Distributed Binary Detection With Lossy Data Compression

Gil Katz, Pablo Piantanida, Mérouane Debbah

Distributed Binary Detection with Lossy Data Compression

Gil Katz, *Student Member, IEEE,* Pablo Piantanida, *Senior Member, IEEE,* and Mérouane Debbah, *Fellow, IEEE*

***Abstract*—Consider the problem where a statistician in a two-node system receives rate-limited information from a transmitter about marginal observations of a memoryless process generated from two possible distributions. Using its own observations, this receiver is required to first identify the legitimacy of its sender by declaring the joint distribution of the process, and then depending on such authentication it generates the adequate reconstruction of the observations satisfying an average per-letter distortion. The performance of this setup is investigated through the corresponding *rate-error-distortion region* describing the trade-off between: the communication rate, the error exponent induced by the detection and the distortion incurred by the source reconstruction. In the special case of *testing against independence*, where the alternative hypothesis implies that the sources are independent, the optimal rate-error-distortion region is characterized. An application example to binary symmetric sources is given subsequently and the explicit expression for the rate-error-distortion region is provided as well. The case of "general hypotheses" is also investigated. A new achievable rate-error-distortion region is derived based on the use of non-asymptotic *binning*, improving the quality of communicated descriptions. Further improvement of performance in the general case is shown to be possible when the requirement of source reconstruction is relaxed, which stands in contrast to the case of general hypotheses.**

***Index Terms*—Data compression; error statistics; signal detection; asymptotic performance; central detector; discrete spatially dependent observations; distributed detection; error exponent; multiterminal detection; multiterminal source coding; side information; lossy source coding; type-I error rate; type-II error rate.**

I. Introduction

THE problem of Hypothesis Testing (HT) is very familiar in statistics. Presented with a list of n independent and identically distributed (i.i.d) realizations of some random variable (RV) X, a statistician attempts to determine the probability distribution that governs the RV, out of a known list of possible distributions. One popular special case is Binary HT, where only two possible hypotheses exist, usually referred to as H_0 and H_1. Readers interested in an overview of HT problems can consult [3] and references therein.

The problem of Binary HT is formally defined by two types of error probabilities which are commonly referred to as Type I and II probabilities. Denote by α_n the first type error probability given by the probability that H_1 is chosen despite H_0 being true, while the error probability of the second type β_n is defined to be the probability that H_0 is chosen while H_1 is true. Although the trade-off between the two error events can be investigated in many ways, one common path is to investigate the exponential rate of decay of the error probability of the second type, i.e., $-\lim_{n\to\infty}\frac{1}{n}\log\beta_n^\star(\epsilon)$, while imposing a fixed constraint over the error probability of the first type, i.e., $\alpha_n \leq \epsilon$ ($\epsilon > 0$). *Stein's Lemma* [3], [4] provides a closed-form expression for the optimal error exponent in this case,

$$-\lim_{n\to\infty}\frac{1}{n}\log\beta_n^\star(\epsilon) = \mathcal{D}(P_0\|P_1)\ , \tag{1}$$

where P_0 and P_1 are the probability distributions implied by hypotheses H_0 and H_1, respectively, and $\mathcal{D}(\cdot\|\cdot)$ is the *Kullback-Leibler divergence* provided that the measure P_0 is *absolutely continuous* resp. to P_1, i.e., $P_0 \ll P_1$. It is worth to emphasize that, the optimal exponential rate of decay of the error probability of the second type does not depend asymptotically on the specific constraint over the error probability ϵ of the first type.

The situation is substantially more complicated in the case of a distributed detection. If it were possible to transmit all signals to some central location with negligible cost and delay, then the previous theory is in principle applicable. However, due to practical considerations such as energy cost, reliability, survivability, communication bandwidth, compartmentalization, there is never total centralization of information in practice [5]. In this paper, we focus on the problem of distributed hypothesis testing where it is assumed that realizations of different memoryless sources of finite alphabets are observed at different physical locations and thus, nodes are subject to satisfy different types of communication constraints. This work attempts a modest step in the direction of a theory for distributed testing based on lossy data compression which seems to offer a formidable mathematical complexity (see [6] and references therein).

A. Related Work

Ahlswede & Csiszar [7] and then Han [8] investigated the two-node distributed binary HT problem, where only one-sided communication is allowed, with rate R [bits/sample] (see Fig. 1 for a representation of a similar system). Both works

This research has been supported by the ERC Grant 305123 MORE (Advanced Mathematical Tools for Complex Network Engineering). The material in this paper was presented in part in the 52nd Annual Allerton Conference on Communication, Control and Computing 2014 [1], and at the 2015 IEEE International Symposium on Information Theory (ISIT) [2].

Gil Katz is with Large Networks and Systems Group (LANEAS), CentraleSupélec, 91192 Gif-sur-Yvette, France. Email: gil.katz@centralesupelec.fr.

P. Piantanida are with Laboratoire de Signaux et Systèmes (L2S, UMR8506), CentraleSupélec-CNRS-Université Paris-Sud, 91192 Gif-sur-Yvette, France. Email: pablo.piantanida@centralesupelec.fr.

M. Debbah is with Large Networks and Systems Group (LANEAS), CentraleSupélec, 91192 Gif-sur-Yvette, France. Email: merouane.debbah@centralesupelec.fr.

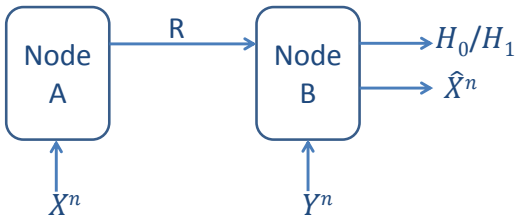

Figure 1. Communication model for joint distributed detection and source reconstruction.

offer similar approaches to derive achievable rate-exponent rates for this problem, while the results are derived based on somewhat different tools. Although optimality is proven in [7] for the special case of "testing against independence", an optimality result for the general case remains elusive.

While testing against independence is a particular case that assumes $P_{1,XY} = P_X P_Y$ and $P_{0,XY} = P_{XY}$, it is important in many scenarios where checking the *relevance* of information being transmitted is of interest. This scenario resembles the known case of transmitting information where side information may be absent [9], [10], but is rendered more complex by the fact that even the receiver is unaware of the relevance of the side information. An equivalent setting, namely vector Gaussian source coding with decoder side information under mutual information and distortion constraints, has been investigated in [11], and benefits of successive refinement for testing against independence are studied in [12]. The problem of testing against independence is approached for the scenario where reciprocal communication is allowed between the two nodes in [13]. Benefits of a two-way communication system were demonstrated through a coding scheme inspired by the seminal work of Kaspi [14].

Considering the general HT scenario described in Fig. 1, the problem faced in this paper shares common roots with the seminal works in [7], [8]. Here, however, we are not interested *only* in distributed testing but also in achieving source reconstruction. This also connects to the lossy source coding problem by Heegard & Berger [15], where two decoders have to reconstruct the same source based on different side informations and the setup investigated in [11]. Along the line of the technical tools used in the present work, authors in [16] suggested the use of "binning" as a possible approach to improve performance of distributed HT by reducing the coding rate. We shall study this approach which, however, brings forth different difficulties, stemming from the fact that the worth of the side information at the decoder is unknown before a decision is made about the state of the system. That is because reliable decoding of the "bin index" is required in presence of side information uncertainty (e.g. similarly to problems under channel uncertainty [17]), which is also met and contended with in our present framework. Binning was also shown to be useful in [18], where a multi-node system composed of several decentralized encoders that send limited-rate messages to a decoder about their observations was investigated for the case of testing against conditional independence.

In this work, we consider another dimension of the problem, as represented in Fig. 1. An authentication system prevents the unauthorized injection of messages into a public channel, on which security is inadequate for the needs of its users since it may be threatened with eavesdropping or injection or both [19]. This threat of compromise of the receiver's authentication data is motivated by situations in multiuser networks –such as automatic fault diagnosis– where the receiver is often the system itself which cannot be treated by conventional cryptography, and which require recourse to new techniques (e.g. image authentication [20], [21] and Smart Grids [22], [23]). Having divided the problem into that of authentication and communication, decoding of a message at the receiver (node B) requires first a reliable identification of the legitimacy of its sender (node A) and then a lossy reconstruction of the underlying feature vector $X = (X_1, \ldots, X_n)$, with an average per-letter distortion depending on the decision made. In a sense, this problem combines the general distributed HT problem studied in [7] and [8] with the classical Heegard & Berger [15].

B. Main Contributions

The paper is divided into three parts. In the first part, we focus on the case of testing against independence where the alternative hypothesis H_1 is a disjoint "version" of H_0 that leads to $\mathbf{X}^n$ and $\mathbf{Y}^n = (Y_1, \ldots, Y_n)$ to be independent from each other while sharing the same marginal distributions as under H_0. By relying on the techniques introduced in [8], we offer an achievable (single-letter) expression for the tradeoff between the coding rate, the error exponent and the average per-letter distortion, referred to as *rate-error-distortion region*. In this setting, we simply assume that reconstruction is only attempted when H_0 is decided, since no effective side-information is available at the decoder when H_1 is the true hypothesis.

Interestingly, it is shown that the optimal rate-error-distortion region is attained by using *layered coding*, where the first layer performs HT, and the second layer uses well-known results for source coding with side information at the decoder [24], while ignoring the information received by node B at the HT stage. This result is quite surprising, as in general there is no reason to believe that such a separation between the two aspects of the problem should be optimal. We explicitly evaluate the rate-error-distortion region for uniform Binary Sources where a Binary Symmetric Channel (BSC) is assumed between X and Y, and plot the resulting tradeoffs between the three quantities of interest.

In the second part, we derive an achievable rate-error-distortion region for the same system, under no specific assumptions on the two hypotheses. To this end, we allow the use of binning not only for source reconstruction but also for the testing purpose. The resulting rate-error-distortion achievable region is in fact a quadruplet, comprised of the rate of communication, the error exponent for an error of the second type, subject to a maximum probability of error of the first type, and the average distortion corresponding to each hypothesis. The techniques required for this analysis are inspired by previous work on distributed HT [8] and recent work [25] on the study of the error exponent for the problem of lossy source coding with side information at the receiver.

It should be mentioned here that although the use of binning for HT was first suggested in [16] as a possible approach to improve performance, the benefits of this were never demonstrated. Along this line, Rahman and Wagner [18] show that binning is optimal for HT when under H_1 the involved variables are assumed to be *conditionally independent* given some additional variable, known at the decoder side. While this work played a big part in inspiring a binning approach for HT, it turns out that using Y as the side information available to the receiver does not necessarily improve testing performance, as the exact value of side information is unknown.

In the third part of this paper, we concentrate on distributed HT without reconstruction constraints. We show that for the case of two general hypotheses, unlike the case of testing against independence, our previous two-stage coding approach leads to significant loss in performance. We do so by suggesting a new approach for testing without requiring the decoding of the involved descriptions. This turns out to be superior to the previous one in terms of error exponent, but prevents the decoder of providing a lossy reconstruction of the source. As the performance of the previous approach for general distributed hypotheses testing is lower-bounded by the known result of [8], the new approach we introduce may also lead to a significant gain in performance, when compared to this non-binned option.

The rest of this paper is organized as follows. Section II presents the optimal rate-error-distortion region for the case of testing against independence. Optimality is also shown for a specific example of a binary symmetric channel (BSC) between X and Y, and numerical results are given. The rate-error-distortion region for the general HT case is given in Section III. In Section IV, we offer a different approach for HT only. The performance of the two previously presented approaches are compared through numerical results. Finally, concluding remarks are given in Section V.

Notation and Conventions

We use upper-case letters to denote random variables (RVs) and lower-case letters to denote realizations of RVs. Vectors are denoted by boldface letters, with their length as a superscript, emitted when it is clear from the context. Let $\mathbf{X}_i^j$ denote the vector $\mathbf{X}$, from position i to position j, i.e., $\mathbf{X}_i^j = (X_i, X_{i+1}, \ldots, X_{j-1}, X_j)$. $\mathcal{P}(\mathcal{X})$ denotes the set of all possible probability distributions on $\mathcal{X}$, while $p_X \in \mathcal{P}(\mathcal{X})$ is a member of this set. $Q_{\mathbf{x}^n}$ denotes the empirical distribution, referred to as *the type*, of the vector $\mathbf{x}^n = (x_1, \ldots, x_n)$. $\mathcal{P}_n(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$ denotes the set of all possible atomic probability distributions (or types) on the alphabet $\mathcal{X}$. The set of all vectors $\mathbf{x}^n \in \mathcal{X}^n$ with a specific type Q is denoted by $\mathcal{T}(Q) = \mathcal{T}_{[Q]}$, while the set of all vectors that are δ-typical (in the usual sense, as defined in Appendix A) is denoted by $\mathcal{T}^n_{[Q]\delta}$. Using Csiszár's notation [26], we let $H(P_X) = \mathbb{E}\left[-\log p_X(X)\right]$ denote the entropy of a RV distributed according to p, and distinguish the binary entropy function by $H_2(x) = -x\log_2 x - (1-x)\log_2(1-x)$. $I(X;Y)$ denotes the mutual information between X and Y while assuming that $p_X p_{Y|X}$ governs the pair, and $\mathcal{D}(P_X \| P'_X)$ the KL divergence between the distributions p and p'. All exponents and logarithms in this paper are base 2, unless stated otherwise. We denote the scalar convolution function by $a \star b \triangleq a(1-b) + b(1-a)$. Finally, known definitions and properties of typical sequences are given in Appendix A.

II. Testing Against Independence

A. Definitions

In this section, we give a more rigorous formulation of the context depicted in Fig. 1 for the case of testing again independence. Let $\mathcal{X}$ and $\mathcal{Y}$ be two finite sets. Nodes A and B observe sequences of random variables $(X_i)_{i\in\mathbb{N}^\star}$ and $(Y_i)_{i\in\mathbb{N}^\star}$ respectively, which take values on $\mathcal{X}$ and $\mathcal{Y}$, resp. For each $i \in \mathbb{N}^\star$, random samples (x_i, y_i) are distributed according to one of two possible joint distributions:

$$\begin{cases} H_0: & p_0(x,y) = P_{XY}(x,y)\ , \\ H_1: & p_1(x,y) = P_{\bar{X}\bar{Y}}(x,y) = P_X(x)P_Y(y)\ . \end{cases} \tag{2}$$

on $\mathcal{X}\times\mathcal{Y}$. Assume that the pairs (X_i, Y_i) are independent across time i.

Let $d : \mathcal{X}\times\hat{\mathcal{X}} \to [0\,; d_{\max}]$ be a finite distortion measure *i.e.*, such that $0 \le d_{\max} < \infty$. We also denote by d the component-wise mean distortion on $\mathcal{X}^n \times \hat{\mathcal{X}}^n$, *i.e.*, for each $(\mathbf{x}^n, \hat{\mathbf{x}}^n) \in \mathcal{X}^n \times \hat{\mathcal{X}}^n$, $d(\mathbf{x}^n, \hat{\mathbf{x}}^n) \triangleq \frac{1}{n}\sum_{i=1}^n d(x_i, \hat{x}_i)$. We assume that node A can send information to node B over an error-free link with rate R bits per source-symbol. Having received the information from node A, node B is then required to make a decision between the two possible hypotheses. After having decided between the two hypotheses, node B attempts to reconstruct the sequence X, with minimum distortion, for some additive distortion measure, that may depend on the actual probability distribution in place. While recovering the sequence seen by node A under hypothesis H_1 may still be possible, it becomes less relevant, as in this case the sequence seen by node B is completely independent and does not constitute as side information. Furthermore, it is very likely that in realistic cases where testing against independence arises, deciding H_1 implies that the information seen by node A is irrelevant to node B. Thus, for the case of testing against independence, we assume node B attempts to decode only if it has decided H_0. In the general hypotheses case, decoding is attempted under any of the two hypotheses.

Definition 1 (Code)**.** *An (n, R)-code for testing against independence in this setup is defined by*

- *An encoding function at node A denoted by $f_n : \mathcal{X}^n \to \{1, \ldots, \|f_n\|\}$;*
- *A decision region $\mathcal{A}_n \subset \{1, \ldots, \|f_n\|\} \times \mathcal{Y}^n$, such that if $(f_n(\mathbf{x}^n), \mathbf{y}^n) \in \mathcal{A}_n$ the decoder declares H_0 and otherwise H_1 ;*
- *A reconstruction function at node B denoted by $g_n : \{1, \ldots, \|f_n\|\} \times \mathcal{Y}^n \to \hat{\mathcal{X}}^n$.*

Definition 2 (Rate-exponent-distortion region)**.** *A tuple $(R, E, D, \epsilon) \in \mathbb{R}_+^4$ is said to be* achievable *if, for any $\delta > 0$*

and for n large enough, there exists an $(n, R+\delta)$-code $(f_n, \mathcal{A}_n, g_n)$ such that:

$$\begin{aligned} n^{-1}\log\|f_n\| &\leq R+\delta \ , \\ \mathbb{E}_0\big[d\big(\mathbf{X}^n, g_n(f_n(\mathbf{X}^n), \mathbf{Y}^n)\big)\big] &\leq D+\delta \ , \\ -\frac{1}{n}\log\beta_n(\mathcal{A}_n) &\geq E-\delta \ , \\ \alpha_n(\mathcal{A}_n) &\leq \epsilon \ , \end{aligned}$$

where $\beta_n(\mathcal{A}_n) = \Pr\big(\mathcal{A}_n | XY \sim p_1(x,y)\big)$ and $\alpha_n(\mathcal{A}_n) = \Pr\big(\mathcal{A}_n^c | XY \sim p_0(x,y)\big)$, and $\mathbb{E}_0$ denotes that distortion is measured under the condition that node B correctly decides H_0. The set of all such achievable tuples is denoted by $\mathcal{R}^\star$ and is referred to as the rate-exponent-distortion region*.*

In [7] and later on in [8], the authors show that when testing against independence, the optimal approach at node B is to apply Stein's Lemma over the common distribution of $\mathbf{Y}^n$ and the encoded descriptions $f_n(\mathbf{X}^n)$. More specifically, by optimizing over all decision regions $\mathcal{A}_n \subset \{1,\ldots,\|f_n\|\} \times \mathcal{Y}^n$, the smallest probability of error of the second type β_n asymptotically behaves as: $\beta_n \approx \exp\left(-nE(R)\right)$ with n large enough, for a fixed constraint on the error probability of the first type $\alpha_n \leq \epsilon$, and the exponent $E(R)$ satisfies [7, Lemma 1.a]:

$$E(R) = \sup_{n\geq 1} E_n(R) \ , \tag{3}$$

where

$$E_n(R) = \sup_{f_n}\left\{ \frac{1}{n} I\left(f_n(\mathbf{X}^n); \mathbf{Y}^n\right) \;\middle|\; \log\|f_n\| \leq nR \right\} . \tag{4}$$

This asymptotic equivalence implies a strong converse property that, much like in the single-node HT setup, the optimal exponential decay of β_n is not dependent upon the chosen constraint $0<\epsilon<1$ on the error probability of the first type α_n (e.g. see [11] for a proof based on image sets).

B. Single-Letter Rate-Error-Distortion-Region

We now state the optimal rate-error-distortion region for testing against independence, which provides a single-letter expression for the rate-error-distortion region for testing against independence, defined in that in Definition 2.

Proposition 1 (Rate-error-distortion region)**.** *A tuple $(R,E,D) \in \mathbb{R}_+^3$ is achievable for the two-node detection and reconstruction problem when testing against independence, as defined in Definition 2, if and only if two random variables $U \in \mathcal{U}$ and $V \in \mathcal{V}$, as well as a reconstruction mapping $g : \mathcal{U}\times\mathcal{V}\times\mathcal{Y} \to \hat{\mathcal{X}}$, can be found, such that*

$$I(U;X) + I(V;X|UY) \leq R \ , \tag{5}$$

$$I(U;Y) \geq E \ , \tag{6}$$

$$\mathbb{E}_0\left[d\big(X, g(UVY)\big)\right] \leq D \ , \tag{7}$$

with (U,V) being two random variables satisfying $U \multimap V \multimap X \multimap Y$ form a Markov chain with $(X,Y) \sim p_0(x,y)$, and $\|\mathcal{U}\| \leq \|\mathcal{X}\|+2$, $\|\mathcal{V}\| \leq \|\mathcal{X}\|\|\mathcal{U}\|+1$.

Proof: The proof of Proposition 1 is given in Appendix B. ∎

Remark 1. *Observe that on one hand, the expression for the rate can be evaluated as follows:*

$$\begin{aligned} R &\geq I(U;X) + I(V;X|U) - I(V;Y|U) \\ &= I(U;Y) + \left[I(V;X) - I(V;Y)\right] \ , \end{aligned} \tag{8}$$

where the final equality stems from the Markov chain formed by the RVs and on the other hand, from the fact that $U \multimap V \multimap X \multimap Y$ form a Markov chain, it is easy to see that

$$\mathbb{E}_0\left[d\big(X, g'(VY)\big)\right] \leq \mathbb{E}_0\left[d\big(X, g(UVY)\big)\right] \leq D \ , \tag{9}$$

for some mapping g' and any g. Note that the rate can now be seen as comprised of two different parts. The first part of the resulting expression in (8) *is dedicated to detection since it only affects the error exponent, and is in fact identical to the expression of the error exponent given in* (6) *in agreement with previous results [7], [8]. The second part of the rate is dedicated only to source reconstruction and therefore, the rate-error-distortion region can be seen as being equivalent to two uncoupled problems that share a common rate. In the following sections, we will see that this is not the case when general hypotheses are considered.*

Remark 2. *Note that while the assumption that distortion is only measured in case the detection of hypothesis H_0 is convenient, it is not necessary. As we assume that the distortion measure is bound from above, the distortion under the decision H_0 (which may or may not be correct) may be expressed as follows:*

$$\begin{aligned} &\mathbb{E}_0\left[d\big(X, g(UVY)\big)\right] \\ &= \mathbb{E}_0\left[d\big(X, g(UVY)\big)\text{, ``correct detection''}\right] \\ &\qquad \times \Pr\{\text{``correct detection''}\} \\ &+ \mathbb{E}_0\left[d\big(X, g(UVY)\big)\text{, ``incorrect detection''}\right] \\ &\qquad \times \Pr\{\text{``incorrect detection''}\} \\ &\leq \mathbb{E}_0\left[d\big(X, g(UVY)\big)|H_0\text{, ``correct detection''}\right] \\ &\qquad + \beta_n d_{max} \ , \end{aligned} \tag{10}$$

where d_{max} is assumed to be that maximal value that the distortion function $d(\cdot,\cdot)$ takes. As $\beta_n d_{max} \to 0$ when $n\to\infty$ the relaxation of the assumption that the distortion is only measured under correct detection does not change the optimal rate-error-distortion region. Note that the assumption that estimation is only done under the decision H_0 was not relaxed, only the fact that distortion is not measured under incorrect detection.

C. Binary Symmetric Source

In some cases, the region defined by Proposition 1 can be calculated analytically. We present such an example here. Consider the following statistical model: Let $X \sim \text{Bern}\left(\frac{1}{2}\right)$, and

$$\begin{cases} H_0: & Y = X + Z, \quad Z \sim \text{Bern}(p) \perp X \\ H_1: & Y \sim \text{Bern}\left(\frac{1}{2}\right) \perp X \ , \end{cases} \tag{11}$$

with $\text{Bern}(p)$ being a *Bernoulli* RV with probability p for being 1, and $\perp$ signifying that X and Z are independent of each other in the case of hypothesis 0, and X and Y are independent

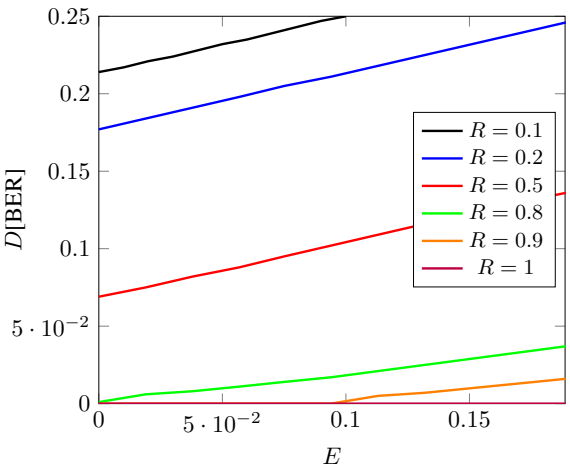

Figure 2. Numerical results of the optimal average distortion as a function of the desired error exponent of the second type, for different amounts of available rate and for $p = 0.25$.

under the premises of hypothesis 1. Under both hypotheses, the marginal distributions of both X and Y are equal. Thus, a decision can be reached only through cooperation between the nodes. In the next proposition, the rate-error-distortion region for this problem is characterized by optimizing over all involved random variables in Proposition 1.

Proposition 2 (Rate-Error-Distortion region for Binary Symmetric Sources)**.** *The rate-error-distortion region for BSS and testing against independence is given by*

$$R \geq 1 - H_2(\alpha * \beta * p) + \theta \left[H_2(\alpha * p) - H_2(\alpha) \right] \ , \quad (12a)$$

$$E \leq 1 - H_2(\alpha * \beta * p) \ , \quad (12b)$$

$$D \geq \theta\alpha - (1-\theta)\, p \ , \quad (12c)$$

for any $0 \leq \alpha, \beta \leq \frac{1}{2}$, $0 \leq \theta \leq 1$.

Proof: The proof is given in Appendix C. ∎

D. Numerical Results

We now present numerical results for the Binary Symmetric Source (BSS) case of testing against independence. Fig. 2 shows six curves, each representing the trade-off between user authentification and source reconstruction, expressed by the desired error exponent (second type) and the resulting average distortion of the source estimation, for a fixed value of available rate and for $p = 0.25$. Unsurprisingly, all curves are non-decreasing, meaning that when the probability of error is exponentially smaller, the amount of rate left for source reconstruction is smaller, resulting in a more crude estimation.

Assuming that both sources $\mathbf{X}^n$ and $\mathbf{Y}^n$ are available at a single location, Stein's Lemma yields an error exponent $E_{\max} = I(X;Y) = 1 - H_2(p) \approx 0.1887$. Obviously, this value constitutes an upper bound –uniform over the rate– on the achievable exponent in the distributed setup presented here. It can be seen that when $R < E_{\max}$, the average distortion reaches its maximal value $D_{\max} = p = 0.25$ for some $E < E_{\max}$. Any exponent bigger than the value for which this happens is unachievable with this rate, since the desired exponent would demand more rate than available. When $R > E_{\max}$, further enlarging the rate allows for better distortion, for the same values of error exponent.

Note especially the curves for the rate values: $R = 0.9$ and $R = 1$, which comply with $R > H_2(p)$. According to Slepian-Wolf coding (see e.g. [4]), this rate is enough to transmit $\mathbf{x}^n$ to node B without distortion, when no detection is necessary. Indeed, it can be seen that for any choice of error exponent that ensures enough available rate for estimation, zero-distortion is achievable. The curve for $R = 1$ is thus almost invisible, as in this case enough rate is available for source reconstruction, for any achievable choice of error exponent.

III. General Hypothesis Testing

We now focus on the general case, where both hypotheses can be general distributions of two variables. Note that now, unlike the case of testing against independence, the performance of the system is measured by four quantities, namely the rate, the error exponent and two distortions, as source reconstruction is attempted under both hypotheses. Nevertheless, distortion is still measured under the assumption that the detection step was completed successfully. Unlike the case of testing against independence, optimality results for general distributed HT remain elusive. An achievable region [8] was inspired by the approach taken for testing against independence. We propose here an achievable region for the general hypothesis testing problem with source reconstruction constraints that makes use of binning for both purposes. The proposed region, while not necessarily optimal in general, aims at improving on known results for the testing part while also adding the reconstruction of the source.

A. Definitions

As before, we suppose that the statistician observes $\mathbf{Y}^n$ samples directly and can be informed about $\mathbf{X}^n$ samples indirectly, via an encoding function $f_n : \mathcal{X}^n \to \{1, \ldots, \|f_n\|\}$ of rate $n^{-1} \log \|f_n\| \leq R$. The code definition remains the same as in Definition 1 with two reconstructions functions $g_{n,i} : \{1, \ldots, \|f_n\|\} \times \mathcal{Y}^n \to \hat{\mathcal{X}}_i^n$. For each $i \in \mathbb{N}^\star$, random samples (x_i, y_i) are distributed according to one of two general joint distributions:

$$\begin{cases} H_0 : & p_0(x,y) = P_{XY}(x,y) \ , \\ H_1 : & p_1(x,y) = P_{\bar{X}\bar{Y}}(x,y) \ , \end{cases} \quad (13)$$

on $\mathcal{X} \times \mathcal{Y}$. Moreover, these samples are independent across time $i = \{1, \ldots, n\}$, and we assume throughout this section that $P_X(x) = P_{\bar{X}}(x)$ and $P_Y(y) = P_{\bar{Y}}(y)$, $\forall (x,y) \in \mathcal{X} \times \mathcal{Y}$.

Definition 3 (Rate-exponent-distortion region)**.** *A tuple* $(R, E, D_0, D_1, \epsilon) \in \mathbb{R}_+^5$ *is said to be* achievable *if, for any* $\delta > 0$*, there exists an* $(n, R + \delta)$*-code* $(f_n, \mathcal{A}_n, g_{n,0}, g_{n,1})$ *such that:*

$$\begin{gathered} n^{-1} \log \|f_n\| \leq R + \delta \ , \\ \mathbb{E}_i \left[d_i\big(\mathbf{X}^n, g_{n,i}(f_n(\mathbf{X}^n), \mathbf{Y}^n)\big) \right] \leq D_i + \delta \ , \ i = 0, 1 \\ -\frac{1}{n} \log \beta_n(\mathcal{A}_n) \geq E - \delta \ , \end{gathered} \quad (14)$$

where $\beta_n(\mathcal{A}_n) = \Pr\big(\mathcal{A}_n|XY \sim p_1(x,y)\big)$ *and* $\alpha_n(\mathcal{A}_n) = \Pr\big(\mathcal{A}_n^c|XY \sim p_0(x,y)\big)$*, and distortion is measured under the condition that node B correctly detects the correct hypothesis. The set of all such achievable tuples is denoted by* $\mathcal{R}^\star$ *and is referred to as the* rate-exponent-distortion region*.*

Remark 3. *Note the slight abuse of notation in the distortion argument of Definition 3: The fact that we assume the distortion is measured only in case the detection phase was completed correctly means that for each distortion argument the "correct" RVs are assumed to be used. Thus,* $\mathbb{E}_0\big[d_0\big(\mathbf{X}^n, g_{n,0}(f_n(\mathbf{X}^n), \mathbf{Y}^n)\big)\big] \leq D_0 + \delta$ *is the correct expression for the distortion under* H_0*, while* $\mathbb{E}_1\big[d_1\big(\bar{\mathbf{X}}^n, g_{n,1}(f_n(\bar{\mathbf{X}}^n), \bar{\mathbf{Y}}^n)\big)\big] \leq D_1 + \delta$ *is the corresponding expression under hypothesis* 1*.*

B. Achievable Rate-Error-Distortion Region

We now state our main result for the general joint distributed detection and reconstruction problem, which is a new achievable rate-error-distortion region. This region is inspired by the one offered for the special case of testing against independence. In a similar manner to the approach taken in Proposition 1, we derive an achievable region based on the separation of two distinguishable steps, namely user authentication and source reconstruction. The statistician first decodes the description needed to perform testing, and then reconstruct the samples sent by the encoder. However, the decision step requires two phases, as summarized in the corresponding constraints present in the error exponent of the next proposition.

Proposition 3 (Achievable rate-error-distortion region)**.** *A tuple* $(R, E, D_0, D_1) \in \mathbb{R}_+^4$*, is achievable for the distributed joint detection and reconstruction problem with general hypotheses, if there exists a positive rate* R' *satisfying:*

$$
\begin{aligned}
&R \geq R' + I\big(X; V_0|UY\big) + I\big(X; V_1|\bar{U}\bar{Y}\big)\ , \\
&E \leq \inf_{Q_X \in \mathcal{P}(\mathcal{X})} \sup_{Q^\star_{U|X}(Q_X) \in \mathcal{P}(\mathcal{U})} \inf_{Q_Y \in \mathcal{P}(\mathcal{Y})} \\
&\qquad \inf_{\substack{Q_{UXY} \in \mathcal{P}(\mathcal{U}\times\mathcal{X}\times\mathcal{Y}) \\ Q_{U|X} = Q^\star_{U|X}}} \Big\{ \min\big[G(Q_{UXY}, Q_X, Q_Y, R'), \\
&\qquad\qquad \min_{\tilde{U}\tilde{X}\tilde{Y} \in \mathcal{L}(Q^\star_{UX}, Q^\star_{UY})} \mathcal{D}\big(P_{\tilde{U}\tilde{X}\tilde{Y}} \| P_{\bar{U}\bar{X}\bar{Y}}\big)\big]\Big\} \\
&D_0 \geq \mathbb{E}_0\left[d_0\big(X, \hat{X}_0(UYV_0)\big)\right]\ , \\
&D_1 \geq \mathbb{E}_1\left[d_1\big(\bar{X}, \hat{X}_1(\bar{U}\bar{Y}V_1)\big)\right]\ .
\end{aligned}
\tag{15}
$$

Here, U *and* $\bar{U}$ *are auxiliary RVs such that* $Q_{U|X}(u|x) = Q_{\bar{U}|\bar{X}}(u|x)\ , \forall (u,x) \in \mathcal{U}\times\mathcal{X}$*,* V_0 *and* V_1 *are auxiliary random variables verifying the Markov chains* $U - V_0 - X - Y$ *and* $\bar{U} - V_1 - \bar{X} - \bar{Y}$ *(along with* U *and* $\bar{U}$ *respectively);* $\mathcal{L}(Q^\star_{UX}, Q^\star_{UY})$ *is the following set of random variables:*

$$
\begin{aligned}
\mathcal{L}(Q^\star_{UX}, Q^\star_{UY}) = \Big\{ & P_{\tilde{U}\tilde{X}\tilde{Y}} \in \mathcal{P}(\mathcal{U}\times\mathcal{X}\times\mathcal{Y}) \big| \\
& P_{\tilde{U}\tilde{X}}(u,x) = Q^\star_{UX}(u,x), \\
& P_{\tilde{U}\tilde{Y}}(u,y) = Q^\star_{UY}(u,y),\ \forall (u,x,y) \Big\}\ ,
\end{aligned}
\tag{16}
$$

where $Q^\star_{UX}, Q^\star_{UY}$ *are joint distributions implied by* Q_X *and the chosen maximizer* $Q^\star_{U|X}$*, and the function* G *appears in* (17)*, at the top pf the next page, with* P_{UXY_i} *defined to be* $P_{UXY_0} = P_{UXY} = P_{XY}Q_{U|X}$ *in the case of hypothesis* 0 *and* $P_{UXY_1} = P_{\bar{U}\bar{X}\bar{Y}} = P_{\bar{X}\bar{Y}}Q_{\bar{U}|\bar{X}}$ *for hypothesis* 1*.*

Proof: The proof is relegated to Appendix D. ∎

We emphasize that when a binning approach is taken, the expression (15) for the error exponent E encapsulates the innate tension between two error events: decoding the description and testing based on it. Provided that a good representation $\mathbf{u}^n$ of the observed samples $\mathbf{x}^n$ at node A is reliably decoded at node B, the statistician is able to perform detection with a very large probability of success. However, such a good representation would also induce a very large size for the codebook, which for a given R would cause each bin to be very large in order to satisfy the rate constraint, making likely errors will appear during the decoding process of the right sequence from the specific bin. On the other hand, when a crude description is chosen, the codebook is smaller and thus so is each bin –if binning is at all necessary. The binning process is therefore not likely to significantly hurt performance, whereas the retrieved representation is much less valuable for the sake of performing the test because of the crude nature of the description supplied by this representation about samples $\mathbf{x}^n$.

In order to ensure the achievability of the error exponent introduced in Proposition 3, we will take a "worst-case" approach. The minimization and maximization operators in the expression for E can thus be read as follows: For every possible vector $\mathbf{x}^n$, the encoder is allowed to choose its strategy of transmission (this is achieved by taking the supremum over $Q^\star_{U|X}$). Having chosen the distribution to generate the codebook, the proposed approach should apply for any type of observed vector $\mathbf{y}^n$, as well as for any joint type $(\mathbf{u}^n, \mathbf{x}^n, \mathbf{y}^n)$, as long as $Q^\star_{U|X}$ is respected. Much like the case of testing against independence, achievability is proven by dividing the problem into two distinct parts: hypothesis testing and source reconstruction. First, a common message –designed to allow detection– is communicated from node A to node B and is then used regardless of the probability distribution in effect which is still unknown at this stage. In order to do so, we choose a decoder based on the empirical entropy, similar to the *Empirical Mutual Information* (MMI) decoder used in compound models (e.g. see [17] and references therein). Two private messages are then transposed upon this common message, each intended to be used (together with the common message) under each of the possible hypotheses. It should be emphasized that dividing the communication in two different phases may well be a suboptimal choice. However, we will see such a choice introduces significant gains in the error exponent.

Remark 4. *Much like in the case of testing against independence (see Remark 2), the assumption that distortion is only measured when correct detection has occurred is convenient but not necessary for the achievability of the region proposed in Proposition 3.*

$$G(Q_{UXY},Q_X,Q_Y,R')=\begin{cases}\min\limits_{i=\{0,1\}}\mathcal{D}\big(Q_{UXY}||P_{UXY_i}\big)+\big[R'-I\big(X;U\big)+I\big(Y;U\big)\big]^+ & I\big(U;X\big)>R'\\ +\infty & \text{else}\ ,\end{cases} \tag{17}$$

IV. Focusing on Hypothesis Testing Only

In this section, we focus on the detection part of the problem only, while still assuming general hypotheses. Although we will show that significant gains can be obtained by introducing binning as suggested in Proposition 3, we next show that the performance of detection can be further improved if source reconstruction is not required by the statistician. We start with the following proposition that uses a different approach for testing without source reconstruction.

Proposition 4 (Improved error exponent for general hypotheses)**.** *A pair (R,E) is an achievable rate and exponent pair for general hypothesis testing, without source reconstruction, provided that:*

$$E\leq\sup_{Q^\star_{U|X}\in\mathcal{P}(\mathcal{U})}\Big\{\min\big\{\hat{G}(Q_{UXY},R)\\ \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{L}(Q^\star_{UX},Q^\star_{UY})}\mathcal{D}\big(P_{\tilde{U}\tilde{X}\tilde{Y}}\|P_{\bar{U}\bar{X}\bar{Y}}\big)\big\}\Big\}\ , \tag{18}$$

where

$$\hat{G}(Q_{UXY},R)=R-\big[I\big(X;U\big)-I\big(U;Y\big)\big] \tag{19}$$

and the set $\mathcal{L}(Q^\star_{UX},Q^\star_{UY})$ is defined by (16)*. It is worth emphasizing that $I\big(U;Y\big)$ in* (18) *is a direct consequence of the choice $Q^\star_{U|X}$ and the distribution implied by H_0, P_{XY}.*

Proof: The proof of this proposition is relegated to Appendix E. ∎

The proof is very similar to that of Proposition 3. We basically derive the probability of error for a specific triplet of sequences $(\mathbf{x}^n,\mathbf{y}^n,\mathbf{u}^n)$, and then calculate the total probability of error by summing over all possible types and corresponding sequences included within each type. The main difference is that now source reconstruction is not required. Thus, instead of first selecting a sequence from within the bin and only then performing the test, we let node B operate over the entirety of the bin. The chosen strategy consists of going over all sequences within the bin. For each sequence $\mathbf{u}_i^n$ for $\{1,\ldots,2^{nR}\}$, we assume it is the correct one and perform the test by checking the typicality of the pair $(\mathbf{u}_i^n,\mathbf{y}^n)$ with relation to the hypothesis H_0. If a sequence is found in a bin such that $(\mathbf{u}_i^n,\mathbf{y}^n)\in T^n_{[UY]_\delta}$, the decoder declares H_0. Otherwise, if no such sequence is found it declares H_1.

As was the case in Proposition 3, Proposition 4 implies that the resulting error exponent is the output of a trade-off between the exponents of the probabilities of two error events. In this case, the trade-off that controls $\beta_n\approx\exp(-nE)$ is between: the probability of erroneous detection while using the right sequence; and the probability of having a different sequence in the bin that is jointly typical with $\mathbf{y}^n$ and thus would make the decoder declare H_0. It turns out, that this trade-off is much preferable to the one offered by Proposition 3, as we can bound the set of sequences that might "confuse" the decoder in a manner that is not dependent on the type of $\mathbf{y}^n$. For instance, the minimizations over Q_X, Q_Y and Q_{UXY} (as seen in Proposition 3) are no longer necessary. This issue has a positive effect on behaviour of the error exponent. Indeed, this new approach takes advantage of the random nature of the binning process. By randomly allocating sequences into bins we allow for bigger codebooks, which provide better descriptions to the original sequence. As long as the size of the bins are not too large, this does not come at a major price (in terms of the chance of "confusing" the decoder), and thus improving significantly the result of [8] in some cases, as can be seen in the example given subsequently. However, the fact that the original sequence sent by the encoder is not retrieved implies that this strategy is not adapted for the joint problem of detection and source reconstruction.

Remark 5. *Another advantage of this strategy over the one given in Proposition 3 is that while knowledge over the probability distribution implied by $P_{\bar{X}\bar{Y}}$ is required in order to analyze performance, such knowledge is not needed in order to perform the test. This stems from the fact that here, the system only tests if H_0 is true or not rather than testing H_0 against H_1.*

A. Binary Symmetric Source

Having proposed two new approaches for distributed testing with general hypotheses, one that allows source reconstruction (Proposition 3) and the other that does not (Proposition 4), it is still not clear whether binning is strictly beneficial for such problems. As was demonstrated in Section II, binning for testing is not necessary to achieve optimality in the case of testing against independence. One may further argue that as binning introduces additional error events, it is not clear whether or not it would be beneficial at all in the case of general hypotheses.

In the following, we investigate the benefits of binning through a Binary Symmetric Source (BSS). While it is analytically clear that detection through the strategy offered by Proposition 4 is superior to the one offered in Proposition 3, we show that for some specific cases both approaches may result in performance gain relative to non-binning approaches. For the sake of simplicity, we consider the following lower bound over the performance, throughout the following numerical analysis [8]:

$$\min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{L}(Q^\star_{UX},Q_{UY})}\mathcal{D}(P_{\tilde{U}\tilde{X}\tilde{Y}}\|P_{\bar{U}\bar{X}\bar{Y}})\geq\mathcal{D}(P_{UY}\|P_{\bar{U}\bar{Y}})\ . \tag{20}$$

Consider the following statistical model: Let $X\sim\text{Bern}\left(\frac{1}{2}\right)$, and

$$\begin{cases}H_0: & Y=X+Z_0, \quad Z_0\sim\text{Bern}(p)\perp X\\ H_1: & Y=X+Z_1, \quad Z_1\sim\text{Bern}(q)\perp X\ ,\end{cases} \tag{21}$$

where $1 \geq q > p \geq 0$. Note that while H_1 does not imply independence between X and Y, the marginal distribution of Y is equal for both hypotheses, making a decision without cooperation impossible. This model was studied first in Wyner-Ziv [24] for source reconstruction. The optimal rate-distortion region (asymptotic regime) was shown to be

$$\begin{cases} R(D) = \inf\limits_{\theta,\delta} \left[\theta \left(H_2(p \star \delta) - H_2(\delta)\right)\right] \ , \\ D = \theta\delta + (1-\theta)p \ , \end{cases} \tag{22}$$

where p is the crossover probability between the source X and the side information Y, and $p \star \delta$ is the binary convolution of p and δ. The parameters satisfy $0 \leq \theta \leq 1$ and $0 \leq \delta \leq 0.5$. The achievability of this region was shown by using time-sharing between two strategies - in the first the auxiliary RV U is the result of passing X through a Binary Symmetry Channel (BSC) with transition probability δ, while in the second U is degenerate.

We now apply Proposition 3 to this setup, we choose to consider only distributions in which Q_X is a BSS, and U is the result of passing X through a BSC with crossover probability δ. While this is not necessarily an optimal choice, it can be justified as an optimal approach for the asymptotic regime, at least. To evaluate the resulting error exponent, we need to calculate two values. The first is given by:

$$\inf_{Q_Y} \inf_{\substack{Q_{UXY} \\ Q_{U|X}=Q^\star_{U|X}}} G(Q_{UXY}, R) \ , \tag{23}$$

as a function of $Q^\star_{U|X}$ (which, under our assumptions, boils down to be a function of δ). This expression encapsulates the error exponent of the event where the wrong sequence is chosen from the bin. The second quantity to calculate is given by:

$$\min_{\tilde{U}\tilde{X}\tilde{Y} \in \mathcal{L}(Q^\star_{UX}, Q_{UY})} \mathcal{D}(P_{\tilde{U}\tilde{X}\tilde{Y}} \| P_{\bar{U}\bar{X}\bar{Y}}) \geq \mathcal{D}(P_{UY} \| P_{\bar{U}\bar{Y}}) \ , \tag{24}$$

also as a function of $Q^\star_{U|X}$. This expression represents the error exponent of the event where, while using the right sequence, an error occurs during the detection process. Having calculated these two functions, we can pick $Q^\star_{U|X}$ such that the "minimum" between the two is "maximized".

The results implied by Proposition 4 can be calculated in a very similar fashion. Now, the trade-off is between the same curve representing the error while using the correct sequence as was mentioned in (24), and the curve implied by $\hat{G}$, representing the event of an error caused through the testing of a different sequence.

B. Numerical Results

A visualization of the performance achieved by each of the proposed methods for general hypotheses is plotted in Fig. 3, for the above discussed statistical model. We choose to consider only distributions in which Q_X is a BSS and $Q^\star_{U|X}$ represents a BSC with transition probability δ, as explained above. The "hypothesis testing" curve represents the error exponent of the probability of the event where a mistake is made in detection, when the correct sequence is used from the bin. This curve is relevant for both methods of detection, namely Proposition 3 and Proposition 4.

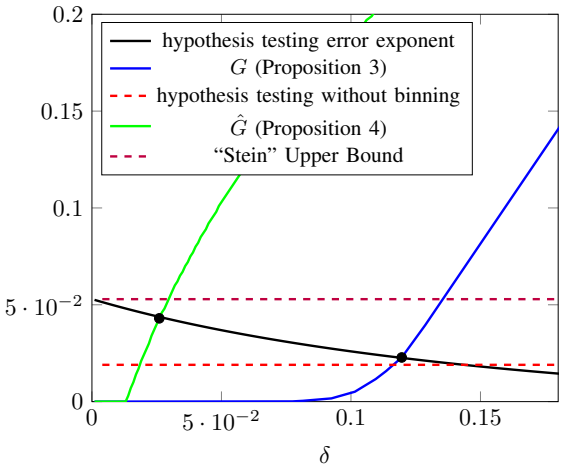

Figure 3. Error exponents for both error events in the BSC case with $p = 0.1$, $q = 0.2$, $R = 0.4$, under the strategies implied by Propositions 3 and 4. The resulting error exponent for each δ is the minimum between the two. Performance with a non-binned codebook is represented by a dashed line.

The interesting tension that exists between the two error events, denoted by either G (Proposition 3) or $\hat{G}$ (Proposition 4) and an error exponent corresponding to testing, is represented by the worst case between those curves. When δ is very small, a sequence $\mathbf{u}^n$ can be found with high probability, such that $\mathbf{x}^n$ is very well described, and the codebook contains many sequences $\mathbf{u}^n$. Thus, given the right sequence $\mathbf{u}^n$, the error event during the test is not likely, and the error exponent of the event where the test fails is high. However, since the rate of communication is fixed, each bin has to contain many sequences in case δ is small, increasing the error probability in decoding the right sequence. When δ grows, the accuracy of the description of $\mathbf{x}^n$ by $\mathbf{u}^n$ is lower, making the probability of error of the test, while using the correct sequence, higher. The codebook, however, is smaller, making the task of choosing the right sequence in the bin easier. Note that the error exponent for choosing the sequence from within the bin has a threshold, under which it is zero. This threshold in this case is roughly $\delta \approx 0.08$, which is the value implied by [24] as the minimal value for the binning approach, in the asymptotic regime.

Similarly, the trade-off between the two error events represented by Proposition 4 is apparent through the curve of the error exponent related to the testing errors, along with the "binning error exponent" denoted by the curve $\hat{G}$. Now, the additional error event –other than committing an error while using the correct sequence which turns out to be the same as before– is the event where a different sequence in the bin "confuses" the decoder by being jointly typical with $\mathbf{y}^n$. While this curve is lower bounded by the curve representing G for all cases, it can be seen that in the present case this approach is largely superior. As under both approaches we are allowed to select the strategy $Q^\star_{U|X}$ (in this specific case δ) freely, the optimal approach under each of the propositions would be to choose the corresponding intersection point between the curve representing G or $\hat{G}$ and the curve entitled "Hypothesis Testing Error Exponent" in Fig. 3. These two points are marked in

Fig. 3 with black dots.

In addition, a lower bound can be found in Fig. 3. We emphasize that this bound is not drawn as a function of δ but rather depicts the best possible performance under the assumptions detailed above, when binning is not performed, as was done in [8]. Thus, δ is chosen to be the smallest possible, such that the size of the codebook would not exceed the available rate of communication. A trivial upper bound is also drawn by providing $\mathbf{x}^n$ to node B and then applying Stein's Lemma.

V. Summary and Discussion

We studied the joint problem of distributed detection and lossy compression with side information. This scenario arises when an authentication system prevents the unauthorized injection of messages into a public channel, assuring the receiver of a message of the legitimacy of its sender. In this setup a user (referred to as node A) is required to communicate a lossy description of a memoryless source to a statistician (referred to as node B) whose task is to verify that the encoding user is the individual he claims to be and then according to its identity to reconstruct the message based on the adequate distortion measure, much like in [9], [10]. However, in the setup considered here the receiver is unaware of the value of its information as well, which leads to a two-step approach where first a decision has to be made about the identity of node A before source reconstruction can take place.

When testing against independence, this two-step approach turns out to be optimal. In this case, detection can be performed optimally as in [7], while source remonstration is performed à la Wyner-Ziv [24], and the two-step approach does not induce performance degradation. An application example to a binary symmetric source was also shown for which the optimal region was explicitly derived, emphasizing an interesting tension between the error exponent corresponding to the (second type) error probability and the average distortion measure.

When testing with general hypotheses, a similar, albeit more involved, approach produced a new achievable rate-error-distortion region. Here, optimality may be hard to reach, as optimality results stay elusive even in the case where the receiver is aware of the value of the side information (see [27] and references therein). Nevertheless, we showed that the two-step approach, which was optimal in the case of testing against independence, induces in the general case a significant loss in performance. It was shown that when source reconstruction is not required, valuable information for testing can be compressed much further than in the opposite case, improving significantly the performance of detection.

Although there are several other threats to authentication systems which require recourse to more sophisticated models and techniques than the ones investigated here, this work attempts a modest step in the direction of a theory for distributed testing based on lossy data compression which seems to offer a formidable mathematical complexity.

Appendix A
Typical Sequences and Related Results

In this appendix we introduce standard notions in information theory, suited for the mathematical developments and proofs needed in this work. The results presented can be easily derived from the standard formulations provided in [26], [28], [29]. Let $\mathcal{X}$ and $\mathcal{Y}$ be finite alphabets and $(\mathbf{x}^n, \mathbf{y}^n) \in \mathcal{X}^n \times \mathcal{Y}^n$. With $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ we denote the set of all joint probability distributions on $\mathcal{X} \times \mathcal{Y}$. We define the *$\delta$-typical* sets, with relation to the pmf $p_X \in \mathcal{P}$, as:

Definition 4 (Typical set)**.** *Consider $p \in \mathcal{P}(\mathcal{X})$ and $\delta > 0$. We say that $\mathbf{x}^n \in \mathcal{X}^n$ is δ- typical if $\mathbf{x}^n \in \mathcal{T}^n_{[X]\delta}$ with:*

$$\mathcal{T}^n_{[X]\delta} = \Big\{ \mathbf{x}^n \in \mathcal{X}^n : \left| Q_{\mathbf{x}^n}(a) - p_X(a) \right| \leq \delta \ , \quad \forall a \in \mathcal{X} \text{ such that } p(a) \neq 0 \Big\}, \tag{25}$$

where $Q_{\mathbf{x}^n}(a) = n^{-1} N(a|\mathbf{x}^n)$ is the type of $\mathbf{x}^n$ and $N(a|\mathbf{x}^n)$ denotes de number of occurrences of $a \in \mathcal{X}$ in $\mathbf{x}^n$.

Definition 5 (Joint and conditional typical sets)**.** *In a similar manner to Definition 4, given $p_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ we can construct the set of δ-jointly typical sequences as:*

$$\mathcal{T}^n_{[XY]\delta} = \Big\{ (\mathbf{x}^n, \mathbf{y}^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \left| Q_{\mathbf{x}^n \mathbf{y}^n}(a,b) - p_{XY}(a,b) \right| \leq \delta, \quad \forall (a,b) \in \mathcal{X} \times \mathcal{Y} \ \text{ such that } \ p_{Y|X}(b|a) Q_{\mathbf{x}^n}(a) \neq 0 \Big\} \ . \tag{26}$$

We also define the conditional *typical sequences. In precise terms, given $\mathbf{x}^n \in \mathcal{X}^n$ we consider the set:*

$$\mathcal{T}^n_{[Y|X]\delta}(\mathbf{x}^n) = \Big\{ \mathbf{y}^n \in \mathcal{Y}^n : \left| Q_{\mathbf{x}^n \mathbf{y}^n}(a,b) - p_{Y|X}(b|a) Q_{\mathbf{x}^n}(a) \right| \leq \delta, \quad \forall (a,b) \in \mathcal{X} \times \mathcal{Y} \text{ such that } p_{Y|X}(b|a) Q_{\mathbf{x}^n}(a) \neq 0 \Big\} \ .$$

We present the following lemmas without proof.

Lemma 1 (Properties of typical sets [29])**.** *The following statements hold:*

1) *Consider $(\mathbf{x}^n, \mathbf{y}^n) \in \mathcal{T}^n_{[XY]\epsilon}$. Then, $\mathbf{x}^n \in \mathcal{T}^n_{[X]\epsilon}$, $\mathbf{y}^n \in \mathcal{T}_{[Y]\epsilon}$, $\mathbf{x}^n \in \mathcal{T}^n_{X|Y\epsilon}(\mathbf{y}^n)$ and $\mathbf{y}^n \in \mathcal{T}^n_{[Y|X]\epsilon}(\mathbf{x}^n)$.*
2) *Be $(\mathbf{X}^n, \mathbf{Y}^n) \sim \prod_{t=1}^n p_{XY}(x_t, y_t)$. If $\mathbf{x}^n \in \mathcal{T}^n_{[X]\epsilon}$ we have*

$$\exp\{-n(H(X) + \delta(\epsilon))\} \leq p_{\mathbf{X}^n}(\mathbf{x}^n) \leq \exp\{-n(H(X) - \delta(\epsilon))\} \tag{27}$$

with $\delta(\epsilon) \to 0$ when $\epsilon \to 0$. Similarly, if $\mathbf{y}^n \in \mathcal{T}^n_{[Y|X]\epsilon}(\mathbf{x}^n)$:

$$\exp\{-n(H(Y|X) + \delta'(\epsilon))\} \leq p_{\mathbf{Y}^n|\mathbf{X}^n}(\mathbf{y}^n|\mathbf{x}^n) \leq \exp\{-n(H(Y|X) - \delta'(\epsilon))\} \tag{28}$$

with $\delta'(\epsilon) \to 0$ when $\epsilon \to 0$.

Proof: See [29, Chapter 2.5]. ∎

Lemma 2 (Conditional typicality lemma [29]). *Consider the product measure $\prod_{t=1}^{n} p_{XY}(x_t, y_t)$, the following result hold true*

$$\Pr\left\{\mathcal{T}^n_{[X]_\epsilon}\right\} \geq 1 - \mathcal{O}\left(\frac{1}{n\epsilon^2}\right),$$
$$\Pr\left\{\mathcal{T}^n_{[Y|X]_\epsilon}(\mathbf{x}^n)|\mathbf{x}^n\right\} \geq 1 - \mathcal{O}\left(\frac{1}{n\epsilon^2}\right),$$
$$\text{for every } \mathbf{x}^n \in \mathcal{X}^n,$$

where $(n\epsilon^2) \to \infty$ when $\epsilon \to 0$ and $n \to \infty$.

Proof: See [29, Chapter 2.5]. ∎

Lemma 3 (Size of typical sets [26]). *For any type $Q \in \mathcal{P}_n(\mathcal{X})$*

$$|\mathcal{P}_n(\mathcal{X})|^{-1} \exp\big(nH(Q)\big) \leq |\mathcal{T}^n_Q| \leq \exp\big(nH(Q)\big) .$$

The size of the set of all empirical distributions (or types) of X and of length n can be calculated to be

$$|\mathcal{P}_n(\mathcal{X})| = \binom{n + |\mathcal{X}| - 1}{|\mathcal{X}| - 1} \leq (n+1)^{|\mathcal{X}|} ,$$

yielding the following bound

$$(n+1)^{-|\mathcal{X}|} \exp\big(nH(Q)\big) \leq |\mathcal{T}^n_Q| \leq \exp\big(nH(Q)\big) .$$

Lemma 4. *For every probability measure $P_X \in \mathcal{P}(\mathcal{X})$ and stochastic mapping $W : \mathcal{X} \mapsto \mathcal{P}(\mathcal{Y})$, there exist sequences $(\varepsilon_n)_{n\in\mathbb{N}_+}, (\varepsilon'_n)_{n\in\mathbb{N}_+} \to 0$ as $n \to \infty$ satisfying:*

$$\begin{aligned} &\left|\frac{1}{n}\log|\mathcal{T}_{[X]_\epsilon}| - H(X)\right| \leq \varepsilon_n , \\ &\left|\frac{1}{n}\log|\mathcal{T}_{[Y|X]_\epsilon}(\mathbf{x})| - H(Y|X)\right| \leq \varepsilon_n , \end{aligned} \tag{29}$$

for each $\mathbf{x} \in \mathcal{T}_{[X]_\epsilon}$ where $\varepsilon_n \equiv \mathcal{O}(n^{-1}\log n)$, and

$$\begin{aligned} &P^n_X\big(\mathcal{T}_{[X]_\epsilon}\big) \geq 1 - \varepsilon'_n , \\ &W^n\big(\mathcal{T}_{[Y|X]_\epsilon}(\mathbf{x})|X^n = \mathbf{x}\big) \geq 1 - \varepsilon'_n , \end{aligned} \tag{30}$$

for all $\mathbf{x} \in \mathcal{X}^n$ where $\varepsilon'_n \equiv \mathcal{O}\left(\frac{1}{n\epsilon^2}\right)$, provided that n is sufficiently large.

Proof: Refer to reference [29, Lemma 2.13] ∎

Lemma 5 (Set of sequences with small empirical entropy [25]). *For any pair of strings of length n, denoted by $(\mathbf{x}^n, \mathbf{y}^n)$, let*

$$\mathcal{S}(\mathbf{x}^n, \mathbf{y}^n) = \left\{(\tilde{\mathbf{x}}^n, \tilde{\mathbf{y}}^n) \in \mathcal{X}^n \times \mathcal{Y}^n \,\middle|\, H(\tilde{\mathbf{x}}^n, \tilde{\mathbf{y}}^n) \leq H(\mathbf{x}^n, \mathbf{y}^n)\right\} ,$$

with $H(\mathbf{x}^n, \mathbf{y}^n)$ being the empirical entropy of the sequences,

$$H(\mathbf{x}^n, \mathbf{y}^n) = -\sum_{a\in\mathcal{X}, b\in\mathcal{Y}} Q_{\mathbf{x}^n\mathbf{y}^n}(a,b)\log Q_{\mathbf{x}^n\mathbf{y}^n}(a,b) .$$

Then

$$|\mathcal{S}(\mathbf{x}^n, \mathbf{y}^n)| \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|}\exp\big[nH(\mathbf{x}^n, \mathbf{y}^n)\big] .$$

Let

$$\mathcal{S}(\mathbf{x}^n|\mathbf{y}^n) = \left\{\tilde{\mathbf{x}}^n \in \mathcal{X}^n \,|\, H(\tilde{\mathbf{x}}^n|\mathbf{y}^n) \leq H(\mathbf{x}^n|\mathbf{y}^n)\right\} ,$$

then

$$|\mathcal{S}(\mathbf{x}^n|\mathbf{y}^n))| \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|}\exp\big[H(\mathbf{x}^n|\mathbf{y}^n)\big] .$$

Lemma 6 (Generalized Markov Lemma [30]). *Let $p_{UXY} \in \mathcal{P}(\mathcal{U}\times\mathcal{X}\times\mathcal{Y})$ be a probability measure that satisfies: $U \multimap X \multimap Y$. Consider $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}^n_{[XY]_{\epsilon'}}$ and random vectors $\mathbf{U}^n$ generated according to:*

$$\Pr\left\{\mathbf{U}^n = \mathbf{u}\middle|U^n \in \mathcal{T}^n_{[U|X]_{\epsilon''}}(\mathbf{x}), \mathbf{x}, \mathbf{y}\right\} = \frac{\mathbb{1}\left\{\mathbf{u}^n \in \mathcal{T}^n_{[U|X]_{\epsilon''}}(\mathbf{x})\right\}}{\left|\mathcal{T}^n_{[U|X]_{\epsilon''}}(\mathbf{x})\right|} . \tag{31}$$

For sufficiently small $\epsilon, \epsilon', \epsilon'' > 0$,

$$\begin{aligned} &\Pr\left\{\mathbf{U}^n \notin \mathcal{T}^n_{[U|XY]_\epsilon}(\mathbf{x}, \mathbf{y})\middle|\mathbf{U}^n \in \mathcal{T}^n_{[U|X]_{\epsilon''}}(\mathbf{x}), \mathbf{x}, \mathbf{y}\right\} \\ &\quad\equiv \mathcal{O}\left(c^{-n}\right) \end{aligned} \tag{32}$$

holds uniformly on $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}^n_{[XY]_{\epsilon'}}$ where $c > 1$.

Lemma 7 (Joint Typicality Lemma [28]). *Let $(X, Y, Z) \sim p(x, y, z)$ and $\epsilon' < \epsilon$. Then there exist $\delta(\epsilon) > 0$ that tends to 0 as $\epsilon \to 0$ such that the following statements hold:*

1) *If $(\mathbf{x}, \mathbf{y}^n)$ is a pair of arbitrary sequences and $\mathbf{Z}^n \sim \prod_{i=1}^{n} p_{Z|X}(z_i|x_i)$ then*

$$\begin{aligned} &\Pr\{(\mathbf{x}^n, \mathbf{y}^n, \mathbf{Z}^n) \in \mathcal{T}^n_{[XYZ]\epsilon}\} \leq \\ &\qquad \exp\{-n(I(Y;Z|X) - \delta(\epsilon))\} . \end{aligned} \tag{33}$$

2) *If $(\mathbf{x}^n, \mathbf{y}^n) \in \mathcal{T}^n_{[XY]\epsilon'}$ and $\mathbf{Z}^n \sim \prod_{i=1}^{n} p_{Z|X}(z_i|x_i)$, then for n sufficiently large*

$$\begin{aligned} &\Pr\{(\mathbf{x}^n, \mathbf{y}^n, \mathbf{Z}^n) \in \mathcal{T}^n_{[XYZ]\epsilon}\} \leq \\ &\qquad \exp\{-n(I(Y;Z|X) - \delta(\epsilon))\} . \end{aligned} \tag{34}$$

Appendix B
Proof of Proposition 1

In this appendix, we prove the achievability and converse to Proposition 1.

Achievability proof

Codebook generation: Fix a conditional probability distribution $Q_{VU|XY} = Q_{V|UX}Q_{U|X}P_{XY}$ such that $U \multimap V \multimap X \multimap Y$ form a Markov chain. Let $Q_U(u) = \sum_{x\in\mathcal{X}} P_X(x)Q_{U|X}(u|x)$ and $Q_{V|U}(v|u) = \sum_{x\in\mathcal{X}} Q_{V|UX}(v|u,x)$. Let the total available rate of communication R be divided into two, such that the parts are dedicated to U and V, which represent the different parts of the message. Denote the rate dedicated to the transmission of U by $\hat{R}$, while the rate dedicated to the transmission of V is denoted by R'. Randomly and independently generate $\exp(n\hat{R})$ sequences $\mathbf{u}$ through the i.i.d. pmf $Q_U(u)$, with replacement, such that $\mathbf{u}(s_1) \in \mathcal{T}_{[U]\delta}$, $\forall s_1$, with $s_1 \in [1 : \exp(n\hat{R})]$. For each codeword $\mathbf{u}(s_1)$, randomly and independently generate $\exp(nS_2)$ sequences denoted by $\mathbf{v}^n(s_1, s_2)$ and indexed with $s_2 \in [1 : \exp(nS_2)]$ by using the conditional pmf $Q_{V|U}(\cdot|\mathbf{u}(s_1))$, with replacement, such that $\mathbf{v}(s_1, s_2) \in \mathcal{T}_{[V|U]\delta}(\mathbf{u}(s_1))$. Divide theses sequences

into $\exp[nR']$ bins, such that each bin contains roughly $\exp[n(S_2 - R')]$ sequences.

Encoding: Assuming that the source sequence $\mathbf{x}^n$ is produced from X, look for the first codeword in U's codebook such that $(\mathbf{u}^n(s_1), \mathbf{x}^n) \in \mathcal{T}^n_{[UX]\delta}$. Then, look for the first codeword $\mathbf{v}^n(s_1, s_2)$ s.t. $(\mathbf{v}^n(s_1, s_2), \mathbf{x}^n) \in \mathcal{T}^n_{[VX|U]\delta}(\mathbf{u}(s_1))$. Let b be the bin of $\mathbf{v}^n(s_1, s_2)$. Send the message $f(\mathbf{x}^n) = (s_1, b)$ to node B.

Decoding: Given $\mathbf{u}(s_1), b$ and $\mathbf{y}^n$, the decoder first checks if $(\mathbf{u}^n(s_1), \mathbf{y}^n) \in \mathcal{T}^n_{[UY]\delta}$. If so, it declares H_0 and otherwise it declares H_1. If the decoder decides H_0, it then attempts to decode the message (with average distortion D) based on $\mathbf{v}(s_1, s_2)$. This codeword is first recovered by looking in the bin b for the unique codeword such that $\mathbf{v}^n(s_1, s_2) \in \mathcal{T}^n_{[V|UY]\delta}(\mathbf{u}(s_1), \mathbf{y}^n)$. Then, a per-letter function $g(\cdot)$ is applied over the entire available information (U, V and Y) in order to produce a reconstruction of the source.

Error events and constraints: We start with the HT part, and the relation between the expression $I(U;X)$ and the achievable error exponent. Denoting by $\mathcal{B}_0$ the event "an error occurred during encoding" (of the HT part U), we expand its probability as $\Pr(\mathcal{B}_0) \leq \Pr(\mathcal{B}_1) + \Pr(\mathcal{B}_2)$ with:

$$\begin{aligned}\Pr(\mathcal{B}_1) &\triangleq \Pr\{\mathbf{X}^n \notin \mathcal{T}^n_{[X]\delta}\} \ , \\ \Pr(\mathcal{B}_2) &\triangleq \Pr\{\nexists s_1 \text{ s.t. } (\mathbf{u}(s_1), \mathbf{X}^n) \in \mathcal{T}^n_{[UX]\delta} | \\ &\qquad\qquad \mathbf{X}^n \in T^n_{[X]\delta}\} \ ,\end{aligned} \tag{35}$$

being the probabilities that the source X produces a non-typical sequence, and that (for a typical source sequence) the codebook doesn't contain an appropriate codeword, respectively. From the Asymptotic Equipartition Property (AEP), $\Pr(\mathcal{B}_1) \leq \eta^{(1)}_n \underset{n\to\infty}{\longrightarrow} 0$. As for $\Pr(\mathcal{B}_2)$:

$$\begin{aligned}\Pr(\mathcal{B}_2) &= && (36a)\\ &= \Big(\Pr\{(\mathbf{U}^n, \mathbf{X}^n) \notin \mathcal{T}^n_{[UX]\delta} | && (36b)\\ &\qquad \mathbf{U}^n \in \mathcal{T}^n_{[U]\delta}, \mathbf{X}^n \in \mathcal{T}^n_{[X]\delta}\}\Big)^{\exp(n\hat{R})} \\ &= \Big(1 - \Pr\{(\mathbf{U}^n, \mathbf{X}^n) \in \mathcal{T}^n_{[UX]\delta} | && (36c)\\ &\qquad \mathbf{U}^n \in \mathcal{T}^n_{[U]\delta}, \mathbf{X}^n \in \mathcal{T}^n_{[X]\delta}\}\Big)^{\exp(n\hat{R})} \\ &\leq \exp[-\exp(n\hat{R})\Pr\{(\mathbf{U}^n, \mathbf{X}^n) \in \mathcal{T}^n_{[UX]\delta} | \\ &\qquad \mathbf{U}^n \in \mathcal{T}^n_{[U]\delta}, \mathbf{X}^n \in T^n_{[X]\delta}\}] && (36d)\\ &\leq \exp[-\exp(n\hat{R})\exp(-n\big(I(U;X) + \eta^{(2)}_n\big))] && (36e)\\ &= \exp\{-\exp[-n\big(I(U;X) - \hat{R} + \eta^{(2)}_n\big)]\} \ . && (36f)\end{aligned}$$

Here, inequality (36d) is due to the inequality $(1-a)^n \leq \exp(an)$ [4]. Since $\eta^{(2)}_n \underset{n\to\infty}{\longrightarrow} 0$, $\Pr(\mathcal{B}_2) \to 0$ if $\hat{R} > I(U;X)$.

Analysis of α_n: Calculating the probability of error of the first type, α_n, boils down to the following:

$$\begin{aligned}\alpha_n &= \Pr(H_1 | XY \sim P_{XY}) && (37a)\\ &\leq \Pr(\mathcal{B}_0) && (37b)\\ &\quad + \Pr\{(\mathbf{U}^n, \mathbf{Y}^n) \notin \mathcal{T}^n_{[UY]\delta} | \\ &\qquad \mathbf{U}^n \in \mathcal{T}^n_{[U]\delta}, (\mathbf{U}^n, \mathbf{X}^n) \in \mathcal{T}^n_{[UX]\delta}, XY \sim P_{XY}\} && (37c)\\ &\leq \Pr(\mathcal{B}_0) + \eta^{(3)} \ . && (37d)\end{aligned}$$

Here, (37c) is due to the fact that when calculating the probability of error of Type I, we may assume that the true distribution controlling the RVs is the one implied by hypothesis 0. (37d), with $\eta^{(3)} \to 0$, is due to the Generalized Markov Lemma (see Lemma 6 in Appendix A). Thus, it may be concluded that $\alpha_n \to 0$ when $n \to \infty$, and thus $\alpha_n \leq \epsilon$ for any constraint $\epsilon > 0$ and n large enough.

Analysis of β_n: Next, we look at the achievable error exponent of Type II with the proposed encoding scheme. For the sake of this analysis, we can assume that hypothesis H_1 is the correct one. We will follow steps similar to the ones used in [13]:

$$\beta_n = \Pr(H_0 | XY \sim P_X P_Y) = \Pr(\mathcal{B}_1^c)\Pr(\mathcal{B}_0^c | \mathcal{B}_1^c) \ , \tag{38}$$

where the event $\mathcal{B}_0$ is defined by

$$\mathcal{B}_0 = \{(\mathbf{U}(s_1), \mathbf{Y}) \notin \mathcal{T}^n_{[UY]\delta'}\} \tag{39}$$

to be the event that the *chosen* sequence $\mathbf{U}(s_1)$ is not jointly typical with the observed sequence $\mathbf{Y}$. The term $\Pr(\mathcal{B}_1^c)$ goes to 1 when n is large thanks to Lemma 2. Note that this also means that with large probability an index s_1 is chosen out of the codebook, thanks to the Covering Lemma [28] and the fact that we enforce $\hat{R} \geq I(U;X)$. Moreover, note that even if a sequence s_1 cannot be found in the codebook, this *does not constitute a problem* for the analysis of β_n, as in this case the decoder declares H_1.

The term $\Pr(\mathcal{B}_0^c | \mathcal{B}_1^c)$ can be developed through the Joint Typicality Lemma (see Lemma 7) as follows:

$$\Pr(\mathcal{B}_0^c | \mathcal{B}_1^c) \leq \exp\{-n(I(U;Y) - \epsilon(\delta'))\} \ , \tag{40}$$

for n large enough and with $\epsilon(\delta') \to 0$ as $\delta' \to 0$. Thus $-\lim \frac{1}{n}\log\beta_n \geq I(U;Y) - \epsilon(\delta')$ when n is large enough, which completes the achievability of the desired error exponent.

Analysis of the Estimation Phase: Finally, we show that given a (correct) decision H_0, the RV V can be used to decode $\mathbf{X}^n$ with the desired distortion: Denoting by $\mathcal{B}_3$ the event "an error occurred during encoding or decoding" (of V), we expand its probability as follows $\Pr(\mathcal{B}_3) \leq \Pr(\mathcal{B}_4) + \Pr(\mathcal{B}_5)$, with $\Pr(\mathcal{B}_4)$ being the probability that no codeword $\mathbf{v}(s_1, s_2)$ could be found in the codebook for the given sequence $\mathbf{x}^n$ and the chosen codeword $\mathbf{u}(s_1)$, and $\Pr(\mathcal{B}_5)$ being the probability

that a different codeword in the same bin b is compatible with $\mathbf{y}^n$ and $\mathbf{u}(s_1)$.

$$\begin{aligned}
&\Pr(\mathcal{B}_4) \\
&\triangleq \Pr\{\nexists s_2 \text{ s.t. } (\mathbf{v}^n(s_1,s_2),\mathbf{x}^n)\in\mathcal{T}^n_{[VX|U]\delta}(\mathbf{u}^n(s_1))\} \\
&= \Big[\Pr\{(\mathbf{V}^n,\mathbf{X}^n)\notin\mathcal{T}^n_{[VX|U]\delta}(\mathbf{u}(s_1))| \\
&\qquad V^n\in\mathcal{T}^n_{[V|U]\delta}(\mathbf{u}(s_1)),\mathbf{X}^n\in\mathcal{T}^n_{[X]\delta}(\mathbf{u}(s_1))\}\Big]^{\exp(nS_2)} \\
&\le \exp\Big\{-\exp(nS_2)\exp[-n\Big(I(V;X|U)+\eta_n^{(6)}\Big)]\Big\} \\
&= \exp\Big\{-\exp[-n\Big(I(V;X|U)-S_2+\eta_n^{(6)}\Big)]\Big\}\ .
\end{aligned} \tag{41}$$

Thus, $\Pr(\mathcal{B}_4)\underset{n\to\infty}{\longrightarrow}0$ if $S_2>I(V;X|U)$. Finally,

$$\begin{aligned}
\Pr(\mathcal{B}_5)\triangleq\Pr\{&\exists s_2'\in b \\
&\text{s.t. } \mathbf{v}^n(s_1,s_2')\in\mathcal{T}^n_{[V|UY]\delta}(\mathbf{u}^n(s_1),\mathbf{y}^n)\}\ ,
\end{aligned} \tag{42}$$

with b being the bin sent to node B.

$$\begin{aligned}
\Pr(\mathcal{B}_5) &\le \exp[n\,(S_2-R'+\epsilon)] \\
&\quad\times\Pr\{\mathbf{V}^n\in\mathcal{T}^n_{[V|UY]\delta}(\mathbf{u}^n(s_1),\mathbf{y}^n)| \\
&\qquad V^n\in\mathcal{T}^n_{[V|U]\delta}(\mathbf{u}^n(s_1))\} \\
&\le \exp[n\,(S_2-R'+\epsilon)] \\
&\quad\times\exp[-n\Big(I(V;Y|U)+\eta_n^{(7)}\Big)] \\
&= \exp\Big[-n\Big(I(V;Y|U)-(S_2-R')+\eta_n^{(7)}-\epsilon\Big)\Big]\ .
\end{aligned} \tag{43}$$

Thus, $\Pr(\mathcal{B}_5)\underset{n\to\infty}{\longrightarrow}0$ if $S_2-R'<I(V;Y|U)$, or equivalently

$$\begin{aligned}
R' &> S_2-I(V;Y|U)>I(V;X|U)-I(V;Y|U) && \text{(44a)}\\
&= I(V;XY|U)-I(V;Y|U)=I(V;X|UY)\ , && \text{(44b)}
\end{aligned}$$

where equality (44b) stems from the Markov chain $U\multimap V\multimap X\multimap Y$. Thus, since the total rate R is composed of $\hat{R}$ and R', we conclude that our scheme is achievable if $R>I(U;X)+I(V;X|UY)$.[1]

We now know that our scheme allows the decoding of $\mathbf{v}^n$ with high probability when the rate is large enough. It remains to be shown that V (together with U and Y, which are also known at node B) is enough to recover X with average distortion D. We choose a (possibly suboptimal) decoder, that decodes x_i only from (u_i,v_i) and y_i:

$$\begin{aligned}
d\big(\mathbf{x}^n,\hat{\mathbf{x}}^n(\mathbf{u}^n,\mathbf{v}^n,\mathbf{y}^n)\big) &= \frac{1}{n}\sum_{i=1}^n d\big(x_i,\hat{x}(u_i,v_i,y_i)\big) && \text{(45a)}\\
&= \sum_{\forall(x,u,v,y)} d\big(x,\hat{x}(u,v,y)\big)Q_{\mathbf{x}^n\mathbf{u}^n\mathbf{v}^n\mathbf{y}^n}(x,u,v,y) && \text{(45b)}\\
&\le \mathbb{E}_0\Big[d(X,\hat{X}(UVY))\Big] && \text{(45c)}\\
&\quad+\sum_{\forall(x,u,v,y)}|Q_{\mathbf{x}^n\mathbf{u}^n\mathbf{v}^n\mathbf{y}^n}(x,u,v,y)-p(x,u,v,y)| && \text{(45d)}\\
&\le \mathbb{E}_0\Big[d(X,\hat{X}(UVY))\Big]+d_{\max}|\mathcal{X}||\mathcal{U}||\mathcal{V}||\mathcal{Y}|\delta_n\ , && \text{(45e)}
\end{aligned}$$

[1]We explicitly ignored an additional error event, which is that $\mathbf{y}^n$ is not typical. The probability of this event goes to 0 much like $\Pr(\mathcal{B}_1)$, thanks to the AEP.

where the summation in (45b) and (45d) is over all the possible letters in the respective alphabets of the RVs $(x,u,v,y)\in\mathcal{X}\times\mathcal{U}\times\mathcal{V}\times\mathcal{Y}$ and inequality (45e) holds since $(\mathbf{x}^n,\mathbf{u}^n,\mathbf{v}^n,\mathbf{y}^n)\in\mathcal{T}^n_{[XUVY]\delta}$. Since $\delta_n\underset{n\to\infty}{\longrightarrow}0$, the condition $D>\mathbb{E}_0\Big[d\big(X,\hat{X}(UVY)\big)\Big]$ is sufficient to achieve distortion $D+\epsilon$ at node B. This concludes the proof of achievability.

Converse proof

For this part of the proof we use the *multi-letter* converse result in [7], which states that when no estimation is required,

$$\limsup_{n\to\infty}\frac{1}{n}\log\|f_n\|\le R\ , \tag{46}$$

$$\liminf_{n\to\infty}\frac{1}{n}I\left(f_n(\mathbf{X}^n);\mathbf{Y}^n\right)\ge E\ . \tag{47}$$

Clearly, this rate-error relationship cannot be beat when *an additional constraint* (in this case, relating to the estimation requirement) is put on the system.

Denote by $W=f(\mathbf{X}^n)$ the message sent from node A to node B. The rate can be bounded as follows:

$$\begin{aligned}
nR &\ge I(W;\mathbf{X}^n) && \text{(48a)}\\
&= I(W;\mathbf{X}^n,\mathbf{Y}^n)=I(W;\mathbf{Y}^n)+I(W;\mathbf{X}^n|\mathbf{Y}^n) && \text{(48b)}\\
&= \sum_{i=1}^n I(W,\mathbf{Y}^{i-1};Y_i)+\sum_{i=1}^n I(W;X_i|\mathbf{Y}^n,\mathbf{X}^{i-1}) && \text{(48c)}\\
&= \sum_{i=1}^n I(W,\mathbf{Y}^{i-1};Y_i) \\
&\quad+\sum_{i=1}^n I(W;X_i|Y_i,\mathbf{Y}_{i+1}^n,\mathbf{Y}^{i-1},\mathbf{X}^{i-1}) && \text{(48d)}\\
&= \sum_{i=1}^n\big[I(W,\mathbf{Y}^{i-1};Y_i) \\
&\quad+I(W,\mathbf{Y}_{i+1}^n,\mathbf{Y}^{i-1},\mathbf{X}^{i-1};X_i|Y_i)\big] && \text{(48e)}\\
&= \sum_{i=1}^n\big[I(W,\mathbf{Y}^{i-1};Y_i)+I(W,\mathbf{Y}^{i-1};X_i|Y_i) \\
&\quad+I(\mathbf{Y}_{i+1}^n,\mathbf{X}^{i-1};X_i|Y_i,\mathbf{Y}^{i-1},W)\big] && \text{(48f)}\\
&= \sum_{i=1}^n\big[I(W,\mathbf{Y}^{i-1};Y_i,X_i) \\
&\quad+I(\mathbf{Y}_{i+1}^n,\mathbf{X}^{i-1};X_i|Y_i,\mathbf{Y}^{i-1},W)\big] && \text{(48g)}\\
&= \sum_{i=1}^n\big[I(W,\mathbf{Y}^{i-1};X_i) \\
&\quad+I(\mathbf{Y}_{i+1}^n,\mathbf{X}^{i-1};X_i|Y_i,\mathbf{Y}^{i-1},W)\big]\ . && \text{(48h)}
\end{aligned}$$

Here, (48b) and (48h)are due to the Markov chains $W-\mathbf{X}^n-\mathbf{Y}^n$ and $W-X_i-Y_i$, respectively. (48e) stems from the fact that both sources X and Y are assumed to be jointly i.i.d. Defining $U_i\triangleq(W,\mathbf{Y}^{i-1})$ and $V_i\triangleq(U_i,\mathbf{Y}_{i+1}^n,\mathbf{X}^{i-1})$ the Markov chain $U_i-V_i-X_i-Y_i$ is satisfied since the sources X and Y are assumed to be jointly i.i.d, and the bound over the rate becomes

$$\begin{aligned}
R &\ge \frac{1}{n}\sum_{i=1}^n\left[I(U_i;X_i)+I(V_i;X_i|U_i,Y_i)\right] \\
&= I(U;X)+I(V;X|UY)\ ,
\end{aligned} \tag{49}$$

with U and V defined through time-sharing as is subsequently shown in (52).

The error exponent can now be expressed as follows:

$$\begin{aligned} I(W;\mathbf{Y}^n) &= \sum_{i=1}^{n} I(W,\mathbf{Y}^{i-1};Y_i) \\ &= \sum_{i=1}^{n} I(U_i;Y_i) = nI(U;Y) \ , \end{aligned} \tag{50}$$

with the same definition of U_i. Thus, the converse over the error exponent is proved with equality.

Finally, the distortion at node B can be bounded as follows. Define the function $\hat{X}_i$ as the i-th coordinate of the estimate in node B:

$$\hat{X}_i(U_i,V_i,Y_i) \triangleq g_i(W,\mathbf{Y}^{i-1},Y_i,\mathbf{Y}_{i+1}^n) \ . \tag{51}$$

The component-wise mean distortion thus verifies

$$\begin{aligned} D+\epsilon &\geq \mathbb{E}_0\left[d\big(\mathbf{X}^n,g(W,\mathbf{Y}^n)\big)\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_0\left[d\big(X_Q,\hat{X}_Q(U_Q,V_Q,Y_Q)\big)|Q=i\right] \\ &= \mathbb{E}_0\left[d\big(X_Q,\hat{X}_Q(U_Q,V_Q,Y_Q)\big)\right] \\ &= \mathbb{E}_0\left[d\big(X,\hat{X}(U,V,Y)\big)\right] \ . \end{aligned} \tag{52}$$

For the sake of this calculation, we use the fact that any U_i and V_i, as they were defined for this converse, contain the entire message W, as well as the past and future of Y. This concludes the converse proof in Proposition 1.

Cardinality bounds

It remains to establish that the cardinality bounds specified by the conditions in Proposition 1 do not affect the minimization. Toward that end we invoke the support lemma [29, p. 310] in order to deduce that $\mathcal{U}$ must have $\|\mathcal{X}\|-1$ letters in order to ensure preservation of $p(x|u)$ plus three more to preserve the constraints on D, $I(U;X)$ and $I(U;Y)$, so $\|\mathcal{U}\| \leq \|\mathcal{X}\|+2$ suffices. Similarly, $\mathcal{V}$ must have $\|\mathcal{X}\|\|\mathcal{U}\|-1$ letters in order to ensure preservation of $p(x,u|v)$ plus two more to preserve D, and $I(X;V|UY)$. Thus, it suffices to have $\|\mathcal{V}\| \leq \|\mathcal{X}\|\|\mathcal{U}\|+1$.

Appendix C
Proof of Proposition 2

Achievability proof

In order to achieve the region proposed in Theorem 2, choose V as the output of a Binary Symmetric Channel (BSC) with cross-over probability α when the input is X. Choose U as the output of another BSC, with cross-over probability β, when the input is V:

$$\begin{aligned} V &= X+W_1, \quad W_1 \sim \text{Bern}\left(\alpha\right) \ , \\ U &= V+W_2, \quad W_2 \sim \text{Bern}\left(\beta\right) \ . \end{aligned} \tag{53}$$

Calculating the expression for the error exponent, U and Y can be thought of as connected through a BSC with cross-over probability $\alpha \star \beta \star p$, which yields:

$$I(U;Y) = H(U) - H(U|Y) = 1 - H_2(\alpha \star \beta \star p) \ . \tag{54}$$

This complies with the expression proposed in Theorem 2. The relation between the second term in the expression for the rate and the amount of distortion expected can be calculated through the following two steps, inspired by the approach taken in [24], for the case of source estimation with side information, jointly distributed according to a BSC (without uncertainty in the probability distribution of the sources):

a) Setting $\hat{X} = g(Y,V) = V$, we have $\mathbb{E}_0\left[d(X,\hat{X})\right] = \alpha$. Note that all expectations henceforth are taken over the distribution imposed by H_0, and under the assumption that the decision H_0 was correct. Y and V can be thought of as being connected through a BSC with cross-over probability $\alpha \star p$. Thus (8) results in

$$\begin{aligned} R_a &= I(U;Y) + [I(V;X) - I(V;Y)] \\ &= 1 - H_2(\alpha \star \beta \star p) + [H_2(\alpha \star p) - H_2(\alpha)] \ . \end{aligned} \tag{55}$$

b) In this part, we let V be degenerate and $\hat{X} = g(Y,V) = Y$. We then have $\mathbb{E}_0\left[d\left(X,\hat{X}\right)\right] = p$. Since in this case $I(V;X) - I(V;Y) = 0$, we have

$$R_b = I(U;Y) = 1 - H_2(\alpha \star \beta \star p) \ . \tag{56}$$

Now let $0 \leq D \leq p$ be given and say that θ, α are such that $D = \theta\alpha + (1-\theta)p$. Since $R(D)$ is convex (for a given error exponent E),

$$\begin{aligned} R(E,D) &= R(\theta\alpha + (1-\theta)p) \\ &\leq \theta R(\alpha) + (1-\theta)R(p) \\ &= \theta R_a + (1-\theta)R_b \\ &\leq 1 - H_2(\alpha \star \beta \star p) + \theta\left[H_2(\alpha \star p) - H_2(\alpha)\right] \ . \end{aligned} \tag{57}$$

Thus, any triplet (R,E,D) that complies with Theorem 2 is achievable through this scheme, and the proof of achievability is complete.

Converse proof

Theorem 1, along with the development in (8), implies that the optimal region, for any specific example of hypothesis testing against independence, is comprised of two RVs, such that the Markov chain $U \multimap V \multimap X \multimap Y$ is respected. Moreover, it implies that with these optimal auxiliary RVs, the required rate is comprised of two independent parts – one part dedicated to detection and the other to estimation. Thus, the proof of the converse to Theorem 2 can be divided, much like the proof of achievability, into two separate parts - one defining the trade-off between the rate and the error exponent, while the other defines the trade-off between the rate and the distortion.

Starting with the relation between the rate and the error exponent, Theorem 1 implies that

$$E \leq I(U;Y) = H(Y) - H(Y|U) = 1 - A \ , \tag{58}$$

while

$$R \geq 1 - A + \theta\left[I(V;X) - I(V;Y)\right] \ , \tag{59}$$

with A defined as $A \triangleq H(Y|U)$. Ignoring the second term in the expression for the rate, the trade-off between rate and

error exponent is clear, and is given through A. Obviously, $A \leq H(Y) = 1$. In addition,

$$A \geq H_2\left(H_2^{-1}\left(H(X|U)\right) \star p\right) \ , \tag{60}$$

which stems from Ms. Gerber's Lemma (see e.g. [28]). In order to allow the exploration of the entire region defined by the bounds over A, we define $\gamma \triangleq H_2^{-1}\left(H(X|U)\right)$. Thus, the trade-off between rate and error exponent becomes

$$\begin{aligned} &E \leq 1 - H_2(\gamma \star p) \ , \\ &R \geq 1 - H_2(\gamma \star p) + \theta\left[I(V;X) - I(V;Y)\right] \ . \end{aligned} \tag{61}$$

In the second part of the proof, it needs to be demonstrated that, once the decision H_0 has been (correctly) made, the optimal estimation region, defined by the rate-distortion relation $\min_{\mathbb{E}\left[d(X,\hat{X})\right] \leq D}\left[I(V;X) - I(Y;X)\right]$, is in agreement with Theorem 2. This proof has already been given in [24] and is thus omitted from this work. Defining V as the output of a BSC with cross-over probability α when X is in the input of the channel, as was shown to be optimal in [24], and keeping in mind the Markov chain implied by Theorem 1, it is clear that $\gamma = H^{-1}\left(H(X|U)\right) \geq \alpha$. Thus, γ can be expressed as $\gamma = \alpha \star \beta$ for some $0 \leq \beta \leq \frac{1}{2}$, which completes the proof.

Appendix D
Proof of Proposition 3

We now prove the achievability of the region offered in Proposition 3 for the joint detection and lossy compression problem, with general hypotheses. We start by describing the codebook, as well as encoding and decoding strategies, and followed by an analysis of error events under the proposed strategy.

Encoding and decoding strategy

Codebook Construction: For a given block-length n we operate on a type-by-type basis. For each type $Q_X \in \mathcal{P}_n(\mathcal{X})$, fix a conditional type $Q^{\star}_{U|X}(Q_X) \in \mathcal{P}_n(\mathcal{U})$. Randomly and uniformly choose a set of codewords denoted by $\mathcal{C}_U^n(Q_X)$, from the resulting marginal type class $\mathcal{T}^n_{Q^{\star}_U}(Q_X)$ which is induced by Q_X and $Q^{\star}_{U|X}(Q_X)$. The size of $\mathcal{C}_U^n(Q_X)$ is an integer satisfying:

$$\begin{aligned} \exp\left[nI\big(Q_X; Q^{\star}_{U|X}(Q_X)\big)\right] &+ (|\mathcal{U}||\mathcal{X}| + 2)\log(n+1) \\ &\leq |\mathcal{C}_U^n(Q_X)| \leq \\ \exp\left[nI\big(Q_X; Q^{\star}_{U|X}(X)\big)\right] &+ (|\mathcal{U}||\mathcal{X}| + 4)\log(n+1) \ , \end{aligned} \tag{62}$$

where $\mathcal{C}_U^n(Q_X)$ is the codebook of the common message for source type Q_X. Define $f_U : \mathcal{T}^n_{Q_X} \to \mathcal{C}_U^n(Q_X)$, i.e., a function $f_U(\mathbf{x}^n)$ that determines the codeword sent by the encoder (node A) to the decoder (node B), as subsequently explained. We define $\mathbf{U}^n \triangleq f_U(\mathbf{X}^n)$. In addition, assign an index: $k(Q_X) : \mathcal{P}_n(\mathcal{X}) \to \{1, \ldots, (n+1)^{|\mathcal{X}|}\}$ to each of the possible types of vectors $\mathbf{x}^n \in \mathcal{X}^n$.

In addition, let V_0 and V_1 be two RVs, designed to transmit a private message to the decoder. After making a decision about the common distribution controlling X and Y, the decoder would use the appropriate private message in order to reconstruct the original sequence $\mathbf{x}$ (with distortion). As was the case when testing against independence as seen in Appendix B, the common distribution $Q_{UV|X} = Q_{U|X}Q_{V|UX}$ is chosen such that the Markov chains $U - V_0 - X - Y$ and $\bar{U} - V_1 - \bar{X} - \bar{Y}$ are respected.

For each codeword $\mathbf{u}^n \in \mathcal{C}_U^n$, randomly generate $\exp\left[nS_0\right]$ sequences $\mathbf{v}_0^n(s_0)$, indexed with $s_0 = [1 : \exp\left(nS_0\right)]$, and $\exp\left[nS_1\right]$ sequences $\mathbf{v}_1^n(s_1)$, indexed with $s_1 = [1 : \exp\left(nS_1\right)]$, from the conditional typical sets $\mathcal{T}^n_{[V_0|U]\delta}(\mathbf{u}^n)$ and $\mathcal{T}^n_{[V_1|\bar{U}]\delta}(\mathbf{u}^n)$, respectively. Divide them into $\exp\left(nR_0\right)$ (respectively $\exp\left(nR_1\right)$) bins, such that each bin contains roughly $\exp\left[n(S_0 - R_0)\right]$ (respectively $\exp\left[n(S_1 - R_1)\right]$) sequences. In the remainder of this proof we only treat source reconstruction in case hypothesis H_0 was chosen, as the complementary case is completely symmetric.

Encoding: Given a sequence $\mathbf{x}^n \in \mathcal{T}^n_{Q_X}$, search for a sequence $\mathbf{u}^n \in \mathcal{C}_U^n(Q_{\mathbf{x}^n})$, i.e., in the codebook that belongs to the type $Q_{\mathbf{x}^n}$, such that $(\mathbf{u}^n, \mathbf{x}^n) \in \mathcal{T}^n_{[UX]\delta}$. As a second step, look for a codeword $\mathbf{v}_0^n(s_0)$ such that $(\mathbf{v}_0^n(s_0), \mathbf{x}^n) \in \mathcal{T}^n_{[V_0X|U]\delta}(\mathbf{u}^n)$ with the typicality measured according to the distribution induced by hypothesis H_0. Let $B_0(\mathbf{v}_0^n(\mathbf{x}^n, \mathbf{u}^n))$ denote the element (or "bin") to which $\mathbf{v}_0^n$ is mapped. Perform the same steps for the case where H_1 is the chosen hypothesis.

The encoder's message then consists of four parts:

$$\begin{aligned} \mathcal{M}_1 &= \{1, 2, \ldots, M_1 \triangleq \exp\left(nR'\right)\} \ , \\ \mathcal{M}_2 &= \left\{1, 2, \ldots, M_2 \triangleq (n+1)^{|\mathcal{X}|}\right\} \ , \\ \mathcal{M}_3 &= \{1, 2, \ldots, M_3 \triangleq \exp\left(nR_0\right)\} \ , \\ \mathcal{M}_4 &= \{1, 2, \ldots, M_3 \triangleq \exp\left(nR_1\right)\} \ , \\ \mathcal{M} &= \mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{M}_3 \times \mathcal{M}_4 \ . \end{aligned} \tag{63}$$

The encoder sends the type of $\mathbf{x}^n$ which requires $|\mathcal{M}_2|$ values but with zero rate, and also $F(f_U(\mathbf{x}^n))$, as well as the respective bins for both private messages, $B_0(\mathbf{v}_0^n(\mathbf{x}^n, \mathbf{u}^n))$ and $B_1(\mathbf{v}_1^n(\mathbf{x}^n, \mathbf{u}^n))$, to be defined subsequently. There are two cases to consider:

1 $\log|\mathcal{C}_U^n(Q_{\mathbf{x}^n})| < nR'$, in which case we can map each member of $C_U^n(Q_{\mathbf{x}^n})$ to an element of $\mathcal{M}_1$ in a one-to-one manner.
2 $\log|\mathcal{C}_U^n(Q_{\mathbf{x}^n})| \geq nR'$, in which case we assign each distinct member of $C_U^n(Q_{\mathbf{x}^n})$ to $\mathcal{M}_1$ uniformly at random.

Let $F(f_U(\mathbf{x}^n))$ denote the element to which $f_U(\mathbf{x}^n)$ is mapped. The encoder can be expressed mathematically as

$$\begin{aligned} \Psi(x) = \big(&F(f_U(\mathbf{x}^n)), k(Q_{\mathbf{x}^n}) \\ &, B_0(\mathbf{v}_0^n(\mathbf{x}^n, \mathbf{u}^n)), B_1(\mathbf{v}_1^n(\mathbf{x}^n, \mathbf{u}^n))\big) \ , \end{aligned} \tag{64}$$

for each $\mathbf{x}^n \in \mathcal{T}^n_{Q_{\mathbf{x}^n}}$.

Decoding: The decoder first attempts to discover the word $\mathbf{u}^n$, by using the information sent from the encoder and the observation vector $\mathbf{y}^n$:

- If $\log|\mathcal{C}_U^n(Q_{\mathbf{x}})| < nR'$ the codeword can be decoded without error;
- Otherwise $\log|\mathcal{C}_U^n(Q_{\mathbf{x}})| \geq nR'$ the decoder receives a bin index and uses side information $\mathbf{y}^n$ to pick the best $\mathbf{u}^n$ in the bin. Given the bin number, the type $Q_{\mathbf{x}^n}$ and the side

information $\mathbf{y}^n$, the decoder uses a minimal empirical entropy decoding[2], that is:

$$\phi(F(f_U(\mathbf{x}^n)), Q_{\mathbf{x}^n}, \mathbf{y}^n) = \hat{\mathbf{u}}^n \ , \tag{65}$$

if $H(\tilde{\mathbf{u}}^n|\mathbf{y}^n) > H(\hat{\mathbf{u}}^n|\mathbf{y}^n)$ for $\hat{\mathbf{u}}^n \in F(f_U(\mathbf{x}^n))$ and all $\tilde{\mathbf{u}}^n \in F(f_U(\mathbf{x}^n))$ with $\tilde{\mathbf{u}}^n \neq \hat{\mathbf{u}}^n$, where

$$H(\hat{\mathbf{u}}^n|\mathbf{y}^n) \triangleq -\sum Q_{\hat{\mathbf{u}}^n\mathbf{y}^n}(a,b)\log Q_{\hat{\mathbf{u}}^n|\mathbf{y}^n}(a|b)$$

is the empirical entropy of the vector $\hat{\mathbf{u}}^n$ given the vector $\mathbf{y}^n$, and the sum is taken over all the letters in the alphabets of U and Y.

As a second step, the decoder uses the private message – either $\mathbf{v}_0^n$ or $\mathbf{v}_1^n$– destined for the case of the current hypothesis in order to estimate $\mathbf{x}^n$, with distortion D_0 or D_1, respectively. Assume hypothesis H_0 is in effect, it searches for a single sequence $\hat{\mathbf{v}}_0^n \in B_0(\mathbf{v}_0^n(\mathbf{x}^n, \mathbf{u}^n))$ such that $\hat{\mathbf{v}}_0^n(s_0) \in \mathcal{T}_{[V_0|UY]\delta}(\mathbf{u}^n\mathbf{y}^n)$. If it finds no such sequence it declares an error during the reconstruction. If it finds more than one, it chooses one sequence at random.

Error probability of the testing step

We now show that, for the detection part, the exponential rate of decay of the error of the second type, under a fixed constraint over the error of the first type, is not smaller than the value claimed by Proposition 3. The analysis of possible errors at the encoder's side stays identical to the one done in the proof of Theorem 1 in Appendix B (note that we assume the $P_X(x) = P_{\bar{X}}(x)$, without which the analysis of the encoder's side, with an emphasis on the codebook construction, might become more involved). Note also that when a problem does arise during encoding, our proposed scheme calls for an error message which prompts node B to declare H_1. Thus, the influence of such errors is only on the error probability of Type I, and not on the error exponent of Type II. We concentrate in this analysis on possible errors at the decoder's side. Define two error events: First, let

$$\mathcal{B}_6 \triangleq \{\mathbf{u}^n \neq F(f_U(\mathbf{x}^n))\} \tag{66}$$

be the event that the chosen sequence from the bin at the decoder is different from the original sequence sent by the encoder. Then, define $\mathcal{B}_7$ to be the event of erroneous detection despite using the correct sequence. We denote the probabilities of events $\mathcal{B}_6$ and $\mathcal{B}_7$ by $P_r^{(n)}$ and $P_d^{(n)}$, respectively. Using the union bound, the probability of error in detection can be bounded by

$$P_e^{(n)} \leq P_r^{(n)} + P_d^{(n)} \ . \tag{67}$$

Evaluation of $P_r^{(n)}$: We evaluate the probability that node B chooses the wrong sequence from the bin under the suggested encoding and decoding schemes. Our evaluation is reliant on the method of types [26], and is specifically inspired by the techniques used in [25, Appendix C]. We first evaluate $P_r^{(n)}$ for a finite block-length n and then use a continuity argument to show that in the limit of $n \to \infty$,

$$-\frac{1}{n}\log P_r^{(n)} \leq G(Q_{UXY}, Q_X, Q_Y, R') \quad , \tag{68}$$

where the function G is the one given in (17).

Since choosing the wrong sequence can only happen in case binning is used, we are only interested in the following subset of the set of all possible sequences:

$$\begin{aligned}\mathcal{A}_n = \Big\{&(\mathbf{u}^n, \mathbf{x}^n, \mathbf{y}^n) \in \mathcal{U}^n \times \mathcal{X}^n \times \mathcal{Y}^n \\ &\big| \ \mathbf{u}^n \in T^n_{Q^\star_{U|X}}(Q_{\mathbf{x}^n}) \, , \, \log|\mathcal{C}_U^n(Q_{\mathbf{x}^n})| \geq nR\Big\} \ .\end{aligned} \tag{69}$$

We first evaluate the probability of choosing the wrong sequence within the set $\mathcal{A}_n$ by using the following lemma.

Lemma 8. *Let* $(\mathbf{u}^n, \mathbf{x}^n, \mathbf{y}^n) \in \mathcal{A}_n$ *and let* $\mathcal{B}_8$ *be the event that* $\mathbf{u}^n \neq \phi(\psi(\mathbf{x}^n), \mathbf{y}^n)$. *Provided that* $\log|\mathcal{C}_U^n(Q_{\mathbf{x}^n})| \geq nR$, *then*

$$\begin{aligned}&\Pr\left(\mathcal{B}_8|\mathbf{U}^n = \mathbf{u}^n, \mathbf{X}^n = \mathbf{x}^n, \mathbf{Y}^n = \mathbf{y}^n\right) \\ &\quad \leq \exp\left[-n\left(R - J(Q_{\mathbf{u}^n\mathbf{x}^n\mathbf{y}^n}) - \delta_n\right)\right] \ ,\end{aligned} \tag{70}$$

with

$$\begin{aligned}&J\big(Q_{\mathbf{u}^n\mathbf{x}^n\mathbf{y}^n}\big) \\ &\triangleq I\big(Q_{\mathbf{x}^n}; Q^\star_{U|X}(Q_{\mathbf{x}^n})\big) - I\big(Q_{\mathbf{u}^n|\mathbf{y}^n}; Q_{\mathbf{y}^n}\big)\end{aligned} \tag{71}$$

and

$$\delta_n \triangleq \frac{1}{n}\log(n+1)^{|\mathcal{U}|(1+|\mathcal{X}|+|\mathcal{Y}|)+4} \ . \tag{72}$$

The probability in (70) *is taken over the choice of the codebook in use.*

Before proving Lemma 8, we recall the following result from [25, Lemma 12].

Lemma 9. *For all strings* $(\mathbf{u}, \mathbf{x})$ *such that* $\mathbf{u} \in T^n_{Q^\star_U}$,

$$\begin{aligned}&\Pr(\mathbf{u} \in \mathcal{C}_U^n(Q_{\mathbf{x}^n})) \\ &\quad \leq (n+1)^{\|\mathcal{U}\|(1+\|\mathcal{X}\|)+4} \\ &\quad \times \exp\left[n\left(I(Q_{\mathbf{x}^n}; Q^\star_{U|X}(Q_{\mathbf{x}^n})) - H(Q_{\mathbf{u}^n})\right)\right] .\end{aligned} \tag{73}$$

Proof (Lemma 8): Let $\mathcal{S}(\mathbf{u}^n|\mathbf{y}^n)$ be the set that includes all sequences $\tilde{\mathbf{u}}^n$, such that $\tilde{\mathbf{u}}^n$ has the same type as $\mathbf{u}$ and $H(\tilde{\mathbf{u}}^n|\mathbf{y}^n) \leq H(\mathbf{u}^n|\mathbf{y}^n)$. Then

$$\begin{aligned}&\Pr\left(\mathcal{B}_8|\mathbf{U}^n = \mathbf{u}^n, \mathbf{X}^n = \mathbf{x}^n, \mathbf{Y}^n = \mathbf{y}^n\right) \\ &\leq \sum \Pr\big(\tilde{\mathbf{u}}^n \in \mathcal{C}_U^n(Q_{\mathbf{x}^n}), \{F(\tilde{\mathbf{u}}^n) = F(\mathbf{u}^n)\}| \\ &\qquad \mathbf{U}^n = \mathbf{u}^n, \mathbf{X}^n = \mathbf{x}, \mathbf{Y}^n = \mathbf{y}\big) \qquad \text{(74a)} \\ &\leq \sum \Pr\big(\tilde{\mathbf{u}}^n \in \mathcal{C}_U^n(Q_{\mathbf{x}^n})|\mathbf{X}^n = \mathbf{x}^n, \mathbf{Y}^n = \mathbf{y}^n\big) \\ &\qquad \times \Pr\big(\{F(\tilde{\mathbf{u}}^n) = F(\mathbf{u}^n)\}\big) \qquad \text{(74b)} \\ &\leq \sum (n+1)^{|\mathcal{U}|(1+|\mathcal{X}|)+4} \\ &\times \exp\left[n\left(I(Q_{\mathbf{x}^n}; Q^\star_{U|X}(Q_{\mathbf{x}^n})) - H(Q_{\mathbf{u}^n})\right)\right]\frac{1}{M_1} \qquad \text{(74c)}\end{aligned}$$

[2]Note that since our chosen test is over empirical entropies, it does not matter at this stage which hypothesis is the true one, for the sake of choosing the sequence from the bin. After having retrieved a single sequence from the bin, the decoder can continue to perform HT by discarding the rest of the sequences in the bin and only using the chosen sequence.

$$
\begin{aligned}
&\leq (n+1)^{|\mathcal{U}||\mathcal{Y}|} \exp\left[nH(Q_{\mathbf{u}^n|\mathbf{y}^n}|Q_{\mathbf{y}^n})\right] \frac{1}{M_1} \\
&\times (n+1)^{|\mathcal{U}|(1+|\mathcal{X}|)+4} && \text{(75a)} \\
&\times \exp\left[n\left(I(Q_{\mathbf{x}^n}; Q^{\star}_{U|X}(Q_{\mathbf{x}^n})) - H(Q_{\mathbf{u}^n})\right)\right] \\
&= (n+1)^{|\mathcal{U}|(1+|\mathcal{X}|+|\mathcal{Y}|)+4} \\
&\times \exp\Big[-n\Big(R - H(Q_{\mathbf{u}^n|\mathbf{y}^n}|Q_{\mathbf{y}^n}) + H(Q_{\mathbf{u}^n}) \\
&\qquad -I(Q_{\mathbf{x}^n}; Q^{\star}_{U|X}(Q_{\mathbf{x}^n}))\Big)\Big] && \text{(75b)} \\
&= (n+1)^{|\mathcal{U}|(1+|\mathcal{X}|+|\mathcal{Y}|)+4} \\
&\times \exp\Big[-n\Big(R + I(Q_{\mathbf{u}^n|\mathbf{y}^n}; Q_{\mathbf{y}^n})) \\
&\qquad -I(Q_{\mathbf{x}^n}; Q^{\star}_{U|X}(Q_{\mathbf{x}^n}))\Big)\Big] && \text{(75c)} \\
&\triangleq (n+1)^{|\mathcal{U}|(1+|\mathcal{X}|+|\mathcal{Y}|)+4} \\
&\times \exp\left[-n\left(R - J(Q_{\mathbf{u}^n\mathbf{x}^n\mathbf{y}^n})\right)\right] && \text{(75d)} \\
&\leq \exp\left[-n\left(R - J(Q_{\mathbf{u}^n\mathbf{x}^n\mathbf{y}^n}) - \delta_n\right)\right] , && \text{(75e)}
\end{aligned}
$$

with δ_n as defined above, and the sums are all taken over the set $\tilde{\mathbf{u}}^n \in \mathcal{S}(\mathbf{u}^n|\mathbf{y}^n)$, $\tilde{\mathbf{u}}^n \neq \mathbf{u}^n$. Here, the probability $\Pr\left(\tilde{\mathbf{u}}^n \in \mathcal{C}^n_U(Q_{x^n})\right)$ is over the choice of the codebook. Inequality (74b) stems from the codebook construction, which divides sequences into bins randomly and independently. Inequality (74c) is due to Lemma 9, which applies here with slight notation changes (see at the end of this proof), and to the upper bound over the size of $\mathcal{C}^n_U(Q_{\mathbf{x}^n})$, given in (62). Inequality (75a) is due to Lemma 5. Finally, equality (75b) is due to the definition of M_1 and (75e) stems from the fact that $\Pr\left(\mathcal{B}_8|\mathbf{U}^n = \mathbf{u}^n, \mathbf{X}^n = \mathbf{x}^n, \mathbf{Y}^n = \mathbf{y}^n\right) \leq 1$ and the definition of δ_n.

■

We now bound the probability of error in choosing the right sequence in the bin $P_r^{(n)}$, for a finite block-length n:

$$
\begin{aligned}
P_r^{(n)} &= \Pr\left(\{\mathbf{u}^n \neq F(f_U(\mathbf{x}^n))\}\right) && \text{(76a)} \\
&\leq \sum \Pr\left(\mathcal{B}_8|\mathbf{U}^n = \mathbf{u}, \mathbf{X}^n = \mathbf{x}, \mathbf{Y}^n = \mathbf{y}\right) \\
&\quad \times \Pr\left(\mathbf{U} = \mathbf{u}, \mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}\right) && \text{(76b)} \\
&\leq \sum \exp\left[-n\left(R - J(Q_{\mathbf{u}^n\mathbf{x}^n\mathbf{y}^n}) - \delta_n\right)\right] \\
&\quad \times P^n_{XY}(\mathbf{x}^n, \mathbf{y}^n) \frac{1}{|\mathcal{T}^n_{Q^{\star}_{U|X}}(Q_{\mathbf{x}^n})|} . && \text{(76c)}
\end{aligned}
$$

Here, claim (76c) is derived from Lemma 8. Note the slight abuse of notation here, where $P^n_{XY}(\mathbf{x}^n, \mathbf{y}^n)$ in (76c) refers to the *real distribution* controlling the RVs, and can thus actually be, according to the true hypothesis, wither $P^n_{XY}(\mathbf{x}^n, \mathbf{y}^n)$ or $P^n_{\bar{X}\bar{Y}}(\mathbf{x}^n, \mathbf{y}^n)$. The probability of choosing a specific sequence $\mathbf{u}^n$ given both source sequences $\mathbf{x}^n$ and $\mathbf{y}^n$ stems from averaging over the code. We can now change the expression to sum first on types and then on sequences within each type class. In order to transform our summation over a set of sequences $\mathcal{A}_n$ into a summation over a set of types (and only then over the sequences within each type) we define the following set of types:

$$
\begin{aligned}
&\mathcal{D}(Q_X, Q_Y) \\
&= \big\{Q_{UXY} \in \mathcal{P}_n(\mathcal{U} \times \mathcal{X} \times \mathcal{Y}) : Q_{U|X} = Q^{\star}_{U|X}(Q_X), \\
&\qquad \log|\mathcal{C}^n_U(Q_X)| \geq nR\big\} .
\end{aligned} \tag{77}
$$

The probability of error in selecting the sequence can thus be bound by (78), at the top of the next page.

In the case of distributed HT, the probability of the source sequences $(\mathbf{x}^n, \mathbf{y}^n)$ is unknown, since the sequences can be created by one of two possible distributions. We thus bound the probability of the observed sources by

$$
\begin{aligned}
P^n_{XY}(\mathbf{x}^n, \mathbf{y}^n) &\leq \max\{P_{XY}(\mathbf{x}^n, \mathbf{y}^n), P_{\bar{X}\bar{Y}}(\mathbf{x}^n, \mathbf{y}^n)\} \\
&= \max_{i=\{0,1\}} \left\{\exp\left[-n\left(\mathcal{D}(Q_{XY}\|P_{XY_i}) + H(Q_{XY})\right)\right]\right\} \\
&= \exp\left[-n\left(\min_{i=\{0,1\}} \mathcal{D}(Q_{XY}\|P_{XY_i}) + H(Q_{XY})\right)\right] ,
\end{aligned} \tag{79}
$$

where, in accordance to the notation of Proposition 3, we use the subscript i in order to differentiate between P_{XY} and $P_{\bar{X}\bar{Y}}$. Using the following facts detailed in Lemma 3,

$$
\begin{aligned}
|\mathcal{T}^n_{Q_{UXY}}| &\leq \exp\left[n(H(Q_{UXY}))\right] \\
&\leq \exp\left(n\log|\mathcal{U}||\mathcal{X}||\mathcal{Y}|\right) , && \text{(80a)} \\
|\mathcal{T}^n_{Q_{U|X}}| &\geq (n+1)^{-|\mathcal{U}||\mathcal{X}|} \exp\left[n\left(H(Q_{U|X}|Q_X)\right)\right] , && \text{(80b)}
\end{aligned}
$$

we obtain from (78) that

$$
\begin{aligned}
&\leq \sum_{Q_X \in \mathcal{P}_n(\mathcal{X})} \sum_{Q_Y \in \mathcal{P}_n(\mathcal{Y})} \sum_{Q_{UXY} \in \mathcal{D}(Q_X, Q_Y)} \\
&\qquad \exp\left[-n\left(\Gamma + R - J(Q_{UXY}) - \delta_n\right)\right] ,
\end{aligned} \tag{81}
$$

with Γ satisfying:

$$
\begin{aligned}
\Gamma &= \min_{i=\{0,1\}} \mathcal{D}(Q_{XY}\|P_{XY_i}) + H(Q_{XY}) \\
&\quad + H(Q_{U|X}|Q_X) - H(Q_{UXY}) \\
&= \min_{i=\{0,1\}} \mathcal{D}(Q_{XY}\|P_{XY_i}) + H(Q_{U|X}|Q_X) \\
&\quad - H(Q_{U|XY}|Q_{XY}) \\
&= \min_{i=\{0,1\}} \sum_{\substack{x\in\mathcal{X} \\ y\in\mathcal{Y}}} Q_{XY}(x,y) \log \frac{Q_{XY}(x,y)}{P_{XY_i}(x,y)} \\
&\quad - \sum_{\substack{u\in\mathcal{U} \\ x\in\mathcal{X}}} Q_{UX}(u,x) \log \frac{Q_{UX}(u,x)}{Q_X(x)} \\
&\quad + \sum_{\substack{u\in\mathcal{U} \\ x\in\mathcal{X} \\ y\in\mathcal{Y}}} Q_{UXY}(u,x,y) \log \frac{Q_{UXY}(u,x,y)}{Q_{XY}(x,y)} \\
&= \min_{i=\{0,1\}} \mathcal{D}(Q_{UXY}\|P_{XY_i}Q_{U|X}) .
\end{aligned} \tag{82}
$$

The probability of error in bin decoding can thus be concluded to satisfy

$$
\begin{aligned}
P_r^{(n)} \leq &\sum_{Q_X \in \mathcal{P}_n(\mathcal{X})} \sum_{Q_Y \in \mathcal{P}_n(\mathcal{Y})} \sum_{Q_{UXY} \in \mathcal{D}(Q_X, Q_Y)} \\
&\exp\left[-n\left(\min_{i=\{0,1\}} \mathcal{D}(Q_{UXY}\|P_{XY_i}Q_{U|X}) \right.\right. \\
&\left.\left. +R - J(Q_{UXY}) - \delta_n\right)\right] .
\end{aligned} \tag{83}
$$

$$P_r^{(n)} \leq \sum_{Q_X,Q_Y} \left[\sum_{Q_{UXY}\in\mathcal{D}(Q_X,Q_Y)} \sum_{(\mathbf{u}^n,\mathbf{x}^n,\mathbf{y}^n)\in\mathcal{T}^n_{Q_{UXY}}} \frac{P^n_{XY}(\mathbf{x}^n,\mathbf{y}^n)}{|\mathcal{T}^n_{Q^\star_{U|X}}(Q_{\mathbf{x}^n})|} \exp\left[-n\big(R - J(Q_{\mathbf{u}^n\mathbf{x}^n\mathbf{y}^n}) - \delta_n\big)\right] \right] . \tag{78}$$

We may now upper bound the summations by maximizing over the types and optimizing over the choice of the of the test channel $Q^\star_{U|X}$. We optimize to then obtain:

$$P_r^{(n)} \leq |\mathcal{P}_n(\mathcal{X})| \max_{Q_X} \min_{Q^\star_{U|X}} |\mathcal{P}_n(\mathcal{Y})| \max_{Q_Y} |\mathcal{P}_n(\mathcal{U}\times\mathcal{X}\times\mathcal{Y})| \max_{\substack{Q_{UXY} \\ Q_{U|X}=Q^\star_{U|X}}} \exp\left\{-nG_n\left[Q_{UXY},Q_X,Q_Y,R\right]\right\} . \tag{84}$$

Thus,

$$\frac{1}{n}\log P_r^{(n)} \leq - \min_{Q_X\in\mathcal{P}_n(\mathcal{X})} \max_{Q^\star_{U|X}(Q_X)} \min_{Q_Y\in\mathcal{P}_n(\mathcal{Y})} \min_{\substack{Q_{UXY} \\ Q_{U|X}=Q^\star_{U|X}}} G_n\left[Q_{UXY},Q_X,Q_Y,R\right] \times \log\left(|\mathcal{P}_n(\mathcal{X})||\mathcal{P}_n(\mathcal{Y})||\mathcal{P}_n(\mathcal{U}\times\mathcal{X}\times\mathcal{Y})|\right)$$

with the function $G_n\left[Q_{UXY},Q_X,Q_Y,R\right]$ defined in (85) at the top of the next page. The cardinalities can be absorbed inside the exponent and become insignificant as $n\to\infty$. From continuity arguments under discrete alphabets, it is made clear that [25, Lemma 14]:

$$P_r^{(n)} \leq \inf_{Q_X\in\mathcal{P}(\mathcal{X})} \sup_{Q^\star_{U|X}(Q_X)\in\mathcal{P}(\mathcal{U})} \inf_{Q_Y\in\mathcal{P}(\mathcal{Y})} \inf_{\substack{Q_{UXY}\in\mathcal{P}(\mathcal{U}\times\mathcal{X}\times\mathcal{Y}) \\ Q_{U|X}=Q^\star_{U|X}}} G\left[Q_{UXY},Q_X,Q_Y,R\right] , \tag{86}$$

where all the optimization steps are now being taken over *probability distributions*, and G is as defined in Proposition 3.

Evaluation of $P_d^{(n)}$: We now study the Type II error probability of detection, under the assumption that the right sequence has been correctly extracted from the bin. The probability that, given the right sequence $\mathbf{u}^n$, node B makes a wrong decision was investigated in detail in [8], using the method of types [26], as well as properties of types and typical sequences, detailed in Appendix A of this paper. That result, however, is dependent on a specific codebook, conceived to allow detection with high probability. As we use a random codebook in our scheme, it is essential to adapt the method of [8]. We give here a general description of this adaptation.

We propose here a slight modification to [8]. Intuitively, since we investigate the exponential decay of β_n while only enforcing a fixed upper bound on α_n, we show that the penalty of replacing the codebook construction in [8] with random coding can be fully absorbed into α_n, leaving the error exponent result of β_n unmodified. Nevertheless, α_n can still be shown to approach 0 as n grows, which indicates that any constraint $\alpha_n \leq \epsilon$ can be fulfilled, for n large enough and $\epsilon > 0$. For the given codebook, define

$$\mathcal{L}(Q^\star_{UX},Q^\star_{UY}) = \Big\{ P_{\tilde{U}\tilde{X}\tilde{Y}} \in \mathcal{P}(\mathcal{U}\times\mathcal{X}\times\mathcal{Y}) : P_{\tilde{U}\tilde{X}}(u,x) = Q^\star_{UX}(u,x), P_{\tilde{U}\tilde{Y}}(u,y) = Q^\star_{UY}(u,y), \forall\,(u,x,y) \Big\} , \tag{87}$$

to be the set of all triplets of auxiliary RVs such that the marginal distribution of each pair (U,X) and (U,Y) is maintained. Similarly to [8], it is not difficult to show that, for the codebook described above,

$$\theta_L(R) \triangleq \min_{\tilde{U}\tilde{X}\tilde{Y}\in\mathcal{L}(Q^\star_{UX},Q^\star_{UY})} \mathcal{D}(P_{\tilde{U}\tilde{X}\tilde{Y}}\|P_{\bar{U}\bar{X}\bar{Y}}) \tag{88}$$

provides a lower bound to the error probability of the second type, after the correct sequence has been recovered from the bin, and under a fixed error probability of the first type.

From the construction of the codebook (specifically the size of the set $\mathcal{C}^n_U(Q_{\mathbf{x}^n})$), it can be seen that the number of sequences in the codebook *per type of X* complies with $M = \exp\left[n(I(Q_{\mathbf{x}^n};Q^\star_{U|X}(Q_{\mathbf{x}^n}))+\eta)\right]$. Given a sequence $\mathbf{x}^n$, search for a sequence $\mathbf{u}_i$ in the codebook that belongs to the type of $\mathbf{x}^n$, such that $(\mathbf{u}_i^n,\mathbf{x}^n)\in\mathcal{T}^n_{[UX]\delta}$ and send its index (or bin number, depending on the type of $\mathbf{x}^n$) to the receiver. As we only consider here the error event where the wrong hypothesis is chosen despite the correct sequence is used, we ignore errors in choosing the correct sequence from the bin, in case binning has occurred, for the sake of this analysis. If there is more than one such sequence choose randomly. If there is no such sequence in the codebook, send an error message. At the decoder (node B), if $(\mathbf{u}_i^n,\mathbf{y}^n)\in\mathcal{T}^n_{[UY]\delta}$ (notice that typicality here is checked only under hypothesis H_0) declare H_0. In any other case (including the case an error message was received) declare H_1. This choice allows us to "push" the penalty of not using the code proposed in [8, Lemma 4] into α_n (which, when $n\to\infty$ can still be bounded by any fixed $\epsilon>0$), thus leaving the evaluation of β_n unchanged, as shown subsequently.

Evaluation of α_n: An error of the first type occurs if for n i.i.d. samples $(\mathbf{x}^n,\mathbf{y}^n)\sim P_{XY}(x,y)$ (hypothesis H_0 holds) the decoder declares H_1. According to the proposed coding schemes, two possible events can induce the decoder to such an error. The first is given by

$$\text{(i)}\quad \mathcal{B}_9 \triangleq \{\nexists\ i \text{ such that } (\mathbf{u}_i^n,\mathbf{x}^n)\in\mathcal{T}^n_{[UX]\delta}\} . \tag{89}$$

Assuming without loss of generality that the sequence $\mathbf{u}_1^n$ was chosen and sent from node A, the second relevant error event is:

$$\text{(ii)}\quad \mathcal{B}_{10} \triangleq \{H_0 \text{ is true and } (\mathbf{u}_1^n,\mathbf{y}^n)\notin\mathcal{T}^n_{[UY]\delta}\} . \tag{90}$$

From the union bound, it is obvious that:

$$\alpha_n \leq \Pr(\mathcal{B}_9) + \Pr(\mathcal{B}_{10}\cap\mathcal{B}_9^c) . \tag{91}$$

$$G_n\left[Q_{UXY},Q_X,Q_Y,R\right]=\begin{cases}\min\limits_{i=\{0,1\}}\mathcal{D}(Q_{UXY}\|P_{XY_i}Q_{U|X})\\ \quad+\Big[R-I(Q_X;Q^\star_{U|X})+I(Q_Y;Q^\star_{U|Y})\Big]^+ & I(Q_X;Q^\star_{U|X})>R\\ +\infty & \text{else}\ .\end{cases}\tag{85}$$

Through the AEP it is easy to conclude that both of these probabilities approach zero when $n\to\infty$. Thus, for n large enough one can conclude that $\alpha_n\leq\epsilon$ for any fixed $\epsilon>0$.

Evaluation of β_n: The error of the second type can be defined by a single event:

$$\mathcal{B}_{11}\triangleq\{H_1\text{ is true and }(\mathbf{u}_1^n,\mathbf{y}^n)\in\mathcal{T}^n_{[UY]\delta}\}\ .\tag{92}$$

The analysis of β_n is identical to what was done in [8]. One important difference, however, is that by defining

$$\mathscr{C}_i\triangleq\Big\{\mathbf{x}^n\in\mathcal{X}^n:\quad(\mathbf{u}_i^n,\mathbf{x}^n)\in\mathcal{T}^n_{[UX]\delta}\Big\}\ ,\tag{93}$$

the sets $\mathscr{C}_i$ are not necessarily disjoint. This, however, does not change the calculations by following same steps as in [8].

Source reconstruction

As a final step, we demonstrate the achievability of the estimation part in Proposition 3, for the case where hypothesis H_0 is chosen (the case of hypothesis H_1 is symmetric).

Remark 6. *Note that the achievable scheme used here in order to prove Proposition 3 ensures that $\alpha_n\to 0$ when $n\to\infty$, despite this not being a requirement. This is crucial in order for the following analysis, done for hypothesis H_0, to be applicable equivalently also for hypothesis H_1.*

Denoting by $\mathcal{B}_{12}$ the event "an error occurred during encoding or decoding, under the correct decision H_0", we expand its probability as follows: $\Pr(\mathcal{B}_{12})\leq P'+P''$, with P' being the probability that no codeword $\mathbf{v}_0^n(s_0)$ could be found in the codebook for the given sequence $\mathbf{x}^n$ and the chosen sequence $\mathbf{u}^n$, and P'' being the probability that a different codeword in the same bin is compatible with $\mathbf{y}^n$ and $\mathbf{u}^n$.

Using standard arguments, both error probabilities can be bounded as follows:

$$\begin{aligned}P'&\triangleq\Pr\{\nexists\, s_0=[1:\exp(nS_0)]\\&\qquad\text{s.t. }(\mathbf{V}_0^n(s_0),\mathbf{X}^n)\in\mathcal{T}^n_{[V_0X|U]\delta}(\mathbf{u}^n)\}\\&\leq\Pr\{(V_0^n,X^n)\notin\mathcal{T}^n_{[V_0X|U]\delta}(\mathbf{u}^n)|\\&\qquad V^n\in\mathcal{T}^n_{[V_0|U]\delta}(\mathbf{u}^n),X^n\in\mathcal{T}^n_{[X|U]\delta}(\mathbf{u}^n)\}^{\exp(nS_0)}\\&\leq\exp\{-\exp[nS_0]\\&\qquad\times\exp\big[-n(I(X;V_0|U)+\eta_n^{(1)})\big]\Big\}\\&=\exp\Big\{-\exp\big[-n\big(I(X;V_0|U)-S_0+\eta_n^{(1)}\big)\big]\Big\}\ .\end{aligned}\tag{94}$$

Thus, $P'\to 0$ provided that $S_0>I(X;V_0|U)$. Next,

$$\begin{aligned}P''&\triangleq\Pr\{\exists\hat{s}_0\in[1:\exp(nS_0)]\\&\qquad\text{s.t. }\mathbf{V}_0^n(\hat{s}_0)\in\mathcal{T}^n_{[V_0|UY]\delta}(\mathbf{u}^n\mathbf{y}^n),\\&\qquad\quad B_0\big(\mathbf{v}_0^n(s_0)\big)=B_0\big(\mathbf{v}_0^n(\hat{s}_0)\big)\Big\}\\&\leq\exp\left[n(S_0-R_0+\epsilon)\right]\\&\times\Pr\{(\mathbf{V}_0^n,\mathbf{Y}^n)\in\mathcal{T}_{[V_0Y|U]\delta}(\mathbf{u}^n)|\mathbf{V}_0^n\in\mathcal{T}_{[V_0|U]\delta}(\mathbf{u}^n),\\&\qquad\qquad\mathbf{Y}^n\in\mathcal{T}_{[Y|U]\delta}(\mathbf{u}^n)\}\\&\leq\exp\left[n(S_0-R_0+\epsilon)\right]\exp\Big[-n\big(Y;V_0|U)+\eta_n^{(2)}\big)\Big]\\&=\exp\Big\{-n\big[I(Y;V_0|U)-(S_0-R_0)+\eta_n^{(2)}-\epsilon\big]\Big\}\ .\end{aligned}\tag{96}$$

Here, $B_0(\mathbf{v}_0^n(s_0))$ denotes the bin $\mathbf{v}_0^n(s_0)$ belongs to, as defined as part of the encoding strategy. R_0 is the rate dedicated to the estimation part, for the case that H_0 was chosen as the correct hypothesis. Defining R_1 equivalently for hypothesis H_1, the total available rate can be said to be divided, under the proposed achievable scheme, to three parts, such that $R=R'+R_0+R_1$. Thus, $P''\to 0$ if $S_0-R_0<I(X;V_0|U)$, or equivalently

$$\begin{aligned}R_0&>S_0-I(Y;V_0|U)\\&>I(X;V_0U)-I(Y;V_0|U)\\&=I(XY;V_0|U)-I(Y;V_0|U)\\&=I(X;V_0|UY)\ .\end{aligned}\tag{97}$$

Thus, the probability of error related to source reconstruction goes to zero provided that $S_0>I(X;V_0|U)$ and $R_0>I(X;V_0|UY)$. Combining this result with the symmetric case of H_1 and the result for the detection step, the required total rate of communication reads

$$R>R'+I(X;V_0|UY)+I(\bar{X};V_1|\bar{U}\bar{Y})\ .\tag{98}$$

We now know that our scheme allows the decoding of either $\mathbf{v}_0$ and $\mathbf{v}_1$, depending on the case, with high probability, when $n\to\infty$. It remains to be shown that using the sequence $\mathbf{v}_0^n$, it is possible to recover $\mathbf{x}^n$ with distortion D_0. We choose a (possibly suboptimal) decoder, that reconstructs $\mathbf{x}^n$ only from $(\mathbf{u}^n,\mathbf{y}^n,\mathbf{v}_0^n)$:

$$\begin{aligned}d(\mathbf{x}^n,\hat{\mathbf{x}}^n(\mathbf{u}^n,\mathbf{y}^n,\mathbf{v}_0^n))&=\frac{1}{n}\sum_{i=1}^n d\big(x_i,\hat{x}_i(u^n,y^n,v_0^n)\big)\\&=\frac{1}{n}\sum d\big(x,\hat{x}(u,y,v_0)\big)N(x,u,y,v_0|\mathbf{x}^n\mathbf{u}^n\mathbf{y}^n\mathbf{v}_0^n)\\&\leq\mathbb{E}_0\left[d\big(X,\hat{X}(UYV_0)\big)\right]\\&+\sum\left|\frac{1}{n}N(x,u,y,v_0|\mathbf{x}^n\mathbf{u}^n\mathbf{y}^n\mathbf{v}_0^n)-p(x,u,y,v_0)\right|\\&\leq\mathbb{E}_0\left[d\big(X,\hat{X}(UYV_0)\big)\right]+d_{\max}|\mathcal{X}||\mathcal{Y}||\mathcal{U}||\mathcal{V}_0|\delta_n\ ,\end{aligned}\tag{99}$$

where the summation is over all the possible letters in the respective alphabets of the RVs, and the final inequality holds since $(\mathbf{x}^n,\mathbf{y}^n,\mathbf{u}^n,\mathbf{v}_0^n)\in\mathcal{T}^n_{[XYUV_0]\delta}$. Since $\delta_n\to 0$ when $n\to\infty$, any distortion D_0 can be achieved, as long as $D_0>\mathbb{E}_0\left[d\big(X,\hat{X}(UYV_0)\big)\right]$.

APPENDIX E
PROOF OF PROPOSITION 4

We now prove the achievability of the error exponent offered in Proposition 4, for the case where source reconstruction is not required. As the proof is in many ways similar to the proof of Proposition 3, given in Appendix D, we concentrate mainly on the main differences.

Codebook generation and encoding strategy: Both the codebook generation and the encoding strategy in this case are very similar to what was done in the proof of Proposition 3, in the part dedicated to detection. The only difference is that now we choose to only work with δ-typical sequences, for some arbitrary δ. When node A sees a non-typical sequence $\mathbf{x}$, it sends an error message. In the opposite case, encoding is done as before. Note that while we only work with δ-typical sequences, there are still different codebooks for each type *within the set* of δ-typical sequences.

Decoding strategy: In case an error message is received, the decoder declares H_1. This strategy implies that any probability of the error event caused by the encoder not seeing a δ-typical sequence is allocated to α_n, rather than β_n. The probability of this event, however, goes to zero when $n \to \infty$ thanks to the AEP, implying that $\alpha_n \leq \epsilon$ for any $\epsilon > 0$, for $n \geq n_0(\epsilon, \delta)$, thus satisfying the constraint over α_n.

When the encoder does not send an error message, the decoder operates on the entire bin in order to make a decision. Going over the sequences in the bin one by one, the decoder checks for each $\mathbf{u}_i^n$ if $(\mathbf{u}_i^n, \mathbf{y}^n) \in T_{[UY]\delta}^n$. If a sequence in the bin is found, which is jointly typical with $\mathbf{y}^n$, the decoder declares H_0. If no such sequence is found, the decoder declares H_1. Note that under this strategy, the decoder does not attempt to find the original sequence sent by the encoder. Specifically, when the decoder declares H_1 it is completely oblivious to the original codeword.

Probability of error: The analysis of the probability of error in detection under this new strategy is very similar to the analysis given in Appendix D. We separately bound the corresponding error probabilities on the two possible error events.

Analysis of α_n: When analyzing $\alpha_n(\mathcal{A}_n) = \Pr\left(\mathcal{A}_n^c | XY \sim p_0(x,y)\right)$, we assume throughout that the probability measure in effect is p_0. Two scenarios can lead to an event where the decoder erroneously declares H_1:

$$\begin{aligned} \mathcal{B}_{13} &\triangleq \left\{\nexists\, i \in \mathcal{C}_U^n(Q_{\mathbf{x}^n}) \,\middle|\, (\mathbf{x}^n, \mathbf{u}_i^n) \in \mathcal{T}_{[UX]\delta}^n\right\} , \\ \mathcal{B}_{14} &\triangleq \left\{\nexists\, i \in F(f(\mathbf{x}^n)) \,\middle|\, (\mathbf{u}_i^n, \mathbf{y}^n) \in \mathcal{T}_{[UY]\delta}^n\right\} . \end{aligned} \tag{100}$$

In the first event, an error message is sent, as there is no fitting codeword within the codebook for the observed sequence $\mathbf{x}^n$. Whereas for the second event, there is no sequence in the bin that prompts the decoder to decide H_0, despite it being the true hypothesis. The probability of event $\mathcal{B}_{13}$ goes to zero with n, thanks to the AEP and the size of the codebook. As for event $\mathcal{B}_{14}$, assume without loss of generality, that the encoder intended to send the first word in the bin $\mathbf{u}_1^n$, i.e., $\mathbf{u}_1^n = f(\mathbf{x}^n)$. The probability that the decoder declares H_1 can be upper-bounded by

$$\begin{aligned} \Pr(\mathcal{B}_{14}) &= \Pr\left\{\nexists\, i \in F(f(\mathbf{X}^n)) \,\middle|\, (\mathbf{U}_i^n, \mathbf{Y}^n) \in \mathcal{T}_{[UY]\delta}^n\right\} \\ &\leq \Pr\{(\mathbf{U}_1^n, \mathbf{Y}^n) \notin \mathcal{T}_{[UY]\delta}^n\} , \end{aligned} \tag{101}$$

where typicality is measured over the probability measure $p_0 = P_{XY}$. As was already discussed above, this probability tends to 0 with the number of available realizations n. This result is attributed to the AEP, by which $\mathbf{x}$ and $\mathbf{y}$ are jointly typical with high probability, and to the generalized Markov Lemma (Lemma 6). Thus, any fixed constraint over the probability of error of the first type $\alpha \leq \epsilon$ ($\epsilon > 0$), may be satisfied when n is large enough.

Analysis of β_n: As we now turn to analyzing the probability of error of the second type, we assume throughout this part that the real hypothesis is H_1. As was the case in Appendix D, the resulting error exponent is the result of a trade-off between two error events. While the analysis of the event where the correct sequence prompts a wrong decision (i.e. in this case is $(f(\mathbf{x}^n), \mathbf{y}^n) \in \mathcal{T}_{[UY]\delta}^n$) stays the same, the second error event is now different. We thus concentrate in this appendix on calculating the probability of the event that some sequence in the bin $\mathbf{u}^n \neq f(\mathbf{x}^n)$ prompts the decoder to declare H_0. We start by presenting the following lemma:

Lemma 10. *Let $\mathcal{A}_n$ be the set of triplets, such that a binned codebook is necessary:*

$$\begin{aligned} \mathcal{A}_n = \Big\{ &(\mathbf{u}^n, \mathbf{x}^n, \mathbf{y}^n) \in T_{Q_{U|X}^\star}^n \times \mathcal{X}^n \times \mathcal{Y}^n \big| \\ &\log |\mathcal{C}_U^n(Q_{\mathbf{x}^n})| \geq nR \Big\} . \end{aligned} \tag{102}$$

Let $(\mathbf{u}^n, \mathbf{x}^n, \mathbf{y}^n) \in \mathcal{A}_n$ and denote by $\mathcal{B}_{15}$ the event indicating that $(\mathbf{u}^n, \mathbf{y}^n) \in \mathcal{T}_{[UY]\delta}^n$, for some $\mathbf{u}^n \neq f(\mathbf{x}^n)$ in the bin. Then,

$$\begin{aligned} &\Pr\left(\mathcal{B}_{15} | \mathbf{U}^n = \mathbf{u}^n, \mathbf{X}^n = \mathbf{x}^n, \mathbf{Y}^n = \mathbf{y}^n\right) \\ &\leq \exp\left[-n\left(R - \hat{J}(Q_{\mathbf{u}^n\mathbf{x}^n\mathbf{y}^n}) - \delta_n\right)\right] , \end{aligned} \tag{103}$$

with

$$\begin{aligned} \hat{J}(Q_{\mathbf{u}^n\mathbf{x}^n\mathbf{y}^n}) \triangleq &I\big(Q_{\mathbf{x}^n}; Q_{U|X}^\star\big) - H(Q_{\mathbf{u}^n}) \\ &+ H\big(Q_{U|Y} | P_Y\big) \end{aligned} \tag{104}$$

and

$$\delta_n \triangleq \frac{1}{n} \log(n+1)^{|\mathcal{U}|(1+|\mathcal{X}|+|\mathcal{Y}|)+4} + \epsilon_n \tag{105}$$

with $\epsilon_n \to 0$ when $n \to \infty$. Moreover, the probability in (103) *is taken over the choice of the codebook in use.*

Proof: The proof of Lemma 10 is very similar to the one given for Lemma 8. The difference is that now the set of sequences that "confuses" the decoder is simply $\hat{\mathcal{S}}(\mathbf{y}^n) = \mathcal{T}_{[U|Y]\delta}^n(\mathbf{y}^n)$. Bounding the set of conditionally typical sequences by [28]:

$$\begin{aligned} &\left|\mathcal{T}_{[U|Y]\delta}^n(\mathbf{y}^n)\right| \\ &\leq (n+1)^{|\mathcal{U}||\mathcal{Y}|} \exp\left[n(H(Q_{U|Y}|P_Y) + \epsilon_n)\right] , \end{aligned} \tag{106}$$

for each $\mathbf{y}^n \in \mathcal{T}_{[Y]\delta}^n$, completes the proof. ∎

Remark 7. *Note that unlike $J(Q_{\mathbf{u}^n\mathbf{x}^n\mathbf{y}^n})$, the quantity $\hat{J}(Q_{\mathbf{u}^n\mathbf{x}^n\mathbf{y}^n})$ is not dependent on the observed $\mathbf{y}^n$. The quantity $H(Q_{U|Y}|P_Y)$ can be analytically calculated when the type of $\mathbf{x}^n$ and the chosen strategy $Q_{U|X}$ is known, without knowing neither the specific sent sequence $\mathbf{u}^n$ nor the observed sequence $\mathbf{y}^n$.*

Using Lemma 10 and summing over all involved types and sequences within each type as was done in Appendix D, the probability of the event where an unintended sequence in the bin causes an error can be bounded by

$$\begin{aligned}
&\lim_{n\to\infty} -\frac{1}{n}\log\Pr(\mathcal{B}_{15}) \geq \\
&\min_{Q_X\in\mathcal{P}_n(\mathcal{X})}\ \max_{Q^\star_{U|X}(Q_X)\in\mathcal{P}_n(\mathcal{U})}\ \min_{Q_Y\in\mathcal{P}_n(\mathcal{Y})}\ \min_{Q_{UXY}\in\mathcal{P}_n(\mathcal{U}\times\mathcal{X}\times\mathcal{Y})} \\
&\qquad \Big\{\mathcal{D}(Q_{UXY}\|P_{\bar{U}\bar{X}\bar{Y}}) + R - \hat{J}(Q_{UXY})\Big\} \\
&= \min_{Q_X}\ \max_{Q^\star_{U|X}(Q_X)}\ \min_{Q_Y}\ \min_{Q_{UXY}} \\
&\qquad \Big\{\mathcal{D}(Q_{UXY}\|P_{\bar{U}\bar{X}\bar{Y}}) + R \\
&\qquad - I(Q_X;Q^\star_{U|X}) + I(Q^\star_{U|Y};P_Y)\Big\}\ .
\end{aligned}$$

As in this case we only work with δ-typical x-sequences, we may choose δ to be any value, as long as it is strictly positive. Thus, we may force Q_X to be arbitrarily close to P_X by taking $\delta\to 0^+$. The error exponent in question thus becomes

$$\begin{aligned}
&\lim_{n\to\infty} -\frac{1}{n}\log\Pr(\mathcal{B}_{15}) \\
&\quad\geq \max_{Q^\star_{U|X}\in\mathcal{P}(\mathcal{U})}\Big\{R - I(P_X;Q^\star_{U|X}) + I(P_Y;Q^\star_{U|Y}) \\
&\quad + \min_{Q_Y\in\mathcal{P}(\mathcal{Y})}\ \min_{Q_{UXY}\in\mathcal{P}(\mathcal{U}\times\mathcal{X}\times\mathcal{Y})}\mathcal{D}(Q_{UXY}\|P_{\bar{U}\bar{X}\bar{Y}})\Big\} + \hat{\epsilon} \\
&\quad= \max_{Q^\star_{U|X}\in\mathcal{P}(\mathcal{U})}\Big\{R - I(P_X;Q^\star_{U|X}) + I(P_Y;Q^\star_{U|Y})\Big\} + \hat{\epsilon}\ ,
\end{aligned}$$

with $\hat{\epsilon}\to 0$ as $\delta\to 0$. This, along with an analysis of the complementary error event similar to the one given for Proposition 3, completes the proof of Proposition 4.

ACKNOWLEDGMENT

The authors are grateful to Prof. Romain Couillet for his valuable comments at the early stage of this work. They are also grateful to the Associate Editor, and to anonymous reviewers for their constructive and helpful comments on the earlier version of the manuscript.

REFERENCES

[1] G. Katz, P. Piantanida, R. Couillet, and M. Debbah, "Joint estimation and detection against independence," in *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on*, Sept 2014, pp. 1220–1227.

[2] ——, "On the necessity of binning for the distributed hypothesis testing problem," in *Information Theory (ISIT), 2015 IEEE International Symposium on*, June 2015, pp. 2797–2801.

[3] E. Lehmann and J. Romano, *Testing Statistical Hypotheses*, ser. Springer Texts in Statistics. Springer, 2005.

[4] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York: John Wiley & Sons, 1991.

[5] R. Tenney and N. R. Sandell, "Detection with distributed sensors," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. AES-17, no. 4, pp. 501–510, July 1981.

[6] T. Han and S.-I. Amari, "Statistical inference under multiterminal data compression," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2300–2324, Oct 1998.

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

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

[9] 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.

[10] A. Kaspi, "Rate-distortion function when side-information may be present at the decoder," *Information Theory, IEEE Transactions on*, vol. 40, no. 6, pp. 2031–2034, Nov 1994.

[11] C. Tian and J. Chen, "Remote vector gaussian source coding with decoder side information under mutual information and distortion constraints," *Information Theory, IEEE Transactions on*, vol. 55, no. 10, pp. 4676–4680, Oct 2009.

[12] ——, "Successive refinement for hypothesis testing and lossless one-helper problem," *Information Theory, IEEE Transactions on*, vol. 54, no. 10, pp. 4666–4681, Oct 2008.

[13] Y. Xiang and Y.-H. Kim, "Interactive hypothesis testing with communication constraints," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, Oct 2012, pp. 1065–1072.

[14] A. Kaspi, "Two-way source coding with a fidelity criterion," *Information Theory, IEEE Transactions on*, vol. 31, no. 6, pp. 735–740, Nov 1985.

[15] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 727 – 734, Nov. 1985.

[16] H. Shimokawa, T. Han, and S.-I. Amari, "Error bound of hypothesis testing with data compression," in *Inf. Theory, 1994 IEEE International Symposium on (ISIT)*, Jun 1994, p. 114.

[17] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2148–2177, Oct 1998.

[18] S. Rahman and A. Wagner, "On the optimality of binning for distributed hypothesis testing," *Information Theory, IEEE Transactions on*, vol. 58, no. 10, pp. 6282–6303, Oct 2012.

[19] W. Diffie and M. Hellman, "New directions in cryptography," *Information Theory, IEEE Transactions on*, vol. 22, no. 6, pp. 644–654, Nov 1976.

[20] Y.-C. Lin, D. Varodayan, and B. Girod, "Image authentication using distributed source coding," *Image Processing, IEEE Transactions on*, vol. 21, no. 1, pp. 273–283, Jan 2012.

[21] C.-Y. Lin and S.-F. Chang, "A robust image authentication method distinguishing jpeg compression from malicious manipulation," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 11, no. 2, pp. 153–168, Feb 2001.

[22] G. Chaojun, P. Jirutitijaroen, and M. Motani, "Detecting false data injection attacks in ac state estimation," *Smart Grid, IEEE Transactions on*, vol. 6, no. 5, pp. 2476–2483, Sept 2015.

[23] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks," *Smart Grid, IEEE Transactions on*, vol. 4, no. 3, pp. 1244–1253, Sept 2013.

[24] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *Information Theory, IEEE Transactions on*, vol. 22, no. 1, pp. 1–10, Jan 1976.

[25] B. Kelly and A. Wagner, "Reliability in source coding with side information," *Information Theory, IEEE Transactions on*, vol. 58, no. 8, pp. 5086–5111, Aug 2012.

[26] I. Csiszár, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, Oct 1998.

[27] Y. Steinberg and N. Merhav, "On successive refinement for the wyner-ziv problem," *Information Theory, IEEE Transactions on*, vol. 50, no. 8, pp. 1636–1654, Aug 2004.

[28] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.

[29] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[30] P. Piantanida, L. Rey Vega, and A. Hero, "A proof of the generalized markov lemma with countable infinite sources," in *Information Theory Proceedings (ISIT), 2014 IEEE International Symposium on*, July 2014.