On Hypothesis Testing Against Conditional Independence With Multiple Decision Centers

Sadaf Salehkalaibar, *Member, IEEE*, Michèle Wigger, *Senior Member, IEEE*, and Roy Timo

***Abstract*— A distributed binary hypothesis testing problem is studied with one observer and two decision centers. Achievable type-II error exponents are derived for testing against conditional independence when the observer communicates with the two decision centers over one common and two individual noise-free bit pipes and when it communicates with them over a noisy broadcast channel. The results are based on a coding and testing scheme that splits the observations into subblocks, so that transmitter and receivers can independently apply to each subblock either Gray-Wyner coordination coding with side-information or hybrid joint source-channel coding with side-information, followed by a Neyman-Pearson test over the subblocks at the receivers. This approach allows to avoid introducing further error exponents that one would expect from the receivers' decoding operations related to binning or the noisy transmission channel. The derived exponents are shown to be optimal in some special cases when communication is over noise-free links. The results reveal a tradeoff between the type-II error exponents at the two decision centers.**

***Index Terms*— Distributed hypothesis testing, broadcast channel, testing against conditional independence, Gray-Wyner network.**

I. Introduction

CONSIDER the distributed hypothesis testing problem where a transmitter communicates with two receivers that each wishes to decide on the joint probability distribution underlying the observations at the three terminals in Fig. 1. In the scenario we consider, communication from the transmitter to the receivers either takes place over one common and two individual noise-free bit pipes or over a discrete memoryless broadcast channel (BC). For simplicity, we restrict attention to a *binary* hypothesis where either $\mathcal{H} = 0$ or $\mathcal{H} = 1$. The focus of this paper is on the asymptotic regime where the length of the observed sequences n tends to infinity and where both the type-I error probabilities (i.e., the probabilities of deciding on hypothesis 1 when $\mathcal{H} = 0$) and the type-II error probabilities (i.e., the probabilities of deciding on hypothesis 0 when $\mathcal{H} = 1$) vanish. We follow the approach in [1] and [2], and aim to quantify the fastest possible exponential decrease of the type-II error probabilities, while we allow the type-I error probabilities to vanish arbitrarily slowly. Ahlswede and Csiszar [1] and Han [2] studied the problem with only a single receiver and where communication takes place over a noise-free link. They presented general upper and lower bounds on the maximum type-II error exponents, and these bounds match when under $\mathcal{H} = 1$ the joint distribution of the observations X^n at the transmitter and Y^n at the receiver equals the product of the marginal distributions under $\mathcal{H} = 0$. This problem formulation is widely known as *testing against independence*. Rahman and Wagner [4] extended this result to a setup called *testing against conditional independence* where the receiver observes two sequences (Y^n, Z^n): under both hypotheses, sequence Z^n has the same joint distribution with the transmitter's observation X^n and the same joint distribution with Y^n; and under $\mathcal{H} = 1$, observation Y^n is conditionally independent of X^n given Z^n. Similar results were also found for scenarios with multiple transmitters [2], [4], interactive transmitters, interactive multi-round communications between nodes, successive refinement and privacy setups [5]–[8].

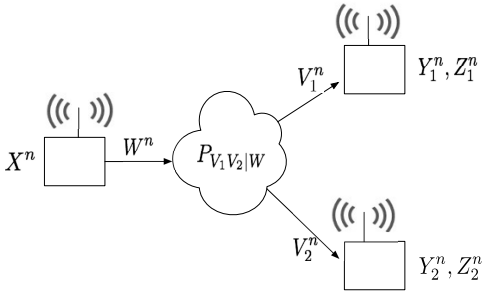

Fig. 1. Multi-terminal hypothesis testing with side information.

When testing against conditional independence, in contrast to the simpler testing against independence, a code construction with binning [3], [4] has to be used to send information from the transmitter to the receiver. The roles of the two receiver observations Z^n and Y^n decouple: Z^n plays the role of side-information for the source-coding scheme and thus reduces the required communication rate by means of binning; Y^n is solely used for hypothesis testing but not for recovering the correct codeword. Generally, the decoding operation at the receiver introduced by binning causes a second competing error exponent compared to the standard scheme

Manuscript received May 20, 2017; revised October 13, 2017 and January 4, 2018; accepted January 10, 2018. Date of publication January 26, 2018; date of current version June 14, 2018. The work of M. Wigger was supported by the ERC Grant through CTO Com. Parts of the material in this paper have been presented at *IEEE SPCOM Systems (ISWCS)*, Bangalore, India, June 2016. The associate editor coordinating the review of this paper and approving it for publication was C. Tian. *(Corresponding author: Sadaf Salehkalaibar.)*

S. Salehkalaibar is with the Department of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran 1417614418, Iran (e-mail: s.saleh@ut.ac.ir).

M. Wigger is with LTCI, Telecom ParisTech, Université Paris-Saclay, 75013 Paris, France (e-mail: michele.wigger@telecom-paristech.fr).

R. Timo is with Ericsson Research, 16483 Stockholm, Sweden (e-mail: roy.timo@ericsson.com).

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

Digital Object Identifier 10.1109/TCOMM.2018.2798659

where the codeword index is directly sent over the channel [3]. In the special case of testing against conditional independence, the second error exponent is however inactive. Rahman and Wagner [4] proposed a multi-letter extension of the binning scheme and an analysis of this scheme that directly proves the final result with the single error exponent.

A similar technique was recently applied also by Sreekuma and Gündüz [9] to derive the optimal error exponent for testing against conditional independence over a discrete memoryless channel (DMC). Their result shows that in this special case, the same error exponent can be achieved as when communication is over a noise-free link of rate equal to the capacity of the DMC. Surprisingly, there is thus no competing error exponent caused by the noisy communication channel. The work in [9] also extends some of the results to a scenario with multiple transmitters.

In contrast to these previous works, here we consider a single transmitter and *multiple receivers with different local observations*. The goal is to understand the tension on the communication channel caused by the receivers being interested in learning different informations from the transmitter.

Multiple receivers with different observations can be used to model a variety of situations:

- *Multiple Decision Centers Deciding on Different Hypotheses:* Multiple decision centers wish to decide on the same binary hypothesis but they have different local informations. This work treats the scenario where communication to the decision centers takes place over a common network.
  Example 1: Consider a road-side sensor which measures road conditions (e.g., wetness) and vehicles parameters (e.g., speed or inter-car distances). Suppose that there are two autonomous cars which measure the same parameters using the on-board sensors. Each of them verifies the accuracy of its own measurements by comparing its data to the data collected at the road-side sensors: if the sets of data are independent, then the car decides that its own data is faulty and raises an alarm (or goes to a predefined mode).
- *Single Decision Center with Uncertain Local Observation:* There is only a single decision center, and the probability distribution of the decision center's observation under each of the two hypotheses is unknown to the transmitter. In this case, the transmitter has to code for both options simultaneously, and our results determine the exponent pairs that are simultaneously achievable for the two options.
  Example 2: Consider an earthquake alert system with a remote sensor and a single local decision center that also senses ground vibrations. At unknown times of the day, there is heavy traffic close to the decision center and thus the sensed vibrations follow a different distribution. In this scenario, the information communicated from the sensor to the decision center needs to be useful under both traffic conditions. Testing against (conditional) independence can be used to distinguish vibrations that are independent at the sensor and the decision center from larger-scale seismic activities.
- *Single Decision Center Performing Two Simultaneous Tests:* Assume there is a single decision center with two sets of observations (Y_1^n, Z_1^n) and (Y_2^n, Z_2^n) that wishes to decide on two hypotheses and it suffices to take each decision only based on one of the two sets of observations. For example, because (Y_2^n, Z_2^n) is irrelevant for the first hypothesis test given (Y_1^n, Z_1^n) and the opposite holds for the second hypothesis test.
  Example 3: Consider a remote combined temperature and humidity sensor and a local weather station that also senses these two phenomena but can well separate the two measurements. For simplicity, the local station might then choose to decide on the temperature to forecast based only on its temperature measurement and to predict the humidity only based on the humidity measurement.

A main feature of the scenario that we consider is that the observer is interested in extracting and transmitting information about its observation X^n that is useful to both receivers. There is thus an inherent tradeoff in the problem, in that some information might be more beneficial for Receiver 1 than for Receiver 2 and vice versa. The goal of this paper is to shed light on this tradeoff when testing against conditional independence. As will be explained shortly, we consider communications of positive rates. Interestingly, for zero-rate communication, such a tradeoff never exists. That means, there is a single strategy at the transmitter that is optimal for both decision centers. This optimal strategy is simply the strategy from [2] and [3] where the transmitter sends a single bit indicating whether its observation is typical with respect to the distribution under $\mathcal{H} = 0$, irrespective of the distribution of the receiver observation.

One of the main contributions of this paper is to propose and analyze a coding and testing scheme for testing against conditional independence with two receivers either over a source coding network with a common and two individual noise-free bit-pipes or over a discrete memoryless BC. In both scenarios, there is a single type-II error exponent as in the scenario with a single receiver. Moreover, the decoding operations at the receivers only limit the rate of communication and the bin sizes that one is allowed to choose, but do not introduce a second competing error exponent. In our scheme, each terminal splits its observation into many subblocks and then applies either a Gray-Wyner coordination coding scheme with side-information [10], [11] or a hybrid source-channel coding scheme [14] to each subblock, and each receiver performs a Neyman-Pearson test over all these subblocks to decide on the underlying hypothesis. The idea of using block coding followed by a Neyman-Pearson test is inspired by [4] and [9]. However, here we use different block codings compared to the works in [4] and [9], as these latter only consider a single decision center. Moreover, we perform the Neyman-Pearson test over the reconstructed codeword sequences and not directly over the transmitted messages or channel outputs. This approach allows to simplify the analysis compared to an analysis that closely follows the steps proposed in [4] for the single-decision center scenario.

The second main contribution of the paper is to show that the proposed schemes achieve the optimal type-II error

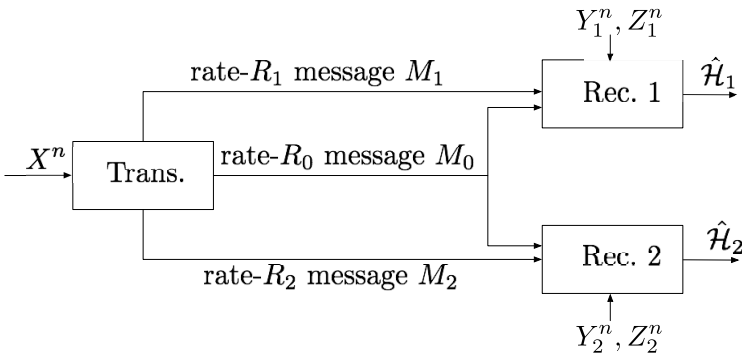

Fig. 2. Hypothesis testing over a Gray-Wyner network with side information.

exponents when: *testing against independence* over a common and two individual noise-free bit-pipes and when *testing against conditional independence* only over a common pipe under some less-noisy assumptions on the side-informations. For this latter result, a Gaussian example is presented that clearly illustrates the tradeoff on the communication channel stemming from the presence of two decision centers.

A. Notation

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 realizations, and $\mathcal{X}^n$ and $\mathcal{Y}^n$ for the corresponding n-fold Cartesian products. Sequences of random variables $(X_i, \ldots, X_j)$ and realizations $(x_i, \ldots, x_j)$ are abbreviated by X_i^j and x_i^j. When $i = 1$, then we also use the notations X^j and x^j instead of X_1^j and x_1^j.

The probability mass function (pmf) of a finite random variable X is written as P_X; the conditional pmf of X given Y is written as $P_{X|Y}$. Entropy, conditional entropy, and mutual information of random variables X and Y are denoted by $H(X)$, $H(X|Y)$, and $I(X;Y)$. Differential entropy and conditional differential entropy of continuous random variables X and Y are indicated by $h(X)$ and $h(X|Y)$. All entropies and mutual informations in this paper are meant with respect to the distribution under hypothesis $\mathcal{H} = 0$. The term $D(P||Q)$ stands for the Kullback-Leibler divergence between two pmfs P and Q over the same alphabet.

For a given P_X and a constant $\mu > 0$, let $\mathcal{T}_\mu^n(P_X) = \{x^n : |\#\{i : x_i = x\}/n - P_X(x)| \leq \mu P_X(x), \forall x \in \mathcal{X}\}$ be the set of μ-typical sequences in $\mathcal{X}^n$ [16]. Similarly, $\mathcal{T}_\mu^n(P_{X,Y})$ stands for the set of jointly μ-typical sequences.

The expectation operator is written as $\mathbb{E}[.]$. A Gaussian distribution with mean a and variance σ^2 is written as $\mathcal{N}(a, \sigma^2)$. We abbreviate *independent and identically distributed* by *i.i.d.*. Finally, the log(.)-function is taken with respect to base 2.

II. Hypothesis Testing Over a Gray-Wyner Netwrok with Side Information

Consider the distributed hypothesis testing problem with one transmitter and two receivers in Fig. 2. The transmitter observes the sequence X^n, and Receivers 1 and 2 observe Y_1^n and Y_2^n, respectively. In this model, for $i \in \{1, 2\}$, Receiver i additionally also observes a side information Z_i^n whose *pairwise* distribution with X^n and with Y_i^n does not depend on the hypothesis $\mathcal{H}$. In fact, under the null hypothesis

$$\mathcal{H} = 0\colon (X^n, Y_1^n, Y_2^n, Z_1^n, Z_2^n) \sim \text{i.i.d. } P_{XY_1Y_2Z_1Z_2}, \quad (1)$$

and under the alternative hypothesis,

$$\mathcal{H} = 1\colon (X^n, Y_1^n, Y_2^n, Z_1^n, Z_2^n) \sim \text{i.i.d. } P_{XZ_1Z_2}P_{Y_1|Z_1}P_{Y_2|Z_2}. \quad (2)$$

Here $P_{XY_1Y_2Z_1Z_2}$ is a given joint distribution over a finite product alphabet $\mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \mathcal{Z}_1 \times \mathcal{Z}_2$, and $P_{XZ_1Z_2}$, $P_{Y_1|Z_1}$ and $P_{Y_2|Z_2}$ denote its conditional marginals, i.e.,

$$\begin{aligned}
&P_{XZ_1Z_2}(x, z_1, z_2) \\
&\quad = \sum_{y_1 \in \mathcal{Y}_1, y_2 \in \mathcal{Y}_2} P_{XZ_1Z_2Y_1Y_2}(x, z_1, z_2, y_1, y_2), \\
&\qquad\qquad (x, z_1, z_2) \in \mathcal{X} \times \mathcal{Z}_1 \times \mathcal{Z}_2, \\
&P_{Y_1|Z_1}(y_1|z_1) \\
&\quad = \sum_{x \in \mathcal{X}, y_2 \in \mathcal{Y}_2, z_2 \in \mathcal{Z}_2} P_{XY_1Y_2Z_2|Z_1}(x, y_1, y_2, z_2|z_1), \\
&\qquad\qquad (y_1, z_1) \in \mathcal{Y}_1 \times \mathcal{Z}_1, \\
&P_{Y_2|Z_2}(y_2|z_2) \\
&\quad = \sum_{x \in \mathcal{X}, y_1 \in \mathcal{Y}_1, z_1 \in \mathcal{Z}_1} P_{XY_1Y_2Z_1|Z_2}(x, y_1, y_2, z_1|z_2), \\
&\qquad\qquad (y_2, z_2) \in \mathcal{Y}_2 \times \mathcal{Z}_2.
\end{aligned}$$

The test here is "against conditional independence" because Z_i has the same joint distribution with the source X under both hypotheses and because under $\mathcal{H} = 1$, Y_i is conditionally independent of X given Z_i.

The transmitter communicates with the two receivers over 1 common and 2 individual noise-free bit pipes. Specifically, it computes messages $(M_0, M_1, M_2) = \phi^{(n)}(X^n)$, using a possibly stochastic encoding function $\phi^{(n)}$ of the form $\phi^{(n)} : \mathcal{X}^n \to \{0, \ldots, 2^{nR_0}\} \times \{0, \ldots, 2^{nR_1}\} \times \{0, \ldots, 2^{nR_2}\}$, and sends message M_0 over the common pipe and messages M_1 and M_2 over the two individual pipes. For $i \in \{1, 2\}$, Receiver i observes messages M_0 and M_i and decides on the hypothesis $\mathcal{H} \in \{0, 1\}$ by means of a decoding function

$g_i^{(n)}\colon \mathcal{Y}_i^n \times \mathcal{Z}_i^n \times \{0,\dots,2^{nR_0}\} \times \{0,\dots,2^{nR_i}\} \to \{0,1\}$. It produces $\hat{\mathcal{H}}_i = g_i^{(n)}(Y_i^n, Z_i^n, M_0, M_i)$.

Definition 1: For each $\epsilon \in (0,1)$, an exponents-rates tuple $(\theta_1, \theta_2, R_0, R_1, R_2)$ is called *ϵ-achievable over the Gray-Wyner network with side information* if there exists a sequence of encoding and decoding functions $\{(\phi^{(n)}, g_1^{(n)}, g_2^{(n)})\}_{n=1}^{\infty}$ such that for $i \in \{1,2\}$ and all positive integers n, the corresponding sequences of type-I error probabilities

$$\alpha_{i,n} \stackrel{\Delta}{=} \Pr[\hat{\mathcal{H}}_i = 1 | \mathcal{H} = 0], \tag{3}$$

and type-II error probabilities

$$\beta_{i,n} \stackrel{\Delta}{=} \Pr[\hat{\mathcal{H}}_i = 0 | \mathcal{H} = 1], \tag{4}$$

satisfy

$$\alpha_{i,n} \le \epsilon,$$

and

$$-\varlimsup_{n\to\infty} \frac{1}{n} \log \beta_{i,n} \ge \theta_i.$$

Definition 2: Given nonnegative rates (R_0, R_1, R_2), define the exponents region $\mathcal{E}_{\mathrm{GW}}^{\mathrm{SI}}(R_0, R_1, R_2)$ as the closure of all non-negative exponent pairs (θ_1, θ_2) for which $(\theta_1, \theta_2, R_0, R_1, R_2)$ is ϵ-achievable over the Gray-Wyner network with side information for every $\epsilon \in (0,1)$.

Remark 1: The exponents region $\mathcal{E}_{\mathrm{GW}}^{\mathrm{SI}}(R_0, R_1, R_2)$ only depends on the marginal distributions $P_{XZ_1Z_2}$, $P_{XY_1|Z_1}$ and $P_{XY_2|Z_2}$ under both hypotheses.

A. Coding and Testing Scheme

We propose to split the block of n transmissions into B subblocks of k consecutive transmissions each such that $n = kB$. So, for each $b \in \{1,\dots,B\}$, let

$$X_b^k := (X_{(b-1)k+1}, \dots, X_{bk}), \tag{5}$$
$$Y_{i,b}^k := (Y_{i,(b-1)k+1}, \dots, Y_{i,bk}), \quad i \in \{1,2\}, \tag{6}$$
$$Z_{i,b}^k := (Z_{i,(b-1)k+1}, \dots, Z_{i,bk}), \quad i \in \{1,2\}. \tag{7}$$

For each of the subblocks, we propose to apply an independent instance of the coordination code for the Gray-Wyner network with side-information in [10], where the receivers only account for side-informations Z_1^n and Z_2^n but not for Y_1^n and Y_2^n. More specifically, choose a small real number $\mu > 0$, as well as auxiliary alphabets $\mathcal{U}_0$, $\mathcal{U}_1$, and $\mathcal{U}_2$, and a conditional joint probability distribution $P_{U_0\ U_1\ U_2|X}$ over $\mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2$ so that

$$R_0 + R_1 \ge I(U_0, U_1; X|Z_1) + \mu, \tag{8}$$
$$R_0 + R_2 \ge I(U_0, U_2; X|Z_2) + \mu, \tag{9}$$
$$R_0 + R_1 + R_2 \ge \max_{i\in\{1,2\}} I(U_0; X|Z_i) + I(U_1; X|U_0, Z_1) + I(U_2; X|U_0, Z_2) + \mu. \tag{10}$$

Construct for each block a coordination code as described in [10, Sec. V-B1)] for suitably chosen auxiliary rates $R_{0,0}, R_{0,1}, R_{0,2}, R_{1,0}, R_{1,1}, R_{2,0}, R_{2,2}, R_0', R_1', R_2' > 0$ satisfying $R_0' > \max\{R_{1,0}, R_{2,0}\}$ and Constraints (50) in [10, Appendix B].

Codebook Generation: Let P_{U_0}, $P_{U_1|U_0}$ and $P_{U_2|U_0}$ be the marginal and conditional marginal pmfs of $P_X \cdot P_{U_0U_1U_2|X}$.

For each block $b \in \{1,\dots,B\}$, generate three codebooks $\mathsf{C}_{0,b}, \mathsf{C}_{1,b}(.), \mathsf{C}_{2,b}(.)$ independently of each other in the following way. Codebook $\mathsf{C}_{0,b}$ consists of $2^{kR_{0,0}}$ superbins, each containing $2^{kR_0'}$ length-k codewords whose entries are randomly and independently generated according to the law P_{U_0}.

We make two partitions of the codewords in each superbin. In the first partition, the codewords of each superbin are assigned to $2^{kR_{1,0}}$ subbins, each containing $2^{k(R_0'-R_{1,0})}$ codewords; in the second partition they are assigned to $2^{kR_{2,0}}$ subbins, each containing $2^{k(R_0'-R_{2,0})}$ codewords. There are thus two different ways to refer to a specific codeword in $\mathsf{C}_{0,b}$. When we consider the first partition, we denote the codewords in the $m_{1,0,b} \in \{1,\dots,2^{kR_{1,0}}\}$-*th* subbin of superbin $m_{0,0,b} \in \{1,\dots,2^{kR_{0,0}}\}$ by

$$\{u_{0,b}^k(1; m_{0,0,b}, m_{1,0,b}, \ell_{1,0,b})\}_{\ell_{1,0,b}=1}^{2^{k(R_0'-R_{1,0})}};$$

when we consider the second partition, we denote the codewords in the $m_{2,0,b} \in \{1,\dots,2^{kR_{2,0}}\}$-*th* subbin of superbin $m_{0,0,b} \in \{1,\dots,2^{kR_{0,0}}\}$ by

$$\{u_{0,b}^k(2; m_{0,0,b}, m_{2,0,b}, \ell_{2,0,b})\}_{\ell_{2,0,b}=1}^{2^{k(R_0'-R_{2,0})}}.$$

Thus, here the first index indicates whether the last two indices refer to the first or the second partition of the superbins.

For $i \in \{1,2\}$, Codebook $\mathsf{C}_{i,b}(.)$ consists of $2^{kR_{0,i}}$ superbins each containing $2^{kR_{i,i}}$ subbins with $2^{kR_i'}$ codewords of length k, where all entries of all codewords are randomly and independently drawn according to P_{U_i}. For $m_{i,i,b} \in \{1,\dots,2^{kR_{i,i}}\}$, we denote the codewords in the $m_{i,i,b}$-*th* subbin of superbin $m_{0,i,b} \in 2^{kR_{0,i}}$ by

$$\{u_{i,b}^k(m_{0,i,b}, m_{i,i,b}, \ell_{i,b})\}_{\ell_{i,b}=1}^{2^{kR_i'}}.$$

All codebooks are revealed to the sender, and codebooks $\{\mathsf{C}_{0,b}, \mathsf{C}_{i,b}(.)\}$ are revealed to Receiver $i \in \{1,2\}$.

Transmitter: The transmitter first decomposes the observed source sequence $X^n = x^n$ into B blocks, each consisting of k consecutive symbols, $x_1^k, \dots, x_B^k$. For each block $b \in \{1,\dots,B\}$, it then forms a list of all the tuples of indices $(m_{0,0,b}, m_{1,0,b}, \ell_{1,0,b}, m_{0,1,b}, m_{1,1,b}, \ell_{1,b}, m_{0,2,b}, m_{2,2,b}, \ell_{2,b})$ so that the triplet of codewords $u_{0,b}^k(1; m_{0,0,b}, m_{1,0,b}, \ell_{1,0}) \in \mathsf{C}_{0,b}$, $u_{1,b}^k(m_{0,1,b}, m_{1,1,b}, \ell_{1,b}) \in \mathsf{C}_{1,b}(.)$, $u_{2,b}^k(m_{0,2,b}, m_{2,2,b}, \ell_{2,b}) \in \mathsf{C}_{2,b}(.)$ satisfies

$$(x_b^k, u_{0,b}^k(1; m_{0,0,b}, m_{1,0,b}, \ell_{1,0,b}), u_{i,b}^k(m_{0,i,b}, m_{i,i,b}, \ell_{i,b})) \in \mathcal{T}_{\mu/2}^k(P_{XU_0U_i}), \quad \in \{1,2\}. \tag{11}$$

If for some block b this list is empty, the transmitter sends the messages $m_0 = 0$, $m_1 = 0$ and $m_2 = 0$ over the bit pipes. Otherwise, it chooses for each block b the tuple $(m_{0,0,b}^\star, m_{1,0,b}^\star, \ell_{1,0,b}^\star, m_{0,1,b}^\star, m_{1,1,b}^\star, \ell_{1,b}^\star, m_{0,2,b}^\star, m_{2,2,b}^\star, \ell_{2,b}^\star)$ uniformly at random over the generated list, and sends the following messages over the bit pipes

$$m_0 = (m_{0,0,1}^\star, \dots, m_{0,0,B}^\star, m_{0,1,1}^\star, \dots, m_{0,1,B}^\star, m_{0,2,1}^\star, \dots, m_{0,2,B}^\star), \tag{12}$$
$$m_1 = (m_{1,0,1}^\star, \dots, m_{1,0,B}^\star, m_{1,1,1}^\star, \dots, m_{1,1,B}^\star), \tag{13}$$
$$m_2 = (m_{2,0,1}^\star, \dots, m_{2,0,B}^\star, m_{2,2,1}^\star, \dots, m_{2,2,B}^\star), \tag{14}$$

where for each $b \in \{1, \ldots, B\}$, the pair $(m^\star_{2,0,b}, \ell^\star_{2,0,b})$ is chosen so that $u^k_{0,b}(1; m^\star_{1,0,b}, \ell^\star_{1,0,b})$ is the same codeword as $u^k_{0,b}(2; m^\star_{2,0,b}, \ell^\star_{2,0,b})$.

Receiver i: Assume that Receiver i observes messages $M_0 = m_0$ and $M_i = m_i$ and source sequences $Y^n_i = y^n_i$ and $Z^n_i = z^n_i$. If $m_0 = m_i = 0$, Receiver i declares $\hat{\mathcal{H}}_i = 1$. Otherwise, it decomposes its observations into B blocks

$$\left\{\left(m_{0,b}, m_{i,b}, y^k_{i,b}, z^k_{i,b}\right)\right\}_{b=1}^{B}. \tag{15}$$

It further parses the common message $m_{0,b}$ as $(m_{0,0,b}, m_{0,1,b}, m_{0,2,b})$ and its private message $m_{i,b}$ as $m_{i,b} = (m_{i,0,b}, m_{i,i,b})$. Then, it seeks a codeword $u^k_{0,b}(i; m_{0,0,b}, m_{i,0,b}, \ell_{i,0,b})$ in codebook $\mathsf{C}_{0,b}$ and a codeword $u^k_{i,b}(m_{0,i,b}, m_{i,i,b}, \ell_{i,b})$ in codebook $\mathsf{C}_{i,b}(.)$ such that

$$\left(u^k_{0,b}(i; m_{0,0,b}, m_{i,0,b}, \ell_{i,0,b}), u^k_{i,b}(m_{0,i,b}, m_{i,i,b}, \ell_{i,b}), z^k_{i,b}\right) \in \mathcal{T}^k_\mu(P_{U_0U_iZ_i}). \tag{16}$$

If exactly one such pair of codewords exists, Receiver i produces the coordination sequence $\hat{u}^k_{i,b} = u^k_{i,b}(m_{0,i,b}, m_{i,i,b}, \ell_{i,b})$. Otherwise, it randomly chooses a triplet $(m^*_{0,i,b}, m^*_{i,i,b}, \ell^*_{i,b})$ and produces the coordination sequence $\hat{u}^k_{i,b} = u^k_{i,b}(m^*_{0,i,b}, m^*_{i,i,b}, \ell^*_{i,b})$. Finally, it applies a Neyman-Pearson test to decide on hypothesis $\mathcal{H}$ based on the i.i.d. sequence of tuples

$$\left\{\left(\hat{u}^k_{i,b}, y^k_{i,b}, z^k_{i,b}\right)\right\}_{b=1}^{B}, \tag{17}$$

in a way that the type-I error probability does not exceed ϵ.

B. Result on Exponents Region

The scheme described in the previous section gives the following achievable exponents region.

Let $\mathcal{E}^{\mathrm{SI,in}}_{\mathrm{GW}}(R_0, R_1, R_2)$ be given by the following:

$$\mathcal{E}^{\mathrm{SI,in}}_{\mathrm{GW}}(R_0, R_1, R_2) := \bigcup_{\substack{(U_0,U_1,U_2):\\ (U_0,U_1,U_2)\to X\\ \to(Y_1,Y_2,Z_1,Z_2)\\ R_0+R_1+R_2\geq \max_{i\in\{1,2\}}\\ I(U_0;X|Z_i)\\ +I(U_1;X|U_0,Z_1)\\ +I(U_2;X|U_0,Z_2)\\ R_0+R_1\geq I(U_1,U_0;X|Z_1)\\ R_0+R_2\geq I(U_0,U_2;X|Z_2)}} \left\{\begin{array}{l}(\theta_1,\theta_2)\colon \theta_1\geq 0, \theta_2\geq 0,\\ \theta_1 \leq I(U_1;Y_1|Z_1)\\ \theta_2 \leq I(U_2;Y_2|Z_2)\end{array}\right\}.$$

Notice that, to evaluate $\mathcal{E}^{\mathrm{SI,in}}_{\mathrm{GW}}(R_0, R_1, R_2)$ it suffices to consider auxiliary random variables U_0, U_1, U_2 over alphabets $\mathcal{U}_0$, $\mathcal{U}_1$, and $\mathcal{U}_2$ whose sizes satisfy the following three conditions: $|\mathcal{U}_0| \leq |\mathcal{X}|+3$, $|\mathcal{U}_1| \leq |\mathcal{X}|\cdot|\mathcal{U}_0|+1$, and $|\mathcal{U}_2| \leq |\mathcal{X}|\cdot|\mathcal{U}_0|+1$.

Theorem 1: The set $\mathcal{E}^{\mathrm{SI,in}}_{\mathrm{GW}}(R_0, R_1, R_2)$ is achievable, i.e.,

$$\mathcal{E}^{\mathrm{SI,in}}_{\mathrm{GW}}(R_0, R_1, R_2) \subseteq \mathcal{E}^{\mathrm{SI}}_{\mathrm{GW}}(R_0, R_1, R_2). \tag{18}$$

Proof: See Appendix A. ∎

The two next-following results show that the exponents region $\mathcal{E}^{\mathrm{SI,in}}_{\mathrm{GW}}$ coincides with the optimal exponents region $\mathcal{E}^{\mathrm{SI}}_{\mathrm{GW}}$ in some special cases.

Let

$$\mathcal{E}_{\mathrm{GW}}(R_0, R_1, R_2) := \bigcup_{\substack{(U_0,U_1,U_2):\\ (U_0,U_1,U_2)\to X\to(Y_1,Y_2)\\ R_0\geq I(U_0;X)\\ R_1\geq I(U_1;X|U_0)\\ R_2\geq I(U_2;X|U_0)}} \left\{\begin{array}{l}(\theta_1,\theta_2)\colon \theta_1\geq 0, \theta_2\geq 0,\\ \theta_1 \leq I(U_1;Y_1)\\ \theta_2 \leq I(U_2;Y_2)\end{array}\right\}. \tag{19}$$

Theorem 2: When there is no side-information, i.e., Z_1 and Z_2 are constants,

then

$$\mathcal{E}^{\mathrm{SI}}_{\mathrm{GW}}(R_0, R_1, R_2) = \mathcal{E}^{\mathrm{SI,in}}_{\mathrm{GW}}(R_0, R_1, R_2) = \mathcal{E}_{\mathrm{GW}}(R_0, R_1, R_2). \tag{20}$$

Proof: Achievability follows by specializing Theorem 1 to Z_1 and Z_2 constant. The converse can be obtained from the converse in [17] where one has to include U_0 into U_1. ∎

In the above Theorem 2 it suffices to consider auxiliary random variables U_0, U_1, and U_2 over alphabets $\mathcal{U}_0$, $\mathcal{U}_1$, and $\mathcal{U}_2$ whose sizes satisfy:

$$|\mathcal{U}_0| \leq |\mathcal{X}| + 2, \tag{21}$$

$$|\mathcal{U}_j| \leq |\mathcal{X}|\cdot|\mathcal{U}_0| + 1, \qquad j \in \{1, 2\}. \tag{22}$$

This follows by simple applications of Caratheodory's theorem.

Theorem 3: Let Z_2 be a constant and Z_1 less noisy than Y_2, i.e., let for all auxiliary random variables U satisfying the Markov chain $U \to X \to (Y_1, Y_2, Z_1)$ the following inequality hold:

$$I(U; Z_1) \geq I(U; Y_2). \tag{23}$$

Then:

$$\mathcal{E}^{\mathrm{SI}}_{\mathrm{GW}}(R_0, R_1 = 0, R_2 = 0) = \mathcal{E}^{\mathrm{SI,in}}_{\mathrm{GW}}(R_0, R_1 = 0, R_2 = 0). \tag{24}$$

Proof: Achievability follows by Theorem 1. The converse is proved in Appendix B. ∎

C. An Example

Theorem 3 was stated for discrete memoryless sources. It can be shown that it remains valid also when sources are memoryless and jointly Gaussian [16, Ch. 3].

Consider the following scenario. Under both hypotheses, $X \sim \mathcal{N}(0, 1)$ and $Z_1 = X + N_z$, where $N_z \sim \mathcal{N}(0, \sigma_z^2)$ is independent of X. Moreover, under hypothesis

$$\mathcal{H} = 0\colon \quad Y_1 = X + Z_1 + N_1, \tag{25}$$

$$Y_2 = Z_1 + N_2, \tag{26}$$

where $N_1 \sim \mathcal{N}(0, \sigma_1^2)$ and $N_2 \sim \mathcal{N}(0, \sigma_2^2)$ are independent of each other and of (X, Z_1), and under hypothesis

$$\mathcal{H} = 1\colon \quad Y_1 = X' + \frac{2 + \sigma_z^2}{1 + \sigma_z^2}\cdot Z_1 + N_1, \tag{27}$$

$$Y_2 = Z_1' + N_2, \tag{28}$$

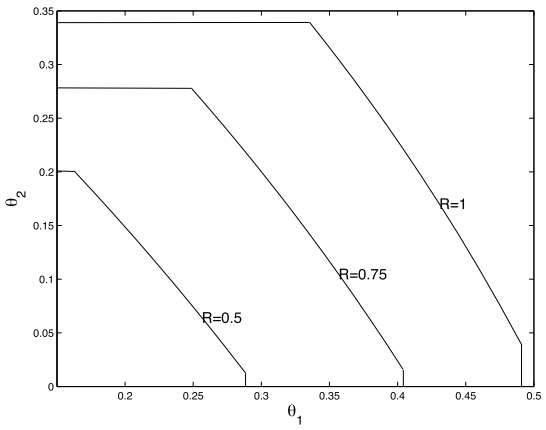

Fig. 3. Exponents region for $\sigma_z^2 = 0.7$, $\sigma_1^2 = 0.2$ and $\sigma_2^2 = 0.3$.

where $X' \sim \mathcal{N}(0, \frac{\sigma_z^2}{1+\sigma_z^2})$ and $Z_1' \sim \mathcal{N}(0, 1+\sigma_z^2)$ are independent of each other and of the tuple (X, Z_1, N_1, N_2).

The described scenario satisfies the less noisy condition in (23). By Theorem 3, when restricting to $R_1 = R_2 = 0$, for this example, the region $\mathcal{E}_{\text{GW}}^{\text{SI}}$ equals $\mathcal{E}_{\text{GW}}^{\text{SI,in}}$. As is proved in Appendix IV, the exponents region $\mathcal{E}_{\text{GW}}^{\text{SI}}(R_0, R_1 = 0, R_2 = 0)$ evaluates to the set of all nonnegative exponent pairs (θ_1, θ_2) that satisfy

$$\theta_1 \le \frac{1}{2}\log\left(\frac{\sigma_z^2 + \sigma_1^2(1+\sigma_z^2)}{2^{2\tilde{\alpha}}\sigma_z^2 + \sigma_1^2(1+\sigma_z^2)}\right), \tag{29a}$$

$$\theta_2 \le \frac{1}{2}\log\left(\frac{1+\sigma_z^2+\sigma_2^2}{2^{-2(\tilde{\alpha}+R_0)}(1+\sigma_z^2)+\sigma_2^2}\right), \tag{29b}$$

for some $\tilde{\alpha} \in [-R_0, 0]$.

The boundary of the exponents region $\mathcal{E}_{\text{GW}}^{\text{SI}}(R_0, R_1 = 0, R_2 = 0)$ is illustrated in Fig. 3 for different values of the rate R_0. Generally, on this boundary $\theta_1 > \theta_2$, because Receiver 1 has the additional side-information Z_1. One observes a trade-off between the two exponents θ_1 and θ_2, which is captured by the parameter $\tilde{\alpha}$ in (29). In other words, having a larger exponent θ_1 comes at the expense of a smaller exponent θ_2, and vice versa.

III. Hypothesis Testing over Noisy Channels

This section considers hypothesis testing over a discrete memoryless BC $(\mathcal{W}, \mathcal{V}_1, \mathcal{V}_2, P_{V_1V_2|W})$, where $\mathcal{W}$ denotes the finite channel input alphabet, $\mathcal{V}_1$ and $\mathcal{V}_2$ the finite channel output alphabets at Receivers 1 and 2, and $P_{V_1V_2|W}$ the BC transition pmf. The setup is illustrated in Fig. 4. The transmitter observes a sequence X^n and produces its channel inputs $W^n := (W_1, \ldots, W_n)$ as $W^n = \Phi^{(n)}(X^n)$ by means of a possibly stochastic encoding function $\Phi^{(n)}\colon \mathcal{X}^n \to \mathcal{W}^n$. Receivers 1 and 2 observe the corresponding channel outputs $V_1^n := (V_{1,1}, \ldots, V_{1,n})$ and $V_2^n := (V_{2,1}, \ldots, V_{2,n})$, as well as the source sequences (Y_1^n, Z_1^n) and (Y_2^n, Z_2^n) defined in the previous section. For $i \in \{1, 2\}$, Receiver i decides on the hypothesis $\mathcal{H} \in \{0, 1\}$ by means of a decoding function $g_i^{(n)}\colon \mathcal{Y}_i^n \times \mathcal{Z}_i^n \times \mathcal{V}_i^n \to \{0, 1\}$. It produces $\hat{\mathcal{H}}_i = g_i^{(n)}(Y_i^n, Z_i^n, V_i^n)$.

As in the previous section, assume that under hypothesis

$$\mathcal{H} = 0\colon (X^n, Y_1^n, Y_2^n, Z_1^n, Z_2^n) \sim \text{i.i.d. } P_{XY_1Y_2Z_1Z_2}, \tag{30}$$

and under hypothesis

$$\mathcal{H} = 1\colon \quad (X^n, Y_1^n, Y_2^n, Z_1^n, Z_2^n) \sim \text{i.i.d. } P_{XZ_1Z_2}P_{Y_1|Z_1}P_{Y_2|Z_2}. \tag{31}$$

Definition 3: For each $\epsilon \in (0, 1)$, an exponent pair (θ_1, θ_2) is called *ϵ-achievable over a BC with side information* if there exists a sequence of encoding and decoding functions $\{(\Phi^{(n)}, g_1^{(n)}, g_2^{(n)})\}_{n=1}^{\infty}$ such that for $i \in \{1, 2\}$ and all positive integers n, the corresponding sequences of type-I and type-II error probabilities satisfy

$$\alpha_{i,n} \le \epsilon,$$

and

$$-\overline{\lim_{n\to\infty}}\frac{1}{n}\log\beta_{i,n} \ge \theta_i,$$

where $\alpha_{i,n}$ and $\beta_{i,n}$ are defined in (3) and (4).

Definition 4: Define the exponents region $\mathcal{E}_{BC}^{\text{SI}}$ as the closure of all non-negative exponent pairs (θ_1, θ_2) for which (θ_1, θ_2) is ϵ-achievable over the BC with side information for every $\epsilon \in (0, 1)$.

A. Coding and Testing Scheme

Fix $\mu > 0$, sufficiently large positive integers k and B, and a joint conditional distribution $P_{U_0U_1U_2|X}$ over finite auxiliary alphabets $\mathcal{U}_0$, $\mathcal{U}_1$ and $\mathcal{U}_2$. Consider also nonnegative rates R_0, R_1, R_2 that satisfy

$$R_0 + R_1 \le I(U_1, U_0; V_1, Z_1), \tag{32}$$

$$R_0 + R_2 \le I(U_2, U_0; V_2, Z_2), \tag{33}$$

$$R_1 \le I(U_1; V_1, Z_1|U_0), \tag{34}$$

$$R_2 \le I(U_2; V_2, Z_2|U_0), \tag{35}$$

$$R_0 > I(U_0; X), \tag{36}$$

$$R_1 > I(U_1; X|U_0), \tag{37}$$

$$R_2 > I(U_2; X|U_0), \tag{38}$$

$$R_1 + R_2 > I(U_1, U_2; X|U_0) + I(U_1; U_2|U_0). \tag{39}$$

Finally, fix a function $f\colon \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{X} \to \mathcal{W}$.

Code Construction: For each block $b \in \{1, \ldots, B\}$, randomly generate a codebook $\mathsf{C}_{0,b} = \{U_{0,b}^k(m_{0,b}) : m_{0,b} \in \{1, \ldots, 2^{kR_0}\}\}$ by drawing each entry of the n-length codeword $U_{0,b}^k(m_{0,b})$ i.i.d. according to the pmf P_{U_0}. Moreover, for each index $m_{0,b}$ and $i \in \{1, 2\}$, randomly generate a codebook $\mathsf{C}_{i,b}(m_{0,b}) := \{U_{i,b}^k(m_{i,b}|m_{0,b})\colon m_{i,b} \in \{1, \ldots, 2^{kR_i}\}\}$ by drawing each entry of the k-length codeword $U_{i,b}^k(m_{i,b}|m_{0,b})$ i.i.d. according to the conditional pmf $P_{U_i|U_0}(.|U_{0,b,j}(m_{0,b}))$, where $U_{0,b,j}(m_{0,b})$ denotes the *j-th* symbol of $U_{0,b}^k(m_{0,b})$. Reveal the realizations $\{\mathcal{C}_{0,b}\}$, $\{\mathcal{C}_{1,b}(\cdot)\}$ and $\{\mathcal{C}_{2,b}(\cdot)\}$ of the randomly generated codebooks to all terminals.

Transmitter: It observes a source sequence x^n and splits it into B subblocks $x^n = (x_1^k, \ldots, x_B^k)$ as in (5). For each

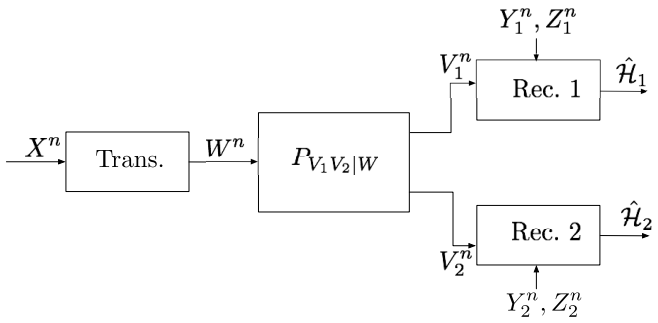

Fig. 4. Hypothesis testing over a BC.

block b, it looks for a triple of indices $(m_{0,b}, m_{1,b}, m_{2,b}) \in \{1, \ldots, 2^{kR_0}\} \times \{1, \ldots, 2^{kR_1}\} \times \{1, \ldots, 2^{kR_2}\}$ such that

$$(x_b^k, u_{0,b}^k(m_{0,b}), u_{1,b}^k(m_{1,b}|m_{0,b}), u_{2,b}^k(m_{2,b}|m_{0,b})) \in \mathcal{T}_{\mu/2}^k(P_{XU_0U_1U_2}), \quad (40)$$

where $u_{0,b}^k(m_{0,b})$, $u_{1,b}^k(m_{1,b}|m_{0,b})$ and $u_{2,b}^k(m_{2,b}|m_{0,b})$ are codewords from the chosen codebooks $\mathcal{C}_{0,b}$, $\{\mathcal{C}_{1,b}(\cdot)\}$ and $\{\mathcal{C}_{2,b}(\cdot)\}$. If the typicality test is successful, the transmitter picks one of the triples satisfying the test at random. Otherwise, it picks a triple $(m_{0,b}, m_{1,b}, m_{2,b})$ uniformly at random over $\{1, \ldots, 2^{kR_0}\} \times \{1, \ldots, 2^{kR_1}\} \times \{1, \ldots, 2^{kR_2}\}$. It finally sends the k inputs

$$w_{(b-1)k+j} = f(u_{0,b,j}(m_{0,b}), u_{1,b,j}(m_{1,b}|m_{0,b}), u_{2,b,j}(m_{2,b}|m_{0,b}), x_{(b-1)k+j}), \quad j \in \{1, \ldots, k\}, \quad (41)$$

over the channel.

Receiver $i \in \{1, 2\}$: Assume that it observes the sequence of channel outputs $v_{i,b}^n$ and the source sequences $y_{i,b}^n$ and $z_{i,b}^n$. It looks for a pair of indices $(\hat{m}_{0,b}, \hat{m}_{i,b})$ such that

$$(u_{i,b}^k(\hat{m}_{i,b}|\hat{m}_{0,b}), v_{i,b}^k, z_{i,b}^k) \in \mathcal{T}_\mu^k(P_{U_iV_iZ_i}), \quad (42)$$

and picks one of these pairs at random. If no such pair can be found, pick $(\hat{m}_{0,b}, \hat{m}_{i,b})$ uniformly over $\{1, \ldots, 2^{kR_0}\} \times \{1, \ldots, 2^{kR_1}\}$. For the chosen $(\hat{m}_{0,b}, \hat{m}_{i,b})$, set

$$\hat{u}_{i,b}^k := u_{i,b}^k(\hat{m}_{0,b}, \hat{m}_{i,b}). \quad (43)$$

Receiver i then decomposes its observations $(y_{i,b}^k, z_{i,b}^k)$ as in (6) and (7) and performs a Neyman-Pearson test on the B i.i.d. blocks,

$$\left\{(\hat{u}_{i,b}^k, v_{i,b}^k, y_{i,b}^k, z_{i,b}^k)\right\}_{b=1}^B,$$

in a way that the type-I error probability does not exceed ϵ.

B. Exponents Region

Let $\mathcal{E}_{BC}^{\text{hyb}}$ be given by the following:

$$\mathcal{E}_{BC}^{\text{hyb}} = \bigcup_{(U_0,U_1,U_2)} \left\{ \begin{array}{r} (\theta_1, \theta_2) \colon \theta_1 \geq 0, \ \ \theta_2 \geq 0, \\ \theta_1 \leq I(U_1; Y_1|Z_1) \\ \theta_2 \leq I(U_2; Y_2|Z_2) \end{array} \right\},$$

where the union is taken over all pmfs $P_{U_0U_1U_2W|X}$ that satisfy the following Markov chains

$$(U_0, U_1, U_2) \to X \to (Y_1, Y_2, Z_1, Z_2), \quad (44)$$

$$(Y_1, Y_2, Z_1, Z_2) \to (U_0, U_1, U_2, X) \to W \to (V_1, V_2), \quad (45)$$

and the mutual information constraints

$$I(U_1, U_0; X|Z_1) \leq I(U_1, U_0; V_1|Z_1), \quad (46a)$$

$$I(U_2, U_0; X|Z_2) \leq I(U_2, U_0; V_2|Z_2), \quad (46b)$$

$$I(U_1; X|Z_1, U_0) \leq I(U_1; V_1|Z_1, U_0), \quad (46c)$$

$$I(U_2; X|Z_2, U_0) \leq I(U_2; V_2|Z_2, U_0), \quad (46d)$$

$$I(U_0, U_1; X|Z_1) + I(U_2; X|Z_2, U_0) + I(U_1; U_2|U_0) \leq I(U_0, U_1; V_1|Z_1) + I(U_2; V_2|Z_2, U_0), \quad (46e)$$

$$I(U_0, U_2; X|Z_2) + I(U_1; X|Z_1, U_0) + I(U_1; U_2|U_0) \leq I(U_1; V_1|Z_1, U_0) + I(U_0, U_2; V_2|Z_2), \quad (46f)$$

$$I(U_1; X|Z_1, U_0) + I(U_2; X|Z_2, U_0) + I(U_1; U_2|U_0) \leq I(U_1; V_1|Z_1, U_0) + I(U_2; V_2|Z_2, U_0), \quad (46g)$$

$$I(U_1, U_0; X|Z_1) + I(U_2, U_0; X|Z_2) + I(U_1; U_2|U_0) \leq I(U_1, U_0; V_1|Z_1) + I(U_2, U_0; V_2|Z_2), \quad (46h)$$

for some function $f : \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 \times \mathcal{X} \to \mathcal{W}$ where $W = f(U_0, U_1, U_2, X)$.

Theorem 4: The exponents region $\mathcal{E}_{BC}^{\text{hyb}}$ is achievable, i.e.,

$$\mathcal{E}_{BC}^{\text{hyb}} \subseteq \mathcal{E}_{BC}^{\text{SI}}.$$

Proof: The region is achieved by the coding and testing scheme described in the previous subsection. This is proved in Appendix D. ∎

To evaluate the region $\mathcal{E}_{BC}^{\text{hyb}}$, it suffices to consider auxiliaries whose alphabets satisfy the following two conditions: $|\mathcal{U}_0| \leq |\mathcal{X}| + 8$, $|\mathcal{U}_1| \leq |\mathcal{X}| \cdot |\mathcal{U}_0| + 3$ and $|\mathcal{U}_2| \leq |\mathcal{X}| \cdot |\mathcal{U}_0| + 3$.

The exponents region $\mathcal{E}_{BC}^{\text{hyb}}$ is achieved by means of hybrid joint source-channel coding with side-information. The constraints in (44) ensure that the receivers can decode their intended hybrid coding codewords; a U_0-codeword is decoded at both receivers and a U_i-codeword at Receiver i only.

These codewords are then used at the receivers for testing against conditional independence, see the exponents expression in (44). Notice that hybrid joint source-channel coding also includes separate source-channel coding as a special case [14]. In fact, the separate scheme's exponents region can be derived by considering $U_0 = (W_0, \tilde{U}_0)$ and $U_i = (W_i, \tilde{U}_i)$, for $i \in \{1, 2\}$, where $(\tilde{U}_0, \tilde{U}_1, \tilde{U}_2, W_0, W_1, W_2)$ are auxiliary random variables which satisfy the Markov chains $(\tilde{U}_0, \tilde{U}_1, \tilde{U}_2) \to X \to (Z_1, Z_2)$ and $(W_0, W_1, W_2) \to W \to (V_1, V_2)$ and the tuple (W_0, W_1, W_2) is independent of $(\tilde{U}_0, \tilde{U}_1, \tilde{U}_1, X, Y_1, Z_1, Y_2, Z_2)$.

This theorem recovers the optimal error exponent for hypothesis testing over a point-to-point channel found in [9]. It can be verified that the optimal error exponent of [9] for the discrete memoryless channel from W to V_1 can be recovered by specializing Theorem 4 to U_0, U_2 constants and $U_1 = (\tilde{U}, W)$ with W independent of $(\tilde{U}, X, Y_1, Z_1)$.

C. An Example

We investigate the achievable exponent region of Theorem 4 by means of an example. Reconsider the first example in Section II-C, but where now communication takes place over a Gaussian BC. Since the exponents region depends on the BC transition law only through the conditional marginals $P_{V_1|W}$ and $P_{V_2|W}$, we assume that the Gaussian BC is degraded and described as follows:

$$V_1 = W + T_1, \tag{47}$$

$$V_2 = V_1 + T_2, \tag{48}$$

where T_1 and T_2 are independent Gaussian random variables of variances r_1^2 and $r_2^2 - r_1^2$ ($r_2^2 \geq r_1^2$). The input W is subject to an expected power constraint $\mathbb{E}[|W|^2] \leq 1$.

Likewise to the first example, we choose the auxiliaries U_0 and U_1 jointly Gaussian with X so that $X = U_1 + Q_1$ and $U_1 = U_0 + Q_0$, and we choose $U_2 = U_0$. Due to the degradedness of the channel, for such a choice of auxiliaries (i.e., when $U_2 = U_0$) constraints (46) simplify to the two constraints

$$I(U_0; X) \leq I(U_0; V_2), \tag{49}$$

$$I(U_1; X|Z_1, U_0) \leq I(U_1; V_1|Z_1, U_0). \tag{50}$$

Let Q_0, Q_1, U_0 be independent zero-mean Gaussian random variables of variances $\sigma_{q_0}^2$, $\sigma_{q_1}^2$, and $1 - \sigma_{q_0}^2 - \sigma_{q_1}^2$ so that $X = Q_0 + Q_1 + U_0$ and $U_1 = U_0 + Q_0$. Then, set the channel input to $W = \alpha U_0 + \beta U_1$ for some parameters $\alpha, \beta \geq 0$ satisfying

$$(\alpha + \beta)^2(1 - \sigma_{q_0}^2 - \sigma_{q_1}^2) + \beta^2\sigma_{q_0}^2 = 1. \tag{51}$$

Specializing the achievable exponents region $\mathcal{E}_{BC}^{\text{hyb}}$ to the proposed choices, proves achievability of all nonnegative pairs (θ_1, θ_2) that satisfy

$$\theta_1 \leq \frac{1}{2}\log\left(\left(\sigma_1^2 + \frac{\sigma_z^2}{1+\sigma_z^2}\right)\cdot\left(\frac{\sigma_{q_1}^2 + \sigma_z^2}{\sigma_{q_1}^2(\sigma_1^2 + \sigma_z^2) + \sigma_1^2\sigma_z^2}\right)\right), \tag{52}$$

$$\theta_2 \leq \frac{1}{2}\log\left(\frac{1 + \sigma_z^2 + \sigma_2^2}{\sigma_{q_0}^2 + \sigma_{q_1}^2 + \sigma_z^2 + \sigma_2^2}\right), \tag{53}$$

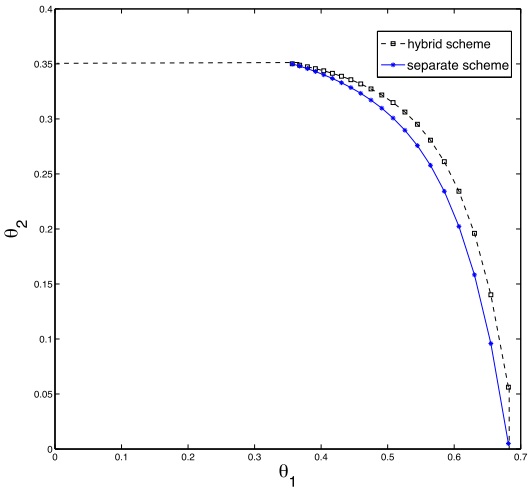

Fig. 5. Achievable exponents regions for $\sigma_z^2 = 0.7$, $\sigma_1^2 = 0.2$, $\sigma_2^2 = 0.3$, $r_1^2 = 0.1$, $r_2^2 = 0.3$.

for some $\sigma_{q_0}^2, \sigma_{q_1}^2 \in [0, 1]$, $\beta > 0$ so that

$$\sigma_{q_0}^2 + \sigma_{q_1}^2 \leq 1, \tag{54}$$

and

$$\frac{1}{\sigma_{q_0}^2 + \sigma_{q_1}^2} \leq \frac{1 + r_2^2}{\beta^2\sigma_{q_0}^2 + r_2^2}, \tag{55}$$

$$\frac{1 + \frac{\sigma_z^2}{\sigma_{q_1}^2}}{1 + \frac{\sigma_z^2}{\sigma_{q_0}^2 + \sigma_{q_1}^2}} \leq 1 + \frac{\sigma_{q_0}^2}{\sigma_{q_1}^2 + \sigma_z^2} + \frac{\beta^2\sigma_{q_0}^2}{r_1^2}. \tag{56}$$

The boundary of the achievable exponents region $\mathcal{E}_{BC}^{\text{hyb}}$ is illustrated in Fig. 5 for a setup parametrized by $\sigma_z^2 = 0.7$, $\sigma_1^2 = 0.2$, $\sigma_2^2 = 0.3$, $r_1^2 = 0.1$ and $r_2^2 = 0.3$. One observes a trade-off between the two exponents θ_1 and θ_2. Comparing this exponents region with the region shown in Figure 3 for the noiseless channel, we observe that the asymmetric channel (different noise variances at the different receivers) changes the nature of this tradeoff. The second line shown in Fig. 5 depicts the boundary of the exponents region that is achieved by a separation based scheme that combines the Gray-Wyner coordination coding with side-information from the previous section with a superposition code for the Gaussian broadcast channel. As it can be seen, the exponents region achieved by this separate coding and testing scheme is strictly smaller than the exponents region of our joint coding and testing scheme.

IV. Conclusion and Discussion

This paper considers a distributed binary hypothesis testing problem in a one-observer, two-decision center setup. Achievable error exponents are presented for *testing against conditional independence* when communication from the observer to the centers is over one common and two individual noise-free bit-pipes and when communication is over a BC. To this end, we presented coding and testing schemes where:

- all terminals split their observations into many subblocks;
- transmitter and receivers apply a Gray-Wyner coordination code with side-information [10] or hybrid joint source-channel coding with side-information for a BC;

- the receivers apply a Neyman-Pearson test to the i.i.d. subblocks of side-information and reconstructed source sequences.

Similarly to [4] and [9], in the above approach, the "multi-letter" decision over subblocks avoids introducing a competing error exponent due to the binning or the channel decoding procedure.

The derived type-II error exponents are optimal when *testing against independence* over a common and two individual noise-free bit pipes, and when *testing against conditional independence* over a single noise-free bit pipe if some of the receiver side-informations are less noisy. An explicit characterization of this latter optimal error exponent is given for a Gaussian example. This characterization clearly reveals a tradeoff between the error exponents achieved at the two decision centers.

Appendix A
Proof of Theorem 1

The proof is based on the scheme in Section II-A which we analyze in the following.

Analysis: From the way we constructed the Neyman-Pearson tests, it immediately follows that the type-I error probabilities at the two receivers cannot exceed ϵ. We turn our attention to the type-II error probabilities. Notice that the analysis in [10, Th. 2] is easily modified to show that for each $b \in \{1, \ldots, B\}$ and $i \in \{1, 2\}$:

$$\Pr[(X_b^k, \hat{U}_{i,b}^k, Z_{i,b}^k) \in \mathcal{T}_\mu^k(P_{U|X}P_{XZ})] > 1 - \mu, \qquad (57)$$

for sufficiently large k. In fact, it suffices to add the sequence $Z_{i,b}^k$ into the typicality test defining event $\mathcal{E}_{3,i}$ in [10, Appendix B]. Thus, by the conditional typicality lemma [16], under the null-hypothesis $\mathcal{H} = 0$, also

$$\Pr[(X_b^k, \hat{U}_{i,b}^k, Z_{i,b}^k, Y_{i,b}^k) \in \mathcal{T}_\mu^k(P_{U|X}P_{XYZ})] > 1 - \mu. \quad (58)$$

Now, recall that each Receiver i only declares $\hat{\mathcal{H}}_i = 0$ if the applied Neyman-Pearson test produces 0. Since for each $i \in \{1, 2\}$:

$$\begin{aligned} &\text{under } \mathcal{H} = 0\text{:} \\ &\{\hat{U}_{i,b}^k, Y_{i,b}^k, Z_{i,b}^k\}_{b=1}^B \text{ is i.i.d.} \sim P_{\hat{U}_i^k Y_i^k Z_i^k}, \end{aligned} \qquad (59a)$$

and

$$\begin{aligned} &\text{under } \mathcal{H} = 1\text{:} \\ &\{\hat{U}_{i,b}^k, Y_{i,b}^k, Z_{i,b}^k\}_{b=1}^B \text{ is i.i.d.} \sim P_{\hat{U}_i^k Z_i^k} P_{Y_i^k|Z_i^k}, \end{aligned} \qquad (59b)$$

the Chernoff-Stein Lemma [22] can be applied to bound the probabilities of type-II error. Thus, for sufficiently large k:

$$\begin{aligned} -\frac{1}{n}\log\beta_{i,n} &\geq \frac{1}{k} D\big(P_{\hat{U}_i^k Y_i^k Z_i^k|\mathcal{H}=0} \,\big\|\, P_{\hat{U}_i^k Y_i^k Z_i^k|\mathcal{H}=1}\big) - \mu \\ &\overset{(a)}{=} \frac{1}{k} I\big(\hat{U}_i^k; Y_i^k \big| Z_i^k\big) - \mu \\ &= H\big(Y_i \big| Z_i\big) - \frac{1}{k} H\big(Y_i^k \big| \hat{U}_i^k, Z_i^k\big) - \mu, \end{aligned} \qquad (60)$$

where mutual informations and entropies have to be computed according to the joint pmf $P_{\hat{U}_i^k Y_i^k Z_i^k}$ under $\mathcal{H} = 0$, and Equality (a) holds by (59). We continue by defining the event

$$\mathcal{E}_{V,i} \triangleq \{(\hat{U}_i^k, Y_i^k, Z_i^k) \in \mathcal{T}_\mu^k(P_{U_i Y_i Z_i})\},$$

and let $\mathbb{1}_V$ be the indicator function of $\mathcal{E}_{V,i}$.

The second term on the RHS of (60) can then be upper bounded as:

$$\begin{aligned} &H(Y_i^k|Z_i^k, \hat{U}_i^k) \\ &\quad= H(Y_i^k, \mathbb{1}_V|Z_i^k, \hat{U}_i^k) \\ &\quad= H(Y_i^k|Z_i^k, \hat{U}_i^k, \mathbb{1}_V) + H(\mathbb{1}_V|Z_i^k, \hat{U}_i^k) \\ &\quad\overset{(a)}{\leq} H(Y_i^k|Z_i^k, \hat{U}_i^k, \mathbb{1}_V) + 1 \\ &\quad\overset{(b)}{\leq} H(Y_i^k|Z_i^k, \hat{U}_i^k, \mathbb{1}_V = 1) + k\log|\mathcal{Y}_i|\cdot\mu + 1 \\ &\quad= \sum_{\substack{(u_i^k, z_i^k) \\ \in \mathcal{T}_\mu^k(P_{U_i Z_i})}} \Big[\Pr[Z_i^k = z_i^k, \hat{U}_i^k = u_i^k|\mathbb{1}_V = 1] \\ &\qquad\qquad \cdot H(Y_i^k|Z_i^k = z_i^k, \hat{U}_i^k = u_i^k, \mathbb{1}_V = 1)\Big] \\ &\qquad + k\log|\mathcal{Y}_i|\cdot\mu + 1 \\ &\quad\overset{(c)}{\leq} \sum_{\substack{(u_i^k, z_i^k) \\ \in \mathcal{T}_\mu^k(P_{U_i Z_i})}} \Big[\Pr[Z_i^k = z_i^k, \hat{U}_i^k = u_i^k|\mathbb{1}_V = 1] \\ &\qquad\qquad \cdot \log(|\mathcal{T}_\mu^k(Y_i^k|u_i^k, z_i^k)|)\Big] \\ &\qquad + k\log|\mathcal{Y}_i|\cdot\mu + 1 \\ &\quad\overset{(d)}{\leq} \sum_{\substack{(u_i^k, z_i^k) \\ \in \mathcal{T}_\mu^k k(P_{U_i Z_i})}} \Big[\Pr[Z_i^k = z_i^k, \hat{U}_i^k = u_i^k|\mathbb{1}_V = 1] \\ &\qquad\qquad \cdot (kH(Y_i|Z_i, U_i) + k\delta(\mu))\Big] \\ &\qquad + k\log|\mathcal{Y}_i|\cdot\mu + 1 \\ &\quad= kH(Y_i|Z_i, U_i) + k\delta(\mu) + k\log|\mathcal{Y}_i|\cdot\mu + 1. \end{aligned} \qquad (61)$$

The steps leading to (61) are justified as follows:

- (a) follows from the fact that $H(\mathbb{1}_V|Z_1^k, \hat{U}_1^k) \leq 1$ because $\mathbb{1}_V$ is a binary random variable;
- (b) follows by (58), because $\Pr[\mathbb{1}_V = 1] \leq 1$, and because $H(Y_1^k|Z_1^k, \hat{U}_1^k, \mathbb{1}_V = 0) \leq k\log|\mathcal{Y}_1|$;
- (c) follows because entropy is maximized by a uniform distribution,
- (d) follows by bounding the size of the typical set [16] where $\delta(\mu)$ is a function that goes to 0 as $\mu \to 0$.

We combine (60) with (61) to obtain that for any choice of $\mu > 0$ and sufficiently large k, B:

$$-\frac{1}{n}\log\beta_{i,n} \geq I(U_i; Y_i|Z_i) - \delta'(\mu), \qquad i \in \{1, 2\}, \quad (62)$$

where $\delta'(\mu)$ is a function that tends to 0 as $\mu \to 0$. Taking $\mu \to 0$ proves Theorem 1.

Appendix B
Converse Proof to Theorem 3

Fix a sequence of encoding and decoding functions $\{\phi^{(n)}, g_1^{(n)}, g_2^{(n)}\}$ so that the inequalities in Definition 1 hold

for sufficiently large blocklengths n. Fix also such a sufficiently large n. Then, define $U_{0,t} \stackrel{\Delta}{=} (M_0, Z_1^{t-1})$ and $U_{1,t} \stackrel{\Delta}{=} (X^{t-1}, Z_{1,t+1}^n)$. Following similar steps as in [23], it can be shown that

$$D(P_{M_0 Y_1^n Z_1^n|\mathcal{H}=0}||P_{M_0 Y_1^n Z_1^n|\mathcal{H}=1}) \geq -(1-\epsilon)\log\beta_{1,n}.$$

Therefore, the type-II error probability at Receiver 1 can be upper bounded as

$$\begin{aligned}
&-\frac{1}{n}\log\beta_{1,n}\\
&\leq \frac{1}{n(1-\epsilon)} D(P_{M_0 Y_1^n Z_1^n|\mathcal{H}=0}||P_{M_0 Y_1^n Z_1^n|\mathcal{H}=1})\\
&\stackrel{(a)}{=} \frac{1}{n(1-\epsilon)} I(M_0; Y_1^n|Z_1^n)\\
&= \frac{1}{n(1-\epsilon)} \sum_{t=1}^n I(M_0; Y_{1,t}|Y_1^{t-1}, Z_1^n)\\
&\stackrel{(b)}{\leq} \frac{1}{n(1-\epsilon)} \sum_{t=1}^n I(M_0, Y_1^{t-1}, Z_1^{t-1}, Z_{1,t+1}^n; Y_{1,t}|Z_{1,t})\\
&\stackrel{(c)}{\leq} \frac{1}{n(1-\epsilon)} \sum_{t=1}^n I(M_0, X^{t-1}, Z_1^{t-1}, Z_{1,t+1}^n; Y_{1,t}|Z_{1,t})\\
&= \frac{1}{n(1-\epsilon)} \sum_{t=1}^n I(U_{0,t}, U_{1,t}; Y_{1,t}|Z_{1,t}),
\end{aligned}$$

where (a) follows because under hypothesis $\mathcal{H} = 1$ and given Z_1^n, the sequence Y_1^n and message M_0 are independent; (b) follows from the memoryless property of the sources; (c) follows from the Markov chain $(Y_{1,t}, Z_{1,t}) \to (M_0, X^{t-1}, Z_1^{t-1}, Z_{1,t+1}^n) \to Y_1^{t-1}$. For the type-II error probability at Receiver 2, one obtains:

$$\begin{aligned}
-\frac{1}{n}\log\beta_{2,n} &\leq \frac{1}{n(1-\epsilon)} D(P_{M_0 Y_2^n|\mathcal{H}=0}||P_{M_0 Y_2^n|\mathcal{H}=1})\\
&= \frac{1}{n(1-\epsilon)} I(M_0; Y_2^n)\\
&= \frac{1}{n(1-\epsilon)} \sum_{t=1}^n I(M_0; Y_{2,t}|Y_{2,t+1}^n)\\
&= \frac{1}{n(1-\epsilon)} \sum_{t=1}^n \Big[I(M_0, Z_1^{t-1}; Y_{2,t}|Y_{2,t+1}^n)\\
&\qquad - I(Z_1^{t-1}; Y_{2,t}|M_0, Y_{2,t+1}^n)\Big]\\
&\stackrel{(b)}{=} \frac{1}{n(1-\epsilon)} \sum_{t=1}^n \Big[I(M_0, Z_1^{t-1}, Y_{2,t+1}^n; Y_{2,t})\\
&\qquad - I(Z_1^{t-1}; Y_{2,t}|M_0, Y_{2,t+1}^n)\Big]\\
&\stackrel{(c)}{=} \frac{1}{n(1-\epsilon)} \sum_{t=1}^n \Big[I(M_0, Z_1^{t-1}, Y_{2,t+1}^n; Y_{2,t})\\
&\qquad - I(Y_{2,t+1}^n; Z_{1,t}|M_0, Z_1^{t-1})\Big]\\
&\stackrel{(d)}{\leq} \frac{1}{n(1-\epsilon)} \sum_{t=1}^n \Big[I(M_0, Z_1^{t-1}, Y_{2,t+1}^n; Y_{2,t})\\
&\qquad - I(Y_{2,t+1}^n; Y_{2,t}|M_0, Z_1^{t-1})\Big]\\
&= \frac{1}{n(1-\epsilon)} \sum_{t=1}^n I(M_0, Z_1^{t-1}; Y_{2,t})\\
&= \frac{1}{n(1-\epsilon)} \sum_{t=1}^n I(U_{0,t}; Y_{2,t}),
\end{aligned}$$

where (b) follows from the memoryless property of the sources; (c) follows from Csiszar and Körner's sum identity [16]; and (d) follows from the less noisy assumption and the Markov chain $(M_0, Y_{2,t+1}^n, Z_1^{t-1}) \to X_t \to (Y_{1,t}, Y_{2,t}, Z_{1,t})$ which holds by the memoryless property of the sources and because M_0 is a function of X^n. For the rate R_0, one finds:

$$\begin{aligned}
nR_0 &\geq H(M_0) \geq I(M_0; X^n, Z_1^n)\\
&= I(M_0; X^n|Z_1^n) + I(Z_1^n; M_0)\\
&= \sum_{t=1}^n [I(M_0; X_t|X^{t-1}, Z_1^n) + I(M_0; Z_{1,t}|Z_1^{t-1})]\\
&= \sum_{t=1}^n \Big[I(M_0, X^{t-1}, Z_1^{t-1}, Z_{1,t+1}^n; X_t|Z_{1,t})\\
&\qquad + I(M_0, Z_1^{t-1}; Z_{1,t})\Big]\\
&= \sum_{t=1}^n \Big[I(X^{t-1}, Z_{1,t+1}^n; X_t|M_0, Z_{1,t}, Z_1^{t-1})\\
&\qquad + I(M_0, Z_1^{t-1}; X_t|Z_{1,t}) + I(M_0, Z_1^{t-1}; Z_{1,t})\Big]\\
&= \sum_{t=1}^n \Big[I(X^{t-1}, Z_{1,t+1}^n; X_t|M_0, Z_{1,t}, Z_1^{t-1})\\
&\qquad + I(M_0, Z_1^{t-1}; Z_{1,t}, X_t)\Big]\\
&\geq \sum_{t=1}^n \Big[I(X^{t-1}, Z_{1,t+1}^n; X_t|M_0, Z_{1,t}, Z_1^{t-1})\\
&\qquad + I(M_0, Z_1^{t-1}; X_t)\Big]\\
&= \sum_{t=1}^n [I(U_{1,t}; X_t|Z_{1,t}, U_{0,t}) + I(U_{0,t}; X_t)].
\end{aligned}$$

Notice that by the memoryless property of the sources and because M_0 is a function of X^n, the Markov chain $(M_0, Z_{1,t+1}^n, Z_1^{t-1}, X^{t-1}) \to X_t \to (Y_{1,t}, Y_{2,t}, Z_t)$ holds, and thus $(U_{0,t}, U_{1,t}) \to X_t \to (Y_{1,t}, Y_{2,t}, Z_t)$. The proof is then concluded by combining these observations with standard time-sharing arguments which require introducing the auxiliary random variables $T \in \{1, \ldots, n\}$, $U_0 \stackrel{\Delta}{=} (U_{0,T}, T)$, $U_1 \stackrel{\Delta}{=} U_{1,T}$, $X \stackrel{\Delta}{=} X_T$, $Y_1 \stackrel{\Delta}{=} Y_{1,T}$, $Y_2 \stackrel{\Delta}{=} Y_{2,T}$, and $Z_1 \stackrel{\Delta}{=} Z_{1,T}$.

Appendix C
Evaluation of $\mathcal{E}_{\mathrm{GW}}^{\mathrm{SI}}(R_0, R_1 = 0, R_2 = 0)$ for the example in Section II-C

That the exponent pairs in (29) lie in $\mathcal{E}_{\mathrm{GW}}^{\mathrm{SI}}(R_0, R_1 = 0, R_2 = 0)$ can be seen by evaluating (19) for auxiliaries

U_0 and U_1 that are jointly Gaussian with X and so that $X = U_1 + W_1$ and $U_1 = U_0 + W_0$ for independent zero-mean Gaussians W_1, W_0 and U_0 that are of variances $\frac{\sigma_z^2}{(\sigma_z^2+1)2^{-2\tilde{\alpha}}-1}$, $(\sigma_z^2+1)2^{-2(\tilde{\alpha}+R_0)} - \sigma_z^2(1 + \frac{1}{(\sigma_z^2+1)2^{-2\tilde{\alpha}}-1})$ and $(1+\sigma_z^2)(1 - 2^{-2(\tilde{\alpha}+R_0)})$, respectively.

That $\mathcal{E}_{\mathrm{GW}}^{\mathrm{SI}}(R_0, R_1 = 0, R_2 = 0)$ is no larger than the region in (29) is proved as follows. By the EPI:

$$h(Y_2|U_0) \geq \frac{1}{2}\log\left(2^{2h(Z_1|U_0)} + 2^{2h(N_2)}\right),$$

$$h(Y_1|U_0, U_1, Z_1) \geq \frac{1}{2}\log\left(2^{2h(X|U_0,U_1,Z_1)} + 2^{2h(N_1)}\right). \quad (63)$$

Moreover, rate-constraint on R_0 is equivalent to

$$\begin{aligned} R_0 &\geq I(U_0; X) + I(U_1; X|U_0, Z_1) \\ &= h(X) - h(X|U_0) + h(X|U_0, Z_1) \\ &\quad - h(X|U_0, U_1, Z_1) \\ &= h(X) - I(X; Z_1|U_0) - h(X|U_0, U_1, Z_1) \\ &= h(X) - h(Z_1|U_0) + h(Z_1|X, U_0) \\ &\quad - h(X|U_0, U_1, Z_1) \\ &= h(X, Z_1) - h(Z_1|U_0) - h(X|U_0, U_1, Z_1), \quad (64) \end{aligned}$$

where the last equality follows from the Markov chain $U_0 \to X \to Z_1$.

Defining now

$$\alpha := h(X|U_0, U_1, Z_1) \quad \text{and } \beta := h(Z_1|U_0), \quad (65)$$

above inequalities show that $\mathcal{E}_{\mathrm{GW}}^{\mathrm{SI}}(R_0, R_1 = 0, R_2 = 0)$ is included in the set of all pairs (θ_1, θ_2) that satisfy

$$\theta_1 \leq h(Y_1|Z_1) - \frac{1}{2}\log\left(2^{2\alpha} + 2^{2h(N_1)}\right), \quad (66)$$

$$\theta_2 \leq h(Y_2) - \frac{1}{2}\log\left(2^{2\beta} + 2^{2h(N_2)}\right), \quad (67)$$

for some choice of parameters $\alpha \leq h(X|Z_1)$ and $\beta \leq h(Z_1)$ so that

$$(\alpha - h(X|Z_1)) + (\beta - h(Z_1)) \geq -R_0. \quad (68)$$

Now, since the right-hand sides of (66) and (67) are decreasing in the parameters α and β, these parameters should be chosen so that the rate-constraint (68) is satisfied with equality. In other words, for fixed α, the optimal β is obtained by solving (68) under the equality constraint. Defining $\tilde{\alpha} := (\alpha - h(X|Z_1)) \leq 0$ and expressing the optimal β in terms of $\tilde{\alpha}$ then establishes the desired inclusion of $\mathcal{E}_{\mathrm{GW}}^{\mathrm{SI}}(R_0, R_1 = 0, R_2 = 0)$ in the set of pairs (θ_1, θ_2) given in (29).

Appendix D
Proof of Theorem 4

We analyze the probability of error of the scheme in Section III-A. It immediately follows that the type-I error probabilities at the two receivers cannot exceed ϵ from the way the Neyman-Pearson test is designed. Now, we consider the type-II error probabilities. They can be upper bounded using the Chernoff-Stein lemma. Thus, for sufficiently large k:

$$\begin{aligned} -\frac{1}{n}\log\beta_{i,n} &\geq \frac{1}{k} D\left(P_{\hat{U}_i^k Y_i^k Z_i^k|\mathcal{H}=0} \,\|\, P_{\hat{U}_i^k Y_i^k Z_i^k|\mathcal{H}=1}\right) - \mu \\ &\overset{(a)}{=} \frac{1}{k} I\left(\hat{U}_i^k; Y_i^k \middle| Z_i^k\right) - \mu \\ &\geq H\left(Y_i \middle| Z_i\right) - \frac{1}{k} H\left(Y_i^k \middle| Z_i^k, \hat{U}_i^k\right) - \mu, \end{aligned}$$

where mutual informations and entropies have to be computed according to the joint pmf $P_{\hat{U}_i^k Y_i^k Z_i^k}$ under $\mathcal{H} = 0$, and Equality (a) follows because under $\mathcal{H} = 1$, the joint distribution of the variables decomposes as $P_{\hat{U}_i^k Z_i^k} P_{Y_i^k|Z_i^k}$. As shown in detail in [14], for sufficiently large values of k, the rate constraints in (32)–(39) ensure that

$$\Pr\left[(\hat{U}_{i,b}^k, Y_{i,b}^k, Z_{i,b}^k) \in \mathcal{T}_\mu^k(P_{U_i Y_i Z_i})\right] > 1 - \mu. \quad (69)$$

Following similar steps as the ones leading to (61), one obtains:

$$H\left(Y_i^k \middle| \hat{U}_i^k, Z_i^k\right) \leq H(Y_i|Z_i, U_i) + \log|\mathcal{Y}_i| \cdot \mu + \frac{1}{k} + \delta(\mu), \quad (70)$$

for a function $\delta(\mu)$ that tends to 0 as $\mu \to 0$. Thus, we get

$$-\frac{1}{n}\log\beta_{i,n} \geq I(U_i; Y_i|Z_i) - \log|\mathcal{Y}_i| \cdot \mu - \frac{1}{k} - \delta(\mu). \quad (71)$$

Taking $\mu \to 0$ and $k \to \infty$ proves the theorem.

Acknowledgement

M. Wigger wishes to thank O. Shayevitz for helpful discussions.

References

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

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

[3] H. Shimokawa, T. Han, and S. I. Amari, "Error bound of hypothesis testing with data compression," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 1994, p. 114.

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

[5] W. Zhao and L. Lai, "Distributed testing against independence with conferencing encoders," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Jeju, South Korea, Oct. 2015, pp. 19–23.

[6] Y. Xiang and Y.-H. Kim, "Interactive hypothesis testing against independence," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jun. 2013, pp. 2840–2844.

[7] G. Katz, P. Piantanida, and M. Debbah. (Apr. 2016). "Collaborative distributed hypothesis testing." [Online]. Available: https://arxiv.org/abs/1604.01292

[8] J. Liao, L. Sankar, F. P. Calmon, and V. Y. F. Tan, "Hypothesis testing under maximal leakage privacy constraints," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017, pp. 779–783.

[9] S. Sreekuma and D. Gunduz. (2017). "Distributed hypothesis testing over noisy channels." [Online]. Available: https://arxiv.org/abs/1704.01535

[10] O. Shayevitz and M. Wigger, "On the capacity of the discrete memoryless broadcast channel with feedback," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1329–1345, Mar. 2013.

[11] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell Syst. Tech. J.*, vol. 48, pp. 1681–1721, Nov. 1974.

[12] A. Kaspi and T. Berger, "Rate-distortion for correlated sources with partially separated encoders," *IEEE Trans. Inf. Theory*, vol. 28, no. 6, pp. 828–840, Nov. 1982.
[13] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 727–734, Nov. 1985.
[14] P. Minero, S. H. Lim, and Y.-H. Kim, "A unified approach to hybrid coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1509–1523, Apr. 2015.
[15] P. W. Cuff, H. H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4181–4206, Sep. 2010.
[16] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[17] M. Wigger and R. Timo, "Testing against independence with multiple decision centers," (Invited Paper), in *Proc. SPCOM*, Bangalore, India, Jun. 2016, pp. 1–5.
[18] I. Csiszár, "Linear codes for sources and source networks: Error exponents, universal coding," *IEEE Trans. Inf. Theory*, vol. 28, no. 4, pp. 585–592, Jul. 1982.
[19] B. G. Kelly and A. B. Wagner, "Improved source coding exponents via Witsenhausen's rate," *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 5615–5633, Sep. 2011.
[20] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 5–12, Jan. 1981.
[21] E. Tuncel, "Slepian-Wolf coding over broadcast channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1469–1482, Apr. 2006.
[22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1991.
[23] S. Salehkalaibar, M. Wigger, and L. Wang. (2017). "Hypothesis testing in multi-hop networks." [Online]. Available: https://arxiv.org/abs/1708.05198

Sadaf Salehkalaibar (M'14) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2008, 2010, and 2014, respectively. She was a Post-Doctoral Fellow with Telecom ParisTech, Paris, France, in 2015 and 2017, respectively. She is currently an Assistant Professor with the Electrical and Computer Engineering Department of, University of Tehran. Her research interests include network information theory and fundamental limits of secure communication with emphasis on information-theoretic security.

Michèle Wigger (S'05–M'09–SM'14) received the M.Sc. degree (Hons.) in electrical engineering and the Ph.D. degree in electrical engineering from ETH Zurich in 2003 and 2008, respectively. In 2009, she was a Post-Doctoral Fellow with the University of California, San Diego, CA, USA. She then joined Telecom Paris Tech, Paris, France, where she is currently an Associate Professor. She has held visiting professor appointments with the Technion–Israel Institute of Technology and ETH Zurich. She has served as an Associate Editor of the IEEE COMMUNICATION LETTERS and is currently an Associate Editor for Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY. She is currently serving on the Board of Governors of the IEEE Information Theory Society. Her research interests include multi-terminal information theory, in particular in distributed source coding and in capacities of networks with states, feedback, user cooperation, or caching.

Roy Timo received the B.E. and Ph.D. degrees from The Australian National University in 2005 and 2009, respectively. He was an Alexander von Humboldt Research Fellow with the Institute for Communications Engineering, Technische Universitat Munchen from 2014 to 2016, a Research Fellow with the Institute for Telecommunications Research, University of South Australia, from 2008 to 2013, and a Post-Doctoral Researcher with the Department of Communications and Electronics, Telecom ParisTech, from 2013 to 2014. He is currently an Experienced Researcher with Ericsson Research, Stockholm, Sweden.