1,X^{i-1})+I(M_2;X_i|M_1,X^{i-1},Y^{i-1}).
\end{aligned}$$

and (b) follows since

$$\begin{aligned}
&I(M_2;Y^{i-1}|M_1,X^{i-1})-I(M_2;Y^{i-1}|M_1,X^{i-1},X_i)\\
&=H(Y^{i-1}|M_1,X^{i-1})-H(Y^{i-1}|M_1,X^{i-1},X_i)\\
&-H(Y^{i-1}|M_1,M_2,X^{i-1})+H(Y^{i-1}|M_1,M_2,X^{i-1},X_i)\\
&=I(Y^{i-1};X_i|M_1,X^{i-1})-I(Y^{i-1};X_i|M_1,M_2,X^{i-1})\\
&\overset{(c)}{\le}0.
\end{aligned}$$

$$\lim_{\epsilon \to 0} \theta_2(R_1, R_2, \epsilon) \leq \max_{\substack{p(u_1|x), p(u_2|u_1,y): \\ R_1 \geq I(U_1;X), \\ R_2 \geq I(U_2;Y|U_1)}} I(U_1;Y) + I(U_2;X|U_1). \tag{4}$$

Here, the inequality (c) holds since

$$X_i \to (M_1, X^{i-1}) \to Y^{i-1}$$

forms a Markov chain. Identifying $U_{1i} = (M_1, X^{i-1}, Y^{i-1})$ and $U_{2i} = M_2$ and note that $U_{1i} \to X_i \to Y_i$ and $U_{2i} \to (U_{1i}, Y_i) \to X_i$ form two Markov chains. Thus, for

$$\begin{aligned} nR_1 &\geq \sum_{i=1}^{n} I(M_1, X^{i-1}, Y^{i-1}; X_i) \\ &= \sum_{i=1}^{n} I(U_{1i}; X_i) \end{aligned}$$

and

$$\begin{aligned} nR_2 &\geq \sum_{i=1}^{n} I(M_2; Y_i | M_1, X^{i-1}, Y^{i-1}) \\ &= \sum_{i=1}^{n} I(U_{2i}; Y_i | U_{1i}), \end{aligned}$$

we have

$$\begin{aligned} n\theta &\leq \sum_{i=1}^{n} I(M_1, X^{i-1}, Y^{i-1}; Y_i) \\ &\qquad + I(M_2; X_i | M_1, X^{i-1}, Y^{i-1}) \\ &= \sum_{i=1}^{n} I(U_{1i}; Y_i) + I(U_{2i}; Y_i | U_{1i}). \end{aligned}$$

Define the time-sharing random variable Q to be uniformly distributed over $[1 : n]$ and independent of (M_1, M_2, X^n, Y^n), and identify $U_1 = (Q, U_{1Q})$, $U_2 = (Q, U_{2Q})$, $X = X_Q$, and $Y = Y_Q$. Clearly, $U_1 \to X \to Y$ and $U_2 \to (U_1, Y) \to X$ form two Markov chains. Thus we have (4) at the top of the page. Finally, the cardinality bounds on U_1 and U_2 follow the standard technique, in particular, the one used in the 2-round interactive lossy source coding problem [10]. This completes the converse proof.

IV. Discussion

Consider the interactive lossy source coding problem first studied by Kaspi [10], as depicted in Fig. 8. Here two nodes interactively communicate with each other so that each node can reconstruct the source observed by the other node with prescribed distortions. Kaspi [10] established the general q-round rate distortion region. Ma and Ishwar [13] provided an ingenious example showing that interactive communication can strictly outperform one-way communication. In this section, we compare the two-round interactive hypothesis testing problem with the two-round interactive lossy source coding problem. For the formal definition of the latter, refer to [10] or [11, Section 20.3].

We recall the optimal tradeoff between communication constraints and distortion constraints.

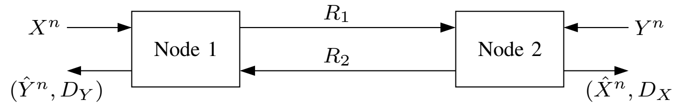

Fig. 8. Interactive lossy source coding.

Theorem 3 (Kaspi [10]): The two-round rate–distortion region is the set of all rate pairs (R_1, R_2) such that

$$\begin{aligned} R_1 &\geq I(U_1; X) - I(U_1; Y), \\ R_2 &\geq I(U_2; Y | U_1) - I(U_2; X | U_1) \end{aligned}$$

for some $p(u_1|x)p(u_2|u_1, y)$ with $|\mathcal{U}_1| \leq |\mathcal{X}| + 1$ and $|\mathcal{U}_2| \leq |\mathcal{Y}| \cdot |\mathcal{U}_1| + 1$ and functions $\hat{x}(u_1, u_2, y)$ and $\hat{y}(u_1, u_2, x)$ that satisfy $\mathsf{E}(d(X, \hat{X})) \leq D_X$ and $\mathsf{E}(d(Y, \hat{Y})) \leq D_Y$.

Achievability is established by performing Wyner–Ziv coding [14] in each round, i.e., joint typicality encoding followed by binning. By contrast, the scheme we used for the interactive hypothesis testing problem is joint typicality encoding in each round (without binning). The excessive communication rates caused by not using binning pay back with the type II error exponent; see Remark 3.

It turns out, however, that this distinction between binning and no binning is not fundamental. By using Wyner–Ziv coding in the interactive hypothesis testing problem, we can establish the following tradeoff between communication constraints and the testing performance.

Proposition 1: The rate–exponent region for two-round interactive hypothesis testing is the set of rate–exponent triples (R_1, R_2, θ) such that

$$\begin{aligned} \theta &\leq I(U_1; Y) + I(U_2; X | U_1) \\ R_1 &\geq I(U_1; X) - I(U_1, Y) \\ R_2 &\geq I(U_2; Y | U_1) - I(U_2; X | U_1) \\ R_1 + R_2 - \theta &\geq I(U_1; X) + I(U_2; Y | U_1) \\ &\qquad - I(U_1; Y) - I(U_2; X | U_1) \end{aligned}$$

for some $p(u_1|x)p(u_2|u_1, y)$.

It can be shown that the region in Proposition 1 is equivalent to the region in Remark 3 (and the optimal error exponent in Theorem 2). As pointed out by Rahman and Wagner [9] in the one-way setup, binning never hurts [9]. Therefore, the coding scheme for two-round interactive lossy source coding leads to an essentially identical scheme for two-round interactive hypothesis testing, which is optimal! This equivalence can be extended to the general q-round interactive hypothesis testing and lossy source coding problems. We will explore this connection further in a subsequent publication elsewhere.

$$
\begin{aligned}
&\mathsf{P}\{U_2^n(M_2) = u_2^n | U_1^n = u_1^n, Y^n = y^n\} \\
&= \mathsf{P}\{U_2^n(M_2) = u_2^n, U_2^n \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n) | U_1^n = u_1^n, Y^n = y^n\} \\
&= \mathsf{P}\{U_2^n(M_2) \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n) | U_1^n = u_1^n, Y^n = y^n\} \cdot \mathsf{P}\{U_2^n(m_2) = u_2^n | U_1^n = u_1^n, Y^n = y^n, U_2^n \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n)\} \\
&\le \mathsf{P}\{U_2^n(M_2) = u_2^n | U_1^n = u_1^n, Y^n = y^n, U_2^n \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n)\} \\
&= \sum_{m_2} \mathsf{P}\{U_2^n(M_2) = u_2^n, M_2 = m_2 | U_1^n = u_1^n, Y^n = y^n, U_2^n \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n)\} \\
&= \sum_{m_2} \mathsf{P}\{M_2 = m_2 | U_1^n = u_1^n, Y^n = y^n, U_2^n \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n)\} \\
&\qquad\qquad \cdot \mathsf{P}\{U_2^n(m_2) = u_2^n | U_1^n = u_1^n, Y^n = y^n, U_2^n \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n), M_2 = m_2\} \\
&\stackrel{(a)}{=} \sum_{m_2} \mathsf{P}\{M_2 = m_2 | U_1^n = u_1^n, Y^n = y^n, U_2^n \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n)\} \cdot \mathsf{P}\{U_2^n(m_2) = u_2^n | U_2^n \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n)\} \\
&\stackrel{(b)}{\le} \sum_{m_2} \mathsf{P}\{M_2 = m_2 | U_1^n = u_1^n, Y^n = y^n, U_2^n \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n)\} \cdot 2^{-n(H(U_2|U_1,Y) - \delta(\epsilon'))} \\
&= 2^{-n(H(U_2|U_1,Y) - \delta(\epsilon'))}.
\end{aligned}
$$

APPENDIX
PROOF OF LEMMA 1

For every $u_2^n \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n)$, $\mathsf{P}\{U_2^n(M_2) = u_2^n | U_1^n = u_1^n, Y^n = y^n\}$ is upper bounded at the top of this page, where (a) follows since $U_2^n(m_2)$ is independent of $(Y^n$, $U_1(m_1))$ and $U_2^n(m_2')$ for $m_2' \neq m_2$ and is conditionally independent of M_2 given $(Y^n$, $U_1(m_1))$ and the indicator variables of the event $U_2^n(m_2) \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n)$, $m_2 \in [1 : 2^{nR_2}]$, which implies that the event $\{U_2^n(m_2) = u_2^n\}$ is conditionally independent of $\{Y^n, U_1(m_1), M_2 = m_2\}$ given $U_2^n(m_2) \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n)$. Step (b) follows from the properties of typical sequences. Similarly, for every $u_2^n \in \mathcal{T}_{\epsilon'}^{(n)}(U_2^n|u_1^n, y^n)$ and n sufficiently large,

$$
\begin{aligned}
&\mathsf{P}\{U_2^n(M_2) = u_2^n | U_1^n = u_1^n, Y^n = y^n\} \\
&\qquad\qquad \ge (1 - \epsilon') 2^{-n(H(U_2|U_1,Y) + \delta(\epsilon'))}.
\end{aligned}
$$

This completes the proof of Lemma 1.

REFERENCES

[1] T. Berger, "Decentralized estimation and decision theory," in *Proc. IEEE Inf. Theory Workshop*, Mt. Kisco, NY, Sep. 1979.

[2] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, no. 4, pp. 533–542, 1986.

[3] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 33, no. 6, pp. 759–772, 1987.

[4] H. Shimokawa, T. S. Han, and S. Amari, "Error bound of hypothesis testing with data compression," in *Proc. IEEE Internat. Symp. Inf. Theory*, Jun. 1994, p. 29.

[5] S.-Y. Tung, "Multiterminal source coding," Ph.D. Thesis, Cornell University, Ithaca, NY, 1978.

[6] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications*, G. Longo, Ed. New York: Springer-Verlag, 1978.

[7] T. S. Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2300–2324, Oct. 1998.

[8] C. Tian and J. Chen, "Successive refinement for hypothesis testing and lossless one-helper problem," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4666–4681, Oct. 2008.

[9] M. S. Rahman and A. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6282–6303, Oct. 2012.

[10] A. H. Kaspi, "Two-way source coding with a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 735–740, 1985.

[11] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge: Cambridge University Press, 2011.

[12] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, 2001.

[13] N. Ma and P. Ishwar, "Some results on distributed source coding for interactive function computation," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 6180–6195, Sep. 2011.

[14] A. D. Wyner and J. Ziv, "The rate–distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, 1976.