Strong Converses using Change of Measure and Asymptotic Markov Chains

Mustapha Hamad
LTCI, Telecom Paris, IP Paris
91120 Palaiseau, France
mustapha.hamad@telecom-paris.fr

Michèle Wigger
LTCI, Telecom Paris, IP Paris
91120 Palaiseau, France
michele.wigger@telecom-paris.fr

Mireille Sarkiss
SAMOVAR, Telecom SudParis, IP Paris
91011 Evry, France
mireille.sarkiss@telecom-sudparis.eu

***Abstract*—The main contribution of this paper is a strong converse result for K-hop distributed hypothesis testing against independence with multiple (intermediate) decision centers under a Markov condition. Our result shows that the set of type-II error exponents that can simultaneously be achieved at all the terminals does not depend on the maximum permissible type-I error probabilities. Our strong converse proof is based on a change of measure argument and on the asymptotic proof of specific Markov chains. This proof method seems to be useful also in other applications, and is appealing because it does not require resorting to variational characterizations or blowing-up methods as in previous related proofs.**

***Index Terms*—Strong converse, change of measure, hypothesis testing, K hops.**

I. Introduction

Strong converse results have a rich history in information theory. They refer to proofs showing that the fundamental performance limit (such as minimum compression rate or capacity) of a specific system does not depend on its asymptotically allowed error (or excess) probability (as long it is not 1). For example, Wolfowitz' strong converse [1] established that the capacity of a discrete-memoryless channel remains unchanged when positive asymptotic decoding error probabilities are tolerated. For source coding, the strong converse establishes that irrespective of the allowed reconstruction error probabilities, a discrete-memoryless source cannot be compressed with a rate below the entropy of the source. Similar results were also established for generalized network scenarios [2]–[4], i.e., for memoryless multi-user channels and distributed compression systems [5]–[9].

Our main interest in this paper is in distributed hypothesis testing problems where multiple terminals observe memoryless source sequences whose underlying joint distribution depends on a binary hypothesis $\mathcal{H} \in \{0, 1\}$. Multiple decision centers wish to decide on the value of $\mathcal{H}$ based on their local source sequences and the communicated bits. Information-theorists showed great interest in Stein-setups where the type-I error probability (the probability of error under the null hypothesis $\mathcal{H} = 0$) is required to stay asymptotically below a given threshold in the infinite blocklength regime, while the type-II error probability (the probability of error under $\mathcal{H} = 1$) has to decay to 0 exponentially fast with largest possible exponent [10]–[30]. For the two-terminal setup with a single sensor communicating to a single decision center over a rate-limited link, Ahlswede and Csiszár [10] proved the strong converse result that the largest possible type-II error exponent is independent of the admissible type-I error probability threshold. A similar strong converse result was also shown in the special case called "testing against independence" for the two-hop hypothesis testing problem (see Figure 1 for $K = 2$) over rate-limited communication links where the last two terminals produce a guess on the binary hypothesis [28] and assuming that the source sequences satisfy certain Markov chains. This latter strong converse result is based on a change of measure and hyper-contractivity arguments [31].

In this paper, we generalize the strong converse result of [28] to an arbitrary number of K hops. We thus show that the set of possible type-II error exponents that are simultaneously achievable at the various decision centers when testing against independence in a K-hop system satisfying a given set of Markov chains, does not depend on the permissible type-I error probabilities and equals the region under vanishing type-I error probabilities determined in [27]. The proof method applied in this paper relies on a similar change of measure argument as in [4]–[6], where we also restrict to jointly typical source sequences as [5]. No variational characterizations, blowing-up lemma [32], or hypercontractivity arguments are required. Instead, we rely on arguments showing that certain Markov chains hold in an asymptotic regime of infinite blocklengths. Our proof method seems to extend also to other applications. We show a warm-up version to establish the well-known strong converse for lossless compression with side-information at the decoder. This simplified version does not require the proof of Markov chains. In the extended version of this paper [33], we also illustrate our proof method to show the strong converse result for the Wyner-Ziv source coding problem, which involves the proof of Markov chains. Due to space constraints, this probably more related proof is omitted from this conference version.

Notation: We follow the notation in [34] and use sans serif font for bit-strings: e.g., m for a deterministic and M for a random bit-string. We also use $\mathrm{len}(\mathsf{m})$ to denote the length of a bit-string. Finally, $\mathcal{T}_{\mu}^{(n)}(\cdot)$ denotes the strongly typical set as defined in [35, Definition 2.8].

II. Lossless Source Coding with Side-Information

A. Setup and Known Results

Consider two terminals, an encoder observing the source sequence X^n and a decoder observing the related side-information sequence Y^n, where we assume that

$$(X^n, Y^n) \text{ i.i.d. } \sim P_{XY}, \tag{1}$$

for a given probability mass function (pmf) P_{XY} on the product alphabet $\mathcal{X} \times \mathcal{Y}$. The encoder uses a function $\phi^{(n)}$ to compress the sequence X^n into a bit-string message M,

$$\mathsf{M} = \phi^{(n)}(X^n) \tag{2}$$

of length nR bits, for a given rate $R > 0$,

$$\text{len}(\mathsf{M}) = nR. \tag{3}$$

Based on this message and its own observation Y^n, the decoder is supposed to reconstruct the source sequence X^n with small probability of error. Thus, the decoder applies a decoding function $g^{(n)}$ to (M, Y^n) to produce the reconstruction sequence $\hat{X}^n \in \mathcal{X}^n$:

$$\hat{X}^n = g^{(n)}(\mathsf{M}, Y^n). \tag{4}$$

Definition 1: Given $\epsilon \in [0, 1)$. Rate $R > 0$ is said ϵ-achievable if there exist sequences (in n) of encoding and reconstruction functions $\phi^{(n)}$ and $g^{(n)}$ such that

$$\varlimsup_{n\to\infty} \Pr[X^n \neq \hat{X}^n] \leq \epsilon. \tag{5}$$

A standard result in information theory says

Theorem 1: All rates $R > H(X|Y)$ are ϵ-achievable for all $\epsilon \in [0, 1)$ and all rates $R < H(X|Y)$ are not ϵ-achievable for any $\epsilon \in [0, 1)$.

In the following subsection we show a new converse proof. The goal is to illustrate (some of) the tools that we employ to prove our main result, Theorem 4 ahead.

B. Alternative Strong Converse Proof

Fix a sequence of encoding and decoding functions $\{\phi^{(n)}, g^{(n)}\}_{n=1}^{\infty}$ satisfying (5). We perform a similar change of measure argument as in [4], [5] where we restrict to typical sequences. Define $\mu_n := n^{-1/3}$ and the set

$$\mathcal{D} := \left\{(x^n, y^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{XY})\colon g^{(n)}\left(\phi^{(n)}(x^n), y^n\right) = x^n\right\}, \tag{6}$$

i.e., the set of all typical (x^n, y^n)-sequences for which the reconstructed sequence $\hat{X}^n$ coincides with the source sequence X^n. Let $\Delta := \Pr[(X^n, Y^n) \in \mathcal{D}]$ and notice that by (5) and [35, Remark to Lemma 2.12]:

$$\Delta \geq 1 - \epsilon - \frac{|\mathcal{X}||\mathcal{Y}|}{4\mu_n^2 n}, \tag{7}$$

and thus

$$\varliminf_{n\to\infty} \Delta \geq 1 - \epsilon. \tag{8}$$

Let further $(\tilde{X}^n, \tilde{Y}^n)$ be random variables of joint pmf

$$P_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n) = \frac{P_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n)}{\Delta} \cdot \mathbb{1}\{(x^n, y^n) \in \mathcal{D}\}. \tag{9}$$

Let also $\tilde{\mathsf{M}} = \phi^{(n)}(\tilde{X}^n)$ and T be uniform over $\{1, \ldots, n\}$ independent of $(\tilde{X}^n, \tilde{Y}^n, \tilde{\mathsf{M}})$, and define $\tilde{X} := \tilde{X}_T$ and $\tilde{Y} := \tilde{Y}_T$.

Notice the following sequence of equalities:

$$\frac{1}{n} H(\tilde{X}^n, \tilde{Y}^n)$$
$$= -\frac{1}{n} \sum_{(x^n,y^n)\in\mathcal{D}} P_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n) \log P_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n) \tag{10}$$
$$= -\frac{1}{n} \sum_{(x^n,y^n)\in\mathcal{D}} P_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n) \log \frac{P_{X^nY^n}(x^n, y^n)}{\Delta} \tag{11}$$
$$= -\frac{1}{n} \sum_{i=1}^{n} \sum_{(x^n,y^n)\in\mathcal{D}} P_{\tilde{X}^n\tilde{Y}^n}(x^n, y^n) \log P_{XY}(x_i, y_i) + \frac{1}{n}\log\Delta \tag{12}$$
$$= -\frac{1}{n} \sum_{i=1}^{n} \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{\tilde{X}_i\tilde{Y}_i}(x, y) \log P_{XY}(x, y) + \frac{1}{n}\log\Delta \tag{13}$$
$$= -\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{\tilde{X}\tilde{Y}}(x, y) \log P_{XY}(x, y) + \frac{1}{n}\log\Delta. \tag{14}$$

Since $(\tilde{X}^n, \tilde{Y}^n) \in \mathcal{T}_{\mu_n}^{(n)}(P_{XY})$, we have:

$$|P_{\tilde{X}\tilde{Y}}(x, y) - P_{XY}(x, y)| \leq \mu_n, \tag{15}$$

and thus as $n \to \infty$ (because $\mu_n = n^{-1/3}$ and Δ is bounded away from 0, see (8)):

$$\lim_{n\to\infty} \frac{1}{n} H(\tilde{X}^n\tilde{Y}^n) = H(XY). \tag{16}$$

In a similar manner one can obtain

$$\lim_{n\to\infty} \frac{1}{n} H(\tilde{Y}^n) = H(Y), \tag{17}$$

and thus combining (16) and (17), by the chain rule:

$$\lim_{n\to\infty} \frac{1}{n} H(\tilde{X}^n|\tilde{Y}^n) = H(X|Y). \tag{18}$$

The strong converse is then easily obtained by this Limit (18), using the same steps as in the weak converse:

$$R \geq \frac{1}{n} H(\tilde{\mathsf{M}}) = \frac{1}{n} H(\tilde{\mathsf{M}}|\tilde{Y}^n) \tag{19}$$
$$= \frac{1}{n} I(\tilde{\mathsf{M}}; \tilde{X}^n|\tilde{Y}^n) = \frac{1}{n} H(\tilde{X}^n|\tilde{Y}^n), \tag{20}$$

and letting $n \to \infty$. Here, the last equality holds because by the definition of the set $\mathcal{D}$, the new source sequence $\tilde{X}^n$ can be obtained as a function of $\tilde{\mathsf{M}}$ and $\tilde{Y}^n$.

III. Testing Against Independence in a K-Hop Network

Consider a system with a transmitter T_0 observing the source sequence Y_0^n, $K-1$ relays labelled $\mathrm{R}_1, \ldots, \mathrm{R}_{K-1}$ and observing sequences $Y_1^n, \ldots, Y_{K-1}^n$, respectively, and a receiver R_K observing sequence Y_K^n.

The source sequences $(Y_0^n, Y_1^n, \ldots, Y_K^n)$ are distributed according to one of two distributions depending on a binary hypothesis $\mathcal{H} \in \{0, 1\}$:

$$\text{if } \mathcal{H} = 0: (Y_0^n, Y_1^n, \ldots, Y_K^n) \text{ i.i.d.} \sim P_{Y_0Y_1\cdots Y_K}; \tag{21a}$$
$$\text{if } \mathcal{H} = 1: (Y_0^n, Y_1^n, \ldots, Y_K^n) \text{ i.i.d.} \sim P_{Y_0} \cdot P_{Y_1} \cdots P_{Y_K}. \tag{21b}$$

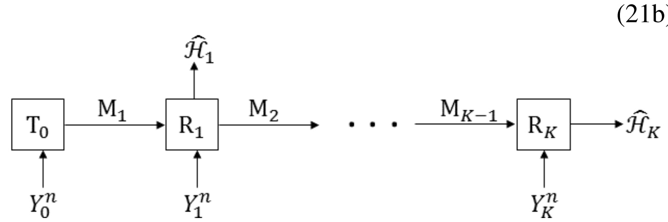

Fig. 1: Cascaded K-hop setup with K decision centers.

Communication takes place over K hops as illustrated in Figure 1. The transmitter T_0 sends a message $\mathsf{M}_1 = \phi_0^{(n)}(Y_0^n)$ to the first relay R_1, which sends a message $\mathsf{M}_2 = \phi_1^{(n)}(Y_1^n, \mathsf{M}_1)$ to the second relay and so on. The communication is thus described by encoding functions

$$\phi_0^{(n)} \colon \mathcal{Y}_0^n \to \{0,1\}^\star \tag{22}$$
$$\phi_k^{(n)} \colon \mathcal{Y}_k^n \times \{0,1\}^\star \to \{0,1\}^\star, \quad k \in \{1,\dots,K-1\}, \tag{23}$$

so that the produced message strings

$$\mathsf{M}_1 = \phi_0^{(n)}(\mathcal{Y}_0^n) \tag{24}$$
$$\mathsf{M}_{k+1} = \phi_k^{(n)}(Y_k^n, \mathsf{M}_k), \quad k \in \{1,\dots,K-1\}, \tag{25}$$

satisfy the maximum rate constraints

$$\operatorname{len}(\mathsf{M}_k) \le nR_k, \quad k \in \{1,\dots,K\}. \tag{26}$$

Each relay $\mathsf{R}_1, \dots, \mathsf{R}_{K-1}$ as well as the receiver R_K, produces a guess of the hypothesis $\mathcal{H}$. These guesses are described by guessing functions

$$g_k^{(n)} \colon \mathcal{Y}_k^n \times \{0,1\}^\star \to \{0,1\}, \quad k \in \{1,\dots,K\}, \tag{27}$$

where we request that the guesses

$$\hat{\mathcal{H}}_{k,n} = g_k^{(n)}(Y_k^n, \mathsf{M}_k), \quad k \in \{1,\dots,K\}, \tag{28}$$

have type-I error probabilities

$$\alpha_{k,n} \triangleq \Pr[\hat{\mathcal{H}}_k = 1 | \mathcal{H} = 0], \quad k \in \{1,\dots,K\}, \tag{29}$$

not exceeding given thresholds $\epsilon_1, \epsilon_2, \dots, \epsilon_K > 0$, and type-II error probabilities

$$\beta_{k,n} \triangleq \Pr[\hat{\mathcal{H}}_k = 0 | \mathcal{H} = 1], \quad k \in \{1,\dots,K\}, \tag{30}$$

decaying to 0 exponentially fast with largest possible exponents.

Definition 2: Given maximum type-I error probabilities $\epsilon_1, \epsilon_2, \dots, \epsilon_K \in [0,1)$ and rates $R_1, R_2, \dots, R_K \ge 0$. The exponent tuple $(\theta_1, \theta_2, \dots, \theta_K)$ is called $(\epsilon_1, \epsilon_2, \dots, \epsilon_K)$-*achievable* if there exists a sequence of encoding and decision functions $\{\phi_0^{(n)}, \phi_1^{(n)}, \dots, \phi_{K-1}^{(n)}, g_1^{(n)}, g_2^{(n)}, \dots, g_K^{(n)}\}_{n\ge1}$ satisfying for each $k \in \{1,\dots,K\}$:

$$\operatorname{len}(\mathsf{M}_k) \le nR_k, \tag{31a}$$
$$\varlimsup_{n\to\infty} \alpha_{k,n} \le \epsilon_k, \tag{31b}$$
$$\varliminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_{k,n}} \ge \theta_k. \tag{31c}$$

Definition 3: The fundamental exponents region $\mathcal{E}^*(R_1, R_2, \dots, R_K, \epsilon_1, \epsilon_2, \dots, \epsilon_K)$ is defined as the closure of the set of all $(\epsilon_1, \epsilon_2, \dots, \epsilon_K)$-achievable exponent pairs $(\theta_1, \theta_2, \dots, \theta_K)$ for given rates $R_1, \dots, R_K \ge 0$.

A. Previous Results on K-Hop Hypothesis Testing

The K-hop hypothesis testing setup of Figure 1 and Equations (21) was also considered in [27] in the special case $\epsilon_1 = \cdots = \epsilon_K = 0$, for which the fundamental exponents region was determined. The result of [27] is based on the following definition and presented in Theorem 4 ahead.

Definition 4: For any $\ell \in \{1,\dots,K\}$, define the function

$$\eta_\ell \colon \mathbb{R}_0^+ \to \mathbb{R}_0^+ \tag{32}$$

$$R \mapsto \max_{\substack{P_{U|Y_{\ell-1}}:\\ R \ge I(U;Y_{\ell-1})}} I(U;Y_\ell). \tag{33}$$

Theorem 2 (Proposition 5 in [27]): The fundamental exponents region satisfies

$$\mathcal{E}^*(R_1,\dots,R_K,0,\dots,0) = \left\{ (\theta_1,\dots,\theta_K) \colon \theta_k \le \sum_{\ell=1}^{k} \eta_\ell(R_\ell),\ k \in \{1,\dots,K\} \right\} \tag{34}$$

For $K = 2$, Cao, Zhou, and Tan [28] also established the following strong converse result.

Theorem 3 (Theorem 1 [28]): For $K = 2$ and arbitrary $\epsilon_1, \epsilon_2 \in [0,1)$:

$$\mathcal{E}^*(R_1,R_2,\epsilon_1,\epsilon_2) = \left\{ (\theta_1,\theta_2) \colon \theta_k \le \sum_{\ell=1}^{k} \eta_\ell(R_\ell),\ k \in \{1,2\} \right\} \tag{35}$$

Our main result is a generalization of above Theorem 3 to an arbitrary number of $K \ge 2$ hops. That is, we prove the strong converse to Theorem 2.

Theorem 4: For $K \ge 2$ and arbitrary $\epsilon_1, \dots, \epsilon_K \ge 0$:

$$\mathcal{E}^*(R_1,\dots,R_K,\epsilon_1,\dots,\epsilon_K) = \mathcal{E}^*(R_1,\dots,R_K,0,\dots,0) \tag{36}$$
$$= \left\{ (\theta_1,\dots,\theta_K) \colon \theta_k \le \sum_{\ell=1}^{k} \eta_\ell(R_\ell),\ k \in \{1,\dots,K\} \right\} \tag{37}$$

Proof: See the following Section IV. ■

IV. Strong Converse Proof of Theorem 4

Fix an exponent-tuple $(\theta_1,\dots,\theta_K)$ in the exponents region $\mathcal{E}^*(R_1,\dots,R_K,\epsilon_1,\dots,\epsilon_K)$, and a sequence (in n) of encoding and decision functions $\{(\phi_0^{(n)}, \phi_1^{(n)}, \dots, \phi_{K-1}^{(n)}, g_1^{(n)}, \dots, g_K^{(n)})\}_{n\ge1}$ achieving this tuple, i.e., satisfying constraints (31).

Fix an arbitrary $k \in \{1,\dots,K\}$ and set $\mu_n = n^{-1/3}$. Let $\mathcal{A}_k$ denote the acceptance region of R_k, i.e.,

$$\mathcal{A}_k := \{(y_0^n,\dots,y_k^n) \colon g_k^{(n)}(\mathsf{m}_k, y_k^n) = 0\}, \tag{38}$$

where we define recursively $\mathsf{m}_1 := \phi_0^{(n)}(y_0^n)$ and

$$\mathsf{m}_\ell := \phi_{\ell-1}^{(n)}(\mathsf{m}_{\ell-1}, y_{\ell-1}), \ell \in \{2,\dots,k\}. \tag{39}$$

Define also the intersection of this acceptance region with the typical set:

$$\mathcal{D}_k \triangleq \mathcal{A}_k \cap \mathcal{T}_{\mu_n}^{(n)}(P_{Y_0\cdots Y_k}). \tag{40}$$

By [35, Remark to Lemma 2.12] and the type-I error probability constraints in (31b),

$$\Delta_k := P_{Y_0^n Y_1^n \cdots Y_k^n}(\mathcal{D}_k) \ge 1 - \epsilon_k - \frac{|\mathcal{Y}_0|\cdots|\mathcal{Y}_k|}{4\mu_n^2 n}, \tag{41}$$

and thus $\varliminf_{n\to\infty} \Delta_k \ge 1 - \epsilon_k > 0$ as $n \to \infty$.

Let $(\tilde{Y}_0^n, \tilde{Y}_1^n, \dots, \tilde{Y}_k^n)$ be random variables of joint pmf

$$P_{\tilde{Y}_0^n,\tilde{Y}_1^n,\dots,\tilde{Y}_k^n}(\tilde{y}_0^n,\tilde{y}_1^n,\dots,\tilde{y}_k^n) = \frac{P_{\tilde{Y}_0^n,\tilde{Y}_1^n,\dots,\tilde{Y}_k^n}(\tilde{y}_0^n,\tilde{y}_1^n,\dots,\tilde{y}_k^n)}{\Delta} \cdot \mathbb{1}\{(\tilde{y}_0^n,\tilde{y}_1^n,\dots,\tilde{y}_k^n) \in \mathcal{D}_k\}. \tag{42}$$

Let also $\tilde{\mathsf{M}}_\ell = \phi_{\ell-1}^{(n)}(\tilde{\mathsf{M}}_{\ell-1}, \tilde{Y}_{\ell-1}^n)$ and T be uniform over $\{1,\ldots,n\}$ independent of $(\tilde{Y}_0^n, \tilde{Y}_1^n, \ldots, \tilde{Y}_k^n, \tilde{\mathsf{M}}_1, \ldots, \tilde{\mathsf{M}}_k)$, and define $\tilde{Y}_\ell := \tilde{Y}_{\ell,T}$ for $\ell \in \{1,\ldots,k\}$.

At the end of this section, we prove the following Lemma 1.

Lemma 1: There exist random variables $\{U_1,\ldots,U_k\}$ satisfying the (in)equalities

$$nR_\ell \geq H(\tilde{\mathsf{M}}_\ell) \geq nI(U_\ell;\tilde{Y}_{\ell-1}) + \log \Delta_k, \quad \ell \in \{1,\ldots,k\}, \tag{43a}$$

$$I(U_\ell;\tilde{Y}_\ell|\tilde{Y}_{\ell-1}) = \o_{1,\ell}(n), \tag{43b}$$

and

$$-\frac{1}{n}\log \Pr[\hat{\mathcal{H}}_k = 0 | \mathcal{H}=1, (Y_0^n,\ldots,Y_k^n) \in \mathcal{D}_k] \leq \sum_{\ell=1}^k I(U_\ell;\tilde{Y}_\ell) + \o_2(n), \tag{43c}$$

where the functions $\{\o_{1,\ell}(n)\}_{\ell=1}^k$ and $\o_2(n)$ all tend to 0 as $n \to \infty$.

The desired bound on θ_k in (37) is then obtained from above lemma by taking $n \to \infty$, as we explain in the following. By Carathéodory's theorem [34, Appendix C], for each n there must exist random variables $U_1,\ldots,U_k$ satisfying (43) over alphabets of sizes

$$|\mathcal{U}_\ell| \leq |\mathcal{Y}_{\ell-1}| \cdot |\mathcal{Y}_\ell| + 2, \qquad \ell \in \{1,\ldots,k\}. \tag{44}$$

We thus restrict to random variables of above (bounded) supports and invoke the Bolzano-Weierstrass theorem to conclude the existence of a pmf $P^{(\ell)}_{Y_{\ell-1}Y_\ell U_\ell}$ over $\mathcal{Y}_{\ell-1} \times \mathcal{Y}_\ell \times \mathcal{U}_\ell$, also abbreviated as $P^{(\ell)}$, and an increasing subsequence of positive numbers $\{n_i\}_{i=1}^\infty$ satisfying

$$\lim_{i\to\infty} P_{\tilde{Y}_{\ell-1}\tilde{Y}_\ell U_\ell; n_i} = P^{(\ell)}_{Y_{\ell-1}Y_\ell U_\ell}, \quad \ell \in \{1,\ldots,k\}, \tag{45}$$

where $P_{\tilde{Y}_{\ell-1}\tilde{Y}_\ell U_\ell; n_i}$ denotes the pmf at blocklength n_i.

By the monotone continuity of mutual information over finite pmfs, we can then deduce that

$$R_\ell \geq I_{P^{(\ell)}}(U_\ell; Y_{\ell-1}), \quad \ell \in \{1,\ldots,k\}, \tag{46}$$

$$\theta_k \leq \sum_{\ell=1}^k I_{P^{(\ell)}}(U_\ell; Y_\ell), \tag{47}$$

where the subscripts indicate that mutual informations should be computed according to the indicated pmfs.

Since for any blocklength n_i the pair $\left(\tilde{Y}_{\ell-1}^{n_i}, \tilde{Y}_\ell^{n_i}\right)$ lies in the jointly typical set $\mathcal{T}^{(n_i)}_{\mu_{n_i}}(P_{Y_{\ell-1}Y_\ell})$, we have $\left|P_{Y_{\ell-1}Y_\ell;n_i} - P_{Y_{\ell-1}Y_\ell}\right| \leq \mu_{n_k}$ and thus the limiting pmfs satisfy $P^{(\ell)}_{Y_{\ell-1}Y_\ell} = P_{Y_{\ell-1}Y_\ell}$. By similar continuity considerations and by (43b), for all $\ell \in \{1,\ldots,k\}$ the Markov chain

$$U_\ell \to Y_{\ell-1} \to Y_\ell, \tag{48}$$

holds under $P^{(\ell)}_{Y_{\ell-1}Y_\ell U_\ell}$.

By the definitions of the functions $\{\eta_\ell(\cdot)\}$ and by (46)–(48):

$$\theta_k \leq \sum_{\ell=1}^k \eta_\ell(R_\ell), \tag{49}$$

which concludes the proof.

A. Proof of Lemma 1

Define $\tilde{U}_{\ell,t} \triangleq (\tilde{\mathsf{M}}_\ell, \tilde{Y}_0^{t-1}, \ldots, \tilde{Y}_k^{t-1})$ for $\ell \in \{1,\ldots,k\}$ and notice:

$$\begin{aligned}
H(\tilde{\mathsf{M}}_\ell) &= I(\tilde{\mathsf{M}}_\ell; \tilde{Y}_0^n \cdots \tilde{Y}_k^n) && (50)\\
&= H(\tilde{Y}_0^n \cdots \tilde{Y}_k^n) - H(\tilde{Y}_0^n \cdots \tilde{Y}_k^n | \tilde{\mathsf{M}}_\ell) && (51)\\
&= nH(\tilde{Y}_{0,T} \cdots \tilde{Y}_{k,T}) + \log \Delta_k + \o_1(n) \\
&\quad - \sum_{t=1}^n H(\tilde{Y}_{0,t} \cdots \tilde{Y}_{k,t} | \tilde{U}_{\ell,t}) && (52)\\
&= nH(\tilde{Y}_{0,T} \cdots \tilde{Y}_{k,T}) \log \Delta_k + \o_1(n) \\
&\quad - nH(\tilde{Y}_{0,T} \cdots \tilde{Y}_{k,T} | \tilde{U}_{\ell,T}, T)] && (53)\\
&= n[I(\tilde{Y}_0 \cdots \tilde{Y}_k; U_\ell)] + \log \Delta_k + \o_1(n) && (54)\\
&\geq n\left[I(\tilde{Y}_{\ell-1}; U_\ell) + \frac{1}{n}\log \Delta_k\right] + \o_1(n). && (55)
\end{aligned}$$

Here, (52) holds by similar steps to (10)–(14), where $\o_1(n)$ is a function that tends to 0 as $n \to \infty$, by the chain rule, by the definition of $\tilde{U}_{\ell,t}$, and by defining T uniform over $\{1,\ldots,n\}$ independent of all other random variables; and (54) holds by defining $U_\ell \triangleq (\tilde{U}_{\ell,T}, T)$ and $\tilde{Y}_\ell \triangleq \tilde{Y}_{\ell,T}$ for all $\ell \in \{0,\ldots,k\}$. This proves Inequality (43a) in the lemma.

We next upper bound the type-II error exponent θ_k. Define:

$$Q_{\tilde{\mathsf{M}}_k}(\mathsf{m}_k) \triangleq \sum_{y_0^n, y_1^n, \ldots, y_{k-1}^n} P_{\tilde{Y}_0^n}(y_0^n) \cdots P_{\tilde{Y}_{k-1}^n}(y_{k-1}^n) \cdot \mathbb{1}\{\mathsf{m}_k = \phi_k(\phi_{k-1}(\cdots(\phi_1(y_0^n)\cdots)), y_{k-1}^n)\}, \tag{56}$$

and

$$Q_{\mathsf{M}_k}(\mathsf{m}_K) \triangleq \sum_{y_0^n, y_1^n, \ldots, y_{k-1}^n} P_{Y_0^n}(y_0^n) \cdots P_{Y_{k-1}^n}(y_{k-1}^n) \cdot \mathbb{1}\{\mathsf{m}_k = \phi_{k-1}(\phi_{k-2}(\cdots(\phi_0(y_0^n)\cdots)), y_{k-1}^n)\}. \tag{57}$$

and notice that

$$Q_{\tilde{\mathsf{M}}_k} P_{\tilde{Y}_k^n}(\mathcal{A}_k) \leq Q_{\mathsf{M}_k} P_{Y_k^n}(\mathcal{A}_k)\, \Delta_k^{-(k+1)} = \beta_{k,n} \Delta_k^{-(k+1)} \tag{58}$$

Notice that by (38), the probability $P_{\tilde{\mathsf{M}}_k \tilde{Y}_k^n}(\mathcal{A}_k) = 1$, and thus by (58) and standard inequalities (see [21, Lemma 1]):

$$-\frac{1}{n}\log \beta_{k,n} \leq -\frac{1}{n}\log\left(Q_{\tilde{\mathsf{M}}_k} P_{\tilde{Y}_k^n}(\mathcal{A}_k)\right) - \frac{(k+1)}{n}\log \Delta_k \tag{59}$$

$$\leq \frac{1}{n} D(P_{\tilde{\mathsf{M}}_k \tilde{Y}_k^n} || Q_{\tilde{\mathsf{M}}_k} P_{\tilde{Y}_k^n}) + \delta_n' \tag{60}$$

where $\delta_n' \triangleq -\frac{(k+1)}{n}\log \Delta_k + \frac{1}{n}$ and tends to 0 as $n \to \infty$.

We continue to upper bound the divergence term as

$$\begin{aligned}
&D(P_{\tilde{\mathsf{M}}_k \tilde{Y}_k^n} || Q_{\tilde{\mathsf{M}}_k} P_{\tilde{Y}_k^n}) \\
&\quad = I(\tilde{\mathsf{M}}_k; \tilde{Y}_k^n) + D(P_{\tilde{\mathsf{M}}_k} || Q_{\tilde{\mathsf{M}}_k}) && (61)\\
&\quad \leq I(\tilde{\mathsf{M}}_k; \tilde{Y}_k^n) + D(P_{\tilde{Y}_{k-1}^n \tilde{\mathsf{M}}_{k-1}} || P_{\tilde{Y}_{k-1}^n} Q_{\tilde{\mathsf{M}}_{k-1}}) && (62)\\
&\quad \leq I(\tilde{\mathsf{M}}_k; \tilde{Y}_k^n) + I(\tilde{\mathsf{M}}_{k-1}; \tilde{Y}_{k-1}^n) \\
&\qquad + D(P_{\tilde{Y}_{k-2}^n \tilde{\mathsf{M}}_{k-2}} || P_{\tilde{Y}_{k-2}^n} Q_{\tilde{\mathsf{M}}_{k-2}}) && (63)\\
&\quad \vdots
\end{aligned}$$

$$\leq \sum_{\ell=1}^{k} I(\tilde{\mathsf{M}}_\ell;\tilde{Y}_\ell^n) \tag{64}$$

$$\leq \sum_{\ell=1}^{k}\sum_{t=1}^{n} I(\tilde{\mathsf{M}}_\ell \tilde{Y}_0^{t-1}\cdots\tilde{Y}_k^{t-1};\tilde{Y}_{\ell,t}) \tag{65}$$

$$= \sum_{\ell=1}^{k}\sum_{t=1}^{n} I(\tilde{U}_{\ell,t};\tilde{Y}_{\ell,t}) \tag{66}$$

$$\leq n\sum_{\ell=1}^{k} I(U_\ell;\tilde{Y}_\ell). \tag{67}$$

Here (62) is obtained by the data processing inequality for KL-divergence and (66)–(67) by the definitions of $\tilde{U}_{\ell,t}, U_\ell, \tilde{Y}_\ell$ and T.

Combined with (60) this establishes Inequality (43c).

Finally, we proceed to prove that for any $\ell \in \{1,\ldots,k\}$ the Markov chain $U_\ell \to \tilde{Y}_{\ell-1} \to \tilde{Y}_\ell$ holds in the limit as $n\to\infty$. We start by noticing the Markov chain $\tilde{\mathsf{M}}_1 \to \tilde{Y}_0^n \to (\tilde{Y}_1^n,\cdots,\tilde{Y}_k^n)$, and thus:

$$\begin{aligned}
0 &= I(\tilde{\mathsf{M}}_1;\tilde{Y}_1^n\cdots\tilde{Y}_k^n|\tilde{Y}_0^n) && (68)\\
&= H(\tilde{Y}_1^n\cdots\tilde{Y}_k^n|\tilde{Y}_0^n) - H(\tilde{Y}_1^n\cdots\tilde{Y}_k^n|\tilde{Y}_0^n\tilde{\mathsf{M}}_1) && (69)\\
&= nH(\tilde{Y}_{1,T}\cdots\tilde{Y}_{k,T}|\tilde{Y}_{0,T}) + \log\Delta_k + \tilde{\phi}_1(n)\\
&\quad -H(\tilde{Y}_1^n\cdots\tilde{Y}_k^n|\tilde{Y}_0^n\tilde{\mathsf{M}}_1) && (70)\\
&\geq nH(\tilde{Y}_{1,T}\cdots\tilde{Y}_{k,T}|\tilde{Y}_{0,T}) + \log\Delta_k + \tilde{\phi}_1(n)\\
&\quad -nH(\tilde{Y}_{1,T}\cdots\tilde{Y}_{k,T}|\tilde{Y}_{0,T}\tilde{Y}_0^{T-1}\cdots\tilde{Y}_k^{T-1}\tilde{Y}_{0,T+1}^n\tilde{\mathsf{M}}_1 T) && (71)\\
&= nI(\tilde{Y}_{1,T}\cdots\tilde{Y}_{k,T};\tilde{Y}_0^{T-1}\cdots\tilde{Y}_k^{T-1}\tilde{Y}_{0,T+1}^n\tilde{\mathsf{M}}_1 T|\tilde{Y}_{0,T})\\
&\quad +\log\Delta_k + \tilde{\phi}_1(n) && (72)\\
&\geq nI(\tilde{Y}_1\cdots\tilde{Y}_k;U_1|\tilde{Y}_0) + \log\Delta_k + \tilde{\phi}_1(n), && (73)
\end{aligned}$$

for some function $\tilde{\phi}_1(n)$ so that $\frac{1}{n}\tilde{\phi}_1(n)$ tends to 0 as $n\to\infty$, and where (70) can be shown in a similar manner to (18) and (73) by the definitions of $\tilde{Y}_\ell$, $\tilde{Y}_0$, $\tilde{U}_{1,t}$, and U_1 for all $\ell \in \{1,\ldots,k\}$.

Since Δ_k is bounded, $\frac{1}{n}\log\Delta_k$ tends to 0 as $n\to\infty$, and we can conclude that

$$\lim_{n\to\infty} I(\tilde{Y}_1\cdots\tilde{Y}_k;\tilde{U}_1|\tilde{Y}_0) = 0, \tag{74}$$

thus proving (43b) for $\ell = 1$.

Notice next that for any $\ell \in \{2,\ldots,k\}$:

$$\begin{aligned}
I(U_\ell;\tilde{Y}_\ell|\tilde{Y}_{\ell-1}) &\leq I(U_\ell\tilde{Y}_0\cdots\tilde{Y}_{\ell-2};\tilde{Y}_\ell|\tilde{Y}_{\ell-1}) && (75)\\
&= I(U_\ell;\tilde{Y}_\ell|\tilde{Y}_0\cdots\tilde{Y}_{\ell-1})\\
&\quad +I(\tilde{Y}_0\cdots\tilde{Y}_{\ell-2};\tilde{Y}_\ell|\tilde{Y}_{\ell-1}). && (76)
\end{aligned}$$

In the following we show that both quantities $I(U_\ell;\tilde{Y}_\ell|\tilde{Y}_0\cdots\tilde{Y}_{\ell-1})$ and $I(\tilde{Y}_0\cdots\tilde{Y}_{\ell-2};\tilde{Y}_\ell|\tilde{Y}_{\ell-1})$ tend to 0 as $n\to\infty$, which establishes (43b) for $\ell\in\{2,\ldots,k\}$.

To prove that $I(\tilde{Y}_0\cdots\tilde{Y}_{\ell-2};\tilde{Y}_\ell|\tilde{Y}_{\ell-1})$ tends to 0 as $n\to\infty$, we notice that for any $\ell\in\{2,\ldots,k\}$:

$$\begin{aligned}
D(P_{\tilde{Y}_0\cdots\tilde{Y}_k}||P_{Y_0\cdots Y_k}) &\geq D(P_{\tilde{Y}_0\cdots\tilde{Y}_\ell}||P_{Y_0\cdots Y_\ell}) && (77)\\
&= D(P_{\tilde{Y}_0\cdots\tilde{Y}_\ell}||P_{Y_0\cdots Y_{\ell-1}}P_{Y_\ell|Y_{\ell-1}}) && (78)\\
&= D(P_{\tilde{Y}_0\cdots\tilde{Y}_{\ell-1}}||P_{\tilde{Y}_0\cdots\tilde{Y}_{\ell-1}}P_{\tilde{Y}_\ell|\tilde{Y}_{\ell-1}})\\
&\quad +\mathbb{E}_{P_{\tilde{Y}_{\ell-1}}}\left[D(P_{\tilde{Y}_\ell|\tilde{Y}_{\ell-1}}||P_{Y_\ell|Y_{\ell-1}})\right]\\
&\quad +D(P_{\tilde{Y}_0\cdots\tilde{Y}_{\ell-1}}||P_{Y_0\cdots Y_{\ell-1}}) && (79)\\
&\geq D(P_{\tilde{Y}_0\cdots\tilde{Y}_\ell}||P_{\tilde{Y}_0\cdots\tilde{Y}_{\ell-1}}P_{\tilde{Y}_\ell|\tilde{Y}_{\ell-1}}) && (80)\\
&\geq I(\tilde{Y}_0\cdots\tilde{Y}_{\ell-2};\tilde{Y}_\ell|\tilde{Y}_{\ell-1}). && (81)
\end{aligned}$$

Since $(\tilde{Y}_0\cdots\tilde{Y}_k)$ lie in the jointly typical set $\mathcal{T}_{\mu_n}^{(n)}(P_{Y_0\cdots Y_k})$:

$$|P_{\tilde{Y}_0\cdots\tilde{Y}_k} - P_{Y_0\cdots Y_k}| \leq \mu_n. \tag{82}$$

Recalling that $\mu_n \downarrow 0$ as $n\to\infty$, and by the continuity of the KL-divergence, we conclude that $D(P_{\tilde{Y}_0\cdots\tilde{Y}_k}||P_{Y_0\cdots Y_k})$ tends to 0 as $n\to\infty$, and thus by (81) and the nonnegativity of mutual information:

$$\lim_{n\to\infty} I(\tilde{Y}_0\cdots\tilde{Y}_{\ell-2};\tilde{Y}_\ell|\tilde{Y}_{\ell-1}) = 0. \tag{83}$$

Following similar steps to (68)–(73), we further obtain:

$$\begin{aligned}
0 &= I(\tilde{\mathsf{M}}_\ell;\tilde{Y}_\ell^n\cdots\tilde{Y}_k^n|\tilde{Y}_0^n\cdots\tilde{Y}_{\ell-1}^n)\\
&= H(\tilde{Y}_\ell^n\cdots\tilde{Y}_K^n|\tilde{Y}_0^n\cdots\tilde{Y}_{k-1}^n)\\
&\quad -H(\tilde{Y}_k^n\cdots\tilde{Y}_k^n|\tilde{Y}_0^n\cdots\tilde{Y}_{\ell-1}^n\tilde{\mathsf{M}}_\ell) && (84)\\
&= nH(\tilde{Y}_{\ell,T}\cdots\tilde{Y}_{k,T}|\tilde{Y}_{0,T}\cdots\tilde{Y}_{\ell-1,T}) + \log\Delta_k + \tilde{\phi}_\ell(n)\\
&\quad -H(\tilde{Y}_\ell^n\cdots\tilde{Y}_k^n|\tilde{Y}_0^n\cdots\tilde{Y}_{\ell-1}^n\tilde{\mathsf{M}}_\ell) && (85)\\
&\geq nH(\tilde{Y}_{\ell,T}\cdots\tilde{Y}_{k,T}|\tilde{Y}_{0,T}\cdots\tilde{Y}_{\ell-1,T}) + \log\Delta_k + \tilde{\phi}_\ell(n)\\
&\quad -\sum_{t=1}^{n} H(\tilde{Y}_{\ell,t}\cdots\tilde{Y}_{k,t}|\tilde{Y}_{0,t}\cdots\tilde{Y}_{\ell-1,t}\\
&\qquad \tilde{Y}_0^{t-1}\cdots\tilde{Y}_k^{t-1}\tilde{Y}_{0,t+1}^n\cdots\tilde{Y}_{\ell-1,t+1}^n\tilde{\mathsf{M}}_\ell) && (86)\\
&= nH(\tilde{Y}_{\ell,T}\cdots\tilde{Y}_{k,T}|\tilde{Y}_{0,T}\cdots\tilde{Y}_{\ell-1,T}) + \log\Delta_k + \tilde{\phi}_\ell(n)\\
&\quad -nH(\tilde{Y}_{\ell,T}\cdots\tilde{Y}_{k,T}|\tilde{Y}_{0,T}\cdots\tilde{Y}_{\ell-1,T}\\
&\qquad \tilde{Y}_0^{T-1}\cdots\tilde{Y}_k^{T-1}\tilde{Y}_{0,T+1}^n\cdots\tilde{Y}_{\ell-1,T+1}^n\tilde{\mathsf{M}}_\ell T) && (87)\\
&\geq nI(\tilde{Y}_{\ell,T}\cdots\tilde{Y}_{k,T};\tilde{Y}_0^{T-1}\cdots\tilde{Y}_k^{T-1}\tilde{\mathsf{M}}_\ell T|\tilde{Y}_{0,T}\cdots\tilde{Y}_{\ell-1,T})\\
&\qquad +\log\Delta_k + \tilde{\phi}_\ell(n) && (88)\\
&= nI(\tilde{Y}_\ell\cdots\tilde{Y}_k;U_\ell|\tilde{Y}_0\cdots\tilde{Y}_{\ell-1}) + \log\Delta_k + \tilde{\phi}_\ell(n), && (89)
\end{aligned}$$

where $\tilde{\phi}_\ell(n)$ is a function so that $\frac{1}{n}\tilde{\phi}_\ell(n)$ tends to 0 as $n\to\infty$. Since Δ_k is bounded, $\frac{1}{n}\log\Delta_k$ tends to 0 as $n\to\infty$, we can conclude that

$$\lim_{n\to\infty} I(\tilde{Y}_\ell;\tilde{U}_\ell|\tilde{Y}_0\cdots\tilde{Y}_{\ell-1}) = 0, \tag{90}$$

Combined with (76), (83), and the nonnegativity of mutual information, this proves (43b) for $\ell\in\{2,\ldots,k\}$.

V. Conclusions and Outlook

We derived the strong converse result for testing against independence over a K-hop network with K decision centers and under a Markov chain assumption regarding the source sequences observed at the terminals. Our strong converse proof is based on a change of measure argument similar to Gu and Effros [5], [6] and to Tyagi and Watanabe [4]. However, to obtain the desired Markov chain, we did not rely on the variational characterization of the weak converse result, as suggested by Tyagi and Watanabe [4], nor did we use the blowing-up lemma or hypercontractivity arguments as in the proof for $K = 2$ [28]. Instead, an easier proof is proposed that relies on showing the validity of the Markov chains in the limit of infinite blocklengths. Our method can also be used for related scenarios, for example to establish the well-known strong converse for the Wyner-Ziv source coding problem, as we show in the extended version of this paper [33].

References

[1] J. Wolfowitz, *Coding Theorems of Information Theory*. Springer Berlin Heidelberg, 1978.

[2] S. L. Fong and V. Y. Tan, "A proof of the strong converse theorem for gaussian multiple access channels," *IEEE Transactions on Information Theory*, vol. 62, no. 8, pp. 4376–4394, 2016.

[3] S. L. Fong and V. Y. Tan, "A proof of the strong converse theorem for gaussian broadcast channels via the gaussian poincaré inequality," *IEEE Transactions on Information Theory*, vol. 63, no. 12, pp. 7737–7746, 2017.

[4] H. Tyagi and S. Watanabe, "Strong converse using change of measure arguments," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 689–703, 2019.

[5] W. Gu and M. Effros, "A strong converse for a collection of network source coding problems," in *2009 IEEE International Symposium on Information Theory*, pp. 2316–2320, IEEE, 2009.

[6] W. Gu and M. Effros, "A strong converse in source coding for super-source networks," in *2011 IEEE International Symposium on Information Theory Proceedings*, pp. 395–399, IEEE, 2011.

[7] Y. Oohama, "Exponential strong converse for source coding with side information at the decoder," *Entropy*, vol. 20, no. 5, p. 352, 2018.

[8] Y. Oohama, "Exponential strong converse for one helper source coding problem," *Entropy*, vol. 21, no. 6, p. 567, 2019.

[9] O. Kosut and J. Kliewer, "Strong converses are just edge removal properties," *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3315–3339, 2018.

[10] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, pp. 533–542, Jul. 1986.

[11] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 33, pp. 759–772, Nov. 1987.

[12] H. Shalaby and A. Papamarcou, "Multiterminal detection with zero-rate data compression," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 254–267, 1992.

[13] H. Shimokawa, T. Han, and S. I. Amari, "Error bound for hypothesis testing with data compression," in *Proc. ISIT*, p. 114, Jul. 1994.

[14] M. S. Rahman and A. B. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 58, pp. 6282–6303, Oct. 2012.

[15] Y. Xiang and Y. H. Kim, "Interactive hypothesis testing against independence," in *Proc. ISIT*, pp. 2840–2844, Jun. 2013.

[16] G. Katz, P. Piantanida, and M. Debbah, "Distributed binary detection with lossy data compression," *IEEE Transactions on Information Theory*, vol. 63, no. 8, pp. 5207–5227, 2017.

[17] S. Sreekumar and D. Gündüz, "Distributed hypothesis testing over noisy channels," in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 983–987, 2017.

[18] N. Weinberger, Y. Kochman, and M. Wigger, "Exponent trade-off for hypothesis testing over noisy channels," in *2019 IEEE International Symposium on Information Theory (ISIT)*, pp. 1852–1856, 2019.

[19] S. Sreekumar and D. Gündüz, "Strong converse for testing against independence over a noisy channel," in *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 1283–1288, 2020.

[20] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing based on unequal-error protection codes," *IEEE Trans. Inf. Theory*, vol. 66, pp. 4150–41820, Jul. 2020.

[21] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing with variable-length coding," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 3, pp. 681–694, 2020.

[22] W. Zhao and L. Lai, "Distributed testing with cascaded encoders," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7339–7348, 2018.

[23] S. Salehkalaibar and M. Wigger, "Distributed hypothesis testing over multi-access channels," in *2018 Information Theory and Applications Workshop (ITA)*, pp. 1–5, 2018.

[24] M. Hamad, M. Wigger, and M. Sarkiss, "Cooperative multi-sensor detection under variable-length coding," in *2020 IEEE Information Theory Workshop (ITW)*, pp. 1–5, 2021.

[25] P. Escamilla, M. Wigger, and A. Zaidi, "Distributed hypothesis testing: cooperation and concurrent detection," *IEEE Transactions on Information Theory*, vol. 66, no. 12, pp. 7550–7564, 2020.

[26] M. Hamad, M. Sarkiss, and M. Wigger, "Benefits of rate-sharing for distributed hypothesis testing." [Online]. Available: https://arxiv.org/pdf/2202.02282.pdf, 2022.

[27] S. Salehkalaibar, M. Wigger, and L. Wang, "Hypothesis testing in multi-hop networks." [Online]. Available: https://arxiv.org/abs/1708.05198v1, 2017.

[28] D. Cao, L. Zhou, and V. Y. F. Tan, "Strong converse for hypothesis testing against independence over a two-hop network," *Entropy (Special Issue on Multiuser Information Theory II)*, vol. 21, Nov. 2019.

[29] M. Hamad, M. Wigger, and M. Sarkiss, "Two-hop network with multiple decision centers under expected-rate constraints," in *2021 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2021.

[30] M. Hamad, M. Wigger, and M. Sarkiss, "Optimal exponents in cascaded hypothesis testing under expected rate constraints," in *2021 IEEE Information Theory Workshop (ITW)*, pp. 1–6, 2021.

[31] J. Liu, R. Van Handel, and S. Verdú, "Beyond the blowing-up lemma: Sharp converses via reverse hypercontractivity," in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 943–947, IEEE, 2017.

[32] K. Marton, "A simple proof of the blowing-up lemma," *IEEE Trans. Inf. Theory*, vol. 32, pp. 445–446, May 1986.

[33] M. Hamad, M. Wigger, and M. Sarkiss, "Strong converses using change of measure and asymptotic Markov chains." [Online]. Available: https://arxiv.org/, May 2022.

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

[35] I. Csiszár and J. Körner, *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press, 2011.