Hypothesis Testing In Multi-Hop Networks

Sadaf Salehkalaibar, *IEEE Member*, Michèle Wigger, *IEEE Senior Member*, Ligong Wang, *IEEE Member*

Abstract

Coding and testing schemes for binary hypothesis testing over three kinds of multi-hop networks are presented and their achievable type-II error exponents as functions of the available communication rates are characterized. The schemes are based on cascade source coding techniques and *unanimous-decision forwarding*, where terminals decide only on the null hypothesis if all previous terminals have decided on this hypothesis, and where they forward their decision to the next hop. The achieved exponent-rate region is analyzed by extending Han's approach to account for the unanimous-decision forwarding strategy and for the more complicated code constructions. The proposed coding and testing schemes are shown to attain the optimal type-II error exponent region for various instances of testing against independence on the single-relay multi-hop network, one instance of the K-relay multi-hop network, and one instance of a network with two parallel multi-hop networks that share a common receiver.

For the basic single-relay multi-hop network, the proposed scheme is further improved by means of binning. This improved scheme is again analyzed by extending Han's approach, and is shown to be optimal when testing against conditional independence under some Markov condition. For completeness, the paper also presents the previously missing analysis of the Shimokawa, Han and Amari binning scheme for the point-to-point hypothesis testing setup.

I. Introduction

Decreasing costs of sensors and improved hardware technologies facilitate a rapid increase of sensor applications, in particular as part of the future *Internet of Things (IoT)*. One of the major theoretical challenges in this respect are sensor networks with multiple sensors collecting correlated data that they communicate to one or multiple decision centers. Of special practical and theoretical interest here is the study of the tradeoff between the quality of the decisions taken at the centers and the available communication resources. This is the topic of this work, where we follow the approach in [1], [2]. That means decision centers have to decide on a binary hypothesis $\mathcal{H}=0$ or $\mathcal{H}=1$ that determines the underlying joint probability mass function (pmf) of all the terminals' observations. The goal is to characterize the set of possible *type-II error exponents* (i.e., the error exponent for deciding $\hat{\mathcal{H}}=0$ when in fact $\mathcal{H}=1$) as functions of the available communication rate so that the *type-I error probabilities* (i.e., error probabilites for deciding $\hat{\mathcal{H}}=1$ when in fact $\mathcal{H}=0$) vanish as the lengths of the observations grow. Previous works on this *exponent-rate region* considered communication scenarios over dedicated noise-free bit-pipes from one or many transmitters to a single decision center [1], [3], [4] or from a single transmitter to two decision centers [5]–[7]. The hypothesis testing problem from a signal processing perspective has been studied in several works [8]–[11]. Recently, simple interactive communication scenarios were also considered [7], [12], [13], as are dedicated *noisy communication channels* [5], [14]. Generally, the problem of identifying the exact type-II error exponents region is open; exact solutions were found for instances of *testing against independence* [1] and of *testing against conditional independence* [4]. Testing against independence refers to a scenario where under $\mathcal{H}=1$ the observations' joint pmf is the product of the marginal pmfs under $\mathcal{H}=0$, and testing against conditional independence refers to a scenario where the independence holds only conditional on some sequence that is observed at the receiver and that has the same distribution under both hypotheses.

The focus of this paper is on *multi-hop networks*, where some of the terminals cannot communicate directly but their communication needs to be relayed by other terminals. The simplest example of a multi-hop network is shown in Fig. 1, where communication from the transmitter to the receiver has to pass through the relay. Multi-hop networks are challenging mainly because relaying terminals have to find tradeoffs between the information they convey about their own observed data and that about the information they received from other terminals. In some cases, they must send complicated functions of their observations and the received information.

In this paper, we propose coding and testing schemes for the basic single-relay multi-hop network in Fig. 1, for its generalization with $K\geq 2$ relays in Fig. 2, and for the network with two parallel single-relay multi-hop channels that share a common receiver in Fig. 3. Our coding and testing schemes are inspired by Han's work on the point-to-point setup [2] and based on the following ideas. The transmitter and the relays quantize their observation sequences and convey these quantization indices to the subsequent relays and the final receiver. In contrast to Han's scheme, here we use advanced code constructions that have previously been proposed for source coding over cascade networks [15]. Moreover, in the second part of the paper, we exploit correlation between the terminals' observations to reduce communication rates by means of binning, similarly to the work by Shimokawa, Han, and Amari [3] for the point-to-point setup. In all our schemes, all relays apply an *unanimous-decision forwarding strategy*. This means they decide on the null-hypothesis $\mathcal{H}=0$ if, and only if, *all preceding relays* have also decided on $\mathcal{H}=0$, and the reconstructed quantization sequences are jointly typical with their own observations,

S. Salehkalaibar is with the Department of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran, s.saleh@ut.ac.ir,
M. Wigger is with LTCI, Telecom ParisTech, Université Paris-Saclay, 75013 Paris, michele.wigger@telecom-paristech.fr,
L. Wang is with ETIS, Université Paris Seine, Université de Cergy-Pontoise, ENSEA, CNRS, ligong.wang@ensea.fr.
Parts of the material in this paper will be presented at *IEEE Information Theory Workshop (ITW), Kaohsiung, Taiwan, November 2017.*

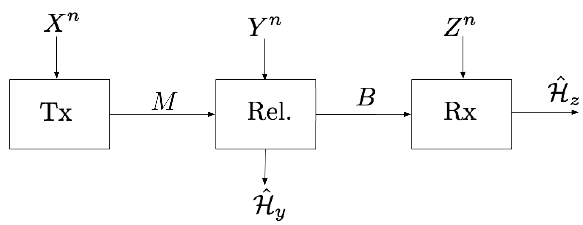

Fig. 1. Hypothesis testing over a single-relay multi-hop channel

and they forward this decision to the next-following terminal. The final receiver applies exactly the same unanimous-decision strategy: it only decides on the null hypothesis if all relays decided on $\mathcal{H} = 0$ and if the quantization sequences it reconstructs are jointly typical with its own observation. Applying the unanimous-decision forwarding strategy allows accumulating error exponents along the path from the initial transmitter to the receiver. We present a family of basic multi-hop networks (see Proposition 2) where the unanimous-decision forwarding strategy allows to increase the error exponent at the receiver by the error exponent at the relay.

For each of the three considered networks, i.e., for the basic single-relay network, the K-relay network, and the parallel network with a common receiver, a converse is presented that, for some instances of testing against independence, matches the exponent-rate region achieved by this scheme. These exact results allow for some interesting conclusions. For example, we elaborate on two examples for the basic single-relay network. In these examples, the optimal coding and testing scheme is rather simple when the communication link from the relay to the receiver is of high rate, but when this rate is more stringent, more involved cascade source coding techniques are needed. The examples allow also to obtain interesting operational explanations of the roles of the various terminals. Specifically, even when the observation at the transmitter is itself not useful at the receiver, not even jointly with the relay's observation, the transmitter can still improve the receiver's error exponent by coordinating the relay's communication to the receiver.

The binning extension is presented and analyzed only for the basic single-relay multi-hop network, and for reasons of completeness also for the point-to-point setup studied in [3]. In fact, [3] describes the scheme based on binning, but does not provide an analysis, which is presented here for completeness. The binning extension is simple conceptually and applies also to the K-relay multi-hop network and the parallel multi-hop network with a common receiver. However, we do not present these extensions here, because their analysis and statement of the achieved exponent-rate region are rather cumbersome. In fact, already for the basic single-relay network the resulting exponent-rate region has a rather complicated form. Nonetheless, we provide a converse proof showing that the attained exponent-rate region is optimal for an instance of testing against conditional independence.

The analyses presented in this paper differ from Han's analysis for the point-to-point setup [2] because of the more complicated code constructions and because of the unanimous decision strategy. In particular, to obtain our desired results, in our code construction we have to ensure that a given joint type is not "over-represented" among tuples of codewords, and in the analysis, we have to properly account for the correlation between the decisions taken at the various terminals. The modified Han analysis technique that we propose applies both to multi-hop relay networks with an arbitrary number of K relays, and to the set of two parallel multi-hop networks with a common receiver, and both in the presence and absence of binning. Moreover, the so obtained exponent-rate regions have a simple and succint description, which leads us to conjecture that the analysis and the structure of the result extend to more complicated networks.

We conclude this introduction with a summary of the main contributions, remarks on notation, and an outline of the paper.

A. Contributions

The main contributions of this article are:

- A first coding and testing scheme without binning is presented for the basic single-relay multi-hop network, and the achieved exponent-rate region is analyzed (see Section III, in particular Theorem 1).
- Matching converse results are proved for instances of testing against independence (Corollaries 1 and 2 and Propositions 2–4 in Section III).
- Examples are given where the optimal coding and testing schemes simplify considerably when there are abundant communication resources; however, they are complicated when resources are scarce (Examples 1 and 2 in Section III).
- The initial coding and testing scheme and its analysis are extended to multi-hop networks with an arbitrary number of K relays and to a network with two parallel single-relay multi-hop channels that share a common receiver (Section IV, in particular Theorems 2 and 3). The resulting exponent-rate regions are optimal in special cases of testing against conditional independence (Propositions 5 and 6).
- A detailed analysis of the Shimokawa-Han-Amari [3] coding and testing scheme with binning over a point-to-point channel is presented (Theorem 4 in Section VI).

- An improved coding scheme based on binning is proposed and analyzed for the single-relay multi-hop network (Section VI, and in particular Theorem 5). The exponent-rate region is shown to be optimal for a special instance of testing against conditional independence (Proposition 7).

B. Notation

We mostly follow the notation in [16]. Random variables are denoted by capital letters, e.g., X, Y, and their realizations by lower-case letters, e.g., x, y. Script symbols such as $\mathcal{X}$ and $\mathcal{Y}$ stand for alphabets of random variables, and $\mathcal{X}^n$ and $\mathcal{Y}^n$ for the corresponding n-fold Cartesian products. Sequences of random variables $(X_i, ..., X_j)$ and realizations $(x_i, \ldots, x_j)$ are abbreviated by X_i^j and x_i^j. When $i = 1$, then we also use X^j and x^j instead of X_1^j and x_1^j.

Generally, we write the probability mass function (pmf) of a discrete random variable X as P_X; to indicate the pmf under hypothesis $\mathcal{H} = 1$, we also use Q_X. The conditional pmf of X given Y is written as $P_{X|Y}$, or as $Q_{X|Y}$ when $\mathcal{H} = 1$. The term $D(P\|Q)$ stands for the Kullback-Leibler (KL) divergence between two pmfs P and Q over the same alphabet. We use $\mathrm{tp}(\cdot)$ to denote the *joint type* of a tuple. For a joint type π_{ABC} over alphabet $\mathcal{A} \times \mathcal{B} \times \mathcal{C}$, we denote by $I_{\pi_{ABC}}(A; B|C)$ the mutual information assuming that the random triple (A, B, C) has pmf π_{ABC}; similarly for the entropy $H_{\pi_{ABC}}(A)$ and the conditional entropy $H_{\pi_{ABC}}(A|B)$. Sometimes we abbreviate π_{ABC} by π. Also, when π_{ABC} has been defined and is clear from the context, we write π_A or π_{AB} for the corresponding subtypes. When the type π_{ABC} coincides with the actual pmf of a triple (A, B, C), we omit the subscript and simply write $H(A)$, $H(A|B)$, and $I(A; B|C)$.

For a given P_X and a constant $\mu > 0$, let $\mathcal{T}_\mu^n(P_X)$ be the set of *μ-typical sequences* in $\mathcal{X}^n$:

$$\mathcal{T}_\mu^n(P_X) = \left\{ x^n \colon \left| \frac{|\{i \colon x_i = x\}|}{n} - P_X(x) \right| \le \mu P_X(x), \quad \forall x \in \mathcal{X} \right\}, \tag{1}$$

where $|\{i \colon x_i = x\}|$ is the number of positions where the sequence x^n equals x. Similarly, $\mathcal{T}_\mu^n(P_{XY})$ stands for the set of *jointly μ-typical sequences* whose definition is as in (1) where X needs to be replaced by the pair (X, Y).

The expectation operator is written as $\mathbb{E}[\cdot]$. We abbreviate *independent and identically distributed* by *i.i.d.*. The notation $\mathcal{U}\{a, \ldots, b\}$ is used to indicate a uniform distribution over the set $\{a, \ldots, b\}$; for the uniform distribution over $\{0, 1\}$ we also use $\mathcal{B}(1/2)$. Finally, the $\log$ function is taken with base 2.

C. Paper Outline

The remainder of the paper is organized as follows. Section II describes the hypothesis testing setup over the basic multi-hop network. The following Section III presents a coding and testing scheme for this setup and the corresponding exponent-rate region. It also presents special cases where this exponent-rate region is optimal. Section IV presents similar results for the multi-hop network with an arbitrary number of relays and Section V presents results for the network consisting of two parallel multi-hop networks with a common receiver. Section VI presents coding and testing schemes for hypothesis testing with binning over the point-to-point channel and the basic multi-hop network. It also presents an optimality result. The paper is concluded in Section VII and by technical appendices.

II. Basic Multi-Hop Setting

Consider the multi-hop hypothesis testing problem with three terminals in Fig. 1. The first terminal in the system, the *transmitter*, observes the sequence $X^n \triangleq (X_1, \ldots, X_n)$, the second terminal, the *relay*, observes the sequence $Y^n \triangleq (Y_1, \ldots, Y_n)$, and the third terminal, the *receiver*, observes the sequence $Z^n = (Z_1, \ldots, Z_n)$. Under the null hypothesis

$$\mathcal{H} = 0\colon \quad (X^n, Y^n, Z^n) \sim \text{i.i.d. } P_{XYZ} \tag{2}$$

whereas under the alternative hypothesis

$$\mathcal{H} = 1\colon \quad (X^n, Y^n, Z^n) \sim \text{i.i.d. } Q_{XYZ} \tag{3}$$

for two given pmfs P_{XYZ} and Q_{XYZ}.

The problem encompasses a noise-free bit-pipe of rate R from the transmitter to the relay and a noise-free bit pipe of rate T from the relay to the receiver. That means, after observing X^n, the transmitter computes the message $M = \phi^{(n)}(X^n)$ using a possibly stochastic encoding function $\phi^{(n)} \colon \mathcal{X}^n \to \{0, ..., \lfloor 2^{nR} \rfloor\}$ and sends it over the noise-free bit pipe to the relay. The relay, after observing Y^n and receiving M, computes the message $B = \phi_y^{(n)}(M, Y^n)$ using a possibly stochastic encoding function $\phi_y^{(n)} \colon \mathcal{Y}^n \times \{0, ..., \lfloor 2^{nR} \rfloor\} \to \{0, ..., \lfloor 2^{nT} \rfloor\}$ and sends it over the noise-free bit pipe to the receiver.

The goal of the communication is that, based on their own observations and based on the received messages, the relay and the receiver can decide on the hypothesis $\mathcal{H}$. The relay thus produces the guess

$$\hat{\mathcal{H}}_y = g_y^{(n)}(Y^n, M), \tag{4}$$

using a decoding function $g_y^{(n)}\colon \mathcal{Y}^n \times \{0,....,\lfloor 2^{nR}\rfloor\} \to \{0,1\}$, and the receiver produces the guess

$$\hat{\mathcal{H}}_z = g_z^{(n)}(Z^n, B), \tag{5}$$

using a decoding function $g_z^{(n)}\colon \mathcal{Z}^n \times \{0,...,\lfloor 2^{nT}\rfloor\} \to \{0,1\}$.

Definition 1: For each $\epsilon \in (0,1)$, we say that the exponent-rate tuple (η,θ,R,T) is ϵ-achievable if there exists a sequence of encoding and decoding functions $(\phi^{(n)},\phi_y^{(n)},g_y^{(n)},g_z^{(n)})$, $n=1,2,\ldots$, such that the corresponding sequences of type-I and type-II error probabilities at the relay

$$\gamma_n \triangleq \Pr[\hat{\mathcal{H}}_y = 1|\mathcal{H}=0], \tag{6}$$

$$\zeta_n \triangleq \Pr[\hat{\mathcal{H}}_y = 0|\mathcal{H}=1], \tag{7}$$

and at the receiver

$$\alpha_n \triangleq \Pr[\hat{\mathcal{H}}_z = 1|\mathcal{H}=0], \tag{8}$$

$$\beta_n \triangleq \Pr[\hat{\mathcal{H}}_z = 0|\mathcal{H}=1], \tag{9}$$

satisfy

$$\alpha_n \le \epsilon, \tag{10}$$

$$\gamma_n \le \epsilon, \tag{11}$$

and

$$-\varlimsup_{n\to\infty} \frac{1}{n}\log\beta_n \ge \theta, \tag{12}$$

$$-\varlimsup_{n\to\infty} \frac{1}{n}\log\zeta_n \ge \eta. \tag{13}$$

Definition 2: For given rates (R,T), we define the exponent-rate region $\mathcal{E}(R,T)$ as the closure of all non-negative pairs (η,θ) for which (η,θ,R,T) is ϵ-achievable for every $\epsilon \in (0,1)$.

III. A Simple Coding and Testing Scheme without Binning

A. Preliminaries

Fix $\mu > 0$, an arbitrary blocklength n, and joint conditional pmfs $P_{SU|X}$ and $P_{V|SUY}$ over finite auxiliary alphabets $\mathcal{S}$, $\mathcal{U}$, and $\mathcal{V}$. Define the joint pmf $P_{SUVXYZ} = P_{XYZ}P_{SU|X}P_{V|SUY}$ and the following nonnegative rates, which are calculated according to the chosen distribution and μ:

$$R_s \triangleq I(X;S) + \mu, \tag{14}$$

$$R_u \triangleq I(U;X|S) + \mu, \tag{15}$$

$$R_v \triangleq I(V;Y,U|S) + \mu. \tag{16}$$

Later, we shall choose the joint distributions in such a way that $R = R_u + R_s$ and $T = R_s + R_v$.

Lemma 1: For every sufficiently large blocklength n, it is possible to create codebooks

$$\mathcal{C}_S = \left\{ s^n(i)\colon i \in \{1,...,\lfloor 2^{nR_s}\rfloor\}\right\} \tag{17}$$

and for each $i \in \{1,\ldots,\lfloor 2^{nR_s}\rfloor\}$:

$$\mathcal{C}_U(i) = \left\{ u^n(j|i)\colon j \in \{1,...,\lfloor 2^{nR_u}\rfloor\}\right\} \tag{18}$$

$$\mathcal{C}_V(i) = \left\{ v^n(k|i)\colon k \in \{1,...,\lfloor 2^{nR_v}\rfloor\}\right\} \tag{19}$$

such that the following three properties hold:

1) Codebooks $\mathcal{C}_S$ and $\{\mathcal{C}_U(i), i \in \{1,...,\lfloor 2^{nR_s}\rfloor\}\}$ cover the source sequence X^n with high probability:

$$\Pr\left[\exists (i,j)\colon (s^n(i),u^n(j|i),X^n) \in \mathcal{T}^n_{\mu/4}(P_{SUX})\right] > 1-\epsilon/4. \tag{20}$$

2) If codebooks $\mathcal{C}_S$ and $\mathcal{C}_U(i)$ cover Y^n for some i, then also codebooks $\mathcal{C}_S$, $\mathcal{C}_U(i)$, and $\mathcal{C}_V(i)$ cover Y^n with high probability:

$$\Pr\left[\exists k\colon (s^n(i),u^n(j|i),v^n(k|i),Y^n) \in \mathcal{T}^n_{\mu/2}(P_{SUVY}) \middle| (s^n(i),u^n(j|i),Y^n) \in \mathcal{T}^n_{3\mu/8}(P_{SUY})\right] > 1-\epsilon/4. \tag{21}$$

3) No joint type π_{SUV} is "over-represented" in the triple of codebooks $\mathcal{C}_S$, $\{\mathcal{C}_U\}$, and $\{\mathcal{C}_V\}$ in the sense that, for all π_{SUV},

$$\left|\left\{(i,j,k)\colon \quad \mathrm{tp}(s^n(i),u^n(j|i),v^n(k|i)) = \pi_{SUV}\right\}\right| \le 2^{n(R_s+R_u+R_v-I_{\pi_{SUV}}(U;V|S)+\mu)}. \tag{22}$$

Proof: A standard random code construction satisfies these properties *on average*. It can then be argued that a deterministic code construction satisfying this property must exist. A detailed proof of the lemma is given in Appendix A. ■

B. Description of Coding and Testing Scheme

Choose a codebook satisfying the three conditions in Lemma 1.

Transmitter: Given that it observes the sequence x^n, the transmitter looks for a pair of indices (i,j) such that

$$(s^n(i), u^n(j|i), x^n) \in \mathcal{T}^n_{\mu/4}(P_{SUX}). \tag{23}$$

If successful, it picks one such pair uniformly at random and sends

$$m = (i,j) \tag{24}$$

over the noise-free bit pipe. Otherwise, it sends $m = 0$.

Relay: Assume that the relay observes the sequence y^n and receives the message m. If $m = 0$, it declares $\hat{\mathcal{H}}_y = 1$ and sends $b = 0$ over the noise-free bit pipe to the receiver. Otherwise, it looks for an index k such that

$$(s^n(i), u^n(j|i), v^n(k|i), y^n) \in \mathcal{T}^n_{\mu/2}(P_{SUVY}). \tag{25}$$

If such an index k exists, the relay sends the pair

$$b = (i,k) \tag{26}$$

over the noise-free bit pipe to the receiver. Otherwise, it sends the message $b = 0$.

Receiver: Assume that the receiver observes z^n and receives message b from the relay. If $b = 0$, the receiver declares $\hat{\mathcal{H}}_z = 1$. Otherwise, it checks whether

$$(s^n(i), v^n(k|i), z^n) \in \mathcal{T}^n_{\mu}(P_{SVZ}). \tag{27}$$

If the typicality check is successful, the receiver declares $\hat{\mathcal{H}}_z = 0$. Otherwise, it declares $\hat{\mathcal{H}}_z = 1$.

Remark 1: Compared to Han's original scheme for the point-to-point setup [2], our scheme for the multi-hop setup contains three main novelties:

- *Cascade source coding:* Point-to-point source coding is replaced by source coding for cascade channels [15], [17]–[19].
- *Unanimous-decision forwarding:* Terminals decide on the null hypothesis only if all previous terminals decided on this hypothesis, and they forward their decision to the following terminal.
- *Codebooks with regular joint types:* In the chosen codebooks, a limited number of *codeword triples* $(s^n(i), u^n(j|i), v^n(k|i))$ have the same joint type. Without this assumption, in the following Theorem 1, the right-hand side of (29) would be decreased by $I(U;V|S)$.

C. Achievable Exponent Region

Theorem 1: The scheme described in the previous subsection shows that the exponent-rate region $\mathcal{E}(R,T)$ contains all nonnegative pairs (η,θ) that satisfy

$$\eta \leq \min_{\substack{\tilde{P}_{SUXY}: \\ \tilde{P}_{SUX}=P_{SUX} \\ \tilde{P}_{SUY}=P_{SUY}}} D(\tilde{P}_{SUXY} \| P_{SU|X} Q_{XY}), \tag{28}$$

$$\theta \leq \min_{\substack{\tilde{P}_{SUVXYZ}: \\ \tilde{P}_{SUX}=P_{SUX} \\ \tilde{P}_{SVUY}=P_{SVUY} \\ \tilde{P}_{SVZ}=P_{SVZ}}} D(\tilde{P}_{SUVXYZ} \| P_{SU|X} P_{V|SUY} Q_{XYZ}), \tag{29}$$

for some auxiliary random variables (U,S,V) satisfying the Markov chains $(U,S) \to X \to (Y,Z)$ and $V \to (S,U,Y) \to Z$ and the rate constraints

$$R \geq I(U,S;X), \tag{30}$$

$$T \geq I(X;S) + I(V;Y,U|S). \tag{31}$$

Proof: See Appendix B. ∎

The following two subsections discuss Theorem 1 for some interesting special cases.

D. Optimality Results and Examples when $X \to Y \to Z$ under both Hypotheses

We start with a setup where both pmfs P_{XYZ} and Q_{XYZ} satisfy the Markov chain $X \to Y \to Z$. Assume thus that the pmfs P_{XYZ} and Q_{XYZ} decompose as

$$P_{XYZ} = P_X \cdot P_{Y|X} \cdot P_{Z|Y}, \tag{32}$$
$$Q_{XYZ} = Q_X \cdot Q_{Y|X} \cdot Q_{Z|Y}. \tag{33}$$

In this case, also our achievable exponent-rate region decomposes.

Proposition 1: When (32) and (33) hold, the exponent-rate region $\mathcal{E}(R,T)$ contains all nonnegative pairs (η,θ) that satisfy

$$\eta \leq \min_{\substack{\tilde{P}_{UXY}: \\ \tilde{P}_{UX}=P_{UX} \\ \tilde{P}_{UY}=P_{UY}}} D(\tilde{P}_{UXY} \| P_{U|X} Q_{XY}), \tag{34}$$
$$\theta \leq \min_{\substack{\tilde{P}_{UXY}: \\ \tilde{P}_{UX}=P_{UX} \\ \tilde{P}_{UY}=P_{UY}}} D(\tilde{P}_{UXY} \| P_{U|X} Q_{XY}) + \min_{\substack{\tilde{P}_{VYZ}: \\ \tilde{P}_{VY}=P_{VY} \\ \tilde{P}_{VZ}=P_{VZ}}} D(\tilde{P}_{VYZ} \| P_{V|Y} P_Y Q_{Z|Y}), \tag{35}$$

for some auxiliary random variables (U,V) satisfying the Markov chains $U \to X \to Y$ and $V \to Y \to Z$ and the rate constraints

$$R \geq I(U;X), \tag{36}$$
$$T \geq I(V;Y). \tag{37}$$

In other words, under assumptions (32) and (33), in Theorem 1 one can restrict without loss in optimality to auxiliaries (S,U,V) so that $S=0$ and the Markov chain $V \to Y \to U$ holds.

Proof: See Appendix C. ∎

Notice that the error exponent θ at the receiver (35) equals the sum of the error exponent at the relay (34) and the error exponent achieved by Han's point-to-point scheme applied to the isolated relay-receiver network (without the transmitter terminal) when the relay's observation is modified to being i.i.d. according to P_Y under both hypotheses. This accumulation of error exponents is due to the applied unanimous-decision forwarding strategy.

The exponent-rate region in Proposition 1 is exact in some special cases, as we discuss in the following.

1) Special Case 1: Same P_{XY} under both Hypotheses:

Corollary 1: Assume (32) and

$$Q_{XYZ} = P_{XY} \cdot P_Z. \tag{38}$$

Under these conditions, the exponent-rate region $\mathcal{E}(R,T)$ is the set of all nonnegative pairs (η,θ) that satisfy

$$\eta = 0, \tag{39}$$
$$\theta \leq I(V;Z), \tag{40}$$

for an auxiliary random variable V satisfying the Markov chain $V \to Y \to Z$ and the rate constraint

$$T \geq I(V;Y). \tag{41}$$

(No constraint is imposed on the rate R.)

Proof: Achievability follows by specializing Proposition 1 to $U=0$ (deterministically). The converse is standard: Fix sequences of encoding and decoding functions $\{\phi^{(n)}, \phi_y^{(n)}, g_y^{(n)}, g_z^{(n)}\}$, and notice that for any $\epsilon>0$ and sufficiently large n:

$$-\frac{1}{n}\log \eta_n \leq \frac{1}{n} D(P_{MY^n|\mathcal{H}=0} \| P_{MY^n|\mathcal{H}=1}) + \epsilon = \epsilon$$
$$-\frac{1}{n}\log \beta_n \leq \frac{1}{n} D(P_{BZ^n|\mathcal{H}=0} \| P_{BZ^n|\mathcal{H}=1}) + \epsilon = I(B;Z^n) + \epsilon = \frac{1}{n}\sum_{t=1}^{n} I(B,Z^{t-1};Z_t) + \epsilon \leq \frac{1}{n}\sum_{t=1}^{n} I(B,Y^{t-1};Z_t) + \epsilon,$$

and

$$T = \frac{1}{n}H(B) \geq \frac{1}{n} I(B;Y^n) = \frac{1}{n}\sum_{t=1}^{n} I(B,Y^{t-1};Y_t). \tag{42}$$

The proof is finalized by introducing appropriate auxiliary random variables and taking $\epsilon \to 0$. ∎

2) Special Case 2: Same P_{YZ} under both Hypotheses:

Corollary 2: Assume (32) and

$$Q_{XYZ} = P_X \cdot P_{YZ}. \tag{43}$$

Under these conditions, the exponent-rate region $\mathcal{E}(R,T)$ is the set of all nonnegative pairs (η,θ) that satisfy

$$\eta \leq I(U;Y) \tag{44}$$

$$\theta \leq I(U;Y), \tag{45}$$

for an auxiliary random variable V satisfying the Markov chain $V \to Y \to Z$ and the rate constraint

$$R \geq I(U;X). \tag{46}$$

(No constraint is imposed on the rate T.)

Proof: Achievability follows by specializing Proposition 1 to $V = 0$. The converse is similar to the proof of the converse of Corollary 1. ∎

3) Special Case 3: Testing against Independence: Consider now the "testing against-independence" version of (32) and (33).

Proposition 2: Assume (32) and

$$Q_{XYZ} = P_X \cdot P_Y \cdot P_Z. \tag{47}$$

Under these conditions, the exponent-rate region $\mathcal{E}(R,T)$ is the set of all nonnegative pairs (η,θ) that satisfy

$$\eta \leq I(U;Y), \tag{48}$$

$$\theta \leq I(U;Y) + I(V;Z), \tag{49}$$

for a pair of auxiliary random variables U and V satisfying the Markov chains $U \to X \to Y$ and $V \to Y \to Z$ and the rate constraints

$$R \geq I(U;X) \tag{50}$$

$$T \geq I(V;Y). \tag{51}$$

Proof: Achievability follows from Proposition 1. The converse is proved in Appendix D. ∎

The error exponent at the receiver (49) equals the sum of the error exponent at the relay and the error exponent attained at the receiver if the relay did not forward its decision.

E. Optimality Results and Examples when $X \to Z \to Y$ under both Hypotheses

We treat two special cases: 1) Same P_{YZ} under both hypotheses, and 2) Same P_{XZ} under both hypotheses. Combined with the Markov chain $X \to Z \to Y$, these assumptions imply that the receiver cannot improve its error exponent by learning information about the observations at the relay (for case 1)) or about the observations at the transmitter (for case 2). As the following results and discussions show, these observations can nevertheless be useful in both cases.

1) Special Case 1: Same P_{YZ} under both Hypotheses: Consider first the setup where the pmfs P_{XYZ} and Q_{XYZ} decompose as

$$P_{XYZ} = P_{X|Z} \cdot P_{YZ}, \tag{52}$$

$$Q_{XYZ} = P_X \cdot P_{YZ}. \tag{53}$$

Since the pair of sequences (Y^n, Z^n) has the same joint distribution under both hypotheses, no positive error-exponent θ is possible when the message B sent from the relay to the receiver is only a function of Y^n but not of the incoming message M. The structure of (52) and (53) might even suggest that Y^n is not useful at the receiver and the optimal strategy for the relay is to simply forward a function of its incoming message M. Proposition 3 shows that this strategy is optimal when $T \geq R$, i.e., when the relay can forward the entire message to the receiver. On the other hand, Example 1 shows that it can be suboptimal when $T < R$.

Proposition 3: Assume conditions (52) and (53) and

$$T \geq R. \tag{54}$$

Then the exponent-rate region $\mathcal{E}(R,T)$ is the set of all nonnegative pairs (η,θ) that satisfy

$$\eta \leq I(S;Y), \tag{55}$$

$$\theta \leq I(S;Z), \tag{56}$$

for some auxiliary random variable S satisfying the Markov chain $S \to X \to (Y,Z)$ and the rate constraint

$$R \geq I(S;X). \tag{57}$$

Proof: For achievability, specialize Theorem 1 to $S = U = V$ and simplify. The converse is proved in Appendix E. ■

We next consider an example that satisfies assumptions (52) and (53), but not (54). We focus on $R = H(X)$, which implies that the transmitter can reliably describe the sequence X^n to the relay.

Example 1: Let under both hypotheses $\mathcal{H} = 0$ and $\mathcal{H} = 1$:

$$X \sim \mathcal{B}(1/2) \quad \text{and} \quad Y \sim \mathcal{U}\{0,1,2\}$$

independent of each other. Also, let $N_0 \sim \mathcal{B}(1/3)$ and $N_1 \sim \mathcal{B}(1/2)$ be independent of each other and of the pair (X,Y), and

$$Z = (Z', Y) \qquad \text{and} \qquad Z' = \begin{cases} X \oplus N_0 & \text{if } Y = 0 \text{ and } \mathcal{H} = 0 \\ N_1 & \text{otherwise} \end{cases}.$$

Notice that the triple (X,Y,Z) satisfies conditions (52) and (53). Moreover, since the joint pmf of the pair of sequences (X^n, Y^n) is the same under both hypotheses, the error exponent η cannot be larger than zero, and we focus on the error exponent θ achievable at the receiver. Notice also that the joint pmf (X_t, Y_t, Z_t) is the same under both hypotheses except for positions $t \in \{1, \ldots, n\}$ for which $Y_t = 0$. The idea of our scheme is thus that the relay describes only the symbols

$$\{X_t \colon Y_t = 0, \text{ for } \; t = 1, \ldots, n\}$$

to the receiver. In other words, we specialize the scheme in Subsection III-B to the choice of random variables

$$\begin{aligned} S &= 0 \\ U &= X \\ V &= \begin{cases} X & \text{if } Y = 0 \\ 0 & \text{otherwise} \end{cases}. \end{aligned}$$

Evaluating Theorem 1 for this choice proves achievability of the error exponent at the receiver

$$\begin{aligned} \min_{\substack{\tilde{P}_{VXYZ}: \\ \tilde{P}_{VXY} = P_{VXY} \\ \tilde{P}_{VZ} = P_{VZ}}} D(\tilde{P}_{VXYZ} \| P_{V|XY} P_X P_{YZ}) &\overset{(a)}{=} \min_{\substack{\tilde{P}_{VXYZ}: \\ \tilde{P}_{VXY} = P_{VXY} \\ \tilde{P}_{VZ} = P_{VZ}}} \mathbb{E}_{P_{VXY}}[D(\tilde{P}_{Z|VXY} \| P_{Z|Y})] \\ &\overset{(b)}{=} P_Y(0) \cdot \min_{\substack{\tilde{P}_{XZ|Y=0}: \\ \tilde{P}_{XZ} = P_{XZ}}} \mathbb{E}_{\tilde{P}_{X|Y=0}}[D(\tilde{P}_{Z|XY=0} \| P_{Z|Y=0})] \\ &\overset{(c)}{=} P_Y(0) \cdot \min_{\substack{\tilde{P}_{XZ'|Y=0}: \\ \tilde{P}_{XZ'Y} = P_{XZ'Y}}} \mathbb{E}_{\tilde{P}_{X|Y=0}}[D(\tilde{P}_{Z'|XY=0} \| P_{Z'|Y=0})] \\ &= P_Y(0) \cdot I(X; Z'|Y=0) = 1/3(1 - H_{\mathrm{b}}(1/3)), \end{aligned} \tag{58}$$

where (a) holds because $\tilde{P}_{VXY} = P_{VXY}$; (b) holds because unless $Y = 0$, the choice $\tilde{P}_{Z|VXY} = P_{Z|Y}$ is valid and thus the minimum value of the KL-divergence is 0, and because if $Y = 0$, then $V = X$; and (c) holds because $Z = (Z', Y)$.

The scheme requires rates

$$R = H(X) = 1$$

and

$$T = I(V; XY) = H(V) = H_{\mathrm{b}}(1/6).$$

By Proposition 3, the error exponent in (58) coincides with the largest error exponent that is achievable when $R \geq H(X) = 1$ and $T \geq R$. We argue in the following that this error exponent cannot be achieved when the relay simply forwards a function of the message M to the receiver. By the optimal error exponent for point-to-point setups, such a strategy can only achieve exponents of the form:

$$\begin{aligned} \max_{P_{S|X} \colon T \geq I(S;X)} I(S;Z) &\overset{(a)}{=} \max_{P_{S|X} \colon T \geq I(S;X)} I(S; Z'|Y) \\ &= 1 - \min_{P_{S|X} \colon T \geq I(S;X)} H(Z'|Y, S) \end{aligned}$$

$$
\begin{aligned}
&= 1 - \min_{P_{S|X}\colon H_\mathrm{b}(X|S)\geq 1-T} \frac{1}{3}\cdot H_\mathrm{b}(X\oplus N_0|S) - \frac{2}{3} \\
&= \frac{1}{3}\left(1 - \min_{P_{S|X}\colon H_\mathrm{b}(X|S)\geq 1-T} H_\mathrm{b}(X\oplus N_0|S)\right) \\
&\overset{(b)}{\leq} \frac{1}{3}\left(1 - \min_{P_{S|X}\colon H_\mathrm{b}(X|S)\geq 1-T} H_\mathrm{b}\left(H_\mathrm{b}^{-1}(H_\mathrm{b}(X|S))\star\frac{1}{3}\right)\right) \\
&\overset{(c)}{\leq} \frac{1}{3}\left(1 - H_\mathrm{b}\left(H_\mathrm{b}^{-1}(1-T)\star\frac{1}{3}\right)\right)
\end{aligned}
\tag{59}
$$

where $\star$ denotes the operation

$$
a\star b = ab + (1-a)(1-b), \quad a,b\in[0,1]. \tag{60}
$$

In the above, (a) follows because $Z=(Z',Y)$ and because (X,S) are independent of Y; (b) follows by Ms. Gerber's lemma [16]; and (c) follows because the function $x\mapsto H_\mathrm{b}(H_\mathrm{b}^{-1}(x)\star 1/3)$ is increasing in $x\in[0,1]$. Since the right-hand side is strictly increasing in $T\in[0,1]$, the error exponent in (59) is strictly smaller than $1/3(1-H_\mathrm{b}(1/3))$ whenever $T<1$, which proves the desired result.

2) Special Case 2: Same P_{XZ} under both Hypotheses: Consider next a setup where

$$
P_{XYZ} = P_{XZ}\cdot P_{Y|Z}, \tag{61}
$$

$$
Q_{XYZ} = P_{XZ}\cdot P_{Y}. \tag{62}
$$

Notice that the pair of sequences (X^n,Z^n) has the same joint pmf under both hypotheses. Thus, when the relay simply fowards the incoming message M without conveying additional information about its observation Y^n to the receiver, no positive error exponent θ is possible. In contrary, as the following proposition shows, when

$$
T\geq H(Y), \tag{63}
$$

then forwarding message M to the receiver is useless and it suffices that the message B sent from the relay to the receiver describes Y^n. In other words, under constraint (63), the optimal error exponent θ in the present multi-hop setup coincides with the optimal error exponent of a point-to-point system that consists only of the relay and the receiver. The three-terminal multi-hop setup with a transmitter observing X^n can however achieve larger error exponents θ than the point-to-point system when $T<H(Y)$, see Example 2 ahead.

Proposition 4: Assume (61)–(63). Under these assumptions, the exponent-rate region $\mathcal{E}(R,T)$ is the set of all nonnegative pairs (η,θ) that satisfy

$$
\eta \leq I(U;Y), \tag{64}
$$

$$
\theta \leq I(Y;Z), \tag{65}
$$

for some auxiliary random variable U satisfying the Markov chain $U\to X\to(Y,Z)$ and the rate constraint

$$
R\geq I(U;X). \tag{66}
$$

Proof: Achievability follows by specializing Theorem 1 to $S=U$ and $V=Y$. For the converse, see Appendix F. ■

We next consider an example where assumptions (61) and (62) hold, but not necessarily (63).

Example 2: Let under both hypotheses $\mathcal{H}=0$ and $\mathcal{H}=1$:

$$
X\sim\mathcal{U}\{0,1,2\} \quad \text{and} \quad Y\sim\mathcal{B}(1/2)
$$

independent of each other. Also, let $N_0\sim\mathcal{B}(1/3)$ and $N_1\sim\mathcal{B}(1/2)$ be independent of each other and of the pair (X,Y), and

$$
Z=(Z',X) \qquad \text{and} \qquad Z'=\begin{cases} Y\oplus N_0 & \text{if } X=0 \text{ and } \mathcal{H}=0 \\ N_1 & \text{otherwise}\end{cases}.
$$

The described triple (X,Y,Z) satisfies conditions (61) and (62). Moreover, since the pmf of the sequences (X^n,Y^n) is the same under both hypotheses, the error exponent η is zero, and we focus on the receiver's error exponent θ. By Proposition 4, the largest error exponent θ that is achievable is

$$
\theta^\star = I(Y;Z) = I(Y;Z'|X) = 1/3(1-H_\mathrm{b}(1/3)). \tag{67}
$$

As we argue in the following, the same maximum error exponent $\theta^\star$ is achievable with a relay-to-receiver rate of only $T = H_{\text{b}}(1/3)$. To see this, notice that the joint pmf of the triple (X_i, Y_i, Z_i) is the same under both hypotheses, except for positions $i \in \{1, \ldots, n\}$ where $X_i = 0$. It thus suffices that the relay conveys the values of its observations $\{Y_i \colon X_i = 0 \text{ for } i = 1, \ldots, n\}$ to the receiver. This is achieved by specializing the coding and testing scheme of Subsection III-B to the choice of auxiliaries

$$\begin{aligned} S &= 0 \\ U &= \begin{cases} 0 & \text{if } X = 0 \\ 1 & \text{otherwise} \end{cases} \\ V &= \begin{cases} Y & \text{if } U = 0 \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

By Theorem 1, the scheme requires rates

$$R = I(U; X) = H(U) = H_{\text{b}}(1/3)$$

and

$$T = I(V; Y, U) = H(V) = H_{\text{b}}(1/6)$$

and achieves the optimal error exponent $\theta^\star$ in (67):

$$\begin{aligned}
&\min_{\substack{\tilde{P}_{UVXYZ}: \\ \tilde{P}_{UX} = P_{UX} \\ \tilde{P}_{UVY} = P_{UVY} \\ \tilde{P}_{VZ} = P_{VZ}}} D(\tilde{P}_{UVXYZ} \| P_{UX} P_Y P_{V|UY} P_{Z|X}) \\
&\overset{(a)}{=} \mathbb{E}_{P_{UX}} \left[\min_{\substack{\tilde{P}_{UVXYZ}: \\ \tilde{P}_{VY|U} = P_{VY|U} \\ \tilde{P}_{VZ} = P_{VZ}}} D(\tilde{P}_{VYZ|UX} \| P_Y P_{V|UY} P_{Z|X}) \right] \\
&\overset{(b)}{=} P_{UX}(0,0) \cdot \min_{\substack{\tilde{P}_{UVXYZ}: \\ \tilde{P}_{VY|U=0} = P_{VY|U=0} \\ \tilde{P}_{VZ} = P_{VZ}}} D(\tilde{P}_{VYZ|U=X=0} \| P_Y P_{V|Y,U=0} P_{Z|X=0}) \\
&\overset{(c)}{=} 1/3 \cdot \min_{\substack{\tilde{P}_{YZ}: \\ \tilde{P}_{YZ} = P_{YZ}}} D(\tilde{P}_{YZ|U=X=0} \| P_Y P_{Z|X=0}) \\
&\overset{(d)}{=} 1/3 \cdot D(P_{YZ'|X=0} \| P_Y P_{Z'|X=0}) \\
&\overset{(e)}{=} 1/3 \cdot I(Y; Z|X=0) = 1/3(1 - H_{\text{b}}(1/3)),
\end{aligned} \tag{68}$$

where (a) holds because $\tilde{P}_{UX} = P_{UX}$; (b) holds because unless $U = X = 0$, the choice $\tilde{P}_{VYZ|U=X=0} = P_Y P_{V|Y,U=0} P_{Z|X=0}$ is valid and thus the minimum value of the KL-divergence is 0; and (c) holds because $\tilde{P}_{VY|U=0} = P_{VY|U=0}$ implies that $V = Y$ also under $\tilde{P}$ when $U = 0$; (d) holds because $Z = (Z', X)$ and by eliminating U; and (e) holds because $P_Y = P_{Y|X=0}$.

Using similar arguments as in Example 1, it can be shown that the optimal error exponent $\theta^\star$ in (67) cannot be achieved without the transmitter's help when $T < 1$.

IV. Extension to a K-Relay Multi-hop Network

The coding and testing scheme described in the previous section and our analysis extend to related scenarios. In this section, we discuss such an extension to the multi-hop channel with an arbitrary number of $K \geq 2$ relays. The extension to a network with two parallel multi-hop channels that share a common receiver is discussed in the subsequent Section V. Interestingly, the resulting error exponents in Theorems 2 and 3 ahead have the same structure as the error exponents in the previous Theorem 1. Specifically, the exponents are obtained by minimizing KL-divergences between an auxiliary distribution on all involved codewords and observations and the true distribution under $\mathcal{H} = 1$, where the minimization relates the auxiliary distribution to the distribution under $\mathcal{H} = 0$ through the testing conditions. The imposed rate constraints come purely from the applied source codes. To achieve this performance, similarly to the basic single-relay multi-hop network, it is important to limit the number of codewords across codebooks that have the same joint type; otherwise the exponents will be decreased. In fact, the loss in exponent grows with the number of hops in the network, and thus would be more important for the two networks that we treat in this and the following section.

We extend the single-relay setup of the previous section to a setup with an arbitrary number of $K \geq 2$ relays. As before the transmitter observes X^n, the receiver observes Z^n, and Relay j observes Y_j^n, $j = 1, \dots, K$. Depending on the value of the hypothesis $\mathcal{H}$, the tuple $(X^n, Y_1^n, \dots, Y_K^n, Z^n)$ is i.i.d. according to one of the following two distributions:

$$\text{under } \mathcal{H} = 0: \quad (X^n, Y_1^n, \dots, Y_K^n, Z^n) \text{ i.i.d. } \sim P_{XY_1 \cdots Y_K Z}, \tag{69}$$

$$\text{under } \mathcal{H} = 1: \quad (X^n, Y_1^n, \dots, Y_K^n, Z^n) \text{ i.i.d. } \sim Q_{XY_1 \cdots Y_K Z}. \tag{70}$$

Communication takes place over $K+1$ noise-free rate-limited bit-pipes: a rate-R pipe from the transmitter to the first relay; a rate-L_j pipe from Relay j to Relay $j+1$, for $j = 1, \dots, K-1$; and a rate-T pipe from Relay K to the receiver. See Fig. 2 for an illustration.

1) A Simple Coding and Testing Scheme: Fully extending the scheme in the previous subsection to this K-relay scenario requires cumbersome notation. For every pair $(j_1, j_2) \in \{1, \dots, K\}$ with $j_1 < j_2$ one has to generate a codeword that sends information from Relay j_1 to Relay j_2. Moreover, in order not to degrade the performance, some of these codewords have to be superpositioned on others. For simplicity, here we only consider codewords that are sent from one terminal to the following terminal, that means either from the transmitter to the first Relay 1, from Relay $j \in \{1, \dots, K-1\}$ to Relay $j+1$, or from the last Relay K to the receiver. As we will show later in this subsection, this scheme is optimal when testing against conditional independence under some Markov condition.

We now describe the proposed coding and testing scheme for this K-relay scenario. Fix $\mu > 0$, an arbitrary blocklength n, and joint distributions $P_{U_0|X}$ and $P_{U_j|Y_j U_{j-1}}$ for $j \in \{1, ..., K\}$. Define the following nonnegative rates

$$R \triangleq I(U_0; X) + \mu, \tag{71}$$

$$L_j \triangleq I(U_j; Y_j, U_{j-1}) + \mu, \qquad j \in \{1, ..., K-1\}, \tag{72}$$

$$T \triangleq I(U_K; Y_K, U_{K-1}) + \mu. \tag{73}$$

<u>*Codebook Generation*</u>: For $j \in \{1, ..., K-1\}$, randomly generate a codebook $\mathcal{C}_{U_j} = \{U_j^n(\ell_j) \colon \ell_j \in \{1, ..., \lfloor 2^{nL_j} \rfloor\}\}$ by selecting each entry of the n-length codeword $U_j^n(\ell_j)$ i.i.d. according to P_{U_j}. Also, randomly generate codebooks $\mathcal{C}_{U_0} = \{U_0^n(\ell_0) \colon \ell_0 \in \{1, ..., \lfloor 2^{nR} \rfloor\}\}$ and $\mathcal{C}_{U_K} = \{U_K^n(\ell_K) \colon \ell_K \in \{1, ..., \lfloor 2^{nT} \rfloor\}\}$ by selecting each entry of the n-length codewords $U_0^n(\ell_0)$ and $U_K^n(\ell_K)$ i.i.d. according to P_{U_0} and P_{U_K}, respectively.

<u>*Transmitter*</u>: Given that it observes the sequence x^n, the transmitter looks for an index ℓ_0 such that

$$(u_0^n(\ell_0), x^n) \in \mathcal{T}_{\mu/4}^n(P_{U_0 X}). \tag{74}$$

If successful, it picks one of these indices uniformly at random and sends

$$m_0 = \ell_0 \tag{75}$$

over the noise-free bit pipe. Otherwise, it sends $m_0 = 0$.

<u>*Relay $j \in \{1, ..., K\}$*</u>: It observes the sequence y_j^n and receives the message m_{j-1}. If $m_{j-1} = 0$, it declares $\hat{\mathcal{H}}_{y_j} = 1$ and sends $m_j = 0$ over the noise-free bit pipe to the receiver. Otherwise, it looks for an index ℓ_j such that

$$(u_j^n(\ell_j), u_{j-1}^n(\ell_{j-1}), y_j^n) \in \mathcal{T}_{\mu/2}^n(P_{U_j U_{j-1} Y_j}). \tag{76}$$

If such an index exists, Relay j chooses one of these indices at random and sends

$$m_j = \ell_j \tag{77}$$

over the noise-free bit pipe to the receiver. Otherwise, it sends the message $m_j = 0$.

<u>*Receiver*</u>: Assume that the receiver observes z^n and receives message m_K from relay K. If $m_K = 0$, the receiver declares $\hat{\mathcal{H}}_z = 1$. Otherwise, it checks whether

$$(u_K^n(\ell_K), z^n) \in \mathcal{T}_{\mu}^n(P_{U_K Z}). \tag{78}$$

If the typicality check is successful, the receiver declares $\hat{\mathcal{H}}_z = 0$. Otherwise, it declares $\hat{\mathcal{H}}_z = 1$.

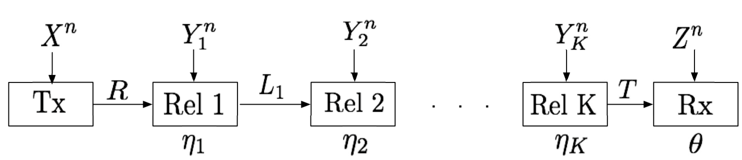

Fig. 2. Hypothesis testing over a multi-hop network with K relays.

Analysis: Similar to the analysis of the scheme in Section III-B, see Appendix B. However, here the code construction is random, whereas in Section III-B a deterministic code construction is used that satisfies the conditions in Lemma 1. One of the main steps in the analysis of this coding and testing scheme is thus to bound the number of codeword tuples with a given joint type. Specifically, one has to show that for each $j \in \{1, \ldots, K\}$, the number of indices (ℓ_j, ℓ_{j-1}) such that $(u_j^n(\ell_j), u_{j-1}^n(\ell_{j-1}))$ have the same joint type $\pi_{U_j U_{j-1}}$ is upper bounded by $2^{n(R_j + R_{j-1} - I_{\pi_{U_j U_{j-1}}}(U_j; U_{j-1}) + \mu)}$. Without this bounding step, the resulting error exponent at the receiver would be reduced by the sum $I(U_1; U_0) + I(U_1; U_2) + \ldots + I(U_K; U_{K-1})$.

2) Achievable Exponent-Rate Region and an Optimality Result: The following achievability result follows from the scheme described in the preceeding subsection.

Theorem 2: The exponent-rate region $\mathcal{E}(R, L_1, ..., L_{K-1}, T)$ includes all nonnegative pairs $(\eta_1, ..., \eta_K, \theta)$ that for all $j \in \{1, ..., K\}$ satisfy

$$\eta_j \leq \min_{\substack{\tilde{P}_{U_0 \ldots U_{j-1} X Y_1 \ldots Y_j}: \\ \tilde{P}_{U_0 X} = P_{U_0 X} \\ \tilde{P}_{U_{j'} U_{j'-1} Y_{j'}} = P_{U_{j'} U_{j'-1} Y_{j'}} \text{ for } j'=1,\ldots,j-1}} D(\tilde{P}_{U_0 \ldots U_{j-1} X Y_1 \ldots Y_j} \| P_{U_0|X} P_{U_1|Y_1 U_0} \ldots P_{U_{j-1}|Y_{j-1} U_{j-2}} Q_{X Y_1 \ldots Y_j}), \tag{79}$$

and

$$\theta \leq \min_{\substack{\tilde{P}_{U_0 \ldots U_{j-1} X Y_1 \ldots Y_K Z}: \\ \tilde{P}_{U_0 X} = P_{U_0 X} \\ \tilde{P}_{U_{j'} U_{j'-1} Y_{j'}} = P_{U_{j'} U_{j'-1} Y_{j'}} \text{ for } j'=1,\ldots,K, \\ \tilde{P}_{U_K Z} = P_{U_K Z}}} D(\tilde{P}_{U_0 \ldots U_{j-1} X Y_1 \ldots Y_K Z} \| P_{U_0|X} P_{U_1|Y_1 U_0} \ldots P_{U_K|Y_K U_{K-1}} Q_{X Y_1 \ldots Y_K Z}), \tag{80}$$

for some auxiliary random variables $(U_0, U_1, ..., U_K)$ satisfying the Markov chains

$$U_0 \to X \to Y_1 \tag{81}$$
$$U_j \to (Y_j, U_{j-1}) \to Y_{j+1}, \qquad j \in \{1, ..., K-1\}, \tag{82}$$
$$U_K \to (U_{K-1}, Y_K) \to Z, \tag{83}$$

and the rate constraints

$$R \geq I(U_0; X), \tag{84}$$
$$L_j \geq I(U_j; Y_j, U_{j-1}), \qquad j \in \{1, ..., K-1\}, \tag{85}$$
$$T \geq I(U_K; Y_K, U_{K-1}). \tag{86}$$

Proof: Based on the scheme described in the previous subsection and similar to the proof of Theorem 1. Details are omitted. ■

The next proposition shows that the exponent-rate region of Theorem 2 is optimal when the pmfs $P_{XY_1 \ldots Y_K Z}$ and $Q_{XY_1 \ldots Y_K Z}$ decompose as

$$P_{XY_1 \ldots Y_K Z} = P_X \cdot P_{Y_1|X} \cdot P_{Y_2|Y_1} \cdot \ldots \cdot P_{Y_K|Y_{K-1}} \cdot P_{Z|Y_K}, \tag{87}$$
$$Q_{XY_1 \ldots Y_K Z} = P_X \cdot P_{Y_1} \cdot P_{Y_2} \ldots \cdot P_{Y_K} \cdot P_Z. \tag{88}$$

Proposition 5: If (87) and (88) hold, then the exponent-rate region $\mathcal{E}_K(R, L_1, ..., L_K, T)$ is the set of all nonnegative tuples $(\eta_1, ..., \eta_K, \theta)$ that satisfy

$$\eta_j \leq \sum_{m=1}^{j} I(U_{m-1}; Y_m), \qquad j \in \{1, ..., K\}, \tag{89}$$
$$\theta \leq \sum_{m=1}^{K} I(U_{m-1}; Y_m) + I(U_K; Z), \tag{90}$$

for some auxiliary random variables $(U_0, U_1, ..., U_K)$ satisfying the Markov chains

$$U_0 \to X \to Y_1 \tag{91}$$
$$U_j \to Y_j \to Y_{j+1}, \qquad j \in \{1, ..., K-1\}, \tag{92}$$
$$U_K \to Y_K \to Z \tag{93}$$

and the rate constraints

$$R \geq I(U_0; X), \tag{94}$$
$$L_j \geq I(U_j; Y_j), \qquad j = 1, ..., K-1, \tag{95}$$
$$T \geq I(U_K; Y_K). \tag{96}$$

Proof: Achievability follows by specializing Theorem 2 to auxiliaries that satisfy the Markov chains (91)–(93). The converse is similar to the converse proof of Proposition 2 in Appendix D and omitted for brevity. ■

V. Extension to a K-Relay Multi-hop Network

Consider the two-transmitter, two-relay setup in Fig. 3. Transmitters 1 and 2 observe sequences $X_1^n \triangleq (X_{1,1}, \ldots, X_{1,n})$ and $X_2^n \triangleq (X_{2,1}, \ldots, X_{2,n})$, Relays 1 and 2 observe sequences $Y_1^n \triangleq (Y_{1,1}, \ldots, Y_{1,n})$ and $Y_2^n \triangleq (Y_{2,1}, \ldots, Y_{2,n})$, and the common receiver observes $Z^n \triangleq (Z_1, \ldots, Z_n)$. Depending on the value of the hypothesis $\mathcal{H}$, the tuple $(X_1^n, X_2^n, Y_1^n, Y_2^n, Z^n)$ is i.i.d. according to one of the following two distributions:

$$\text{under } \mathcal{H} = 0: \quad (X_1^n, X_2^n, Y_1^n, Y_2^n, Z^n) \sim \text{ i.i.d. } P_{X_1 X_2 Y_1 Y_2 Z}, \tag{97}$$
$$\text{under } \mathcal{H} = 1: \quad (X_1^n, X_2^n, Y_1^n, Y_2^n, Z^n) \sim \text{ i.i.d. } Q_{X_1 X_2 Y_1 Y_2 Z}. \tag{98}$$

Each Transmitter $i \in \{1, 2\}$ can communicate with its corresponding Relay i over a noise-free bit pipe of rate R_i, and each Relay $i \in \{1, 2\}$ in its turn can communicate with the final receiver over an individual noise-free bit pipe of rate T_i. Encoding and detection functions, and probabilities of type-I and type-II errors are defined in a similar way as for the single multi-hop channel in Section II. Let then η_1 and η_2 denote the type-II error exponents at Relays 1 and 2, and θ the type-II error exponent at the receiver. As in Section II, the interest is in identifying the set of all exponent triples (η_1, η_2, θ) so that the type-I error probabilities vanish as the observation length tends to infinity. Define thus the exponent-rate region $\mathcal{E}_{\text{par-2}}(R_1, R_2, T_1, T_2)$ for the setup here analogously to Definition 2.

1) Coding and Testing Scheme: For this setup with only 5 terminals, we extend the scheme in Subsection III-B in its full generality.

Fix $\mu > 0$, an arbitrary blocklength n, and a joint conditional distribution $P_{S_l, U_l | X_l}$ and $P_{V_l | S_l, U_l, Y_l}$ for $l = 1, 2$. Define the following nonnegative rates:

$$R_{u_l} \triangleq I(U_l; X_l | S_l) + \mu, \tag{99}$$
$$R_{s_l} \triangleq I(X_l; S_l) + \mu, \tag{100}$$
$$R_{v_l} \triangleq I(V_l; Y_l, U_l | S_l) + \mu. \tag{101}$$

Also, define $R_l \triangleq R_{u_l} + R_{s_l}$ and $T_l \triangleq R_{s_l} + R_{v_l}$.

Codebook Generation: For $l \in \{1, 2\}$, randomly generate a codebook $\mathcal{C}_S \triangleq \{S_l^n(i_l), i_l \in \{1, ..., \lfloor 2^{nR_{s_l}} \rfloor\}\}$ by selecting each entry of the n-length codeword $S_l^n(i_l)$ i.i.d. according to P_{S_l}. For each index i_l, randomly generate a U_l-codebook $\{U_l^n(j_l | i_l) : j_l \in \{1, ..., \lfloor 2^{nR_{u_l}} \rfloor\}\}$ by selecting the t-th entry of the n-length codeword $U_l^n(j_l | i_l)$ according to $P_{U_l | S_l}(. | S_{l,t}(i_l))$, $S_{l,t}(i_l)$ denotes the t-entry of $S_l^n(i_l)$. For each index i_l, randomly generate a V_l-codebook $\{V_l^n(k_l | i_l),\, k_l \in \{1, ..., \lfloor 2^{nR_{v_l}} \rfloor\}\}$. The codebooks are revealed to all terminals.

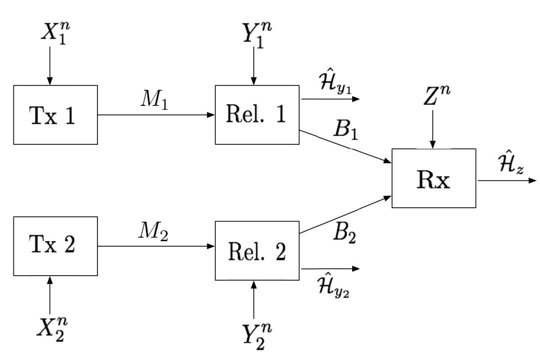

Fig. 3. Hypothesis testing over a parallel multi-hop network.

<u>*Transmitter* $l \in \{1,2\}$:</u> Given that it observes the sequence $X^n = x^n$, Transmitter l looks for a pair of indices (i_l, j_l) such that

$$(x_l^n, s_l^n(i_l), u^n(j_l|i_l)) \in \mathcal{T}_\mu^n(P_{X_l S_l U_l}). \tag{102}$$

If successful, the transmitter picks one of these pairs uniformly at random and sends the corresponding index pair $M_l \triangleq (I_l, J_l)$ to Relay l. If no such pair exists, it sends $M_l = 0$ to Relay l.

<u>*Relay* $l \in \{1,2\}$:</u> Assume that it observes the sequence $Y_l^n = y_l^n$ and receives message M_l from Transmitter l. If $m_l = 0$, it declares $\hat{\mathcal{H}}_{y_l} = 1$ and sends $B_l = 0$ to the receiver. Otherwise, it parses $m_l = (i_l, j_l)$ and looks for an k_l such that

$$(s_l^n(i_l), u_l^n(j_l|i_l), v_l^n(k_l|i_l), y_l^n) \in \mathcal{T}_\mu^n(P_{S_l U_l V_l Y_l}). \tag{103}$$

If successful the relay picks one of these indices uniformly at random and sends $B_l \triangleq (I_l, K_l)$ to the receiver. Otherwise, it declares $\hat{\mathcal{H}}_{y_l} = 1$ and sends $B_l = 0$ to the receiver.

<u>*Receiver*:</u> Assume that the receiver observes $Z^n = z^n$ and Message $B_1 = b_1$ from Relay 1 and Message $B_2 = b_2$ from Relay 2. If $b_1 = 0$ or $b_2 = 0$, the receiver declares $\hat{\mathcal{H}}_z = 1$. Otherwise, it parses $b_1 = (i_1, k_1)$ and $b_2 = (i_2, k_2)$ and checks whether

$$(s_1^n(i_1), s_2^n(i_2), v_1^n(k_1|i_1), v_2^n(k_2|i_2), z^n) \in \mathcal{T}_\mu^n(P_{S_1 S_2 V_1 V_2 Z}).$$

If the check is successful, it declares $\hat{\mathcal{H}}_z = 0$; otherwise, it declares $\hat{\mathcal{H}}_z = 1$.

<u>*Analysis*</u>: As for the K-relay scenario, here the code construction is random, and thus it is crucial to bound the number of codeword tuples that are of a given joint type. In particular, it is important to show that, the number of indices $(i_1, j_1, k_1, i_2, j_2, k_2)$ for which $(s_1^n(i_1), u_1^n(j_1|i_1), v_1^n(k_1|i_1), s_2^n(i_2), u_2^n(j_2|i_2), v_2^n(k_2|i_2))$ have the same joint type $\pi_{S_1 U_1 V_1 S_2 U_2 V_2}$ is upper bounded by

$$2^{n(R_{s_1}+R_{u_1}+R_{v_1}+R_{s_2}+R_{u_2}+R_{v_2}-I_{\pi_{S_1U_1V_1}}(U_1;V_1|S_1)-I_{\pi_{S_2U_2V_2}}(U_2;V_2|S_2)-I_{\pi_{S_1U_1V_1S_2U_2V_2}}(S_1,U_1,V_1;S_2,U_2,V_2)+\mu)}.$$

Without this bounding step, the error exponent at the receiver will be decreased by the sum $I(U_1;V_1|S_1) + I(U_2;V_2|S_2) + I(S_1,U_1,V_1;S_2,U_2,V_2)$. The rest of the analysi is similar to the analysis of the scheme in Subsection III-B.

2) Achievable Exponent-Rate Region and an Optimality Result: The following achievability result follows from the scheme described in the preceeding section.

Theorem 3: The exponent-rate region $\mathcal{E}_{\text{par-2}}(R_1, R_2, T_1, T_2)$ includes all nonnegative triples (η_1, η_2, θ) that satisfy for $l \in \{1,2\}$:

$$\eta_l \leq \min_{\substack{\tilde{P}_{S_l U_l X_l Y_l} \\ \tilde{P}_{S_l U_l X_l} = P_{S_l U_l X_l} \\ \tilde{P}_{S_l U_l Y_l} = P_{S_l U_l Y_l}}} D(\tilde{P}_{S_l U_l X_l Y_l} \| P_{S_l U_l | X_l} Q_{X_l Y_l}), \tag{104}$$

$$\theta \leq \min_{\substack{\tilde{P}_{S_1 U_1 V_1 S_2 U_2 V_2 X_1 Y_1 X_2 Y_2 Z}: \\ \tilde{P}_{S_l U_l X_l} = P_{S_l U_l X_l} \\ \tilde{P}_{S_l V_l U_l Y_l} = P_{S_l V_l U_l Y_l} \\ \tilde{P}_{S_1 S_2 V_1 V_2 Z} = P_{S_1 S_2 V_1 V_2 Z}}} D(\tilde{P}_{S_1 U_1 V_1 S_2 U_2 V_2 X_1 Y_1 X_2 Y_2 Z} \| P_{S_1 U_1|X_1} P_{S_2 U_2|X_2} P_{V_1|S_1 U_1 Y_1} P_{V_2|S_2 U_2 Y_2} Q_{X_1 X_2 Y_1 Y_2 Z}), \tag{105}$$

for some auxiliary random variables $(S_1, S_2, U_1, U_2, V_1, V_2)$ satisfying the rate constraints

$$R_l \geq I(U_l, S_l; X_l), \tag{106}$$

$$T_l \geq I(X_l; S_l) + I(V_l; Y_l, U_l | S_l), \tag{107}$$

and so that their joint pmf with (X_1, X_2, Y_1, Y_2, Z) factorizes as

$$P_{S_1 S_2 U_1 U_2 V_1 V_2 X_1 X_2 Y_1 Y_2 Z} = P_{X_1 X_2 Y_1 Y_2 Z} P_{S_1 U_1|X_1} P_{S_2 U_2|X_2} P_{V_1|S_1 U_1 Y_1} P_{V_2|S_2 U_2 Y_2}.$$

Proof: Follows similar lines as the proof of Theorem 1 in Appendix B. Details omitted. ■

The achievable exponent-rate region of Theorem 3 is optimal when the pmfs $P_{X_1 X_2 Y_1 Y_2 Z}$ and $Q_{X_1 X_2 Y_1 Y_2 Z}$ decompose as

$$P_{X_1 X_2 Y_1 Y_2 Z} = P_{X_1|Y_1} \cdot P_{X_2|Y_2} \cdot P_{Y_1 Y_2} \cdot P_{Z|Y_1 Y_2}, \tag{108}$$

$$Q_{X_1 X_2 Y_1 Y_2 Z} = P_{X_1} \cdot P_{X_2} \cdot P_{Y_1 Y_2} \cdot P_Z. \tag{109}$$

Proposition 6: If (108)–(109) hold, then the exponent-rate region $\mathcal{E}_{\text{par-2}}(R_1, R_2, T_1, T_2)$ is the set of all nonnegative triples (η_1, η_2, θ) that satisfy

$$\eta_l \leq I(U_l; Y_l), \qquad l \in \{1,2\}, \tag{110}$$

$$\theta \leq I(U_1; Y_1) + I(U_2; Y_2) + I(V_1, V_2; Z), \tag{111}$$

for some auxiliary random variables (U_1, U_2, V_1, V_2) satisfying the rate constraints

$$R_l \geq I(U_l; X_l), \tag{112}$$

$$T_l \geq I(V_l; Y_l), \tag{113}$$

and so that their joint pmf with (X_1, X_2, Y_1, Y_2, Z) factorizes as

$$P_{U_1U_2V_1V_2X_1X_2Y_1Y_2Z} = P_{X_1X_2Y_1Y_2Z}P_{U_1|X_1}P_{U_2|X_2}P_{V_1|Y_1}P_{V_2|Y_2}.$$

Proof: Achievability follows by specializing Theorem 3 to $S_1 = S_2 = 0$ and to V_1 and V_2 satisfying the Markov chains $V_l \to Y_l \to U_l$, for $l \in \{1, 2\}$ and is given in Appendix G. The converse is also proved in Appendix G. ■

VI. Coding and Testing Schemes with Binning

In source coding, it is well known that binning can decrease the required rate of communication when observations at different terminals are correlated. This is also the case here, provided that the observations at different terminals are correlated under both hypotheses. In this section, we extend our coding scheme for the basic single-relay multi-hop network to include binning. Before doing so, for completeness, we provide a detailed proof of the Shimokawa, Han, and Amari error exponent [3] for hypothesis testing in a point-to-point system. In their conference paper [3], they describe a coding and testing scheme based on binning, however they do not include a detailed proof of the resulting error exponent. Recently Katz, Piantanida, and Debbah [21] analyzed a similar binning scheme; however, their obtained exponent-rate function is smaller than the one in [3].

A. Point-to-Point System

1) Problem Setup: Consider the point-to-point HT with two terminals in Fig. 4. The transmitter observes the sequence $X^n \triangleq (X_1, ..., X_n)$ and the receiver observes the sequence $Y^n \triangleq (Y_1, ..., Y_n)$. Under the null hypothesis

$$\mathcal{H} = 0: \quad (X^n, Y^n) \sim \text{ i.i.d } P_{XY}, \tag{114}$$

whereas under the alternative hypothesis

$$\mathcal{H} = 1: \quad (X^n, Y^n) \sim \text{ i.i.d. } Q_{XY}. \tag{115}$$

for two given pmfs P_{XY} and Q_{XY}. The transmitter computes a message $M = \phi^{(n)}(X^n)$ and sends it to the receiver over a noise-free pipe of rate R. Based on M and its observation Y^n, the receiver produces its guess $\hat{\mathcal{H}} = g^{(n)}(Y^n, M)$.

Definition 3: For each $\epsilon \in (0, 1)$, we say that the exponent-rate pair (θ, R) is ϵ-achievable if for each sufficiently large blocklength n, there exists a sequence of encoding and decoding functions $(\phi^{(n)}, g^{(n)})$ such that the corresponding type-I and type-II error probabilities at the receiver

$$\alpha_n \triangleq \Pr[\hat{\mathcal{H}} = 1 | \mathcal{H} = 0], \tag{116}$$

$$\beta_n \triangleq \Pr[\hat{\mathcal{H}} = 0 | \mathcal{H} = 1], \tag{117}$$

satisfy

$$\alpha_n \leq \epsilon, \tag{118}$$

and

$$- \overline{\lim_{n \to \infty}} \frac{1}{n} \log \beta_n \geq \theta. \tag{119}$$

Definition 4: The exponent-rate function $\theta^*(R)$ is the supremum of all ϵ-achievable error exponents for given rate R, i.e.,

$$\theta^*(R) \triangleq \sup\{\theta \geq 0 \colon (\theta, R) \text{ is } \epsilon\text{-achievable } \forall \epsilon > 0\}. \tag{120}$$

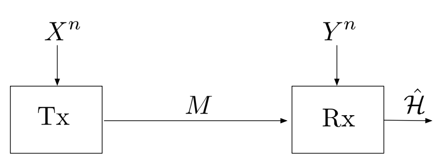

Fig. 4. Hypothesis testing over a point-to-point channel

2) Coding and Testing Scheme: Fix $\mu > 0$, an sufficiently large blocklength n, and the conditional pmf $P_{S|X}$ over a finite auxiliary alphabet $\mathcal{S}$. Define joint pmfs $P_{SXY} = P_{XY}P_{S|X}$, $Q_{SXY} = Q_{XY}P_{S|X}$ and the nonnegative rate R' such that

$$R + R' = I(X;S) + \mu, \tag{121}$$

$$R' < I(Y;S). \tag{122}$$

Code Construction: Construct a random codebook

$$\mathcal{C}_S = \{s^n(m,\ell)\colon m \in \{1,...,\lfloor 2^{nR}\rfloor\}, \ell \in \{1,...,\lfloor 2^{nR'}\rfloor\}\}, \tag{123}$$

by drawing all entries of all codewords i.i.d. according to the chosen distribution P_S.

Transmitter: Given that it observes the sequence x^n, the transmitter looks for indices (m,ℓ) such that

$$(s^n(m,\ell), x^n) \in \mathcal{T}^n_{\mu/2}(P_{SX}). \tag{124}$$

If successful, it picks one of these indices uniformly at random and sends the index m over the noise-free bit pipe. Otherwise, it sends $m = 0$.

Receiver: Assume that the receiver observes y^n and receives the message m from the transmitter. If $m = 0$, the receiver declares $\hat{\mathcal{H}} = 1$. Otherwise, it first looks for an index $\ell' \in \{1,...,\lfloor 2^{nR'}\rfloor\}$ such that

$$H_{\text{tp}(s^n(m,\ell'),y^n)}(S|Y) \leq H_{\text{tp}(s^n(m,\tilde{\ell}),y^n)}(S|Y), \qquad \forall \tilde{\ell} \in \{1,\ldots,\lfloor 2^{nR}\rfloor\}. \tag{125}$$

If no such index exists, the receiver declares $\hat{\mathcal{H}} = 1$. Otherwise, it checks whether

$$(s^n(m,\ell'), y^n) \in \mathcal{T}^n_{\mu}(P_{SY}),$$

and declares $\hat{\mathcal{H}} = 0$ if this typicality check is successful and $\hat{\mathcal{H}} = 1$ otherwise.

3) Result on the Error Exponent: The scheme described in the previous subsection yields the following set of achievable exponent-rate pairs.

Theorem 4 (Theorem 1 in [3]): Given a rate $R > 0$, for any choice of the auxiliary random variable S that satisfies the Markov chain $S \to X \to Y$ and the rate constraint

$$R \geq I(S;X|Y),$$

the exponent-rate function $\theta^*(R)$ is lower bounded as

$$\theta^*(R) \geq \min\left\{ \min_{\substack{\tilde{P}_{SXY}:\\ \tilde{P}_{SX}=P_{SX}\\ \tilde{P}_{SY}=P_{SY}}} D(\tilde{P}_{SXY}\|P_{S|X}Q_{XY}), \quad \min_{\substack{\tilde{P}_{SXY}:\\ \tilde{P}_{SX}=P_{SX}\\ \tilde{P}_{Y}=P_{Y}\\ H(S|Y)\leq H_{\tilde{P}_{SY}}(S|Y)}} D(\tilde{P}_{SXY}\|P_{S|X}Q_{XY}) + R - I(S;X|Y) \right\}, \tag{126}$$

Proof: See Appendix H. ■

The inequality in Theorem 4 holds with equality in the special cases of *testing against independence* [1], where $Q_{XY} = P_X \cdot P_Y$[1], and of *testing against conditional independence* [4] where Y decomposes as $Y = (Y', Z)$ and $Q_{XY'Z} = P_{XZ}P_{Y'|Z}$.

B. Single-Relay Multi-Hop Network

We turn back to the basic single-realy multi-hop scenario and propose an improved coding and testing scheme based on binnning.

[1]There is no need to apply the coding scheme with binning to attain the optimal error exponent in this case, see [1].

1) Coding and Testing Scheme: Fix $\mu > 0$, an arbitrary blocklength n, and joint conditional pmfs $P_{SU|X}$ and $P_{V|SUY}$ over finite auxiliary alphabets $\mathcal{S}, \mathcal{U}$, and $\mathcal{V}$. Define the joint pmf $P_{SUVXYZ} = P_{XYZ}P_{SU|X}P_{V|SUY}$ and the following nonnegative rates, which are calculated according to the chosen distribution,

$$R_s + R'_s = I(X;S) + \mu, \tag{127a}$$
$$R_u + R'_u = I(U;X|S) + \mu, \tag{127b}$$
$$R_v + R'_v = I(V;Y,U|S) + \mu, \tag{127c}$$
$$R'_s \le \min\{I(S;Y), I(S;Z)\}, \tag{127d}$$
$$R'_u \le I(U;Y|S), \tag{127e}$$
$$R'_v \le I(V;Z|S). \tag{127f}$$

The joint distributions are chosen in a way that

$$R = R_u + R_s \tag{128}$$
$$T = R_s + R_v. \tag{129}$$

Code Construction: Construct a random codebook

$$\mathcal{C}_S = \{S^n(i,d) \colon \quad i \in \{1, ..., \lfloor 2^{nR_s} \rfloor\}, d \in \{1, ..., \lfloor 2^{nR'_s} \rfloor\}\}$$

by selecting each entry of the n-length codeword $S^n(i,d)$ in an i.i.d. manner according to the pmf P_S. Then, for each pair (i,d), generate random codebooks

$$\mathcal{C}_U(i,d) = \{U^n(j,e|i,d) \colon \quad j \in \{1, ..., \lfloor 2^{nR_u} \rfloor\}, \ e \in \{1, ..., \lfloor 2^{nR'_u} \rfloor\}\}$$

and

$$\mathcal{C}_V(i,d) = \{V^n(k,f|i,d) \colon \quad k \in \{1, ..., \lfloor 2^{nR_v} \rfloor\}, \ f \in \{1, ..., \lfloor 2^{nR'_v} \rfloor\}\}$$

by selecting each entry of the n-length codewords $U^n(j,e|i,d)$ and $V^n(k,f|i,d)$ in a memoryless manner using the conditional pmfs $P_{U|S}(\cdot|S_t(i,d))$ and $P_{V|S}(\cdot|S_t(i,d))$, where $S_t(i,d)$ denotes the t-th component of the random codeword $S^n(i,d)$. All codebooks are revealed to all parties.

Transmitter: Given that the transmitter observes the sequence x^n, it looks for indices (i,d,j,e) such that

$$(s^n(i,d), u^n(j,e|i,d), x^n) \in \mathcal{T}^n_{\mu/4}(P_{SUX}). \tag{130}$$

If successful, it picks one such pair uniformly at random, and sends

$$m = (i,j), \tag{131}$$

over the noise-free bit pipe. Otherwise, it sends $m = 0$.

Relay: Assume that the relay observes the sequence y^n and receives the message m. If $m = 0$, it declares $\hat{\mathcal{H}}_y = 1$ and sends $b = 0$ over the noise-free bit pipe to the receiver. Otherwise, it first looks for indices (d', e') such that

$$H_{\text{tp}(s^n(i,d'),u^n(j,e'|i,d'),y^n)}(S,U|Y) \le H_{\text{tp}(s^n(i,\tilde{d}),u^n(j,\tilde{e}|i,\tilde{d}),y^n)}(S,U|Y), \quad \forall \tilde{d} \in \{1, ..., \lfloor 2^{nR'_s} \rfloor\}, \ \tilde{e} \in \{1, ..., \lfloor 2^{nR'_u} \rfloor\}, \tag{132}$$

and then looks for indices (k,f) such that

$$(s^n(i,d'), u^n(j,e'|i,d'), v^n(k,f|i,d'), y^n) \in \mathcal{T}^n_{\mu/2}(P_{SUVY}). \tag{133}$$

If successful, it declares $\hat{\mathcal{H}}_y = 0$ and picks one of these index pairs uniformly at random. It then sends the corresponding indices

$$b = (i,k) \tag{134}$$

to the receiver. Otherwise, it declares $\hat{\mathcal{H}}_y = 1$ and sends the message $b = 0$ to the receiver.

Receiver: Assume that the receiver observes z^n and receives message b from the relay. If $b = 0$, the receiver declares $\hat{\mathcal{H}}_z = 1$. Otherwise, it first looks for indices (d'', f') such that

$$H_{\text{tp}(s^n(i,d''),v^n(k,f'|i,d''),z^n)}(S,V|Z) \le H_{\text{tp}(s^n(i,\tilde{d}),v^n(k,f|i,\tilde{d}),z^n)}(S,V|Z), \quad \forall \tilde{d} \in \{1, ..., \lfloor 2^{nR'_s} \rfloor\}, \ \tilde{f} \in \{1, ..., \lfloor 2^{nR'_v} \rfloor\}. \tag{135}$$

Then, it checks whether

$$(s^n(i,d''), v^n(k,f'|i,d''), z^n) \in \mathcal{T}^n_{\mu}(P_{SVZ}). \tag{136}$$

If successful, the receiver declares $\hat{\mathcal{H}}_z = 0$. Otherwise, it declares $\hat{\mathcal{H}}_z = 1$.

2) Result on the Exponent-Rate Region: The coding scheme in the previous subsection establishes the following achievability result.

Theorem 5: The exponent-rate region $\mathcal{E}(R,T)$ contains all nonnegative pairs (η,θ) that satisfy

$$\eta \leq \min\{\eta_1,\eta_2,\eta_3\}, \tag{137}$$

$$\theta \leq \min\{\theta_i\colon i=1,...,10\}, \tag{138}$$

where

$$\eta_1 = \min_{\substack{\tilde{P}_{SUXY}:\\ \tilde{P}_{SUX}=P_{SUX}\\ \tilde{P}_{SUY}=P_{SUY}}} D(\tilde{P}_{SUXY}\|P_{SU|X}Q_{XY}), \tag{139a}$$

$$\eta_2 = \min_{\substack{\tilde{P}_{SUXY}:\\ \tilde{P}_{SUX}=P_{SUX}\\ \tilde{P}_{Y}=P_{Y}\\ H(S,U|Y)\leq H_{\tilde{P}}(S,U|Y)\\ H(U|S,Y)\leq H_{\tilde{P}}(U|S,Y)}} D(\tilde{P}_{SUXY}\|P_{SU|X}Q_{XY})+R-I(S,U;X)+I(S,U;Y), \tag{139b}$$

$$\eta_3 = \min_{\substack{\tilde{P}_{SUXY}:\\ \tilde{P}_{SUX}=P_{SUX}\\ \tilde{P}_{SY}=P_{SY}\\ H(U|S,Y)\leq H_{\tilde{P}}(U|S,Y)}} D(\tilde{P}_{SUXY}\|P_{SU|X}Q_{XY})+R_u-I(U;X|S)+I(U;Y|S) \tag{139c}$$

$$\theta_1 = \min_{\substack{\tilde{P}_{SUVXYZ}:\\ \tilde{P}_{SUX}=P_{SUX}\\ \tilde{P}_{SUVY}=P_{SUVY}\\ \tilde{P}_{SVZ}=P_{SVZ}}} D(\tilde{P}_{SUVXYZ}\|P_{SU|X}P_{V|SUY}Q_{XYZ}) \tag{139d}$$

$$\theta_2 = \min_{\substack{\tilde{P}_{SUVXYZ}:\\ \tilde{P}_{SUX}=P_{SUX}\\ \tilde{P}_{SUVY}=P_{SUVY}\\ \tilde{P}_{SZ}=P_{SZ}\\ H(V|S,Z)\leq H_{\tilde{P}}(V|S,Z)}} D(\tilde{P}_{SUVXYZ}\|P_{SU|X}P_{V|SUY}Q_{XYZ})+I(V;Z|S)+R_v-I(V;Y,U|S) \tag{139e}$$

$$\begin{aligned}\theta_3 = \min_{\substack{\tilde{P}_{SUVXYZ}:\\ \tilde{P}_{SUX}=P_{SUX}\\ \tilde{P}_{SVY}=P_{SVY}\\ \tilde{P}_{SZ}=P_{SZ}\\ H(U|S,Y)\leq H_{\tilde{P}}(U|S,Y)\\ H(V|S,Z)\leq H_{\tilde{P}}(V|S,Z)}} & D(\tilde{P}_{SUVXYZ}\|P_{SU|X}P_{V|SUY}Q_{XYZ})+I(V,Y;U|S)+I(V;Z|S)+R_u+R_v\\ &\quad -I(U;X|S)-I(V;Y,U|S)\end{aligned} \tag{139f}$$

$$\theta_4 = \min_{\substack{\tilde{P}_{SUVXYZ}:\\ \tilde{P}_{SUX}=P_{SUX}\\ \tilde{P}_{SVY}=P_{SVY}\\ \tilde{P}_{SVZ}=P_{SVZ}\\ H(U|S,Y)\leq H_{\tilde{P}}(U|S,Y)}} D(\tilde{P}_{SUVXYZ}\|P_{SU|X}P_{V|SUY}Q_{XYZ})+I(U;Y,V|S)+R_u-I(U;X|S) \tag{139g}$$

$$\begin{aligned}\theta_5 = \min_{\substack{\tilde{P}_{SUVXYZ}:\\ \tilde{P}_{SUX}=P_{SUX}\\ \tilde{P}_{VY}=P_{VY}\\ \tilde{P}_{Z}=P_{Z}\\ H(U|S,Y)\leq H_{\tilde{P}}(U|S,Y)\\ H(U,S|Y)\leq H_{\tilde{P}}(U,S|Y)\\ H(V|S,Z)\leq H_{\tilde{P}}(V|S,Z)\\ H(V,S|Z)\leq H_{\tilde{P}}(V,S|Z)}} & D(\tilde{P}_{SUVXYZ}\|P_{SU|X}P_{V|SUY}Q_{XYZ})+I(S,V;Z)+I(S,U;V,Y)\\ &\quad +2R_s+R_u+R_v-I(S;X)-I(U,S;X)-I(V;Y,U|S)\end{aligned} \tag{139h}$$

$$\begin{aligned}\theta_6 = \min_{\substack{\tilde{P}_{SUVXYZ}:\\ \tilde{P}_{SUX}=P_{SUX}\\ \tilde{P}_{VY}=P_{VY}\\ \tilde{P}_{Z}=P_{Z}\\ H(U|S,Y)\leq H_{\tilde{P}}(U|S,Y)\\ H(U,S|Y)\leq H_{\tilde{P}}(U,S|Y)\\ H(V|S,Z)\leq H_{\tilde{P}}(V|S,Z)\\ H(V,S|Z)\leq H_{\tilde{P}}(V,S|Z)}} & D(\tilde{P}_{SUVXYZ}\|P_{SU|X}P_{V|SUY}Q_{XYZ})+I(S,U;Y,V)+I(V;Z|S)\\ &\quad +R_s+R_u+R_v-I(U,S;X)-I(V;Y,U|S)\end{aligned} \tag{139i}$$

$$\theta_7 = \min_{\substack{\tilde{P}_{SUVXYZ}:\\ \tilde{P}_{SUX}=P_{SUX}\\ \tilde{P}_{VY}=P_{VY}\\ \tilde{P}_{SVZ}=P_{SVZ}\\ H(U|S,Y)\leq H_{\tilde{P}}(U|S,Y)\\ H(U,S|Y)\leq H_{\tilde{P}}(U,S|Y)}} D(\tilde{P}_{SUVXYZ}\|P_{SU|X}P_{V|SUY}Q_{XYZ}) + I(U,S;V,Y) + R_s + R_u - I(S,U;X) \tag{139j}$$

$$\theta_8 = \min_{\substack{\tilde{P}_{SUVXYZ}:\\ \tilde{P}_{SUX}=P_{SUX}\\ \tilde{P}_{VY}=P_{VY}\\ \tilde{P}_{SZ}=P_{SZ}\\ H(U|S,Y)\leq H_{\tilde{P}}(U|S,Y)\\ H(U,S|Y)\leq H_{\tilde{P}}(U,S|Y)\\ H(V|S,Z)\leq H_{\tilde{P}}(V|S,Z)}} D(\tilde{P}_{SUVXYZ}\|P_{SU|X}P_{V|SUY}Q_{XYZ}) + I(V;Z|S) + I(S,U;Y|V) + R_s + R_u + R_v - I(S,U;X) - I(V;Y,U|S) \tag{139k}$$

$$\theta_9 = \min_{\substack{\tilde{P}_{SUVXYZ}:\\ \tilde{P}_{SUX}=P_{SUX}\\ \tilde{P}_{SUVY}=P_{SUVY}\\ \tilde{P}_{Z}=P_{Z}\\ H(V|S,Z)\leq H_{\tilde{P}}(V|S,Z)\\ H(V,S|Z)\leq H_{\tilde{P}}(V,S|Z)}} D(\tilde{P}_{SUVXYZ}\|P_{SU|X}P_{V|SUY}Q_{XYZ}) + I(V,S;Z) + R_s + R_v - I(S;X) - I(V;Y,U|S) \tag{139l}$$

$$\theta_{10} = \min_{\substack{\tilde{P}_{SUVXYZ}:\\ \tilde{P}_{SUX}=P_{SUX}\\ \tilde{P}_{SVY}=P_{SVY}\\ \tilde{P}_{Z}=P_{Z}\\ H(U,S|Y)\leq H_{\tilde{P}}(U,S|Y)\\ H(V|S,Z)\leq H_{\tilde{P}}(V|S,Z)\\ H(V,S|Z)\leq H_{\tilde{P}}(V,S|Z)}} D(\tilde{P}_{SUVXYZ}\|P_{SU|X}P_{V|SUY}Q_{XYZ}) + I(V,S;Z) + I(U;Y,V|S) + R_s + R_u + R_v - I(S,U;X) - I(V;Y,U|S). \tag{139m}$$

for some choice of the auxiliary random variables (U,S,V) and auxiliary rates R_s, R_u, R_v satisfying the Markov chains

$$(U,S) \to X \to (Y,Z) \tag{140}$$

$$V \to (S,Y,U) \to Z \tag{141}$$

and the rate constraints $R = R_s + R_u$ and $T = R_s + R_v$ and

$$R \geq I(U,S;X), \tag{142}$$

$$T \geq I(S;X) + I(V;Y,U|S), \tag{143}$$

$$R_s \geq I(S;X) - \min\{I(S;Y), I(S;Z)\}, \tag{144}$$

$$R_u \geq I(U;X|S) - I(U;Y|S), \tag{145}$$

$$R_v \geq I(V;Y,U|S) - I(V;Z|S). \tag{146}$$

Proof: See Appendix I. ■

The exponent-rate region achieved with the binning scheme is optimal when testing against conditional independence. In fact, consider a setup where the relay's and the receiver's observations decompose as

$$Y = (\bar{Y}, W) \tag{147}$$

$$Z = (\bar{Z}, W, Q) \tag{148}$$

and

$$\text{under } \mathcal{H}=0: \qquad (X^n, \bar{Y}^n, \bar{Z}^n, W^n, Q^n) \text{ i.i.d. } \sim P_{X|\bar{Y}W} \cdot P_{\bar{Y}\bar{Z}QW}, \tag{149}$$

$$\text{under } \mathcal{H}=1: \qquad (X^n, \bar{Y}^n, \bar{Z}^n, W^n, Q^n) \text{ i.i.d. } \sim P_{X|W} \cdot P_{\bar{Y}|WQ} \cdot P_{\bar{Z}WQ}. \tag{150}$$

Proposition 7: If (147)–(150) hold, the exponent-rate region $\mathcal{E}(R,T)$ is the set of all nonnegative pairs (η,θ) that satisfy

$$\eta \leq I(U;Y|W), \tag{151}$$

$$\theta \leq I(U;Y|W) + I(V;Z|Q,W), \tag{152}$$

for some auxiliary random variables (U,V) satisfying the Markov chains $U \to X \to Y$ and $V \to Y \to Z$ and the rate constraints

$$R \geq I(U;X|W), \tag{153}$$

$$T \geq I(V;Y|W,Q), \tag{154}$$

Proof: Achievability follows by specializing Theorem 5 to $S=0$. For the converse, see Appendix J. ■

VII. Concluding Remarks

The paper presents coding and testing schemes for a basic multi-hop network with a single relay, for its K-relay extension, and for a network with two parallel multi-hop channels that share a common receiver, and analyzes their exponent-rate region. The schemes are similar to Han's scheme in [2] for the point-to-point channel but also introduce coding techniques for cascade source coding and an unanimous-decision forwarding strategy where any terminal only decides on the null-hypothesis $\mathcal{H}=0$ if all preceding terminals have decided on this null-hypothesis. The more complicated code construction and the unanimous-decision forwarding strategy complicate the analysis compared to the point-to-point setup [2] and to the previously considered many-to-one [1], [3], [4], two-way [7], [12], [13], or broadcast scenarios [5]–[7]. In particular, they require imposing an additional structure on the codebooks: each joint type should not be over-represented on the set of codeword tuples in these codebooks. For each of the three considered networks, the proposed scheme is optimal for some instances of testing against independence.

When combined with binning, the scheme for the single-relay multi-hop network is also optimal for an instance of testing against conditional independence. The binning extensions for the other networks are not detailed out due to their involved forms. Instead, for completeness, an analysis of the Shimokawa-Han-Amari exponent-rate function [3] for the point-to-point setup is presented.

The derived exponent-rate regions for the schemes without binning all have the same structure, see Theorems 1, 2, and 3: the exponents are obtained by minimizing the KL-divergence $D(\tilde{P}\|Q)$ with Q being the joint pmf on the codewords and observations under $\mathcal{H}=1$ and $\tilde{P}$ being a joint pmf over the same quantities and the optimization variables that is subject to decoding/testing constraints and rate-constraints stemming from the source codes. Future work encompasses showing that similar exponent-rate regions are achievable also in other setups.

Another line of potential future research is to find coding and testing schemes beyond the random code constructions discussed in this paper. This is motivated by the following example, for which the current coding and testing schemes are not sufficient. Let X and Y be independent Bernoulli-$1/2$ random variables under both hypotheses, and let $Z=X\oplus Y$ when $\mathcal{H}=0$ and $Z=X\oplus Y\oplus 1$ when $\mathcal{H}=1$. In this example, the receiver can have an infinite error exponent even when $R=T=0$. In fact, it suffices that the transmitter sends its first observed symbol X_1 to the relay, which then forwards $Y_1\oplus X_1$ to the receiver. If $Z_1=Y_1\oplus X_1$, then the relay decides on $\hat{H}_z=0$ and otherwise on $\hat{\mathcal{H}}_z=1$. With the described coding and testing scheme the probability of error at any given blocklength n is 0. This corresponds to an unbounded error exponent.

VIII. Acknowledgement

The authors would like to thank Pierre Escamilla and Abdellatif Zaidi for helpful discussions.

References

[1] A. Ahlswede and I. Csiszar, "Hypothesis testing with communication constraints," *IEEE Trans. on Info. Theory*, vol. 32, no. 4, pp. 533–542, Jul. 1986.
[2] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. on Info. Theory*, vol. 33, no. 6, pp. 759–772, Nov. 1987.
[3] H. Shimokawa, T. Han and S. I. Amari, "Error bound for hypothesis testing with data compression," in *Proc. IEEE Int. Symp. on Info. Theory*, Jul. 1994, p. 114.
[4] M. S. Rahman and A. B. Wagner, "On the Optimality of binning for distributed hypothesis testing," *IEEE Trans. on Info. Theory*, vol. 58, no. 10, pp. 6282–6303, Oct. 2012.
[5] S. Salehkalaibar, M. Wigger and R. Timo, "On hypothesis testing against independence with multiple decision centers," *Submitted to IEEE Trans. Comm*, May 2017.
[6] M. Wigger and R. Timo, "Testing against independence with multiple decision centers," in *Proc. of SPCOM*, Bangalore, India, Jun. 2016.
[7] P. Escamilla, A. Zaidi, and M. Wigger, "Distributed hypothesis testing with collaborative detection," *Submitted to Information Theory Workshop (ITW)*, 2017.
[8] K. R. Varshney and L. R. Varshney, "Quantization of prior probabilities for hypothesis testing," *IEEE Trans. Signal Proc.*, vol. 56, no. 10, pp. 4553–4562, Oct. 2008.
[9] Y. Li, S. Nitinawarat and V. V. Veeravalli, "Universal outlier hypothesis testing," in *Proc. IEEE Int. Symp on Info. Theory*, Istanbul, Turkey, Jul. 2013, pp. 2666–2670.
[10] M. Naghshvar and T. Javidi, "Active M-ary sequential hypothesis testing," in *Proc. IEEE Int. Symp on Info. Theory*, Austin, Texas, Jun 2010, pp. 1623–1627.
[11] H. V. Poor, *An introduction to signal detection and estimation*, Springer, 1994.
[12] W. Zhao and L. Lai, "Distributed testing against independence with multiple terminals," in *Proc. 52nd Allerton Conf. Comm, Cont. and Comp.*, IL, USA, pp. 1246–1251, Oct. 2014.
[13] Y. Xiang and Y. H. Kim, "Interactive hypothesis testing against independence," in *Proc. IEEE Int. Symp. on Info. Theory*, Istanbul, Turkey, pp. 2840–2844, Jun. 2013.
[14] S. Sreekuma and D. Gunduz, "Distributed hypothesis testing over noisy channels," available at: https://arxiv.org/abs/1704.01535.
[15] P. Cuff, H. I. Su and A. El Gamal, "Cascade multiterminal source coding," in *Proc. IEEE Int. Symp. on Info. Theory*, Seoul, Korea, pp. 1199–1203, Jun.-Jul. 2009.
[16] A. El Gamal and Y. H. Kim, *Network information theory*, Cambridge Univ. Press, 2011.
[17] Y.-K. Chia, H. H. Permuter, and T. Weissman, "Cascade, Triangular, and Two-Way Source Coding With Degraded Side Information at the Second User," *IEEE Trans. on Info. Theory*, vol. 58, no. 1, pp. 189–206, Jan. 2012.

[18] H. H. Permuter and T. Weissman, "Cascade and triangular source coding with side information at the first two nodes," *IEEE Trans. on Info. Theory*, vol. 58, no. 8, pp. 3339–3339, June 2012.
[19] R. Tandon, S. Mohajer, and H. V. Poor, "Cascade source coding with erased side information, *in Proc. IEEE Int. Symp. on Info. Theory*, Austin, Texas, pp. 2944–2948, Jun.-Jul. 2009.
[20] I. Csiszar and J. Korner, *Information theory: coding theorems for discrete memoryless systems*, Cambridge Univ. Press, 2nd ed., 2011.
[21] G. Katz, P. Piantanida and M. Debbah, "Collaborative distributed hypothesis testing," *arXiv*, 1604.01292, Apr. 2016.

APPENDIX A
PROOF LEMMA 1

We prove the existence of the desired codebooks using probabilistic arguments. Construct a random codebook $\mathcal{C}_S = \{S^n(i)\colon i = 1, \ldots, \lfloor 2^{nR_s} \rfloor\}$ by selecting each entry of the n-length codeword $S^n(i)$ in an i.i.d. manner according to the pmf P_S. For each index i, generate a random codebook $\mathcal{C}_U(i) = \{U^n(j|i)\colon j \in \{1, ..., \lfloor 2^{nR_u} \rfloor\}\}$ by selecting each entry of the n-length codeword $U^n(j|i)$ in an i.i.d. and memoryless manner using the conditional pmf $P_{U|S}(\cdot|S_t(i))$, where $S_t(i)$ denotes the t-th component of the random codeword $S^n(i)$. For each index i, generate a random codebook $\mathcal{C}_V(i) = \{V^n(k|i)\colon k \in \{1, ..., \lfloor 2^{nR_v} \rfloor\}\}$ by selecting the t-th entry of the n-length codeword $V^n(k|i)$ in a memoryless manner using the conditional pmf $P_{V|S}(\cdot|S_t(i))$.

By the covering lemma and rate expressions (14)–(16), when averaged over the choice of the random codebooks, properties 1) and 2) are satisfied for all sufficiently large n. Also, property 3) holds averaged over the choice of the random codebooks from the following sequence of inequalities. For any joint type π_{SUV} and $\mu' > 0$:

$$\begin{aligned}
&\mathbb{E}_{\mathcal{C}}\left[\left|\left\{(i,j,k)\colon \quad \text{tp}(S^n(i), U^n(j|i), V^n(k|i)) = \pi_{SUV}\right\}\right|\right] \\
&= \sum_{i,j,k} \Pr[\text{tp}(S^n(i), U^n(j|i), V^n(k|i)) = \pi_{SUV}] \\
&= \sum_{i,j,k} \sum_{\substack{(s^n,u^n,v^n)\colon \\ \text{tp}(s^n,u^n,v^n)=\pi_{SUV}}} \Pr[(S^n(i), U^n(j|i), V^n(k|i)) = (s^n, u^n, v^n)] \\
&\overset{(a)}{\leq} \sum_{i,j,k} \sum_{\substack{(s^n,u^n,v^n)\colon \\ \text{tp}(s^n,u^n,v^n)=\pi_{SUV}}} 2^{-n(H_\pi(S)+D(\pi_S\|P_S))} \times 2^{-n(H_\pi(U|S)+D(\pi_{U|S}\|P_{U|S}))} \times 2^{-n(H_\pi(V|S)+D(\pi_{V|S}\|P_{V|S}))} \\
&\leq \sum_{i,j,k} \sum_{\substack{(s^n,u^n,v^n)\colon \\ \text{tp}(s^n,u^n,v^n)=\pi_{SUV}}} 2^{-n(H_\pi(S)+H_\pi(U|S)+H_\pi(V|S)-\mu')} \\
&\overset{(b)}{\leq} \sum_{i,j,k} 2^{nH_\pi(USV)} 2^{-n(H_\pi(S)+H_\pi(U|S)+H_\pi(V|S)-\mu')} \\
&\leq 2^{n(R_s+R_u+R_v)} \cdot 2^{-n(I_\pi(U;V|S)-\mu')},
\end{aligned} \tag{155}$$

where (a) follows from [2, Lemma 3d], (b) follows from [2, Lemma 3c]. Therefore, we have shown that for sufficiently large n and all $\mu, \epsilon > 0$, we get

$$\Pr\left[\exists (i,j)\colon (S^n(i), U^n(j|i), X^n) \in \mathcal{T}^n_{\mu/4}(P_{SUX})\right] > 1 - \epsilon/32, \tag{156}$$

$$\Pr\left[\exists k\colon (S^n(i), U^n(j|i), V^n(k|i), Y^n) \in \mathcal{T}^n_{\mu/2}(P_{SUVY}) \middle| (S^n(i), U^n(j|i), Y^n) \in \mathcal{T}^n_{3\mu/8}(P_{SUY})\right] > 1 - \epsilon/32, \tag{157}$$

$$\mathbb{E}_{\mathcal{C}}\left[\left|\left\{(i,j,k)\colon \text{tp}(S^n(i), U^n(j|i), V^n(k|i)) = \pi_{SUV}\right\}\right|\right] \leq 1/8 \cdot 2^{n(R_s+R_u+R_v)} \cdot 2^{-n(I_\pi(U;V|S)-\mu)}. \tag{158}$$

Now, we must show that there exists at least one codebook that satisfies properties (20), (21), and (22) in Lemma 1. We first order the sets of codebooks according to property (20), and then restrict to the subset of best codebooks that have total probability at least $1/2$. We then repeat this step by ordering the codebooks according to properties (21), and then (22). In this way, we construct a nonempty subset of codebooks $\mathcal{C}_S$, $\{\mathcal{C}_U\}$, and $\{\mathcal{C}_V\}$ so that we have

$$\Pr\left[\exists (i,j)\colon (s^n(i), u^n(j|i), X^n) \in \mathcal{T}^n_{\mu/4}(P_{SUX})\right] > 1 - \epsilon/4, \tag{159}$$

$$\Pr\left[\exists k\colon (s^n(i), u^n(j|i), v^n(k|i), Y^n) \in \mathcal{T}^n_{\mu/2}(P_{SUVY}) \middle| (S^n(i), U^n(j|i), Y^n) \in \mathcal{T}^n_{3\mu/8}(P_{SUY})\right] > 1 - \epsilon/4, \tag{160}$$

$$\left|\left\{(i,j,k)\colon \text{tp}(S^n(i), U^n(j|i), V^n(k|i)) = \pi_{SUV}\right\}\right| \leq 2^{n(R_s+R_u+R_v)} \cdot 2^{-n(I_\pi(U;V|S)-\mu)}, \tag{161}$$

which completes the proof of lemma.

Appendix B
Proof of Theorem 1

The analysis of the error probabilities at the relay are standard. We therefore focus on the error probabilities at the receiver. If $M \neq 0$ and $B \neq 0$, let I, J, K be the random indices sent over the noise-free bit pipes and define the following events:

$$\begin{aligned}\mathcal{E}_{\text{Relay}}: &\quad \big\{(s^n(I), u^n(J|I), Y^n) \notin \mathcal{T}^n_{3\mu/8}(P_{SUY}\big\},\\ \mathcal{E}_{\text{Rx}}: &\quad \big\{(s^n(I), u^n(J|I), v^n(K|I), Z^n) \notin \mathcal{T}^n_{\mu}(P_{SUVZ})\big\}.\end{aligned}$$

The type-I error probability can be bounded as:

$$\begin{aligned}\alpha_n &\leq \Pr[M=0 \text{ or } B=0 \text{ or } \mathcal{E}_{\text{Relay}} \text{ or } \mathcal{E}_{\text{Rx}}]\\ &\leq \Pr[M=0] + \Pr[B=0 \text{ or } \mathcal{E}_{\text{Relay}}|M\neq 0] + \Pr[\mathcal{E}_{\text{Rx}}|M\neq 0, B\neq 0] \qquad (162)\\ &\overset{(a)}{\leq} \epsilon/4 + \Pr[\mathcal{E}_{\text{Relay}}|M\neq 0] + \Pr[B=0|M\neq 0, \mathcal{E}^c_{\text{Relay}}] + \epsilon/4 \qquad (163)\\ &\overset{(b)}{\leq} \epsilon/4+\epsilon/4+\epsilon/4+\epsilon/4 \qquad (164)\\ &= \epsilon, \qquad (165)\end{aligned}$$

where (a) holds because the chosen code construction satisfies (20) and by the Markov lemma, and (b) holds because the code construction satisfies (21) and by the Markov lemma.

We now bound the probability of type-II error at the receiver. Let $\mathcal{P}^n$ be the set of all types over the product alphabets $\mathcal{S}^n \times \mathcal{U}^n \times \mathcal{V}^n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$. Also, let $\mathcal{P}^n_\mu$ be the subset of types $\pi_{SUVXYZ} \in \mathcal{P}^n$ that simultaneously satisfy the following three conditions:

$$\begin{aligned}|\pi_{SUX} - P_{SUX}| &\leq \mu/4, \qquad (166)\\ |\pi_{SUVY} - P_{SUVY}| &\leq \mu/2, \qquad (167)\\ |\pi_{SVZ} - P_{SVZ}| &\leq \mu. \qquad (168)\end{aligned}$$

The type-II error probability can then be written as:

$$\beta_n \leq \Pr\big[(X^n, Y^n, Z^n) \in \mathcal{A}_{\text{Rx},n}|\mathcal{H}=1\big], \qquad (169)$$

where $\mathcal{A}_{\text{Rx},n}$ includes the acceptance region at the receiver:

$$\mathcal{A}_{\text{Rx},n} \triangleq \bigcup_{i,j,k} \left\{(x^n, y^n, z^n) \colon \text{tp}(s^n(i), u^n(j|i), v^n(k|i), x^n, y^n, z^n) \in \mathcal{P}^n_\mu\right\}. \qquad (170)$$

We thus have

$$\beta_n \leq \sum_{\pi_{SUVXYZ} \in \mathcal{P}^n_\mu} \sum_{\substack{(i,j,k):\\ \text{tp}(s^n(i), u^n(j|i), v^n(k|i)) = \pi_{SUV}}} \sum_{\substack{(x^n,y^n,z^n):\\ \text{tp}(s^n(i),u^n(j|i),v^n(k|i),x^n,y^n,z^n)=\pi_{SUVXYZ}}} \Pr[X^n=x^n, Y^n=y^n, Z^n=z^n|\mathcal{H}=1]. \qquad (171)$$

Notice that for a triple (x^n, y^n, z^n) of type π_{XYZ}:

$$\Pr[X^n=x^n, Y^n=y^n, Z^n=z^n|\mathcal{H}=1] = 2^{-n(H_\pi(X,Y,Z)+D(\pi_{XYZ}\|Q_{XYZ}))}. \qquad (172)$$

Moreover, by standard arguments, for any joint type π_{SUVXYZ} and any triple of sequences $(s^n,\, u^n, v^n)$ of matching subtype π_{SUV}:

$$\big|\big\{(x^n, y^n, z^n)) \colon \text{tp}(s^n, u^n, v^n, x^n, y^n, z^n) = \pi_{SUVXYZ}\big\}\big| \leq 2^{nH_\pi(X,Y,Z|S,U,V)}. \qquad (173)$$

Also,

$$|\mathcal{P}^n_\mu| \leq |\mathcal{P}^n| \leq (n+1)^{|\mathcal{S}|\cdot|\mathcal{U}|\cdot|\mathcal{V}|\cdot|\mathcal{X}|\cdot|\mathcal{Y}|\cdot|\mathcal{Z}|}. \qquad (174)$$

Combining (171)–(174) with Property (22) in Lemma 1 yields the following upper bound:

$$\begin{aligned}\beta_n \leq &(n+1)^{|\mathcal{S}|.|\mathcal{U}|.|\mathcal{V}|.|\mathcal{X}|.|\mathcal{Y}|.|\mathcal{Z}|}\\ &\times \max_{\pi_{USVXYZ}\in\mathcal{P}^n_\mu} \Big[2^{n(R_s+R_u+R_v-I_\pi(U;V|S)+\mu)} \cdot 2^{nH_\pi(X,Y,Z|S,U,V)} \cdot 2^{-n(H_\pi(XYZ)+D(\pi_{XYZ}\|Q_{XYZ}))}\Big]. \qquad (175)\end{aligned}$$

Plugging the rate expressions (14)–(16) into (175) and simplifying, results in the following upper bound, we obtainL:

$$\beta_n \leq (n+1)^{|\mathcal{S}|.|\mathcal{U}|.|\mathcal{V}|.|\mathcal{X}|.|\mathcal{Y}|.|\mathcal{Z}|} \cdot 2^{-n\theta_\mu}. \tag{176}$$

where

$$\theta_\mu \triangleq \min_{\pi_{USVXYZ} \in \mathcal{P}_\mu^n} \big[H_\pi(XYZ) + D(\pi_{XYZ} \| Q_{XYZ}) - H_\pi(XYZ|SUV) + I_\pi(U;V|S) \big] - I(X;SU) - I(YU;V|S). \tag{177}$$

Now, we let $\mu \to 0$ and $n \to \infty$. By continuity properties, after performing some simple manipulations, we obtain that $\theta_\mu \to \theta$ as $\mu \to 0$, where

$$\theta \triangleq \min_{\substack{\tilde{P}_{SUVXYZ}: \\ \tilde{P}_{SUX} = P_{SUX} \\ \tilde{P}_{SVUY} = P_{SVUY} \\ \tilde{P}_{SVZ} = P_{SVZ}}} D(\tilde{P}_{SUVXYZ} \| P_{SU|X} P_{V|SUY} Q_{XYZ}). \tag{178}$$

Appendix C
Proof of Proposition 1

Fix a pair of probability distributions $P_{SU|X}$ and $P_{V|SUY}$. We will show that there exists a new pair of probability distributions $P_{\bar{U}|X}$ and $P_{\bar{V}|Y}$ so that

$$\min_{\substack{\tilde{P}_{SUVXYZ}: \\ \tilde{P}_{SUX} = P_{SU|X} P_X \\ \tilde{P}_{SUVY} = P_{SUVY} \\ \tilde{P}_{SVZ} = P_{SVZ}}} D(\tilde{P}_{SUVXYZ} \| P_{SU|X} P_{V|SUY} Q_{XY} Q_{Z|Y}) \geq \min_{\substack{\tilde{P}_{\bar{U}\bar{V}XYZ}: \\ \tilde{P}_{\bar{U}X} = P_{\bar{U}X} \\ \tilde{P}_{\bar{U}\bar{V}Y} = P_{\bar{U}\bar{V}Y} \\ \tilde{P}_{\bar{V}Z} = P_{\bar{V}Z}}} D(\tilde{P}_{SUVXYZ} \| P_{\bar{U}|X} P_{\bar{V}|Y} Q_{XY} Q_{Z|Y}). \tag{179}$$

This will conclude the proof of the proposition.

Consider the following sequence of inequalities:

$$\begin{aligned}
&\min_{\substack{\tilde{P}_{SUVXYZ}: \\ \tilde{P}_{SUX} = P_{SUX} \\ \tilde{P}_{SUVY} = P_{SUVY} \\ \tilde{P}_{SVZ} = P_{SVZ}}} D(\tilde{P}_{SUVXYZ} \| P_{SU|X} P_{V|SUY} Q_{XY} Q_{Z|Y}) \\
&\overset{(a)}{=} \min_{\substack{\tilde{P}_{SUVXYZ}: \\ \tilde{P}_{SUX} = P_{SUX} \\ \tilde{P}_{SUVY} = P_{SUVY} \\ \tilde{P}_{SVZ} = P_{SVZ}}} D(\tilde{P}_{SUXY} \| P_{SU|X} Q_{XY}) + \mathbb{E}_{\tilde{P}_{SUXY}} \big[D(\tilde{P}_{VZ|SUXY} \| P_{V|SUY} Q_{Z|Y}) \big] \\
&\overset{(b)}{\geq} \min_{\substack{\tilde{P}_{\bar{U}XY}: \\ \tilde{P}_{\bar{U}X} = P_{\bar{U}X} \\ \tilde{P}_{\bar{U}Y} = P_{\bar{U}Y}}} D(\tilde{P}_{\bar{U}XY} \| P_{\bar{U}|X} Q_{XY}) + \min_{\substack{\tilde{P}_{SUVYZ}: \\ \tilde{P}_{SUVY} = P_{SUVY} \\ \tilde{P}_{SVZ} = P_{SVZ}}} \mathbb{E}_{\tilde{P}_{SUY}} [D(\tilde{P}_{VZ|SUY} \| P_{V|SUY} Q_{Z|Y})] \qquad (180) \\
&\overset{(c)}{=} \min_{\substack{\tilde{P}_{\bar{U}XY}: \\ \tilde{P}_{\bar{U}X} = P_{\bar{U}X} \\ \tilde{P}_{\bar{U}Y} = P_{\bar{U}Y}}} D(\tilde{P}_{\bar{U}XY} \| P_{\bar{U}|X} Q_{XY}) \\
&\qquad + \min_{u,s} \min_{\substack{\tilde{P}_{VYZ|S=s,U=u}: \\ \tilde{P}_{VY|S=s,U=s} = P_{VY|S=s,U=s} \\ \tilde{P}_{VZ|S=s} = P_{VZ|S=s}}} \mathbb{E}_{\tilde{P}_{Y|S=s,U=u}} [D(\tilde{P}_{VZ|Y,S=s,U=u} \| P_{V|Y,S=s,U=u} Q_{Z|Y})] \qquad (181) \\
&\overset{(d)}{\geq} \min_{\substack{\tilde{P}_{\bar{U}XY}: \\ \tilde{P}_{\bar{U}X} = P_{\bar{U}X} \\ \tilde{P}_{\bar{U}Y} = P_{\bar{U}Y}}} D(\tilde{P}_{UXY} \| P_{U|X} Q_{XY}) + \min_{\substack{\tilde{P}_{\bar{V}YZ}: \\ \tilde{P}_{\bar{V}Y} = P_{\bar{V}Y} \\ \tilde{P}_{\bar{V}Z} = P_{\bar{V}Z}}} \mathbb{E}_{\tilde{P}_Y} [D(\tilde{P}_{\bar{V}Z|Y} \| P_{\bar{V}|Y} Q_{Z|Y})] \qquad (182)
\end{aligned}$$

where (a) follows by the chain rule for KL-divergences; (b) follows by introducing $\bar{U} := (U,S)$, because for any functions g_1 and g_2 it holds that $\min_x(g_1(x) + g_2(x)) \geq \min_x(g_1(x) + \min_x(g_2(x))$, and by the data-processing inequality for KL-divergences; and (c) holds because the minimum cannot be larger than the average and thus the minimum will be achieved for a degenerate distribution $\tilde{P}_{US}$; and (d) by defining $\bar{V}$ so that for all pairs (v,y):

$$P_{\bar{V}|Y}(v|y) = P_{V|SUY}(v|s^*,u^*,y)$$

where s^*, u^* denotes the minimizing (s,u)-pair in (181).

APPENDIX D
PROOF OF PROPOSITION 2

Fix a sequence of encoding and decoding functions $\{\phi^{(n)}, \phi_y^{(n)}, g_y^{(n)}, g_z^{(n)}\}$ so that the inequalities of Definition 1 hold for sufficiently large blocklengths n. Fix also such a sufficiently large n and define for each $t \in \{1, \ldots, n\}$:

$$\begin{aligned} U_t &\triangleq (M, X^{t-1}) \\ V_t &\triangleq (B, Y^{t-1}). \end{aligned} \tag{183}$$

Define also $U \triangleq (U_Q, Q)$; $V \triangleq (V_Q, Q)$; $X \triangleq X_Q$; $Y \triangleq Y_Q$; and $Z \triangleq Z_Q$; for $Q \sim \mathcal{U}\{1, \ldots, n\}$ independent of the tuples $(S^n, U^n, V^n, X^n, Y^n, Z^n)$. Notice the Markov chains $U \to X \to Y$ and $V \to Y \to Z$, where the latter holds in particular because P_{XYZ} decomposes as $P_X P_{Y|X} P_{Z|Y}$.

We start by lower bounding rate R:

$$\begin{aligned} R &= \frac{1}{n} H(M) \\ &\geq \frac{1}{n} I(M; X^n) \\ &= \frac{1}{n} \sum_{i=1}^{n} I(M; X_t | X^{t-1}) \\ &\stackrel{(a)}{=} \frac{1}{n} \sum_{t=1}^{n} I(M, X^{t-1}; X_t) \\ &= \frac{1}{n} \sum_{t=1}^{n} I(U_t; X_t) \\ &\stackrel{(b)}{=} I(U; X), \end{aligned} \tag{184}$$

where (a) holds because the sources are i.i.d, and (b) holds by the definition of the random variables Q, U, X and because X is independent of Q. Following similar steps, rate T can be lower bounded as:

$$\begin{aligned} T &= \frac{1}{n} H(B) \\ &\geq \frac{1}{n} I(B; Y^n) \\ &= \frac{1}{n} \sum_{i=1}^{n} I(B, Y^{t-1}; Y_t) \\ &= I(V; Y). \end{aligned}$$

The type-II error probability at the relay can be bounded as:

$$\begin{aligned} -\frac{1}{n} \log \zeta_n &\leq \frac{1}{n} D(P_{MY^n|\mathcal{H}=0} \| P_{MY^n|\mathcal{H}=1}) + \epsilon \\ &\stackrel{(a)}{=} \frac{1}{n} I(M; Y^n) + \epsilon \\ &= \frac{1}{n} \sum_{t=1}^{n} I(M; Y_t | Y^{t-1}) + \epsilon \\ &= \frac{1}{n} \sum_{t=1}^{n} I(M, Y^{t-1}; Y_t) + \epsilon \\ &\stackrel{(b)}{\leq} \frac{1}{n} \sum_{t=1}^{n} I(M, X^{t-1}; Y_t) + \epsilon \\ &= \frac{1}{n} \sum_{t=1}^{n} I(U_t; Y_t) + \epsilon \\ &= I(U; Y) + \epsilon \end{aligned} \tag{185}$$

where (a) follows because under hypothesis $\mathcal{H} = 1$, M and Y^n are independent, and (b) follows from the Markov chain $Y^{t-1} \to (M, X^{t-1}) \to Y_t$.

We finally turn to the type-II error probability at the receiver:

$$
\begin{aligned}
-\frac{1}{n}\log\beta_n &\le \frac{1}{n}D(P_{BZ^n|\mathcal{H}=0}\|P_{BZ^n|\mathcal{H}=1})+\epsilon\\
&\overset{(a)}{=} \frac{1}{n}D(P_{B|\mathcal{H}=0}\|P_{B|\mathcal{H}=1})+\frac{1}{n}\mathbb{E}_B\left[D(P_{Z^n|B,\mathcal{H}=0}\|P_{Z^n|B,\mathcal{H}=1})\right]+\epsilon\\
&\overset{(b)}{\le} \frac{1}{n}D(P_{MY^n|\mathcal{H}=0}\|P_{MY^n|\mathcal{H}=1})+\mathbb{E}_B\left[D(P_{Z^n|B,\mathcal{H}=0}\|P_{Z^n|\mathcal{H}=1})\right]+\epsilon\\
&\overset{(c)}{=} \frac{1}{n}I(M;Y^n)+\frac{1}{n}I(B;Z^n)+\epsilon\\
&\le \frac{1}{n}\sum_{t=1}^{n}I(U_t;Y_t)+\frac{1}{n}\sum_{t=1}^{n}I(V_t;Z_t)+\epsilon\\
&= I(U;Y)+I(V;Z)+\epsilon,
\end{aligned}
\tag{186}
$$

where (a) follows by the chain rule for KL-divergences; (b) follows on one hand because B is a function of (M,Y^n) and by the data processing inequality for KL-divergences, and on the other hand because under $\mathcal{H}=1$ observation Z^n is independent of B with same marginal distribution as under $\mathcal{H}=0$; (c) follows because under $\mathcal{H}=1$ observation Y^n is independent of M with same marginal distributions as under $\mathcal{H}=0$; and the last two inequalities can be proved by proceeding along similar lines as the proof of (185) and exploiting the Markov chains $Y^{t-1}\to(M,X^{t-1})\to Y_t$ and $Z^{t-1}\to(B,Y^{t-1})\to Z_t$.

Appendix E
Proof of Proposition 3

We fix a sufficiently large n and a sequence of encoding and decoding functions such that the properties of Definition 1 hold. Also, define $S_t\overset{\Delta}{=}(M,X^{t-1},Z^{t-1})$. First, consider the rate R:

$$
\begin{aligned}
nR &= H(M)\\
&\ge I(M;X^n,Z^n)\\
&= \sum_{t=1}^{n}I(M;X_t,Z_t|X^{t-1},Z^{t-1})\\
&= \sum_{t=1}^{n}I(M,X^{t-1},Z^{t-1};X_t,Z_t)\\
&\ge \sum_{t=1}^{n}I(M,X^{t-1},Z^{t-1};X_t)\\
&= \sum_{t=1}^{n}I(S_t;X_t)
\end{aligned}
$$

Now, consider the error exponent at the relay:

$$
\begin{aligned}
-\frac{1}{n}\log\zeta_n &\le \frac{1}{n}D(P_{MY^n|\mathcal{H}=0}\|P_{MY^n|\mathcal{H}=1})+\epsilon\\
&\overset{(a)}{=} \frac{1}{n}I(M;Y^n)+\epsilon\\
&= \frac{1}{n}\sum_{t=1}^{n}I(M;Y_t|Y^{t-1})+\epsilon\\
&\overset{(b)}{=} \frac{1}{n}\sum_{t=1}^{n}I(M,Y^{t-1};Y_t)+\epsilon\\
&\le \frac{1}{n}\sum_{t=1}^{n}I(M,X^{t-1},Y^{t-1},Z^{t-1};Y_t)+\epsilon\\
&\overset{(c)}{=} \frac{1}{n}\sum_{t=1}^{n}I(M,X^{t-1},Z^{t-1};Y_t)+\epsilon\\
&= \frac{1}{n}\sum_{t=1}^{n}I(S_t;Y_t)+\epsilon
\end{aligned}
$$

where (a) follows because under hypothesis $\mathcal{H}=1$, M and Y^n are independent, (b) follows from the memoryless property of the channel, and (c) follows from the Markov chain $Y^{t-1} \to (M, X^{t-1}) \to Y_t$. Next, consider the error exponent at the receiver:

$$\begin{aligned}
-\frac{1}{n}\log\beta_n &\leq \frac{1}{n}D(P_{BZ^n|\mathcal{H}=0}\|P_{BZ^n|\mathcal{H}=1})+\epsilon \\
&\overset{(a)}{\leq} \frac{1}{n}D(P_{MY^nZ^n|\mathcal{H}=0}\|P_{MY^nZ^n|\mathcal{H}=1})+\epsilon \\
&\overset{(b)}{=} \frac{1}{n}I(M;Y^n,Z^n)+\epsilon \\
&\overset{(c)}{=} \frac{1}{n}I(M;Z^n)+\epsilon \\
&= \frac{1}{n}\sum_{t=1}^{n} I(M;Z_t|Z^{t-1})+\epsilon \\
&= \frac{1}{n}\sum_{t=1}^{n} I(M,Z^{t-1};Z_t)+\epsilon \\
&\leq \frac{1}{n}\sum_{t=1}^{n} I(M,X^{t-1},Z^{t-1};Z_t)+\epsilon \\
&\leq \frac{1}{n}\sum_{t=1}^{n} I(S_t;Z_t)+\epsilon
\end{aligned}$$

where (a) follows from the data processing inequality, (b) follows because M and (Y^n, Z^n) are independent under hypothesis $\mathcal{H}=1$, and (c) follows from the Markov chain $M \to X^n \to Z^n \to Y^n$. The proof of the converse is finally concluded by defining a time-sharing random variable $Q \sim \mathcal{U}\{1,...,n\}$ and $S \triangleq (S_Q, Q)$, $X \triangleq X_Q$, $Y \triangleq Y_Q$ and $Z \triangleq Z_Q$.

Appendix F
Proof of Proposition 4

Fix a sequence of encoding and decoding functions that satisfy properties of Definition 1, and then a sufficiently large n. Define $U_t \triangleq (M, X^{t-1})$, and consider first the rate R:

$$\begin{aligned}
nR &= H(M) \\
&\geq I(M;X^n) \\
&= \sum_{t=1}^{n} I(M;X_t|X^{t-1}) \\
&= \sum_{t=1}^{n} I(M,X^{t-1};X_t) \\
&= \sum_{t=1}^{n} I(U_t;X_t)
\end{aligned}$$

Now, consider the type-II error probability at the relay:

$$\begin{aligned}
-\frac{1}{n}\log\zeta_n &\leq \frac{1}{n}D(P_{MY^n|\mathcal{H}=0}\|P_{MY^n|\mathcal{H}=1})+\epsilon \\
&= \frac{1}{n}I(M;Y^n)+\epsilon \\
&= \frac{1}{n}\sum_{t=1}^{n} I(M;Y_t|Y^{t-1})+\epsilon \\
&= \frac{1}{n}\sum_{t=1}^{n} I(M,Y^{t-1};Y_t)+\epsilon \\
&\leq \frac{1}{n}\sum_{t=1}^{n} I(M,X^{t-1};Y_t)+\epsilon \\
&= \frac{1}{n}\sum_{t=1}^{n} I(U_t;Y_t)+\epsilon,
\end{aligned}$$

and the type-II error probability at the receiver:

$$\begin{aligned}
-\frac{1}{n}\log\beta_n &\leq \frac{1}{n}D(P_{BZ^n|\mathcal{H}=0}\|P_{BZ^n|\mathcal{H}=1})+\epsilon \\
&\leq \frac{1}{n}D(P_{MY^nZ^n|\mathcal{H}=0}\|P_{MY^nZ^n|\mathcal{H}=1})+\epsilon \\
&\overset{(a)}{=} \frac{1}{n}I(M,Z^n;Y^n) \\
&= \frac{1}{n}I(M;Y^n|Z^n)+I(Y^n;Z^n) \\
&\overset{(b)}{=} \frac{1}{n}I(Z^n;Y^n) \\
&\leq \frac{1}{n}\sum_{t=1}^{n}I(Z_t;Y_t)+\epsilon, \qquad (187)
\end{aligned}$$

where here (a) follows because under hypothesis $\mathcal{H}=1$, (M,Z^n) are independent of Y^n, and (b) follows from the Markov chain $M\to X^n\to Z^n\to Y^n$.

The proof of the converse is finally concluded by defining a time-sharing random variable as $Q\sim\mathcal{U}\{1,...,n\}$ and $U\triangleq(U_Q,Q)$, $X\triangleq X_Q$, $Y\triangleq Y_Q$ and $Z\triangleq Z_Q$.

Appendix G
Proof of Proposition 6

Achievability: Consider the error exponent at the receiver:

$$\begin{aligned}
&\min_{\substack{\tilde{P}_{U_1V_1U_2V_2X_1Y_1X_2Y_2Z}:\\ \tilde{P}_{U_lX_l}=P_{U_lX_l}\\ \tilde{P}_{V_lU_lY_l}=P_{V_lU_lY_l}\\ \tilde{P}_{V_1V_2Z}=P_{V_1V_2Z}}} D(\tilde{P}_{U_1V_1U_2V_2X_1Y_1X_2Y_2Z}||P_{U_1|X_1}P_{U_2|X_2}P_{V_1|U_1Y_1}P_{V_2|U_2Y_2}P_{X_1}P_{X_2}P_{Y_1Y_2}P_Z) \\
&= \min_{\substack{\tilde{P}_{U_1V_1U_2V_2Y_1Y_2Z}:\\ \tilde{P}_{V_lU_lY_l}=P_{V_lU_lY_l}\\ \tilde{P}_{V_1V_2Z}=P_{V_1V_2Z}}} D(\tilde{P}_{U_1V_1U_2V_2Y_1Y_2Z}||P_{U_1}P_{U_2}P_{V_1|U_1Y_1}P_{V_2|U_2Y_2}P_{Y_1Y_2}P_Z) \\
&= \min_{\substack{\tilde{P}_{U_1V_1U_2V_2Y_1Y_2Z}:\\ \tilde{P}_{V_lU_lY_l}=P_{V_lU_lY_l}\\ \tilde{P}_{V_1V_2Z}=P_{V_1V_2Z}}} \Bigg[D(\tilde{P}_{Y_1Y_2}||P_{Y_1Y_2})+\mathbb{E}_{\tilde{P}_{Y_1Y_2}}[D(\tilde{P}_{U_1|Y_1Y_2}||P_{U_1})] \\
&\qquad +\mathbb{E}_{\tilde{P}_{U_1Y_1Y_2}}[D(\tilde{P}_{U_2|U_1Y_1Y_2}||P_{U_2})]+\mathbb{E}_{\tilde{P}_{U_1U_2Y_1Y_2}}[D(\tilde{P}_{V_1|U_1U_2Y_1Y_2}||P_{V_1|U_1Y_1})] \\
&\qquad +\mathbb{E}_{\tilde{P}_{U_1U_2V_1Y_1Y_2}}[D(\tilde{P}_{V_2|U_1U_2V_1Y_1Y_2}||P_{V_2|U_2Y_2})]+\mathbb{E}_{\tilde{P}_{U_1U_2V_1V_2Y_1Y_2}}[D(\tilde{P}_{Z|U_1U_2V_1V_2Y_1Y_2}||P_Z)]\Bigg] \\
&\geq I(Z;V_1V_2)+I(U_1;Y_1)+I(U_2;Y_2), \qquad (188)
\end{aligned}$$

where the last step holds because under conditions

$$\tilde{P}_{V_lU_lY_l}=P_{V_lU_lY_l}, \qquad l\in\{1,2\}, \qquad (189)$$

$$\tilde{P}_{V_1V_2Z}=P_{V_1V_2Z}, \qquad (190)$$

the following inequalities hold:

$$\begin{aligned}
D(\tilde{P}_{Y_1Y_2}||P_{Y_1Y_2}) &\geq 0 \\
\mathbb{E}_{\tilde{P}_{Y_1Y_2}}[D(\tilde{P}_{U_1|Y_1Y_2}||P_{U_1})] &= I_{\tilde{P}}(U_1;Y_1Y_2)\geq I_{\tilde{P}}(U_1;Y_1)=I(U_1;Y_1) \\
\mathbb{E}_{\tilde{P}_{U_1Y_1Y_2}}[D(\tilde{P}_{U_2|U_1Y_1Y_2}||P_{U_2})] &= I_{\tilde{P}}(U_2;Y_1Y_2U_1)\geq I_{\tilde{P}}(U_2;Y_2)=I(U_2;Y_2) \\
\mathbb{E}_{\tilde{P}_{U_1U_2Y_1Y_2}}[D(\tilde{P}_{V_1|U_1U_2Y_1Y_2}||P_{V_1|U_1Y_1})] &\geq 0 \\
\mathbb{E}_{\tilde{P}_{U_1U_2V_1Y_1Y_2}}[D(\tilde{P}_{V_2|U_1U_2V_1Y_1Y_2}||P_{V_2|U_2Y_2})] &\geq 0 \\
\mathbb{E}_{\tilde{P}_{U_1U_2V_1V_2Y_1Y_2}}[D(\tilde{P}_{Z|U_1U_2V_1V_2Y_1Y_2}||P_Z)] &= I_{\tilde{P}}(Z;U_1U_2V_1V_2Y_1Y_2)\geq I_{\tilde{P}}(Z;V_1V_2)=I(Z;V_1V_2). \qquad (191)
\end{aligned}$$

Moreover, each of these inequalities holds with equality for the choice

$$\tilde{P}_{U_1X_1U_2X_2Y_1Y_2V_1V_2Z}=P_{Y_1Y_2}\cdot P_{U_1|Y_1}\cdot P_{U_2|Y_2}\cdot P_{V_1|U_1Y_1}\cdot P_{V_2|U_2Y_2}\cdot P_{X_1|U_1}\cdot P_{X_2|U_2}, \qquad (192)$$

which represents a feasible choice, i.e., it satisfies both conditions (189) and (190). In particular, for all triples (v_1, v_2, z), the choice in (192) satisfies

$$\begin{aligned}\tilde{P}_{V_1V_2Z}(v_1,v_2,z) &= \sum_{u_1u_2y_1y_2} P_{Y_1Y_2}(y_1,y_2)P_{U_1|Y_1}(u_1|y_1)P_{U_2|Y_2}(u_2|y_2)P_{V_1|U_1Y_1}(v_1|u_1,y_1)P_{V_2|U_2Y_2}(v_2|u_2,y_2)P_{Z|V_1V_2}(z|v_1,v_2)\\ &= P_{V_1V_2Z}(v_1,v_2,z). \end{aligned} \tag{193}$$

Combining all these arguments shows that (188) holds with equality, which proves the achievability of the desired error exponent at the receiver.

The error exponents at the relays can be obtained similarly. Details omitted.

Converse: Define for each $t \in \{1,\dots,n\}$ and for $l \in \{1,2\}$:

$$\begin{aligned} U_{l,t} &\triangleq (M_l, X_l^{t-1})\\ V_{l,t} &\triangleq (B_l, Y_l^{t-1}). \end{aligned} \tag{194}$$

Then, consider rate

$$\begin{aligned} R_l &= \frac{1}{n}H(M_l)\\ &\geq \frac{1}{n}I(M_l;X_l^n)\\ &= \frac{1}{n}\sum_{t=1}^{n} I(M_l;X_{l,t}|X_l^{t-1})\\ &= \frac{1}{n}\sum_{t=1}^{n} I(M_l,X_l^{t-1};X_{l,t})\\ &= \frac{1}{n}\sum_{t=1}^{n} I(U_{l,t};X_{l,t}), \end{aligned}$$

and rate

$$\begin{aligned} T_l &= \frac{1}{n}H(B_l)\\ &\geq \frac{1}{n}I(B_l;Y_l^n)\\ &= \frac{1}{n}\sum_{t=1}^{n} I(B_l;Y_{l,t}|Y_l^{t-1})\\ &= \frac{1}{n}\sum_{t=1}^{n} I(B_l,Y_l^{t-1};Y_{l,t})\\ &= \frac{1}{n}\sum_{t=1}^{n} I(V_{l,t};Y_{l,t}). \end{aligned}$$

Now, consider the error exponent at Relay $l \in \{1,2\}$:

$$\begin{aligned} -\frac{1}{n}\log\zeta_{l,n} &\leq \frac{1}{n}D(P_{M_lY_l^n|\mathcal{H}=0}\|P_{M_lY_l^n|\mathcal{H}=1})+\epsilon\\ &= \frac{1}{n}I(M_l;Y_l^n)+\epsilon\\ &= \frac{1}{n}\sum_{t=1}^{n} I(M_l;Y_{l,t}|Y_l^{t-1})+\epsilon\\ &= \frac{1}{n}\sum_{t=1}^{n} I(M_l,Y_l^{t-1};Y_{l,t})+\epsilon\\ &\leq \frac{1}{n}\sum_{t=1}^{n} I(M_l,Y_l^{t-1},X_l^{t-1};Y_{l,t})+\epsilon\\ &\overset{(a)}{=} \frac{1}{n}\sum_{t=1}^{n} I(M_l,X_l^{t-1};Y_{l,t})+\epsilon \end{aligned}$$

$$= \frac{1}{n}\sum_{t=1}^{n} I(U_{l,t};Y_{l,t}) + \epsilon$$

where (a) follows from the Markov chain $Y_l^{t-1} \to (M_l, X_l^{t-1}) \to Y_{l,t}$. Similarly, for the error exponent at the receiver:

$$\begin{aligned}
-\frac{1}{n}\log\beta_n &\leq \frac{1}{n}D(P_{B_1B_2Z^n|\mathcal{H}=0}\|P_{B_1B_2Z^n|\mathcal{H}=1}) + \epsilon \\
&\overset{(a)}{=} \frac{1}{n}D(P_{B_1B_2|\mathcal{H}=0}\|P_{B_1B_2|\mathcal{H}=1}) + \frac{1}{n}\mathbb{E}_{B_1B_2}\big[D(P_{Z^n|B_1B_2,\mathcal{H}=0}\|P_{Z^n|B_1B_2,\mathcal{H}=1})\big] + \epsilon \\
&\overset{(b)}{\leq} \frac{1}{n}D(P_{M_1M_2Y_1^nY_2^n|\mathcal{H}=0}\|P_{M_1M_2Y_1^nY_2^n|\mathcal{H}=1}) + \frac{1}{n}\mathbb{E}_{B_1B_2}\big[D(P_{Z^n|B_1B_2,\mathcal{H}=0}\|P_{Z^n|B_1B_2,\mathcal{H}=1})\big] + \epsilon \\
&\overset{(c)}{=} \frac{1}{n}\mathbb{E}_{Y_1^nY_2^n}\big[D(P_{M_1M_2|Y_1^nY_2^n,\mathcal{H}=0}\|P_{M_1M_2|Y_1^n,Y_2^n,\mathcal{H}=1})\big] + \mathbb{E}_{B_1B_2}\big[D(P_{Z^n|B_1,B_2,\mathcal{H}=0}\|P_{Z^n|\mathcal{H}=1})\big] + \epsilon \\
&\overset{(d)}{=} \frac{1}{n}\mathbb{E}_{Y_1^n}\big[D(P_{M_1|Y_1^n,\mathcal{H}=0}\|P_{M_1|\mathcal{H}=1})\big] + \frac{1}{n}\mathbb{E}_{Y_2^n}\big[D(P_{M_2|Y_2^n,\mathcal{H}=0}\|P_{M_2|\mathcal{H}=1})\big] \\
&\quad + \mathbb{E}_{B_1B_2}\big[D(P_{Z^n|B_1,B_2,\mathcal{H}=0}\|P_{Z^n|\mathcal{H}=1})\big] + \epsilon \\
&= \frac{1}{n}I(M_1;Y_1^n) + \frac{1}{n}I(M_2;Y_2^n) + \frac{1}{n}I(B_1,B_2;Z^n) + \epsilon \\
&\overset{(e)}{\leq} \frac{1}{n}\sum_{t=1}^{n} I(U_{1,t};Y_{1,t}) + \frac{1}{n}\sum_{t=1}^{n} I(U_{2,t};Y_{2,t}) + \frac{1}{n}\sum_{t=1}^{n} I(V_{1,t},V_{2,t};Z_t) + \epsilon
\end{aligned} \tag{195}$$

where (a) holds by the chain rule for KL-divergences; (b) holds because B_1 and B_2 are functions of (M_1, Y_1^n) and (M_2, Y_2^n) and by the data processing inequality for KL-divergences; (c) holds on one hand because (Y_1^n, Y_2^n) have same joint distribution under both hypothesis, and on the other hand because under $\mathcal{H} = 1$ observation Z^n is independent of the two messages (B_1, B_2); (d) holds because under $\mathcal{H} = 1$, M_1 and M_2 are independent of each other and of Y_1^n, Y_2^n and under $\mathcal{H} = 0$ the Markov chains $M_1 \to Y_1^n \to (M_2, Y_2^n)$ and $M_2 \to Y_2^n \to (M_1, Y_1^n)$ hold; and (e) follows because of the Markov chains $Y_1^{t-1} \to (M_1, X^{t-1}) \to Y_{1,t}$; $Y_2^{t-1} \to (M_2, X^{t-1}) \to Y_{2,t}$; and $Z^{t-1} \to (B_1, B_2, Y_1^{t-1}, Y_2^{t-1}) \to Z_t$.

The proof is finalized by introducing appropriate auxiliary random variables and taking $\epsilon \to 0$.

Appendix H
Proof of Theorem 4

We analyze the probabilities of error of the coding and testing scheme described in Subsection VI-A2 averaged over the random code constructions. By standard arguments (successively eliminating the worst half of the codewords from the codebooks as detailed out in Appendix A) the desired result can be proved for a set of deterministic codebooks.

Fix an arbitrary $\epsilon > 0$ and the scheme's parameter $\mu > 0$. For a fixed blocklength n, let $\mathcal{P}^n$ be the set of all types over the product alphabets $\mathcal{S}^n \times \mathcal{S}^n \times \mathcal{X}^n \times \mathcal{Y}^n$. Also, let $\mathcal{P}^n_\mu$ be the subset of types $\pi_{SS'XY} \in \mathcal{P}^n$ that simultaneously satisfy the following conditions:

$$\forall (s,s',x,y) \in \mathcal{S}\times\mathcal{S}\times\mathcal{X}\times\mathcal{Y}: \qquad \begin{aligned} &|\pi_{SX}(s,x) - P_{SX}(s,x)| \leq \mu/2, \\ &|\pi_{S'Y}(s',y) - P_{SY}(s,y)| \leq \mu, \end{aligned} \tag{196}$$

and

$$H_{\pi_{S'Y}}(S|Y) \leq H_{\pi_{SY}}(S|Y). \tag{197}$$

Notice that when $\mu \to 0$ and $n \to \infty$, then

$$\mathcal{P}^n_\mu \to \mathcal{P}^* \qquad \text{as} \;\; \mu \to 0 \quad \text{and} \;\; n \to \infty, \tag{198}$$

where

$$\mathcal{P}^* \triangleq \big\{\tilde{P}_{SXY} \colon \tilde{P}_{SX} = P_{SX} \text{ and } \tilde{P}_{S'Y} = P_{SY} \text{ and } H_{\tilde{P}_{SY}}(S|Y) \geq H(S|Y)\big\}. \tag{199}$$

We first analyze the type-I error probability at the receiver. For the case of $M \neq 0$, let L be the index chosen at the transmitter. Define events

$$\mathcal{E}^{(1)}_{\text{Rx}} \colon \big\{(S^n(M,1), Y^n) \notin \mathcal{T}^n_\mu(P_{SY})\big\}, \tag{200}$$

$$\mathcal{E}^{(2)}_{\text{Rx}} \colon \big\{\exists \ell' \neq L \colon \quad H_{\text{tp}(S^n(M,\ell'),Y^n)}(S|Y) \leq \min_{\tilde{\ell}} H_{\text{tp}(S^n(M,\tilde{\ell}),Y^n)}(S|Y)\big\}. \tag{201}$$

For all sufficiently large n, the type-I error probability can be bounded as:

$$\alpha_n \leq \Pr[M = 0 \text{ or } \mathcal{E}^{(1)}_{\text{Rx}} \text{ or } \mathcal{E}^{(2)}_{\text{Rx}}]$$

$$
\begin{aligned}
&\leq \Pr[M=0] + \Pr[\mathcal{E}_{\text{Rx}}^{(1)}|M\neq 0] + \Pr[\mathcal{E}_{\text{Rx}}^{(2)}|M\neq 0, \mathcal{E}_{\text{Rx}}^{(1)c}] \\
&\overset{(a)}{\leq} \epsilon/3 + \Pr[\mathcal{E}_{\text{Rx}}^{(1)}|M\neq 0] + \Pr[\mathcal{E}_{\text{Rx}}^{(2)}|M\neq 0, \mathcal{E}_{\text{Rx}}^{(1)c}] \\
&\overset{(b)}{\leq} \epsilon/3 + \epsilon/3 + \Pr[\mathcal{E}_{\text{Rx}}^{(2)}|M\neq 0, \mathcal{E}_{\text{Rx}}^{(1)c}] \\
&\overset{(c)}{\leq} \epsilon/3 + \epsilon/3 + \epsilon/3 \\
&= \epsilon,
\end{aligned} \tag{202}
$$

where for sufficiently large n inequality (a) follows by the standard covering lemma [16] and by the rate constraint in (121); (b) follows from the Markov lemma [16]; and (c) from the following sequence of inequalities:[2]

$$
\begin{aligned}
&\Pr\left[\mathcal{E}_{\text{Rx}}^{(2)}|M\neq 0, \mathcal{E}_{\text{Rx}}^{(1)c}\right] \\
&= \Pr\left[H_{\text{tp}(S^n(M,L),Y^n)}(S|Y) \geq \min_{\tilde{\ell}\neq L} H_{\text{tp}(S^n(M,\tilde{\ell}),Y^n)}(S|Y) \;\middle|\; (S^n(M,L),Y^n)\in\mathcal{T}_{\mu}^n(P_{SY}),\ (S^n(M,L),X^n)\in\mathcal{T}_{\mu/2}^n(P_{SX})\right] \\
&\overset{(d)}{\leq} \Pr\left[H_{\text{tp}(S^n(1,1),Y^n)}(S|Y) \geq \min_{\tilde{\ell}\geq 1} H_{\text{tp}(S^n(2,\tilde{\ell}),Y^n)}(S|Y) \;\middle|\; (S^n(1,1),Y^n)\in\mathcal{T}_{\mu}^n(P_{SY}), (S^n(1,1),X^n)\in\mathcal{T}_{\mu/2}^n(P_{SX})\right] \\
&= \sum_{\pi_{SS'Y}\in\mathcal{P}_\mu^n} \sum_{\tilde{\ell}=1}^{\lfloor 2^{nR'}\rfloor} \sum_{\substack{s^n,y^n,s'^n:\\ \text{tp}(s^n,y^n,s'^n)=\pi_{SYS'}}} \Pr\left[S^n(1,1)=s^n, Y^n=y^n, S^n(2,\tilde{\ell})=s'^n \;\middle|\right. \\
&\qquad\qquad\qquad\qquad\qquad \left.(S^n(1,1),Y^n)\in\mathcal{T}_{\mu}^n(P_{SY}, (S^n(1,1),X^n)\in\mathcal{T}_{\mu/2}^n(P_{SX}))\right] \\
&\overset{(e)}{\leq} \sum_{\pi_{SS'Y}\in\mathcal{P}_\mu^n} \sum_{\tilde{\ell}=1}^{\lfloor 2^{nR'}\rfloor} \sum_{\substack{s^n,y^n,s'^n:\\ \text{tp}(s^n,y^n,s'^n)=\pi_{SS'Y}}} \Pr\left[S^n(1,1)=s^n, Y^n=y^n \;\middle|\; (S^n(1,1),Y^n)\in\mathcal{T}_{\mu}^n(P_{SY}), (S^n(1,1),X^n)\in\mathcal{T}_{\mu/2}^n(P_{SX})\right] \\
&\qquad\qquad\qquad\qquad \cdot \Pr[S^n(2,\tilde{\ell})=s'^n] \\
&\overset{(f1)}{\leq} \sum_{\pi_{SS'Y}\in\mathcal{P}_\mu^n} \sum_{\tilde{\ell}=1}^{\lfloor 2^{nR'}\rfloor} \sum_{\substack{s^n,y^n,s'^n:\\ \text{tp}(s^n,y^n,s'^n)=\pi_{SS'Y}}} 2^{-nH_\pi(S,Y)}\cdot 2^{-nH_\pi(S')} \\
&\overset{(f2)}{\leq} \sum_{\pi_{SS'Y}\in\mathcal{P}_\mu^n} \sum_{\tilde{\ell}=1}^{\lfloor 2^{nR'}\rfloor} 2^{nH_\pi(S,S',Y)}\cdot 2^{-nH_\pi(S,Y)}\cdot 2^{-nH_\pi(S')} \\
&= \sum_{\pi_{SS'Y}\in\mathcal{P}_\mu^n} 2^{n(R'-I_\pi(S';Y,S))} \\
&\leq \sum_{\pi_{SS'Y}\in\mathcal{P}_\mu^n} 2^{n(R'-I_\pi(S';Y))} \\
&\overset{(g)}{\leq} (n+1)^{|\mathcal{S}|^2\cdot|\mathcal{Y}|}\cdot \max_{\pi_{SS'Y}\in\mathcal{P}_\mu^{(n)}} 2^{n(R'-I(S;Y)+\delta_n(\mu))} \\
&\overset{(h)}{\leq} \epsilon/3,
\end{aligned} \tag{203}
$$

where $\delta_n(\mu)$ is a function that tends to 0 as $\mu\to 0$ and $n\to\infty$. The inequalities are justified as follows:

- (d): (Our proof is inspired by a proof on the Wyner-Ziv source coding result in [16, Appendix to Chapter 11].) Inequality (d) follows from the following sequence of equalities and from the law of total probability applied over the pair (M,L). For any $(m,\ell)\in\{1,\ldots,\lfloor 2^{nR}\rfloor\}\times\{1,\ldots,\lfloor 2^{nR'}\rfloor\}$:

$$
\begin{aligned}
&\Pr\left[H_{\text{tp}(S^n(M,L),Y^n)}(S|Y) \geq \min_{\tilde{\ell}\neq L} H_{\text{tp}(S^n(M,\tilde{\ell}),Y^n)}(S|Y) \;\middle|\right. \\
&\qquad\qquad \left.(S^n(M,L),Y^n)\in\mathcal{T}_{\mu}^n(P_{SY}), (S^n(M,L),X^n)\in\mathcal{T}_{\mu/2}^n(P_{SX}), M=m, L=\ell\right]
\end{aligned}
$$

[2]For ease of exposition we slightly abuse notation: by $\pi_{SS'Y}\in\mathcal{P}_\mu^{(n)}$ we mean that $\pi_{SS'Y}$ takes value in the projection of $\mathcal{P}_\mu^{(n)}$ onto $\mathcal{S}^n\times\mathcal{S}^n\times\mathcal{Y}^n$.

$$\overset{(i)}{=} \Pr\Big[H_{\text{tp}(S^n(M,L),Y^n)}(S|Y) \geq \min_{\tilde{\ell}\neq L} H_{\text{tp}(S^n(2,\tilde{\ell}),Y^n)}(S|Y) \ \Big| \ (S^n(M,L),Y^n) \in \mathcal{T}_\mu^n(P_{SY}), (S^n(M,L),X^n) \in \mathcal{T}_{\mu/2}^n(P_{SX}), M=m, L=\ell\Big]$$
$$\overset{(ii)}{=} \Pr\Big[H_{\text{tp}(S^n(1,1),Y^n)}(S|Y) \geq \min_{\tilde{\ell}\neq L} H_{\text{tp}(S^n(2,\tilde{\ell}),Y^n)}(S|Y) \ \Big| \ (S^n(1,1),Y^n) \in \mathcal{T}_\mu^n(P_{SY}), (S^n(1,1),X^n) \in \mathcal{T}_{\mu/2}^n(P_{SX}), M=m, L=\ell\Big]$$
$$\overset{(iii)}{\leq} \Pr\Big[H_{\text{tp}(S^n(1,1),Y^n)}(S|Y) \geq \min_{\tilde{\ell}} H_{\text{tp}(S^n(2,\tilde{\ell}),Y^n)}(S|Y) \ \Big| \ (S^n(1,1),Y^n) \in \mathcal{T}_\mu^n(P_{SY}), (S^n(1,1),X^n) \in \mathcal{T}_{\mu/2}^n(P_{SX}), M=m, L=\ell\Big], \tag{204}$$

where (i) holds because the picked codeword $S^n(M,L)$ has the same joint pmf with any set of J other codewords in the codebook; (ii) holds because the transmitter uniformly picks one of the pairs of indices (m,l) that satisfy the typicality check (124); (iii) holds because I adding an element to a minimization can only decrease this minimization.

- (e): This inequality holds because the codebook's codewords are drawn independently of each other.
- $(f1)$ and $(f2)$: These two inequalities hold by standard arguments on types.
- (g): This inequality holds by the polynomial number of types, because $H_\pi(S'|Y) \leq H_\pi(S|Y)$, and by (198) and continuity of the entropy function.
- (h): holds for sufficiently small $\mu > 0$ so that $R' - I(S;Y) + \delta_n(\mu) < 0$ and sufficiently large n.

We now bound the probability of type-II error at the receiver.

$$\beta_n \leq \Pr[(X^n,Y^n) \in \mathcal{A}_{\text{Rx},n} | \mathcal{H}=1], \tag{205}$$

where $\mathcal{A}_{\text{Rx},n}$ includes the receiver's acceptance region:

$$\mathcal{A}_{\text{Rx},n} \triangleq \bigcup_m \bigcup_{\ell,\ell'} \{(x^n,y^n) : \text{tp}(s^n(m,\ell), s^n(m,\ell'), x^n, y^n) \in \mathcal{P}_\mu^n\}. \tag{206}$$

We thus have

$$\begin{aligned}\beta_n &\leq \sum_{\pi_{SS'XY}\in\mathcal{P}_\mu^n} \sum_m \sum_{\substack{(\ell,\ell'):\\ \text{tp}(s^n(m,\ell),s^n(m,\ell'))=\pi_{SS'}}} \sum_{\substack{(x^n,y^n):\\ \text{tp}(s^n(m,\ell),s^n(m,\ell'),x^n,y^n)=\pi_{SS'XY}}} \Pr[X^n=x^n, Y^n=y^n|\mathcal{H}=1] \\ &= \sum_{\pi_{SXY}\in\mathcal{P}_\mu^n} \sum_m \sum_{\substack{\ell\\ \text{tp}(s^n(m,\ell))=\pi_S}} \sum_{\substack{(x^n,y^n):\\ \text{tp}(s^n(m,\ell),x^n,y^n)=\pi_{SXY}}} \Pr[X^n=x^n, Y^n=y^n|\mathcal{H}=1] \\ &+ \sum_{\pi_{SS'XY}\in\mathcal{P}_\mu^n} \sum_m \sum_{\substack{(\ell,\ell'),\ell\neq\ell':\\ \text{tp}(s^n(m,\ell),s^n(m,\ell'))=\pi_{SS'}}} \sum_{\substack{(x^n,y^n):\\ \text{tp}(s^n(m,\ell),s^n(m,\ell'),x^n,y^n)=\pi_{SS'XY}}} \Pr[X^n=x^n, Y^n=y^n|\mathcal{H}=1].\end{aligned} \tag{207}$$

By standard arguments on types and by constraint (121), we get the following upper bounds:

$$\begin{aligned}&\sum_{\pi_{SXY}\in\mathcal{P}_\mu^n} \sum_m \sum_{\substack{\ell\\ \text{tp}(s^n(m,\ell))=\pi_S}} \sum_{\substack{(x^n,y^n):\\ \text{tp}(s^n(m,\ell),x^n,y^n)=\pi_{SXY}}} \Pr[X^n=x^n, Y^n=y^n|\mathcal{H}=1] \\ &\leq (n+1)^{|\mathcal{S}|.|\mathcal{X}|.|\mathcal{Y}|} \cdot 2^{n(I(S;X)+\mu)} \cdot 2^{nH_\pi(X,Y|S)} \cdot 2^{-n(H_\pi(X,Y)+D(\pi_{XY}\|Q_{XY}))},\end{aligned} \tag{208}$$

and similarly

$$\begin{aligned}&\sum_{\pi_{SS'XY}\in\mathcal{P}_\mu^n} \sum_m \sum_{\substack{(\ell,\ell'),\ell\neq\ell':\\ \text{tp}(s^n(m,\ell),s^n(m,\ell'))=\pi_{SS'}}} \sum_{\substack{(x^n,y^n):\\ \text{tp}(s^n(m,\ell),s^n(m,\ell'),x^n,y^n)=\pi_{SS'XY}}} \Pr[X^n=x^n, Y^n=y^n|\mathcal{H}=1] \\ &\leq \sum_{\pi_{SS'XY}\in\mathcal{P}_\mu^n} \sum_m \sum_{\substack{(\ell,\ell'),\ell\neq\ell':\\ \text{tp}(s^n(m,\ell),s^n(m,\ell'))=\pi_{SS'}}} 2^{nH_\pi(X,Y|S,S')} \cdot 2^{-n(H_\pi(X,Y)+D(\pi_{XY}\|Q_{XY}))} \\ &\leq (n+1)^{|\mathcal{S}|^2.|\mathcal{X}|.|\mathcal{Y}|} \cdot 2^{nR} \cdot 2^{n(2R'-I_\pi(S;S'))} \cdot 2^{nH_\pi(X,Y|S,S')} \cdot 2^{-n(H_\pi(X,Y)+D(\pi_{XY}\|Q_{XY}))},\end{aligned} \tag{209}$$

where we used that for any joint type $\pi_{SS'}$:

$$\mathbb{E}_\mathcal{C}[|(\ell,\ell') : \ell\neq\ell', \text{tp}(S^n(m,\ell), S^n(m,\ell')) = \pi_{SS'}|] = \sum_{\substack{\ell,\ell'\\ \ell\neq\ell'}} \sum_{\substack{(s^n,s'^n):\\ \text{tp}(s^n,s'^n)=\pi_{SS'}}} \Pr[(S^n(m,\ell), S^n(m,\ell')) = (s^n, s'^n)]$$

$$\leq \sum_{\substack{\ell,\ell' \\ \ell\neq\ell'}} \sum_{\substack{(s^n,s'^n): \\ \mathrm{tp}(s^n,s'^n)=\pi_{SS'}}} 2^{-n(H_\pi(S)+D(\pi_S\|P_S))} \cdot 2^{-n(H_\pi(S')+D(\pi_{S'}\|P_S))}$$
$$\leq \sum_{\substack{\ell,\ell' \\ \ell\neq\ell'}} 2^{nH_\pi(S,S')} \cdot 2^{-n(H_\pi(S)+H_\pi(S'))}$$
$$\leq 2^{2nR'} \cdot 2^{-nI_\pi(S;S')}. \tag{210}$$

Define now

$$\theta_{1,\mu} \triangleq \min_{\pi_{SXY}\in\mathcal{P}_\mu^n} [H_\pi(X,Y) + D(\pi_{XY}\|Q_{XY}) - H_\pi(X,Y|S) - I(S;X)], \tag{211}$$
$$\theta_{2,\mu} \triangleq \min_{\pi_{SS'XY}\in\mathcal{P}_\mu^n} H_\pi(X,Y) + D(\pi_{XY}\|Q_{XY}) - H_\pi(X,Y|S,S') - R - 2R' + I_\pi(S;S'), \tag{212}$$

and note that when $\mu \to 0$ and thus $n \to \infty$, then $\mathcal{P}_\mu^n \to \mathcal{P}^*$, where $\mathcal{P}^*$ is defined in (199). Thus, letting $\mu \to 0$ and thus $n \to \infty$, by (207), (211), and (212) the error exponent achieved by the described scheme exceeds the minimum

$$\min\{\tilde{\theta}_1, \tilde{\theta}_2\}, \tag{213}$$

where

$$\tilde{\theta}_1 \triangleq \min_{\substack{\tilde{P}_{SXY}: \\ \tilde{P}_{SX}=P_{SX} \\ \tilde{P}_{SY}=P_{SY}}} H_{\tilde{P}}(X,Y) + D(\tilde{P}_{XY}\|Q_{XY}) - H_{\tilde{P}}(X,Y|S) - I(S;X)$$
$$= \min_{\substack{\tilde{P}_{SXY}: \\ \tilde{P}_{SX}=P_{SX} \\ \tilde{P}_{SY}=P_{SY}}} D(\tilde{P}_{SXY}\|P_{S|X}Q_{XY}) \tag{214}$$

and

$$\tilde{\theta}_2 \triangleq \min_{\substack{\tilde{P}_{SS'XY}: \\ \tilde{P}_{SX}=P_{SX} \\ \tilde{P}_{S'Y}=P_{SY} \\ H_{SY}(S|Y)\leq H_{\tilde{P}_{SY}}(S|Y)}} D(\tilde{P}_{SXY}\|P_{S|X}Q_{XY}) + R - I(S;X) + I_{\tilde{P}}(S';X,Y,S)$$
$$= \min_{\substack{\tilde{P}_{SXY}: \\ \tilde{P}_{SX}=P_{SX} \\ \tilde{P}_{Y}=P_{Y} \\ H(S|Y)\leq H_{\tilde{P}_{SY}}(S|Y)}} D(\tilde{P}_{SXY}\|P_{S|X}Q_{XY}) + R - I(S;X) + I(S;Y), \tag{215}$$

where the last equality holds because $I_{\tilde{P}}(S';X,Y,S) \geq I_{\tilde{P}}(S';Y) = I(S;Y)$ and because this inequality holds with equality when $\tilde{P}_{S'|XYX} = \tilde{P}_{S'|Y} = P_{S|Y}$, which represents a valid choice.

This concludes the proof.

Appendix I
Proof of Theorem 5

We analyze the probabilities of error of the coding and testing scheme described in Subsection VI-B1 averaged over the random code constructions. By standard arguments (successively eliminating the worst half of the codewords from the codebooks as detailed out in Appendix A) the desired result can be proved for a set of deterministic codebooks.

Fix an arbitrary $\epsilon > 0$ and the parameter of the scheme μ sufficiently close to 0 as will become clear in the sequel. Fix also a blocklength n.

If $M \neq 0$, let I, J be the indices sent over the noise-free bit pipe from the transmitter to the relay. If both $B \neq 0$ and $M \neq 0$, let K denote the second index sent over the noise-free bit-pipe from the relay to the receiver.

We first analyze the type-I error probability at the receiver. Define events:

$$\mathcal{E}_{\text{Rx}}^{(1)}\colon \big\{(X^n(I,D), V^n(K,F|I,D), Z^n) \notin \mathcal{T}_\mu^n(P_{SVZ})\big\}, \tag{216}$$
$$\mathcal{E}_{\text{Rx}}^{(2)}\colon \big\{\exists (d'',f') \neq (D,F)\colon H_{\mathrm{tp}(S^n(I,d''),V^n(K,f'|I,d),Z^n)}(S,V|Z) \leq \min_{(\tilde{d},\tilde{f})} H_{\mathrm{tp}(S^n(I,\tilde{d}),V^n(K,\tilde{f}|I,\tilde{d}),Z^n)}(S,V|Z)\big\} \tag{217}$$

The type-I error probability can then be bounded as follows:

$$\alpha_n \leq \Pr[M = 0 \text{ or } B = 0 \text{ or } \mathcal{E}_{\text{Rx}}^{(1)} \text{ or } \mathcal{E}_{\text{Rx}}^{(2)} \text{ or } \mathcal{E}_{\text{Rx}}^{(3)}]$$

$$
\begin{aligned}
&\leq \Pr[M=0] + \Pr[B=0|M\neq 0] + \Pr\left[\mathcal{E}_{\text{Rx}}^{(1)}\middle|M\neq 0, B\neq 0\right] + \Pr\left[\mathcal{E}_{\text{Rx}}^{(2)}\middle|M\neq 0, B\neq 0, \mathcal{E}_{\text{Rx}}^{(1)c}\right] \\
&\overset{(a)}{\leq} \epsilon/4 + Pr[B=0|M\neq 0] + \Pr\left[\mathcal{E}_{\text{Rx}}^{(1)}\middle|M\neq 0, B\neq 0\right] + \Pr\left[\mathcal{E}_{\text{Rx}}^{(2)}\middle|M\neq 0, B\neq 0, \mathcal{E}_{\text{Rx}}^{(1)c}\right] \\
&\overset{(b)}{\leq} \epsilon/2 + \Pr\left[\mathcal{E}_{\text{Rx}}^{(1)}\middle|M\neq 0, B\neq 0\right] + \Pr\left[\mathcal{E}_{\text{Rx}}^{(2)}\middle|M\neq 0, B\neq 0, \mathcal{E}_{\text{Rx}}^{(1)c}\right] \\
&\overset{(c)}{\leq} 3\epsilon/4 + \Pr\left[\mathcal{E}_{\text{Rx}}^{(2)}\middle|M\neq 0, B\neq 0, \mathcal{E}_{\text{Rx}}^{(1)c}\right] \\
&\overset{(d)}{\leq} \epsilon.
\end{aligned}
\tag{218}
$$

where (a) holds by the covering lemma and the rate-constraints in (127); (b) and (d) can be proved following similar lines as the type-I error analysis in Appendix H; and (c) follows from the Markov lemma.

We now bound the probability of type-II error at the receiver. Let $\mathcal{P}^n$ be the set of all types over the product alphabets $\mathcal{S}^n \times \mathcal{S}^n \times \mathcal{S}^n \times \mathcal{U}^n \times \mathcal{U}^n \times \mathcal{V}^n \times \mathcal{V}^n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$. Also, let $\mathcal{P}_\mu^n$ be the subset of types $\pi_{SS'S''UU'VV'XYZ} \in \mathcal{P}^n$ that simultaneously satisfy the following three conditions:

$$
\begin{aligned}
\forall (s,s',s'',u,u',v,v',x,y,z) \in \mathcal{S}\times\mathcal{S}\times\mathcal{S}\times\mathcal{U}\times\mathcal{U}\times\mathcal{V}\times\mathcal{V}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}: \\
|\pi_{SUX}(s,u,x) - P_{SUX}(s,u,x)| \leq \mu/4, \\
|\pi_{S'U'VY}(s',u',v,y) - P_{SUVY}(s',u',v,y)| \leq \mu/2, \\
|\pi_{S''V'Z}(s'',v',z) - P_{SVZ}(s'',v',z)| \leq \mu,
\end{aligned}
\tag{219}
$$

and the following minimal entropy constraints

$$
H_{\pi_{S'U'Y}}(S,U|Y) \leq H_{\pi_{SUY}}(S,U|Y), \tag{220}
$$

$$
H_{\pi_{S''V'Z}}(V,S|Z) \leq H_{\pi_{SVZ}}(V,S|Z). \tag{221}
$$

The type-II error probability at the receiver can be bounded as:

$$
\beta_n \leq \Pr[(X^n, Y^n, Z^n) \in \mathcal{A}_{\text{Rx},n} | \mathcal{H}=1], \tag{222}
$$

where $\mathcal{A}_{\text{Rx},n}$ includes the acceptance region at the receiver:

$$
\mathcal{A}_{\text{Rx},n} \triangleq \bigcup_{i,j,k} \bigcup_{d,d',d'',e,e',f,f'} \mathcal{S}^{(n)}(i,j,k,d,d',d'',e,e',f,f') \tag{223}
$$

where $\mathcal{S}^{(n)}(i,j,k,d,d',d'',e,e',f,f') \subseteq \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$ is defined as:

$$
\begin{aligned}
&\mathcal{S}^{(n)}(i,j,k,d,d',d'',e,e',f,f') := \\
&\left\{(x^n,y^n,z^n) : \text{tp}(s^n(i,d), s^n(i,d'), s^n(i,d''), u^n(j,e|i,d), u^n(j,e'|i,d'), v^n(k,f|i,d'), v^n(k,f'|i,d''), x^n, y^n, z^n) \in \mathcal{P}_\mu^n\right\}
\end{aligned}
\tag{224}
$$

We thus have

$$
\beta_n \leq \sum_{(d,d',d'',e,e',f,f')} \sum_{(i,j,k)} \Pr\big[(X^n,Y^n,Z^n) \in \mathcal{S}^{(n)}(i,j,k,d,d',d'',e,e',f,f')\big]. \tag{225}
$$

We split up the outermost summation into 20 partial summations as indicated in Table I, and we analyze the exponents for each of these partial sums. The exponent of the total sum is then given by the smallest exponent corresponding to any of the partial sums.

Each of the partial sums is analyzed by now standard arguments as used in Appendices B and H. Remark that for any partial sum corresponding to an index tuple with

$$
d = d' = d'',
$$

one is only interested in the subset of $\mathcal{P}_\mu^n$ where the subtype $\pi_{SS'S''}$ corresponds to $S = S' = S''$. Similarly, for any partial sum corresponding to $d = d'$ and $e = e'$ one is only interested in the subset of $\mathcal{P}_\mu^n$ where the subtype $\pi_{SS'UU'}$ corresponds to $S = S'$ and $U = U'$. See again Table I for an illustration which types have to be considered for which partial sums.

We detail out the derivation of the upper bound on the exponent of partial sum 4; the other partial sums can be treated similarly. Let $\mathcal{P}^{4,n}$ be the subset of types obtained from $\mathcal{P}^n$ by imposing the additional constraints $S = S' = S''$ and $V = V'$:

$$
\begin{aligned}
\mathcal{P}^{4,n} := \Bigg\{ \pi_{SUU'VXYZ}: \quad & \forall (s,u,u',v,x,y,z) \in \mathcal{S}\times\mathcal{U}\times\mathcal{U}\times\mathcal{V}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}: \\
& |\pi_{SUX}(s,u,x) - P_{SUX}(s,u,x)| \leq \mu/4, \\
& |\pi_{SU'VY}(s,u',v,y) - P_{SUVY}(s,u',v,y)| \leq \mu/2, \\
& |\pi_{SVZ}(s,v,z) - P_{SVZ}(s,v,z)| \leq \mu,
\end{aligned}
\tag{226}
$$

$$H_{\pi_{SU'Y}}(U|S,Y) \le H_{\pi_{SUY}}(U|S,Y). \Big\} \tag{227}$$

Then

$$\sum_{\substack{(d,e,e',f):\\ e\neq e'}} \sum_{(i,j,k)} \Pr\big[(X^n,Y^n,Z^n)\in \mathcal{S}^{(n)}(i,j,k,d,d,d,e,e',f,f)\big]$$
$$= \sum_{\pi_{SUU'VXYZ}\in\mathcal{P}_\mu^{4,n}} \sum_{(i,j,k)} \sum_{\substack{(d,e,e',f):\\ e\neq e',\\ \text{tp}(s^n(i,d),u^n(j,e|i,d),u^n(j,e'|i,d),v^n(k,f|i,d))\\ =\pi_{SUU'V}}} \sum_{\substack{(x^n,y^n,z^n):\\ \text{tp}(s^n(i,d),u^n(j,e|i,d),u^n(j,e'|i,d),v^n(k,f|i,d),x^n,y^n,z^n)\\ =\pi_{SUU'VXYZ}}} \Pr\big[(X^n,Y^n,Z^n)=(x^n,y^n,z^n)\big]$$
$$= \sum_{\pi_{SUU'VXYZ}\in\mathcal{P}_\mu^{4,n}} \sum_{(i,j,k)} \sum_{\substack{(d,e,e',f):\\ e\neq e'\\ \text{tp}(s^n(i,d),u^n(j,e|i,d),u^n(j,e'|i,d),v^n(k,f|i,d))=\pi_{SUU'V}}} 2^{nH_\pi(X,Y,Z|S,U,U',V)}\cdot 2^{-n(H_\pi(X,Y,Z)+D(\pi_{XYZ}\|Q_{XYZ}))}$$
$$\le (n+1)^{|\mathcal{S}|.|\mathcal{U}|.|\mathcal{V}|.|\mathcal{X}|.|\mathcal{Y}|.|\mathcal{Z}|}\cdot 2^{n(R_s+R_u+R_v)}\cdot 2^{n(R_s'+2R_u'+R_v'-I_\pi(U';V|S)-I_\pi(U;U',V|S))}\cdot 2^{nH_\pi(X,Y,Z|S,U,U',V)}$$
$$\cdot 2^{-n(H_\pi(X,Y,Z)+D(\pi_{XYZ}\|Q_{XYZ}))}, \tag{228}$$

where the last step follows by bounding the number of types and because for any joint type $\pi_{SUU'V}$:

$$\Big|\Big\{(d,e,e',f):\quad e\neq e' \quad \text{and} \quad \text{tp}(s^n(i,d),u^n(j,e|i,d),u^n(j,e'|i,d),v^n(k,f|i,d))=\pi_{SUU'V}\Big\}\Big|$$
$$\le 2^{n(R_s'+2R_u'+R_v'-I_\pi(U';V|S)-I_\pi(U;U',V|S)+\mu)}, \tag{229}$$

Following similar lines, one can find upper bounds on the exponents of the other partial sums. Notice that due to the superposition code construction, different partial sums yield the same exponent. Table I presents the error exponent that we found on each partial sum. The exponents $\theta_{1,\mu},\ldots,\theta_{10,\mu}$ are defined as follows, where the minimizations are over appropriate subsets of $\mathcal{P}_\mu^n$:

$$\theta_{1,\mu} \triangleq \min_{\pi_{SUV}} -R_s-R_u-R_v-R_s'-R_u'-R_v'+I_\pi(U;V|S)$$
$$-H_\pi(X,Y,Z|S,U,V)+H_\pi(X,Y,Z)+D(\pi_{XYZ}\|Q_{XYZ}), \tag{230}$$
$$\theta_{2,\mu} \triangleq \min_{\pi_{SUVV'}} -R_s-R_u-R_v-R_s'-R_u'-2R_v'+I_\pi(U;V,V'|S)+I_\pi(V;V'|S)$$
$$-H_\pi(X,Y,Z|S,U,V,V')+H_\pi(X,Y,Z)+D(\pi_{XYZ}\|Q_{XYZ}), \tag{231}$$
$$\theta_{3,\mu} \triangleq \min_{\pi_{SUU'VV'}} -R_s-R_u-R_v-R_s'-2R_u'-2R_v'+I_\pi(U;V,V'|S)+I_\pi(V;V'|S)$$
$$+I_\pi(U';U,V,V'|S)-H_\pi(X,Y,Z|S,U,U',V,V')+H_\pi(X,Y,Z)+D(\pi_{XYZ}\|Q_{XYZ}), \tag{232}$$
$$\theta_{4,\mu} \triangleq \min_{\pi_{SUU'V}} -R_s-R_u-R_v-R_s'-2R_u'-R_v'+I_\pi(U;V,U'|S)+I_\pi(V;U'|S)$$
$$-H_\pi(X,Y,Z|S,U,U',V)+H_\pi(X,Y,Z)+D(\pi_{XYZ}\|Q_{XYZ}), \tag{233}$$
$$\theta_{5,\mu} \triangleq \min_{\pi_{SS'S''UU'VV'}} -R_s-R_u-R_v-3R_s'-2R_u'-2R_v'+I_\pi(S,U;S',U')$$
$$+I_\pi(S'',V';S,S',U,U')+I_\pi(V;S,S'',U,U'|S')-H_\pi(X,Y,Z|S,S',S'',U,U',V,V')$$
$$+H_\pi(X,Y,Z)+D(\pi_{XYZ}\|Q_{XYZ}), \tag{234}$$
$$\theta_{6,\mu} \triangleq \min_{\pi_{SS'UU'VV'}} -R_s-R_u-R_v-2R_s'-2R_u'-2R_v'+I_\pi(S,U;S',U')+I_\pi(V;V'|S')+I_\pi(V,V';S,U,U'|S')$$
$$-H_\pi(X,Y,Z|S,S',U,U',V,V')+H_\pi(X,Y,Z)+D(\pi_{XYZ}\|Q_{XYZ}), \tag{235}$$
$$\theta_{7,\mu} \triangleq \min_{\pi_{SS'UU'V}} -R_s-R_u-R_v-2R_s'-2R_u'-R_v'+I_\pi(S,U;S',U')+I_\pi(V;S,U,U'|S')$$
$$-H_\pi(X,Y,Z|S,S',U,U',V)+H_\pi(X,Y,Z)+D(\pi_{XYZ}\|Q_{XYZ}), \tag{236}$$
$$\theta_{8,\mu} \triangleq \min_{\pi_{SS'UU'VV'}} -R_s-R_u-R_v-2R_s'-2R_u'-2R_v'+I_\pi(S,U;S',U')+I_\pi(V;S,U,U'|S')$$

Partial sum	Assumption on index tuples	Relevant type	Corresponding error exponent
1	$d=d'=d''$ and $e=e'$ and $f=f'$	π_{SUV} ($S=S'=S''$ and $U=U'$ and $V=V'$)	$\theta_{1,\mu}$
2	$d=d'=d''$ and $e=e'$ and $f\neq f'$	$\pi_{SUVV'}$ ($S=S'=S''$ and $U=U'$)	$\theta_{2,\mu}$
3	$d=d'=d''$ and $e\neq e'$ and $f\neq f'$	$\pi_{SUU'VV'}$ ($S=S'=S''$)	$\theta_{3,\mu}$
4	$d=d'=d''$ and $e\neq e'$ and $f=f'$	$\pi_{SUU'V}$ ($S=S'=S''$ and $V=V'$)	$\theta_{4,\mu}$
5	$d\neq d'\neq d''$ and $e\neq e'$ and $f\neq f'$	$\pi_{SS'S''UU'VV'}$	$\theta_{5,\mu}$
6	$d\neq d'\neq d''$ and $e=e'$ and $f\neq f'$	$\pi_{SS'S''UVV'}$ ($U=U'$)	$\theta_{5,\mu}$
7	$d\neq d'\neq d''$ and $e=e'$ and $f=f'$	$\pi_{SS'S''UV}$ ($U=U'$ and $V=V'$)	$\theta_{5,\mu}$
8	$d\neq d'\neq d''$ and $e\neq e'$ and $f=f'$	$\pi_{SS'S''UU'V}$ ($V=V'$)	$\theta_{5,\mu}$
9	$d'=d''\neq d$ and $e\neq e'$ and $f\neq f'$	$\pi_{SS'UU'VV'}$ ($S'=S''$)	$\theta_{6,\mu}$
10	$d'=d''\neq d$ and $e=e'$ and $f\neq f'$	$\pi_{SS'UVV'}$ ($S'=S''$ and $U=U'$)	$\theta_{6,\mu}$
11	$d'=d''\neq d$ and $e\neq e'$ and $f=f'$	$\pi_{SS'UU'V}$ ($S'=S''$ and $V=V'$)	$\theta_{7,\mu}$
12	$d'=d''\neq d$ and $e=e'$ and $f=f'$	$\pi_{SS'UV}$ ($S'=S''$ and $U=U'$ and $V=V'$)	$\theta_{7,\mu}$
13	$d=d''\neq d'$ and $e\neq e'$ and $f\neq f'$	$\pi_{SS'UU'VV'}$ ($S=S''$)	$\theta_{8,\mu}$
14	$d=d''\neq d'$ and $e=e'$ and $f\neq f'$	$\pi_{SS'UVV'}$ ($S=S''$ and $U=U'$)	$\theta_{8,\mu}$
15	$d=d''\neq d'$ and $e=e'$ and $f=f'$	$\pi_{SS'UV}$ ($S=S''$ and $U=U'$ and $V=V'$)	$\theta_{8,\mu}$
16	$d=d''\neq d'$ and $e\neq e'$ and $f=f'$	$\pi_{SS'UU'V}$ ($S=S''$ and $V=V'$)	$\theta_{8,\mu}$
17	$d=d'\neq d''$ and $e=e'$ and $f\neq f'$	$\pi_{SS''UVV'}$ ($S=S'$ and $U=U'$)	$\theta_{9,\mu}$
18	$d=d'\neq d''$ and $e=e'$ and $f=f'$	$\pi_{SS''UV}$ ($S=S'$ and $U=U'$ and $V=V'$)	$\theta_{9,\mu}$
19	$d=d'\neq d''$ and $e\neq e'$ and $f\neq f'$	$\pi_{SS''UU'VV'}$ ($S=S'$)	$\theta_{10,\mu}$
20	$d=d'\neq d''$ and $e\neq e'$ and $f=f'$	$\pi_{SS''UU'V}$ ($S=S'$ and $V=V'$)	$\theta_{10,\mu}$

TABLE I
CORRESPONDENCE OF INDEX TUPLES AND POSSIBLE TYPES.

$$+ I_\pi(V';S',V,U,U'|S) - H_\pi(X,Y,Z|S,S',U,U',V,V') + H_\pi(X,Y,Z) + D(\pi_{XYZ}\|Q_{XYZ}), \quad (237)$$

$$\theta_{9,\mu} \triangleq \min_{\pi_{SS''UVV'}} -R_s - R_u - R_v - 2R'_s - R'_u - 2R'_v + I_\pi(S,V;S'',V') + I_\pi(U;S'',V,V'|S)$$
$$- H_\pi(X,Y,Z|S,S'',U,V,V') + H_\pi(X,Y,Z) + D(\pi_{XYZ}\|Q_{XYZ}), \quad (238)$$

$$\theta_{10,\mu} \triangleq \min_{\pi_{SS''UU'VV'}} -R_s - R_u - R_v - 2R'_s - 2R'_u - 2R'_v + I_\pi(S,U;S'',V') + I_\pi(U';S'',U,V,V'|S)$$
$$+ I_\pi(V;S'',U,V'|S) - H_\pi(X,Y,Z|S,S'',U,U',V,V') + H_\pi(X,Y,Z) + D(\pi_{XYZ}\|Q_{XYZ}), \quad (239)$$

Reassembling the upper bounds on the 20 partial sums, yields

$$\beta_n \leq (n+1)^{|\mathcal{S}|.|\mathcal{U}|.|\mathcal{V}|.|\mathcal{X}|.|\mathcal{Y}|}\cdot 2^{-n\theta_{1,\mu}} + (n+1)^{|\mathcal{S}|.|\mathcal{U}|.|\mathcal{V}|^2.|\mathcal{X}|.|\mathcal{Y}|}\cdot 2^{-n\theta_{2,\mu}} + (n+1)^{|\mathcal{S}|.|\mathcal{U}|^2.|\mathcal{V}|^2.|\mathcal{X}|.|\mathcal{Y}|}\cdot 2^{-n\theta_{3,\mu}}$$
$$+ (n+1)^{|\mathcal{S}|.|\mathcal{U}|^2.|\mathcal{V}|.|\mathcal{X}|.|\mathcal{Y}|}\cdot 2^{-n\theta_{4,\mu}} + (n+1)^{|\mathcal{S}|^3.|\mathcal{U}|^2.|\mathcal{V}|^2.|\mathcal{X}|.|\mathcal{Y}|}\cdot 2^{-n\theta_{5,\mu}} + (n+1)^{|\mathcal{S}|^2.|\mathcal{U}|^2.|\mathcal{V}|^2.|\mathcal{X}|.|\mathcal{Y}|}\cdot 2^{-n\theta_{6,\mu}}$$
$$+ (n+1)^{|\mathcal{S}|^2.|\mathcal{U}|^2.|\mathcal{V}|.|\mathcal{X}|.|\mathcal{Y}|}\cdot 2^{-n\theta_{7,\mu}} + (n+1)^{|\mathcal{S}|^2.|\mathcal{U}|^2.|\mathcal{V}|^2.|\mathcal{X}|.|\mathcal{Y}|}\cdot 2^{-n\theta_{8,\mu}} + (n+1)^{|\mathcal{S}|^2.|\mathcal{U}|.|\mathcal{V}|^2.|\mathcal{X}|.|\mathcal{Y}|}\cdot 2^{-n\theta_{9,\mu}}$$
$$+ (n+1)^{|\mathcal{S}|^2.|\mathcal{U}|^2.|\mathcal{V}|^2.|\mathcal{X}|.|\mathcal{Y}|}\cdot 2^{-n\theta_{10,\mu}}.$$

Taking $\mu \to 0$ and $n \to \infty$, and eliminating some auxiliary random variables in the minimizations at the exponents, it can be shown that for any $i \in \{1,\dots,10\}$ the error exponent $\theta_{i,\mu} \to \theta_i$, where $\theta_1,\dots,\theta_{10}$ are defined in (139). More specifically, notice for example that $\theta_{4,\mu} \to \tilde{\theta}_4$ as $\mu \to 0$ and $n \to \infty$, where

$$\tilde{\theta}_4 := \min_{\substack{\tilde{P}_{SUVXYZ}: \\ \tilde{P}_{SUX}=P_{SUX} \\ \tilde{P}_{SU'VY}=P_{SUVY} \\ \tilde{P}_{SVZ}=P_{SVZ} \\ H_{\tilde{P}}(U'|S,Y)\leq H_{\tilde{P}}(U|S,Y)}} -R_s - R_u - R_v - R_s' - 2R_u' - R_v' + I_{\tilde{P}}(U;V,U'|S) + I_{\tilde{P}}(V;U'|S) - H_{\tilde{P}}(X,Y,Z|S,U,U',V) + H_{\tilde{P}}(X,Y,Z) + D(\tilde{P}_{XYZ}\|Q_{XYZ}) \tag{240}$$

Moreover,

$$\begin{aligned}
\tilde{\theta}_4 &= \min_{\substack{\tilde{P}_{SUVXYZ}: \\ \tilde{P}_{SUX}=P_{SUX} \\ \tilde{P}_{SU'VY}=P_{SUVY} \\ \tilde{P}_{SVZ}=P_{SVZ} \\ H_{\tilde{P}}(U'|S,Y)\leq H_{\tilde{P}}(U|S,Y)}} -R_s - R_u - R_v - R_s' - 2R_u' - R_v' + I_{\tilde{P}}(U;V,U'|S) + I_{\tilde{P}}(V;U'|S) \\
&\qquad -H_{\tilde{P}}(X,Y,Z|S,U,U',V) + H_{\tilde{P}}(X,Y,Z) + D(\tilde{P}_{XYZ}\|Q_{XYZ}) \\
&\overset{(a)}{=} \min_{\substack{\tilde{P}_{SUVXYZ}: \\ \tilde{P}_{SUX}=P_{SUX} \\ \tilde{P}_{SU'VY}=P_{SUVY} \\ \tilde{P}_{SVZ}=P_{SVZ} \\ H_{\tilde{P}}(U'|S,Y)\leq H_{\tilde{P}}(U|S,Y)}} -R_u' - I(U,S;X) - I(V;Y,U|S) + I_{\tilde{P}}(U;V,U'|S) + I_{\tilde{P}}(V;U'|S) \\
&\qquad -H_{\tilde{P}}(X,Y,Z|S,U,U',V) + H_{\tilde{P}}(X,Y,Z) + D(\tilde{P}_{XYZ}\|Q_{XYZ}) \\
&= \min_{\substack{\tilde{P}_{SUVXYZ}: \\ \tilde{P}_{SUX}=P_{SUX} \\ \tilde{P}_{SU'VY}=P_{SUVY} \\ \tilde{P}_{SVZ}=P_{SVZ} \\ H_{\tilde{P}}(U'|S,Y)\leq H_{\tilde{P}}(U|S,Y)}} -R_u' - I(U,S;X) - I(V;Y,U|S) + I_{\tilde{P}}(U;V,U'|S) + I_{\tilde{P}}(V;U'|S) \\
&\qquad -H_{\tilde{P}}(X,Y,Z|S,U,V) + I_{\tilde{P}}(X,Y,Z;U'|S,U,V) + H_{\tilde{P}}(X,Y,Z) + D(\tilde{P}_{XYZ}\|Q_{XYZ}) \\
&= \min_{\substack{\tilde{P}_{SUVXYZ}: \\ \tilde{P}_{SUX}=P_{SUX} \\ \tilde{P}_{SU'VY}=P_{SUVY} \\ \tilde{P}_{SVZ}=P_{SVZ} \\ H_{\tilde{P}}(U'|S,Y)\leq H_{\tilde{P}}(U|S,Y)}} -R_u' + D(\tilde{P}_{SUVXYZ}||P_{SU|X}P_{V|SUY}Q_{XYZ}) \\
&\qquad +I_{\tilde{P}}(U;U'|V,S) + I_{\tilde{P}}(V;U'|S) + I_{\tilde{P}}(X,Y,Z;U'|S,U,V) \\
&= \min_{\substack{\tilde{P}_{SUVXYZ}: \\ \tilde{P}_{SUX}=P_{SUX} \\ \tilde{P}_{SU'VY}=P_{SUVY} \\ \tilde{P}_{SVZ}=P_{SVZ} \\ H_{\tilde{P}}(U'|S,Y)\leq H_{\tilde{P}}(U|S,Y)}} -R_u' + D(\tilde{P}_{SUVXYZ}||P_{SU|X}P_{V|SUY}Q_{XYZ}) + I_{\tilde{P}}(X,Y,Z,U,V;U'|S) \\
&\overset{(b)}{=} \min_{\substack{\tilde{P}_{SUVXYZ}: \\ \tilde{P}_{SUX}=P_{SUX} \\ \tilde{P}_{SVY}=P_{SVY} \\ \tilde{P}_{SVZ}=P_{SVZ} \\ H(U|S,Y)\leq H_{\tilde{P}}(U|S,Y)}} -R_u' + D(\tilde{P}_{SUVXYZ}||P_{SU|X}P_{V|SUY}Q_{XYZ}) + I(Y,V;U|S) \\
&\overset{(c)}{=} \min_{\substack{\tilde{P}_{SUVXYZ}: \\ \tilde{P}_{SUX}=P_{SUX} \\ \tilde{P}_{SVY}=P_{SVY} \\ \tilde{P}_{SVZ}=P_{SVZ} \\ H(U|S,Y)\leq H_{\tilde{P}}(U|S,Y)}} R_u - I(U;X|S) + D(\tilde{P}_{SUVXYZ}||P_{SU|X}P_{V|SUY}Q_{XYZ}) + I(Y,V;U|S) \\
&= \theta_4,
\end{aligned} \tag{241}$$

where (a) follows from the rate constraints (127a)–(127c), (b) follows from the fact that $I_{\tilde{P}}(X,Y,Z,U,V;U'|S) \geq I_{\tilde{P}}(Y,V;U'|S)$ with equality when $\tilde{P}_{SUU'VXYZ} = \tilde{P}_{SUVXYZ} \cdot \tilde{P}_{U'|SVY}$, (c) follows from the rate constraint (127b). The other limits can be proved in analogous manners. This concludes the achievability proof for the error exponent achieved at the receiver.

The error exponent at the relay can be derived in a similar way. Details are omitted.

Appendix J
Proof of Converse to Proposition 7

Fix a sequence of encoding and decoding functions $\{\phi^{(n)}, \phi_y^{(n)}, g_y^{(n)}, g_z^{(n)}\}$ so that the inequalities of Definition 1 hold for sufficiently large blocklengths n. Fix also such a sufficiently large n and define for each $t \in \{1,\dots,n\}$:

$$\begin{aligned}
U_t &\triangleq (M, X^{t-1}, W^{t-1}, W_{t+1}^n) \\
V_t &\triangleq (B, \bar{Y}^{t-1}, Q^{t-1}, Q_{t+1}^n, W^{t-1}, W_{t+1}^n).
\end{aligned} \tag{242}$$

Define also $U \triangleq (U_T, T)$; $V \triangleq (V_T, T)$; $X \triangleq X_T$; $Y \triangleq Y_T$; $W \triangleq W_T$; and $Z \triangleq Z_T$; for $T \sim \mathcal{U}\{1, \ldots, n\}$ independent of the tuples $(U^n, V^n, X^n, Y^n, Z^n, W^n)$. Notice the Markov chains $U \to X \to Y$ and $V \to Y \to Z$. First, consider the rate R:

$$\begin{aligned}
R &= \frac{1}{n} H(M) \\
&\geq \frac{1}{n} I(M; X^n | W^n) \\
&= \frac{1}{n} \sum_{t=1}^{n} I(M; X_t | X^{t-1}, W^n) \\
&= \frac{1}{n} \sum_{t=1}^{n} I(M, X^{t-1}, W^{t-1}, W_{t+1}^n; X_t | W_t) \\
&= \frac{1}{n} \sum_{t=1}^{n} I(U_t; X_t | W_t) \\
&= I(U; X | W).
\end{aligned} \tag{243}$$

Similarly, for the rate T:

$$\begin{aligned}
T &= \frac{1}{n} H(B) \\
&\geq \frac{1}{n} I(B; Y^n | W^n, Q^n) \\
&= \frac{1}{n} \sum_{t=1}^{n} I(B; Y_t | Y^{t-1}, W^n, Q^n) \\
&= \frac{1}{n} \sum_{t=1}^{n} I(B, Y^{t-1}, Q^{t-1}, Q_{t+1}^n, W^{t-1}, W_{t+1}^n; Y_t | W_t, Q_t) \\
&= \frac{1}{n} \sum_{t=1}^{n} I(V_t; Y_t | W_t, Q_t) \\
&= I(V; Y | W, Q).
\end{aligned} \tag{244}$$

The type-II error probability at the relay can be bounded as

$$\begin{aligned}
-\frac{1}{n} \log \zeta_n &\leq \frac{1}{n} D(P_{M\bar{Y}^n W^n | \mathcal{H}=0} \| P_{M\bar{Y}^n W^n | \mathcal{H}=1}) + \epsilon \\
&= \frac{1}{n} D(P_{M\bar{Y}^n W^n} \| P_{M|W^n} P_{\bar{Y}^n|W^n} P_{W^n}) + \epsilon \\
&= \frac{1}{n} I(M; \bar{Y}^n | W^n) + \epsilon \\
&= \frac{1}{n} \sum_{t=1}^{n} I(M; \bar{Y}_t | \bar{Y}^{t-1}, W^n) + \epsilon \\
&= \frac{1}{n} \sum_{t=1}^{n} I(M, \bar{Y}^{t-1}, W^{t-1}, W_{t+1}^n; \bar{Y}_t | W_t) + \epsilon \\
&\overset{(a)}{\leq} \frac{1}{n} \sum_{t=1}^{n} I(M, X^{t-1}, W^{t-1}, W_{t+1}^n; \bar{Y}_t | W_t) + \epsilon \\
&= \frac{1}{n} \sum_{t=1}^{n} I(U_t; \bar{Y}_t | W_t) + \epsilon \\
&= I(U; Y | W) + \epsilon,
\end{aligned} \tag{245}$$

where (a) follows from the Markov chain $\bar{Y}^{i-1} \to (M, X^{i-1}, W^n) \to \bar{Y}_i$. Finally, consider the type-II error probability at the receiver:

$$\begin{aligned}
-\frac{1}{n} \log \beta_n &\leq \frac{1}{n} D(P_{B\bar{Z}^n Q^n W^n | \mathcal{H}=0} \| P_{B\bar{Z}^n Q^n W^n | \mathcal{H}=1}) + \epsilon \\
&= \frac{1}{n} \mathbb{E}_{Q^n W^n} \left[D(P_{B\bar{Z}^n | Q^n W^n, \mathcal{H}=0} \| P_{B\bar{Z}^n | Q^n W^n, \mathcal{H}=1}) \right] + \epsilon \\
&= \frac{1}{n} \mathbb{E}_{Q^n W^n} \left[D(P_{B | Q^n W^n, \mathcal{H}=0} \| P_{B | Q^n W^n, \mathcal{H}=1}) \right] + \frac{1}{n} \mathbb{E}_{BQW^n} \left[D(P_{\bar{Z}^n | BQ^n W^n, \mathcal{H}=0} \| P_{\bar{Z}^n | BQ^n W^n, \mathcal{H}=1}) \right] + \epsilon
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{n}\mathbb{E}_{Q^nW^n}\big[D(P_{M\bar{Y}^n|Q^nW^n,\mathcal{H}=0}\|P_{M\bar{Y}^n|Q^nW^n,\mathcal{H}=1})\big] + \frac{1}{n}\mathbb{E}_{BQW^n}\big[D(P_{\bar{Z}^n|BQ^nW^n}\|P_{\bar{Z}^n|Q^nW^n})\big] + \epsilon \\
&= \frac{1}{n}\mathbb{E}_{\bar{Y}^nQ^nW^n}\big[D(P_{M|W^n\bar{Y}^n}\|P_{M|W^n})\big] + I(B;\bar{Z}^n|Q^n,W^n) + \epsilon \\
&= \frac{1}{n}I(M;\bar{Y}^n|W^n) + I(B;Z^n|Q^n,W^n) + \epsilon \\
&= \frac{1}{n}\sum_{t=1}^{n} I(M,\bar{Y}^{t-1},W^{t-1},W_{t+1}^{n};\bar{Y}_t|W_t) + \frac{1}{n}\sum_{t=1}^{n} I(B,\bar{Z}^{t-1},Q^{t-1},Q_{t+1}^{n},W^{t-1},W_{t+1}^{n};\bar{Z}_t|Q_t,W_t) + \epsilon \\
&\overset{(a)}{\leq} \frac{1}{n}\sum_{t=1}^{n} I(U_t;\bar{Y}_t|W_t) + \frac{1}{n}\sum_{t=1}^{n} I(V_t;Z_t|Q_t,W_t) + \epsilon \\
&= I(U;Y|W) + I(V;Z|Q,W) + \epsilon,
\end{aligned} \tag{246}
$$

where (a) follows from the Markov chain $Z^{t-1} \to (B, Y^{t-1}, Q^n, W^n) \to Z_t$.