


Article

A Strong Converse Theorem for Hypothesis Testing Against Independence over a Two-Hop Network [†]

Daming Cao ¹, Lin Zhou ^{2,*}  and Vincent Y. F. Tan ³

¹ School of Information Science and Engineering, Southeast University, Nanjing 210000, China; dmcao@seu.edu.cn

² Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA

³ Department of Electrical and Computer Engineering and Department of Mathematics, National University of Singapore, Singapore 119077, Singapore; vtan@nus.edu.sg

* Correspondence: linzhou@umich.edu.

† This paper was partially presented at ISIT 2019, Paris, France, 7–12 July 2019.

Received: 29 October 2019; Accepted: 28 November 2019; Published: 29 November 2019



Abstract: By proving a strong converse theorem, we strengthen the weak converse result by Salehkalaibar, Wigger and Wang (2017) concerning hypothesis testing against independence over a two-hop network with communication constraints. Our proof follows by combining two recently-proposed techniques for proving strong converse theorems, namely the strong converse technique via reverse hypercontractivity by Liu, van Handel, and Verdú (2017) and the strong converse technique by Tyagi and Watanabe (2018), in which the authors used a change-of-measure technique and replaced hard Markov constraints with soft information costs. The techniques used in our paper can also be applied to prove strong converse theorems for other multiterminal hypothesis testing against independence problems.

Keywords: strong converse; hypothesis testing with communication constraints; testing against independence; two-hop network; relay

1. Introduction

Motivated by situations where the source sequence is *not* available directly and can only be obtained through limited communication with the data collector, Ahlswede and Csiszár [1] proposed the problem of hypothesis testing with a communication constraint. In the setting of [1], there is one encoder and one decoder. The encoder has access to one source sequence X^n and transmits a compressed version of it to the decoder at a limited rate. Given the compressed version and the available source sequence Y^n (side information), the decoder knows that the pair of sequences (X^n, Y^n) is generated i.i.d. from one of the two distributions and needs to determine which distribution the pair of sequences is generated from. The goal in this problem is to study the tradeoff between the compression rate and the exponent of the type-II error probability under the constraint that the type-I error probability is either vanishing or non-vanishing. For the special case of testing against independence, Ahlswede and Csiszár provided an exact characterization of the rate-exponent tradeoff. They also derived the so-called strong converse theorem for the problem. This states that the rate-exponent tradeoff cannot be improved even when one is allowed a non-vanishing type-I error probability. However, the characterization the rate-exponent tradeoff for the general case (even in the absence of a strong converse) remains open till date.

Subsequently, the work of Ahlswede and Csiszár was generalized to the distributed setting by Han in [2] who considered hypothesis testing over a Slepian-Wolf network. In this setting, there are

two encoders, each of which observes one source sequence and transmits a compressed version of the source to the decoder. The decoder then performs a hypothesis test given these two compression indices. The goal in this problem is to study the tradeoff between the coding rates and the exponent of type-II error probability, under the constraint that the type-I error probability is either vanishing or non-vanishing. Han derived an inner bound to the rate-exponent region. For the special case of zero-rate communication, Shalaby and Papamarcou [3] applied the blowing-up lemma [4] judiciously to derive the exact rate-exponent region and a strong converse theorem. Further generalizations of the work of Ahlswede and Csiszár can be categorized into two classes: non-interactive models where encoders do not communicate with one another [5–8] and the interactive models where encoders do communicate [9,10].

We revisit one such interactive model as shown in Figure 1. This problem was considered by Salehkalaibar, Wigger and Wang in [11] and we term the problem as *hypothesis testing over a two-hop network*. The two-hop model considered here has potential applications in the Internet of Things (IoT) and sensor networks. In these scenarios, direct communication from the transmitter to the receiver might *not* be possible due to power constraints that result from limited resources such as finite battery power. However, it is conceivable in such a scenario to assume that there are relays—in our setting, there is a single relay—that aid in the communication or other statistical inference tasks (such as hypothesis testing) between the transmitter and receiver.

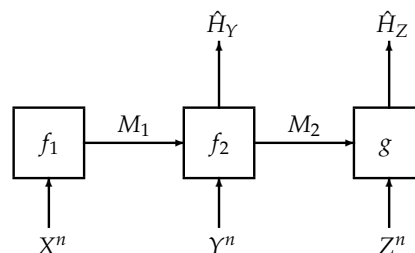


Figure 1. System model for hypothesis testing over a two-hop network

The main task in this problem is to construct two hypothesis tests between two joint distributions P_{XYZ} and Q_{XYZ} . One of these two distributions governs the law of (X^n, Y^n, Z^n) where each copy (X_i, Y_i, Z_i) is generated independently either from P_{XYZ} and Q_{XYZ} . As shown in Figure 1, the first terminal has knowledge of a source sequence X^n and sends an index M_1 to the second terminal, which we call the relay; the relay, given side information Y^n and compressed index M_1 , makes a guess of the hypothesis \hat{H}_Y and sends another index M_2 to the third terminal; the third terminal makes another guess of the hypothesis \hat{H}_Z based on M_2 and its own side information Z^n . The authors in [11] derived an inner bound for the rate-exponent region and showed that the bound is tight for several special cases, including the case of testing against independence in which $Q_{XYZ} = P_X P_Y P_Z$. However, even in this simpler case of testing against independence, which is our main concern in this paper, the authors in [11] *only* established a *weak* converse.

In this paper, we strengthen the result by Salehkalaibar, Wigger and Wang in [11] by deriving a *strong* converse for the case of testing against independence. Our proof follows by combining two recently proposed strong converse techniques by Liu et al. in [12] and by Tyagi and Watanabe in [13]. In [12], the authors proposed a framework to prove strong converse theorems based on functional inequalities and reverse hypercontractivity of Markov semigroups. In particular, they applied their framework to derive strong converse theorems for a collection of problems including the hypothesis testing with communication constraints problem in [1]. In [13], the authors proposed another framework for strong converse proofs, where they used a change-of-measure technique and replaced hard Markov constraints with soft information costs. They also leveraged variational formulas for various information-theoretic quantities; these formulas were introduced by Oohama in [14–16].

Notation

Random variables and their realizations are in upper (e.g., X) and lower case (e.g., x) respectively. All sets are denoted in calligraphic font (e.g., \mathcal{X}). We use \mathcal{X}^c to denote the complement of \mathcal{X} . Let $X^n := (X_1, \dots, X_n)$ be a random vector of length n and x^n its realization. Given any x^n , we use \hat{P}_{x^n} to denote its type (empirical distribution). All logarithms are base e . We use \mathbb{R}_+ and \mathbb{N} to denote the set of non-negative real numbers and natural numbers respectively. Given any positive integer $a \in \mathbb{N}$, we use $[a]$ to denote $\{1, \dots, a\}$. We use $1\{\cdot\}$ to denote the indicator function and use standard asymptotic notation such as $O(\cdot)$. The set of all probability distributions on a finite set \mathcal{X} is denoted as $\mathcal{P}(\mathcal{X})$. Given any two random variables (X, Y) and any realization of x , we use $P_{Y|x}(\cdot)$ to denote the conditional distribution $P_{Y|X}(\cdot|x)$. Given a distribution $P \in \mathcal{P}(\mathcal{X})$ and a function $f : \mathcal{X} \rightarrow \mathcal{R}$, we use $P(f)$ to denote $E_P[f(X)]$. For information-theoretic quantities, we follow [17]. In particular, when the joint distribution of (X, Y) is $P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, we use $I_{P_{XY}}(X; Y)$ and $I_P(X; Y)$ interchangeably. Throughout the paper, for ease of notation, we drop the subscript for distributions when there is no confusion. For example, when the joint distribution of (X, Y, Z) is P_{XYZ} , we use $I_P(X; Y|Z)$ and $I_{P_{XYZ}}(X; Y|Z)$ interchangeably. For any $(p, q) \in [0, 1]^2$, let $D_b(p||q)$ denote the binary divergence function, i.e., $D_b(p||q) = p \log(p/q) + (1 - p) \log((1 - p)/(1 - q))$.

2. Problem Formulation and Existing Results

2.1. Problem Formulation

Fix a joint distribution $P_{XYZ} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$ satisfying the Markov chain $X - Y - Z$, i.e.,

$$P_{XYZ}(x, y, z) = P_X(x)P_{Y|X}(y|x)P_{Z|Y}(z|y). \tag{1}$$

Let P_X, P_Y and P_Z be induced marginal distributions of P_{XYZ} . As shown in Figure 1, we consider a two-hop hypothesis testing problem with three terminals. The first terminal, which we term the transmitter, observes a source sequence X^n and sends a compression index M_1 to the second terminal, which we term the relay. Given M_1 and side information Y^n , the relay sends another compression index M_2 to the third terminal, which we term the receiver. The main task in this problem is to construct hypothesis tests at both the relay and the receiver to distinguish between

$$H_0 : (X^n, Y^n, Z^n) \sim P_{XYZ}^n = P_X P_{Y|X}^n P_{Z|Y}^n, \tag{2}$$

$$H_1 : (X^n, Y^n, Z^n) \sim P_X^n P_Y^n P_Z^n. \tag{3}$$

For subsequent analyses, we formally define a code for hypothesis testing over a two-hop network as follows.

Definition 1. An (n, N_1, N_2) -code for hypothesis testing over a two-hop network consists of

- Two encoders:

$$f_1 : \mathcal{X}^n \rightarrow \mathcal{M}_1 := \{1, \dots, N_1\}, \tag{4}$$

$$f_2 : \mathcal{M}_1 \times \mathcal{Y}^n \rightarrow \mathcal{M}_2 := \{1, \dots, N_2\}, \text{ and} \tag{5}$$

- Two decoders

$$g_1 : \mathcal{M}_1 \times \mathcal{Y}^n \rightarrow \{H_0, H_1\}, \tag{6}$$

$$g_2 : \mathcal{M}_2 \times \mathcal{Z}^n \rightarrow \{H_0, H_1\}. \tag{7}$$

Given an (n, N_1, N_2) -code with encoding and decoding functions (f_1, f_2, g_1, g_2) , we define acceptance regions for the null hypothesis H_0 at the relay and the receiver as

$$\mathcal{A}_{Y,n} := \{(m_1, y^n) : g_1(m_1, y^n) = H_0\}, \tag{8}$$

$$\mathcal{A}_{Z,n} := \{(m_2, z^n) : g_2(m_2, z^n) = H_0\} \tag{9}$$

respectively. We also define conditional distributions

$$P_{M_1|X^n}(m_1|x^n) := 1\{f_1(x_1^n) = m_1\}, \tag{10}$$

$$P_{M_2|Y^n M_1}(m_2|y^n, m_1) := 1\{f_2(m_1, y^n) = m_2\}. \tag{11}$$

Thus, for a (n, N_1, N_2) -code characterized by (f_1, f_2, g_1, g_2) , the joint distribution of random variables $(X^n, Y^n, Z^n, M_1, M_2)$ under the null hypothesis H_0 is given by

$$P_{X^n Y^n Z^n M_1 M_2}(x^n, y^n, z^n, m_1, m_2) = P_{XYZ}^n(x^n, y^n, z^n) P_{M_1|X^n}(m_1|x^n) P_{M_2|Y^n M_1}(m_2|y^n, m_1), \tag{12}$$

and under the alternative hypothesis H_1 is given by

$$\bar{P}_{X^n Y^n Z^n M_1 M_2}(x^n, y^n, z^n, m_1, m_2) = P_X^n(x^n) P_Y^n(y^n) P_Z^n(z^n) P_{M_1|X^n}(m_1|x^n) P_{M_2|Y^n M_1}(m_2|y^n, m_1). \tag{13}$$

Now, let $P_{Y^n M_1}$ and $P_{Z^n M_2}$ be marginal distributions induced by $P_{X^n Y^n Z^n M_1 M_2}$ and let $\bar{P}_{Y^n M_1}$ and $\bar{P}_{Z^n M_2}$ be marginal distributions induced by $\bar{P}_{X^n Y^n Z^n M_1 M_2}$. Then, we can define the type-I and type-II error probabilities at the relay as

$$\beta_1 := P_{M_1 Y^n}(\mathcal{A}_{Y,n}^c), \tag{14}$$

$$\beta_2 := \bar{P}_{M_1 Y^n}(\mathcal{A}_{Y,n}) \tag{15}$$

respectively and at the receiver as

$$\eta_1 := P_{M_2 Z^n}(\mathcal{A}_{Z,n}^c), \tag{16}$$

$$\eta_2 := \bar{P}_{M_2 Z^n}(\mathcal{A}_{Z,n}) \tag{17}$$

respectively. Clearly, β_1, β_2, η_1 , and η_2 are functions of n but we suppress these dependencies for brevity.

Given above definitions, the achievable rate-exponent region for the hypothesis testing problem in a two-hop network is defined as follows.

Definition 2. Given any $(\varepsilon_1, \varepsilon_2) \in (0, 1)^2$, a tuple (R_1, R_2, E_1, E_2) is said to be $(\varepsilon_1, \varepsilon_2)$ -achievable if there exists a sequence of (n, N_1, N_2) -codes such that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log N_i \leq R_i, \quad \forall i \in \{1, 2\}, \tag{18}$$

$$\limsup_{n \rightarrow \infty} \beta_1 \leq \varepsilon_1, \tag{19}$$

$$\limsup_{n \rightarrow \infty} \eta_1 \leq \varepsilon_2, \tag{20}$$

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \log \beta_2 \geq E_1, \tag{21}$$

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \log \eta_2 \geq E_2. \tag{22}$$

The closure of the set of all (ϵ_1, ϵ_2) -achievable rate-exponent tuples is called the (ϵ_1, ϵ_2) -rate-exponent region and is denoted as $\mathcal{R}(\epsilon_1, \epsilon_2)$. Furthermore, define the rate-exponent region as

$$\mathcal{R} := \mathcal{R}(0, 0). \tag{23}$$

2.2. Existing Results

In the following, we recall the exact characterization of \mathcal{R} given by Salehkalaibar, Wigger and Wang ([11] (Corollary 1)). For this purpose, define the following set of joint distributions

$$\mathcal{Q} := \{Q_{XYZUV} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \mathcal{U} \times \mathcal{V}) : Q_{XYZ} = P_{XYZ}, U - X - Y, V - Y - Z\}. \tag{24}$$

Given $Q_{XYZUV} \in \mathcal{Q}$, define the following set

$$\mathcal{R}(Q_{XYZUV}) := \{(R_1, R_2, E_1, E_2) : R_1 \geq I_Q(U; X), R_2 \geq I_Q(V; Y), E_1 \leq I_Q(U; Y), E_2 \leq I_Q(U; Y) + I_Q(V; Z)\} \tag{25}$$

Finally, let

$$\mathcal{R}^* := \bigcup_{Q_{XYZUV} \in \mathcal{Q}} \mathcal{R}(Q_{XYZUV}). \tag{26}$$

Theorem 1. *The rate-exponent region \mathcal{R} for the hypothesis testing over a two-hop network problem satisfies*

$$\mathcal{R} = \mathcal{R}^*. \tag{27}$$

In the following, inspired by Oohama’s variational characterization of rate regions for multiuser information theory [14–16], we provide an alternative characterization of \mathcal{R}^* . For this purpose, given any $(b, c, d) \in \mathbb{R}_+^3$ and any $Q_{XYZUV} \in \mathcal{Q}$, let

$$R_{b,c,d}(Q_{XYZUV}) := -I_Q(U; Y) + bI_Q(U; X) - c(I_Q(U; Y) + I_Q(V; Z)) + dI_Q(V; Y). \tag{28}$$

be a linear combination of the mutual information terms in (25). Furthermore, define

$$R_{b,c,d} := \min_{Q_{XYZUV} \in \mathcal{Q}} R_{b,c,d}(Q_{XYZUV}). \tag{29}$$

An alternative characterization of \mathcal{R}^* is given by

$$\mathcal{R}^* = \bigcap_{(b,c,d) \in \mathbb{R}_+^3} \{(R_1, R_2, E_1, E_2) : -E_1 + bR_1 - cE_2 + dR_2 \geq R_{b,c,d}\}. \tag{30}$$

3. Strong Converse Theorem

3.1. The Case $\epsilon_1 + \epsilon_2 < 1$

Theorem 2. *Given any $(\epsilon_1, \epsilon_2) \in (0, 1)^2$ such that $\epsilon_1 + \epsilon_2 < 1$ and any $(b, c, d) \in \mathbb{R}_+^3$, for any (n, N_1, N_2) -code such that $\beta_1 \leq \epsilon_1, \eta_1 \leq \epsilon_2$, we have*

$$\log \beta_2 + b \log N_1 + c \log \eta_2 + d \log N_2 \geq nR_{b,c,d} + \Theta(n^{3/4} \log n). \tag{31}$$

The proof of Theorem 2 is given in Section 4. Several remarks are in order.

First, using the alternative expression of the rate-exponent region in (30), we conclude that for any $(\epsilon_1, \epsilon_2) \in (0, 1)^2$ such that $\epsilon_1 + \epsilon_2 < 1$, we have $\mathcal{R}(\epsilon_1, \epsilon_2) = \mathcal{R}^*$. This result significantly strengthens the weak converse result in ([11] (Corollary 1)) in which it was shown that $\mathcal{R}(0, 0) = \mathcal{R}^*$.

Second, it appears difficult to establish the strong converse result in Theorem 2 using existing classical techniques including image-size characterizations (a consequence of the blowing-up lemma) [4,6] and the perturbation approach [18]. In Section 4, we combine two recently proposed strong converse techniques by Liu, van Handel, and Verdú [12] and by Tyagi and Watanabe [13]. In particular, we use the strong converse technique based on reverse hypercontractivity in [12] to bound the exponent of the type-II error probability at the receiver and the strong converse technique in [13], which leverages an appropriate change-of-measure technique and replaces hard Markov constraints with soft information costs, to analyze the exponent of type-II error probability at the relay. Finally, inspired by the single-letterization steps in ([19] (Lemma C.2)) and [13], we single-letterize the derived multi-letter bounds from the previous steps to obtain the desired result in Theorem 2.

Third, we briefly comment on the apparent necessity of combining the two techniques in [12,13] instead of applying just one of them to obtain Theorem 2. The first step to apply the technique in [13] is to construct a “truncated source distribution” which is supported on a smaller set (often defined in terms of the decoding region) and is not too far away from the true source distribution in terms of the relative entropy. For our problem, the source satisfies the Markov chain $X^n - Y^n - Z^n$. If we naïvely apply the techniques in [13], the Markovian property would not hold for the truncated source $(\tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n)$. On the other hand, it appears rather challenging to extend the techniques in [12] to the hypothesis testing over a multi-hop network problem since the techniques therein rely heavily on constructing semi-groups and it is difficult to devise appropriate forms of such semi-groups to be used and analyzed in this multi-hop setting. Therefore, we carefully combine the two techniques in [12,13] to ameliorate the aforementioned problems. In particular, we first use the technique in [13] to construct a truncated source $(\tilde{X}^n, \tilde{Y}^n)$ and then let the conditional distribution of \tilde{Z}^n given $(\tilde{X}^n, \tilde{Y}^n)$ be given by the true conditional source distribution $P_{Z|Y}^n$ to maintain the Markovian property of the source (see (56)). Subsequently, in the analysis of error exponents, we use the technique in [12] to analyze the exponent of type-II error probability at the receiver to circumvent the need to construct new semi-groups.

Finally, we remark that the techniques (or a subset of the techniques) used to prove Theorem 2 can also be used to establish a strong converse result for other multiterminal hypothesis testing against independence problems, e.g., hypothesis testing over the Gray-Wyner network [7], the interactive hypothesis testing problem [9] and the cascaded hypothesis testing problem [10].

3.2. The Case $\varepsilon_1 + \varepsilon_2 > 1$

In this subsection, we consider the case where the sum of type-I error probabilities at the relay and the receiver is upper bounded by a quantity strictly greater than one. For ease of presentation of our results, let

$$\begin{aligned} \mathcal{Q}_2 := \{Q_{XYZU_1U_2V} \in \mathcal{Q}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{V}) : \\ Q_{XYZ} = P_{XYZ}, U_1 - X - Y, U_2 - X - Y, V - Y - Z\}. \end{aligned} \tag{32}$$

Given any $Q_{XYZU_1U_2V} \in \mathcal{Q}_2$, define the following set of rate-exponent tuples

$$\begin{aligned} \tilde{\mathcal{R}}(Q_{XYZU_1U_2V}) := \left\{ (R_1, R_2, E_1, E_2) : R_1 \geq \max\{I_Q(U_1; X), I_Q(U_2; X)\}, R_2 \geq I_Q(V; X), \right. \\ \left. E_1 \leq I_Q(U_1; Y), E_2 \leq I_Q(U_2; Y) + I_Q(V; Z) \right\}. \end{aligned} \tag{33}$$

Furthermore, define

$$\tilde{\mathcal{R}} := \bigcup_{Q_{XYZU_1U_2V}} \tilde{\mathcal{R}}(Q_{XYZU_1U_2V}). \tag{34}$$

Given any $Q_{XYZU_1U_2V} \in \mathcal{Q}_2$ and $(b_1, b_2, c, d) \in \mathbb{R}_+^4$, define the following linear combination of the mutual information terms

$$\begin{aligned} \tilde{\mathcal{R}}_{b_1, b_2, c, d}(Q_{XYZU_1U_2V}) \\ := -I_Q(U_1; Y) + b_1 I_Q(U_1; X) + b_2 I_Q(U_2; X) - c(I_Q(U_2; Y) + I_Q(V; Z)) + d I_Q(V; Y), \end{aligned} \quad (35)$$

and let

$$\tilde{\mathcal{R}}_{b_1, b_2, c, d} := \min_{Q_{XYZU_1U_2V}} \tilde{\mathcal{R}}_{b_1, b_2, c, d}(Q_{XYZU_1U_2V}). \quad (36)$$

Then, based on [14–16], an alternative characterization of $\tilde{\mathcal{R}}$ is given by

$$\tilde{\mathcal{R}} = \bigcup_{(b_1, b_2, c, d) \in \mathbb{R}_+^4} \{(R_1, R_2, E_1, E_2) : -E_1 + b_1 R_1 + b_2 R_1 - c E_2 + d R_2 \geq \tilde{\mathcal{R}}_{b_1, b_2, c, d}\}. \quad (37)$$

Analogously to Theorem 2, we obtain the following result.

Theorem 3. Given any $(\varepsilon_1, \varepsilon_2) \in (0, 1)^2$ and any $(b_1, b_2, c, d) \in \mathbb{R}_+^4$, for any (n, N_1, N_2) -code such that $\beta_1 \leq \varepsilon_1, \eta_1 \leq \varepsilon_2$, we have

$$\log \beta_2 + b_1 \log N_1 + b_2 \log N_1 + c \log \eta_2 + d \log N_2 \geq n \tilde{\mathcal{R}}_{b_1, b_2, c, d} + \Theta(n^{3/4} \log n). \quad (38)$$

The proof of Theorem 3 is similar to that of Theorem 2 and thus omitted for simplicity.

To prove Theorem 3, we need to analyze two special cases (cf. Figure 2) of our system model separately:

- (i) Firstly, we consider the first hop, which involves the transmitter and the relay only. The first hop itself is a hypothesis testing problem with a communication constraint [1]. Using the techniques either in [13] or [12], we can obtain bounds on a linear combination of the rate of the first encoder and the type-II error exponent of the relay, (i.e., $\log \beta_2 + b_1 \log N_1$ for any $b_1 \in \mathbb{R}_+$) for any type-I error probability $\beta_1 \in (0, 1)$ at the relay.
- (ii) Secondly, we study the second special case in which the relay does not make a decision. Using similar steps to the proof of Theorem 2, we can obtain a lower bound on a linear combination of the rate at the transmitter, the rate at the relay and the type-II exponent at the receiver (i.e., $b_2 \log N_1 + c \log \eta_2 + d \log N_2$ for any $(b_2, c, d) \in \mathbb{R}_+^3$) for any type-I error probability $\eta_1 \in (0, 1)$ at the receiver.
- (iii) Finally, combining the results obtained in the first two steps, we obtain a lower bound on the linear combination of rates and type-II exponents (as shown in Theorem 3). The proof is completed by using standard single-letterization steps and the variational formula in Equation (37).

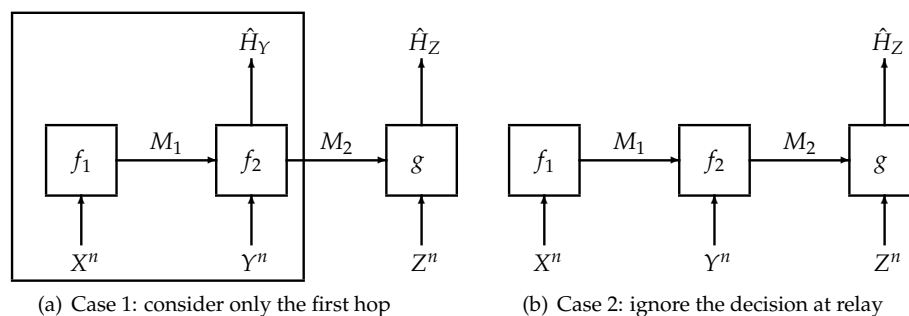


Figure 2. Illustration of the proof sketch of Theorem 3.

Using Theorem 3, we obtain the following proposition.

Proposition 1. For any $(\varepsilon_1, \varepsilon_2) \in (0, 1)^2$ such that $\varepsilon_1 + \varepsilon_2 > 1$, we have

$$\mathcal{R}(\varepsilon_1, \varepsilon_2) = \tilde{\mathcal{R}}. \quad (39)$$

The converse proof of Proposition 1 follows from Theorem 3 and the alternative characterization of $\tilde{\mathcal{R}}$ in (37). The achievability proof is inspired by ([6] (Theorem 5)) and is provided in Appendix A. The main idea is that we can time-share between two close-to optimal coding schemes, each of which corresponds to one special case of the current problem as mentioned after Theorem 3.

Recall that in the first remark of Theorem 2, we provide an exact characterization of the rate-exponent region for any $(\varepsilon_1, \varepsilon_2) \in (0, 1)^2$ such that $\varepsilon_1 + \varepsilon_2 < 1$. The converse proof follows from Theorem 2 and the achievability part was given in ([20] (Corollary 1)). Combining the first remark of Theorem 2 and Proposition 1, we provide an exact characterization of $\mathcal{R}(\varepsilon_1, \varepsilon_2)$ for any $(\varepsilon_1, \varepsilon_2) \in (0, 1)^2$ such that $\varepsilon_1 + \varepsilon_2 \neq 1$. We remark the case in which $\varepsilon_1 + \varepsilon_2 = 1$ was also excluded in the analysis of the successive refinement of hypothesis testing with communication constraints problem studied by Tian and Chen [6]. In fact, our converse result in Theorem 3 holds for any $(\varepsilon_1, \varepsilon_2) \in (0, 1)^2$ including the case $\varepsilon_1 + \varepsilon_2 = 1$. However, the achievability result presented in Appendix A holds only when $\varepsilon_1 + \varepsilon_2 > 1$ and thus we are unable to characterize $R(\varepsilon_1, \varepsilon_2)$ when $\varepsilon_1 + \varepsilon_2 = 1$. Because of the need to propose an achievability scheme which uses completely different techniques to handle the case in which $\varepsilon_1 + \varepsilon_2 = 1$, which does not dovetail with the main message and contribution of this paper, we omit this case in this paper.

4. Proof of Theorem 2

4.1. Preliminaries

Before presenting the proof of Theorem 2, in this subsection, we briefly review the two strong converse techniques that we judiciously combine in this work, namely the change-of-measure technique by Tyagi and Watanabe [13] and the hypercontractivity technique by Liu et al. [12].

The critical step in the strong converse technique by Tyagi and Watanabe [13] is to construct a truncated source distribution, which is supported over a small set related to the decoding regions. Furthermore, the constructed truncated distribution should satisfy the following conditions:

- (i) The truncated distribution is *close* to the original source distribution in terms of the KL divergence;
- (ii) Under the truncated distribution, the (type-I) error probability is *small*.

Subsequent steps proceed similarly as the weak converse analysis of the problem and lead to bounds on the rates and (type-II) exponents. We then single-letterize the obtained bounds (using classical techniques in information theory without the memoryless property, e.g., [21]). Finally, we relate the single-letterized results to the the variational characterization [14,16] of the fundamental limit of the problem, which uses the idea of replacing hard Markov constraints with soft information costs.

The advantage of the Tyagi-Watanabe technique lies in its simplicity and similarity to weak converse analyses. In contrast, the disadvantage of the technique is that the structure of the source distribution (e.g., Markovian) is potentially *lost* in the constructed truncated distribution. As we have illustrated briefly after Theorem 2, this disadvantage prevents us from solely using the Tyagi-Watanabe technique to prove the strong converse theorem for our setting.

On the other hand, the key technique in the strong converse technique by Liu et al. [12] is the use of ideas from reverse hypercontractivity. In particular, one needs to use the variational formula of the KL divergence ([22] (Chapter 12)) and carefully construct Markov semigroups. The operation of applying a Markov semigroup is similar to a soft version of blowing up of decoding sets [4] for the discrete memoryless case. The advantage of the strong converse technique by Liu et al. lies

in its wide applicability (beyond discrete settings) and its versatile performance (beyond showing strong converses it can be used to show that the second order terms scale as $O(\sqrt{n})$). However, the construction of appropriate Markov semigroups is problem-specific, which limits its applicability to other information-theoretic problems in the sense that one has to construct specific semigroups for each problem. Fortunately, in our setting this construction and combination with Tyagi-Watanabe’s technique, is feasible.

4.2. Summary of Proof Steps

In the rest of this section, we present the proof of strong converse theorem for the hypothesis testing over the two-hop network. The proof follows by combining the techniques in [12,13] and is separated into three main steps. First, we construct a truncated source distribution $P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n}$ and show that this truncated distribution is not too different from $P_{X^n Y^n Z^n}^n$ in terms of the relative entropy. Subsequently, we analyze the exponents of type-II error probabilities at the relay and the receiver under the constraint that their type-I error probabilities are non-vanishing. Finally, we single-letterize the constraints on rate and error exponents to obtain desired result in Theorem 2.

To begin with, let us fix an (n, N_1, N_2) -code with functions (f_1, f_2, g_1, g_2) such that the type-I error probabilities are bounded above by $\epsilon_1 \in (0, 1)$ and $\epsilon_2 \in (0, 1)$ respectively, i.e., $\beta_1 \leq \epsilon_1$ and $\eta_1 \leq \epsilon_2$. We note from (19) and (20) that $\beta_1 \leq \epsilon_1 + o(1)$ and $\beta_2 \leq \epsilon_2 + o(1)$. Since the $o(1)$ terms are immaterial in the subsequent analyses, they are omitted for brevity.

4.3. Step 1: Construction of a Truncated Distribution

Paralleling the definitions of acceptance regions in (8) and (9), we define the following acceptance regions at the relay and the receiver as

$$\mathcal{D}_{Y,n} = \{(x^n, y^n) : g_1(f_1(x^n), y^n) = H_0\}, \tag{40}$$

$$\mathcal{D}_{Z,n} = \{(x^n, y^n, z^n) : g_2(f_2(f_1(x^n), y^n), z^n) = H_0\}, \tag{41}$$

respectively. Note that the only difference between $\mathcal{A}_{Y,n}$ and $\mathcal{D}_{Y,n}$ lies in whether we consider the compression index m_1 or the original source sequence x^n . Recalling the definitions of the type-I error probabilities for the relay denoted by β_1 in (14) and for the receiver denoted by η_1 in (16), and using (40) and (41), we conclude that

$$P_{XY}^n(\mathcal{D}_{Y,n}) = 1 - \beta_1, \tag{42}$$

$$P_{XYZ}^n(\mathcal{D}_{Z,n}) = 1 - \eta_1. \tag{43}$$

For further analysis, given any $m_2 \in \mathcal{M}_2$, define a conditional acceptance region at the receiver (conditioned on m_2) as

$$\mathcal{G}(m_2) := \{z^n : g_2(m_2, z^n) = H_0\}. \tag{44}$$

For ease of notation, given any $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$, we use $\mathcal{G}(x^n, y^n)$ and $\mathcal{G}(f_2(f_1(x^n), y^n))$ (here $f_2(f_1(x^n), y^n)$ plays the role of m_2 in (44)) interchangeably and define the following set

$$\mathcal{B}_n := \left\{ (x^n, y^n) : P_{ZY}^n(\mathcal{G}(x^n, y^n) | y^n) \geq \frac{1 - \epsilon_1 - \epsilon_2}{1 + 3\epsilon_2 - \epsilon_1} \right\}. \tag{45}$$

Combining (41), (43) and (44), we obtain

$$1 - \varepsilon_2 \leq P_{XYZ}^n(\mathcal{D}_{Z,n}) \tag{46}$$

$$= \sum_{(x^n, y^n) \in \mathcal{B}_n} P_{XY}^n(x^n, y^n) P_{Z|Y}^n(\mathcal{G}(x^n, y^n) | y^n) + \sum_{(x^n, y^n) \notin \mathcal{B}_n} P_{XY}^n(x^n, y^n) P_{Z|Y}^n(\mathcal{G}(x^n, y^n) | y^n) \tag{47}$$

$$\leq P_{XY}^n(\mathcal{B}_n) + (1 - P_{XY}^n(\mathcal{B}_n)) \frac{1 - \varepsilon_1 - \varepsilon_2}{1 + 3\varepsilon_2 - \varepsilon_1}. \tag{48}$$

Thus, we have

$$P_{XY}^n(\mathcal{B}_n) \geq \frac{3 - 3\varepsilon_2 + \varepsilon_1}{4}. \tag{49}$$

For subsequent analyses, let

$$\mu := \left(\min_{y: P_Y(y) > 0} P_Y(y) \right)^{-1}, \tag{50}$$

$$\theta_n := \sqrt{\frac{3\mu}{n} \log \frac{8|\mathcal{Y}|}{1 - \varepsilon_1 - \varepsilon_2}}, \tag{51}$$

and define the typical set $\mathcal{T}_n(P_Y)$ as

$$\mathcal{T}_n(P_Y) = \{y^n : |\hat{P}_{y^n}(y) - P_Y(y)| \leq \theta_n P_Y(y) \quad \forall y \in \mathcal{Y}\}. \tag{52}$$

Using the Chernoff bound, we conclude that when n is sufficiently large,

$$P_Y^n(\mathcal{T}_n(P_Y)) \geq 1 - \frac{1 - \varepsilon_1 - \varepsilon_2}{4}. \tag{53}$$

Now, define the following set

$$\mathcal{C}_n := \mathcal{B}_n \cap \mathcal{D}_{Y,n} \cap (\mathcal{X}^n \times \mathcal{T}_n(P_Y)). \tag{54}$$

Then, combining (42), (49) and (53), we conclude that when n is sufficiently large,

$$P_{XY}^n(\mathcal{C}_n) \geq 1 - P_{XY}^n(\mathcal{B}_n^c) - P_{XY}^n(\mathcal{D}_{Y,n}^c) - P_Y^n(\mathcal{T}_n^c(P_Y)) \geq \frac{1 - \varepsilon_1 - \varepsilon_2}{2}. \tag{55}$$

Let the truncated distribution $P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n}$ be defined as

$$P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n}(x^n, y^n, z^n) := \frac{P_{XY}^n(x^n, y^n) 1\{(x^n, y^n) \in \mathcal{C}_n\}}{P_{XY}^n(\mathcal{C}_n)} P_{Z|Y}^n(z^n | y^n). \tag{56}$$

Note that under our constructed truncated distribution $P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n}$, the Markov chain $\tilde{X}^n - \tilde{Y}^n - \tilde{Z}^n$ holds. In other words, the Markovian property of the original source distribution P_{XYZ}^n is retained for the truncated distribution $P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n}$, which appears to be necessary to obtain a tight result if one wishes to use weak converse techniques. This is critical for our subsequent analyses.

Using the result in (55), we have that the marginal distribution $P_{\tilde{X}^n}$ satisfies that for any $x^n \in \mathcal{X}^N$,

$$P_{\tilde{X}^n}(x^n) = \sum_{y^n, z^n} P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n}(x^n, y^n, z^n) \tag{57}$$

$$\leq \frac{P_X^n(x^n)}{P_{XY}^n(\mathcal{C}_n)} \leq \frac{2P_X^n(x^n)}{1 - \varepsilon_1 - \varepsilon_2}. \tag{58}$$

Analogously to (58), we obtain that

$$P_{\tilde{Y}^n}(y^n) \leq \frac{2P_Y^n(y^n)}{1 - \varepsilon_1 - \varepsilon_2}, \quad \forall y^n \in \mathcal{Y}^n, \tag{59}$$

$$P_{\tilde{Z}^n}(z^n) \leq \frac{2P_Z^n(z^n)}{1 - \varepsilon_1 - \varepsilon_2}, \quad \forall z^n \in \mathcal{Z}^n. \tag{60}$$

Finally, note that

$$D(P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n} \| P_{\tilde{X}^n \tilde{Y}^n}^n \| P_{\tilde{X}^n \tilde{Y}^n}^n) = D(P_{\tilde{X}^n \tilde{Y}^n} \| P_{\tilde{X}^n \tilde{Y}^n}^n) \tag{61}$$

$$= \log \frac{1}{P_{\tilde{X}^n \tilde{Y}^n}^n(\mathcal{C}_n)} \tag{62}$$

$$\leq \log \frac{2}{1 - \varepsilon_1 - \varepsilon_2}. \tag{63}$$

4.4. Step 2: Analyses of the Error Exponents of Type-II Error Probabilities

4.4.1. Type-II Error Probability β_2 at the Relay

Let \tilde{M}_1 and \tilde{M}_2 be the outputs of encoders f_1 and f_2 respectively when the tuple of source sequences $(\tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n)$ is distributed according to $P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n}$ defined in (56). Thus, recalling the definitions in (10), (11) and (56), we find that the joint distribution of $(\tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n, \tilde{M}_1, \tilde{M}_2)$ is given by

$$P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n \tilde{M}_1 \tilde{M}_2}(x^n, y^n, z^n, m_1, m_2) = P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n}(x^n, y^n, z^n) P_{M_1|X^n}(m_1|x^n) P_{M_2|Y^n M_1}(m_2|y^n, m_1). \tag{64}$$

Let $P_{\tilde{M}_1 \tilde{Y}^n}$ be induced by $P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n \tilde{M}_1 \tilde{M}_2}$. Combining (8) and (56), we conclude that

$$P_{\tilde{M}_1 \tilde{Y}^n}(\mathcal{A}_{Y,n}) = \sum_{\substack{x^n, y^n, z^n, m_1, m_2: \\ g_1(m_1, y^n) = H_0}} P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n \tilde{M}_1 \tilde{M}_2}(x^n, y^n, z^n, m_1, m_2) \tag{65}$$

$$= \sum_{x^n, y^n: g_1(f_1(x^n), y^n) = H_0} \frac{P_{\tilde{X}^n \tilde{Y}^n}^n(x^n, y^n) 1\{(x^n, y^n) \in \mathcal{C}_n\}}{P_{\tilde{X}^n \tilde{Y}^n}^n(\mathcal{C}_n)} \tag{66}$$

$$= \sum_{x^n, y^n} \frac{P_{\tilde{X}^n \tilde{Y}^n}^n(x^n, y^n) 1\{(x^n, y^n) \in \mathcal{C}_n\}}{P_{\tilde{X}^n \tilde{Y}^n}^n(\mathcal{C}_n)} \tag{67}$$

$$= 1. \tag{68}$$

where (67) follows from the definition of $\mathcal{D}_{Y,n}$ in (40) and the fact that $\mathcal{D}_{Y,n} \subseteq \mathcal{C}_n$.

Thus, using the data processing inequality for the relative entropy and the definition of β_2 in (15), we obtain that

$$D(P_{\tilde{M}_1 \tilde{Y}^n} \| P_{M_1} P_Y^n) \geq D_b(P_{\tilde{M}_1 \tilde{Y}^n}(\mathcal{A}_{Y,n}) \| P_{M_1} P_Y^n(\mathcal{A}_{Y,n})) \tag{69}$$

$$= -\log(P_{M_1} P_Y^n(\mathcal{A}_{Y,n})) \tag{70}$$

$$= -\log \beta_2. \tag{71}$$

Furthermore, recalling that M_1 denotes the output of encoder f_1 when $(X^n, Y^n, Z^n) \sim P_{XYZ}^n$ and \tilde{M}_1 denotes the output of encoder f_1 when $(X^n, Y^n, Z^n) \sim P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n}$, and using the result in (58), we conclude that

$$P_{\tilde{M}_1}(m_1) = \sum_{x^n, y^n, z^n: f_1(x^n)=m_1} P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n}(x^n, y^n, z^n) \tag{72}$$

$$= \sum_{x^n: f_1(x^n)=m_1} P_{\tilde{X}^n}(x^n) \tag{73}$$

$$\leq \sum_{x^n: f_1(x^n)=m_1} \frac{2P_X^n(x^n)}{1 - \varepsilon_1 - \varepsilon_2} \tag{74}$$

$$\leq \frac{2P_{M_1}(m_1)}{1 - \varepsilon_1 - \varepsilon_2}, \tag{75}$$

for any $m_1 \in \mathcal{M}_1$. Thus, combining (59), (71) and (75), we have

$$-\log \beta_2 \leq D(P_{\tilde{M}_1 \tilde{Y}^n} \| P_{M_1} P_Y^n) \tag{76}$$

$$= D(P_{\tilde{M}_1 \tilde{Y}^n} \| P_{\tilde{M}_1} P_{\tilde{Y}^n}) + E_{P_{\tilde{M}_1 \tilde{Y}^n}} \left[\log \frac{P_{\tilde{M}_1}(\tilde{M}_1) P_{\tilde{Y}^n}(\tilde{Y}^n)}{P_{M_1}(\tilde{M}_1) P_Y^n(\tilde{Y}^n)} \right] \tag{77}$$

$$\leq D(P_{\tilde{M}_1 \tilde{Y}^n} \| P_{\tilde{M}_1} P_{\tilde{Y}^n}) + E_{P_{\tilde{M}_1 \tilde{Y}^n}} \left[\log \frac{\frac{2P_{M_1}(\tilde{M}_1)}{1-\varepsilon_1-\varepsilon_2} \frac{2P_Y^n(\tilde{Y}^n)}{1-\varepsilon_1-\varepsilon_2}}{P_{M_1}(\tilde{M}_1) P_Y^n(\tilde{Y}^n)} \right] \tag{78}$$

$$= I(\tilde{M}; \tilde{Y}^n) + 2 \log \frac{2}{1 - \varepsilon_1 - \varepsilon_2}. \tag{79}$$

4.4.2. Type-II Error Probability η_2 at the Receiver

In this subsection, we analyze the error exponent of the type-II error probability at the receiver. For this purpose, we make use of the method introduced in [12] based on reverse hypercontractivity. We define the following additional notation:

- Give $P_{YZ} \in \mathcal{P}(\mathcal{Y} \times \mathcal{Z})$, define

$$\alpha := \max_{y,z} \frac{P_{Z|Y}(z|y)}{P_Z(z)} \in (1, \infty). \tag{80}$$

In the subsequent analysis, we only consider the case when $\alpha > 1$. When $\alpha = 1$, choosing $t = \frac{1}{\sqrt{n}}$ instead of the choice in (101), we can obtain a similar upper bound for $-\log \eta_2$ as in (102), where the only difference is that $\Psi(n, \varepsilon_1, \varepsilon_2)$ should be replaced by another term scaling in order $\Theta(\sqrt{n})$.

- Given any $(\varepsilon_1, \varepsilon_2) \in (0, 1)^2$ such that $\varepsilon_1 + \varepsilon_2 < 1$, let

$$\Psi(n, \varepsilon_1, \varepsilon_2) := 2 \sqrt{n(\alpha - 1) \log \frac{1 + 3\varepsilon_2 - \varepsilon_1}{1 - \varepsilon_1 - \varepsilon_2}}. \tag{81}$$

- Give any $m_2 \in \mathcal{M}_2$ and $z^n \in \mathcal{Z}^n$, let

$$h(m_2, z^n) := 1\{z^n \in \mathcal{G}(m_2)\}. \tag{82}$$

- Two operators in ([12] (Equations (25), (26), (29)))

$$\Lambda_{\alpha,t} = (\exp(-t) + \alpha(1 - \exp(-t))P_Z)^{\otimes n}, \tag{83}$$

$$T_{y^n,t} = \prod_{i=1}^n (\exp(-t) + (1 - \exp(-t))P_{Z|y_i}). \tag{84}$$

Note that in (84), we use the convenient notation $P_{Z|y}(z) = P_{Z|Y}(z|y)$. The two operators in (83) and (84) will be used to lower bound $D(P_{\tilde{Z}^n \tilde{M}_2} \| P_Z^n \bar{P}_{M_2})$ via a variational formula of the relative entropy (cf. ([12] (Section 4))).

Let $P_{\tilde{Z}^n \tilde{M}_2}, P_{\tilde{Z}^n | \tilde{M}_2}, P_{\tilde{Z}^n | \tilde{Y}^n}$ be induced by the joint distribution $P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n \tilde{M}_1 \tilde{M}_2}$ in (64) and let \bar{P}_{M_2} be induced by the joint distribution $\bar{P}_{X^n Y^n Z^n M_1 M_2}$ in (13). Invoking the variational formula for the relative entropy ([22] (Equation (2.4.67))) and recalling the notation $P(f) = E_P[f]$, we have

$$D(P_{\tilde{Z}^n \tilde{M}_2} \| P_Z^n \bar{P}_{M_2}) \geq P_{\tilde{Z}^n \tilde{M}_2}(\log \Lambda_{\alpha,t} h(\tilde{M}_2, \tilde{Z}^n)) - \log((P_Z^n \bar{P}_{M_2})(\Lambda_{\alpha,t} h(M_2, Z^n))). \tag{85}$$

Given any $m_2 \in \mathcal{M}_2$, similar to ([12] (Equations (18)–(21))), we obtain

$$\begin{aligned} &P_Z^n(\Lambda_{\alpha,t} h(m_2, Z^n)) \\ &= P_Z^n((\exp(-t) + \alpha(1 - \exp(-t))P_Z)^{\otimes n} h(m_2, Z^n)) \end{aligned} \tag{86}$$

$$= (\exp(-t) + \alpha(1 - \exp(-t)))^n P_Z^n(h(m_2, Z^n)) \tag{87}$$

$$\leq \exp((\alpha - 1)nt) P_Z^n(h(m_2, Z^n)). \tag{88}$$

Thus, averaging over m_2 with distribution \bar{P}_{M_2} on both sides of (88), we have

$$\begin{aligned} &(P_Z^n \bar{P}_{M_2})(\Lambda_{\alpha,t} h(M_2, Z^n)) \\ &\leq \exp((\alpha - 1)nt) (\bar{P}_{M_2} P_Z^n)(h(M_2, Z^n)) \end{aligned} \tag{89}$$

$$= \exp((\alpha - 1)nt) \eta_2, \tag{90}$$

where (90) follows from the definition of η_2 in (17).

Furthermore, given any $\tilde{m}_2 \in \mathcal{M}_2$, we obtain

$$P_{\tilde{Z}^n | \tilde{m}_2}(\log \Lambda_{\alpha,t} h(\tilde{m}_2, \tilde{Z}^n)) \tag{91}$$

$$= \left(\sum_{\tilde{y}^n} P_{\tilde{Z}^n | \tilde{y}^n} P_{\tilde{Y}^n | \tilde{M}_2}(\tilde{y}^n | \tilde{m}_2) \right) (\log \Lambda_{\alpha,t} h(\tilde{m}_2, \tilde{Z}^n)) \tag{92}$$

$$= \sum_{\tilde{y}^n} P_{\tilde{Y}^n | \tilde{M}_2}(\tilde{y}^n | \tilde{m}_2) P_{\tilde{Z}^n | \tilde{y}^n}(\log \Lambda_{\alpha,t} h(\tilde{m}_2, \tilde{Z}^n)) \tag{93}$$

$$\geq \sum_{\tilde{y}^n} P_{\tilde{Y}^n | \tilde{M}_2}(\tilde{y}^n | \tilde{m}_2) P_{\tilde{Z}^n | \tilde{y}^n}(\log T_{y^n,t} h(\tilde{m}_2, \tilde{Z}^n)) \tag{94}$$

$$\geq \sum_{\tilde{y}^n} P_{\tilde{Y}^n | \tilde{M}_2}(\tilde{y}^n | \tilde{m}_2) \left(1 + \frac{1}{t} \right) \log P_{\tilde{Z}^n | \tilde{y}^n}(h(\tilde{m}_2, \tilde{Z}^n)) \tag{95}$$

$$= \left(1 + \frac{1}{t} \right) \left(\sum_{\tilde{y}^n} P_{\tilde{Y}^n | \tilde{M}_2}(\tilde{y}^n | \tilde{m}_2) \log P_{\tilde{Z}^n | \tilde{y}^n}(\mathcal{G}(\tilde{m}_2)) \right). \tag{96}$$

where (94) follows from ([12] (Lemma 4)) and (95) follows similarly to ([12] (Equations (14)–(17))).

Thus, averaging on both sides of (96) over \tilde{m}_2 with distribution $P_{\tilde{M}_2}$ and using the definition of the joint distribution $P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n \tilde{M}_1 \tilde{M}_2}$ in (64), we obtain that

$$P_{\tilde{Z}^n \tilde{M}_2}(\log \Lambda_{\alpha,t} h(\tilde{M}_2, \tilde{Z}^n)) \geq \left(1 + \frac{1}{t}\right) \left(\sum_{\tilde{y}^n, \tilde{m}_2} P_{\tilde{Y}^n \tilde{M}_2}(\tilde{y}^n, \tilde{m}_2) \log P_{\tilde{Z}^n | \tilde{y}^n}(\mathcal{G}(\tilde{m}_2))\right) \tag{97}$$

$$= \left(1 + \frac{1}{t}\right) \sum_{\tilde{x}^n, \tilde{y}^n, \tilde{m}_1, \tilde{m}_2} \left(P_{\tilde{X}^n \tilde{Y}^n}(\tilde{x}^n, \tilde{y}^n) 1\{\tilde{m}_1 = f_1(\tilde{x}^n), \tilde{m}_2 = f_2(\tilde{m}_1, \tilde{y}^n)\} \times \log \left(\sum_{\tilde{z}^n: g_2(\tilde{z}^n, \tilde{m}_2) = H_0} P_{\tilde{Z}^n | \tilde{Y}^n}^n(\tilde{z}^n | \tilde{y}^n)\right)\right) \tag{98}$$

$$= \left(1 + \frac{1}{t}\right) \left(\sum_{\tilde{x}^n, \tilde{y}^n} \frac{P_{\tilde{X}^n \tilde{Y}^n}^n(\tilde{x}^n, \tilde{y}^n) 1\{(\tilde{x}^n, \tilde{y}^n) \in \mathcal{C}_n\}}{P_{\tilde{X}^n \tilde{Y}^n}^n(\mathcal{C}_n)} \log P_{\tilde{Z}^n | \tilde{Y}^n}^n(\mathcal{G}(\tilde{x}^n, \tilde{y}^n) | \tilde{y}^n)\right) \tag{99}$$

$$\geq \left(1 + \frac{1}{t}\right) \log \frac{1 - \varepsilon_1 - \varepsilon_2}{1 + 3\varepsilon_2 - \varepsilon_1}, \tag{100}$$

where (100) follows from the definitions of \mathcal{B}_n in (45) and \mathcal{C}_n in (54).

Therefore, combining (85), (90) and (100) and choosing

$$t = \sqrt{\frac{1}{n(\alpha - 1)} \log \frac{1 + 3\varepsilon_2 - \varepsilon_1}{1 - \varepsilon_1 - \varepsilon_2}}, \tag{101}$$

via simple algebra, we obtain that

$$-\log \eta_2 \leq D(P_{\tilde{Z}^n \tilde{M}_2} \| P_{\tilde{Z}^n} \tilde{P}_{\tilde{M}_2}) + \Psi(n, \varepsilon_1, \varepsilon_2) - \log \frac{1 - \varepsilon_1 - \varepsilon_2}{1 + 3\varepsilon_2 - \varepsilon_1}. \tag{102}$$

In the following, we further upper bound $D(P_{\tilde{Z}^n \tilde{M}_2} \| P_{\tilde{Z}^n} \tilde{P}_{\tilde{M}_2})$. For this purpose, define the following distribution

$$\tilde{P}_{\tilde{M}_2}(m_2) := \sum_{y^n, m_1} P_{\tilde{M}_1}(m_1) P_{\tilde{Y}^n}(y^n) 1\{m_2 = f_2(m_1, y^n)\}. \tag{103}$$

Combining the results in (59) and (75), and recalling that $\tilde{P}_{\tilde{M}_2}$ is induced by joint distribution $\tilde{P}_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n \tilde{M}_1 \tilde{M}_2}$ in (13), for any $m_2 \in \mathcal{M}_2$, we have

$$\tilde{P}_{\tilde{M}_2}(m_2) \leq \left(\frac{2}{1 - \varepsilon_1 - \varepsilon_2}\right)^2 \left(\sum_{y^n, m_1} P_{\tilde{M}_1}(m_1) P_{\tilde{Y}^n}^n(y^n) 1\{m_2 = f_2(f_1(x^n), y^n)\}\right) \tag{104}$$

$$= \frac{4\tilde{P}_{\tilde{M}_2}(m_2)}{(1 - \varepsilon_1 - \varepsilon_2)^2}. \tag{105}$$

Thus, combining (60) and (105), we have

$$D(P_{\tilde{Z}^n \tilde{M}_2} \| P_{\tilde{Z}^n}^n \tilde{P}_{\tilde{M}_2}) = D(P_{\tilde{Z}^n \tilde{M}_2} \| P_{\tilde{Z}^n} \tilde{P}_{\tilde{M}_2}) + E_{P_{\tilde{Z}^n \tilde{M}_2}} \left[\log \frac{P_{\tilde{Z}^n}(\tilde{Z}^n) \tilde{P}_{\tilde{M}_2}(\tilde{M}_2)}{P_{\tilde{Z}^n}^n(\tilde{Z}^n) \tilde{P}_{\tilde{M}_2}(\tilde{M}_2)} \right] \tag{106}$$

$$\leq D(P_{\tilde{Z}^n \tilde{M}_2} \| P_{\tilde{Z}^n} \tilde{P}_{\tilde{M}_2}) + E_{P_{\tilde{Z}^n \tilde{M}_2}} \left[\log \frac{2P_{\tilde{Z}^n}^n(\tilde{Z}^n) 4\tilde{P}_{\tilde{M}_2}(\tilde{M}_2)}{\frac{1 - \varepsilon_1 - \varepsilon_2}{(1 - \varepsilon_1 - \varepsilon_2)^2} \tilde{P}_{\tilde{M}_2}(\tilde{M}_2)} \right] \tag{107}$$

$$= D(P_{\tilde{Z}^n \tilde{M}_2} \| P_{\tilde{Z}^n} \tilde{P}_{\tilde{M}_2}) + 3 \log \frac{2}{1 - \varepsilon_1 - \varepsilon_2}. \tag{108}$$

Therefore, combining (102) and (108), we have

$$-\log \eta_2 \leq D(P_{\tilde{Z}^n \tilde{M}_2} \| P_{\tilde{Z}^n \tilde{P}_{\tilde{M}_2}}) + \Psi(n, \varepsilon_1, \varepsilon_2) - \log \frac{1 - \varepsilon_1 - \varepsilon_2}{1 + 3\varepsilon_2 - \varepsilon_1} - 3 \log \frac{1 - \varepsilon_1 - \varepsilon_2}{2}. \quad (109)$$

4.5. Step 3: Analyses of Communication Constraints and Single-Letterization Steps

For any (n, N_1, N_2) -code, since $\tilde{M}_i \in \mathcal{M}_i$ for $i \in \{1, 2\}$, we have that

$$\log N_1 \geq H(\tilde{M}_1) \geq I(\tilde{M}_1; \tilde{X}^n, \tilde{Y}^n), \quad (110)$$

$$\log N_2 \geq H(\tilde{M}_2) \geq I(\tilde{M}_2; \tilde{Y}^n). \quad (111)$$

Furthermore, from the problem setting (see (64)), we have

$$I(\tilde{M}_1; \tilde{Y}^n | \tilde{X}^n) = 0, \quad (112)$$

For subsequent analyses, given any $(b, c, d, \gamma) \in \mathbb{R}_+^4$, define

$$\begin{aligned} R_{b,c,d,\gamma}^{(n)} := & -I(\tilde{M}_1; \tilde{Y}^n) + bI(\tilde{M}_1; \tilde{X}^n, \tilde{Y}^n) - cD(P_{\tilde{Z}^n \tilde{M}_2} \| P_{\tilde{Z}^n \tilde{P}_{\tilde{M}_2}}) + dI(\tilde{M}_2; \tilde{Y}^n) + \gamma I(\tilde{M}_1; \tilde{Y}^n | \tilde{X}^n) \\ & + (b + d + \gamma)D(P_{\tilde{X}^n \tilde{Y}^n} \| P_{X^n Y^n}). \end{aligned} \quad (113)$$

Combining the results in (63), (79), (109) to (112), for any $\gamma \in \mathbb{R}_+$, we obtain

$$\begin{aligned} & \log \beta_2 + b \log N_1 + c \log \eta_2 + d \log N_2 + c\Psi(n, \varepsilon_1, \varepsilon_2) \\ & \geq R_{b,c,d,\gamma}^{(n)} + \log \frac{1 - \varepsilon_1 - \varepsilon_2}{1 + 3\varepsilon_2 - \varepsilon_1} + (b + d + \gamma + 5) \log \frac{1 - \varepsilon_1 - \varepsilon_2}{2}. \end{aligned} \quad (114)$$

The proof of Theorem 2 is complete by the two following lemmas which provide a single-letterized lower bound for $R_{b,c,d,\gamma}^{(n)}$ and relate the derived lower bound to $R_{b,c,d}$. For this purpose, recalling the definition of θ_n in (51), we define the following set of joint distributions

$$\begin{aligned} \mathcal{Q}_1 := & \{Q_{XYZUV} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \times \mathcal{U} \times \mathcal{V}) : \\ & Q_{Z|Y} = P_{Z|Y}, X - Y - Z, V - Y - Z, \\ & |Q_Y(y) - P_Y(y)| \leq \theta_n P_Y(y), \quad \forall y \in \mathcal{Y}\}. \end{aligned} \quad (115)$$

Given $Q_{XYZUV} \in \mathcal{Q}_1$, define

$$\Delta_{b,d,\gamma}(Q_{XYZUV}) := (b + \gamma)D(Q_{XY} \| P_{XY}) + dD(Q_Y \| P_Y) + \gamma I_Q(U; Y|X). \quad (116)$$

Recall the definition of $R_{b,c,d}(Q_{XYZUV})$ in (28). Define

$$R_{b,c,d,\gamma} := \min_{Q_{XYZUV} \in \mathcal{Q}_1} (R_{b,c,d}(Q_{XYZUV}) + \Delta_{b,d,\gamma}(Q_{XYZUV})). \quad (117)$$

The following lemma presents a single-letterized lower bound for $R_{b,c,d,\gamma}^{(n)}$.

Lemma 1. For any $(b, c, d, \gamma) \in \mathbb{R}_+^4$,

$$R_{b,c,d,\gamma}^{(n)} \geq nR_{b,c,d,\gamma}. \quad (118)$$

The proof of Lemma 1 is inspired by ([13] (Prop. 2)) and provided in Appendix B.

Combining the results in (114) and Lemma 1, we obtain the desired result and this completes the proof of Theorem 2.

Lemma 2. *Choosing $\gamma = \sqrt{n}$, we have*

$$nR_{b,c,d,\gamma} + \log \frac{1 - \varepsilon_1 - \varepsilon_2}{1 + 3\varepsilon_2 - \varepsilon_1} + (b + d + \gamma + 5) \log \frac{1 - \varepsilon_1 - \varepsilon_2}{2} \geq nR_{b,c,d} + \Theta(n^{3/4} \log n). \quad (119)$$

The proof of Lemma 2 is inspired by ([19] (Lemma C.2)) and provided in Appendix C.

5. Discussion and Future Work

We strengthened the result in ([11] (Corollary 1)) by deriving a strong converse theorem for hypothesis testing against independence over a two-hop network with communication constraints (see Figure 1). In our proof, we combined two recently proposed strong converse techniques [12,13]. The apparent necessity of doing so comes from the Markovian requirement in the source distribution (recall (1)) and is reflected in the construction of a truncated distribution in (56) to ensure the Markovian structure of the source sequences is preserved. Subsequently, due to this constraint, the application of the strong converse technique by Tyagi and Watanabe in [13] was only amenable in analyzing the type-II error exponent at the relay. On the other hand, to analyze the type-II error exponent at the receiver, we need to carefully adapt the strong converse technique based on reverse hypercontractivity by Liu, van Handel and Verdú in [12]. Furthermore, to complete the proof, we carefully combine the single-letterization techniques in [12,13].

Another important take-home message is the techniques (or a subset of the techniques) used in this paper can be applied to strengthen the results of other multiterminal hypothesis testing against independence problems. If the source distribution has no Markov structure, it is usually the case that one can directly apply the technique by Tyagi and Watanabe [13] to obtain strong converse theorems. Such examples include [7–9]. On the other hand, if the source sequences admit Markovian structure, then it appears necessary to combine techniques in [12,13] to obtain strong converse theorems, just as it was done in this paper.

Finally, we discuss some avenues for future research. In this paper, we only derived the strong converse but not a second-order converse result as was done in ([12] (Section 4.4)) for the problem of hypothesis testing against independence with a communication constraint [1]. Thus, in the future, one may refine the proof in the current paper by deriving second-order converse or exact second-order asymptotics. Furthermore, one may also consider deriving strong converse theorems or simplifying existing strong converse proofs for hypothesis testing problems with both communication and privacy constraints such as that in [23] by using the techniques in the current paper. It is also interesting to explore whether current techniques can be applied to obtain strong converse theorems for hypothesis testing with zero-rate compression problems [3].

Author Contributions: Formal analysis, D.C. and L.Z.; Supervision, V.Y.F.T.; Writing—original draft, D.C. and L.Z.; Writing—review & editing, V.Y.F.T.

Funding: D.C. is supported by the China Scholarship Council with No. 201706090064 and the National Natural Science Foundation of China under Grant 61571122. L.Z. was supported by NUS RSB grants (C-261-000-207-532 and C-261-000-005-001).

Acknowledgments: The authors acknowledge Sadaf Salehkalaibar (University of Tehran) for drawing our attention to ([11] (Corollary 1)) and providing helpful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Achievability Proof of Proposition 1

Fix any joint distribution $Q_{XYZU_1U_2V} \in \mathcal{Q}_2$. Let (f'_1, g'_1) be an encoder-decoder pair with rate $R_1 = I_Q(U_1; X)$ for the hypothesis testing with communication constraint problem [1] (i.e., no receiver in Figure 1) such that the type-II error probability decays exponentially fast at speed no smaller than $E_1 = I_Q(U_1; Y)$ and the type-I error probability is vanishing, i.e., $\log N'_1 \leq nR_1$, $\beta'_2 \leq \exp(-nE_1)$ and $\beta'_1 \leq \varepsilon'_1$ for any $\varepsilon'_1 > 0$. Furthermore, let $(f''_1, f''_2, g''_1, g''_2)$ be a tuple of encoders and decoders with rates

$(R_1, R_2) = (I_Q(U_2; X), I_Q(V; Y))$ for the problem in Figure 1 such that the type-II error probability at the receiver decays exponentially fast at speed no smaller $E_2 = I_Q(V; Z)$ and type-I error probability at the receiver is vanishing, i.e., $\log N_1'' \leq nR_1$, $\log N_2'' \leq nR_2$, $\eta_2'' \leq \exp(-nE_2)$ and $\eta_1'' \leq \varepsilon_2'$ for any $\varepsilon_2' > 0$. Such tuples of encoders and decoders exist as proved in [1,11]. Furthermore, let $\mathcal{A}'_1 \subseteq \mathcal{X}^n \times \mathcal{Y}^n$ be the acceptance region associated with (f'_1, g'_1) at the relay and let $\mathcal{A}'_2 \subseteq \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$ be the acceptance region associated with $(f''_1, f''_2, g''_1, g''_2)$ at the receiver.

Now, let us partition the source space \mathcal{X}^n into two disjoint sets \mathcal{X}_1^n and \mathcal{X}_2^n such that $\mathcal{X}_1^n \cup \mathcal{X}_2^n = \mathcal{X}^n$, $P_X^n(\mathcal{X}_1^n) > 1 - \varepsilon_1$ and $P_X^n(\mathcal{X}_2^n) > 1 - \varepsilon_2$. We construct an (n, N_1, N_2) -code as follows. Given a source sequence X^n , if $X^n \in \mathcal{X}_1^n$, then encoder f'_1 is used; and if otherwise, the encoder f''_1 is used. Furthermore, an additional bit indicating whether $X^n \in \mathcal{X}_1^n$ is also sent to the relay and further forwarded to the receiver by the relay. Given encoded index M_1 , if $X^n \in \mathcal{X}_1^n$, the relay uses decoder g'_1 to make the decision; otherwise, if $X^n \in \mathcal{X}_2^n$, the relay declares hypothesis H_1 to be true. Furthermore, in both cases, the relay transmits an index M_2 using encoder f''_2 . Given the index M_2 , if $X^n \in \mathcal{X}_1^n$, the receiver declares hypothesis H_1 to be true; otherwise, the receiver uses decoder g''_2 to make the decision.

The performance of the constructed (n, N_1, N_2) -code is as follows. In terms of rates, we have

$$\log N_1 \leq nR_1 + 1, \tag{A1}$$

$$\log N_2 \leq nR_2 + 1. \tag{A2}$$

The type-I error probability at the relay satisfies that

$$1 - \beta_1 = P_{XY}^n\{\mathcal{A}'_1 \cap (\mathcal{X}_1^n \times \mathcal{Y}^n)\} \tag{A3}$$

$$\geq P_X^n\{\mathcal{X}_1^n\} - P_{XY}^n\{(\mathcal{A}'_1)^c\} \tag{A4}$$

$$\geq 1 - \varepsilon_1, \tag{A5}$$

where (A5) follows when n is sufficiently large and thus ε_1' can be made arbitrarily close to zero. Furthermore, the type-II error probability at the relay can be upper bounded as follows

$$\beta_2 = P_X^n P_Y^n\{\mathcal{A}'_1 \cap (\mathcal{X}_1^n \times \mathcal{Y}^n)\} \tag{A6}$$

$$\leq P_X^n P_Y^n\{\mathcal{A}'_1\} \tag{A7}$$

$$= \beta_2' \tag{A8}$$

$$\leq \exp(-nE_1'). \tag{A9}$$

Similarly, for n sufficiently large, the error probabilities at the receiver can be upper bounded as follows

$$\eta_1 = 1 - P_{XYZ}^n\{\mathcal{A}''_2 \cap (\mathcal{X}_2^n \times \mathcal{Y}^n \times \mathcal{Z}^n)\} \tag{A10}$$

$$\leq 1 - P_X^n(\mathcal{X}_2^n) + P_{XYZ}^n((\mathcal{A}''_2)^c) \tag{A11}$$

$$\leq \varepsilon_2, \tag{A12}$$

and

$$\eta_2 = P_X^n P_Y^n P_Z^n\{\mathcal{A}''_2 \cap (\mathcal{X}_2^n \times \mathcal{Y}^n \times \mathcal{Z}^n)\} \tag{A13}$$

$$\leq P_X^n P_Y^n P_Z^n\{\mathcal{A}''_2\} \tag{A14}$$

$$\leq \exp(-nE_2''). \tag{A15}$$

The achievability proof of Proposition 1 is now complete.

Appendix B. Proof of Lemma 1

Recall the definition of distribution $\bar{P}_{\tilde{M}_2}$ (see (103)). Noting that $P_{\tilde{M}_2}$ is the marginal distribution induced by $P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n \tilde{M}_1 \tilde{M}_2}$ (see (64)), we have that for any $\tilde{m}_2 \in \mathcal{M}_2$

$$P_{\tilde{M}_2}(\tilde{m}_2) = \sum_{y^n, m_1} P_{\tilde{Y}^n \tilde{M}_1}(y^n, m_1) 1\{\tilde{m}_2 = f_2(m_1, y^n)\}. \tag{A16}$$

Thus, applying the data processing inequality for the relative entropy, we have that

$$I(\tilde{M}_1; \tilde{Y}^n) = D(P_{\tilde{Y}^n \tilde{M}_1} \| P_{\tilde{Y}^n} P_{\tilde{M}_1}) \tag{A17}$$

$$\geq D(P_{\tilde{M}_2} \| \bar{P}_{\tilde{M}_2}). \tag{A18}$$

Using (A18) and following similar steps to the proof of weak converse in ([11] (Equationation (186))), we obtain

$$D(P_{\tilde{Z}^n \tilde{M}_2} \| P_{\tilde{Z}^n} \bar{P}_{\tilde{M}_2}) = I(\tilde{M}_2; \tilde{Z}^n) + D(P_{\tilde{M}_2} \| \bar{P}_{\tilde{M}_2}) \tag{A19}$$

$$\leq I(\tilde{M}_2; \tilde{Z}^n) + I(\tilde{M}_1; \tilde{Y}^n). \tag{A20}$$

Using (A20) and the definition of $R_{b,c,d,\gamma}^{(n)}$ in (113), we have the following lower bound for $R_{b,c,d,\gamma}^{(n)}$

$$\begin{aligned} R_{b,c,d,\gamma}^{(n)} &\geq -I(\tilde{M}_1; \tilde{Y}^n) + b(D(P_{\tilde{X}^n \tilde{Y}^n} \| P_{\tilde{X}^n \tilde{Y}^n}) + H(\tilde{X}^n, \tilde{Y}^n) - H(\tilde{X}^n, \tilde{Y}^n | \tilde{M}_1)) \\ &\quad - c(I(\tilde{M}_2; \tilde{Z}^n) + I(\tilde{M}_1; \tilde{Y}^n)) + d(D(P_{\tilde{X}^n \tilde{Y}^n} \| P_{\tilde{X}^n \tilde{Y}^n}) + H(\tilde{Y}^n) - h(\tilde{Y}^n | \tilde{M}_2)) \\ &\quad + \gamma(D(P_{\tilde{X}^n \tilde{Y}^n} \| P_{\tilde{X}^n \tilde{Y}^n}) + H(\tilde{Y}^n | \tilde{X}^n) - H(\tilde{Y}^n | \tilde{X}^n, \tilde{M}_1)). \end{aligned} \tag{A21}$$

The rest of the proof concerns single-letterizing each term in (A21). For this purpose, for each $j \in [n]$, we define two auxiliary random variables $U_j := (\tilde{M}_1, \tilde{X}^{j-1}, \tilde{Y}^{j-1})$ and $V_j := (\tilde{M}_2, \tilde{Y}^{j-1})$ and let J be a random variable which is distributed uniformly over the set $[n]$ and is independent of all other random variables.

Using standard single-letterization techniques as in [21], we obtain

$$I(\tilde{M}_1; \tilde{Y}^n) = \sum_{j \in [n]} I(\tilde{M}_1; \tilde{Y}_j | \tilde{Y}^{j-1}) \tag{A22}$$

$$\leq \sum_{j \in [n]} I(\tilde{M}_1, \tilde{Y}^{j-1}; \tilde{Y}_j) \tag{A23}$$

$$\leq \sum_{j \in [n]} I(\tilde{M}_1, \tilde{X}^{j-1}, \tilde{Y}^{j-1}; \tilde{Y}_j) \tag{A24}$$

$$= nI(U_J, J; \tilde{Y}_J), \tag{A25}$$

and

$$H(\tilde{X}^n, \tilde{Y}^n | \tilde{M}_1) = nH(\tilde{X}_J, \tilde{Y}_J | U_J, J). \tag{A26}$$

Furthermore, analogous to ([13] (Prop. 1)), we obtain that

$$H(\tilde{X}^n, \tilde{Y}^n) + D(P_{\tilde{X}^n \tilde{Y}^n} \| P_{\tilde{X}Y}^n) = \sum_{x^n, y^n} P_{\tilde{X}^n \tilde{Y}^n}(x^n, y^n) \log \frac{1}{P_{\tilde{X}Y}^n(x^n, y^n)} \tag{A27}$$

$$= \sum_{x^n, y^n} P_{\tilde{X}^n \tilde{Y}^n}(x^n, y^n) \sum_{j \in [n]} \log \frac{1}{P_{XY}(x_j, y_j)} \tag{A28}$$

$$= \sum_{j \in [n]} P_{\tilde{X}_j \tilde{Y}_j}(x_j, y_j) \log \frac{1}{P_{XY}(x_j, y_j)} \tag{A29}$$

$$= n(H(\tilde{X}_J, \tilde{Y}_J) + D(P_{\tilde{X}_J \tilde{Y}_J} \| P_{XY})). \tag{A30}$$

Subsequently, we can single-letterize $I(\tilde{M}_2; \tilde{Z}^n)$ as follows:

$$I(\tilde{M}_2; \tilde{Z}^n) = \sum_{j \in [n]} I(\tilde{M}_2; \tilde{Z}_j | \tilde{Z}^{j-1}) \tag{A31}$$

$$\leq \sum_{j \in [n]} I(\tilde{M}_2, \tilde{Z}^{j-1}, \tilde{Y}^{j-1}; \tilde{Z}_j) \tag{A32}$$

$$= \sum_{j \in [n]} I(\tilde{M}_2, \tilde{Y}^{j-1}; \tilde{Z}_j) \tag{A33}$$

$$= nI(V_J, J; \tilde{Z}_J), \tag{A34}$$

where (A33) follows from the Markov chain $\tilde{Z}^{j-1} - \tilde{M}_2 \tilde{Y}^{j-1} - \tilde{Z}_j$ implied by the joint distribution of $(\tilde{X}^n, \tilde{Y}^n, \tilde{Z}^n, \tilde{M}_1, \tilde{M}_2)$ in (64). Furthermore, using similar proof techniques to ([13] (Prop. 1)) and standard single-letterization techniques (e.g., in [4] or [21]), we obtain that

$$H(\tilde{Y}^n | \tilde{X}^n) + D(P_{\tilde{X}^n \tilde{Y}^n} \| P_{\tilde{X}Y}^n) \geq n(H(\tilde{Y}_J | \tilde{X}_J) + D(P_{\tilde{X}_J \tilde{Y}_J} \| P_{XY})), \tag{A35}$$

$$H(\tilde{Y}^n) + D(P_{\tilde{X}^n \tilde{Y}^n} \| P_{\tilde{X}Y}^n) \geq n(H(\tilde{Y}_J) + D(P_{Y_J} \| P_Y)), \tag{A36}$$

$$H(\tilde{Y}^n | \tilde{M}_2) = nH(\tilde{Y}_J | V_J, J), \tag{A37}$$

$$H(\tilde{Y}^n | \tilde{M}_1, \tilde{X}^n) \leq nH(\tilde{Y}_J | X_J, U_J, J). \tag{A38}$$

Let $U := (U_J, J)$, $V := (V_J, J)$, $X' := \tilde{X}_J$, $Y' := \tilde{Y}_J$ and $Z' := \tilde{Z}_J$. Using the joint distribution $P_{\tilde{X}^n \tilde{Y}^n \tilde{Z}^n \tilde{M}_1 \tilde{M}_2}$ in (64), we conclude that the joint distribution of random variables (X', Y', Z', U, V) , denoted by $Q_{X'Y'Z'UV}$, belongs to the set \mathcal{Q}_1 defined in (115). The proof of Lemma 1 is complete by combining (A21) to (A38) and noting that $I_Q(X', Y'; U) \geq I_Q(X'; U)$.

Appendix C. Proof of Lemma 2

Given any $\gamma \in \mathbb{R}_+$, let $Q_{XYZUV}^{(\gamma)}$ achieve the minimum in (117). Recall the definition of θ_n in (51) and define a new alphabet $\tilde{\mathcal{V}} := \mathcal{V} \cup \{v^*\}$. We then define a joint distribution $P_{Y\tilde{\mathcal{V}}}^{(\gamma)}$ by specifying the following (conditional) marginal distributions

$$P_{\tilde{\mathcal{V}}}^{(\gamma)}(v) := \frac{1}{1 + \theta_n} Q_V^{(\gamma)}(v) 1\{v \neq v^*\} + \frac{\theta_n}{1 + \theta_n} 1\{v = v^*\}, \tag{A39}$$

$$P_{Y|\tilde{\mathcal{V}}}^{(\gamma)}(y|v) := Q_{Y|V}^{(\gamma)}(y|v) 1\{v \neq v^*\} + \left(\frac{1 + \theta_n}{\theta_n} P_Y(y) - \frac{1}{\theta_n} Q_Y^{(\gamma)}(y) \right) 1\{v = v^*\}. \tag{A40}$$

Thus, the induced marginal distribution $P_Y^{(\gamma)}$ satisfies

$$P_Y^{(\gamma)}(y) = \sum_{v \in \mathcal{V}} P_{\tilde{V}}^{(\gamma)}(v) P_{Y|\tilde{V}}^{(\gamma)}(y|v) \tag{A41}$$

$$= \left(\sum_{v \in \mathcal{V}} \frac{1}{1 + \theta_n} Q_V^{(\gamma)}(v) Q_{Y|V}^{(\gamma)}(y|v) \right) + \left(P_Y(y) - \frac{1}{1 + \theta_n} Q_Y^{(\gamma)}(y) \right) \tag{A42}$$

$$= P_Y(y). \tag{A43}$$

Furthermore, let $P_{\tilde{V}|Y}^{(\gamma)}$ be induced by $P_{Y\tilde{V}}^{(\gamma)}$ and define the following distribution

$$P_{XYZU\tilde{V}}^{(\gamma)} = P_{XYZ} Q_{U|X}^{(\gamma)} P_{\tilde{V}|Y}^{(\gamma)}. \tag{A44}$$

Recall the definition of $R_{b,c,d}(\cdot)$ in (28). The following lemma lower bounds the difference between $R_{b,c,d}(Q_{XYZUV}^{(\gamma)})$ and $R_{b,c,d}(P_{XYZU\tilde{V}}^{(\gamma)})$ and is critical in the proof of Lemma 2.

Lemma A1. *When $\gamma = \sqrt{n}$, we have*

$$R_{b,c,d}(Q_{XYZUV}^{(\gamma)}) - R_{b,c,d}(P_{XYZU\tilde{V}}^{(\gamma)}) \geq \Theta\left(\frac{\log n}{n^{1/4}}\right). \tag{A45}$$

The proof of Lemma A1 is deferred to Appendix D.

Now, using the assumption that $Q_{XYZUV}^{(\gamma)}$ is a minimizer for $R_{b,c,d,\gamma}$ in (117), the fact that $\Delta_{b,d,\gamma}(Q_{XYZUV}^{(\gamma)}) \geq 0$ (see (116)) and the result in (A45), we conclude that when $\gamma = \sqrt{n}$,

$$R_{b,c,d,\gamma} = R_{b,c,d}(Q_{XYZUV}^{(\gamma)}) + \Delta_{b,d,\gamma}(Q_{XYZUV}^{(\gamma)}) \tag{A46}$$

$$\geq R_{b,c,d}(P_{XYZU\tilde{V}}^{(\gamma)}) + \Theta\left(\frac{\log n}{n^{1/4}}\right) \tag{A47}$$

$$\geq R_{b,c,d} + \Theta\left(\frac{\log n}{n^{1/4}}\right), \tag{A48}$$

where (A48) follows from the definition of $R_{b,c,d}$ in (29) and the fact that $P_{XYZU\tilde{V}}^{(\gamma)} \in \mathcal{Q}$ (see (24)).

The proof of Lemma 2 is complete by using (A48) and noting that when $\gamma = \sqrt{n}$,

$$\log \frac{1 - \varepsilon_1 - \varepsilon_2}{1 + 3\varepsilon_2 - \varepsilon_1} + (b + d + \gamma + 5) \log \frac{1 - \varepsilon_1 - \varepsilon_2}{2} = \Theta(\sqrt{n}). \tag{A49}$$

Appendix D. Proof of Lemma A1

In subsequent analyses, all distributions indicated by $P^{(\gamma)}$ are induced by $P_{XYZU\tilde{V}}^{(\gamma)}$. We have

$$D(Q_{XYU}^{(\gamma)} \| P_{XYU}^{(\gamma)}) = D(Q_{XY}^{(\gamma)} \| P_{XY}^{(\gamma)}) + I_{Q^{(\gamma)}}(U; Y|X). \tag{A50}$$

Recalling the definitions of $R_{b,c,d}$ in (29) and $R_{b,c,d,\gamma}$ in (117), we conclude that for any $\gamma \in \mathbb{R}_+$,

$$R_{b,c,d,\gamma} \leq R_{b,c,d} \leq b \log |\mathcal{X}| + d \log |\mathcal{Y}| =: a'. \tag{A51}$$

Using the definition of $\Delta_{b,d,\gamma}(Q_{XYZUV})$ in (116) and recalling that $Q_{XYZUV}^{(\gamma)}$ is a minimizer for $R_{b,c,d,\gamma}$, we have

$$\gamma D(Q_{XYU}^{(\gamma)} \| P_{XYU}^{(\gamma)}) \leq \Delta_{b,d,\gamma}(Q_{XYZUV}^{(\gamma)}) \tag{A52}$$

$$= R_{b,c,d,\gamma} - R_{b,c,d}(Q_{XYZUV}^{(\gamma)}) \tag{A53}$$

$$\leq a' + (c + 1) \log |\mathcal{Y}| + c \log |\mathcal{Z}| =: a. \tag{A54}$$

We can now upper bound $I_{P^{(\gamma)}}(\tilde{V}; Y)$ as follows:

$$I_{P^{(\gamma)}}(\tilde{V}; Y) = D(P_{Y|\tilde{V}}^{(\gamma)} \| P_Y^{(\gamma)} | P_{\tilde{V}}^{(\gamma)}) \tag{A55}$$

$$= D(P_{Y|\tilde{V}}^{(\gamma)} \| P_Y | P_{\tilde{V}}^{(\gamma)}) \tag{A56}$$

$$= \frac{1}{1 + \theta_n} D(Q_{Y|\tilde{V}}^{(\gamma)} \| P_Y | Q_V^{(\gamma)}) + \frac{\theta_n}{1 + \theta_n} D\left(\frac{1 + \theta_n}{\theta_n} P_Y - \frac{1}{\theta_n} Q_Y^{(\gamma)} \parallel P_Y\right) \tag{A57}$$

$$= \frac{1}{1 + \theta_n} (D(Q_{Y|\tilde{V}}^{(\gamma)} \| Q_Y^{(\gamma)} | Q_V^{(\gamma)}) + D(Q_Y^{(\gamma)} \| P_Y)) + \frac{\theta_n}{1 + \theta_n} D\left(\frac{1 + \theta_n}{\theta_n} P_Y - \frac{1}{\theta_n} Q_Y^{(\gamma)} \parallel P_Y\right) \tag{A58}$$

$$\leq \frac{1}{1 + \theta_n} I_{Q^{(\gamma)}}(V; Y) + \frac{1}{1 + \theta_n} \frac{a}{\gamma} + \frac{\theta_n}{1 + \theta_n} \log \mu, \tag{A59}$$

where (A56) follows from (A43), and (A59) follows from the result in (A54), the fact that $D(Q_Y^{(\gamma)} \| P_Y) \leq D(Q_{XYU}^{(\gamma)} \| P_{XYU}^{(\gamma)})$ and the definition of μ in (50). Thus, when $\gamma = \sqrt{n}$, recalling the definition of θ_n in (51), we have

$$I_{Q^{(\gamma)}}(V; Y) \geq I_{P^{(\gamma)}}(\tilde{V}; Y) - \frac{a}{\gamma} - \theta_n \log \mu \tag{A60}$$

$$= I_{P^{(\gamma)}}(\tilde{V}; Y) + \Theta\left(\frac{1}{\sqrt{n}}\right). \tag{A61}$$

Similar to (A59), we obtain

$$I_{P^{(\gamma)}}(\tilde{V}; Z) = D(P_{Z|\tilde{V}}^{(\gamma)} \| P_Z^{(\gamma)} | P_{\tilde{V}}^{(\gamma)}) \tag{A62}$$

$$= D(P_{Z|\tilde{V}}^{(\gamma)} \| P_Z | P_{\tilde{V}}^{(\gamma)}) \tag{A63}$$

$$= \frac{1}{1 + \theta_n} D(Q_{Z|\tilde{V}}^{(\gamma)} \| P_Z | Q_V^{(\gamma)}) + \frac{\theta_n}{1 + \theta_n} D\left(\frac{1 + \theta_n}{\theta_n} P_Z - \frac{1}{\theta_n} Q_Z^{(\gamma)} \parallel P_Z\right) \tag{A64}$$

$$= \frac{1}{1 + \theta_n} (D(Q_{Z|\tilde{V}}^{(\gamma)} \| Q_Z^{(\gamma)} | Q_V^{(\gamma)}) + D(Q_Z^{(\gamma)} \| P_Z)) + \frac{\theta_n}{1 + \theta_n} D\left(\frac{1 + \theta_n}{\theta_n} P_Z - \frac{1}{\theta_n} Q_Z^{(\gamma)} \parallel P_Z\right) \tag{A65}$$

$$\geq \frac{1}{1 + \theta_n} I_{Q^{(\gamma)}}(V; Z), \tag{A66}$$

where (A64) follows since $Q^{(\gamma)} \in \mathcal{Q}_1$ (see (115)) implies that $Q_{Z|Y}^{(\gamma)} = P_{Z|Y}$ and the Markov chains $Z - Y - X$ and $V - Y - Z$ holds and thus using (A39) to (A40), we have

$$P_{Z|\tilde{V}}^{(\gamma)}(z|v) = \frac{\sum_y P_{Z|Y}(z|y) P_{\tilde{V}}(v) P_{Y|\tilde{V}}(y|v)}{P_{\tilde{V}}(v)} \tag{A67}$$

$$= \frac{\sum_y Q_{Z|Y}^{(\gamma)}(z|y) Q_V^{(\gamma)}(v) Q_{Y|V}^{(\gamma)}(y|v)}{Q_V^{(\gamma)}(v)} \tag{A68}$$

$$= Q_{Z|V}^{(\gamma)}(z|v), \tag{A69}$$

and

$$P_{Z|\tilde{V}}^{(\gamma)}(z|v^*) = \frac{\sum_y P_{Z|Y}(z|y)P_{\tilde{V}}(v^*)P_{Y|\tilde{V}}(y|v^*)}{P_{\tilde{V}}(v^*)} \tag{A70}$$

$$= \sum_y Q_{Z|Y}^{(\gamma)}(z|y) \left(\frac{1+\theta_n}{\theta_n} P_Y(y) - \frac{1}{\theta_n} Q_Y^{(\gamma)}(y) \right) \tag{A71}$$

$$= \frac{1+\theta_n}{\theta_n} P_Z(z) - \frac{1}{\theta_n} Q_Z^{(\gamma)}(z), \tag{A72}$$

Therefore, we have

$$I_{Q^{(\gamma)}}(V; Z) \leq (1 + \theta_n) I_{P^{(\gamma)}}(\tilde{V}; Z) \tag{A73}$$

$$\leq I_{P^{(\gamma)}}(\tilde{V}; Z) + \theta_n \log |\mathcal{Z}| \tag{A74}$$

$$= I_{P^{(\gamma)}}(\tilde{V}; Z) + \Theta \left(\frac{1}{\sqrt{n}} \right). \tag{A75}$$

Let $\|P - Q\|$ be the ℓ_1 norm between P and Q regarded as vectors. Using Pinsker’s inequality, the result in (105), and the data processing inequality for the relative entropy [17], we obtain

$$\|Q_{UX}^{(\gamma)} - P_{UX}^{(\gamma)}\| \leq \sqrt{2 \log 2 \cdot D(Q_{UX}^{(\gamma)} \| P_{UX}^{(\gamma)})} \tag{A76}$$

$$\leq \sqrt{2 \log 2 \cdot D(Q_{XYU}^{(\gamma)} \| P_{XYU}^{(\gamma)})} \tag{A77}$$

$$\leq \sqrt{\frac{2a \log 2}{\gamma}}. \tag{A78}$$

From the support lemma ([21] (Appendix C)), we conclude that the cardinality of U can be upper bounded by a function depending only on $|\mathcal{X}|, |\mathcal{Y}|$ and $|\mathcal{Z}|$ (these alphabets are all finite). Thus, when $\gamma = \sqrt{n}$, invoking ([4] (Lemma 2.2.7)), we have

$$|H(Q_{UX}^{(\gamma)}) - H(P_{UX}^{(\gamma)})| \leq \sqrt{\frac{2a \log 2}{\gamma}} \log \frac{|\mathcal{U}||\mathcal{X}|}{\sqrt{\frac{2a \log 2}{\gamma}}} = \Theta \left(\frac{\log n}{n^{1/4}} \right). \tag{A79}$$

Similar to (A79), we have

$$|I_{Q^{(\gamma)}}(U; X) - I_{P^{(\gamma)}}(U; X)| \leq \Theta \left(\frac{\log n}{n^{1/4}} \right), \tag{A80}$$

$$|I_{Q^{(\gamma)}}(U; Y) - I_{P^{(\gamma)}}(U; Y)| \leq \Theta \left(\frac{\log n}{n^{1/4}} \right). \tag{A81}$$

Combining (A61), (A75), (A80) and (A81), when $\gamma = \sqrt{n}$, using the definition of $R_{b,c,d}(\cdot)$ in (28), we have

$$\begin{aligned} & R_{b,c,d}(Q_{XYZUV}^{(\gamma)}) \\ & \geq -(c+1)I_{Q^{(\gamma)}}(U; Y) + bI_{Q^{(\gamma)}}(U; X) - cI_{Q^{(\gamma)}}(V; Z) + dI_{Q^{(\gamma)}}(V; Y) \end{aligned} \tag{A82}$$

$$\geq -(c+1)I_{P^{(\gamma)}}(U; Y) + bI_{P^{(\gamma)}}(U; X) - cI_{P^{(\gamma)}}(\tilde{V}; Z) + dI_{P^{(\gamma)}}(\tilde{V}; Y) + \Theta \left(\frac{\log n}{n^{1/4}} \right) \tag{A83}$$

$$= R_{b,c,d}(P_{XYZUV}^{(\gamma)}) + \Theta \left(\frac{\log n}{n^{1/4}} \right). \tag{A84}$$

The proof of Lemma A1 is now complete.

References

1. Ahlswede, R.; Csiszár, I. Hypothesis testing with communication constraints. *IEEE Trans. Inf. Theory* **1986**, *32*, 533–542.
2. Han, T. Hypothesis testing with multiterminal data compression. *IEEE Trans. Inf. Theory* **1987**, *33*, 759–772.
3. Shalaby, H.M.; Papamarcou, A. Multiterminal detection with zero-rate data compression. *IEEE Trans. Inf. Theory* **1992**, *38*, 254–267.
4. Csiszár, I.; Körner, J. *Information Theory: Coding Theorems for Discrete Memoryless Systems*; Cambridge University Press: Cambridge, UK, 2011.
5. Han, T.S.; Amari, S. Statistical inference under multiterminal data compression. *IEEE Trans. Inf. Theory* **1998**, *44*, 2300–2324.
6. Tian, C.; Chen, J. Successive refinement for hypothesis testing and lossless one-helper problem. *IEEE Trans. Inf. Theory* **2008**, *54*, 4666–4681.
7. Wigger, M.; Timo, R. Testing against independence with multiple decision centers. In Proceedings of the IEEE SPCOM, Bengaluru, India, 12–15 June 2016; pp. 1–5.
8. Zhao, W.; Lai, L. Distributed detection with vector quantizer. *IEEE Trans. Signal Inf. Process. Netw.* **2016**, *2*, 105–119.
9. Xiang, Y.; Kim, Y.H. Interactive hypothesis testing with communication constraints. In Proceedings of the IEEE 50th Annual Allerton on Communication, Control, and Computing, Monticello, IL, USA, 1–5 October 2012; pp. 1065–1072.
10. Zhao, W.; Lai, L. Distributed testing with cascaded encoders. *IEEE Trans. Inf. Theory* **2018**, *64*, 7339–7348.
11. Salehkalaibar, S.; Wigger, M.; Wang, L. Hypothesis Testing Over the Two-Hop Relay Network. *IEEE Trans. Inf. Theory* **2019**, *65*, 4411–4433. doi:10.1109/TIT.2019.2897698.
12. Liu, J.; van Handel, R.; Verdú, S. Beyond the blowing-up lemma: Sharp converses via reverse hypercontractivity. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 943–947.
13. Tyagi, H.; Watanabe, S. Strong Converse using Change of Measure. *IEEE Trans. Inf. Theory* **2020**, in press.
14. Oohama, Y. Exponential strong converse for source coding with side information at the decoder. *Entropy* **2018**, *20*, 352.
15. Oohama, Y. Exponent function for one helper source coding problem at rates outside the rate region. In Proceedings of the 2015 IEEE International Symposium on Information Theory (ISIT), Hong Kong, China, 14–19 June 2015; pp. 1575–1579.
16. Oohama, Y. Exponential Strong Converse for One Helper Source Coding Problem. *Entropy* **2019**, *21*, 567.
17. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: New York, NY, USA, 2012.
18. Gu, W.; Effros, M. A strong converse for a collection of network source coding problems. In Proceedings of the 2009 IEEE International Symposium on Information Theory, Seoul, Korea, 28 June–3 July 2009; pp. 2316–2320. doi:10.1109/ISIT.2009.5205934.
19. Liu, J.; van Handel, R.; Verdú, S. Beyond the Blowing-Up Lemma: Optimal Second-Order Converses via Reverse Hypercontractivity. Available online: <http://web.mit.edu/jingbo/www/preprints/m-sl-blup.pdf> (accessed on 17 April 2019).
20. Salehkalaibar, S.; Wigger, M.; Wang, L. Hypothesis testing over cascade channels. In Proceedings of the 2017 IEEE Information Theory Workshop (ITW), Kaohsiung, Taiwan, 6–10 November 2017; pp. 369–373. doi:10.1109/ITW.2017.8277994.
21. El Gamal, A.; Kim, Y.H. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2011.
22. Raginsky, M.; Sason, I. Concentration of measure inequalities in information theory, communications, and coding. *Found. Trends® Commun. Inf. Theory* **2013**, *10*, 1–246.
23. Gilani, A.; Amor, S.B.; Salehkalaibar, S.; Tan, V.Y.F. Distributed Hypothesis Testing with Privacy Constraints. *Entropy* **2019**, *21*, 478.

