Hypothesis Testing and Identification Systems

Minh Thanh Vu, Tobias J. Oechtering and Mikael Skoglund
Department of Information Science and Engineering
KTH Royal Institute of Technology

Abstract

We study hypothesis testing problems with fixed compression mappings and with user-dependent compression mappings to decide whether or not an observation sequence is related to one of the users in a database, which contains compressed versions of users' data. We provide optimal characterizations of the exponents of the probability of the second kind of error when the number of users in the database grows exponentially. Additionally we also characterize the identification capacity when different compression mappings that vary among users are used to enroll user sequences into the database. We establish exponentially strong converse equivalence between different settings. Finally we show that an identification scheme can be turned into a multi-user hypothesis testing scheme and vice versa.

Index Terms

Mixture distribution, identification systems, information-spectrum method, strong converse, excess relative information, soft-covering, constructive transformation.

I. Introduction

Membership testing has not been actively considered in existing works on identification systems. It is often assumed that the observation sequence is related to the data inside the system. In this work we put our attention to this important problem. Assume a database that stores *compressed* versions of data sequences of M users $(x^n(m))_{m=1}^M$. An observation sequence y^n is provided to a processing center which has access to these compressed data sequences. The processing center performs a screening step and returns *Yes/No* when y^n is *related to one of the user/ independent of all users* in the system. We call the first case hypothesis H_0 and the second case hypothesis H_1. The case that $M = 1$ was studied in [1] and referred in this work as single-user testing against independence to differentiate it from our multi-user setting.

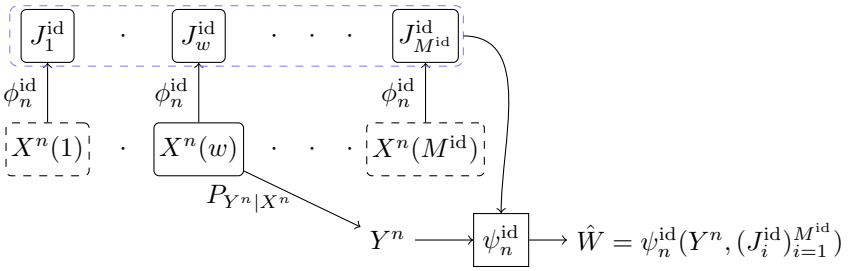

Fig. 1: A simplified model for the identification system. In the *enrollment* phase, observations $(x^n(i))_{i=1}^M$ are mapped into stored indices by a mapping ϕ_n^{id}. In the *identification* phase, one user w is selected uniformly at random. The random selection is described by a random variable W. An observation sequence y^n which is the output of the observation channel $P_{Y^n|X^n}$ with the input $x^n(w)$ is provided to the processing center. The processing center tries to recognize the user by producing an estimate index $\hat{w}$ based on y^n and the database.

Our problem is influenced by the identification system, cf. Fig. 1, investigated in [2], where the task is to identify the correct user based on the observation y^n and the information inside the database $(x^n(i))_{i=1}^M$. Therein, y^n is an output of an observation channel $P_{Y|X}$ where the input sequence is selected uniformly at random from the database $(x^n(i))_{i=1}^M$. The compression of users' data is considered in [3], [4]. In contrast, we are not interested in identifying the true user even if the sequence y^n is indeed related to one user in the system. Instead, we only want to know if the sequence is related to the database.

We first study the hypothesis testing problem when only one compression mapping is used to enroll all users' data. Next, we generalize the setting by allowing that each user can have a different compression mapping. This generalization may arise when separate or distributed databases which use different compression techniques are merged together for the identification purpose. We also examine the related identification problem where each user can have their own compression mapping. Finally, we study the code transformation between the identification systems and our multi-user hypothesis testing problem.

We summarize our contributions in the following.

This paper was presented in parts at ITW 2018 and at ISIT 2019.

- We characterize the error exponent of the hypothesis testing problem with a single compression mapping in both weak converse sense and strong converse sense. Our results complement the one given in [3] by showing the fundamental trade-off between the number of users in the database and the ability to reject a stranger. In both weak and strong converse proofs we use a soft-covering lemma to transform our multi-user setting to a single-user setting. To the best of our knowledge, this step is the first instance of novel use of a soft-covering lemma in converse proofs, which is otherwise mainly used in achievability proofs.
- We establish an exponentially strong converse equivalence between our multi-user setting and the single-user testing against independence via a code transformation argument. Furthermore, we connect the exponentially strong converse for the single-user testing problem with the one of the Wyner-Ahlswede-Körner (WAK) network [5], [6] provided by Oohama [7], [8] by using a similar code transformation argument.
- We show that the optimal characterizations in the strong converse sense of both the multi-user hypothesis testing problem and the identification problem with user-dependent compression mappings under a uniformity condition are the same as the ones for the single compression mappings. The generalized tools used in these proofs are also of independent interest.
- Finally, we show that a pair of mappings for an identification system can be explicitly turned into a pair of mappings for our multi-user hypothesis testing problem and vice versa. This property is highly desirable in practice since one can add a new feature, herein rejecting a stranger, on top of an old feature (recognizing an existing user) without re-designing the whole system from the scratch. Several consequences of the code transformation such as the equivalence of the ϵ-achievability and the equivalence of the second-order achievability are also provided.

A related identification model was studied in [1, Section V] which was an extension of the one in [9]. Ahlswede and Csiszár considered a database $(x^n(i))_{i=1}^M \subset \mathcal{X}^{nM}$ iid generated from $P_X^{\otimes}$ of M users. Under hypothesis H_0, which happens with probability $\pi_0 > 0$ an observation y^n is related to a user i according to $P_{Y|X}^{\otimes n}$. While under hypothesis H_1, which happens with probability $1-\pi_0$, y^n is independent of all $(x^n(i))_{i=1}^M$. The authors approached the problem using the Bayesian framework and used the following *specific* searching procedure to calculate the expected cost. Given y^n, the processing searches for a list of indices j such that $x^n(i) \in \mathcal{G}(y^n) \subset \mathcal{X}^n$. The decision incurs a cost c for each matched index irrespective of whether H_0 or H_1 is true. Additionally, if H_0 is true a cost k is incurred if the true one is not in the list, i.e., $x^n(i) \notin \mathcal{G}(y^n)$. It was shown that the total expected cost can be expressed in terms of error probabilities of type I and II of testing P_{YX} against $P_Y \times P_X$ as $C_n = c(M-\pi_0)\beta_n + \pi_0(k-c)\alpha_n + \pi_0 c$. The objective of [1, Section V] is to obtain the minimum cost $C_n^\star(R_c)$ of the total expected cost C_n when all data sequences $(x^n(i))_{i=1}^M$ are compressed using the same compression mapping at a rate R_c. When the false alarm cost is greater than the miss detection cost $k > c$, an upper bound and a lower bound on $C_n^\star(R_c)$ were given when $R < R_{\max}(R_c)$ and $R > R_{\max}(R_c)$, respectively. The rate R corresponds to the number of user M and $R_{\max}(R_c)$ is the maximum error exponent for the testing against independence problem in [1]. In contrast, we formulate the problem according to the Neyman-Pearson framework in which we do not assume any specific decision making procedure. Additionally, with this approach, each user can have a distinct compression mapping. We characterize the optimal error exponents.

Other hypothesis testing problems related to the identification problem include [10]–[12]. In [10] the hypothesis H_1 was tested against M other hypotheses in the binary setting where the focus was to minimize the overall identification error under a specific decision rule. It was shown that when the rate of M is below a value, which is less than or equal to the uncompressed binary identification capacity, then the overall error probability goes to zero. In [11] the author considered the M-ary hypothesis testing problem with fixed M and studied the large deviation regime. In [12] the decision rule was based on a decoding metric using the hashed data and observation sequences at different lengths. The exponents of the probability of miss and the expected number of incorrect items on the list were provided for a fixed hashed function. Error exponent aspects of the probability of estimating the correct user in the identification systems have been studied in [13]–[15]. Recent developments on distributed hypothesis testing with privacy constraints and relay networks can be found in [16]–[19].

The paper is organized as follows. We present the complete achievable error exponent in the weak converse sense of our multi-user hypothesis testing when the same compression mapping is used for all users in Section II. We also remark that the arguments used in the weak converse proof can be applied for other settings. In Section III we show the optimality of the same compression mapping setting in the strong converse sense and establish the exponentially strong converse equivalence between testing cases. We also show the strong converse proofs for the user-dependent compression mapping settings. Finally, in Section IV we present the code transformation arguments between the single-user testing, the WAK network, and the identification systems as well as their consequences.

II. Preliminaries

We begin with some notational conventions. Random variables, their realizations are denoted by uppercase, lowercase letters, respectively. Sets are denoted by calligraphic letters. The complement of a set $\mathcal{A}$ is denoted by $\mathcal{A}^c$. $\log$ is taken to the natural base. For a mapping ϕ_n, $|\phi_n^{\mathrm{t}}|$ denotes the cardinality of its image. For a measure μ, $\mu^{\otimes n}$ denotes its n-fold product extension. Additionally, we use the $\bar{(\cdot)}$ notation, e.g. $\bar{\alpha}_n$, $\bar{\mathcal{A}}_n$, to emphasize that the single-user scenario is considered.

We assume that the database consists of M users with $\lim_{n\to\infty} \frac{1}{n}\log M = R$. i.e., the number of users grows with the block length n at rate R. We also denote this relation in the sequel by $M \doteq e^{nR}$. In this and the next sections, if not otherwise stated,

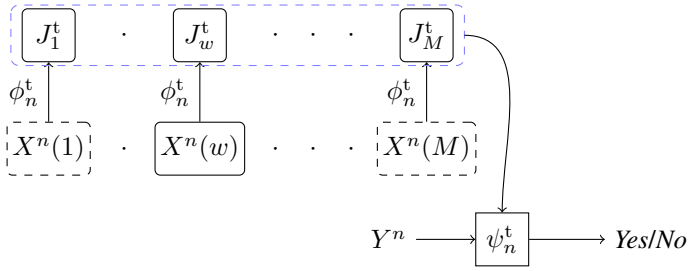

Fig. 2: Screening to find out whether the observation is related to one user inside the system.

we mainly consider the discrete scenario where alphabets $\mathcal{X}$ and $\mathcal{Y}$ are finite. For each i the corresponding data sequence $x^n(i)$ is generated iid from the distribution P_X. Under H_0 the joint distribution of the sequence y^n and sequences $(x^n(m))_{m=1}^M$ is given by

$$P_{H_0} = \sum_{i=1}^{M} \frac{1}{M} P_{Y^n X^n(i)} \times \prod_{k=1, k\neq i}^{M} P_{X^n(k)}, \tag{1}$$

i.e., the sequence y^n is related to one randomly chosen user in the system, where $P_{Y^n X^n(i)} = P_{Y|X}^{\otimes n} \times P_{X^n(i)}$. The joint distribution under H_1 is given by

$$P_{H_1} = P_Y^{\otimes n} \times \prod_{i=1}^{M} P_{X^n(i)}, \tag{2}$$

i.e., the sequence y^n is not related to the information in the database. Note that under both hypotheses the users' data sequences are mutually independent. Additionally, the processes governed by P_{H_0} and P_{H_1} are *non-stationary* since $M \doteq e^{nR}$ grows exponentially with n[1]. We can also view P_{H_0} as the result of mixing M general random processes uniformly where the distributions at instance n are given by $P_{Y^n X^n(i)} \times \prod_{k=1,k\neq i}^{M} P_{X^n(k)}$, $i \in [1:M]$.

We first consider the case that a single compression mapping is used to enroll all users' data into the database. We illustrate the setting in Fig. 2. We state in the following the formal definition of a corresponding testing scheme.

Definition 1. *A testing scheme consists of a compression mapping ϕ_n^{t} which enrolls the users' data sequences of length n into the database according to*

$$\phi_n^{\mathrm{t}} \colon \mathcal{X}^n \to \mathcal{M}_1, \tag{3}$$

and a decision mapping ψ_n^{t} which outputs whether H_0 or H_1 is deemed true

$$\psi_n^{\mathrm{t}} \colon \mathcal{Y}^n \times \mathcal{M}_1^M \to \{0,1\}. \tag{4}$$

Note that the compression mapping for our membership testing scheme ϕ_n^{t} could be different from the compression mapping ϕ_n^{id} in Fig. 1, i.e., the testing scheme might induce additional storage space. In Section IV we show that an identification scheme $(\phi_n^{\mathrm{id}}, \psi_n^{\mathrm{id}})$ can be turned into a testing scheme $(\phi_n^{\mathrm{t}}, \psi_n^{\mathrm{t}})$ and vice versa such that $\phi_n^{\mathrm{t}} = \phi_n^{\mathrm{id}}$, i.e., no additional database for the pre-scan is needed.

For brevity in the following we abbreviate the ensemble of user sequences $(X^n(i))_{i=1}^M$ as $\mathbf{X}^n$ and the ensemble of enrollments $(\phi_n^{\mathrm{t}}(X^n(i)))_{i=1}^M$ as $\phi_n^{\mathrm{t}}(\mathbf{X}^n)$. The bold lower case notations $\mathbf{x}^n$ and $\phi_n^{\mathrm{t}}(\mathbf{x}^n)$ are used to denote the corresponding realizations. For a given compression mapping ϕ_n^{t} the induced distributions under both hypotheses are denoted by

$$H_0 \colon P_{Y^n \phi_n^{\mathrm{t}}(\mathbf{X}^n)}, \quad H_1 \colon P_{Y^n} \times P_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}. \tag{5}$$

The acceptance region of hypothesis H_0 is defined as

$$\mathcal{A}_n = \big\{(y^n, \phi_n^{\mathrm{t}}(\mathbf{x}^n)) \mid \psi_n(y^n, \phi_n^{\mathrm{t}}(\mathbf{x}^n)) = 0\big\}. \tag{6}$$

An error of the first (second) type occurs when y^n is related to one unknown user (independent from all users) in the system but the testing scheme declares otherwise. Accordingly, the probability of first and second type of error are given respectively as

$$\alpha_n = P_{Y^n \phi_n^{\mathrm{t}}(\mathbf{X}^n)}(\mathcal{A}_n^c), \ \beta_n = P_{Y^n} \times P_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}(\mathcal{A}_n). \tag{7}$$

Definition 2. *An error exponent E of type II is achievable given (R, R_c) with $M \doteq e^{nR}$ if there exist a sequence of testing schemes $(\phi_n^{\mathrm{t}}, \psi_n)$ such that*

$$\lim_{n\to\infty} \alpha_n = 0, \ \limsup_{n\to\infty} \frac{1}{n} \log |\mathcal{M}_1| \leq R_c,$$

[1]Under P_{H_0} the process is also *non-ergodic*.

$$\liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_n} \geq E. \tag{8}$$

We define

$$E_{\mathrm{f}}^{\star}(R, R_c) = \sup\{E \mid E \textit{ is achievable given } (R, R_c)\}. \tag{9}$$

Note that the case $M = 1$ in our setting corresponds to the testing against independence setting studied in [1]. We refer to it in the following as the single-user scenario.
Given a compression rate R_c we define the following functions

$$\begin{aligned} R_{\max}(R_c) &= \max_{\substack{U-X-Y,\ |\mathcal{U}|\leq|\mathcal{X}|+1,\\ I(X;U)\leq R_c}} I(Y;U) \\ \theta(R, R_c) &= R_{\max}(R_c) - R, \text{ on } 0 \leq R < R_{\max}(R_c). \end{aligned} \tag{10}$$

Different interpretations of $R_{\max}(R_c)$ appear in previous works. In [1] $R_{\max}(R_c)$ is the maximum error exponent of type II for the single-user testing against independence problem. In [3] $R_{\max}(R_c)$ characterizes the number of users that can be supported at a given compression rate R_c in the identification systems in Fig. 1 with vanishing probability of identification error.

Our first result characterizes the error exponent $E_{\mathrm{f}}^{\star}(R, R_c)$.

Theorem 1. *Given a single-user hypothesis testing scheme $(\bar{\phi}_n^{\mathrm{t}}, \bar{\psi}_n^{\mathrm{t}})$ with probabilities of errors $\bar{\alpha}_n$ and $\bar{\beta}_n$. Using the same compression mapping $\bar{\phi}_n^{\mathrm{t}}$, we can construct a testing scheme $(\bar{\phi}_n^{\mathrm{t}}, \psi_n')$, for the multi-user case that achieves the following multi-user probabilities of error*

$$\alpha_n \leq \bar{\alpha}_n,\ \beta_n \leq \bar{\beta}_n M. \tag{11}$$

Consequently, for $E_{\mathrm{f}}^{\star}(R, R_c)$ defined in Definition 2 we have

$$E_{\mathrm{f}}^{\star}(R, R_c) = \begin{cases} \theta(R, R_c) & \textit{when } R < R_{\max}(R_c), \\ 0 & \textit{otherwise.} \end{cases} \tag{12}$$

For notational brevity, in the sequel we abbreviate $R_{\max}(R_c)$ as $R_{\max}$.

Proof. The achievability part straightforwardly follows from the single-user scenario. An outline is presented here for completeness. Let $\bar{\mathcal{A}}_n$ be the acceptance region for the single-user scenario corresponding to $(\bar{\phi}_n^{\mathrm{t}}, \bar{\psi}_n)$. For each user $i \in [1 : M]$ the compression mapping $\bar{\phi}_n^{\mathrm{t}}$ maps the data sequence $x^n(i)$ into an index that is stored in the database. We define the acceptance region for the multi-user case as follows

$$\mathcal{A}_n = \{(y^n, \bar{\phi}_n^{\mathrm{t}}(\mathbf{x}^n)) \mid (y^n, \bar{\phi}_n^{\mathrm{t}}(x^n(i))) \in \bar{\mathcal{A}}_n \text{ for some } i\}. \tag{13}$$

The probabilities of the first and second type of errors can be bounded accordingly as in (11). We omit the details. The existence of testing schemes that achieve the exponent $R_{\max}(R_c) - \gamma$ for the single-user scenario for any $\gamma > 0$ is well-known, cf. [1], [20]. Thus $E_{\mathrm{f}}^{\star}(R, R_c) \geq \theta(R, R_c)$ for $R < R_{\max}(R_c)$ follows from (11) since $M \doteq e^{nR}$ holds.

We now present the converse for Theorem 1. Since the distributions in our setting are non-stationary, the standard weak converse proof using divergence cannot be straightforwardly extended. We will use in the following the information-spectrum method plus a covering lemma to obtain the weak converse result.

We first consider the case $R < R_{\max}(R_c)$. Suppose that $E > 0$ is an achievable error exponent, i.e. there exists a sequence of compression mappings (ϕ_n^{t}) and a sequence of decision mappings (ψ_n^{t}) such that all the conditions in Definition 1 hold. Then we can view (ψ_n^{t}) as decision mappings for the simple hypothesis testing problem $H_0 : P_{Y^n \phi_n^{\mathrm{t}}(\mathbf{X}^n)}$ versus $H_1 : P_Y^n \times P_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}$. Define two general sources $\mathfrak{C}^1 = \{C_n^1\}_{n=1}^{\infty}$, $\mathfrak{C}^2 = \{C_n^2\}_{n=1}^{\infty}$ in which for each n, $C_n^1 \sim P_{Y^n \phi_n^{\mathrm{t}}(\mathbf{X}^n)}$ and $C_n^2 \sim P_{Y^n} \times P_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}$ hold. By [21, Theorem 4.1.1] we have

$$E \leq \underline{D}(\mathfrak{C}^1 \| \mathfrak{C}^2), \tag{14}$$

where the right-hand side is the *spectral inf-divergence*, which is defined as for two general sources $\mathbf{Z} = \{Z_i\}_{i=1}^{\infty}$ and $\mathbf{Z}' = \{Z_i'\}_{i=1}^{\infty}$ as

$$\underline{D}(\mathbf{Z} \| \mathbf{Z}') = \sup\left\{\alpha \mid \lim_{n\to\infty} \Pr\left\{\frac{1}{n} \log \frac{dP_{Z^n}}{dP_{Z'^n}}(Z^n) < \alpha\right\} = 0\right\}. \tag{15}$$

Using Lemma 1 presented in the next section with $\hat{E} = \underline{D}(\mathfrak{C}^1 \| \mathfrak{C}^2) - \gamma$ for any $\gamma > 0$ small enough such that $\hat{E} > 0$, we obtain

$$\lim_{n\to\infty} \Pr\left\{\frac{1}{n} \log \frac{P_{\bar{Y}^n \phi_n^{\mathrm{t}}(\bar{X}^n)}}{P_{\bar{Y}^n} \times P_{\phi_n^{\mathrm{t}}(\bar{X}^n)}}(\bar{Y}^n, \phi_n^{\mathrm{t}}(\bar{X}^n)) < R + \hat{E} - \gamma\right\} = 0. \tag{16}$$

This implies that

$$R + \hat{E} - \gamma \leq \underline{D}(\bar{\mathfrak{C}}^1 \| \bar{\mathfrak{C}}^2), \tag{17}$$

in which $(\bar{Y}^n, \bar{X}^n) \sim P_{XY}^{\otimes n}$ and $\bar{\mathfrak{C}}^1 = \{\bar{C}_n^1\}_{n=1}^{\infty}$, $\bar{\mathfrak{C}}^2 = \{\bar{C}_n^2\}_{n=1}^{\infty}$ are two general source where for each n, $\bar{C}_n^1 \sim P_{\bar{Y}^n \phi_n^{\mathrm{t}}(\bar{X}^n)}$ and $\bar{C}_n^2 \sim P_{\bar{Y}^n} \times P_{\phi_n^{\mathrm{t}}(\bar{X}^n)}$ hold. From [21, Theorem 3.5.2] the right-hand side of (17) can be upper-bounded further by the following

$$\underline{D}(\bar{\mathfrak{C}}^1 \| \bar{\mathfrak{C}}^2) \leq \liminf_{n\to\infty} \frac{1}{n} I(\bar{Y}^n; \phi_n^{\mathrm{t}}(\bar{X}^n)). \tag{18}$$

Let (n_k) be a sub-sequence of indices such that $\frac{1}{n_k} I(\bar{Y}^{n_k}; \phi_{n_k}^{\mathrm{t}}(\bar{X}^{n_k}))$ converges to the right-hand side of (18). With n_k sufficiently large we have $\frac{1}{n_k} \log |\phi_{n_k}^{\mathrm{t}}| \leq R_c + \gamma$ and

$$\liminf_{n\to\infty} \frac{1}{n} I(\bar{Y}^n; \phi_n^{\mathrm{t}}(\bar{X}^n)) \leq \frac{1}{n_k} I(\bar{Y}^{n_k}; \phi_{n_k}^{\mathrm{t}}(\bar{X}^{n_k})) + \gamma \leq R_{\max}(R_c + \gamma) + \gamma. \tag{19}$$

The last inequality follows from the corresponding entropy characterization of $R_{\max}(R_c)$ provided in [1]. This step can also be carried out using the standard single-letterization approach. In summary we have

$$E - 3\gamma \leq R_{\max}(R_c + \gamma) - R, \quad \stackrel{\gamma\to 0}{\Longrightarrow} \quad E \leq R_{\max}(R_c) - R. \tag{20}$$

Since the right-hand side of the above inequality is positive, our reasoning is valid. This implies that $E_{\mathrm{f}}^{\star}(R, R_c) \leq R_{\max}(R_c) - R$, as E is arbitrary.

Now, assume that $R \geq R_{\max}(R_c)$. Let (ϕ_n^t, ψ_n^t) be a testing scheme such that $\lim_{n\to\infty} \alpha_n = 0$. Suppose further that (16) is still valid, i.e., $\underline{D}(\mathfrak{C}^1 \| \mathfrak{C}^2) > 0$ holds. Then by taking $\gamma \to 0$ we see that (17) is contradicted. Therefore $\underline{D}(\mathfrak{C}^1 \| \mathfrak{C}^2) = 0$ for all sequences (ϕ_n^{t}). Therefore $E_{\mathrm{f}}^{\star}(R, R_c) = 0$ in this case. □

Remark 1. *We note that the proof of Theorem 1 remains valid when P_{XY} is jointly Gaussian. The only step that needs to be changed is* (19) *when an explicit calculation based on the entropy power inequality is needed. We leave the details to the interested reader.*

Remark 2. *Note that in the case of no compression our setting is an instance of hypothesis testing for the mixed source problem. When M is a constant, we obtain a similar result as in [21, Example 4.1.1]. Namely the optimal exponent is given by $E_{\mathrm{f}}^{\star} = I(X;Y)$ and does not depend on M. Therefore, Theorem 1 states that allowing the mixing coefficients, herein $1/M$ in P_{H_0}, to depend on n can lead to a non-trivial reduction in the error exponent of type II.*

III. Strong Converses

In this section we first show that the result derived in the previous section is also tight in the strong converse sense. Next, we generalize our hypothesis testing problem by studying the case where compression mappings can be different from user to user, referred to as the user-dependent compression mapping case. It is shown that the generalization does not increase the optimal error exponent. We choose to present these results separately, since the generality of the user-dependent mapping setting might overshadow several interesting interpretations of derivation steps in the fixed mapping setting. Finally, we study a similar generalized identification system where each user can have a different compression mapping.

Before we begin, let us define the notion of information density which will be used frequently in the subsequence. The information density between two random variables Z and V that are jointly distributed according to P_{ZV} is defined when $P_{ZV} \ll P_Z \times P_V$ as

$$\iota_{ZV}(z; v) = \log \frac{dP_{ZV}}{d(P_Z \times P_V)}(z, v). \tag{21}$$

When P_{ZV} is clear from the context we omit the subscript.

A. Strong converse for fixed compression mappings

Since both distributions are non-stationary, we employ the information spectrum approach on top of the result by Ahlswede and Csiszár in [1] to show the strong converse for the setting in Section II. Similar to Definition 2 we have the following definition for ϵ-achievability.

Definition 3. *Let $\epsilon \in [0, 1)$ be an arbitrarily given constant. An error exponent E of type II is ϵ-achievable given (R, R_c) if there exist compression and decision mappings $(\phi_n^{\mathrm{t}}, \psi_n^{\mathrm{t}})$ such that*

$$\limsup_{n\to\infty} \alpha_n \leq \epsilon, \ \limsup_{n\to\infty} \frac{1}{n} \log |\phi_n^{\mathrm{t}}| \leq R_c,$$
$$\liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_n} \geq E. \tag{22}$$

We define

$$E_{\epsilon,\mathrm{f}}^{\star}(R, R_c) = \sup\{E \mid E \text{ is } \epsilon\text{-achievable given } (R, R_c)\}.$$

We first present a key lemma that relates our multi-user setting to the single-user setting which enables further analysis. It says roughly that when the likelihood ratio test is considered, the probability of type I error in our multi-user setting is greater than or equal to the one of the single-user setting due to the presence of multiple users.

Lemma 1. *For any compression sequence (ϕ_n^{t}), and $\hat{E}, \gamma > 0$, we have for all sufficiently large n*

$$\begin{aligned}&\Pr\left\{\frac{1}{n}\log\frac{P_{Y^n\phi_n^{\mathrm{t}}(\mathbf{X}^n)}(Y^n,\phi_n^{\mathrm{t}}(\mathbf{X}^n))}{P_{Y^n}\times P_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}(Y^n,\phi_n^{\mathrm{t}}(\mathbf{X}^n))} > \hat{E}\right\}\\ &\le \Pr\left\{\iota_{\bar{Y}^n\phi_n^{\mathrm{t}}(\bar{X}^n)}(\bar{Y}^n;\phi_n^{\mathrm{t}}(\bar{X}^n)) > \log M + n(\hat{E}-\gamma)\right\} + \mathcal{O}(\exp(-n\hat{E})),\end{aligned} \tag{23}$$

where the left-hand side of (23) *is evaluated with $(Y^n,\mathbf{X}^n)\sim P_{H_0}$ as given in* (1)*, and $(\bar{Y}^n,\bar{X}^n)\sim P_{YX}^{\otimes n}$.*

The proof of Lemma 1 uses a soft-covering lemma from [22]. Soft-covering is usually used in *achievability* proofs. Its appearance when deriving the *strong converse* is an interesting step for us.

Proof. We denote the LHS of (23) by $L_{n,\mathrm{f}}(\hat{E},\gamma)$ and suppress the dependency on $(\hat{E},\gamma)$ in the proof for notation brevity. Given a compressed tuple $\phi_n^{\mathrm{t}}(\mathbf{x}^n)$ of users' data sequences, we define the following (conditional) distribution on $\mathcal{Y}^n$

$$\hat{P}_{H_0,\phi_n^{\mathrm{t}}(\mathbf{x}^n)}(y^n) = \frac{1}{M}\sum_{i=1}^{M} P_{Y^n|\phi_n^{\mathrm{t}}(X^n)}(y^n|\phi_n^{\mathrm{t}}(x^n(i))).$$

We observe that under hypothesis H_0 the joint distribution induced by the mapping ϕ_n^{t} can be reformulated as

$$P_{Y^n\phi_n^{\mathrm{t}}(\mathbf{X}^n)}(y^n,\phi_n^{\mathrm{t}}(\mathbf{x}^n)) = \hat{P}_{H_0,\phi_n^{\mathrm{t}}(\mathbf{x}^n)}(y^n)\times\prod_{i=1}^{M} P_{\phi_n^{\mathrm{t}}(X^n)}(\phi_n^{\mathrm{t}}(x^n(i))). \tag{24}$$

The corresponding induced joint distribution under hypothesis H_1 is given by

$$P_{Y^n}\times P_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}(y^n,\phi_n^{\mathrm{t}}(\mathbf{x}^n)) = P_{Y^n}(y^n)\times\prod_{k=1}^{M} P_{\phi_n^{\mathrm{t}}(X^n)}(\phi_n^{\mathrm{t}}(x^n(k))). \tag{25}$$

Therefore, since $(Y^n,\mathbf{X}^n)\sim P_{H_0}$ we can rewrite $L_{n,\mathrm{f}}$ as

$$\begin{aligned}L_{n,\mathrm{f}} &= \mathbb{E}_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}\left[\Pr\left\{\frac{\hat{P}_{H_0,\phi_n^{\mathrm{t}}(\mathbf{X}^n)}}{P_{Y^n}}(Y^n) > \eta\middle|\phi_n^{\mathrm{t}}(\mathbf{X}^n)\right\}\right]\\ &= \mathbb{E}_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}F_\eta(\hat{P}_{H_0,\phi_n^{\mathrm{t}}(\mathbf{X}^n)}||P_{Y^n}).\end{aligned} \tag{26}$$

Herein we have

$$F_\eta(P||Q) = \Pr\left\{\frac{dP}{dQ}(X) > \eta\right\}$$

where $X\sim P$ is the excess relative information metric with threshold η, $\eta = e^{n\hat{E}}$, as defined in[2] [22]. For each tuple $\phi_n^{\mathrm{t}}(\mathbf{x}^n)$, which is a realization of $\phi_n^{\mathrm{t}}(\mathbf{X}^n)$, we can view $\hat{P}_{H_0,\phi_n^{\mathrm{t}}(\mathbf{x}^n)}$ as the output distribution induced by selecting one sequence in the tuple uniformly at random and feeding it into the input of the channel $P_{Y^n|\phi_n^{\mathrm{t}}(X^n)}$. The soft-covering lemma for the F_η metric in [22, Theorem 24] states that

$$\begin{aligned}&\mathbb{E}_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}F_\eta(\hat{P}_{H_0,\phi_n^{\mathrm{t}}(\mathbf{X}^n)}||P_{Y^n})\\ &\le \Pr\left[\iota(\bar{Y}^n;\phi_n^{\mathrm{t}}(\bar{X}^n)) > \log(M\sigma)\right] + \frac{1}{\nu}\Pr\left[\iota(\bar{Y}^n;\phi_n^{\mathrm{t}}(\bar{X}^n)) > \log M - \tau\right] + \frac{\exp(-\tau)}{(\eta-1-\nu-\sigma)^2},\end{aligned} \tag{27}$$

where herein $\sigma,\nu>0$ are arbitrarily satisfying $\eta-1>\nu+\sigma$, $\tau\in\mathbb{R}$ and $(\bar{Y}^n,\bar{X}^n)\sim P_{YX}^{\otimes n}$. If we take, $\tau=-n\hat{E}$, $\sigma=\eta/4-1$ and $\nu=\eta/4$, then we obtain

$$\mathbb{E}_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}F_\eta(\hat{P}_{H_0,\phi_n^{\mathrm{t}}(\mathbf{X}^n)}||P_{Y^n}) \le \Pr\left\{\iota(\bar{Y}^n;\phi_n^{\mathrm{t}}(\bar{X}^n)) > \log M + n\hat{E} + \log(1/4-1/\eta))\right\} + 8\exp(-n\hat{E}). \tag{28}$$

The conclusion of the lemma follows. □

Roughly speaking, to provide a strong converse statement we aim to drive $L_{n,\mathrm{f}}$ to 0 as $n\to\infty$ which is explained in the following. For a given compression mapping sequence (ϕ_n^{t}) the inequality [21, Lemma 4.1.2]

$$\alpha_n + e^{n\hat{E}}\beta_n \ge 1 - L_{n,\mathrm{f}} \tag{29}$$

[2]The metric is denoted therein by $\bar{F}_\eta(P||Q)$. We use a slightly different notation herein, since $\bar{(\cdot)}$ has been employed to denote the single-user case.

implies that if for a given threshold $\hat{E}$, $L_{n,\mathrm{f}}$ goes to 0, then the ϵ-achievable error exponent is upper bounded by $\hat{E}$. It can be seen that if $\hat{E}$ is greater than the *spectral-sup mutual information* of the joint process $\{(Y^n, \phi_n^{\mathrm{t}}(\mathbf{X}^n))\}_{n=1}^{\infty}$ then $L_{n,\mathrm{f}}$ always goes to 0. However, the bound is hard to characterize in a single letter form. The following corollary of Lemma 1 shows that there exists a sequence $(L_{n_k,\mathrm{f}})$ which goes to 0. It will be shown later that the conclusion is sufficient for proving a strong converse statement.

Corollary 1. *Given a compression sequence (ϕ_n^{t}) such that $\limsup_{n\to\infty} \frac{1}{n}\log|\phi_n^t| \le R_c$. If $\hat{E} = R_{\max} - R + 3\gamma$ where $R \le R_{\max}$ and $\gamma > 0$ is arbitrary, then there exists a subsequence $(n_k)_{k=1}^{\infty}$ such that*

$$\lim_{k\to\infty} L_{n_k,\mathrm{f}}(\hat{E},\gamma) = 0. \tag{30}$$

Proof. It suffices to show that there exists a subsequence (n_k) such that the first term in the RHS of (23) converges to 0. Define the following acceptance region for the single-user hypothesis testing problem

$$\bar{\mathcal{A}}_n = \left\{(y^n, \phi_n^{\mathrm{t}}(x^n)) \mid \iota(y^n;\phi_n^{\mathrm{t}}(x^n)) > n\tilde{E}\right\}, \tag{31}$$

where $\tilde{E} = R_{\max} + \gamma$. Then it can be seen that

$$\begin{aligned}
\bar{\beta}_n &= P_{\bar{Y}^n} \times P_{\phi_n^{\mathrm{t}}(\bar{X}^n)}(\bar{\mathcal{A}}_n) \\
&= \sum_{(y^n,\phi_n^{\mathrm{t}}(x^n))\in\bar{\mathcal{A}}_n} P_{\bar{Y}^n}(y^n) P_{\phi_n^{\mathrm{t}}(\bar{X}^n)}(\phi_n^{\mathrm{t}}(x^n)) \\
&\le e^{-n\tilde{E}} \sum_{(y^n,\phi_n^{\mathrm{t}}(x^n))\in\bar{\mathcal{A}}_n} P_{\bar{Y}^n\phi_n^{\mathrm{t}}(\bar{X}^n)}(y^n,\phi_n^{\mathrm{t}}(x^n)) \\
&\le e^{-n\tilde{E}},
\end{aligned} \tag{32}$$

thus $\liminf_{n\to\infty} \frac{1}{n}\log\frac{1}{\bar{\beta}_n} \ge \tilde{E}$. Since $\tilde{E} > R_{\max}$, by the strong converse result of Ahlswede and Csiszar [1, Theorem 3] we have

$$\limsup_{n\to\infty} \bar{\alpha}_n = 1, \text{ where } \bar{\alpha}_n = P_{\bar{Y}^n\phi_n^{\mathrm{t}}(\bar{X}^n)}(\bar{\mathcal{A}}_n^c). \tag{33}$$

Then there exists a subsequence $(n_k)_{k=1}^{\infty}$ such that

$$\lim_{k\to\infty} \bar{\alpha}_{n_k} = 1 \Leftrightarrow \lim_{k\to\infty} P_{\bar{Y}^{n_k}\phi_{n_k}^{\mathrm{t}}(\bar{X}^{n_k})}(\bar{\mathcal{A}}_{n_k}) = 0. \tag{34}$$

Additionally, for all sufficiently large n we have

$$\hat{E} - \gamma + \frac{1}{n}\log M > \tilde{E}$$

which implies further that

$$P_{\bar{Y}^n\phi_n^{\mathrm{t}}(\bar{X}^n)}(\bar{\mathcal{A}}_n) \ge \Pr\{\iota(\bar{Y}^n;\phi_n^{\mathrm{t}}(\bar{X}^n)) > \log M + n(\hat{E}-\gamma)\}.$$

By Lemma 1 and (34) we then obtain

$$\lim_{k\to\infty} L_{n_k,\mathrm{f}}(\hat{E},\gamma) = 0. \tag{35}$$

□

We now summarize the above analysis in the following theorem, which is the strong converse statement for the setting introduced in Section II.

Theorem 2. *For $E_{\epsilon,\mathrm{f}}^{\star}(R,R_c)$ defined in Definition 3 we have for all $\epsilon,\ 0 \le \epsilon < 1$,*

$$E_{\epsilon,\mathrm{f}}^{\star}(R,R_c) = \theta(R,R_c) = R_{\max} - R, \text{ if } R \le R_{\max}. \tag{36}$$

Proof. Suppose that there exists a sequence of compression mappings (ϕ_n^{t}) and a sequence of decision mappings (ψ_n^{t}) such that

$$\begin{aligned}
\limsup_{n\to\infty} \alpha_n \le \epsilon,\ \limsup_{n\to\infty} \frac{1}{n}\log|\phi_n^{\mathrm{t}}| \le R_c, \\
\text{and } \liminf_{n\to\infty} \frac{1}{n}\log\frac{1}{\beta_n} \ge E.
\end{aligned} \tag{37}$$

Given $\gamma > 0$ small enough such that $\epsilon + \gamma < 1$, then for all sufficiently large n we have $\alpha_n \leq \epsilon + \gamma$ and $\beta_n \leq e^{-n(E-\gamma)}$. Select $\hat{E} = R_{\max} - R + 3\gamma$ and let (n_k) be the corresponding subsequence such that (30) holds. Then as [21, Lemma 4.1.2], c.f. also [23, Section 13.1], we have

$$1 - \alpha_n - e^{n\hat{E}}\beta_n \leq \Pr\left\{\frac{1}{n}\log\frac{P_{Y^n\phi_n^{\mathrm{t}}(\mathbf{X}^n)}}{P_{Y^n} \times P_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}}(Y^n, \phi_n^{\mathrm{t}}(\mathbf{X}^n)) > \hat{E}\right\},$$

holds for any n. Since, the RHS is actually $L_{n,\mathrm{f}}(\hat{E}, \gamma)$, we obtain for all sufficiently large n_k the following

$$\begin{aligned} 1 - \epsilon - \gamma - e^{n_k\hat{E}}e^{-n_k(E-\gamma)} &\leq L_{n_k,\mathrm{f}} \\ \implies \hat{E} - E + \gamma &\geq \frac{1}{n_k}\log(1 - \epsilon - \gamma - L_{n_k,\mathrm{f}}) \\ \implies R_{\max} - R + 4\gamma \geq E &\overset{\gamma\to 0}{\implies} R_{\max} - R \geq E \\ \implies R_{\max} - R &\geq E^{\star}_{\epsilon,\mathrm{f}}(R, R_c). \end{aligned} \tag{38}$$

The last inequality holds since E is an arbitrary ϵ-achievable exponent. The conclusion follows since $R_{\max} - R \leq E^{\star}_{\epsilon}(R, R_c)$ when $R < R_{\max}$ by Theorem 1. Note that the upper bound still holds even if we define $E^{\star}_{\epsilon}(R, R_c)$ on the closure of ϵ-achievable region of (R_c, E). □

In the next theorem we provide some partial information about the behavior of errors of type I and II when the number of users exceeds the identifiable threshold, i.e., $R > R_{\max}$. The result further implies that $E^{\star}_{\epsilon,\mathrm{f}}(R, R_c) = 0$ for all $\epsilon \in [0, 1)$ when $R > R_{\max}$.

Theorem 3. *Given a sequence of compression mappings (ϕ_n^{t}) such that $\limsup_{n\to\infty}\frac{1}{n}\log|\phi_n^{\mathrm{t}}| \leq R_c$. Consider the case $R = R_{\max}(R_c) + \gamma$ where $\gamma > 0$ is arbitrary, then for any sequence of decision mappings (ψ_n^{t}), the following holds*

$$\limsup_{n\to\infty}(\alpha_n + \beta_n) \geq 1. \tag{39}$$

Moreover, in case of no compression, i.e., correspondingly $R = I(X;Y) + \gamma$, we obtain $\lim_{n\to\infty}(\alpha_n + \beta_n) = 1$.

We interpret the result of Theorem 3 via the receiver operating characteristic curve as follows. In the design process, one aims to attain the highest detection probability for a given false alarm level. Theorem 2 states that the detection probability can be driven to 1 as long as the number of users is below $R_{\max}(R_c)$ and the false alarm level is strictly below 1. However, Theorem 3 says that when the number of users in the system exceeds $R_{\max}(R_c)$, the performance of *any* decision rule is not better than a random guess.

Proof. From (24) and (25), the variational distance between $P_{Y^n\phi_n^{\mathrm{t}}(\mathbf{X}^n)}$ and $P_{Y^n} \times P_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}$ is given by

$$\|P_{Y^n,\phi_n^{\mathrm{t}}(\mathbf{X}^n)} - P_{Y^n} \times P_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}\|_{TV} = \mathbb{E}_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}\big\|\hat{P}_{H_0,\phi_n^{\mathrm{t}}(\mathbf{X}^n)} - P_{Y^n}\big\|_{TV}.$$

By the soft-covering lemma [24, Corollary VII.2], [25, Lemma 2], we obtain that

$$\|P_{Y^n,\phi_n^{\mathrm{t}}(\mathbf{X}^n)} - P_{Y^n} \times P_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}\|_{TV} \leq \Pr\big\{\iota(\bar{Y}^n; \phi_n^{\mathrm{t}}(\bar{X}^n)) > n(R_{\max} + \frac{\gamma}{2})\big\} + \frac{1}{2}\sqrt{\frac{e^{n(R_{\max}+\gamma/2)}}{M}}, \tag{40}$$

where again $(\bar{Y}^n, \bar{X}^n) \sim P_{XY}^{\otimes n}$. From the definition of the total variational distance we obtain

$$|1 - \alpha_n - \beta_n| \leq \sup_{\mathcal{A}}|P_{Y^n,\phi_n^{\mathrm{t}}(\mathbf{X}^n)}(\mathcal{A}) - P_{Y^n} \times P_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}(\mathcal{A})| = \|P_{Y^n,\phi_n^{\mathrm{t}}(\mathbf{X}^n)} - P_{Y^n} \times P_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}\|_{TV}. \tag{41}$$

Let $\bar{\mathcal{A}}_n$ be defined as in (31) with $\tilde{E} = R_{\max} + \gamma/2$ instead. Then by using (40) we obtain[3]

$$\liminf_{n\to\infty}|1 - (\alpha_n + \beta_n)| \leq \liminf_{n\to\infty}\|P_{Y^n,\phi_n^{\mathrm{t}}(\mathbf{X}^n)} - P_{Y^n} \times P_{\phi_n^{\mathrm{t}}(\mathbf{X}^n)}\|_{TV} \leq \liminf_{n\to\infty} P_{\bar{Y}^n\phi_n^{\mathrm{t}}(\bar{X}^n)}(\bar{\mathcal{A}}_n) \overset{(34)}{=} 0, \tag{42}$$

which implies $\limsup_{n\to\infty}(\alpha_n+\beta_n) \geq 1$. In case of no compression we replace the $\liminf_{n\to\infty}(\cdot)$ operation by the $\lim_{n\to\infty}(\cdot)$ operation and use the weak law of large numbers in the last step. □

Combining the results of Theorem 2 and Theorem 3 we obtain $E^{\star}_{\epsilon}(R, R_c) = \max\{R_{\max}(R_c) - R, 0\}$. We establish in the following a reverse statement of Theorem 1.

Proposition 1. *Fix $\hat{E}$ and $\gamma > 0$. Given a multi-user testing scheme $(\phi_n^{\mathrm{t}}, \psi_n^{\mathrm{t}})$ with probabilities of errors (α_n, β_n) we can construct a single-user testing scheme $(\phi_n^{\mathrm{t}}, \psi_n')$ such that the corresponding probabilities of errors are given by*

$$\bar{\alpha}_n \leq \alpha_n + e^{n\hat{E}}\beta_n + \mathcal{O}(\exp(-n\hat{E}))$$

[3]Since $R > R_{\max}$ holds, the last equality in (42) can also be shown by first using (27) and with an arbitrary $\eta > 1$, $\tau = R - R_{\max} > 0$ and suitable ν and σ. Then we can take the limits (in the order) $\liminf_{n\to\infty}$ and $\eta \to 1$. The reader is referred to Remark 25 in [22] for the details.

$$\bar{\beta}_n \leq e^{-n(\hat{E}-\gamma)} \frac{1}{M}, \tag{43}$$

for all sufficiently large n.

Proof. Define a single-user decision region as

$$\bar{\mathcal{A}}_n = \big\{(y^n, \phi_n^{\mathrm{t}}(x^n)) \mid \iota_{\bar{Y}^n \phi_n^{\mathrm{t}}(\bar{X}^n)}(y^n; \phi_n^t(x^n)) > \log M + n(\hat{E}-\gamma)\big\}. \tag{44}$$

By Lemma 1 we obtain

$$\alpha_n + e^{n\hat{E}}\beta_n \geq 1 - L_n(\hat{E}, \gamma) \geq P_{\bar{Y}^n \phi_n^{\mathrm{t}}(\bar{X}^n)}(\bar{\mathcal{A}}_n^c) - 8\exp(-n\hat{E}) = \bar{\alpha}_n - 8\exp(-n\hat{E}). \tag{45}$$

Similar to (32), by the change of measure we also have

$$\bar{\beta}_n \leq \frac{e^{-n(\hat{E}-\gamma)}}{M}. \tag{46}$$

□

As a consequence of Theorem 1 and Proposition 1 we have the following theorem. It shows an equivalence for the exponentially strong converse between our multi-user testing problem and the single-user testing against independence problem.

Theorem 4. *Assume that $R < R_{\max}(R_c)$ then the two following statements are equivalent.*

1) *For any single-user HT scheme $(\bar{\phi}_n, \bar{\psi}_n)$ with $\limsup_{n\to\infty} \frac{1}{n}\log|\bar{\phi}_n^{\mathrm{t}}| \leq R_c$ and $\liminf_{n\to\infty} \frac{1}{n}\log\frac{1}{\bar{\beta}_n} \geq \bar{E}$, if $\bar{E} > R_{\max}(R_c)$, then $\bar{\alpha}_n \to 1$ exponentially fast at a positive convergence rate.*
2) *For any multi-user HT scheme $(\phi_n^{\mathrm{t}}, \psi_n^{\mathrm{t}})$ with $\limsup_{n\to\infty} \frac{1}{n}\log|\phi_n^{\mathrm{t}}| \leq R_c$ and $\liminf_{n\to\infty} \frac{1}{n}\log\frac{1}{\bar{\beta}_n} \geq E$, if $E > R_{\max}(R_c) - R$, then $\alpha_n \to 1$ exponentially fast at a positive convergence rate.*

Proof. Assume that the first statement holds. Let $(\phi_n^{\mathrm{t}}, \psi_n^{\mathrm{t}})$ be an arbitrary multi-user testing scheme such that

$$\limsup_{n\to\infty} \frac{1}{n}\log|\phi_n^{\mathrm{t}}| \leq R_c, \text{ and } \liminf_{n\to\infty} \frac{1}{n}\log\frac{1}{\beta_n} \geq E,$$

holds where $E > R_{\max}(R_c) - R$. Select $\hat{E} = E - 2\gamma$ for $\gamma > 0$ small enough such that $E + R - 4\gamma > R_{\max}(R_c)$. By applying Proposition 1 we obtain a single-user hypothesis testing scheme $(\phi_n^{\mathrm{t}}, \psi_n')$ such that for all sufficiently large n the corresponding single-user false alarm and miss detection probabilities $(\bar{\alpha}_n, \bar{\beta}_n)$ are bounded by

$$\begin{aligned} \bar{\alpha}_n &\leq \alpha_n + e^{-n\gamma} + C(\exp(-n(E-2\gamma)) \\ \bar{\beta}_n &\leq \exp(-n(E+R-4\gamma)), \end{aligned} \tag{47}$$

where C is an absolute constant. Since $E + R - 4\gamma > R_{\max}(R_c)$ holds, the assumption implies that $\bar{\alpha}_n \to 1$ exponentially at a positive rate. Hence $\alpha_n \to 1$ exponentially at a positive rate.

Assume now that the second statement holds. Let $(\bar{\phi}_n, \bar{\psi}_n)$ be an arbitrary single-user hypothesis testing scheme such that

$$\limsup_{n\to\infty} \frac{1}{n}\log|\bar{\phi}_n^{\mathrm{t}}| \leq R_c \text{ and } \liminf_{n\to\infty} \frac{1}{n}\log\frac{1}{\bar{\beta}_n} \geq \bar{E},$$

where $\bar{E} > R_{\max}(R_c)$. Select an arbitrary $\gamma > 0$ small enough such that $\bar{E} - 2\gamma > R_{\max}(R_c)$. Then for all sufficiently large n we obtain by applying Theorem 1 a multi-user testing scheme such that the corresponding multi-user false alarm and miss detection probabilities (α_n, β_n) are upper bounded by

$$\alpha_n \leq \bar{\alpha}_n, \qquad \beta_n \leq e^{-n(\bar{E}-2\gamma-R)}. \tag{48}$$

Since $\bar{E} - R - 2\gamma > R_{\max}(R_c) - R$ the assumption implies that α_n tends to 1 exponentially fast at a positive rate. Hence $\bar{\alpha}_n$ goes to 1 exponentially fast a positive rate. □

In Section IV-A we connect the first statement in Theorem 4 with a known result for the Wyner-Ahlswede-Körner network in [7], [8].

In this paragraph, we present a general result which is a consequence of Theorem 1 and Proposition 1. Suppose that $\mathcal{X}$ and $\mathcal{Y}$ have sufficient structure, for example being Polish spaces. We assume further that in (1) we have $P_{Y^n X^{n(i)}} = P_{Y^n X^n}$ which is not necessarily iid nor discrete. We keep using Definition 3 for the ϵ-achievability for the multi-user testing problem and the supremum ϵ-achievable error exponent of type II, $E_{\epsilon,\mathrm{f}}^{\star}(R, R_c)$. Furthermore, for $M = 1$ we use the notation $E_{\epsilon,\mathrm{s}}^{\star}(R_c)$ to denote the supremum ϵ-achievable error exponent for the single-user case. It can be seen that for a given compression rate R_c, $E_{\epsilon,\mathrm{s}}^{\star}(R_c)$ is finite for similar reasons as in Remark 8 in Section IV-B. Then we have the following theorem.

Theorem 5. *When $\{P_{Y^n X^n}\}_{n=1}^{\infty}$ is a sequence of general distributions, we have for all $\epsilon \in [0, 1)$ and all $(R, R_c) \in \mathbb{R}_+^2$,*

$$E_{\epsilon,\mathrm{f}}^{\star}(R, R_c) = \max\{E_{\epsilon,\mathrm{s}}^{\star}(R_c) - R, 0\}. \tag{49}$$

Proof. Assume that $R < E^{\star}_{\epsilon,\mathrm{s}}(R_c)$ holds. By Theorem 1 we have $E^{\star}_{\epsilon,\mathrm{f}}(R,R_c) \geq E^{\star}_{\epsilon,\mathrm{s}}(R_c) - R$. Therefore for all (R,R_c) we have

$$E^{\star}_{\epsilon,\mathrm{f}}(R,R_c) \geq \max\{E^{\star}_{\epsilon,\mathrm{s}}(R_c) - R, 0\}. \tag{50}$$

Assume for now that E is an ϵ-achievable error exponent of type II for the multi-user testing problem, then there exist a pair of sequences $(\phi^{\mathrm{t}}_n, \psi^{\mathrm{t}}_n)$ such that all the conditions in Definition 3 are fulfilled. Applying Proposition 1 with $\hat{E} = E - \gamma$ we obtain a single-user testing scheme $(\phi^{\mathrm{t}}_n, \psi'_n)$ such that

$$\bar{\alpha}_n \leq \alpha_n + e^{-n\gamma} + C\exp(-n(E-\gamma)),\ \bar{\beta}_n \leq \exp(-n(E-3\gamma+R)), \tag{51}$$

for all sufficiently large n. Consider first the case $R < E^{\star}_{\epsilon,\mathrm{s}}(R_c)$. Assume further that $E > 0$, which is plausible if $E^{\star}_{\epsilon,\mathrm{f}}(R,R_c) > 0$, then with $\gamma > 0$ sufficiently small we obtain

$$\limsup_{n\to\infty} \bar{\alpha}_n \leq \epsilon,\ \liminf_{n\to\infty} \frac{1}{n}\log\frac{1}{\bar{\beta}_n} \geq E - 3\gamma + R. \tag{52}$$

Therefore we have $E - 3\gamma + R \leq E^{\star}_{\epsilon,\mathrm{s}}(R_c)$. Accordingly, we must have $E^{\star}_{\epsilon,\mathrm{f}}(R,R_c) \leq E^{\star}_{\epsilon,\mathrm{s}}(R_c) - R$ for this case.

Next, consider the case that $R \geq E^{\star}_{\epsilon,\mathrm{s}}(R_c)$. Assume that $E^{\star}_{\epsilon,\mathrm{f}}(R,R_c) > 0$ still holds. Then again we have $E - 3\gamma + R \leq E^{\star}_{\epsilon,\mathrm{s}}(R_c)$. By taking $\gamma \to 0$ we see that the last inequality is not valid for $0 < E < E^{\star}_{\epsilon,\mathrm{f}}(R,R_c)$. Therefore $E^{\star}_{\epsilon,\mathrm{f}}(R,R_c) = 0$ in this case. □

B. Hypothesis Testing with user-dependent compression mappings

We now consider the case that each user has its own compression mapping ϕ^{t}_{kn} for $k \in [1:M]$. This setting for instance represents the case where the hypothesis testing database $(\phi^{\mathrm{t}}_{kn}(X^n(k)))$ is a merger of different, distributed sub-databases, which employ different compression mechanisms. For simplicity we denote by $\boldsymbol{\phi}^{\mathrm{t}}_n(\mathbf{X}^n)$ the ensemble $(\phi^{\mathrm{t}}_{1n}(X^n(1)), \ldots, \phi^{\mathrm{t}}_{Mn}(X^n(M)))$. The corresponding realization is denoted by $\boldsymbol{\phi}^{\mathrm{t}}_n(\mathbf{x}^n)$. These mappings induce the following distributions for our hypothesis testing problem.

$$\begin{aligned} H_0\colon P_{Y^n\boldsymbol{\phi}^{\mathrm{t}}_n(\mathbf{X}^n)} &= \frac{1}{M}\sum_{k=1}^{M} P_{Y^n\phi^{\mathrm{t}}_{kn}(X^n(k))} \times \prod_{j\neq k} P_{\phi^{\mathrm{t}}_{jn}(X^n(j))} \\ H_1\colon P_{Y^n} \times P_{\boldsymbol{\phi}^{\mathrm{t}}_n(\mathbf{X}^n)} &= P_{Y^n} \times \prod_{k=1}^{M} P_{\phi^{\mathrm{t}}_{kn}(X^n(k))}. \end{aligned} \tag{53}$$

The decision region is given similarly as

$$\mathcal{A}_{n,\mathrm{var}} = \{(y^n, \boldsymbol{\phi}^{\mathrm{t}}_n(\mathbf{x}^n)) \mid \psi^{\mathrm{t}}_n(y^n, \boldsymbol{\phi}^{\mathrm{t}}_n(\mathbf{x}^n)) = 0\}. \tag{54}$$

The corresponding false alarm and miss detection probabilities are given by

$$\alpha_{n,\mathrm{var}} = P_{Y^n\boldsymbol{\phi}^{\mathrm{t}}_n(\mathbf{X}^n)}(\mathcal{A}^c_{n,\mathrm{var}}),\ \beta_{n,\mathrm{var}} = P_{Y^n} \times P_{\boldsymbol{\phi}^{\mathrm{t}}_n(\mathbf{X}^n)}(\mathcal{A}_{n,\mathrm{var}}). \tag{55}$$

To enable the characterization in this setting we assume a *uniformity condition* that all the mappings ϕ^{t}_{kn} admit the same compression rate. We state in the following the corresponding ϵ-achievable definition.

Definition 4. *An error exponent E of type II is ϵ-achievable for a given pair (R,R_c) if there exist a doubly indexed*[4] *sequence (ϕ^{t}_{kn}) of enrollment mappings, and a sequence of decision mappings (ψ^{t}_n) such that*

$$\begin{aligned} \limsup_{n\to\infty} \max_{k\in[1:M]} \frac{1}{n}\log|\phi^t_{kn}| &\leq R_c, \\ \limsup_{n\to\infty} \alpha_{n,\mathrm{var}} \leq \epsilon,\ \liminf_{n\to\infty} \frac{1}{n}\log\frac{1}{\beta_{n,\mathrm{var}}} &\geq E. \end{aligned} \tag{56}$$

Let $E^{\star}_{\epsilon,\mathrm{var}}(R,R_c)$ be the supremum of all ϵ-achievable error exponent for a given pair (R,R_c).

In the following we proceed to show the strong converse for this user-dependent setting. The proof idea is similar to the one in the previous section. Namely, we reduce the likelihood ratio in the multi-user setting to the one in the single-user setting and carefully control the minimum false alarm probability in the single-user setting when multiple compression mappings are present. For this purpose, we need to derive some generalized lemmas, since the tools used in the previous section are no longer applicable. We first need the following lemma which is a fix for the soft covering using excess relative information metric in [22, Theorem 24] .

[4]Note that in our case for a given n only $M \doteq e^{nR}$ mappings ϕ^{t}_{kn} are active for compression, i.e., k does not run freely to ∞.

Lemma 2. *For each $i \in [1:M]$ define the following distribution $Q_i = P_{ZV_i} \times \prod_{j\neq i} P_{V_j}$. Furthermore, for each i let $(Z^i, V_1^i, \ldots, V_M^i)$ be a tuple of random variables that follows the law Q_i. For brevity we denote an outcome in the corresponding sample space $(z, v_1, \ldots, v_m)$ by $(z, \boldsymbol{v})$. Then we define two distributions of interest as follows*

$$Q_{H_0} = \frac{1}{M}\sum_{i=1}^{M} Q_i \text{ and } Q_{H_1} = P_Z \times \prod_{k=1}^{M} P_{V_k}. \tag{57}$$

Assume that for all $k \in [1:M]$, $P_{ZV_k} \ll P_Z \times P_{V_k}$ holds. Given positive numbers η, σ such that $\eta > \sigma$ we have

$$Q_{H_0}\left(\left\{(z,\boldsymbol{v}) \mid \frac{dQ_{H_0}}{dQ_{H_1}}(z,\boldsymbol{v}) > \eta\right\}\right) \leq \frac{1}{M}\sum_{k=1}^{M} \Pr\left\{\frac{dP_{ZV_k}}{d(P_Z \times P_{V_k})}(Z^k, V_k^k) > M\sigma\right\} + \frac{1}{\eta-\sigma}. \tag{58}$$

For a given tuple $\boldsymbol{v}$, the distribution $\frac{1}{M}\sum_{i=1}^{M} P_{Z|V_i=v_i}$ can be seen as the output distribution of a general channel that depends on which input is picked. Lemma 2 therefore can still be seen as a soft-covering lemma. The proof is given in the following.

Proof. First we have

$$Q_{H_0}\left(\left\{(z,\boldsymbol{v}) \mid \frac{dQ_{H_0}}{dQ_{H_1}}(z,\boldsymbol{v}) > \eta\right\}\right) = \frac{1}{M}\sum_{i=1}^{M} Q_i\left(\left\{(z,\boldsymbol{v}) \mid \frac{dQ_{H_0}}{dQ_{H_1}}(z,\boldsymbol{v}) > \eta\right\}\right). \tag{59}$$

Since

$$\frac{dQ_{H_0}}{dQ_{H_1}}(z,\boldsymbol{v}) = \frac{1}{M}\sum_{i=1}^{M} \frac{dP_{ZV_i}}{d(P_Z \times P_{V_i})}(z, v_i),\ Q_{H_1}\text{-a.s.} \tag{60}$$

We have

$$\begin{aligned} Q_i\left(\left\{(z,\boldsymbol{v}) \mid \frac{dQ_{H_0}}{dQ_{H_1}}(z,\boldsymbol{v}) > \eta\right\}\right) &= Q_i\left(\left\{(z,\boldsymbol{v}) \mid \frac{1}{M}\sum_{i=k}^{M} \frac{dP_{ZV_k}}{d(P_Z \times P_{V_k})}(z, v_k) > \eta\right\}\right) \\ &\leq \Pr\left\{\frac{dP_{ZV_i}}{d(P_Z \times P_{V_i})}(Z^i, V_i^i) > M\sigma\right\} + \Pr\left\{\frac{1}{M}\sum_{k\neq i} \frac{dP_{ZV_k}}{d(P_Z \times P_{V_k})}(Z^i, V_k^i) > \eta - \sigma\right\}. \end{aligned} \tag{61}$$

Since for $k \neq i$ we have $(Z^i, V_k^i) \sim P_Z \times P_{V_k}$, the following equality is valid

$$\mathbb{E}\left[\frac{dP_{ZV_k}}{d(P_Z \times P_{V_k})}(Z^i, V_k^i)\right] = 1. \tag{62}$$

The second term can be upper bounded by Markov's inequality as

$$\Pr\left\{\frac{1}{M}\sum_{k\neq i} \frac{dP_{ZV_k}}{d(P_Z \times P_{V_k})}(Z^i, V_k^i) > \eta - \sigma\right\} \leq \frac{1}{\eta-\sigma}\frac{1}{M}\sum_{k\neq i} \mathbb{E}\left[\frac{dP_{ZV_k}}{d(P_Z \times P_{V_k})}(Z^i, V_k^i)\right] \leq \frac{1}{\eta-\sigma}. \tag{63}$$

Hence everything is done. □

Remark 3. *By closely following the proof of [22, Theorem 24] the inequality stated in Lemma 2 can be generalized into*

$$\begin{aligned} Q_{H_0}\left(\left\{(z,\boldsymbol{v}) \mid \frac{dQ_{H_0}}{dQ_{H_1}}(z,\boldsymbol{v}) > \eta\right\}\right) \leq \frac{1}{M}\sum_{k=1}^{M} \Bigg[&\Pr\left\{\frac{dP_{ZV_k}}{d(P_Z \times P_{V_k})}(Z^k, V_k^k) > M\sigma\right\} \\ &+ \frac{1}{\nu}\Pr\left\{\frac{dP_{ZV_k}}{d(P_Z \times P_{V_k})}(Z^k, V^k) \geq M\exp(-\tau)\right\}\Bigg] + \frac{\exp(-\tau)}{(\eta-\sigma-\nu-1)^2}, \end{aligned} \tag{64}$$

where $\nu > 0$ such that $\eta - \sigma - \nu - 1 > 0$ and $\tau \in \mathbb{R}$ is arbitrary. This generalization is more than what is needed in our setting, hence we choose to present in Lemma 2 a simplified version instead.

Lemma 2 implies that for a given doubly indexed sequence (ϕ_{kn}^{t}) and a given positive pair $(\hat{E}, \gamma)$ the following holds

$$\begin{aligned} &\Pr\left\{\iota_{Y^n\boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)}(Y^n; \boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)) > n\hat{E}\right\} \\ &\quad\leq \frac{1}{M}\sum_{k=1}^{M} \Pr\{\iota_{\bar{Y}^n\phi_{kn}^{\mathrm{t}}(\bar{X}^n)}(\bar{Y}^n; \phi_{kn}^{\mathrm{t}}(\bar{X}^n)) > \log M + n(\hat{E}-\gamma)\} + \mathcal{O}(\exp(-n\hat{E})) \end{aligned} \tag{65}$$

for all sufficiently large n. Similarly, we denote the left-hand side in (65) as $L_{n,\mathrm{var}}$.

Next, we need a *maximal lemma* given in the following. This lemma says that with the uniformity condition given in Definition 4 the minimum false alarm probability is still going to 1. It is a generalization of the arguments in the proof of Corollary 1.

Lemma 3. *Suppose that $N(n)$ is an arbitrary sequence of natural numbers. Let (ϕ_{kn}^{t}) be a doubly indexed sequence of compression mappings on $\mathcal{X}^n$ such that $\limsup_{n\to\infty}\max_{k\in[1:N]}\frac{1}{n}\log|\phi_{kn}^{\mathrm{t}}|\leq R_c$ holds. Further let $\tilde{E}=R_{\max}(R_c)+\gamma$ where $\gamma>0$ is arbitrary, then*[5]

$$\liminf_{n\to\infty}\max_{k\in[1:N]}P_{XY}^{\otimes n}\big\{(x^n,y^n)\mid \iota_{\bar{Y}^n\phi_{kn}^{\mathrm{t}}(\bar{X}^n)}(y^n;\phi_{kn}^{\mathrm{t}}(x^n))>n\tilde{E}\big\}=0 \tag{66}$$

Proof. For each n and $k\in[1:N]$, define the following sets

$$\bar{\mathcal{A}}_{kn}=\{(y^n,x^n)\mid i_{\bar{Y}^n\phi_{kn}^{\mathrm{t}}(\bar{X}^n)}(y^n;\phi_{kn}^{\mathrm{t}}(x^n))>n\tilde{E}\}. \tag{67}$$

Furthermore let $k^\star(n)$ be such that

$$P_{YX}^{\otimes n}(\bar{\mathcal{A}}_{k^\star n}^c)=\min_{k\in[1:N]}P_{YX}^{\otimes n}(\bar{\mathcal{A}}_{kn}^c) \tag{68}$$

We define a sequence of compression mappings (ϕ_n^{t}) for the single-user hypothesis testing against independence problem by $\phi_n^{\mathrm{t}}=\phi_{k^\star n}^{\mathrm{t}}$, $n=1,2,\ldots$. Let also denote the sequence of acceptance regions $(\bar{\mathcal{A}}_n)$ as $\bar{\mathcal{A}}_n=\bar{\mathcal{A}}_{k^\star n}$. Then it can be seen that

$$\limsup_{n\to\infty}\frac{1}{n}\log|\phi_n^{\mathrm{t}}|=\limsup_{n\to\infty}\frac{1}{n}\log|\phi_{k^\star n}^{\mathrm{t}}|\leq\limsup_{n\to\infty}\max_{k\in[1:N]}\frac{1}{n}\log|\phi_{kn}^{\mathrm{t}}|\leq R_c. \tag{69}$$

Furthermore from the definition of $k^\star$ we also have

$$\bar{\beta}_n=P_Y^{\otimes n}\times P_X^{\otimes n}(\bar{\mathcal{A}}_n)=P_Y^{\otimes n}\times P_X^{\otimes n}(\bar{\mathcal{A}}_{k^\star n})\leq e^{-n\tilde{E}}. \tag{70}$$

By the strong converse result in [1] we have

$$\limsup_{n\to\infty}\bar{\alpha}_n=1,\text{where }\bar{\alpha}_n=P_{YX}^{\otimes n}(\bar{\mathcal{A}}_n^c). \tag{71}$$

Therefore we have

$$\liminf_{n\to\infty}\max_{k\in[1:N]}P_{YX}^{\otimes n}(\bar{\mathcal{A}}_{kn})=0. \tag{72}$$

□

Having Lemma 2 and Lemma 3, we obtain a similar consequence as Corollary 1 which will be used in the proof of the strong converse for the current setting. For a fixed $\gamma>0$, choosing $\hat{E}=\tilde{E}-R+3\gamma$ in Lemma 2 where $R\leq R_{\max}(R_c)$ and applying Lemma 3 with $N=M$, we again have

$$\liminf_{n\to\infty}L_{n,\text{var}}=0, \tag{73}$$

where $L_{n,\text{var}}$ denotes the left-hand side of (65). Accordingly, we have the following theorem.

Theorem 6. *For $E_{\epsilon,\text{var}}^\star(R,R_c)$ defined as in Definition 4 and for $R\leq R_{\max}(R_c)$ we have $E_{\epsilon,\text{var}}^\star(R,R_c)=R_{\max}(R_c)-R$ for all $\epsilon\in[0,1)$.*

Proof. Let E be an ϵ-achievable error exponent for a given pair (R,R_c). Let (ϕ_{kn}^{t}) and (ψ_n^{t}) be sequence of mappings such that all the conditions in Definition 4 are fulfilled. Similarly as in the proof of Theorem 2 we select $\hat{E}=R_{\max}-R+3\gamma$ where $\gamma>0$ is arbitrary but small enough. Then we have

$$1-\epsilon-\gamma-e^{n_k\hat{E}}e^{-n_k(E-\gamma)}\leq L_{n_k,\text{var}} \tag{74}$$

for all sufficiently large n_k where (n_k) is a subsequence such that $(L_{n_k,\text{var}})$ converges to $\liminf_{n\to\infty}L_{n,\text{var}}=0$. Therefore we obtain

$$R_{\max}(R_c)-R+4\gamma-E\geq 0. \tag{75}$$

Since γ is arbitrary, we have $R_{\max}(R_c)-R\geq E$, which implies that $R_{\max}-R\geq E_{\epsilon,\text{var}}^\star(R,R_c)$. The other direction follows from Theorem 1. □

The following theorem shows that the error probabilities behave similarly as in Theorem 3 when $R>R_{\max}(R_c)$.

[5]Note that Theorem 10 in Section IV-A indicates that the $\liminf$ operation can be replaced by $\lim$.

Theorem 7. *Assume that $R = R_{\max}(R_c) + \gamma$ where $\gamma > 0$ is arbitrary but given. Given any doubly indexed sequence (ϕ_{kn}^{t}) such that $\limsup_{n\to\infty} \max_{k\in[1:M]} \frac{1}{n} \log |\phi_{kn}^{\mathrm{t}}| \le R_c$ holds. Then for any sequence of decision mappings (ψ_n^{t}) we have*

$$\limsup_{n\to\infty} (\alpha_{n,\mathrm{var}} + \beta_{n,\mathrm{var}}) \ge 1. \tag{76}$$

Proof. By [24, Theorem VII.1] we obtain that

$$\|P_{Y^n \boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)} - P_{Y^n} \times P_{\boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)}\|_{TV} \le \frac{1}{M} \sum_{k=1}^{M} P_{XY}^{\otimes n} \left(\left\{ (y^n, x^n) \mid \iota_{\bar{Y}^n \phi_{kn}^{\mathrm{t}}(\bar{X}^n)}(y^n; \phi_{kn}^{\mathrm{t}}(x^n)) > \log M + \tau \right\} \right) + e^{\tau/2}/2. \tag{77}$$

Pick $\tau = -n\gamma$ where $\gamma > 0$, then applying Lemma 3 with $N = M$ we have

$$\liminf_{n\to\infty} \|P_{Y^n \boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)} - P_{Y^n} \times P_{\boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)}\|_{TV} = 0. \tag{78}$$

Hence the conclusion follows. □

Remark 4. *Assume that instead of using a deterministic mapping ψ_n^{t} to decide whether H_0 or H_1 is true, we use a probabilistic mapping $P_{H_0|y^n,\boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{x}^n)} \mapsto [0,1]$ to represent the probability that H_0 is chosen given $(y^n, \boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{x}^n))$. Then the false alarm and miss-detection probabilities are given by*

$$\begin{aligned} \alpha_{n,\mathrm{var}} &= \sum_{y^n, \boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{x}^n)} P_{Y^n \boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)}(y^n, \boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{x}^n))(1 - P_{H_0|y^n,\boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{x}^n)}) \\ \beta_{n,\mathrm{var}} &= \sum_{y^n, \boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{x}^n)} P_{Y^n} \times P_{\boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)}(y^n, \boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{x}^n)) P_{H_0|y^n,\boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{x}^n)}. \end{aligned} \tag{79}$$

By [23, Lemma 12.2] the inequality

$$\alpha_{n,\mathrm{var}} + e^{n\hat{E}} \beta_{n,\mathrm{var}} \ge \Pr\left\{ \iota_{Y^n \boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)}(Y^n; \boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)) \le n\hat{E} \right\}, \tag{80}$$

is still valid. As (80) *implies* (74)*, the conclusion of Theorem 6 still holds when random decisions are used. In other words, stochastic decisions do not increase the optimal error exponent. Furthermore it can be seen that*

$$\begin{aligned} |1 - \alpha_n - \beta_n| &\le \|P_{H_0|Y^n \boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)} P_{Y^n \boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)} - P_{H_0|Y^n \boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)} P_{Y^n} \times P_{\boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)}\|_{TV} \\ &= \|P_{Y^n \boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)} - P_{Y^n} \times P_{\boldsymbol{\phi}_n^{\mathrm{t}}(\mathbf{X}^n)}\|_{TV}. \end{aligned} \tag{81}$$

Hence, Theorem 7 is still valid for random decisions.

C. Identification Systems with user-dependent compression mappings

Let W be the hidden random variable that characterizes the uniformly chosen user in the identification setting in Fig. 1 which is independent of users sequences $\mathbf{X}^n$. The underlying joint probability distribution is given by

$$P_{Y^n \mathbf{X}^n W}(y^n, \mathbf{x}^n, w) = \frac{1}{M^{\mathrm{id}}} P_{YX}^{\otimes n}(y^n, x^n(w)) \times \prod_{k \ne w} P_X^{\otimes n}(x^n(k)). \tag{82}$$

In the following we study the problem of identifying W with high probability when the compression mappings can be different from user to user. This problem is a generalization of the one in [3] which was depicted in Fig. 1. Similarly we use $\boldsymbol{\phi}_n^{\mathrm{id}}(\mathbf{X}^n)$ to denote the ensemble $(\phi_{1n}^{\mathrm{id}}(X^n(1)), \ldots, \phi_{M^{\mathrm{id}}n}^{\mathrm{id}}(X^n(M^{\mathrm{id}})))$. The decision mapping herein is defined by

$$\psi_n^{\mathrm{id}}(y^n, \boldsymbol{\phi}_n^{\mathrm{id}}(\mathbf{x}^n)) \mapsto \hat{w} \in \{1, \ldots, M^{\mathrm{id}}\} \cup \{e\}, \tag{83}$$

and the probability of identifying an incorrect index is given by

$$\Pr\{W \ne \hat{W}\} = P_{Y^n \boldsymbol{\phi}_n^{\mathrm{id}}(\mathbf{X}^n) W} \left(\left\{ (y^n, \boldsymbol{\phi}_n^{\mathrm{id}}(\mathbf{x}^n), w) \mid \psi_n^{\mathrm{id}}(y^n, \boldsymbol{\phi}_n^{\mathrm{id}}(\mathbf{x}^n)) \ne w \right\} \right). \tag{84}$$

Strong converse proofs for the setting in [3] have been given in [26] and [27]. We show that allowing different compression mappings does not change the optimal performance of the system in the strong converse sense. We first define the ϵ-achievability.

Definition 5. *A pair of compression-identification rates (R^{id}, R_c) is ϵ-achievable if there exists a doubly indexed sequence of compression mappings $(\phi_{kn}^{\mathrm{id}})$, a sequence of identification mappings (ψ_n^{id}) such that*

$$\begin{aligned} &\limsup_{n\to\infty} \max_{k\in[1:M^{\mathrm{id}}]} \frac{1}{n} \log |\phi_{kn}^{\mathrm{id}}| \le R_c, \\ &\liminf_{n\to\infty} \frac{1}{n} \log M^{\mathrm{id}} \ge R^{\mathrm{id}}, \ \limsup_{n\to\infty} \Pr\{W \ne \hat{W}\} \le \epsilon. \end{aligned} \tag{85}$$

Let $R^{\star}_{\epsilon,\mathrm{var}}(R_c)$ be the supremum of all ϵ-achievable identification rate R^{id} at a compression rate R_c.

We point out in the following one important difference between the identification system setting and our previous multi-user hypothesis testing problem. In our multi-user hypothesis testing, the number of users M grows at a fixed rate R, whereas in the identification setting we want to maximize the growing rate R^{id} of the number of supportable users M^{id} in the database. We therefore use two different notations to differentiate the distinction.

To prove the strong converse, we need the following lemma, which relates the probability of correct identifying with the probability that the compressed sequence of the chosen user and the observation sequence are jointly typical.

Lemma 4. *For any given η we have the following inequality*

$$\Pr\{W=\hat{W}\} \le \max_{k\in[1:M^{\mathrm{id}}]} \Pr\{\iota_{\bar{Y}^n\phi^{\mathrm{id}}_{kn}(\bar{X}^n)}(\bar{Y}^n;\phi^{\mathrm{id}}_{kn}(\bar{X}^n)) > \log M^{\mathrm{id}} - \eta\} + e^{-\eta}, \tag{86}$$

where $(\bar{Y}^n,\bar{X}^n)\sim P_{YX}^{\otimes n}$.

Proof. Define for each $k\in[1:M^{\mathrm{id}}]$ the following correctly decodable and jointly typical sets

$$\begin{aligned}\mathcal{D}_k &= \{(y^n,\mathbf{x}^n) \mid k=\psi^{\mathrm{id}}_n(y^n,\boldsymbol{\phi}^{\mathrm{id}}_n(\mathbf{x}^n))\}\\ \mathcal{A}_k &= \{(y^n,\mathbf{x}^n) \mid \iota_{\bar{Y}^n\phi^{\mathrm{id}}_{kn}(\bar{X}^n)}(y^n;\phi^{\mathrm{id}}_{kn}(x^n(k))) > \log M^{\mathrm{id}} - \eta\}.\end{aligned} \tag{87}$$

We observe that the sets $\mathcal{D}_k$ are disjoint. Then we have

$$\begin{aligned}\Pr\{W=\hat{W}\} &= \frac{1}{M^{\mathrm{id}}}\sum_{k=1}^{M^{\mathrm{id}}} P_{Y^n\phi^{\mathrm{id}}_{kn}(X^n(k))}\times\prod_{j\neq k} P_{\phi^{\mathrm{id}}_{jn}(X^n(j))}(\mathcal{D}_k)\\ &= \frac{1}{M^{\mathrm{id}}}\sum_{k=1}^{M^{\mathrm{id}}}\left[P_{Y^n\phi^{\mathrm{id}}_{kn}(X^n(k))}\times\prod_{j\neq k} P_{\phi^{\mathrm{id}}_{jn}(X^n(j))}(\mathcal{D}_k\cap\mathcal{A}_k^c) + P_{Y^n\phi^{\mathrm{id}}_{kn}(X^n(k))}\times\prod_{j\neq k} P_{\phi^{\mathrm{id}}_{jn}(X^n(j))}(\mathcal{D}_k\cap\mathcal{A}_k)\right].\end{aligned} \tag{88}$$

The second term inside the bracket, which is the probability of correct identification as well as the observation and the compressed sequence being jointly typical, can be upper bounded by

$$P_{Y^n\phi^{\mathrm{id}}_{kn}(X^n(k))}\times\prod_{j\neq k} P_{\phi^{\mathrm{id}}_{jn}(X^n(j))}(\mathcal{A}_k) = \Pr\{\iota_{\bar{Y}^n\phi^{\mathrm{id}}_{kn}(\bar{X}^n)}(\bar{Y}^n;\phi^{\mathrm{id}}_{kn}(\bar{X}^n)) > \log M^{\mathrm{id}} - \eta\}. \tag{89}$$

Similarly the first term inside the bracket, which is the probability of correct identification as well as the observation and the compressed sequence being not jointly typical, is upper bounded by

$$M^{\mathrm{id}}e^{-\eta}P_{Y^n}\times P_{\boldsymbol{\phi}^{\mathrm{id}}_n(\mathbf{X}^n)}(\mathcal{D}_k\cap\mathcal{A}_k^c) \le M^{\mathrm{id}}e^{-\eta}P_{Y^n}\times P_{\boldsymbol{\phi}^{\mathrm{id}}_n(\mathbf{X}^n)}(\mathcal{D}_k). \tag{90}$$

Therefore we have

$$\begin{aligned}\Pr\{W=\hat{W}\} &\le \frac{1}{M^{\mathrm{id}}}\sum_{k=1}^{M^{\mathrm{id}}}\Pr\{\iota_{\bar{Y}^n\phi^{\mathrm{id}}_{kn}(\bar{X}^n)}(\bar{Y}^n;\phi^{\mathrm{id}}_{kn}(\bar{X}^n)) > \log M^{\mathrm{id}} - \eta\} + e^{-\eta}P_{Y^n}\times P_{\boldsymbol{\phi}^{\mathrm{id}}_n(\mathbf{X}^n)}\left(\bigcup_k \mathcal{D}_k\right)\\ &\le \max_{k\in[1:M^{\mathrm{id}}]}\Pr\{\iota_{\bar{Y}^n\phi^{\mathrm{id}}_{kn}(\bar{X}^n)}(\bar{Y}^n;\phi^{\mathrm{id}}_{kn}(\bar{X}^n)) > \log M^{\mathrm{id}} - \eta\} + e^{-\eta}.\end{aligned} \tag{91}$$

The conclusion of the lemma follows. □

The following theorem concludes that the performance of an identification systems with user-dependent compression mappings are not better than the one using the same compression mapping for all users.

Theorem 8. *For $R^{\star}_{\epsilon,\mathrm{var}}(R_c)$ defined in Definition 5 we have $R^{\star}_{\epsilon,\mathrm{var}}(R_c) = R_{\max}(R_c)$ for all $\epsilon\in[0,1)$.*

Proof. Assume that the rate pair (R^{id},R_c) is ϵ-achievable. There exist corresponding sequences $(\phi^{\mathrm{id}}_{kn})$ and (ψ^{id}_n) such that all conditions in Definition 5 holds. The ϵ-achievability implies that $\liminf_{n\to\infty}\Pr\{W=\hat{W}\}\ge 1-\epsilon$. Suppose that $R^{\mathrm{id}} = R_{\max}(R_c)+3\gamma$ for some $\gamma>0$. By choosing $\eta=n\gamma$ in Lemma 4 and applying Lemma 3 with $N=M^{\mathrm{id}}$ we obtain that

$$\begin{aligned}\liminf_{n\to\infty}\Pr\{W=\hat{W}\} &\le \liminf_{n\to\infty}\max_{k\in[1:M^{\mathrm{id}}]}\Pr\{\iota_{\bar{Y}^n\phi^{\mathrm{id}}_{kn}(\bar{X}^n)}(\bar{Y}^n;\phi^{\mathrm{id}}_{kn}(\bar{X}^n)) > \log M^{\mathrm{id}} - \eta\}\\ &\le \liminf_{n\to\infty}\max_{k\in[1:M^{\mathrm{id}}]}\Pr\{\iota_{\bar{Y}^n\phi^{\mathrm{id}}_{kn}(\bar{X}^n)}(\bar{Y}^n;\phi^{\mathrm{id}}_{kn}(\bar{X}^n)) > n(R_{\max}(R_c)+\gamma)\} = 0,\end{aligned} \tag{92}$$

which contradicts the assumption of ϵ-achievability. Therefore we must have $R^{\mathrm{id}}\le R_{\max}(R_c)+3\gamma$. Since γ is arbitrary we then have $R^{\mathrm{id}}\le R_{\max}(R_c)$ which leads to $R^{\star}_{\epsilon,\mathrm{var}}(R_c)\le R_{\max}(R_c)$. □

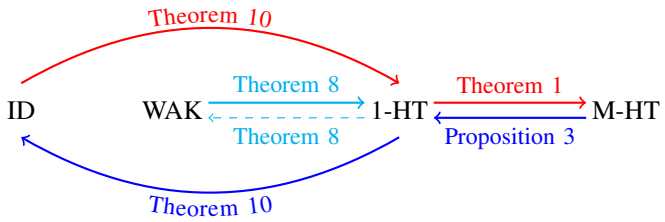

Fig. 3: Code transformations across settings. Solid are dashed lines indicate constructive and non-constructive transformation, respectively.

IV. Equivalence relations

This section is devoted to establishing operational equivalence between different settings: Wyner-Ahlswede-Körner (WAK) network, single-user testing against independence and identification systems. These equivalence relations are established via simple code transformations. We summarize the results of this section in Fig. 3.

Specifically, as a consequence of the operational equivalence we show explicitly that a code for an identification system can be used for the membership testing purpose and vice versa. These are highlighted by red and blue colors in Fig. 3. In other words, we do not need to use two different databases for screening users and recognizing users. This property is highly desirable in practice since one can add a new feature, herein membership testing, on top of an old feature (recognizing an existing user) without re-designing the whole system from the scratch.

Furthermore, we show that the simple code transformation argument between the WAK setting and the single-user HT implies the strong converse, exponentially strong converse equivalence statements. The latter complements the result in Theorem 4 proven previously.

The transformations between the single-user HT and the identification setting imply the equalities of the ϵ-achievable regions and the corresponding second-order quantities. These results connect and generalize existing results and are also of independent interest.

The presentation is given in the order of generality in the assumption of distributions. For simplicity we drop superscripts $(\cdot)^{\mathrm{id}}$ and $(\cdot)^{\mathrm{t}}$ at most of places.

A. Equivalence between single-user HT and WAK problems

Assume that the source in the WAK problem and the hypothesis H_0 in the HT setting are given by $\bar{X}^n\bar{Y}^n \sim P_{XY}^{\otimes n}$. We briefly recap the definitions of a WAK-code and a HT scheme. A WAK-code consists of a pair of encoding mappings (ϕ_{1n}, ϕ_{2n}) and a decoding function ψ_n which are defined as

$$\begin{aligned} \phi_{1n}\colon \mathcal{X}^n \to &\mathcal{M}_1,\ \phi_{2n}\colon \mathcal{Y}^n \to \mathcal{M}_2 \\ &\psi_n\colon \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{Y}^n. \end{aligned} \tag{93}$$

The WAK problem aims to control $\Pr\{\bar{Y}^n \neq \hat{\bar{Y}}^n\}$, where $\hat{\bar{Y}}^n = \psi_n(\phi_{1n}(\bar{X}^n), \phi_{2n}(\bar{Y}^n))$.

The single-user HT setting aims to differentiate whether the observations (x^n, y^n) are generated from hypothesis $H_0\colon P_{XY}^{\otimes n}$ or hypothesis $H_1\colon P_X^{\otimes n} \times P_Y^{\otimes n}$. A single-user HT scheme consists of an encoding mapping ϕ_n and a *stochastic* decision mapping ψ_n, which are defined as follows

$$\phi_n\colon \mathcal{X}^n \to \mathcal{M},\ \psi_n\colon \mathcal{M} \times \mathcal{Y}^n \to \{0,1\}, \tag{94}$$

in which H_0 is chosen with probability $P_{H_0|m,y^n}$. The probabilities of error of type I and II are given as

$$\begin{aligned} \bar{\alpha}_n &= \sum_{y^n,m} P_{\bar{Y}^n\phi_n(\bar{X}^n)}(y^n, m)(1 - P_{H_0|y^n,m}) \\ \bar{\beta}_n &= \sum_{y^n,m} P_{\bar{Y}^n} \times P_{\phi_n(\bar{X}^n)}(y^n, m) P_{H_0|y^n,m}. \end{aligned} \tag{95}$$

We present in the following an equivalent relation between a WAK-code and a single-user HT scheme.

Theorem 9. *Fix an arbitrary $\gamma > 0$. Given a WAK-code $(\phi_{1n}, \phi_{2n}, \psi_n)$, we can construct a single-user hypothesis testing scheme (ϕ_{1n}, ψ'_n) such that the corresponding error probabilities of type I and II are given by*

$$\begin{aligned} \bar{\alpha}_n &\le \Pr\{\bar{Y}^n \neq \hat{\bar{Y}}^n\} + \Pr\{\bar{Y}^n \notin \mathcal{A}_\gamma^n\} \\ \bar{\beta}_n &\le e^{-n(H(\bar{Y})-\gamma)}|\mathcal{M}_2|, \end{aligned} \tag{96}$$

where $\mathcal{A}_\gamma^n$ is the weakly typical set w.r.t P_Y. Conversely, given a testing scheme (ϕ_n, ψ_n) for the single-user hypothesis testing problem there exists a WAK-code $(\phi_n, \phi'_{2n}, \psi'_n)$ such that

$$\Pr\{\hat{\bar{Y}}^n \neq \bar{Y}^n\} \leq \bar{\alpha}_n + e^{n(\hat{E}-\gamma)}\bar{\beta}_n + \Pr\{\bar{Y}^n \notin \mathcal{A}_\gamma^n\} + \frac{e^{n(H(\bar{Y})+2\gamma-\hat{E})}}{|\mathcal{M}_2|}, \tag{97}$$

where $\hat{E}$ is a free parameter that satisfies the condition $H(\bar{Y}) > \hat{E} + 2\gamma$.

Proof. WAK $\Rightarrow$ Single-user HT: For a given $m_1 \in \mathcal{M}_1$, define the following correctly decodable set of the WAK-code

$$\mathcal{D}_{m_1} = \{y^n \mid y^n = \psi_n(m_1, \phi_{2n}(y^n)),\ y^n \in \mathcal{A}_\gamma^n\}. \tag{98}$$

Then, it is clear that for all $m_1 \in \mathcal{M}_1$ we have $|\mathcal{D}_{m_1}| \leq |\mathcal{M}_2|$ as ϕ_{2n} can only take at most $|\mathcal{M}_2|$ values. A decision region for the single-user HT, based on $\mathcal{Y}^n \times \mathcal{M}_1$, is defined as

$$\bar{\mathcal{A}}_n = \bigcup_{m_1} \big(\mathcal{D}_{m_1} \times \{m_1\}\big) \subset \mathcal{Y}^n \times \mathcal{M}_1. \tag{99}$$

The validity of $\bar{\mathcal{A}}_n$, i.e., the existence of a decision mapping ψ_n, follows from the fact that we have full access to the sequence y^n when making a decision. We use the mapping ϕ_{1n} as the compression mapping for $\mathcal{X}^n$ in the single-user HT problem. From (99) the probability of type I of error is bounded by

$$\bar{\alpha}_n = P_{\bar{Y}^n\phi_{1n}(\bar{X}^n)}(\bar{\mathcal{A}}_n^c) \leq \Pr\{\hat{\bar{Y}}^n \neq \bar{Y}^n\} + \Pr\{\bar{Y}^n \notin \mathcal{A}_\gamma^n\}, \tag{100}$$

and the probability of type II of error is bounded above by

$$\begin{aligned}
\bar{\beta}_n &= P_{\bar{Y}^n} \times P_{\phi_{1n}(\bar{X}^n)}(\bar{\mathcal{A}}_n) = \sum_{m_1} P_{\phi_{1n}(\bar{X}^n)}(m_1) P_{\bar{Y}^n}(\mathcal{D}_{m_1}) \\
&\leq \sum_{m_1} P_{\phi_{1n}(\bar{X}^n)}(m_1) |\mathcal{D}_{m_1}| e^{-n(H(Y)-\gamma)} \\
&\leq e^{-n(H(\bar{Y})-\gamma)} |\mathcal{M}_2|.
\end{aligned} \tag{101}$$

WAK $\Leftarrow$ Single-user HT: Given a single-user HT testing scheme (ϕ_n, ψ_n), for each $m_1 \in \mathcal{M}$ we define the set

$$\mathcal{D}_n(m_1) = \left\{ y^n \middle| \frac{1}{n} \log \frac{P_{\bar{Y}^n|\phi_n(\bar{X}^n)}(y^n|m_1)}{P_{\bar{Y}^n}(y^n)} > \hat{E} - \gamma \right\} \cap \mathcal{A}_\gamma^n. \tag{102}$$

$\mathcal{D}_n(m_1)$ plays the role of the conditional typical set in the standard proof of the WAK setting, cf. [28], [29]. We use the mapping ϕ_n as the compression mapping for $\mathcal{X}^n$ in the WAK-problem. From the definition of $\mathcal{D}_n(m_1)$ we obtain

$$\Pr\big\{\bar{Y}^n \notin \mathcal{D}_n(\phi_n(\bar{X}^n))\big\} \overset{(*)}{\leq} \bar{\alpha}_n + e^{n(\hat{E}-\gamma)}\bar{\beta}_n + \Pr\{\bar{Y}^n \notin \mathcal{A}_\gamma^n\}, \tag{103}$$

where $(*)$ follows from [23, Lemma 12.2]. Furthermore, we have $|\mathcal{D}_n(m_1)| \leq e^{n(H(\bar{Y})+2\gamma-\hat{E})}$ for all $m_1 \in \mathcal{M}$ since

$$\begin{aligned}
1 &\geq P_{\bar{Y}|\phi_n(\bar{X}^n)}(\mathcal{D}_n(m_1)|m_1) \\
&= \sum_{y^n \in \mathcal{D}_n(m_1)} P_{\bar{Y}^n|\phi_n(\bar{X}^n)}(y^n|m_1) \geq \sum_{y^n \in \mathcal{D}_n(m_1)} P_{\bar{Y}^n}(y^n) e^{n(\hat{E}-\gamma)} \\
&\geq |\mathcal{D}_n(m_1)| e^{-n(H(Y)+\gamma)} e^{n(\hat{E}-\gamma)}.
\end{aligned} \tag{104}$$

Let m_2 be a uniformly random bin index of y^n and $\mathcal{B}(m_2)$ be the set of all such y^n. The decoder decides that $\hat{y}^n$ is the reconstructed sequence if it is the unique sequence such that $\hat{y}^n \in \mathcal{B}(m_2) \cap \mathcal{D}_n(m_1)$, where m_1 and m_2 are sent messages from Encoder 1 and 2. It then follows that

$$\begin{aligned}
\Pr\{\hat{\bar{Y}}^n \neq \bar{Y}^n\} &\leq \Pr\{\bar{Y}^n \notin \mathcal{B}(M_2) \cap \mathcal{D}_n(\phi_n(\bar{X}^n))\} + \Pr\{\exists \tilde{y}^n \neq \bar{Y}^n, \tilde{y}^n \in \mathcal{D}_n(\phi_n(\bar{X}^n)) \cap \mathcal{B}(M_2)\} \\
&\overset{(a)}{\leq} \Pr\{\bar{Y}^n \notin \mathcal{D}_n(\phi_n(\bar{X}^n))\} + \Pr\{\exists \tilde{y}^n \neq \bar{Y}^n, \tilde{y}^n \in \mathcal{D}_n(\phi_n(\bar{X}^n)) \cap \mathcal{B}(M_2)\} \\
&\overset{(b)}{\leq} \alpha_n + e^{n(\hat{E}-\gamma)}\beta_n + \Pr\{\bar{Y}^n \notin \mathcal{A}_\gamma^n\} + \frac{e^{n(H(\bar{Y})+2\gamma-\hat{E})}}{|\mathcal{M}_2|}.
\end{aligned} \tag{105}$$

(a) is valid since $\bar{Y}^n \in \mathcal{B}(M_2)$. The inequality (b) holds since each $\tilde{y}^n$ is assigned independently to a bin with probability $1/|\mathcal{M}_2|$ and the number of such $\tilde{y}^n$ is bounded by $e^{n(H(\bar{Y})+2\gamma-\hat{E})}$, cf. (104). The existence of deterministic mappings ϕ'_{2n} and ψ'_n follows immediately. □

Remark 5. *Given a compression rate R_c for x^n in both settings, let $R^\star_{2,\epsilon}(R_c)$ and $E^\star_\epsilon(R_c)$ be the minimum ϵ-achievable compression rate for y^n in the WAK problem and the maximum ϵ-achievable error exponent for the single-user HT problem,*

respectively. As a direct consequence, it can be inferred from Theorem 9 that $R^{\star}_{2,\epsilon}(R_c) + E^{\star}_{\epsilon}(R_c) = H(Y)$ holds for all $(R_c, \epsilon) \in \mathbb{R}_+ \times [0, 1)$. In particular it implies that a strong converse for the WAK problem implies a strong converse for the single-user HT problem and vice versa. Strong converse for these two problems have been given in [30], [31] and [32].

An important consequence of Theorem 9, which states an equivalence of exponentially strong converse statements, is given as follows.

Theorem 10. *The following statements are equivalent:*

1) *For any code $(\phi_{1n}, \phi_{2n}, \psi_n)$ which satisfies $\limsup_{n\to\infty} \frac{1}{n} \log |\phi_{1n}| \leq R_c$ and $\limsup_{n\to\infty} \frac{1}{n} \log |\phi_{2n}| \leq R_2$ in the WAK problem, if $R_2 < H(Y) - R_{\max}(R_c)$, then $\Pr\{\bar{Y}^n \neq \hat{\bar{Y}}^n\} \to 1$ exponentially fast at a positive convergence rate.*
2) *For any single-user HT scheme (ϕ_n, ψ_n) with $\limsup_{n\to\infty} \frac{1}{n} \log |\phi_n| \leq R_c$ and $\liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_n} \geq E$, if $E > R_{\max}(R_c)$, then $\bar{\alpha}_n \to 1$ exponentially fast at a positive convergence rate.*

A proof of the first statement is presented in [7], [8].

Proof. Assume that the first statement holds. It suffices to show the second statement when $E < H(Y)$. Let (ϕ_n, ψ_n) be a single-user hypothesis testing scheme such that $\limsup_{n\to\infty} \frac{1}{n} \log |\phi_n| \leq R_c$ and $\liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\beta_n} \geq E$ where $E > R_{\max}(R_c)$. Let $\gamma > 0$ be small enough such that $E - 4\gamma > R_{\max}(R_c)$. By the second part of Theorem 9, there exists a WAK-code $(\phi_n, \phi'_{2n}, \psi'_n)$ such that with $\hat{E} = E - \gamma$ and $|\mathcal{M}_2| = e^{n(H(Y)+4\gamma-E)}$ we have

$$\Pr\{\hat{\bar{Y}}^n \neq \bar{Y}^n\} \leq \bar{\alpha}_n + 2e^{-n\gamma} + \Pr\{\bar{Y}^n \notin \mathcal{A}^n_\gamma\}, \tag{106}$$

for all sufficiently large n. Since the weakly typical set $\mathcal{A}^n_\gamma$ includes the strongly typical set $\mathcal{T}^n_\epsilon$ for a fixed, positive, and small enough ϵ, the last term goes to 0 exponentially with a convergence rate of at least $2\epsilon^2$. Since $H(Y)+4\gamma-E < H(Y)-R_{\max}(R_c)$ the assumption implies that $\Pr\{\bar{Y}^n = \hat{\bar{Y}}^n\}$ goes to 0 exponentially at a rate of $\eta > 0$, we then have $\bar{\alpha}_n \to 1$ exponentially at a positive rate.

Conversely, assume that the second statement holds. Let $(\phi_{1n}, \phi_{2n}, \psi_n)$ be a WAK-code such that

$$\limsup_{n\to\infty} \frac{1}{n} \log |\phi_{1n}| \leq R_c, \ \limsup_{n\to\infty} \frac{1}{n} \log |\phi_{2n}| \leq R_2, \tag{107}$$

where $R_2 < H(Y) - R_{\max}(R_c)$. Let $\gamma > 0$ be small enough such that $R_2 + 2\gamma < H(Y) - R_{\max}(R_c)$. By the first part of Theorem 9, the constructed testing scheme satisfies

$$\begin{aligned} \bar{\alpha}_n &\leq \Pr\{\bar{Y}^n \neq \hat{\bar{Y}}^n\} + \Pr\{\bar{Y}^n \notin \mathcal{A}^n_\gamma\} \\ \bar{\beta}_n &\leq e^{-n(H(Y)-2\gamma-R_2)}, \end{aligned} \tag{108}$$

for all sufficiently large n. Since $H(Y) - 2\gamma - R_2 > R_{\max}(R_c)$ holds the corresponding false alarm probability $\bar{\alpha}_n$ hence goes to 1 exponentially at a rate of $\xi > 0$, or $\Pr\{\bar{Y}^n \neq \hat{\bar{Y}}^n\} \to 1$ exponentially at a rate of $\min\{\xi, 2\epsilon^2\}$. □

B. Equivalence between single-user HT and Identification

Two hypotheses in the single-user HT problem are $H_0 \colon P_{\bar{X}^n\bar{Y}^n}$, $H_1 \colon P_{\bar{X}^n} \times P_{\bar{Y}^n}$. Note that we do not assume either $P_{\bar{X}^n\bar{Y}^n} = P^{\otimes n}_{XY}$, or $\mathcal{X}$ and $\mathcal{Y}$ are finite. Similarly as in Section III-A, it should be assumed that $\mathcal{X}$ and $\mathcal{Y}$ have sufficient structure, for example being Polish spaces. With abuse of terminology and notation we will redefine some terms and notations in the subsequent development.

A testing scheme consists of two compression mappings (ϕ_{1n}, ϕ_{2n}) and a deterministic decision mapping ψ_n where

$$\begin{aligned} \phi_{1n} \colon \mathcal{X}^n \to \mathcal{M}_1, \, \phi_{2n} \colon \mathcal{Y}^n \to \mathcal{M}_2 \\ \psi_n \colon \mathcal{M}_1 \times \mathcal{M}_2 \to \{0, 1\}. \end{aligned} \tag{109}$$

The acceptance region can be defined similarly as in (6). The probabilities of type I and II errors $\bar{\alpha}_n$ and $\bar{\beta}_n$ can also be determined accordingly. For the iid, discrete case, this setup was discussed briefly in [1], for which a single-letter characterization for the optimal achievable error exponent of type II of error is an open question.

The underlying distribution for the identification systems in Fig. 1 is given by

$$P_{Y^n \mathbf{X}^n W}(y^n, \mathbf{x}^n, w) = \frac{1}{M^{\mathrm{id}}} P_{\bar{Y}^n\bar{X}^n}(y^n, x^n(w)) \times \prod_{k\neq w} P_{\bar{X}^n(k)}(x^n(k)). \tag{110}$$

An identification scheme consists of two compression mappings (ϕ_{1n}, ϕ_{2n}) and a decoding mapping ψ_n where

$$\begin{aligned} \phi_{1n} \colon \mathcal{X}^n \to \mathcal{M}_1, \, \phi_{2n} \colon \mathcal{Y}^n \to \mathcal{M}_2 \\ \psi_n \colon \mathcal{M}_1^{M^{\mathrm{id}}} \times \mathcal{M}_2 \to \{1, \ldots, M^{\mathrm{id}}\} \cup \{e\}. \end{aligned} \tag{111}$$

Note that we are back to the case that the same compression mapping is used to enroll users data into a database. This setting reduces to the one given in Fig. 1 if we set ϕ_{2n} to be the identity mapping. In the identification problem one wants to control the probability of incorrect identification $\Pr\{\hat{W} \neq W\}$, where $\hat{W} = \psi_n(\phi_{1n}(\mathbf{X}^n), \phi_{2n}(Y^n))$. For the iid discrete scenario, this setting was studied in [33], where inner bounds and outer bounds on the achievable rate region were derived. A connection between the achievable regions of these two problems has been drawn recently in [34] via the entropy characterization.
We first establish the following useful lemma.

Lemma 5. *For a given $\gamma > 0$ and a given identification scheme $(\phi_{1n}, \phi_{2n}, \psi_n)$, we have*

$$\Pr\{\hat{W} \neq W\} \geq \Pr\left\{\iota_{\phi_{2n}(\bar{Y}^n)\phi_{1n}(\bar{X}^n)}(\phi_{2n}(\bar{Y}^n); \phi_{1n}(\bar{X}^n)) \leq \log M^{\mathrm{id}} - \eta\right\} - e^{-\eta}, \tag{112}$$

where again $(\bar{Y}^n, \bar{X}^n) \sim P_{\bar{Y}^n\bar{X}^n}$

The proof of Lemma 5 can be obtained from the one of Lemma 4 by simply replacing Y^n with $\phi_{2n}(Y^n)$, $\bar{Y}^n$ with $\phi_{2n}(\bar{Y}^n)$ and the user data compression mappings ϕ_{kn}^{id} with ϕ_{1n} for all $k \in [1 : M^{\mathrm{id}}]$. Using Lemma 5 we can establish the code transformation between the identification problem and the single-user hypothesis testing against independence as follows.

Theorem 11. *Fix an arbitrary $\eta > 0$. Given an identification scheme $(\phi_{1n}, \phi_{2n}, \psi_n)$, we can construct a single-user HT scheme $(\phi_{1n}, \phi_{2n}, \psi_n')$ such that the corresponding error probabilities of type I and II are given by*

$$\begin{aligned} \bar{\alpha}_n &\leq \Pr\{\hat{W} \neq W\} + e^{-\eta}, \\ \bar{\beta}_n &\leq \frac{e^{\eta}}{M^{\mathrm{id}}}. \end{aligned} \tag{113}$$

Conversely, given a testing scheme $(\phi_{1n}, \phi_{2n}, \psi_n)$ for the single-user HT problem, we can construct an identification scheme with M^{id} users $(\phi_{1n}, \phi_{2n}, \psi_n')$ such that

$$\Pr\{\hat{W} \neq W\} \leq \bar{\alpha}_n + M^{\mathrm{id}}\bar{\beta}_n. \tag{114}$$

Proof. ID $\Rightarrow$ Single-user HT: From a given identification scheme $(\phi_{1n}, \phi_{2n}, \psi_n)$ we use the same pair of mappings (ϕ_{1n}, ϕ_{2n}) to compress information in the single-user HT. Define an acceptance region for the single-user HT setup as

$$\bar{\mathcal{A}}_n = \{(\phi_{1n}(x^n), \phi_{2n}(y^n)) \mid \iota_{\phi_{2n}(\bar{Y}^n)\phi_{1n}(\bar{X}^n)}(\phi_{2n}(y^n); \phi_{1n}(x^n)) > \log M^{\mathrm{id}} - \eta\}. \tag{115}$$

The probability of type I of error is then given by

$$\bar{\alpha}_n = P_{\phi_{1n}(\bar{X}^n)\phi_{2n}(\bar{Y}^n)}(\bar{\mathcal{A}}_n^c) \leq \Pr\{\hat{W} \neq W\} + e^{-\eta}, \tag{116}$$

where the inequality follows from Lemma 5. By the change of measure we also obtain

$$\bar{\beta}_n = P_{\phi_{1n}(\bar{X}^n)} \times P_{\phi_{2n}(\bar{Y}^n)}(\bar{\mathcal{A}}_n) \leq \sum_{(\phi_{1n}(x^n), \phi_{2n}(y^n)) \in \bar{\mathcal{A}}_n} P_{\phi_{1n}(\bar{X}^n)\phi_{2n}(\bar{Y}^n)}(\phi_{1n}(x^n), \phi_{2n}(y^n)) \frac{e^{\eta}}{M^{\mathrm{id}}} \leq \frac{e^{\eta}}{M^{\mathrm{id}}}. \tag{117}$$

ID $\Leftarrow$ Single-user HT: Given a testing scheme $(\phi_{1n}, \phi_{2n}, \psi_n)$ of the single-user HT and a number of users M^{id}. We use the mapping ϕ_{1n} to compress each user's information and store it into a database and the mapping ϕ_{2n} to compress the observation sequence y^n in the identification setting. We define the decoding rule as follows. We search for a unique $\hat{w}$ such that

$$\psi_n(\phi_{2n}(y^n), \phi_{1n}(x^n(\hat{w}))) = 0. \tag{118}$$

If there exists none or there is more than one of such index, we output e. Define the following error events

$$\begin{aligned} \mathcal{E}_1 &= \{\psi_n(\phi_{2n}(Y^n), \phi_{1n}(X^n(W))) = 1\} \\ \mathcal{E}_2 &= \{\exists \tilde{w} \neq W \mid \psi_n(\phi_{2n}(Y^n), \phi_{1n}(X^n(\tilde{w}))) = 0\}. \end{aligned} \tag{119}$$

Then

$$\Pr\{\hat{W} \neq W\} \leq \Pr\{\mathcal{E}_1\} + \Pr\{\mathcal{E}_2\}. \tag{120}$$

It can be seen that the first term is equal to $\bar{\alpha}_n$ while the second term is upper bounded by $M^{\mathrm{id}}\bar{\beta}_n$. □

Remark 6. *Assume that a stochastic mapping ψ_n with the corresponding transition kernel $P_{\hat{w}|\phi_{1n}(\mathbf{x}^n),\phi_{2n}(y^n)}$ is used to estimate the true user in the identification problem instead. It can be seen that the conclusion of Lemma 5 is still valid. Therefore, Theorem 11 can be generalized with stochastic decision and identification mappings.*

Remark 7. *Suppose that we have an identification system that uses the same compression mapping ϕ_n^{id} to enroll data $\mathbf{x}^n$ from M users into the database where $\lim_{n\to\infty} \frac{1}{n} \log M = R$. Suppose further that the identification mapping ψ_n using uncompressed y^n and $\phi_n^{\mathrm{id}}(\mathbf{x}^n)$ is able to recognize M^{id} users where $\liminf_{n\to\infty} \frac{1}{n} \log M^{\mathrm{id}} \geq R^{\mathrm{id}} > R$. Then by using the first part of Theorem 11 with $\eta = n\gamma$ where $\gamma > 0$ and Theorem 1 we can construct a membership testing scheme using the*

available database $\phi_n^{\mathrm{id}}(\mathbf{x}^n)$ such that the error exponent of type II error is lower bounded by $R^{\mathrm{id}} - R - \gamma > 0$ for small enough γ.

We are now ready to present consequences of the transformation between the single-user HT problem and the identification problem. For that purpose we need some additional definitions, which we state in the following.

Definition 6. *For an arbitrary but fixed $\epsilon \in [0,1)$, define $\mathcal{R}_{\mathrm{ID},\epsilon}$ to be the closure of all tuples $(R_1, R_2, R^{\mathrm{id}})$ such that there exists an identification scheme $(\phi_{1n}, \phi_{2n}, \psi_n)$ which satisfies*

$$\limsup_{n\to\infty} \Pr\{\hat{W} \neq W\} \leq \epsilon, \quad \limsup_{n\to\infty} \frac{1}{n} \log |\phi_{in}| \leq R_i,\ i = 1,2, \tag{121a}$$

$$\liminf_{n\to\infty} \frac{1}{n} \log M^{\mathrm{id}} \geq R^{\mathrm{id}}. \tag{121b}$$

Define the ϵ-identification capacity for a given compression rate pair (R_1, R_2) as

$$R_\epsilon^\star(R_1, R_2) = \sup\{R^{\mathrm{id}} \mid (R_1, R_2, R^{\mathrm{id}}) \in \mathcal{R}_{\mathrm{ID},\epsilon}\}.$$

For an ϵ-achievable identification tuple $(R_1, R_2, R^{\mathrm{id}})$ we say that a second-order rate $\hat{R}$ is achievable if the condition (121b) *is replaced by*

$$\liminf_{n\to\infty} \frac{1}{\sqrt{n}} (\log M^{\mathrm{id}} - nR^{\mathrm{id}}) \geq \hat{R}. \tag{122}$$

Similarly, we define the maximum second-order identification rate $\hat{R}_\epsilon^\star(R_1, R_2, R^{\mathrm{id}})$ for a given rate pair $(R_1, R_2, R^{\mathrm{id}})$ as the supremum of achievable second-order rates $\hat{R}$.

Definition 7. *Let $\mathcal{R}_{\mathrm{HT},\epsilon}$ be the closure of all tuples (R_1, R_2, E) such that there exists a single-user HT scheme $(\phi_{1n}, \phi_{2n}, \psi_n)$ such that*

$$\limsup_{n\to\infty} \bar{\alpha}_n \leq \epsilon, \quad \limsup_{n\to\infty} \frac{1}{n} \log |\phi_{in}| \leq R_i,\ i = 1,2 \tag{123a}$$

$$\liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\bar{\beta}_n} \geq E. \tag{123b}$$

Similarly define the maximum ϵ- achievable error exponent for a given compression rate pair (R_1, R_2) as

$$E_\epsilon^\star(R_1, R_2) = \sup\{E \mid (R_1, R_2, E) \in \mathcal{R}_{\mathrm{HT},\epsilon}\}.$$

For an ϵ-achievable hypothesis testing tuple (R_1, R_2, E) we say that a second-order rate $\hat{E}$ is achievable if the condition (123b) *is replaced by*

$$\liminf_{n\to\infty} \frac{1}{\sqrt{n}} (\log \frac{1}{\bar{\beta}_n} - nE) \geq \hat{E}. \tag{124}$$

We define the maximum second-order HT error exponent $\hat{E}_\epsilon^\star(R_1, R_2, E)$ as the supremum of all achievable $\hat{E}$.

We show in the following theorem a connection between the maximum ϵ-achievable identification rate in the ID setting and the maximum ϵ-achievable error exponent in the single-user HT setting.

Theorem 12. *For all $\epsilon \in [0,1)$ and for all $(R_1, R_2) \in \mathbb{R}_+^2$, the following equality holds $E_\epsilon^\star(R_1, R_2) = R_\epsilon^\star(R_1, R_2)$.*

Proof. Assume that both quantities are finite. Given $\gamma > 0$ there exists an identification scheme $(\phi_{1n}, \phi_{2n}, \psi_n)$ such that all conditions in (121) are satisfied for $(R_1 + \gamma, R_2 + \gamma, R_\epsilon^\star(R_1, R_2) - \gamma)$. This implies that for all sufficiently large n we have $M \geq e^{n(R_\epsilon^\star(R_1,R_2)-2\gamma)}$. Then by the first part of Theorem 11 with $\eta = n\gamma$ the probabilities of error of the corresponding single-user HT scheme are bounded by

$$\limsup_{n\to\infty} \bar{\alpha}_n \leq \epsilon,\ \liminf_{n\to\infty} \frac{1}{n} \log \frac{1}{\bar{\beta}_n} \geq R_\epsilon^\star(R_1, R_2) - 3\gamma. \tag{125}$$

This implies that $E_\epsilon^\star(R_1, R_2) \geq R_\epsilon^\star(R_1, R_2)$, by taking $\gamma \to 0$.
Conversely, there exists a single-user testing scheme $(\phi_{1n}, \phi_{2n}, \psi_n)$ such that all conditions in (123) are satisfied for $(R_1 + \gamma, R_2 + \gamma, E_\epsilon^\star(R_1, R_2) - \gamma)$. This implies that for all sufficiently large n we have $\bar{\beta}_n \leq e^{-n(E_\epsilon^\star(R_1,R_2)-2\gamma)}$. By choosing $M = e^{n(E_\epsilon^\star(R_1,R_2)-3\gamma)}$ we obtain for $\eta = n\gamma$

$$\limsup_{n\to\infty} \Pr\{\hat{W} \neq W\} \leq \epsilon, \tag{126}$$

which implies that $(R_1 + \gamma, R_2 + \gamma, E_\epsilon^\star(R_1, R_2) - 3\gamma) \in \mathcal{R}_{\mathrm{ID},\epsilon}$. Hence, we have $E_\epsilon^\star(R_1, R_2) \leq R_\epsilon^\star(R_1, R_2)$.

Next, if $E_\epsilon^\star(R_1, R_2) = \infty$ for some pair $(R_1, R_2) \in \mathbb{R}_+^2$ and $\epsilon > 0$, then we can modify the proof as follows: Let $\{E_m\}_{m=1}^\infty$ be a sequence such that $E_m < \infty, \forall m$, and $E_m \to \infty$ as $m \to \infty$. Then we replace $E_\epsilon^\star(R_1, R_2)$ with E_m in the last paragraph

to get $(R_1, R_2, E_m) \in \mathcal{R}_{\mathrm{ID},\epsilon}$. This holds for any m, hence $R^\star_\epsilon(R_1, R_2) = \infty$ as well. The case $R^\star_\epsilon(R_1, R_2) = \infty$ can be handled similarly. □

Remark 8. *For any pair of compression mappings (ϕ_{1n}, ϕ_{2n}) such that $\limsup_{n\to\infty} \frac{1}{n} \log |\phi_{in}| \le R_i$ for $i = 1, 2$, define the joint process $(\phi_1(\mathbf{\bar{X}}), \phi_2(\mathbf{\bar{Y}})) = \{(\phi_{1n}(\bar{X}^n), \phi_{2n}(\bar{Y}^n))\}_{n=1}^\infty$. Then it follows from [21, Theorem 3.5.2] that the corresponding spectral sup-mutual information satisfies $\bar{I}(\phi_1(\mathbf{\bar{X}}); \phi_2(\mathbf{\bar{Y}})) \le \min\{R_1, R_2\} + \gamma$ where $\gamma > 0$ is arbitrary. Therefore, Lemma 5 implies that $R \le \min\{R_1, R_2\} + \gamma$ for any ϵ-achievable identification rate, otherwise the right-hand side of* (112) *would go to 1 by the definition of spectral-sup mutual information. Hence both $E^\star_\epsilon(R_1, R_2)$ and $R^\star_\epsilon(R_1, R_2)$ are finite and equal each other.*

Remark 9. *Suppose that ϕ_{2n} is the identity mapping and $P_{X^nY^n} = P_{XY}^{\otimes n}$ on a finite alphabet $\mathcal{X} \times \mathcal{Y}$ in both scenarios. It then can be inferred straightforwardly from Theorem 12 that the strong converse for the identification problem holds.*

In the next theorem we establish a connection between the maximum second order quantities in these two settings.

Theorem 13. *For all $\epsilon \in [0, 1)$ and for all ϵ-achievable pair (R_1, R_2, T) we have $\hat{R}^\star_\epsilon(R_1, R_2, T) = \hat{E}^\star_\epsilon(R_1, R_2, T)$.*

Proof. To show this relation we apply Theorem 11 with $\eta = \sqrt{n}\gamma$ with an arbitrary $\gamma > 0$. Given a second-order identification rate $\hat{R}$ by the first part of Theorem 11 we obtain $\beta_n \le e^{-n(T+(\hat{R}-2\gamma)/\sqrt{n})}$ since we have $M \ge e^{n(T+(\hat{R}-\gamma)/\sqrt{n})}$ for all sufficiently large n. This means that $\hat{R} - 2\gamma$ is a second-order achievable exponent for the hypothesis testing problem.

Given a second-order error exponent $\hat{E}$ by the second part of Theorem 11 we can choose $M = e^{n(T+(\hat{E}-2\gamma)/\sqrt{n})}$ to achieve the desired performance since we have $\beta_n \le e^{-n(T+(\hat{E}-\gamma)/\sqrt{n})}$ for all sufficiently large n. This means that $\hat{E} - 2\gamma$ is a second-order achievable exponent for the identification problem.

Since $\hat{E}$, $\hat{R}$ and γ are arbitrary the theorem follows. □

Remark 10. *For the iid, discrete, and uncompressed scenario such that $V = \mathrm{Var}[\iota(X;Y)] > 0$, Strassen's result [35] implies that $\hat{E}_\epsilon(\log|\mathcal{X}|, \log|\mathcal{Y}|, I(X;Y)) = \sqrt{V}\Phi^{-1}(\epsilon)$ where $\Phi(\cdot)$ is the cumulative distribution function of the standard Gaussian. Recently, Zhou et.al [26, Theorem 8] obtained the second-order identification rate $\hat{R}_\epsilon(\log|\mathcal{X}|, \log|\mathcal{Y}|, I(X;Y)) = \sqrt{V}\Phi^{-1}(\epsilon)$. Our result shows that the equality of these two results are a part of a more general relation.*

REFERENCES

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

[2] F. M. J. Willems, T. Kalker, and J.-P. Linnartz, "On the capacity of a biometrical identification system," in *International Symposium on Information Theory*, 2003.

[3] E. Tuncel, "Capacity/storage tradeoff in high-dimensional identification systems," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2097–2106, 2009.

[4] E. Tuncel and D. Gündüz, "Identification and lossy reconstruction in noisy databases," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 822–831, 2014.

[5] A. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 21, no. 3, pp. 294–300, 1975.

[6] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inf. Theory*, vol. 21, no. 6, pp. 629–637, 1975.

[7] Y. Oohama, "Exponent function for one helper source coding problem at rates outside the rate region," in *2015 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2015, pp. 1575–1579. [Online]. Available: https://arxiv.org/abs/1504.05891

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

[9] H. Chernoff *et al.*, "The identification of an element of a large population in the presence of noise," *The Annals of Statistics*, vol. 8, no. 6, pp. 1179–1197, 1980.

[10] S. Voloshynovskiy, O. Koval, F. Beekhof, F. Farhadzadeh, and T. Holotyak, "Information-theoretical analysis of private content identification," in *IEEE Information Theory Workshop (ITW)*. IEEE, 2010, pp. 1–5.

[11] N. A. Schmid, "Large deviations performance analysis for biometrics recognition," in *40th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2002.

[12] P. Moulin, "Statistical modeling and analysis of content identification," in *Information Theory and Applications Workshop (ITA)*. IEEE, 2010, pp. 1–5.

[13] V. Yachongka and H. Yagi, "Reliability function and strong converse of biomedical identification systems," in *International Symposium on Information Theory and Its Applications (ISITA)*. IEEE, 2016, pp. 547–551.

[14] G. Dasarathy and S. C. Draper, "On reliability of content identification from databases based on noisy queries," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 1066–1070.

[15] N. Merhav, "Reliability of universal decoding based on vector-quantized codewords," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 2696–2709, 2017.

[16] S. Sreekumar, D. Gündüz, and A. Cohen, "Distributed hypothesis testing under privacy constraints," in *2018 IEEE Information Theory Workshop (ITW)*. IEEE, 2018, pp. 1–5.

[17] P. Escamilla, M. Wigger, and A. Zaidi, "Distributed hypothesis testing with concurrent detections," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 166–170.

[18] S. Salehkalaibar, M. Wigger, and L. Wang, "Hypothesis testing over the two-hop relay network," *IEEE Trans. Inf. Theory*, 2019.

[19] D. Cao, L. Zhou, and V. Y. Tan, "Strong converse for hypothesis testing against independence over a two-hop network," *arXiv preprint arXiv:1808.05366*, 2018.

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

[21] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer-Verlag Berlin Heidelberg, 2003.

[22] J. Liu, P. Cuff, and S. Verdú, "E_γ-resolvability," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 2629–2658, 2017.

[23] Y. Polyanskiy and Y. Wu, "Lecture notes on information theory," *MIT (6.441), UIUC (ECE 563)*, 2017.

[24] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.

[25] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1562–1575, 2006.

[26] L. Zhou, V. Y. Tan, L. Yu, and M. Motani, "Exponential strong converse for content identification with lossy recovery," *IEEE Trans. Inf. Theory*, vol. 64, no. 8, pp. 5879–5897, 2018.

[27] M. T. Vu, T. J. Oechtering, M. Skoglund, and H. Boche, "Uncertainty in identification systems," *Submitted*, 2018. [Online]. Available: https://people.kth.se/~mtvu/Ui_jnlp_v3.pdf
[28] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.
[29] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.
[30] R. Ahlswede, P. Gács, and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Probability Theory and Related Fields*, vol. 34, no. 2, pp. 157–177, 1976.
[31] J. Liu, R. van Handel, and S. Verdú, "Second-order converses via reverse hypercontractivity," *arXiv preprint arXiv:1812.10129*, 2018.
[32] S. Watanabe, "A converse bound on wyner-ahlswede-körner network via gray-wyner network," in *2017 IEEE Information Theory Workshop (ITW)*. IEEE, 2017, pp. 81–85.
[33] M. B. Westover and J. A. O'Sullivan, "Achievable rates for pattern recognition," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 299–320, 2008.
[34] G. Pichler, P. Piantanida, and G. Matz, "Distributed information-theoretic clustering," *arXiv preprint arXiv:1602.04605*, 2016.
[35] V. Strassen, "Asymptotische Abschätzungen in Shannons Informationstheorie." *Trans. Third Prague Conference on Information Theory*, 1962.