

Distributed Detection With Vector Quantizer

Wenwen Zhao, *Student Member, IEEE*, and Lifeng Lai, *Member, IEEE*

Abstract—Motivated by distributed inference over big datasets problems, we study multiterminal distributed inference problems when each terminal employs vector quantizer. The use of vector quantizer enables us to relax the conditional independence assumption normally used in the distributed detection with scalar quantizer scenarios. We first consider a case of practical interest in which each terminal is allowed to send zero-rate messages to a decision maker. Subject to a constraint that the error exponent of the type 1 error probability is larger than a certain level, we characterize the best error exponent of the type 2 error probability using basic properties of the r -divergent sequences. We then consider the scenario with positive rate constraints, for which we design schemes to benefit from the less strict rate constraints.

Index Terms—Distributed detection, exponential-type constraints, error exponent, hypothesis testing.

I. INTRODUCTION

INFERRING the relationship among multiple random variables from data plays an important role in machine learning, statistical inference, and wireless sensor network applications. The centralized setting in which all of the data is available at one terminal is well studied. The distributed setting, in which available data is stored/observed at multiple terminals connected by channels with limited capacities, is more challenging and has attracted significant recent research interests [3]–[20].

Of particular interest to this paper is a class of distributed detection problems. In the class of distributed detection problems considered, there are L terminals \mathcal{X}_i , $i = 1, \dots, L$, where terminal \mathcal{X}_i has observations related to random variable X_i . These terminals can send information related to their own data using a limited rate to a decision maker \mathcal{Y} . Based on the messages received from these terminals and its own data related to random variable Y , the decision maker \mathcal{Y} tries to determine which of the following two hypotheses is true:

$$H_0 : P_{X_1 \dots X_L Y} \quad \text{vs} \quad H_1 : Q_{X_1 \dots X_L Y},$$

in which $P_{X_1 \dots X_L Y}$ and $Q_{X_1 \dots X_L Y}$ are two different probability mass functions (PMFs). As the communication rates between the terminals and the decision maker are limited, terminal \mathcal{X}_i has to quantize/compress its observations X_i^n with n

being the number of samples available at terminal \mathcal{X}_i . The focus of the distributed detection is to design the quantizers and decision functions under various resources (e.g., communication cost) and performance (e.g., error probabilities) constraints.

There have been a large number of existing work on distributed detection problems, see [6]–[20] and reference therein. Most of the existing work consider the *scalar quantizer*, in which the quantizer at terminal \mathcal{X}_i quantizes each component of X_i^n one by one. This setup is well suited for certain sensor network applications, as the complexity of the scalar quantizer is low and it incurs minimal decision delay. Under this scalar quantization setup, it is typically assumed that the observations at different terminals are conditionally (conditioned on the hypothesis) independent. The problem become very challenging once this conditional independence assumption is relaxed [6], [7]. Some interesting recent work have made important progress for the case with correlated observations [11], [12], [15].

In this paper, we focus on distributed detection problems with *vector quantizer*, in which the observations are processed in blocks. This setup is not only related to the distributed detection problems motivated by sensor networks, it is also relevant to recent interests in distributed inference/learning over large data set problems [3]–[5]. In these distributed inference over large data set problems, available data is stored in multiple terminals. In these setups, the computational cost at each terminal is less of a concern than the communication cost between terminals. The use of vector quantizer allows us to borrow powerful tools from information theory for distributed detection/inference problems. As we will show in the paper, these tools enable us to make progress in understanding the general problems without the conditional independence assumptions.

We first focus on the zero-rate compression case in which each terminal is only allowed to send messages to the decision maker with a diminishing rate (zero-rate compression). If the decision maker were required to fully recover the data of terminals \mathcal{X}_i s as in the distributed source coding problems [21], [22], this zero-rate compression is not enough. However, in our setup, this zero-rate compression will still be valuable for the decision maker for statistical inference. In addition, we impose an exponential-type constraint on the type 1 error probability (i.e., the type 1 error probability is required to decrease exponentially fast with a certain error exponent). We fully characterize the best achievable error exponent of the type 2 error probability under these zero-rate compression and exponential-type type 1 error probability constraints by providing matching lower and upper bounds. A clear benefit of this zero-rate compression approach is that the terminals only need to consume a limited amount of communication resources. In addition, we show that a very simple scheme in which each terminal only

Manuscript received September 24, 2015; revised January 10, 2016; accepted January 19, 2016. Date of publication February 03, 2016; date of current version April 29, 2016. The work was supported by the Qatar National Research Fund under Grant QNRF-6-1326-2-532. This paper was presented in part at Annual Allerton Conference on Communication, Control, and Computing, Montecello, IL, 2014 [1] and at the IEEE Symposium on Information Theory, Hongkong, China, 2015 [2]. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Vincenzo Matta.

The authors are with the Department of Electrical and Computer Engineering, Worcester Polytechnic Institute, Worcester, MA 01609 USA (e-mail: wzhao@wpi.edu; llai@wpi.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSPN.2016.2524572

sends the empirical distribution (or an approximation of it) is optimal. This implies that the complexity of the optimal scheme employed by sensors in practical detection problem can be very low. Furthermore, we provide an example in which the performance of the scheme with zero-rate compression is very close to that of the centralized case.

We then extend the study to the positive rate constraints case. Compared with the zero-rate compression case, in this scenario, each terminal can convey more information to the decision maker. As the general problem is very complicated, we focus on the special case of testing against independence. The case with (X_1, \dots, X_L) all at one terminal (with Y being at another terminal) was first considered by Ahlswede and Csiszár [23]. In [23], the problem was converted to a source coding with a helper problem. However, this approach may not work for our case, as the corresponding problem will be a source coding with multiple helpers problem, which is still open. In our paper, we use a different approach to exploit the more flexible rate constraints and characterize the corresponding type 2 error exponents. Furthermore, we provide an upper-bound on the best achievable type 2 error exponent using any scheme that satisfies the communication rate and type 1 error exponents.

Our work is also related to several existing interesting works [24]–[30] that study distributed detection/estimation problems using vector quantizers. Of particular relevance, among these contributions, [26], [27] considered a similar setup with the exponential-type constraint on the type 1 error probability and zero-rate constraint on the communication. [26] provided a lower bound on the type 2 error exponent. Later [27] established an upper bound that coincides with the lower bound derived in [26] by converting the exponential-type constraint problems to the constant-type constraint problems considered in [25]. Furthermore, as mentioned above, the problem of testing against independence was first studied by Ahlswede and Csiszár [23], which provided a matching upper and lower bound on the type 2 error exponent. The key difference between the model considered in [23], [26], [27] and our model is that [23], [26], [27] focused on the case where (X_1, \dots, X_L) are all at one terminal (with Y being at another terminal), while in our model these random variables are all at different terminals.

The remainder of the paper is organized as follows. In Section II, we introduce the model studied in this paper. In Section III, we present our results for the zero-rate compression case. In Section IV, we focus on the scenario with positive rate constraints. In Section V-B, we use several numerical examples to illustrate analytical results obtained in this paper. Finally, we offer some concluding remarks in Section VI.

II. MODEL

Consider a system with L terminals: \mathcal{X}_i , $i = 1, \dots, L$ and a decision maker \mathcal{Y} . Each terminal and the decision maker observe a component of the random vector (X_1, \dots, X_L, Y) that take values in a finite set $\mathcal{X}_1 \times \dots \times \mathcal{X}_L \times \mathcal{Y}$ and admit a joint PMF with two possible forms:

$$H_0 : P_{X_1 \dots X_L Y}, \quad H_1 : Q_{X_1 \dots X_L Y}. \quad (1)$$

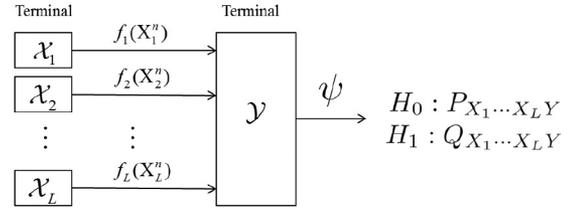


Fig. 1. Model

With a slight abuse of notation, we use \mathcal{X}_i to denote both the terminal and the alphabet set from which the random variable X_i takes values. $(X_1^n, \dots, X_L^n, Y^n)$ are independently and identically generated according to one of the above joint PMFs. In other words, $(X_1^n, \dots, X_L^n, Y^n)$ is generated by either $P_{X_1 \dots X_L Y}^n$ or $Q_{X_1 \dots X_L Y}^n$. In a typical hypothesis testing problem, one determines which hypothesis is true under the assumption that $(X_1^n, \dots, X_L^n, Y^n)$ are fully available at the decision maker. In this paper, we consider a distributed setting in which X_i^n , $i = 1, \dots, L$ and Y^n are at different locations. In particular, terminal \mathcal{X}_i observes only X_i^n and terminal \mathcal{Y} observes only Y^n . Terminals \mathcal{X}_i s are allowed to send messages to the decision maker \mathcal{Y} . Using Y^n and the received messages, \mathcal{Y} determines which hypothesis is true. We denote this system as $S_{X_1 \dots X_L | Y}$. Figure 1 illustrates the system model. In the following, we will use the term “decision maker” and terminal \mathcal{Y} interchangeably. Here, Y^n is used to model any side information available at the decision maker. If \mathcal{Y} is set to be an empty set, then the decision maker does not have side information.

After observing the data sequence $x_i^n \in \mathcal{X}_i^n$, terminal \mathcal{X}_i will use a vector quantizer (will also be called encoder in the sequel) f_i to transform the sequence x_i^n into a message $f_i(x_i^n)$, which takes values from the message set \mathcal{M}_i

$$f_i : \mathcal{X}_i^n \rightarrow \mathcal{M}_i = \{1, 2, \dots, M_i\}, \quad (2)$$

with rate constraint:

$$\frac{1}{n} \log M_i \leq R_i, \quad i = 1, \dots, L. \quad (3)$$

Using messages M_i , $i = 1, \dots, L$ and its side information Y^n , the decision maker will employ a decision function ψ to determine which hypothesis is true:

$$\psi : \mathcal{M}_1 \times \dots \times \mathcal{M}_L \times \mathcal{Y}^n \rightarrow \{H_0, H_1\}. \quad (4)$$

For any given vector quantizers f_i , $i = 1, \dots, L$ and decision function ψ , one can define the acceptance region as

$$\mathcal{A}_n = \{(X_1^n, \dots, X_L^n, Y^n) \in \mathcal{X}_1^n \times \dots \times \mathcal{X}_L^n \times \mathcal{Y}^n : \psi(f_1(X_1^n) \dots f_L(X_L^n) Y^n) = H_0\}. \quad (5)$$

Correspondingly, the type 1 error probability is defined as

$$\alpha_n = P_{X_1 \dots X_L Y}^n(\mathcal{A}_n^c), \quad (6)$$

in which \mathcal{A}_n^c denotes the complement of \mathcal{A}_n , and the type 2 error probability is defined as

$$\beta_n = Q_{X_1 \dots X_L Y}^n(\mathcal{A}_n). \quad (7)$$

Our goal is to design the quantization functions f_i , $i = 1, \dots, L$ and the decision function ψ to maximize the type 2 error exponent under certain type 1 error and communication rate constraints (3).

More specifically, we consider two kinds of type 1 error constraint, namely:

- Constant-type constraint

$$\alpha_n \leq \epsilon \quad (8)$$

for a prefixed $\epsilon > 0$, which implies that the type 1 error probability must be smaller than a given threshold; and

- Exponential-type constraint

$$\alpha_n \leq \exp(-nr) \quad (9)$$

for a given $r > 0$, which implies that the type 1 error probability must decrease exponentially fast with an exponent no less than r . Hence the exponential-type constraint is stricter than the constant-type constraint.

To distinguish these two different type 1 error constraints, we use different notations to denote the corresponding type 2 error exponent.

- Under the constant-type constraint, we define the type 2 error exponent as

$$\theta(R_1, \dots, R_L, \epsilon) = \liminf_{n \rightarrow \infty} \left(-\frac{1}{n} \log \left(\min_{f_1, \dots, f_L, \psi} \beta_n \right) \right),$$

in which the minimization is over all f_i s and ψ satisfying condition (3) and (8).

- Under the exponential-type constraint, we define the type 2 error exponent as

$$\sigma(R_1, \dots, R_L, r) = \liminf_{n \rightarrow \infty} \left(-\frac{1}{n} \log \left(\min_{f_1, \dots, f_L, \psi} \beta_n \right) \right),$$

in which the minimization is over all f_i s and ψ satisfying condition (3) and (9).

Here, we would like to highlight main differences between our work and the large number of existing work in distributed detection [6]–[20]:

- Most of the existing work focus on scalar quantizer that performs observation-by-observation processing, i.e., $f_i(x^n) = (f_{i,1}(x_i(1)), \dots, f_{i,t}(x_i(t)), \dots, f_{i,n}(x_i(n)))$. In other words, at time t , the scalar quantizer quantizes observation $x_i(t)$ into $f_{i,t}(x_i(t))$ without consideration of all other observations. In our work, we consider vector quantizer. The use of vector quantizer allows us to borrow powerful tools from information theory for distributed detection/inference problems. Furthermore, it enables us to consider zero-rate compression problems, which are of practical importance in big data applications.
- In the scalar quantizer problems, it is assumed that the observations at the terminals are conditionally independent (conditioned on the hypothesis), i.e., it is assumed that

$$P_{X_1 \dots X_L Y} = P_{X_1} \dots P_{X_L} P_Y, \quad (10)$$

$$\text{and } Q_{X_1 \dots X_L Y} = Q_{X_1} \dots Q_{X_L} Q_Y. \quad (11)$$

One of the main reasons for this assumption is that, under the scalar quantizer setup, the problem quickly becomes intractable when the observations are not conditionally independent. With the vector quantizer, as shown in the sequel, one can make substantial progress without the conditional independence assumption.

- While most of the existing works [6]–[20] with communication theoretic flavor focus on minimizing the exact error probability (and hence is a challenging task), our work focuses on maximizing the error exponent and hence it is an asymptotic setup.

III. TESTING UNDER ZERO-RATE COMPRESSION WITH EXPONENTIAL-TYPE CONSTRAINTS

In this section, we focus on the “zero-rate” compression, i.e., $R_1 = \dots = R_L = 0$ under the exponential-type constraint. More specifically, we assume

$$\text{as } n \rightarrow \infty, M_i \rightarrow \infty, \quad (12)$$

but

$$R_i = \frac{1}{n} \log M_i \downarrow 0, \quad i = 1, \dots, L. \quad (13)$$

In this case, $\sigma(R_1, \dots, R_L, r)$ will be denoted as $\sigma(0, \dots, 0, r)$. This zero-rate compression is of practical interest, as the normalized (normalized by the length of the data) communication cost is minimal. It is well-known that in the traditional distributed source coding with side information problems [21], [22], whose goal is to recover (X_1^n, \dots, X_L^n) at terminal \mathcal{Y} , this zero-rate information is not useful. However, in our setup, the goal is only to determine which hypothesis is true. This zero-rate information will be very useful.

The scenario with zero rate compression under the constant-type constraint has been considered in [25]. We will discuss the scenario with zero rate compression under the exponential-type constraint (9).

In the following subsections, we first review several concepts that are useful for our development. We then characterize the type 2 error exponent with $L = 2$ before extending the result to the general case.

A. Preliminary

Following [22], for any sequence $x^n = (x(1), \dots, x(n)) \in \mathcal{X}^n$, the relative frequencies (empirical PMF) $\pi(a|x^n) \triangleq n(a|x^n)/n, \forall a \in \mathcal{X}$ of the components of x^n is called the type of x^n and is denoted by $tp(x^n)$. Here $n(a|x^n)$ is the total number of indices t at which $x(t) = a$.

Furthermore, we call a random variable $X^{(n)}$ that has the same distribution as $tp(x^n)$ as the type variable of x^n .

For any given sequence x^n , we can measure how likely this sequence is generated from a PMF P_X using the concept of typical sequence [22] and r -divergent sequence [26]. Roughly speaking, a sequence x^n is said to be typical if the empirical PMF is close to P_X . More precisely, x^n is called to be typical if $|\pi(a|x^n) - P_X(a)| \leq \epsilon P_X(a), \forall a \in \mathcal{X}$ for a given ϵ . We use $T_\epsilon^{(n)}$ to denote the set of typical sequences.

The concept of r -divergent sequences also plays an important role in the following development. Here, we review the definition and some important properties of r -divergent sequences. More details and properties of r -divergent sequences can be found in [26].

Definition 1: ([26]) Let X be a random variable taking values in a finite set \mathcal{X} with PMF P_X , and $r \geq 0$. An n -sequence $x^n = (x_1, \dots, x_n) \in \mathcal{X}^n$ is called r -divergent sequence for X if

$$D(X^{(n)}||X) \leq r, \quad (14)$$

where $X^{(n)}$ is the type variable of x^n and $D(\cdot||\cdot)$ is the Kullback-Leibler (KL) divergence of the two random variables involved. The set of all r -divergent sequences is denoted by $S_r^n(X)$.

In particular, $S_0^n(X)$ (i.e., $r = 0$) represents the set of all x^n sequences such that $tp(x^n) = P_X$. The following lemma from [26] summarizes key properties of r -divergent sequences.

Lemma 1: ([26]) Let $r > 0$ be fixed.

- 1) $P_X^n(S_r^n(X)) \geq 1 - (n+1)^{|\mathcal{X}|} \exp(-nr)$.
- 2) Let $x^n \in \mathcal{X}^n$ and X be a random variable in \mathcal{X} , then

$$\Pr(X^n = x^n) = \exp \left[-n(H(X^{(n)}) + D(X^{(n)}||X)) \right]. \quad (15)$$

- 3) Let \mathcal{A}_n be a subset of \mathcal{X}^n and

$$P_X^n(\mathcal{A}_n) \geq 1 - \exp(-nr) \quad (16)$$

holds. Let $\mathcal{A}_n(X^{(n)}) \triangleq \mathcal{A}_n \cap S_0^n(X^{(n)})$, we have

$$\left| \mathcal{A}_n(X^{(n)}) \right| \geq \left(1 - (n+1)^{|\mathcal{X}|} \exp[-n(r - c_n)] \right) \times \left| S_0^n(X^{(n)}) \right| \quad (17)$$

with $c_n = D(X^{(n)}||X)$.

B. The Case With $L = 2$

In this subsection, to assist the presentation, we first focus on the case with $L = 2$ and provide details on how to characterize $\sigma(0, 0, r)$. We will then discuss the general case in Section III-C.

We first establish an upper bound on the error exponent that any scheme can achieve. We will follow the similar strategy as in [27]. In particular, we will first convert a problem with the exponential-type constraint to a corresponding problem with the constant-type constraint. We then obtain an upper bound on the error exponent using the results in [25] for the constant-type constraint.

Theorem 1: Let $P_{X_1 X_2 Y}$ be arbitrary and $Q_{X_1 X_2 Y} > 0$. For zero-rate compression in $S_{X_1 X_2 | Y}$ with $R_1 = R_2 = 0$, the error exponent satisfies

$$\sigma(0, 0, r) \leq \sigma_{opt}, \quad (18)$$

in which

$$\sigma_{opt} \triangleq \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{H}_r} D(\tilde{P}_{X_1 X_2 Y} || Q_{X_1 X_2 Y}) \quad (19)$$

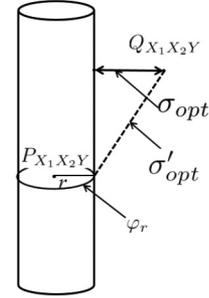


Fig. 2. σ_{opt} for zero-rate hypothesis testing

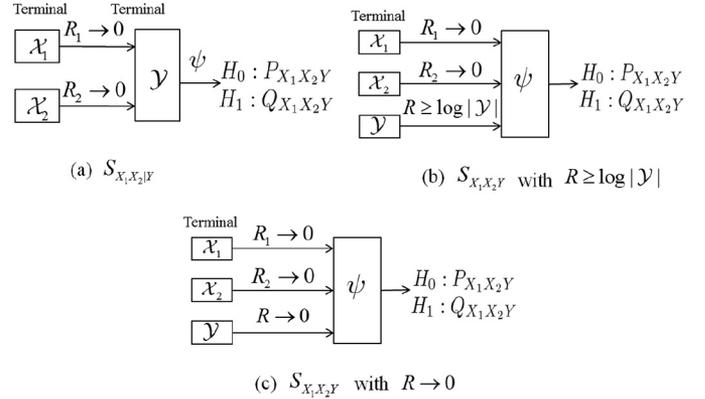


Fig. 3. Model for achievability in zero-rate compression

with

$$\mathcal{H}_r = \left\{ \tilde{P}_{X_1 X_2 Y} : \tilde{P}_{X_1} = \hat{P}_{X_1}, \tilde{P}_{X_2} = \hat{P}_{X_2}, \tilde{P}_Y = \hat{P}_Y \right. \\ \left. \text{for some } \hat{P}_{X_1 X_2 Y} \in \varphi_r \right\}, \quad (20)$$

$$\varphi_r = \left\{ \hat{P}_{X_1 X_2 Y} : D(\hat{P}_{X_1 X_2 Y} || P_{X_1 X_2 Y}) \leq r \right\}. \quad (21)$$

Proof: Please refer to Appendix A. \blacksquare

Figure 2 illustrates a geometric interpretation of σ_{opt} . In a centralized detection problem, X_1^n , X_2^n and Y^n are all available to the decision maker, so the decision maker knows the joint distribution of the observations. Setting the acceptance region as all observations whose empirical joint PMF having a KL-divergence to $P_{X_1 X_2 Y}$ less than or equal to r , expressed by φ_r in Figure 2, then the best type 2 error exponent is the dashed line from $Q_{X_1 X_2 Y}$ to φ_r in Figure 2, denoted as σ'_{opt} . In our distributed setting, different sequences are observed at different terminals and sent to the decision maker using zero-rate compression. Hence, the decision maker only gets the information about the marginal empirical PMF of the observations. Consequently, we should search over all joint distributions that have the same marginal distributions with the ones in φ_r , which is the region \mathcal{H}_r . Therefore, the best type 2 error exponent is the solid line from $Q_{X_1 X_2 Y}$ to \mathcal{H}_r .

Now, we present a scheme that can achieve the type 2 error exponent characterized in Theorem 1. Instead of showing that $\sigma(0, 0, r) \geq \sigma_{opt}$ directly, we show that σ_{opt} is achievable in a transformed model. The original model with $L = 2$ is shown in Figure 3 (a), denoted as $S_{X_1 X_2 | Y}$. In the original model, the decision maker is located in terminal \mathcal{Y} , so it has full access

to Y^n . This model can also be viewed as a scenario in which the decision maker is located in a separate terminal and terminal \mathcal{Y} also sends encoded messages to the decision maker, but its rate R is so large ($R \geq \log |\mathcal{Y}|$) that the decision maker can fully recover Y^n . This new view is shown in Figure 3 (b). Therefore, the two system shown in Figure 3 (a) and (b) are equivalent, resulting in $\sigma(0, 0, r) = \sigma(0, 0, \log |\mathcal{Y}|, r)$. However, if the rate for terminal \mathcal{Y} is not large enough, such as $R = 0$, which is shown in Figure 3 (c), then the decision maker cannot fully recover Y^n , thus it has less information than the decision maker in Figure 3 (b), and yields a larger error probability. Hence, we have $\sigma(0, 0, r) = \sigma(0, 0, \log |\mathcal{Y}|, r) \geq \sigma(0, 0, 0, r)$. We denote the system in Figure 3 (b) and (c) as $S_{X_1 X_2 Y}$. If we can show that $\sigma(0, 0, 0, r) \geq \sigma_{opt}$ in $S_{X_1 X_2 Y}$, then we have $\sigma(0, 0, r) \geq \sigma_{opt}$ in $S_{X_1 X_2 | Y}$.

In the following, we will describe a scheme to show that $\sigma(0, 0, 0, r) \geq \sigma_{opt}$ in $S_{X_1 X_2 Y}$. Before proceeding to the formal proof, we first describe the high level idea of the scheme. After observing x_i^n , terminal \mathcal{X}_i knows the type $tp(x_i^n)$ and sends $tp(x_i^n)$ (or an approximation of it, see below) to the decision maker. Terminal \mathcal{Y} does the same. As there are at most $n^{|\mathcal{X}_i|}$ types [31], the rate required for sending the type from terminal \mathcal{X}_i is $(|\mathcal{X}_i| \log n)/n$, which goes to zero as n increases. After receiving all type information from the terminals, the decision maker will check whether there is a joint type $\tilde{P}_{X_1 X_2 Y} \in \mathcal{H}_r$ such that its marginal types are the same as the information received from the terminals. If yes, the decision maker declares H_0 to be true, otherwise declares H_1 to be true. If the message size M_i is less than $n^{|\mathcal{X}_i|}$, then instead of the exact type information $tp(x_i^n)$, each terminal will send an approximated version. Details on how to approximate the type will be provided in the proof. As long as $M_i \rightarrow \infty$, the approximation will be close (to be made precise in the proof) to the true type, and hence the decision maker can still use the above mentioned decision rule. We will show that this scheme can achieve σ_{opt} in $S_{X_1 X_2 Y}$.

The following theorem provides details about the above mentioned idea.

Theorem 2: For zero-rate compression in $S_{X_1 X_2 Y}$ with $R_1 = R_2 = R = 0$, the error exponent satisfies

$$\sigma(0, 0, 0, r) \geq \sigma_{opt} \quad (22)$$

where σ_{opt} is defined as (19).

Proof: First, define g -distance from any joint distribution to $P_{X_1 X_2 Y}$ as

$$g(\tilde{X}_1, \tilde{X}_2, \tilde{Y}) = \min_{\substack{\hat{P}_{X_1 X_2 Y} \\ \hat{P}_{X_1} = \tilde{P}_{X_1} \\ \hat{P}_{X_2} = \tilde{P}_{X_2} \\ \hat{P}_Y = \tilde{P}_Y}} D(\hat{P}_{X_1 X_2 Y} || P_{X_1 X_2 Y}) \quad (23)$$

which is continuous in $((\tilde{P}_{X_1})_{x_1 \in \mathcal{X}_1}, (\tilde{P}_{X_2})_{x_2 \in \mathcal{X}_2}, (\tilde{P}_Y)_{y \in \mathcal{Y}})$.

Next, divide the $(|\mathcal{X}_1| + |\mathcal{X}_2| + |\mathcal{Y}|)$ dimensional unit cube into equal-sized $M_1 \times M_2 \times M$ small cells with each edge of length κ_1 along the first $|\mathcal{X}_1|$ components, each edge of length κ_2 along the $|\mathcal{X}_2|$ components and each edge of length τ along the $|\mathcal{Y}|$ components, where

$$\kappa_1 = M_1^{-1/|\mathcal{X}_1|}, \quad \kappa_2 = M_2^{-1/|\mathcal{X}_2|}, \quad \tau = M^{-1/|\mathcal{Y}|},$$

in which

$$M_1 \rightarrow \infty, M_2 \rightarrow \infty, M \rightarrow \infty, \quad (24)$$

but $\log M_i/n \rightarrow 0$ for $i = 1, 2$ and $\log M/n \rightarrow 0$, as $n \rightarrow \infty$ (i.e., zero-rate compression for all three terminals).

Choose and fix a representative point in each cell for every set of variables $(\tilde{X}_1, \tilde{X}_2, \tilde{Y})$. Then in a given cell, we make its representative variable set $(\tilde{X}_1, \tilde{X}_2, \tilde{Y})$ correspond in such a way that $((\tilde{P}_{X_1})_{x_1 \in \mathcal{X}_1}, (\tilde{P}_{X_2})_{x_2 \in \mathcal{X}_2}, (\tilde{P}_Y)_{y \in \mathcal{Y}})$ is the representative point of $((\tilde{P}_{X_1})_{x_1 \in \mathcal{X}_1}, (\tilde{P}_{X_2})_{x_2 \in \mathcal{X}_2}, (\tilde{P}_Y)_{y \in \mathcal{Y}})$. For each terminal, after observing its sequence, determines its type and then finds the index of the corresponding edge. Each terminal then sends the index to the decision maker. After receiving all the indexes, the decision maker can determine the cell index. Since we have assumed (24), we see that with any $\eta > 0$

$$|\tilde{P}_{X_1} - \check{P}_{X_1}| < \eta, \quad x_1 \in \mathcal{X}_1, \quad (25)$$

$$|\tilde{P}_{X_2} - \check{P}_{X_2}| < \eta, \quad x_2 \in \mathcal{X}_2, \quad (26)$$

$$|\tilde{P}_Y - \check{P}_Y| < \eta, \quad y \in \mathcal{Y}, \quad (27)$$

for sufficiently large $n \geq n_0(\eta)$. Furthermore, the continuity of $g(\tilde{X}_1, \tilde{X}_2, \tilde{Y})$ in $(\tilde{X}_1, \tilde{X}_2, \tilde{Y})$ yields

$$|g(\tilde{X}_1, \tilde{X}_2, \tilde{Y}) - g(\check{X}_1, \check{X}_2, \check{Y})| < \eta. \quad (28)$$

Denoting by $(\check{X}_1^{(n)}, \check{X}_2^{(n)}, \check{Y}^{(n)})$ the representative point of $(X_1^{(n)}, X_2^{(n)}, Y^{(n)})$ where $X_1^{(n)}, X_2^{(n)}$ and $Y^{(n)}$ are the type variables of $x_1^n \in \mathcal{X}_1^n, x_2^n \in \mathcal{X}_2^n$ and $y^n \in \mathcal{Y}^n$ respectively, we set an acceptance region

$$\mathcal{A}_n = \left\{ (x_1^n, x_2^n, y^n) : g(\check{X}_1^{(n)}, \check{X}_2^{(n)}, \check{Y}^{(n)}) \leq r + 2\eta \right\}.$$

More precisely, our decoding scheme is as follows. Upon receiving (M_1, M_2, M) , find the representative point and its joint distribution. Then calculate the g -distance from this joint distribution to $P_{X_1 X_2 Y}$. If the g -distance is less than or equal to $r + 2\eta$, then we decide H_0 is true and vice versa. In other words, we first find the region of joint distributions which has a g -distance to $P_{X_1 X_2 Y}$ less than or equal to $r + 2\eta$, which is visualized in Figure 4 as $\mathcal{H}_{r+2\eta}$. Then after knowing the joint distribution of the representative point, we can tell whether it is in $\mathcal{H}_{r+2\eta}$ or not. If it is in $\mathcal{H}_{r+2\eta}$, we decide H_0 is true and vice versa. In Figure 4, we use a square region to denote all possible joint distributions of the representative points.

Now we analyze the two types of error probability. For any $\rho > 0$ set

$$\xi_\rho = \{(x_1^n, x_2^n, y^n) : g(X_1^{(n)}, X_2^{(n)}, Y^{(n)}) \leq \rho\};$$

then in view of (28) it is clear that

$$\xi_{r+\eta} \subset \mathcal{A}_n \subset \xi_{r+3\eta} \quad (29)$$

It is easy to see that $(x_1^n, x_2^n, y^n) \in \xi_{r+\eta}$ if $(x_1^n, x_2^n, y^n) \in S_{r+\eta}^{n-1}(X_1 X_2 Y)$, that is $S_{r+\eta}^{n-1}(X_1 X_2 Y) \subset \xi_{r+\eta}$, which yields

$$1 - \alpha_n = P_{X_1 X_2 Y}^n(\mathcal{A}_n) \geq 1 - \exp(-nr)$$

for n large enough. Hence, the constraint (9) is satisfied.

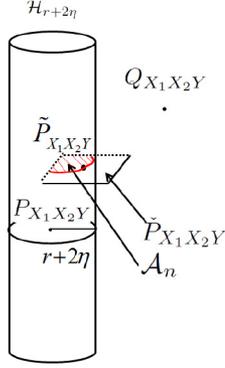


Fig. 4. Visualization of acceptance region

On the other hand, from the second inclusion in (29),

$$\begin{aligned}
\beta_n &= Q_{X_1 X_2 Y}^n(\mathcal{A}_n) \\
&\leq Q_{X_1 X_2 Y}^n(\xi_{r+3\eta}) \\
&\leq \sum_{X_1^{(n)}, X_2^{(n)}, Y^{(n)}} \exp(-nD(X_1^{(n)} X_2^{(n)} Y^{(n)} \| Q_{X_1 X_2 Y})) \\
&\quad g(X_1^{(n)}, X_2^{(n)}, Y^{(n)}) \leq r + 3\eta \\
&\leq (n+1)^{|\mathcal{X}_1| |\mathcal{X}_2| |\mathcal{Y}|} \\
&\quad \max_{X_1^{(n)}, X_2^{(n)}, Y^{(n)}} \exp(-nD(X_1^{(n)} X_2^{(n)} Y^{(n)} \| Q_{X_1 X_2 Y})) \\
&\quad g(X_1^{(n)}, X_2^{(n)}, Y^{(n)}) \leq r + 3\eta \\
&\leq (n+1)^{|\mathcal{X}_1| |\mathcal{X}_2| |\mathcal{Y}|} \\
&\quad \exp \left[-n \left\{ \min_{\substack{\tilde{X}_1, \tilde{X}_2, \tilde{Y} \\ g(\tilde{X}_1, \tilde{X}_2, \tilde{Y}) \leq r + 3\eta}} D(\tilde{X}_1 \tilde{X}_2 \tilde{Y} \| Q_{X_1 X_2 Y}) \right\} \right].
\end{aligned}$$

Therefore,

$$\sigma(0, 0, 0, r) \geq \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{H}_{r+3\eta}} D(\tilde{P}_{X_1 X_2 Y} \| Q_{X_1 X_2 Y}),$$

which establishes (22) if we let $\eta \rightarrow 0$. \blacksquare

As $\sigma(0, 0, r) = \sigma(0, 0, \log |\mathcal{Y}|, r)$. From Theorem 2, we have

$$\begin{aligned}
\sigma(0, 0, r) &= \sigma(0, 0, \log |\mathcal{Y}|, r) \\
&\geq \sigma(0, 0, 0, r) \geq \sigma_{opt}.
\end{aligned}$$

Coupled with Theorem 1, we have:

Theorem 3: Let $P_{X_1 X_2 Y}$ be arbitrary and $Q_{X_1 X_2 Y} > 0$. For zero-rate compression in $S_{X_1 X_2 | Y}$ with $R_1 = R_2 = 0$ and type 1 error constraint (9), the best type 2 error exponent

$$\sigma(0, 0, r) = \sigma_{opt}, \quad (30)$$

where σ_{opt} is defined as (19).

Proposition 4: Given $P_{X_1 X_2 Y}$ and $Q_{X_1 X_2 Y}$, the problem of finding σ_{opt} defined in (19) is a convex optimization problem.

Proof: First, given $Q_{X_1 X_2 Y}$, it is easy to verify that the objective function $D(\tilde{P}_{X_1 X_2 Y} \| Q_{X_1 X_2 Y})$ in (19) is a convex function of $\tilde{P}_{X_1 X_2 Y}$.

Then, we show that the feasible set \mathcal{H}_r defined in (20) is also convex. Suppose $\tilde{P}'_{X_1 X_2 Y} \in \mathcal{H}_r$ and $\tilde{P}''_{X_1 X_2 Y} \in \mathcal{H}_r$, and

$\tilde{P}'_{X_1 X_2 Y}$ has the same marginal PMFs with $\hat{P}'_{X_1 X_2 Y} \in \varphi_r$, and $\tilde{P}''_{X_1 X_2 Y}$ has the same marginal PMFs $\hat{P}''_{X_1 X_2 Y} \in \varphi_r$. Setting

$$\tilde{P}'''_{X_1 X_2 Y} = \pi \tilde{P}'_{X_1 X_2 Y} + (1 - \pi) \tilde{P}''_{X_1 X_2 Y},$$

for $0 \leq \pi \leq 1$, we will show that $\tilde{P}'''_{X_1 X_2 Y} \in \mathcal{H}_r$, i.e. \mathcal{H}_r is a convex set. As we have

$$\tilde{P}'_{X_1} = \pi \tilde{P}'_{X_1} + (1 - \pi) \tilde{P}''_{X_1} = \pi \hat{P}'_{X_1} + (1 - \pi) \hat{P}''_{X_1},$$

and similar results with \tilde{P}'''_{X_2} and \tilde{P}'''_Y , we can conclude that $\tilde{P}'''_{X_1 X_2 Y}$ has the same marginal distribution as $\pi \hat{P}'_{X_1 X_2 Y} + (1 - \pi) \hat{P}''_{X_1 X_2 Y}$. Due to the convexity of $D(\hat{P}_{X_1 X_2 Y} \| P_{X_1 X_2 Y})$ with respect to $\hat{P}_{X_1 X_2 Y}$ for a given $P_{X_1 X_2 Y}$, we have $(\pi \hat{P}'_{X_1 X_2 Y} + (1 - \pi) \hat{P}''_{X_1 X_2 Y}) \in \varphi_r$. This implies that $\tilde{P}'''_{X_1 X_2 Y} \in \mathcal{H}_r$, and hence \mathcal{H}_r is a convex set.

As the result, for any given $P_{X_1 X_2 Y}$ and $Q_{X_1 X_2 Y}$, characterizing σ_{opt} is a convex optimization problem and can be solved efficiently. \blacksquare

C. General Case

The results of the previous section can be extended to the general case with L terminals. We have the following theorem, whose proof follows the similar steps as in those in Section III-B and hence is omitted for conciseness.

Theorem 5: Let $P_{X_1, \dots, X_L Y}$ be arbitrary and $Q_{X_1, \dots, X_L Y} > 0$. For zero-rate compression in $S_{X_1, \dots, X_L | Y}$ with $R_i = 0$, $i = 1, \dots, L$ and type 1 error constraint (9), the best type 2 error exponent

$$\begin{aligned}
\sigma(0, \dots, 0, r) &= \min_{\tilde{P}_{X_1 \dots X_L Y} \in \mathcal{H}_r} D(\tilde{P}_{X_1 \dots X_L Y} \| Q_{X_1 \dots X_L Y}) \quad (31)
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{H}_r &= \left\{ \tilde{P}_{X_1 \dots X_L Y} : \tilde{P}_{X_i} = \hat{P}_{X_i}, \tilde{P}_Y = \hat{P}_Y, i = 1, \dots, L \right. \\
&\quad \left. \text{for some } \hat{P}_{X_1 \dots X_L Y} \in \varphi_r \right\}, \quad (32)
\end{aligned}$$

$$\varphi_r = \left\{ \hat{P}_{X_1 \dots X_L Y} : D(\hat{P}_{X_1 \dots X_L Y} \| P_{X_1 \dots X_L Y}) \leq r \right\}. \quad (33)$$

Similar to (19), characterizing (31) is a convex optimization problem, hence it can be solved efficiently.

IV. TESTING AGAINST INDEPENDENCE WITH CONSTANT-TYPE CONSTRAINTS

In this section, we consider the scenario with positive communication rate constraints, i.e., $R_i > 0$, under the constant-type constraint. As the general case is a very complex problem even for $L = 1$ [27], we focus on the testing against independence case in which we are interested in determining whether X_1, \dots, X_L and Y are independent or not. Hence, the two hypotheses are

$$H_0 : P_{X_1 \dots X_L Y}, \quad H_1 : Q_{X_1 \dots X_L Y} = P_{X_1 \dots X_L} P_Y.$$

To facilitate the presentation, in the following, we only provide details for the $L = 2$ case. The results can be extended to the general L case with proper modifications. For $L = 2$, our goal is to characterize $\theta(R_1, R_2, \epsilon)$ under $\alpha_n \leq \epsilon$ and communication constraints (3).

Compared with the zero-rate compression case discussed in Section III, in this scenario, each terminal can convey more information to the decision maker as the communication rate constraint $R_i > 0$ is less strict. Before presenting the formal proof, we first describe high level ideas on how to exploit the more flexible rate constraints (terms in the following will be made precise in the proof). For a given rate constraint R_i , terminal \mathcal{X}_i first generates a quantization codebook containing 2^{nR_i} quantization sequences. After observing x_i^n , terminal \mathcal{X}_i picks one sequence u_i^n from the quantization codebook to describe x_i^n and sends this sequence to the decision maker. After receiving the descriptions from terminals, the decision maker will declare that the hypothesis H_0 is true if the descriptions from these terminals and the side-information at the decision maker are correlated. Otherwise, the decision maker will declare H_1 . The following theorem provides details of the scheme and error probability analysis.

Theorem 6: In system $S_{X_1 X_2 | Y}$ with $R_i > 0$, $i = 1, 2$, constraint on type 1 error probability (8) and communication constraints (3), the error exponent of the type 2 error probability is lower bounded by

$$\theta(R_1, R_2, \epsilon) \geq \max_{P_{U_1|X_1} P_{U_2|X_2}} I(U_1 U_2; Y), \quad (34)$$

in which the maximization is over $P_{U_i|X_i}$'s such that $I(U_i; X_i) \leq R_i$ and $|U_i| \leq |X_i| + 1$.

Proof: In the following, $\epsilon > \epsilon' > \epsilon'' > \epsilon'''$ are given small numbers.

Quantization codebook generation. Fix a conditional PMF $P_{U_1 U_2 | X_1 X_2 Y} = P_{U_1 | X_1} P_{U_2 | X_2}$ that attains the maximum in (34). Let $P_{U_1}(u_1) = \sum_{x_1} P_{X_1}(x_1) P_{U_1 | X_1}(u_1 | x_1)$ and $P_{U_2}(u_2) = \sum_{x_2} P_{X_2}(x_2) P_{U_2 | X_2}(u_2 | x_2)$. Randomly and independently generate 2^{nR_1} sequences $u_1^n(m_1)$, $m_1 \in \{1, \dots, 2^{nR_1}\}$ each according to $\prod_{i=1}^n P_{U_1}(u_{1i})$. Randomly and independently generate 2^{nR_2} sequences $u_2^n(m_2)$, $m_2 \in \{1, \dots, 2^{nR_2}\}$ each according to $\prod_{i=1}^n P_{U_2}(u_{2i})$. These sequences constitute the codebook \mathcal{C} , which is revealed to all terminals.

Encoding (Quantization). After observing sequence x_1^n , terminal \mathcal{X}_1 finds a $u_1^n(m_1)$ such that $(x_1^n, u_1^n(m_1)) \in T_{\epsilon'''}^{(n)}$, and sends the index m_1 to terminal \mathcal{Y} . 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 $[1, \dots, 2^{nR_1}]$ uniformly at random. Similarly, after observing a sequence x_2^n , terminal \mathcal{X}_2 finds a $u_2^n(m_2)$ such that $(x_2^n, u_2^n(m_2)) \in T_{\epsilon'''}^{(n)}$, then it sends the index m_2 to terminal \mathcal{Y} . 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 $[1, \dots, 2^{nR_2}]$ uniformly at random.

Testing. Upon receiving m_1 and m_2 , terminal \mathcal{Y} sets the acceptance region \mathcal{A}_n for H_0 to

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

where the jointly typical set $T_{\epsilon}^{(n)}$ is defined with respect to $P_{X_1 X_2 Y}$, $P_{U_1 | X_1}$ and $P_{U_2 | X_2}$.

Error probability analysis. Terminal \mathcal{Y} chooses $\hat{H} \neq H_0$ if and only if one or more of the following events occur:

$$\begin{aligned} \varepsilon_1 &= \left\{ (U_1^n(m_1), X_1^n) \notin T_{\epsilon'''}^{(n)} \text{ for all } m_1 \in [1 : 2^{nR_1}] \right\}, \\ \varepsilon_2 &= \left\{ (U_2^n(m_2), X_2^n) \notin T_{\epsilon'''}^{(n)} \text{ for all } m_2 \in [1 : 2^{nR_2}] \right\}, \\ \varepsilon_3 &= \left\{ (U_1^n(M_1), U_2^n(M_2), Y^n) \notin T_{\epsilon}^{(n)} \right\}. \end{aligned}$$

Hence, $\mathcal{A}_n = (\varepsilon_1 \cup \varepsilon_2 \cup \varepsilon_3)^c$.

After tedious and lengthy computation, we can show that

$$\alpha_n = P_{X_1 X_2 Y}^n(\mathcal{A}_n^c) \leq \epsilon, \quad (35)$$

$$\beta_n = Q_{X_1 X_2 Y}^n(\mathcal{A}_n) \leq 2^{-n(I(U_1 U_2; Y) - \delta(\epsilon))}, \quad (36)$$

if the conditions specified in the theorem are satisfied. Details of the error probability analysis can be found in Appendix B. Hence, we have (34). \blacksquare

Finally, we establish an upper bound on the type 2 error exponent that any scheme can achieve.

Theorem 7: In system $S_{X_1 X_2 | Y}$ with $R_i \geq 0$, $i = 1, 2$, constraint on type 1 error probability (8) and communication constraints (3), the best error exponent for type 2 error probability

$$\lim_{\epsilon \rightarrow 0} \theta(R_1, R_2, \epsilon) \leq \max_{U_1 U_2} I(U_1 U_2; Y) \quad (37)$$

in which the maximization is over U_i 's such that $R_i \geq I(U_i; X_i)$, $|U_i| \leq |X_i| + 1$, $U_1 \rightarrow X_1 \rightarrow (X_2, Y)$ and $U_2 \rightarrow X_2 \rightarrow (X_1, Y)$.

Proof: We will show that for any encoding and decoding scheme that satisfies the type 1 error constraint $\alpha_n \leq \epsilon$ and rate constraints (3), the type 2 error exponent must satisfy (37).

First, for any scheme that satisfies the type 1 error and rate constraints, we have

$$\begin{aligned} & D(P_{M_1 M_2 Y^n} \| P_{M_1 M_2} P_{Y^n}) \\ &= \sum_{(m_1, m_2, y^n) \in \mathcal{A}_n} P_{M_1 M_2 Y^n} \log \frac{P_{M_1 M_2 Y^n}}{P_{M_1 M_2} P_{Y^n}} \\ &+ \sum_{(m_1, m_2, y^n) \in \mathcal{A}_n^c} P_{M_1 M_2 Y^n} \log \frac{P_{M_1 M_2 Y^n}}{P_{M_1 M_2} P_{Y^n}} \\ &\stackrel{(a)}{\geq} (1 - \alpha_n) \log \frac{1 - \alpha_n}{\beta_n} + \alpha_n \log \frac{\alpha_n}{1 - \beta_n} \\ &= (1 - \alpha_n) \log \frac{1}{\beta_n} + \alpha_n \log \frac{1}{1 - \beta_n} - H(\alpha_n) \\ &\geq (1 - \alpha_n) \log \frac{1}{\beta_n} - H(\alpha_n) \\ &\stackrel{(b)}{\geq} (1 - \epsilon) \log \frac{1}{\beta_n} - H(\alpha_n). \end{aligned}$$

where $M_i = f_i(X_i^n)$, $i = 1, 2$, α_n and β_n are defined in (6) and (7), and $H(\alpha_n)$ is

$$H(\alpha_n) \triangleq -(1 - \alpha_n) \log(1 - \alpha_n) - \alpha_n \log \alpha_n. \quad (38)$$

In the derivation above, (a) is true due to the log sum inequality [22] and (b) follows by the constraint (8).

Hence we have the following upper bound

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \theta(R_1, R_2, \epsilon) &\leq \lim_{n \rightarrow \infty} \frac{1}{n} D(P_{M_1 M_2 Y^n} \| P_{M_1 M_2} P_{Y^n}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} I(M_1 M_2; Y^n) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} (H(Y^n) - H(Y^n | M_1 M_2)) \\ &= H(Y) - \lim_{n \rightarrow \infty} \frac{1}{n} H(Y^n | M_1 M_2). \quad (39) \end{aligned}$$

If we simplify $\frac{1}{n} H(Y^n | M_1 M_2)$, we obtain the desired bound. In Appendix C, we show that

$$\frac{1}{n} H(Y^n | M_1 M_2) = H(Y | U_1 U_2),$$

for properly chosen $U_1 U_2$ satisfying the conditions specified in the statement of theorem. Combing this with (39), we obtain the desired result. ■

V. APPLICATION

In the section, we provide some examples and numerical results to illustrate the application of the results developed in Section III and Section IV.

A. Classical Examples

In this subsection, we provide two examples to show that the approach established in this paper can be applied to the classical cases with the conditional independence assumption, i.e., testing with hypotheses in (10) and (11).

1) *Example for Testing with Zero-Rate Compression:* We can apply our theory on zero-rate compression under the exponential-type constraint on type 1 error probability to the case with the conditional independence assumption.

Under the conditional independence assumption, $P_{X_1 \dots X_L Y} = P_{X_1} \dots P_{X_L} P_Y$ and $Q_{X_1 \dots X_L Y} = Q_{X_1} \dots Q_{X_L} Q_Y$, which is a special case of the general hypotheses $P_{X_1 \dots X_L Y}$ and $Q_{X_1 \dots X_L Y}$. Hence, Theorem 5 still holds true for testing under conditionally independent hypotheses.

Corollary 8: Let $P_{X_1} \dots P_{X_L} P_Y$ be arbitrary and $Q_{X_1} \dots Q_{X_L} Q_Y > 0$. For zero-rate compression in $S_{X_1 \dots X_L | Y}$ with $R_i = 0, i = 1, \dots, L$ and type 1 error constraint (9), the best type 2 error exponent

$$\begin{aligned} \sigma(0, \dots, 0, r) \\ = \min_{\tilde{P}_{X_1 \dots X_L Y} \in \mathcal{H}_r} D(\tilde{P}_{X_1 \dots X_L Y} \| Q_{X_1} \dots Q_{X_L} Q_Y) \quad (40) \end{aligned}$$

where

$$\mathcal{H}_r = \left\{ \tilde{P}_{X_1 \dots X_L Y} : \tilde{P}_{X_i} = \hat{P}_{X_i}, \tilde{P}_Y = \hat{P}_Y, i = 1, \dots, L \right. \\ \left. \text{for some } \hat{P}_{X_1 \dots X_L Y} \in \varphi_r \right\}, \quad (41)$$

$$\varphi_r = \left\{ \hat{P}_{X_1 \dots X_L Y} : D(\hat{P}_{X_1 \dots X_L Y} \| P_{X_1} \dots P_{X_L} P_Y) \leq r \right\}. \quad (42)$$

We note that the scalar quantizer will send at least one bit for each observation (i.e., the communication rate is at least 1), hence the scalar quantizer cannot handle this zero-rate compression scenario. As shown in Corollary 8, the use of vector quantizer enables us to extend the classic example to the zero-rate compression scenario.

2) *Example for Testing With Positive Communication Rate:* We can apply our coding scheme to the case with a positive communication rate under the conditional independence assumption. To facilitate our presentation, we only provide details for $L = 2$ case.

Corollary 9: Under conditional independence assumption, in system $S_{X_1 X_2 | Y}$ with $R_i \geq 0, i = 1, 2$, constraint on type 1 error probability (8) and communication constraints (3), the error exponent of the type 2 error probability is lower bounded by

$$\begin{aligned} \theta(R_1, R_2, \epsilon) &\geq \max_{P_{U_1 | X_1} P_{U_2 | X_2} \in \varphi_0} \\ &\min_{\tilde{P}_{U_1 U_2 X_1 X_2 Y} \in \xi_0} \left\{ D\left(\tilde{P}_{U_1 U_2 X_1 X_2 Y} \| Q_{U_1 U_2 X_1 X_2 Y}\right) \right\} \quad (43) \end{aligned}$$

where φ_0 is

$$\begin{aligned} \varphi_0 = \{ P_{U_1 | X_1} P_{U_2 | X_2} : R_1 \geq I(U_1; X_1), R_2 \geq I(U_2; X_2), \\ |\mathcal{U}_1| \leq |\mathcal{X}_1| + 1, |\mathcal{U}_2| \leq |\mathcal{X}_2| + 1 \}, \end{aligned}$$

and ξ_0 is

$$\begin{aligned} \xi_0 = \left\{ \tilde{P}_{U_1 U_2 X_1 X_2 Y} : \tilde{P}_{U_1 X_1} = P_{U_1 X_1}, \right. \\ \tilde{P}_{U_2 X_2} = P_{U_2 X_2}, \\ \left. \tilde{P}_{U_1 U_2 Y} = P_{U_1 U_2 Y} \right\}. \quad (44) \end{aligned}$$

Proof: The proof can be obtained by properly modifying the proof of Theorem 6. The details are provided in Appendix D for completeness. ■

We note that the scalar quantizer can be applied to the positive rate case under the conditional independence assumption. In fact, the optimal scalar quantizer that minimizes the error probability is known. However, the optimal scalar quantizer relies on an optimization over thresholds used for the quantization, which needs to be carried out numerically for different problems. Due to this reason, it is challenging to obtain error exponent formulas for the optimal scalar quantizer.

B. Numerical Results

In this section, we provide several numerical examples to illustrate the results obtained in Section III and Section IV.

1) *Numerical Results for Testing With Zero-rate Compression Under Exponential-Type Constraints:* In Figure 5, we illustrate σ_{opt} , namely the optimal type 2 error exponent characterized in Theorem 3, as a function of the type 1 error exponent constraint r . For comparison, we also plot the corresponding curve for the centralized case. In the figure, the solid line represents σ_{opt} and the dashed line is the optimal type 2 error exponent for the centralized case. In generating

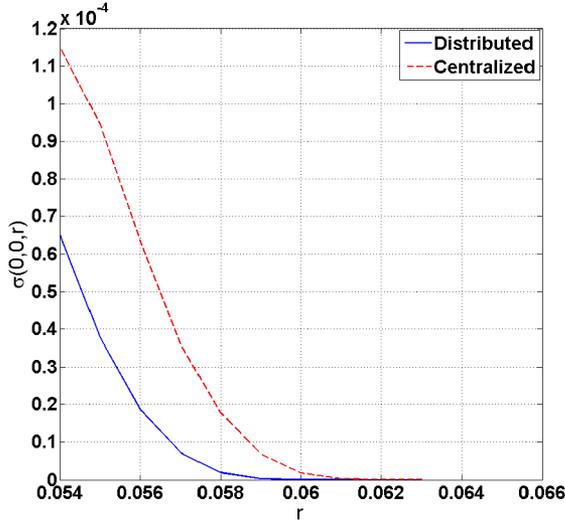

 Fig. 5. $\sigma(0,0,r)$ vs r with $D(P_{X_1 X_2 Y} || Q_{X_1 X_2 Y}) = 0.0624$

Figure 5, we set X_1 , X_2 and Y as binary random variables. Furthermore, we set

$$P_{X_1 X_2 Y} = \left\{ \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8} \right\}$$

and

$$Q_{X_1 X_2 Y} = \left\{ \frac{1}{12}, \frac{1}{12}, \frac{5}{72}, \frac{7}{72}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \right\}.$$

It is easy to verify that $D(P_{X_1 X_2 Y} || Q_{X_1 X_2 Y}) = 0.0624$. From Figure 5, we can see that the type 2 error exponent obtained in the distributed case is smaller than that of the centralized case for every r . This is reasonable as in the centralized case, the decision maker has full access to all observations and hence makes less error. Furthermore, the type 2 error exponents for both settings are close to 0 when $r > 0.062$, which makes sense as when $r > D(P_{X_1 X_2 Y} || Q_{X_1 X_2 Y})$, no matter what observation is observed, the decision maker decides H_0 is true.

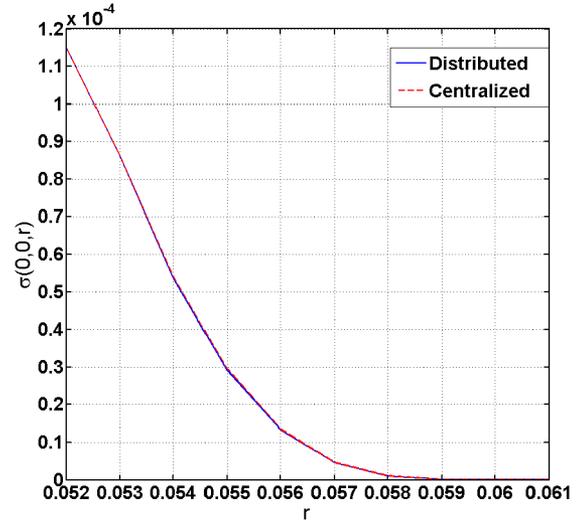
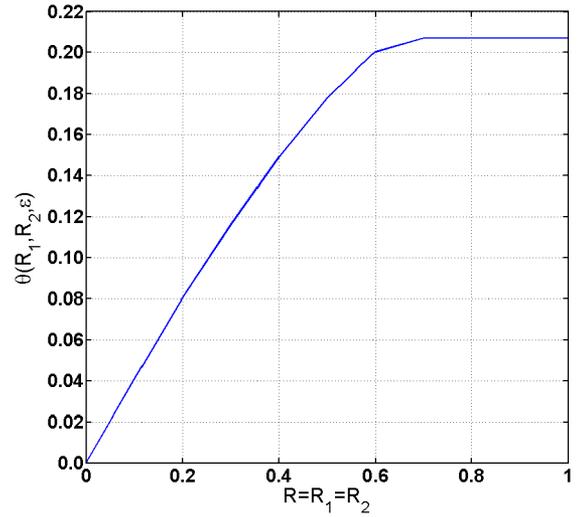
Figure 6 illustrates σ_{opt} for different PMFs. In generating Figure 6, we keep $P_{X_1 X_2 Y}$ same as above, but change $Q_{X_1 X_2 Y}$ to

$$Q_{X_1 X_2 Y} = \left\{ \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \right\}.$$

In this case, $D(P_{X_1 X_2 Y} || Q_{X_1 X_2 Y}) = 0.0588$. From Figure 6, we can see that the type 2 error exponent obtained in the distributed setting is quite close to that of the centralized case. This implies that, for certain PMFs, the distributed setting with a proper zero-rate compression can achieve a performance close to that of the centralized setting.

2) *Numerical Results for Testing Against Independence Under Constant-Type Constraints:* In Figure 7, we illustrate $\theta(R_1, R_2, \epsilon)$ discussed in Theorem 6 as a function of the rate constraints. In generating this figure, we again set X_1 , X_2 and Y to be binary random variables and set

$$P_{X_1 X_2 Y} = \left\{ \frac{1}{6}, \frac{1}{3}, \frac{1}{12}, \frac{1}{6}, 0, 0, \frac{1}{8}, \frac{1}{8} \right\},$$


 Fig. 6. $\sigma(0,0,r)$ vs r with $D(P_{X_1 X_2 Y} || Q_{X_1 X_2 Y}) = 0.0588$

 Fig. 7. $\theta(R_1, R_2, \epsilon)$ vs $R = R_1 = R_2$ with $D(P_{X_1 X_2 Y} || Q_{X_1 X_2 Y}) = 0.2229$

from which one can calculate $Q_{X_1 X_2 Y} = P_{X_1 X_2} P_Y$. Furthermore, to make the computation feasible, we assume $|\mathcal{U}_i| = |\mathcal{X}_i| = 2$ in the simulation. In order to visualize the result better, we make $R_1 = R_2 = R$. Hence, we demonstrate a lower bound on the type 2 error exponent achievable using our scheme.

From Figure 7, we can see that the type 2 error exponent increases as R increases, which makes sense as the constraint is relaxed, the decision maker can get more information about X_1^n and X_2^n , and thus make less error. Furthermore, when R is large enough, the decision maker can fully recover X_1^n and X_2^n , which is then the same as the centralized setting. According to Stein's lemma, in the centralized setting, the type 2 error exponent equals $D(P_{X_1 X_2 Y} || P_{X_1 X_2} P_Y)$. In our simulation, $D(P_{X_1 X_2 Y} || P_{X_1 X_2} P_Y) = 0.2229$, and we can see that the maximum value in Figure 7 is quite close to 0.2229.

VI. CONCLUSION

In this paper, we have discussed distributed inference problems with vector quantizers. Using properties of r -divergence sequences, we have characterized the best error exponent of the type 2 error probability under the zero-rate compression and exponential-type type 1 error probability constraints. Furthermore, we have discussed the problem of testing against independence under the constant-type constraint on the type 1 error probability. We have derived a lower bound and upper-bound on the type 2 error exponent.

APPENDIX A
PROOF OF THEOREM 1

In this appendix, we present the proof of Theorem 1. In this proof, we need to show that for any encoding and decoding scheme that meets the type 1 error constraint, we have (18). Let \mathcal{A}_n be an arbitrary acceptance region such that

$$\alpha_n \leq \exp(-nr), \quad r > 0 \quad (45)$$

where

$$\alpha_n = P_{X_1 X_2 Y}^n(\mathcal{A}_n^c). \quad (46)$$

Equations (45) and (46) imply that

$$P_{X_1 X_2 Y}^n(\mathcal{A}_n) \geq 1 - \exp(-n(r - \gamma)), \quad \forall n \geq n_0, \quad (47)$$

where $\gamma > 0$ is an arbitrarily small constant, and n_0 is a sufficiently large positive integer.

Next, select an arbitrary ‘‘internal point’’ $P_{X_{10} X_{20} Y_0} \in \varphi_r$, where φ_r is specified in (21). Then clearly

$$D(P_{X_{10} X_{20} Y_0} \| P_{X_1 X_2 Y}) < r. \quad (48)$$

Define

$$\hat{T}_n(\delta) = \left\{ \text{joint types } \hat{P}_n \text{ on } \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}^n : \right. \\ \left. D\left(\hat{P}_n \| P_{X_{10} X_{20} Y_0}\right) < \delta \right\} \quad (49)$$

where $\delta > 0$ is an arbitrary constant. Then, in view of (48) and the uniform continuity of the divergence, for all $\hat{P}_n \in \hat{T}_n(\delta)$ it holds that

$$c_n \equiv D\left(\hat{P}_n \| P_{X_1 X_2 Y}\right) < r - 2\gamma, \quad (50)$$

provided that we take $\gamma > 0$ and $\delta > 0$ sufficiently small. Consequently, according to Lemma 1, we have

$$\left| \mathcal{A}_n(\hat{P}_n) \right| \geq (1 - (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} \exp(-n\gamma)) \left| S_0(\hat{P}_n) \right| \quad (51)$$

for all $\hat{P}_n \in \hat{T}_n(\delta)$. Now we define the set

$$T_n(\delta) = \left\{ (x_1^n, x_2^n, y^n) \in \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}^n : \right. \\ \left. X_1^{(n)} X_2^{(n)} Y^{(n)} \in \hat{T}_n(\delta) \right\} \quad (52)$$

and consider an i.i.d. random sequence of length n generated according to the probability distribution $P_{X_{10} X_{20} Y_0}$. Then, from (51), we have

$$P_{X_{10} X_{20} Y_0}^n(\mathcal{A}_n) \\ \geq P_{X_{10} X_{20} Y_0}^n(\mathcal{A}_n \cap T_n(\delta))$$

$$\begin{aligned} &= \sum_{\hat{P}_n \in \hat{T}_n(\delta)} P_{X_{10} X_{20} Y_0}^n(\mathcal{A}_n \cap S_0(\hat{P}_n)) \\ &= \sum_{\hat{P}_n \in \hat{T}_n(\delta)} P_{X_{10} X_{20} Y_0}^n(\mathcal{A}_n(\hat{P}_n)) \\ &= \sum_{\hat{P}_n \in \hat{T}_n(\delta)} \sum_{\substack{tp(x_{10}^n, x_{20}^n, y_0^n) = \hat{P}_n \\ (x_{10}^n, x_{20}^n, y_0^n) \in \mathcal{A}_n}} P_{X_{10} X_{20} Y_0}^n(X_{10}^n = x_{10}^n, \\ &\quad X_{20}^n = x_{20}^n, Y_0^n = y_0^n) \\ &\stackrel{(a)}{=} \sum_{\hat{P}_n \in \hat{T}_n(\delta)} \sum_{\substack{tp(x_{10}^n, x_{20}^n, y_0^n) = \hat{P}_n \\ (x_{10}^n, x_{20}^n, y_0^n) \in \mathcal{A}_n}} \exp \left[-n \left(H \left(X_{10}^{(n)} X_{20}^{(n)} Y_0^{(n)} \right) \right. \right. \\ &\quad \left. \left. + D \left(X_{10}^{(n)} X_{20}^{(n)} Y_0^{(n)} \| X_{10} X_{20} Y_0 \right) \right) \right] \\ &= \sum_{\hat{P}_n \in \hat{T}_n(\delta)} \left| \mathcal{A}_n(\hat{P}_n) \right| \exp \left[-n \left(H \left(X_{10}^{(n)} X_{20}^{(n)} Y_0^{(n)} \right) \right. \right. \\ &\quad \left. \left. + D \left(X_{10}^{(n)} X_{20}^{(n)} Y_0^{(n)} \| X_{10} X_{20} Y_0 \right) \right) \right] \\ &\geq \sum_{\hat{P}_n \in \hat{T}_n(\delta)} \left(1 - (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} \exp(-n\gamma) \right) \left| S_0(\hat{P}_n) \right| \\ &\quad \exp \left[-n \left(H \left(X_{10}^{(n)} X_{20}^{(n)} Y_0^{(n)} \right) \right. \right. \\ &\quad \left. \left. + D \left(X_{10}^{(n)} X_{20}^{(n)} Y_0^{(n)} \| X_{10} X_{20} Y_0 \right) \right) \right] \\ &\geq \left(1 - (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} \exp(-n\gamma) \right) \\ &\quad \sum_{\hat{P}_n \in \hat{T}_n(\delta)} P_{X_{10} X_{20} Y_0}^n(S_0(\hat{P}_n)) \\ &= \left(1 - (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} \exp(-n\gamma) \right) P_{X_{10} X_{20} Y_0}(\hat{T}_n(\delta)) \\ &\geq \left(1 - (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} \exp(-n\gamma) \right) \\ &\quad \times \left(1 - (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} \exp(-n\delta) \right), \quad (53) \end{aligned}$$

where (a) is true due to (15), and the last step is true due to (17).

Now consider the zero-rate ($R_1 = 0, R_2 = 0, R \geq 0$) hypothesis testing problem with

$$H_0 : P_{X_{10} X_{20} Y_0} \quad \text{vs} \quad H_1 : Q_{X_1 X_2 Y}. \quad (54)$$

Then, for this hypothesis testing problem, if we use the same acceptance region \mathcal{A}_n as above, the type 1 error probability

$$\begin{aligned} \alpha_n^{(0)} &= 1 - P_{X_{10} X_{20} Y_0}^n(\mathcal{A}_n) \\ &\leq 1 - \left(1 - (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} \exp(-n\gamma) \right) \\ &\quad \times \left(1 - (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} \exp(-n\delta) \right) \\ &\leq \epsilon, \end{aligned}$$

where ϵ is the constant-type constraint on the type 1 error probability.

Hence, for the hypothesis testing problem (54), the acceptance region \mathcal{A}_n satisfies the constant-type type 1 error probability constraint.

From [25], we know that the type 2 error exponent

$$\theta(0, 0, \epsilon) \leq \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{L}_0} D(\tilde{P}_{X_1 X_2 Y} \| Q_{X_1 X_2 Y}), \quad (55)$$

where

$$\mathcal{L}_0 = \left\{ \tilde{P}_{X_1 X_2 Y} : \tilde{P}_{X_1} = P_{X_{10}}, \tilde{P}_{X_2} = P_{X_{20}}, \tilde{P}_Y = P_{Y_0} \right\}.$$

On the other hand, we note that $P_{X_{10} X_{20} Y_0}$ was arbitrary as long as condition (48) is satisfied. Therefore, in the light of the definition of \mathcal{H}_r , we see that the infimum of the right-hand side in (55) over all possible internal points $P_{X_{10} X_{20} Y_0}$ satisfying (48) coincides with

$$\min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{H}_r} D(\tilde{P}_{X_1 X_2 Y} \| Q_{X_1 X_2 Y}).$$

Thus (55) reduces to

$$\sigma(0, 0, r) \leq \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{H}_r} D(\tilde{P}_{X_1 X_2 Y} \| Q_{X_1 X_2 Y}).$$

APPENDIX B

ERROR PROBABILITY ANALYSIS FOR THEOREM 6

To analyze the type 1 error probability, we have

$$\begin{aligned} \alpha_n &= P_{X_1 X_2 Y}^n(\mathcal{A}_n^c) \\ &= P_{X_1 X_2 Y}^n(\varepsilon_1 \cup \varepsilon_2 \cup \varepsilon_3) \\ &\leq P_{X_1 X_2 Y}^n(\varepsilon_1) + P_{X_1 X_2 Y}^n(\varepsilon_2) + P_{X_1 X_2 Y}^n(\varepsilon_1^c \cap \varepsilon_2^c \cap \varepsilon_3). \end{aligned}$$

We now bound each term.

1) By the covering lemma [22, Section 3.7],

$$P_{X_1 X_2 Y}^n(\varepsilon_1) \rightarrow 0$$

as $n \rightarrow \infty$ if

$$R_1 \geq I(U_1; X_1) + \delta(\epsilon)$$

and

$$P_{X_1 X_2 Y}^n(\varepsilon_2) \rightarrow 0$$

as $n \rightarrow \infty$ if

$$R_2 \geq I(U_2; X_2) + \delta(\epsilon).$$

2) To bound the last term, we need three steps, each of which uses a version of the Markov lemma [22, Section 12.1].

Step 1: Show that $(U_2^n(M_2), X_1^n, X_2^n) \in T_{\epsilon''}^{(n)}$ with a probability tends to 1 as n increases.

Since $X_2^n | \{U_2^n(M_2) = u_2^n, X_1^n = x_1^n\} \sim \prod_{i=1}^n P_{X_2 | X_1}(x_{2i} | x_{1i})$ and $\epsilon'' > \epsilon'''$, by the Markov lemma, $\Pr\{(U_2^n(M_2), X_1^n, X_2^n) \notin T_{\epsilon''}^{(n)}\}$ tends to zero as $n \rightarrow \infty$.

Step 2: Show that $(U_1^n(M_1), U_2^n(M_2), X_1^n, X_2^n) \in T_{\epsilon'}^{(n)}$ with a probability tends to 1 as n increases.

From the distribution we draw $U_1^n(M_1)$ and $U_2^n(M_2)$, we have the Markov chain

$$U_2^n(M_2) \leftrightarrow X_2^n \leftrightarrow X_1^n \leftrightarrow U_1^n(M_1).$$

As $(u_2^n, x_1^n, x_2^n) \in T_{\epsilon''}^{(n)}$ and from the Markov chain we know that

$$\begin{aligned} \Pr\{U_1^n(M_1) = u_1^n | U_2^n(M_2) = u_2^n, X_1^n = x_1^n, X_2^n = x_2^n\} \\ = \Pr\{U_1^n(M_1) = u_1^n | x_1^n\}. \end{aligned}$$

By the covering lemma, $\Pr\{(x_1^n, U_1^n) \in T_{\epsilon''}^{(n)}\}$ converges to 1 as $n \rightarrow \infty$, that is $\Pr\{U_1^n(M_1) = u_1^n | x_1^n\}$ satisfies the first condition in the Markov lemma. Then we show that it also satisfies the second condition in the Markov lemma. For all $u_1^n \in T_{\epsilon''}^{(n)}(U_1 | x_1^n)$,

$$\begin{aligned} \Pr\{U_1^n(M_1) = u_1^n | X_1^n = x_1^n\} \\ &= \Pr\{U_1^n(M_1) = u_1^n, U_1^n(M_1) \in T_{\epsilon''}^{(n)}(U_1 | x_1^n) \\ &\quad | X_1^n = x_1^n\} \\ &= \Pr\{U_1^n(M_1) \in T_{\epsilon''}^{(n)}(U_1 | x_1^n) | X_1^n = x_1^n\} \\ &\quad \times \Pr\{U_1^n(M_1) = u_1^n | U_1^n(M_1) \in T_{\epsilon''}^{(n)}(U_1 | x_1^n), \\ &\quad X_1^n = x_1^n\} \\ &\leq \Pr\{U_1^n(M_1) = u_1^n | U_1^n(M_1) \in T_{\epsilon''}^{(n)}(U_1 | x_1^n), \\ &\quad X_1^n = x_1^n\} \\ &= \sum_{m_1} \Pr\{U_1^n(M_1) = u_1^n, M_1 = m_1 | \\ &\quad U_1^n(M_1) \in T_{\epsilon''}^{(n)}(U_1 | x_1^n), X_1^n = x_1^n\} \\ &= \sum_{m_1} \Pr\{U_1^n(M_1) = u_1^n | U_1^n(M_1) \in T_{\epsilon''}^{(n)}(U_1 | x_1^n), \\ &\quad X_1^n = x_1^n, M_1 = m_1\} \\ &\quad \times \Pr\{M_1 = m_1 | U_1^n(M_1) \in T_{\epsilon''}^{(n)}(U_1 | x_1^n), X_1^n = x_1^n\} \\ &\stackrel{(a)}{=} \sum_{m_1} \Pr\{U_1^n(m_1) = u_1^n | U_1^n(m_1) \in T_{\epsilon''}^{(n)}(U_1 | x_1^n)\} \\ &\quad \times \Pr\{M_1 = m_1 | U_1^n(M_1) \in T_{\epsilon''}^{(n)}(U_1 | x_1^n), X_1^n = x_1^n\} \\ &\stackrel{(b)}{\leq} \sum_{m_1} \Pr\{M_1 = m_1 | U_1^n(M_1) \in T_{\epsilon''}^{(n)}(U_1 | x_1^n), X_1^n = x_1^n\} \\ &\quad 2^{-n(H(U_1 | X_1) - \delta(\epsilon''))} \\ &= 2^{-n(H(U_1 | X_1) - \delta(\epsilon''))}, \end{aligned}$$

where (a) follows since

$$\begin{aligned} \Pr\{U_1^n(M_1) = u_1^n | U_1^n(M_1) \in T_{\epsilon''}^{(n)}(U_1 | x_1^n), \\ X_1^n = x_1^n, M_1 = m_1\} \\ &= \Pr\{U_1^n(m_1) = u_1^n | U_1^n(m_1) \in T_{\epsilon''}^{(n)}(U_1 | x_1^n), \\ &\quad X_1^n = x_1^n, M_1 = m_1\} \\ &= \Pr\{U_1^n(m_1) = u_1^n | U_1^n(m_1) \in T_{\epsilon''}^{(n)}(U_1 | x_1^n)\}. \end{aligned}$$

(b) follows from properties of typical sequences. Similarly, we can also prove that for every $u_1^n \in T_{\epsilon''}^{(n)}(U_1 | x_1^n)$ and n sufficiently large,

$$\begin{aligned} \Pr\{U_1^n(M_1) = u_1^n | X_1^n = x_1^n\} \\ \geq (1 - \epsilon'') 2^{-n(H(U_1 | X_1) + \delta(\epsilon''))}. \end{aligned}$$

Hence, this satisfies the second condition in the Markov Lemma. By the Markov lemma, we have $(U_1^n(M_1), U_2^n(M_2), X_1^n, X_2^n) \in T_{\epsilon'}^{(n)}$.

Step 3: Show that $(Y^n, U_1^n(M_1), U_2^n(M_2)) \in T_{\epsilon}^{(n)}$ with a probability tends to 1 as n increases.

First, $(U_1^n(M_1), U_2^n(M_2)) \leftrightarrow (X_1^n, X_2^n) \leftrightarrow Y^n$ forms a Markov chain as $(U_1^n(M_1), U_2^n(M_2))$ is a function of (X_1^n, X_2^n) . According to *Step 1* and *Step 2*, we have $(U_1^n(M_1), U_2^n(M_2), X_2^n, X_1^n) \in T_{\epsilon'}^{(n)}$, and Y^n is drawn $\sim \prod_{i=1}^n P_{Y|X_1X_2}(y_i|x_{1i}, x_{2i})$, hence, by the Markov lemma, we have $(Y^n, U_1^n(M_1), U_2^n(M_2)) \in T_{\epsilon}^{(n)}$ with a probability tends to 1 as n increases. This implies that $P_{X_1X_2Y}^n(\epsilon_1^c \cap \epsilon_2^c \cap \epsilon_3)$ tends to 0 as n increases.

Combining all steps above, we have that $\alpha_n \downarrow 0$ as n increases, hence the type 1 error probability constraint is satisfied.

For the type 2 error probability, assume in this case that H_1 is true. Then

$$\begin{aligned} \beta_n &= (P_{X_1X_2}^n P_Y^n)(\mathcal{A}_n) \\ &= (P_{X_1X_2}^n P_Y^n)(\epsilon_1^c \cap \epsilon_2^c \cap \epsilon_3^c) \\ &= (P_{X_1X_2}^n P_Y^n)(\epsilon_1^c) \times (P_{X_1X_2}^n P_Y^n)(\epsilon_2^c) \\ &\quad \times (P_{X_1X_2}^n P_Y^n)(\epsilon_3^c | \epsilon_1^c \cap \epsilon_2^c) \end{aligned}$$

We now bound each factor.

- 1) By the covering lemma, $(P_{X_1X_2}^n P_Y^n)(\epsilon_1^c) \rightarrow 1$ as $n \rightarrow \infty$ if

$$R_1 \geq I(U_1; X_1) + \delta(\epsilon)$$

and $(P_{X_1X_2}^n P_Y^n)(\epsilon_2^c) \rightarrow 1$ as $n \rightarrow \infty$ if

$$R_2 \geq I(U_2; X_2) + \delta(\epsilon).$$

2)

$$\begin{aligned} &(P_{X_1X_2}^n P_Y^n)(\epsilon_3^c | \epsilon_1^c \cap \epsilon_2^c) \\ &= \sum_{(u_1^n, u_2^n, y^n) \in T_{\epsilon}^{(n)}} (P_{X_1X_2}^n P_Y^n) \{U_1^n(M_1) = u_1^n, \\ &\quad U_2^n(M_2) = u_2^n, Y^n = y^n | \epsilon_1^c \cap \epsilon_2^c\} \\ &= \sum_{(u_1^n, u_2^n, y^n) \in T_{\epsilon}^{(n)}} P_{X_1X_2}^n \{U_1^n(M_1) = u_1^n, \\ &\quad U_2^n(M_2) = u_2^n | \epsilon_1^c \cap \epsilon_2^c\} \times P_Y^n \{Y^n = y^n | \epsilon_1^c \cap \epsilon_2^c\} \\ &\leq 2^{n(H(U_1U_2Y) + \delta(\epsilon))} 2^{-n(H(U_1U_2) - \delta(\epsilon'))} 2^{-n(H(Y) - \delta(\epsilon'))} \\ &= 2^{-n(I(U_1U_2; Y) - \delta(\epsilon))}. \end{aligned}$$

Combining the bounds on the three factors, we have

$$\beta_n \leq 2^{-n(I(U_1U_2; Y) - \delta(\epsilon))}.$$

In summary, the type 1 error probability averaged over all codebooks is upper bounded by ϵ if

$$R_1 \geq I(U_1; X_1) \text{ and } R_2 \geq I(U_2; X_2),$$

while the type 2 error probability averaged over all codebooks is upper bounded by $2^{-n(I(U_1U_2; Y) - \delta(\epsilon))}$. Therefore, there exists a codebook such that

$$\begin{aligned} \theta(R_1, R_2, \epsilon) &\geq I(U_1U_2; Y), \\ R_1 &\geq I(U_1; X_1), \\ R_2 &\geq I(U_2; X_2). \end{aligned}$$

This completes the achievability proof.

APPENDIX C PROOF OF THEOREM 7

Now we simplify the upper bound in (39) in the following steps. First consider

$$\begin{aligned} nR_1 &\geq H(M_1) \\ &\geq I(M_1; X_1^n) \\ &= \sum_{i=1}^n I(M_1; X_{1i} | X_1^{i-1}) \\ &= \sum_{i=1}^n I(M_1 X_1^{i-1}; X_{1i}) \\ &\stackrel{(a)}{=} \sum_{i=1}^n I(M_1 X_1^{i-1} X_2^{i-1}; X_{1i}) \\ &\stackrel{(b)}{=} \sum_{i=1}^n I(U_{1i}; X_{1i}), \end{aligned}$$

where (a) follows since $X_{1i} \leftrightarrow (M_1, X_1^{i-1}) \leftrightarrow X_2^{i-1}$ forms a Markov chain, which can be derived by the following:

$$\begin{aligned} (X_1^n, X_{1i}) &\leftrightarrow X_1^{i-1} \leftrightarrow X_2^{i-1} \\ &\stackrel{(c)}{\Rightarrow} (M_1, X_{1i}) \leftrightarrow X_1^{i-1} \leftrightarrow X_2^{i-1} \\ &\stackrel{(d)}{\Rightarrow} X_{1i} \leftrightarrow (M_1, X_1^{i-1}) \leftrightarrow X_2^{i-1}, \end{aligned} \quad (56)$$

(c) is true as M_1 is a function of X_1^n and (d) is true due to the weak union property of Markov chain [32]. (b) is true by identifying $U_{1i} = (M_1, X_1^{i-1}, X_2^{i-1})$ and noting that $U_{1i} \leftrightarrow X_{1i} \leftrightarrow (X_{2i}, Y_i)$ forms a Markov chain as

$$\begin{aligned} (X_1^n, X_1^{i-1}, X_2^{i-1}) &\leftrightarrow X_{1i} \leftrightarrow (X_{2i}, Y_i) \\ &\Rightarrow (M_1, X_1^{i-1}, X_2^{i-1}) \leftrightarrow X_{1i} \leftrightarrow (X_{2i}, Y_i). \end{aligned}$$

Following similar steps as above, we have

$$\begin{aligned} nR_2 &\geq \sum_{i=1}^n I(M_2 X_2^{i-1} Y^{i-1}; X_{2i}) \\ &\stackrel{(f)}{=} \sum_{i=1}^n I(U_{2i}; X_{2i}), \end{aligned}$$

where (e) follows since $Y^{i-1} \leftrightarrow (M_2, X_2^{i-1}) \leftrightarrow X_{2i}$; (f) is true by identifying $U_{2i} = (M_2, X_2^{i-1}, Y^{i-1})$ and noting that $U_{2i} \leftrightarrow X_{2i} \leftrightarrow (X_{1i}, Y_i)$.

Finally, we consider

$$\begin{aligned} H(Y^n|M_1M_2) &= \sum_{i=1}^n H(Y_i|M_1M_2Y^{i-1}) \\ &\geq \sum_{i=1}^n H(Y_i|M_1M_2Y^{i-1}X_1^{i-1}X_2^{i-1}) \\ &= \sum_{i=1}^n H(Y_i|U_{1i}U_{2i}). \end{aligned}$$

Define the time-sharing random variable Q to be the uniformly distributed over $[1 : n]$ and independent of $(M_1, M_2, X_1^n, X_2^n, Y^n)$, and identify $U_1 = (U_{1Q}, Q)$, $U_2 = (U_{2Q}, Q)$, $X_1 = X_{1Q}$, $X_2 = X_{2Q}$, and $Y = Y_Q$. Clearly, we have $U_1 \leftrightarrow X_1 \leftrightarrow (X_2, Y)$ and $U_2 \leftrightarrow X_2 \leftrightarrow (X_1, Y)$ forms three Markov chains. Hence we have shown

$$\begin{aligned} R_1 &\geq I(U_1; X_1), \quad R_2 \geq I(U_2; X_2), \\ \lim_{\epsilon \rightarrow 0} \theta(R_1, R_2, \epsilon) &\leq H(Y) - H(Y|U_1U_2) = I(Y; U_1U_2), \end{aligned}$$

for some conditional PMF $P_{U_1|X_1}$ and $P_{U_2|X_2}$.

APPENDIX D PROOF OF COROLLARY 9

In the following, $\epsilon > \epsilon' > \epsilon'' > \epsilon'''$ are given small numbers.

Quantization codebook generation. Fix a joint distribution attaining the maximum of (43), which satisfies $P_{U_1U_2X_1X_2Y} = P_{X_1}P_{X_2}P_Y P_{U_1|X_1}P_{U_2|X_2}$. Let $P_{U_1}(u_1) = \sum_{x_1} P_{X_1}(x_1)P_{U_1|X_1}(u_1|x_1)$, and $P_{U_2}(u_2) = \sum_{x_2} P_{X_2}(x_2)P_{U_2|X_2}(u_2|x_2)$. Randomly and independently generate $M_1 = 2^{n(I(U_1; X_1) + \eta)}$ sequences $u_1^n(m_1)$, $m_1 \in \{1, \dots, M_1\}$ each according to $\prod_{i=1}^n P_{U_1}(u_{1i})$. Randomly and independently generate $M_2 = 2^{n(I(U_2; X_2) + \eta)}$ sequences $u_2^n(m_2)$, $m_2 \in \{1, \dots, M_2\}$ each according to $\prod_{i=1}^n P_{U_2}(u_{2i})$. These sequences constitute the codebook c , which is revealed to all terminals. We use \mathcal{C} to denote the set of all possible codebooks.

Encoding (Quantization). After observing sequence x_1^n , terminal \mathcal{X}_1 finds a $u_1^n(m_1)$ such that $(x_1^n, u_1^n(m_1)) \in T_{\epsilon'''}^{(n)}$, and sends the index m_1 to terminal \mathcal{Y} . 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 $[1, \dots, M_1]$ uniformly at random. Similarly, after observing a sequence x_2^n , terminal \mathcal{X}_2 finds a $u_2^n(m_2)$ such that $(x_2^n, u_2^n(m_2)) \in T_{\epsilon'''}^{(n)}$, then it sends the index m_2 to terminal \mathcal{Y} . 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 $[1, \dots, M_2]$ uniformly at random.

Testing. Upon receiving m_1 and m_2 , terminal \mathcal{Y} sets the acceptance region \mathcal{A}_n for H_0 to

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

where the jointly typical set $T_{\epsilon}^{(n)}$ is defined with respect to $P_{X_1}P_{X_2}P_Y$, $P_{U_1|X_1}$ and $P_{U_2|X_2}$.

Error probability analysis. Terminal \mathcal{Y} chooses $\hat{H} \neq H_0$ if and only if one or more of the following events occur:

$$\begin{aligned} \epsilon_1 &= \{(U_1^n(m_1), X_1^n) \notin T_{\epsilon'''}^{(n)} \text{ for all } m_1 \in [1 : M_1]\}, \\ \epsilon_2 &= \{(U_2^n(m_2), X_2^n) \notin T_{\epsilon'''}^{(n)} \text{ for all } m_2 \in [1 : M_2]\}, \\ \epsilon_3 &= \{(U_1^n(M_1), U_2^n(M_2), Y^n) \notin T_{\epsilon}^{(n)}\}. \end{aligned}$$

Hence, $\mathcal{A}_n = (\epsilon_1 \cup \epsilon_2 \cup \epsilon_3)^c$.

To analyze the type 1 error probability, we have

$$\begin{aligned} \alpha_n &= (P_{X_1}^n P_{X_2}^n P_Y^n)(\mathcal{A}_n^c) \\ &= (P_{X_1}^n P_{X_2}^n P_Y^n)(\epsilon_1 \cup \epsilon_2 \cup \epsilon_3) \\ &\leq (P_{X_1}^n P_{X_2}^n P_Y^n)(\epsilon_1) + (P_{X_1}^n P_{X_2}^n P_Y^n)(\epsilon_2) \\ &\quad + (P_{X_1}^n P_{X_2}^n P_Y^n)(\epsilon_3). \end{aligned}$$

We now bound each term.

1) By the covering lemma [22, Section 3.7],

$$(P_{X_1}^n P_{X_2}^n P_Y^n)(\epsilon_1) \rightarrow 0$$

as $n \rightarrow \infty$ if

$$R_1 \geq I(U_1; X_1) + \delta(\epsilon)$$

and

$$(P_{X_1}^n P_{X_2}^n P_Y^n)(\epsilon_2) \rightarrow 0$$

as $n \rightarrow \infty$ if

$$R_2 \geq I(U_2; X_2) + \delta(\epsilon).$$

2)

$$\begin{aligned} &(P_{X_1}^n P_{X_2}^n P_Y^n)(\epsilon_3^c) \\ &= \sum_{(u_1^n, u_2^n, y^n) \in T_{\epsilon}^{(n)}} (P_{X_1}^n P_{X_2}^n P_Y^n) \{U_1^n(M_1) = u_1^n, \\ &\quad U_2^n(M_2) = u_2^n, Y^n = y^n\} \\ &= \sum_{(u_1^n, u_2^n, y^n) \in T_{\epsilon}^{(n)}} P_{X_1}^n \{U_1^n(M_1) = u_1^n\} \\ &\quad P_{X_2}^n \{U_2^n(M_2) = u_2^n\} P_Y^n \{Y^n = y^n\} \\ &\leq 2^{n(H(U_1U_2Y) + \delta(\epsilon))} 2^{-n(H(U_1) - \delta(\epsilon'))} 2^{-n(H(U_2) - \delta(\epsilon'))} \\ &\quad 2^{-n(H(Y) - \delta(\epsilon'))} \\ &= 2^{-n(H(U_1) + H(U_2) + H(Y) - H(U_1) - H(U_2) - H(Y) - \delta(\epsilon))} \\ &\rightarrow 1. \end{aligned}$$

To calculate the type 2 error probability, assume in this case that H_1 is true. For $m_1 \in [1 : M_1]$, $m_2 \in [1 : M_2]$, and $y^n \in T_{\epsilon}^{(n)}(Y|u_1^n(m_1), u_2^n(m_2))$, define

$$\begin{aligned} S_{m_1, m_2}(x_2^n, y^n) &= \{u_1^n(m_1)\} \times \{u_2^n(m_2)\} \times T_{\epsilon}^{(n)}(X_1| \\ &\quad u_1^n(m_1)) \times T_{\epsilon}^{(n)}(X_2|u_2^n(m_2)) \times \{y^n\}, \end{aligned}$$

and

$$\varphi_n = \bigcup_{m_1=1}^{M_1} \bigcup_{m_2=1}^{M_2} \bigcup_{y^n \in T_{\epsilon}^{(n)}(Y|u_1^n(m_1), u_2^n(m_2))} S_{m_1, m_2}(y^n).$$

Suppose $U_1^{(n)}U_2^{(n)}X_1^{(n)}X_2^{(n)}Y^{(n)}$ is a type variable of $(u_1^n, u_2^n, x_1^n, x_2^n, y^n) \in S_{m_1, m_2}(y^n)$, then

$$\begin{aligned} & Q_{X_1}^n Q_{X_2}^n Q_Y^n(x_1^n, x_2^n, y^n) \\ &= \exp[-n(H(X_1^{(n)}X_2^{(n)}Y^{(n)}) \\ &\quad + D(X_1^{(n)}X_2^{(n)}Y^{(n)}||Q_{X_1}Q_{X_2}Q_Y))]. \end{aligned}$$

Denoting $N(U_1^{(n)}U_2^{(n)}X_1^{(n)}X_2^{(n)}Y^{(n)})$ the number of those elements $(u_1^n, u_2^n, x_1^n, x_2^n, y^n) \in \varphi_n$ that have $(U_1^{(n)}U_2^{(n)}X_1^{(n)}X_2^{(n)}Y^{(n)})$ as their type variable, it follows that

$$\begin{aligned} & N(U_1^{(n)}U_2^{(n)}X_1^{(n)}X_2^{(n)}Y^{(n)}) \\ &\leq \exp \left[n(I(X_1; U_1) + I(X_2; U_2) + H(X_1^{(n)}|U_1^{(n)}Y^{(n)}) \right. \\ &\quad \left. + H(X_2^{(n)}|U_1^{(n)}U_2^{(n)}X_1^{(n)}Y^{(n)}) \right. \\ &\quad \left. + H(Y|U_1U_2) + 2\eta + 2\epsilon \right] \\ &= \exp \left[n(I(X_1; U_1) + I(X_2; U_2) + H(X_1^{(n)}|U_1^{(n)}Y^{(n)}) \right. \\ &\quad \left. + H(X_2^{(n)}|U_1^{(n)}U_2^{(n)}X_1^{(n)}Y^{(n)}) \right. \\ &\quad \left. + H(Y) + 2\eta + 2\epsilon \right]. \end{aligned}$$

Hence,

$$\begin{aligned} \beta_n &= Q_{X_1}^n Q_{X_2}^n Q_Y^n(\mathcal{A}_n) \\ &\leq \sum_{U_1^{(n)}U_2^{(n)}X_1^{(n)}X_2^{(n)}Y^{(n)}} \exp \left[-n \left(k \left(U_1^{(n)}U_2^{(n)}X_1^{(n)}X_2^{(n)}Y^{(n)} \right) \right. \right. \\ &\quad \left. \left. - 2\eta - 2\epsilon \right) \right], \quad (57) \end{aligned}$$

where

$$\begin{aligned} & k \left(U_1^{(n)}U_2^{(n)}X_1^{(n)}X_2^{(n)}Y^{(n)} \right) \\ &= H \left(X_1^{(n)}X_2^{(n)}Y^{(n)} \right) + D \left(X_1^{(n)}X_2^{(n)}Y^{(n)} || Q_{X_1}Q_{X_2}Q_Y \right) \\ &\quad - I(X_1; U_1) - I(X_2; U_2) - H \left(X_1^{(n)}|U_1^{(n)}Y^{(n)} \right) \\ &\quad - H \left(X_2^{(n)}|U_1^{(n)}U_2^{(n)}X_1^{(n)}Y^{(n)} \right) - H(Y), \quad (58) \end{aligned}$$

and the sum is taken over all possible type variables of elements $(u_1^n, u_2^n, x_1^n, x_2^n, y^n) \in \varphi_n$. Hence, we have $(u_1^n(m_1), x_1^n) \in T_\epsilon^{(n)}(U_1X_1)$, $(u_2^n(m_2), x_2^n) \in T_\epsilon^{(n)}(U_2X_2)$, and $(u_1^n(m_1), u_2^n(m_2), y^n) \in T_\epsilon^{(n)}(U_1U_2Y)$. This implies that the sum ranges over all possible type variables $(U_1^{(n)}U_2^{(n)}X_1^{(n)}X_2^{(n)}Y^{(n)})$ such that, for all $u_1 \in \mathcal{U}_1$, $u_2 \in \mathcal{U}_2$, $x_1 \in \mathcal{X}_1$, $x_2 \in \mathcal{X}_2$, and $y \in \mathcal{Y}$,

$$\begin{aligned} & |P_{U_1^{(n)}X_1^{(n)}}(u_1x_1) - P_{U_1X_1}(u_1x_1)| \leq \epsilon P_{U_1X_1}(u_1x_1), \\ & |P_{U_2^{(n)}X_2^{(n)}}(u_2x_2) - P_{U_2X_2}(u_2x_2)| \leq \epsilon P_{U_2X_2}(u_2x_2), \\ & |P_{U_1^{(n)}U_2^{(n)}Y^{(n)}}(u_1u_2y) - P_{U_1U_2Y}(u_1u_2y)| \\ & \leq \epsilon P_{U_1U_2Y}(u_1u_2y). \end{aligned}$$

Thus, we can rewrite (58) as

$$\begin{aligned} & k \left(U_1^{(n)}U_2^{(n)}X_1^{(n)}X_2^{(n)}Y^{(n)} \right) \\ &= H(\tilde{X}_1\tilde{X}_2\tilde{Y}) + D(\tilde{X}_1\tilde{X}_2\tilde{Y}||Q_{X_1}Q_{X_2}Q_Y) \\ &\quad - I(\tilde{X}_1; \tilde{U}_1) - I(\tilde{X}_2; \tilde{U}_2) - H(\tilde{Y}) \\ &\quad - H(\tilde{X}_1|\tilde{U}_1\tilde{Y}) - H(\tilde{X}_2|\tilde{U}_1\tilde{U}_2\tilde{X}_1\tilde{Y}) + \delta(\epsilon), \quad (59) \end{aligned}$$

with some variable $\tilde{U}_1\tilde{U}_2\tilde{X}_1\tilde{X}_2\tilde{Y}$ such that

$$\begin{aligned} \tilde{P}_{U_1X_1} &= P_{U_1X_1}, \quad \tilde{P}_{U_2X_2} = P_{U_2X_2}, \\ \tilde{P}_{U_1U_2Y} &= P_{U_1U_2Y}, \end{aligned}$$

where $\delta(\epsilon) \rightarrow 0$. Through some calculation, we can get

$$k \left(U_1^{(n)}U_2^{(n)}X_1^{(n)}X_2^{(n)}Y^{(n)} \right) \quad (60)$$

$$= D(\tilde{P}_{U_1U_2X_1X_2Y}||Q_{U_1U_2X_1X_2Y}) + \delta(\epsilon), \quad (61)$$

where $Q_{U_1|X_1} = P_{U_1|X_1}$, $P_{U_2|X_2} = Q_{U_2|X_2}$.

Thus, by (57) and (60), we have

$$\begin{aligned} \beta_n &\leq (n+1)^{|\mathcal{U}_1| \cdot |\mathcal{U}_2| \cdot |\mathcal{X}_1| \cdot |\mathcal{X}_2| \cdot |\mathcal{Y}|} \\ &\quad \max_{\tilde{P}_{U_1U_2X_1X_2Y} \in \xi(U_1U_2)} \exp \left[-n \left(D(\tilde{P}_{U_1U_2X_1X_2Y}||Q_{U_1U_2X_1X_2Y}) \right. \right. \\ &\quad \left. \left. + \delta(\epsilon) - 2\eta - 2\epsilon \right) \right]. \end{aligned}$$

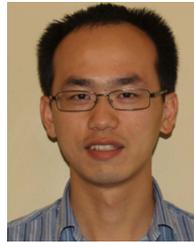
REFERENCES

- [1] W. Zhao and L. Lai, "Distributed testing against independence with multiple terminals," in *Proc. Allerton Conf. Commun. Control Comput.*, Montecello, IL, USA, Oct. 2014, pp. 1246–1251.
- [2] W. Zhao and L. Lai, "Distributed testing with zero-rate compression," in *Proc. IEEE Int. Symp. Inf. Theory*, Hongkong, China, Jun. 2015, pp. 2792–2796.
- [3] M.-F. Balcan, A. Blum, S. Fine, and Y. Mansour, "Distributed learning, communication complexity and privacy," in *Proc. Conf. Learn. Theory*, Princeton, NJ, USA, Jun. 2012, vol. 23, pp. 26.1–26.22.
- [4] Y. Zhang, J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Information-theoretic lower bounds for distributed statistical estimation with communication constraints," in *Proc. Adv. Neural Inf. Process. Syst.*, Stateline, NV, USA, Dec. 2013, pp. 2328–2336.
- [5] O. Shamir and N. Srebro, "Distributed stochastic optimization and learning," in *Proc. Allerton Conf. Commun. Control Comput.*, Montecello, IL, USA, Oct. 2014, pp. 850–857.
- [6] P. Willett, P. F. Swaszek, and R. S. Blum, "The good, bad, and ugly: Distributed detection of a known signal in dependent Gaussian noise," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3266–3279, Dec. 2000.
- [7] J. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing*, H. V. Poor and J. B. Thomas, Eds. Greenwich, CT, USA: JAI Press, 1990.
- [8] P. K. Varshney, *Distributed Detection and Data Fusion*. Berlin, Germany: Springer-Verlag, 1996.
- [9] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors—Part I: Fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.
- [10] R. Niu, B. Chen, and P. K. Varshney, "Fusion of decisions transmitted over Rayleigh fading channels in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1018–1027, Mar. 2006.
- [11] H. Chen, B. Chen, and P. K. Varshney, "A new framework for distributed detection with conditionally dependent observations," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1409–1419, Mar. 2012.
- [12] P. Yang, B. Chen, H. Chen, and P. K. Varshney, "Tandem distributed detection with conditionally dependent observations," in *Proc. Int. Conf. Inf. Fusion*, Singapore, Jul. 2012, pp. 1808–1813.
- [13] B. Kailkhura, S. Brahma, Y. S. Han, and P. K. Varshney, "Optimal distributed detection in the presence of Byzantines," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 2925–2929.
- [14] J.-F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, Feb. 2003.
- [15] J. Unnikrishnan and V. V. Veeravalli, "Decentralized detection with correlated observations," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2007, pp. 381–385.
- [16] J. Rhim, L. Varshney, and V. Goyal, "Quantization of prior probabilities for collaborative distributed hypothesis testing," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4537–4550, Sep. 2012.

- [17] F. Penna and S. Stanczak, "Decentralized eigenvalue algorithms for distributed signal detection in wireless networks," *IEEE Trans. Signal Process.*, vol. 63, no. 2, pp. 427–440, Jan. 2015.
- [18] Z. Quan, W.-K. Ma, S. Cui, and A. Sayed, "Optimal linear fusion for distributed detection via semidefinite programming," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2431–2436, Apr. 2010.
- [19] H. Zhu, G. Giannakis, and A. Cano, "Distributed in-network channel decoding," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3970–3983, Oct. 2009.
- [20] F. Cattivelli and A. Sayed, "Distributed detection over adaptive networks using diffusion adaptation," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 1917–1932, May 2011.
- [21] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.
- [22] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [23] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. IT-32, no. 4, pp. 533–542, Jul. 1986.
- [24] T. Berger, "Decentralized estimation and decision theory," in presented the *IEEE 7th Spring Workshop Inf. Theory*, Mt. Kisco, NY, USA, Sep. 1979.
- [25] H. Shalaby and A. Papamarcou, "Multiterminal detection with zero-rate data compression," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 254–267, Mar. 1992.
- [26] T. Han and K. Kobayashi, "Exponential-type error probabilities for multiterminal hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 2–14, Jan. 1989.
- [27] T. S. Han and S.-I. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2300–2324, Oct. 1998.
- [28] 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.
- [29] M. S. Rahman and A. B. Wagner, "The optimality of binning for distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6282–6303, Oct. 2012.
- [30] Y. Xiang and Y.-H. Kim, "Interactive hypothesis testing against independence," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 2840–2844.
- [31] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2005.
- [32] F. Zhu and B. Chen, "On the sum capacity of the discrete memoryless interference channel with one-sided weak interference and mixed interference," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 2012, pp. 2271–2275.



Wenwen Zhao (S'15) received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2013. She is currently pursuing the Ph.D. degree in Electrical and Computer Engineering from Worcester Polytechnic Institute, Worcester, MA, USA. Her research interests include statistical learning and information theory.



Lifeng Lai (M'07) received the B.E. and M.E. degrees from Zhejiang University, Hangzhou, China, in 2001 and 2004, respectively, and the Ph.D. degree from The Ohio State University, Columbus, OH, USA, in 2007. He was a Postdoctoral Research Associate with Princeton University, Princeton, NJ, USA, from 2007 to 2009, and was an Assistant Professor with the University of Arkansas, Little Rock, AR, USA, from 2009 to 2012. Since August 2012, he has been an Assistant Professor with Worcester Polytechnic Institute, Worcester, MA, USA. His research interests include information theory, stochastic signal processing and their applications in wireless communications, and security and other related areas. He was a Distinguished University Fellow of the Ohio State University, from 2004 to 2007. He served as a Guest Editor for the *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, Special Issue on Signal Processing Techniques for Wireless Physical Layer Security. He is currently serving as an Editor for the *IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS*, and an Associate Editor for the *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY*. He was the corecipient of the Best Paper Award from the IEEE Global Communications Conference (GlobeCom) in 2008, the Best Paper Award from the IEEE Conference on Communications (ICC) in 2011, and the Best Paper Award from the IEEE SMART GRID COMMUNICATIONS (SmartGridComm) in 2012. He was also the recipient of the National Science Foundation CAREER Award in 2011, and Northrop Young Researcher Award in 2012.