On Distributed Hypothesis Testing with Constant-Bit Communication Constraints

Xiangxiang Xu and Shao-Lun Huang
DSIT Research Center
Tsinghua–Berkeley Shenzhen Institute
Shenzhen, China 518055
Email: xiangxiangxu@ieee.org, shaolun.huang@sz.tsinghua.edu.cn

***Abstract*—In this paper, we consider the distributed hypothesis testing (DHT) problem where two nodes are constrained to transmit constant bits to a central decoder. In such cases, we show that in order to achieve the optimal error exponents, it suffices to consider the empirical distributions of observed data sequences and encode them to the transmission bits. With such a coding strategy, we develop a geometric approach in the distribution spaces to show the optimal achievable error exponents and coding scheme for the following cases: (i) both nodes can transmit $\log_2 3$ bits; (ii) one of the nodes can transmit 1 bit, and the other node is not constrained; (iii) the joint distribution of the nodes are conditionally independent given one hypothesis. Our approach essentially reveals new potentials for characterizing the precise error exponents for DHT with general communication constraints.**

I. Introduction

The machine learning problems with distributed data have gained much attentions recently in federated learning [1], where the available data are observed at different nodes. The central problem of distributed learning is to develop an efficient learning scheme under communication or computation constraints, for various learning tasks, such as label inference or feature extraction. On the other hand, such problems have also been extensively studied in information theory [2]–[5], and the behaviors are typically much more complex than their non-distributed counterparts.

In this paper, we investigate the distributed hypothesis testing (DHT) problem as follows. Suppose that there are a pair of random variables X, Y and joint distributions $P_{XY}^{(0)}$ and $P_{XY}^{(1)}$. In addition, there are n samples i.i.d. generated from either $P_{XY}^{(0)}$ or $P_{XY}^{(1)}$, which may correspond to the two hypothesis $\mathsf{H} = 0$ and $\mathsf{H} = 1$ in statistics, or different labels in supervised learning problems. Moreover, in the distributed setup, we assume that there are two nodes, referred to as node N_X and node N_Y, each observes only the n i.i.d. samples of X and the samples of Y, respectively, and each node sends an encoded message to a central decoder. Then, the decoder makes a decision of the hypothesis $\hat{\mathsf{H}}$ according to the received messages. In particular, we focus on the case where both N_X and N_Y are required to encode (compress) the observed length-n sequences to constant number of bits, due to limited communication budgets. Our goal is to design the encoder of each node and the central decoder to minimize the error probability of inferring the label. Specifically, we focus on the asymptotic regime such that n is large, and characterize the error exponent pair for both type-I and type-II errors. The rigorous mathematical formulation is presented in Section II.

The general framework of such multiterminal statistical inference problems was first introduced in [6]. Then, the DHT problem with full side information was formulated and investigated in [7], where the sequence observed by N_Y can be directly transmitted to the center, while N_X can only send messages at some positive rate. Following this work, there have been a series of studies on DHT under different settings of communication constraints, which are typically represented as the communications rates, or equivalently, the compression rates of the encoders. Specifically, the DHT problem with zero-rate compression was first introduced in [2], where the one-bit compression case was also discussed. The achievable error exponent pairs under two-sided one-bit compression were later established in [3]. The DHT problem under zero-rate compression was also investigated in [8], [9]. A comprehensive survey of representative works through this line of researches can be found in [4]. Recently, the studies on DHT are still fairly active, with new analyzing tools and settings considered. For example, the DHT problem under zero-rate communication constraints was revisited in [5] from the perspectives of information-spectrum approach and finite blocklength analysis, and the variant of DHT with transmission noises was investigated in [10]. Despite of such massive studies, the characterizations of DHT under general communication constraints still remain open, except for several special cases, e.g., two-sided one-bit compression (cf. [3]) and zero-rate compression (cf. [4]).

The primary aim of this paper is to investigate the optimal error exponent pairs of DHT with constant-bit communication constraints, with the following main contributions. First, we demonstrate that the optimal encoding scheme depends only on the empirical distributions of the observed sequences, rather than the sequences themselves, as long as the compression rates are zeros. With such coding strategy, we develop a geometric approach in the distribution spaces to characterize the achievable error exponent pairs. Using this approach, we further provide lower bounds for the error exponents, via investigating the performance under a threshold decision rule. In addition, we show that the lower bounds are tight and establish the optimal error exponents, for the following cases:

(i) *two-sided one-trit compression*, where both nodes can transmit one-trit (trinary digit) message; (ii) *one-sided one-bit compression*, where one node can transmit one bit, and the other node is not constraint; (iii) the nodes are conditionally independent given one hypothesis. Our characterization extends previous studies on two-sided one-bit compression (cf. [3], [4]) and provides a novel geometric interpretation, which suggests new potentials for error exponent region characterization of DHT under general communication constraints.

II. Problem Formulation and Preliminaries

In this section, we introduce the mathematical formulation of DHT problem, and provide some useful definitions and notations.

A. Problem Formulation

First, we assume both X and Y are discrete random variables, taking values from finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively. Then, the general setup of DHT can be illustrated as Fig. 1. When $\mathsf{H} = i$, n i.i.d. sample pairs $\{(X_j, Y_j)\}_{j=1}^n$ are generated from the joint distribution $P_{XY}^{(i)}$, with node N_X observing $X^n \triangleq (X_1, \dots, X_n)$ and node N_Y observing $Y^n \triangleq (Y_1, \dots, Y_n)$, respectively. Then, nodes N_X and N_Y compress their observed sequences with the encoders $f_n \colon \mathcal{X}^n \to \mathcal{M}_X^{(n)}$ and $g_n \colon \mathcal{Y}^n \to \mathcal{M}_Y^{(n)}$, respectively, which encode the observed sequences into messages $f_n(X^n)$ and $g_n(Y^n)$. The encoded messages are further sent to a central machine, which makes the decision $\hat{\mathsf{H}} \triangleq \phi_n(f_n(X^n), g_n(Y^n))$, with $\phi_n \colon \mathcal{M}_X^{(n)} \times \mathcal{M}_Y^{(n)} \to \{0, 1\}$ being used as the decoder.

Due to the limited communication budgets in practice, there are typically constraints on the sizes of the message sets $\mathcal{M}_X^{(n)}$ and $\mathcal{M}_Y^{(n)}$. Following the convention introduced in [4], we use $\|f_n\| \triangleq \left|\mathcal{M}_X^{(n)}\right|$ and $\|g_n\| \triangleq \left|\mathcal{M}_Y^{(n)}\right|$ to denote the cardinalities of message sets, and express the constraints on $\|f_n\|$ and $\|g_n\|$ as a pair (R_X, R_Y), referred as the *rate* of encoders f_n and g_n, with $R_X, R_Y \in [0, \infty) \cup \{0_M \colon M \geq 1\}$. Specifically, each $R_X \in [0, \infty)$ indicates the constraint[1]

$$\limsup_{n\to\infty} \frac{1}{n} \log \|f_n\| \leq R_X, \tag{1}$$

and each $R_X = 0_M$ with $M \geq 1$ indicates the constraint

$$\limsup_{n\to\infty} \|f_n\| \leq M, \tag{2}$$

namely, the encoded message $f_n(x^n)$ is allowed to take at most M distinct values[2]. Specifically, we refer to f_n (or g_n) as a *zero-rate* encoder if it satisfies the constraint $R_X = 0$ (or $R_Y = 0$), and the corresponding hypothesis testing setting is called the *zero-rate compression* regime. In this paper, we focus on the DHT problem under constant-bit communication constraints, also referred to as *constant-bit compression* regime, where we have $R_X \in \{0_M \colon M \geq 1\}$ or $R_Y \in \{0_M \colon M \geq 1\}$.

[1]Throughout, the logarithm $\log(\cdot)$ indicates the natural logarithm with base e, unless otherwise specified.

[2]For mathematical convenience, we allow M to take 1, where no information can be transmitted from the node to center.

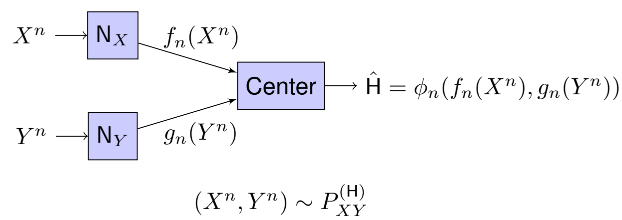


Fig. 1. Distributed Hypothesis Testing Setup

Note that the coding scheme can be characterized as a functional tuple $\mathcal{C}_n = (f_n, g_n, \phi_n)$. For a given coding scheme $\mathcal{C}_n$, the performance can be characterized by the type-I error $\pi_0(\mathcal{C}_n)$ and type-II error $\pi_1(\mathcal{C}_n)$, defined as

$$\pi_i(\mathcal{C}_n) \triangleq \mathbb{P}\left\{\hat{\mathsf{H}} \neq i \middle| \mathsf{H} = i\right\}, \tag{3}$$

for $i = 0, 1$, where $\mathbb{P}\{\cdot\}$ denotes the probability with respect to the corresponding i.i.d. sampling process over n sample pairs.

In particular, we consider the asymptotic regime such that n is large and characterize the achievable error exponents, defined as follows.

Definition 1 (Error Exponent Region): Given a rate pair (R_X, R_Y), an error exponent pair (E_0, E_1) is achievable under (R_X, R_Y), if there exists a sequence of coding schemes $\{\mathcal{C}_n = (f_n, g_n, \phi_n)\}_{n\geq 1}$ such that the encoders f_n and g_n satisfy the rate constraints (R_X, R_Y), and

$$\lim_{n\to\infty} \frac{1}{n} \log \pi_i(\mathcal{C}_n) = -E_i, \quad i = 0, 1. \tag{4}$$

Then, we define the error exponent region $\mathcal{E}(R_X, R_Y)$ as the closure of the set of all achievable error exponent pairs under the rate constraints. Specifically, under constant-bit compression, if the coding schemes $\mathcal{C}_n$'s in (4) have a common decoder ϕ for all $n \geq 1$, we call an error exponent pair (E_0, E_1) is achievable under decoder ϕ. Then, we use $\mathcal{E}[\{\phi\}]$ (or simply $\mathcal{E}[\phi]$) to denote the closure of the set of all such pairs.

Our goal is to characterize the error exponent region under constant-bit compression regime and the coding schemes to achieve the error exponents.

B. Definitions and Notations

Given an alphabet $\mathcal{Z} \in \{\mathcal{X}, \mathcal{Y}, \mathcal{X} \times \mathcal{Y}\}$, we use $\mathcal{P}^{\mathcal{Z}}$ to denote the set of distributions supported on $\mathcal{Z}$. For a given distribution $P_Z \in \mathcal{P}^{\mathcal{Z}}$, we use $(P_Z)^{\otimes n}$ to denote the n-th product of P_Z. In addition, for $P_Z, Q_Z \in \mathcal{P}^{\mathcal{Z}}$, we introduce the metric

$$d_{\max}(P_Z, Q_Z) \triangleq \max_{z\in\mathcal{Z}} |P_Z(z) - Q_Z(z)|. \tag{5}$$

For a joint distribution $Q_{XY} \in \mathcal{P}^{\mathcal{X}\times\mathcal{Y}}$, the corresponding marginal distributions are denoted by $[Q_{XY}]_X \in \mathcal{P}^{\mathcal{X}}$ and $[Q_{XY}]_Y \in \mathcal{P}^{\mathcal{Y}}$. In particular, for each $i = 0, 1$, we denote $P_X^{(i)} \triangleq [P_{XY}^{(i)}]_X, P_Y^{(i)} \triangleq [P_{XY}^{(i)}]_Y$.

A sequence $(z_1, \dots, z_n) \in \mathcal{Z}^n$ is denoted by $\{z_i\}_{i=1}^n$ or simply z^n. With slight abuse of notation, we use (x^n, y^n) or simply $x^n y^n$ to denote the sequence $\{(x_i, y_i)\}_{i=1}^n \in (\mathcal{X}\times\mathcal{Y})^n$, and denote the set

$$\{(x^n, y^n)\colon x^n \in \mathcal{S}_X, y^n \in \mathcal{S}_Y\} \subset (\mathcal{X}\times\mathcal{Y})^n \tag{6}$$

by $\mathcal{S}_X \times \mathcal{S}_Y$, for given $\mathcal{S}_X \subset \mathcal{X}^n$ and $\mathcal{S}_Y \subset \mathcal{Y}^n$.

We also introduce the definition of Hamming$d$-neighborhood as follows.

Definition 2: The *Hamming d-neighborhood* of $\mathcal{S}_Z \subset \mathcal{Z}^n$ is

$$\mathcal{N}_{\mathrm{H}}^d(\mathcal{S}_Z) \triangleq \{z^n \in \mathcal{Z}^n\colon d_{\mathrm{H}}(z^n, \tilde{z}^n) \leq k \text{ for some } \tilde{z}^n \in \mathcal{S}_Z\},$$

where $d_{\mathrm{H}}(z^n, \tilde{z}^n)$ denotes the Hamming distance between $z^n, \tilde{z}^n \in \mathcal{Z}^n$, i.e.,

$$d_{\mathrm{H}}(z^n, \tilde{z}^n) \triangleq \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{z_i \neq \tilde{z}_i\}},$$

and where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function.

For a given sequence $z^n \in \mathcal{Z}^n$, we use $\hat{P}_{z^n} \in \mathcal{P}^{\mathcal{Z}}$ to denote its empirical distribution (type), defined as

$$\hat{P}_{z^n}(z') = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{z_i = z'\}} \text{ for all } z' \in \mathcal{Z}.$$

Then, the set of all empirical distributions of sequences in $\mathcal{Z}^n$ is denote as

$$\hat{\mathcal{P}}_n^{\mathcal{Z}} \triangleq \left\{\hat{P}_{z^n} : z^n \in \mathcal{Z}^n\right\}.$$

Specifically, given a type $Q_Z \in \hat{\mathcal{P}}_n^{\mathcal{Z}}$, we use $\mathcal{T}_{Q_Z}^n$ (or simply $\mathcal{T}_{Q_Z}$) to denote the set of sequences $z^n \in \mathcal{Z}^n$ with the type Q_Z, i.e.,

$$\mathcal{T}_{Q_Z}^n \triangleq \{z^n \in \mathcal{Z}^n\colon \hat{P}_{z^n} = Q_Z\}.$$

For a given $\eta > 0$, we also define

$$\mathcal{T}_{Q_Z;\eta}^n \triangleq \left\{z^n \in \mathcal{Z}^n\colon d_{\max}(\hat{P}_{z^n}, Q_Z) \leq \eta\right\}. \tag{7}$$

Moreover, encoder f_n is called *type-based*, if its output $f_n(x^n)$ relies only on the type $\hat{P}_{x^n}$ of the sequence x^n, for all $x^n \in \mathcal{X}^n$. Similarly, g_n is type-based if $g_n(y^n)$ is a function of $\hat{P}_{y^n}$, for all $y^n \in \mathcal{Y}^n$.

Furthermore, we use $\mathcal{P}_\star \triangleq \mathcal{P}^{\mathcal{X}} \times \mathcal{P}^{\mathcal{Y}}$ to denote the product space of marginal distributions. For each $i = 0, 1$ and $t > 0$, we define the subsets $\mathcal{D}_i(t)$ of $\mathcal{P}_\star$ as

$$\mathcal{D}_i(t) \triangleq \{(Q_X, Q_Y) \in \mathcal{P}_\star\colon D_i^*(Q_X, Q_Y) < t\}, \tag{8}$$

where the function $D_i^*\colon \mathcal{P}_\star \to \mathbb{R}$ is defined as

$$D_i^*(Q_X, Q_Y) \triangleq \min_{\substack{Q_{XY}\colon [Q_{XY}]_X = Q_X\\ [Q_{XY}]_Y = Q_Y}} D(Q_{XY}\|P_{XY}^{(i)}), \tag{9}$$

where $D(\cdot\|\cdot)$ denotes the Kullback-Leibler (KL) divergence between distributions. The following simple fact will be useful in our analyses, of which a proof is provided in Appendix A.

Fact 1: For each $i = 0, 1$, and $t \geq 0$, $\mathcal{D}_i(t)$ is convex.

In addition, several useful operations on $\mathcal{P}_\star$ are defined as follows. For each given $\mathcal{A} \subset \mathcal{P}_\star$, we define its projections $\Pi_X(\mathcal{A})$ on $\mathcal{P}^{\mathcal{X}}$ and $\Pi_Y(\mathcal{A})$ on $\mathcal{P}^{\mathcal{Y}}$, as

$$\Pi_X(\mathcal{A}) \triangleq \{Q_X \in \mathcal{P}^{\mathcal{X}}\colon (Q_X, Q_Y') \in \mathcal{A} \text{ for some } Q_Y' \in \mathcal{P}^{\mathcal{Y}}\},$$
$$\Pi_Y(\mathcal{A}) \triangleq \{Q_Y \in \mathcal{P}^{\mathcal{Y}}\colon (Q_X', Q_Y) \in \mathcal{A} \text{ for some } Q_X' \in \mathcal{P}^{\mathcal{X}}\}.$$

Then, we have the following definition.

Definition 3: The binary operator "$\triangleright$" on $\mathcal{P}_\star$ is defined such that, for $\mathcal{A}, \mathcal{A}' \subset \mathcal{P}_\star$,

$$\mathcal{A} \triangleright \mathcal{A}' \triangleq \{(Q_X, Q_Y) \in \mathcal{A}\colon Q_X \in \Pi_X(\mathcal{A}'), Q_Y \in \Pi_Y(\mathcal{A}')\}.$$

In addition, for each $k \geq 0$, we define the operator "$\triangleright_k$" as

$$\mathcal{A} \triangleright_0 \mathcal{A}' \triangleq \mathcal{A}, \quad \mathcal{A} \triangleright_1 \mathcal{A}' \triangleq \mathcal{A}', \tag{10}$$
$$\mathcal{A} \triangleright_{k+2} \mathcal{A}' \triangleq (\mathcal{A} \triangleright_k \mathcal{A}') \triangleright (\mathcal{A} \triangleright_{k+1} \mathcal{A}') \quad \text{for } k \geq 0. \tag{11}$$

We also define operators "$\overset{X}{\triangleright}$", "$\overset{Y}{\triangleright}$" as

$$\mathcal{A} \overset{X}{\triangleright} \mathcal{A}' \triangleq \{(Q_X, Q_Y) \in \mathcal{A}\colon Q_X \in \Pi_X(\mathcal{A}')\}, \tag{12}$$
$$\mathcal{A} \overset{Y}{\triangleright} \mathcal{A}' \triangleq \{(Q_X, Q_Y) \in \mathcal{A}\colon Q_Y \in \Pi_Y(\mathcal{A}')\}. \tag{13}$$

For sequences $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$, we use $a_n = o(b_n)$ to indicate that $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$. We also define $\lceil M \rfloor \triangleq \{0, \dots, M-1\}$ for $M \geq 1$, and $\bar{\imath} \triangleq 1 - i$ for $i \in \{0, 1\}$.

C. Characterizations on Decoders

Under finite-symbol compression regime, for a given n, suppose that we have $(\|f_n\|, \|g_n\|) = (M_X, M_Y)$ with $M_X, M_Y \geq 1$. Without loss of generality, we assume the corresponding message sets are $\mathcal{M}_X^{(n)} = \lceil M_X \rfloor$ and $\mathcal{M}_Y^{(n)} = \lceil M_Y \rfloor$, respectively. Then, the decoder ϕ_n is a Boolean-valued function on $\lceil M_X \rfloor \times \lceil M_Y \rfloor$, formalized as follows.

1) Decoder Representation and Special Decoders:

Definition 4: Given $M_X, M_Y \geq 1$, an $M_X \times M_Y$ decoder is a function $\phi : \lceil M_X \rfloor \times \lceil M_Y \rfloor \to \{0, 1\}$. For a given decoder ϕ, its complement $\bar{\phi}$ is defined as $\bar{\phi} \triangleq 1 - \phi$. In addition, the *decision matrix* associated with ϕ is defined as an $M_Y \times M_X$ Boolean matrix $\mathbf{A}$ with entries $A(m_Y, m_X) \triangleq \phi(m_X, m_Y)$ for all $(m_X, m_Y) \in \lceil M_X \rfloor \times \lceil M_Y \rfloor$.

Specifically, the one-to-one correspondence between decoder ϕ and its associated decision matrix $\mathbf{A}$ is denoted by $\phi \leftrightarrow \mathbf{A}$.

Several special decoders as then introduced as follows.

We call ϕ *trivial* if $\phi \equiv 0$ or $\phi \equiv 1$, and call ϕ *degenerated* if $\mathbf{A} \leftrightarrow \phi$ has duplicate rows or columns. In addition, we call a decoder ϕ *monotonic*, if $\phi(m_X, m_Y) \leq \phi(m_X', m_Y')$ for all $m_X \leq m_X'$ and $m_Y \leq m_Y'$. As an important example of monotonic decoders, we introduce threshold decoders as follows.

Definition 5: For given $M_X, M_Y \geq 1$, the $M_X \times M_Y$ *threshold decoders* are the $M_X \times M_Y$ decoder φ_{M_X,M_Y} and its complement $\bar{\varphi}_{M_X,M_Y}$, where

$$\varphi_{M_X,M_Y}(m_X, m_Y) \triangleq \mathbb{1}_{\{m_X + m_Y \geq \min\{M_X, M_Y\}\}},$$

for all $(m_X, m_Y) \in \lceil M_X \rfloor \times \lceil M_Y \rfloor$.

For a given threshold decoder φ_{M_X,M_Y}, the inputs $m_X \in \lceil M_X \rfloor, m_Y \in \lceil M_Y \rfloor$ can be regarded as discrete-valued beliefs of node N_X and node N_Y for $\hat{\mathsf{H}} = 1$, and the decision $\varphi_{M_X,M_Y}(m_X, m_Y)$ is obtained, by first using a summation to fuse the beliefs from two nodes, then comparing the fused result $m_X + m_Y$ to the threshold $\min\{M_X, M_Y\}$.

We will sometimes find it convenient to express a decision matrix as filled grids of the same size, with occupied grids and empty grids indicating "1" and "0", respectively. For example, when $M_X = M_Y = 2$, the threshold decoders $\varphi_{2,2}$ and $\bar{\varphi}_{2,2}$ as defined in Definition 5 can be represented as "" and "", respectively.

For two given decoders ϕ, ϕ' with decision matrices $\mathbf{A} \leftrightarrow \phi$ and $\mathbf{A}' \leftrightarrow \phi'$, we call ϕ' a subdecoder of ϕ if $\mathbf{A}'$ is a submatrix of $\mathbf{A}$. In addition, ϕ, ϕ' are called *equivalent*, denoted by $\phi \simeq \phi'$, if $\mathbf{A}'$ can be obtained from $\mathbf{A}$ by row permutations and column permutations.

2) Reduction and Decomposition Operations: Then, we introduce two important operations on decoders.

a) Decoder Reduction: Given a decision matrix $\mathbf{A}$ and $i \in \{0, 1\}$, its *i-dominated rows* (or columns) are defined as the rows (or columns) being all i's. Then, a decoder ϕ is called *reducible* if $\mathbf{A} \leftrightarrow \phi$ has dominated rows or columns, and we introduce reduction operations of ϕ as follows.

Definition 6: Given a non-trivial reducible decoder $\phi \leftrightarrow \mathbf{A}$, if $\mathbf{A}$ has i-dominated columns for $i \in \{0, 1\}$, we define decoder $\omega_X^{(i)}(\phi)$ such that $\omega_X^{(i)}(\phi) \leftrightarrow \mathbf{A}_X^{(i)}$, where $\mathbf{A}_X^{(i)}$ denotes the submatrix of $\mathbf{A}$ obtained by deleting its i-dominated columns; similarly, if $\mathbf{A}$ has i-dominated rows, we define $\omega_Y^{(i)}(\phi)$ such that $\omega_Y^{(i)}(\phi) \leftrightarrow \mathbf{A}_Y^{(i)}$, where $\mathbf{A}_Y^{(i)}$ is the submatrix of $\mathbf{A}$ obtained by deleting i-dominated rows.

We refer to $\omega_X^{(0)}, \omega_X^{(1)}, \omega_Y^{(0)}, \omega_Y^{(1)}$ as elementary reduction operators. Then, we define the elementary reduction operators and their compositions as *reduction operators*. Given ϕ and ϕ', we say ϕ can be reduced to ϕ', if $\phi' = \omega(\phi)$ for some reduction operator ω, or $\phi' = \phi$. Moreover, a decoder ϕ is called *completely reducible* if it can be reduced to trivial decoders.

b) Decoder Decomposition:

Definition 7: Given $M_X, M_Y \geq 1$, an $M_X \times M_Y$ decoder ϕ is called *decomposable* if there exist non-trivial decoders $\phi_0, \phi_1 \in \mathcal{F}_{M_X,M_Y}$ and $i \in \{0, 1\}$, such that for all $(m_X, m_Y) \in \lceil M_X \rfloor \times \lceil M_Y \rfloor$,

$$\phi(m_X, m_Y) = \phi_0(m_X, m_Y) \oplus \phi_1(m_X, m_Y) \oplus \bar{\imath}, \tag{14}$$

and

$$\mathcal{I}_X^{(i)}(\phi_0) \cap \mathcal{I}_X^{(i)}(\phi_1) = \mathcal{I}_Y^{(i)}(\phi_0) \cap \mathcal{I}_Y^{(i)}(\phi_1) = \varnothing, \tag{15}$$

where "$\oplus$" represents the elementwise "exclusive or" operation, and where, for each $M_X \times M_Y$ decoder ϕ and $i \in \{0, 1\}$, we have defined

$$\begin{aligned} \mathcal{I}_X^{(i)}(\phi) &\triangleq \{m_X \in \lceil M_X \rfloor : \exists\, m_Y' \in \lceil M_Y \rfloor, \phi(m_X, m_Y') = i\}, \\ \mathcal{I}_Y^{(i)}(\phi) &\triangleq \{m_Y \in \lceil M_Y \rfloor : \exists\, m_X' \in \lceil M_X \rfloor, \phi(m_X', m_Y) = i\}. \end{aligned} \tag{16}$$

Then, (14) is referred to as a decomposition of ϕ.

3) Decoder Families: We use $\mathcal{F}_{M_X,M_Y}$ to denote the collection of all $M_X \times M_Y$ decoders, and we define

$$\mathcal{F} \triangleq \bigcup_{M_X \geq 1, M_Y \geq 1} \mathcal{F}_{M_X,M_Y} \tag{17}$$

as the collection of all decoders. Each subset of $\mathcal{F}$ is then referred as a *decoder family*. For given $P_{XY}^{(0)}, P_{XY}^{(1)}$ and decoder family $\mathcal{H} \subset \mathcal{F}$, we use $\mathcal{E}[\mathcal{H}]$ to denote the error exponent region associated with $\mathcal{H}$, defined as

$$\mathcal{E}[\mathcal{H}] \triangleq \bigcup_{\phi \in \mathcal{H}} \mathcal{E}[\phi], \quad \text{for all } \mathcal{H} \subset \mathcal{F}.$$

In addition, we have the following definition.

Definition 8: Given $P_{XY}^{(0)}, P_{XY}^{(1)}$ and decoders families $\mathcal{H}, \mathcal{H}' \subset \mathcal{F}$, we use $\mathcal{H} \preceq \mathcal{H}'$ to indicate that $\mathcal{E}[\mathcal{H}] \subset \mathcal{E}[\mathcal{H}']$. Specifically, if $\mathcal{H} \preceq \mathcal{H}'$ for some $\mathcal{H}' \subset \mathcal{H}$, $\mathcal{H}'$ is called *sufficient* for $\mathcal{H}$.

The following simple fact is an immediate consequence of the definition.

Fact 2: The relation "$\preceq$" is transitive, i.e., for all decoder families $\mathcal{H}_0, \mathcal{H}_1$ and $\mathcal{H}_2$, if $\mathcal{H}_0 \preceq \mathcal{H}_1$ and $\mathcal{H}_1 \preceq \mathcal{H}_2$, then $\mathcal{H}_0 \preceq \mathcal{H}_2$. In addition, given $\mathcal{H}_0, \mathcal{H}_1 \subset \mathcal{F}$ with $\mathcal{H}_0 \preceq \mathcal{H}_1$, we have $(\mathcal{H}_0 \cup \mathcal{H}') \preceq (\mathcal{H}_1 \cup \mathcal{H}')$ for all $\mathcal{H}' \subset \mathcal{F}$.

D. Auxiliary Results

Several useful auxiliary results are listed as follows.

Lemma 1 ([13, Lemma 2.6, Lemma 2.12]): Given an alphabet $\mathcal{Z}$ and $n \geq 1$, we have

$$\left|\hat{\mathcal{P}}_n^{\mathcal{Z}}\right| \leq (n+1)^{|\mathcal{Z}|}. \tag{18}$$

In addition, suppose Z^n is i.i.d. generated from some $P_Z \in \mathcal{P}^{\mathcal{Z}}$, then for each $Q_Z \in \hat{\mathcal{P}}_n^{\mathcal{Z}}$, we have

$$\begin{aligned} &(n+1)^{-|\mathcal{Z}|} \cdot \exp(-nD(Q_Z \| P_Z)) \\ &\qquad \leq \mathbb{P}\left\{Z^n \in \mathcal{T}_{Q_Z}^n\right\} \leq \exp(-nD(Q_Z \| P_Z)), \end{aligned} \tag{19}$$

and, for each $\eta > 0$,

$$\mathbb{P}\left\{Z^n \in \mathcal{T}_{P_Z;\eta}^n\right\} \geq 1 - \frac{|\mathcal{Z}|}{4n\eta^2}. \tag{20}$$

Lemma 2 ([13, Lemma 5.1]): Given an alphabet $\mathcal{Z}$, $P_Z \in \operatorname{relint}(\mathcal{P}^{\mathcal{Z}})$, and a sequence $\{d_n\}$ of positive integers with $d_n = o(n)$, there exists a sequence $\epsilon_n = o(1)$, such that for all $\mathcal{S}_Z \subset \mathcal{Z}^n$, the sequence Z^n are i.i.d. generated from P_Z satisfies

$$\mathbb{P}\left\{Z^n \in \mathcal{N}_{\mathrm{H}}^{d_n}(\mathcal{S}_Z)\right\} \leq \mathbb{P}\left\{Z^n \in \mathcal{S}_Z\right\} \cdot \exp(-n\epsilon_n)$$

where $\mathcal{N}_{\mathrm{H}}^{d}(\cdot)$ denotes the Hamming d-neighborhood as defined in Definition 2, and where $\operatorname{relint}(\cdot)$ denotes the relative interior.

Lemma 3 (Blowing up lemma [12], [13, Lemma 5.4]): Given an alphabet $\mathcal{Z}$ and sequence $\epsilon_n = o(1)$, there exist a sequence $\{d_n\}$ of positive integers with $d_n = o(n)$, and a sequence $\nu_n = o(1)$, such that for all given $n \geq 1$, $\mathcal{S}_Z \subset \mathcal{Z}^n$, $P_Z \in \mathcal{P}^{\mathcal{Z}}$, and Z^n i.i.d. generated from P_Z, if

$$\mathbb{P}\left\{Z^n \in \mathcal{S}_Z\right\} \geq \exp(-n\epsilon_n),$$

then

$$\mathbb{P}\left\{Z^n \in \mathcal{N}_{\mathrm{H}}^{d_n}(\mathcal{S}_Z)\right\} \geq 1 - \nu_n.$$

III. Main Results

In this section, we illustrate the geometric structure associated with DHT with constant-bit communication constraints, and provide characterizations of the error exponent region. Throughout our analyses, we assume that all entries of underlying distributions are positive, i.e.,

$$P_{XY}^{(0)}, P_{XY}^{(1)} \in \operatorname{relint}(\mathcal{P}^{\mathcal{X}\times\mathcal{Y}}), \tag{21}$$

where $\operatorname{relint}(\cdot)$ denotes the relative interior.

A. Optimality of Type-based Encoders

We first demonstrate that, the type-based encoders are asymptotically optimal for the broader class of DHT problems with zero-rate communication constraints, i.e., when $\|f_n\|$ and $\|g_n\|$ do not increase exponentially over n. To begin, we introduce the following fact, a proof of which is provided in Appendix B.

Fact 3: For each $i \in \{0,1\}$, under $\mathsf{H} = i$, the probability of observing sequences with marginal types $(Q_X, Q_Y) \in \hat{\mathcal{P}}_n^{\mathcal{X}} \times \hat{\mathcal{P}}_n^{\mathcal{Y}}$ is

$$\begin{aligned}&\mathbb{P}\left\{(\hat{P}_{X^n}, \hat{P}_{Y^n}) = (Q_X, Q_Y)\middle|\mathsf{H} = i\right\}\\&\quad= \exp(-n(D_i^*(Q_X, Q_Y) + o(1))).\end{aligned} \tag{22}$$

In addition, we have the following useful lemma. A proof is provided in Appendix C.

Lemma 4: Given zero-rate encoders f_n and g_n, there exist mappings $\theta_X \colon \mathcal{P}^{\mathcal{X}} \to \mathcal{M}_X^{(n)}$ and $\theta_Y \colon \mathcal{P}^{\mathcal{Y}} \to \mathcal{M}_Y^{(n)}$, such that

$$\begin{aligned}&\mathbb{P}\left\{f_n(X^n) = \theta_X(Q_X), g_n(Y^n) = \theta_Y(Q_Y)|\mathsf{H} = i\right\}\\&\geq \exp(-n \cdot (D_i^*(Q_X, Q_Y) + \epsilon_n)),\end{aligned} \tag{23}$$

for $i \in \{0,1\}$ and $(Q_X, Q_Y) \in \hat{\mathcal{P}}_n^{\mathcal{X}} \times \hat{\mathcal{P}}_n^{\mathcal{Y}}$, with $\epsilon_n = o(1)$, where D_0^* and D_1^* are as defined in (9).

Remark 1: From Fact 3, if $\mathsf{H} = i$, the probability of observing sequences with marginal types $(Q_X, Q_Y) \in \hat{\mathcal{P}}_n^{\mathcal{X}} \times \hat{\mathcal{P}}_n^{\mathcal{Y}}$ is $\exp(-nD_i^*(Q_X, Q_Y) + o(n))$, which corresponds to the right-hand side of (23). Therefore, Lemma 4 states that, each zero-rate encoder pair (f_n, g_n) has similar behaviors as the type-based encoders that map the observed sequences x^n and y^n to $\theta_X(\hat{P}_{x^n})$ and $\theta_Y(\hat{P}_{y^n})$, respectively.

Then, the following result establishes that, the performance of DHT can be improved via replacing original encoders with some type-based encoders, no matter what decoder is used.

Theorem 1: For a given $n \geq 1$ and zero-rate encoders f_n and g_n with ranges $\mathcal{M}_X^{(n)}$ and $\mathcal{M}_Y^{(n)}$, there exist type-based encoders $\tilde{f}_n, \tilde{g}_n$ with the same ranges as f_n, g_n, respectively, such that, for each decoder $\phi_n \colon \mathcal{M}_X^{(n)} \times \mathcal{M}_Y^{(n)} \to \{0,1\}$, we have

$$\pi_i(\tilde{\mathcal{C}}_n) \leq \pi_i(\mathcal{C}_n) \cdot \exp(n\zeta_n), \quad \text{for } i = 0, 1,$$

with $\zeta_n = o(1)$, where we have defined the coding schemes $\mathcal{C}_n \triangleq (f_n, g_n, \phi_n)$ and $\tilde{\mathcal{C}}_n \triangleq (\tilde{f}_n, \tilde{g}_n, \phi_n)$.

Remark 2: The optimality of type-based decision in non-distributed hypothesis testing can be established by a more straightforward argument, see, e.g., [14, Lemma 3.5.3]. Specifically, suppose n i.i.d samples $x^n \in \mathcal{X}^n$ are generated by $P_X^{(\mathsf{H})}$, and $f_n(x^n)$ is used as our decision for $\mathsf{H} \in \{0,1\}$, where $f_n \colon \mathcal{X}^n \to \{0,1\}$. Then, there exists a type-based decision $\tilde{f}_n \colon \mathcal{X}^n \to \{0,1\}$ such that

$$\pi_i(\tilde{f}_n) \leq 2 \cdot \pi_i(f_n), \quad \text{for } i \in \{0,1\},$$

where $\pi_0(\cdot)$ and $\pi_1(\cdot)$ denote the type-I error and type-II error for corresponding decision functions, respectively. It is also easy to verify that both Neyman–Pearson test [15] and Hoeffding's test [16] depend only on the types. In particular, Neyman–Pearson test depends only on the empirical mean of log-likelihood ratio $\log \frac{P_X^{(0)}(x)}{P_X^{(1)}(x)}$, see, e.g., [17, Theorem 11.7.1]. And, when only $P_X^{(0)}$ is available but $P_X^{(1)}$ is unknown, the resulting Hoeffding's test depends only on the KL divergence $D\big(\hat{P}_{x^n}\big\|P_X^{(0)}\big)$, which is also a function of the type $\hat{P}_{x^n}$.

Proof: To begin, note that from Fact 3, there exists some $\epsilon_n = o(1)$, such that for each $i \in \{0,1\}$, we have

$$\begin{aligned}&\mathbb{P}\left\{X^n \in \mathcal{T}_{Q_X}^n, Y^n \in \mathcal{T}_{Q_Y}^n|\mathsf{H} = i\right\}\\&\quad\leq \exp(-n(D_i^*(Q_X, Q_Y) - \epsilon_n)).\end{aligned} \tag{24}$$

In addition, we construct the type-based encoders $\tilde{f}_n$, $\tilde{g}_n$ such that

$$\tilde{f}_n(x^n) \triangleq \theta_X(\hat{P}_{x^n}), \quad \tilde{g}_n(y^n) \triangleq \theta_Y(\hat{P}_{y^n}) \tag{25}$$

for all $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$, where $\theta_X(\cdot)$ and $\theta_Y(\cdot)$ are as defined in Lemma 4. We also define

$$\Gamma_i^n \triangleq \{(Q_X, Q_Y) \in \hat{\mathcal{P}}_n^{\mathcal{X}} \times \hat{\mathcal{P}}_n^{\mathcal{Y}} \colon \phi_n(\theta_X(Q_X), \theta_Y(Q_Y)) \neq i\}$$

for $i = 0, 1$ and $n \geq 1$.

Then, it can be verified that for given sequences $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$, we have $\phi_n(\tilde{f}(x^n), \tilde{g}(y^n)) \neq i$ if and only if $(\hat{P}_{x^n}, \hat{P}_{y^n}) \in \Gamma_i^n$. Therefore, the error of the type-based coding scheme $\tilde{\mathcal{C}}_n$ can be written as

$$\begin{aligned}\pi_i(\tilde{\mathcal{C}}_n) &= \mathbb{P}\left\{\phi_n(\tilde{f}_n(X^n), \tilde{g}_n(Y^n)) \neq i\middle|\mathsf{H} = i\right\}\\&= \mathbb{P}\left\{(\hat{P}_{X^n}, \hat{P}_{Y^n}) \in \Gamma_i^n\middle|\mathsf{H} = i\right\}\\&= \sum_{(Q_X, Q_Y) \in \Gamma_i^n} \mathbb{P}\left\{X^n \in \mathcal{T}_{Q_X}^n, Y^n \in \mathcal{T}_{Q_Y}^n\middle|\mathsf{H} = i\right\}\\&\leq \sum_{(Q_X, Q_Y) \in \Gamma_i^n} \exp(-n \cdot (D_i^*(Q_X, Q_Y) - \epsilon_n)),\end{aligned} \tag{26}$$

where the inequality follows from (24).

If Γ_i^n is empty, then $\pi_i(\tilde{\mathcal{C}}_n) = 0 \leq \pi_i(\mathcal{C}_n)$ is trivially true. Otherwise, for each $n \geq 1$, let us define[3]

$$(Q_X^{(i)}, Q_Y^{(i)}) \triangleq \operatorname*{arg\,min}_{(Q_X, Q_Y) \in \Gamma_i^n} D_i^*(Q_X, Q_Y), \tag{27}$$

[3]For convenience, the dependencies of $Q_X^{(i)}, Q_Y^{(i)}$ on n are omitted from the notations.

and from (26) we have

$$\begin{aligned}\pi_i(\tilde{\mathcal{C}}_n) &\le (n+1)^{|\mathcal{X}|+|\mathcal{Y}|}\exp(-nD_i^*(Q_X^{(i)},Q_Y^{(i)})-\epsilon_n)\\ &= \exp(-n\cdot(D_i^*(Q_X^{(i)},Q_Y^{(i)})-\epsilon_n')),\end{aligned}\tag{28}$$

where the first inequality follows from the fact that

$$\begin{aligned}|\Gamma_i^n| \le \left|\hat{\mathcal{P}}_n^{\mathcal{X}}\times\hat{\mathcal{P}}_n^{\mathcal{Y}}\right| &\le (n+1)^{|\mathcal{X}|}\cdot(n+1)^{|\mathcal{Y}|}\\ &= (n+1)^{|\mathcal{X}|+|\mathcal{Y}|},\end{aligned}\tag{29}$$

and where

$$\epsilon_n' \triangleq \epsilon_n + \frac{(|\mathcal{X}|+|\mathcal{Y}|)\log(n+1)}{n}\tag{30}$$

satisfies $\epsilon_n' = o(1)$.

Moreover, from the definition (27) of $Q_X^{(i)}, Q_Y^{(i)}$, we have

$$\phi_n(\theta_X(Q_X^{(i)}),\theta_Y(Q_Y^{(i)})) \ne i.$$

As a result, from Lemma 4 we can obtain, for $i=0,1$,

$$\begin{aligned}&\pi_i(\mathcal{C}_n)\\ &\quad= \mathbb{P}\left\{\phi_n(f_n(X^n),g_n(Y^n))\ne i|\mathsf{H}=i\right\}\\ &\quad\ge \mathbb{P}\left\{f_n(X^n)=\theta_X(Q_X^{(i)}),g_n(Y^n)=\theta_Y(Q_Y^{(i)})\middle|\mathsf{H}=i\right\}\\ &\quad\ge \exp(-n\cdot(D_i^*(Q_X^{(i)},Q_Y^{(i)})+\xi_n^{(i)})\end{aligned}\tag{31}$$

with $\xi_n^{(i)} = o(1)$. Therefore, from (28) and (31) we have,

$$\pi_i(\tilde{\mathcal{C}}_n) \le \pi_i(\mathcal{C}_n)\cdot\exp(n\zeta_n), \quad \text{for } i=0,1,$$

where

$$\zeta_n \triangleq \epsilon_n + (\xi_n^{(0)}\vee\xi_n^{(1)}) = o(1).$$

■

B. Geometric Characterization

For DHT problem with communication constraints $(0_{M_X}, 0_{M_Y})$, we further illustrate that the error exponent region $\mathcal{E}(0_{M_X}, 0_{M_Y})$ can be characterized as a geometric problem of separating two sets in $\mathcal{P}_\star$. For convenience, in the following discussions we will assume that $M_X \ge M_Y$, and the result for $M_X < M_Y$ can be obtained by symmetry arguments.

To begin, we introduce the following fact, of which a proof is provide in Appendix D.

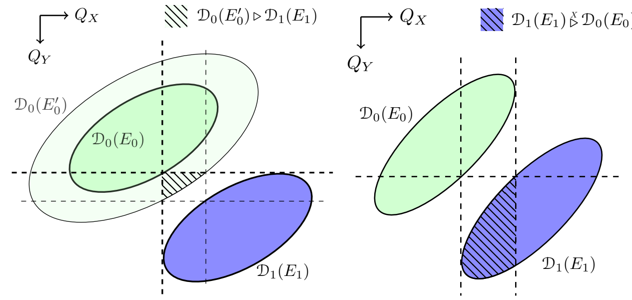
Fact 4: For all $P_{XY}^{(0)}, P_{XY}^{(1)} \in \mathcal{P}^{\mathcal{X}\times\mathcal{Y}}$ and $M_X, M_Y \ge 1$, we have

$$\mathcal{E}(0_{M_X}, 0_{M_Y}) = \mathcal{E}[\mathcal{F}_{M_X,M_Y}] = \mathcal{E}[\mathcal{H}],$$

where $\mathcal{H}$ is any decoder family sufficient for $\mathcal{F}_{M_X,M_Y}$ (cf. Definition 8).

Therefore, it suffices to construct a sufficient decoder family $\mathcal{H}$, and then investigate the region $\mathcal{E}[\phi]$ for each $\phi \in \mathcal{H}$. Before discussing the construction of decoder families, we first characterize the region $\mathcal{E}[\phi]$ for each given ϕ. To this end, the notion of separability on $\mathcal{P}_\star$ will be useful.

Definition 9: Given $M_X, M_Y \ge 1$ and a decoder $\phi \in \mathcal{F}_{M_X,M_Y}$, a pair of disjoint subsets $(\mathcal{A}_0, \mathcal{A}_1)$ of $\mathcal{P}_\star$ is *separable*



(a) $\varphi_{2,2}$ and $\varphi_{3,3}$ (b) $\varphi_{3,2}$

Fig. 2. Geometric interpretation for achievable error exponent pairs under different decoders, with each point representing a pair of marginal distributions $(Q_X, Q_Y) \in \mathcal{P}_\star$.

under ϕ, if there exist mappings $\theta_X\colon \mathcal{P}^{\mathcal{X}} \to \lceil M_X \rfloor$ and $\theta_Y\colon \mathcal{P}^{\mathcal{Y}} \to \lceil M_Y \rfloor$, such that for both $i \in \{0,1\}$,

$$\phi(\theta_X(Q_X),\theta_Y(Q_Y)) = i, \quad \text{for all } (Q_X,Q_Y)\in\mathcal{A}_i.\tag{32}$$

Then, our main result is as follows. A proof is provided in Appendix E.

Theorem 2: For each $\phi \in \mathcal{F}$, we have

$$\mathcal{E}[\phi] = \{(E_0,E_1)\colon (\mathcal{D}_0(E_0),\mathcal{D}_1(E_1)) \text{ is separable under } \phi\},$$

where $\mathcal{D}_0(\cdot)$ and $\mathcal{D}_1(\cdot)$ are as defined in (8). In addition, each exponent pair (E_0, E_1) in the interior of $\mathcal{E}[\phi]$ can be achieved by the coding schemes $\{(f_n, g_n, \phi)\}_{n\ge1}$ with $f_n(x^n) \triangleq \theta_X(\hat{P}_{x^n}), g_n(y^n) \triangleq \theta_Y(\hat{P}_{y^n})$, where θ_X and θ_Y are as defined in Definition 9.

Remark 3: By using a similar argument, we can show that under zero-rate communication constraints $(R_X, R_Y) = (0,0)$, the error exponent region is

$$\mathcal{E}(0,0) = \{(E_0,E_1)\colon \mathcal{D}_0(E_0)\cap\mathcal{D}_1(E_1) = \varnothing\},\tag{33}$$

which coincides with the classical results demonstrated in, e.g., [3, Theorem 6], [4, Theorem 5.5]. Furthermore, note that (33) also corresponds to a limiting case of Theorem 2, and we have

$$\mathcal{E}[\mathcal{F}_{M_X,M_Y}] \to \{(E_0,E_1)\colon \mathcal{D}_0(E_0)\cap\mathcal{D}_1(E_1) = \varnothing\}$$

as $M_X \to \infty, M_Y \to \infty$.

The relation between error exponent pair (E_0, E_1) and the separability of $(\mathcal{D}_0(E_0), \mathcal{D}_1(E_1))$ is illustrated in Fig. 2. In this figure, the x-axis and y-axis represent the marginal distributions of X and Y, respectively, and each point corresponds to a pair of marginal distributions $(Q_X, Q_Y) \in \mathcal{P}_\star$. Let us first consider the DHT problem with one-bit compression, with $\varphi_{2,2} \leftrightarrow$ used as the decoder. Then, it can be easily verified from Fig. 2a that $(\mathcal{D}_0(E_0), \mathcal{D}_1(E_1))$ is separable under $\varphi_{2,2}$. Therefore, it follows from Theorem 2 that $(E_0, E_1) \in \mathcal{E}[\varphi_{2,2}]$, and thus for all $\epsilon > 0$, the error exponent pair $(E_0 - \epsilon, E_1)$ is achievable under $\varphi_{2,2}$. Moreover, for all $\epsilon > 0$, $\mathcal{D}_0(E_0 + \epsilon)$ is a strict superset of $\mathcal{D}_0(E_0)$, making $(\mathcal{D}_0(E_0 + \epsilon), \mathcal{D}_1(E_1))$ inseparable under $\varphi_{2,2}$. Hence, with the type-II error exponent E_1 fixed, E_0 is the optimal type-I error exponent under $\varphi_{2,2}$. In addition, when both nodes are allowed to transmit one-trit messages with $\varphi_{3,3} \leftrightarrow$ used as the decoder, the optimal type-I error exponent can be improved to $E_0' > E_0$, as illustrated in the figure. Compared with the one-bit setting, it can be noted that the two additional symbols are used to encode the hatched area $\mathcal{D}_0(E_0') \triangleright \mathcal{D}_1(E_1)$, such that $(\mathcal{D}_0(E_0'), \mathcal{D}_1(E_1))$ is still separable, where the operator "$\triangleright$" is as defined in Definition 3. Similarly, Fig. 2b illustrates the geometric characterization when two nodes N_X and N_Y have one-trit and one-bit communication budgets, respectively, with $\varphi_{3,2} \leftrightarrow$  used as the decoder.

The above geometric interpretations also suggest a recursive property of the separability under threshold decoders. For example, in Fig. 2b, $(\mathcal{D}_0(E_0), \mathcal{D}_1(E_1))$ is separable under $\varphi_{3,2}$, if and only if $\mathcal{D}_0(E_0)$ and $\mathcal{D}_1(E_1) \overset{x}{\triangleright} \mathcal{D}_0(E_0)$ (shown in hatched) are separable under $\varphi_{2,2}$, where "$\overset{x}{\triangleright}$" is as defined in (12). Such recursive properties can be further generalized as the following proposition, of which a proof is provided in Appendix F.

Proposition 1: Suppose $\mathcal{A}_0, \mathcal{A}_1 \subset \mathcal{P}_\star$, and ϕ is a reducible decoder. For each $i \in \{0, 1\}$, if $\omega_X^{(i)}(\phi)$ exists, then $(\mathcal{A}_0, \mathcal{A}_1)$ is separable under ϕ if and only if $(\mathcal{A}_0 \overset{x}{\triangleright} \mathcal{A}_{\bar{\imath}}, \mathcal{A}_1 \overset{x}{\triangleright} \mathcal{A}_{\bar{\imath}})$ is separable under $\omega_X^{(i)}(\phi)$. Similarly, if $\omega_Y^{(i)}(\phi)$ exists, then $(\mathcal{A}_0, \mathcal{A}_1)$ is separable under ϕ if and only if $(\mathcal{A}_0 \overset{y}{\triangleright} \mathcal{A}_{\bar{\imath}}, \mathcal{A}_1 \overset{y}{\triangleright} \mathcal{A}_{\bar{\imath}})$ is separable under $\omega_Y^{(i)}(\phi)$, where "$\overset{x}{\triangleright}$" and "$\overset{y}{\triangleright}$" are as defined in (12) and (13), respectively.

With Proposition 1, the error exponent region under threshold decoders can be summarized as follows, of which a proof is provided in Appendix G.

Theorem 3: Given $M_X \geq M_Y \geq 1$, the error exponent regions under $M_X \times M_Y$ threshold decoders are

$$\mathcal{E}[\varphi_{M_X,M_Y}] = \{(E_0, E_1) \colon \mathcal{D}_0(E_0) \triangleright_{M_Y} \mathcal{D}_1(E_1) = \varnothing\},$$
$$\mathcal{E}[\bar{\varphi}_{M_X,M_Y}] = \{(E_0, E_1) \colon \mathcal{D}_1(E_1) \triangleright_{M_Y} \mathcal{D}_0(E_0) = \varnothing\}$$

if $M_X = M_Y$, and

$$\begin{aligned}\mathcal{E}[\varphi_{M_X,M_Y}] &= \mathcal{E}[\bar{\varphi}_{M_X,M_Y}] \\ &= \{(E_0, E_1) \colon \mathcal{D}_0(E_0) \triangleright_{M_Y} \big(\mathcal{D}_1(E_1) \overset{x}{\triangleright} \mathcal{D}_0(E_0)\big) = \varnothing\}\end{aligned}$$

if $M_X > M_Y$, where the operators "$\triangleright_k$" and "$\overset{x}{\triangleright}$" are as defined in Definition 3.

From Fact 4 and Theorem 3, we can readily obtain an inner bound of $\mathcal{E}(0_{M_X}, 0_{M_Y})$ as

$$\begin{aligned}\mathcal{E}(0_{M_X}, 0_{M_Y}) &= \mathcal{E}[\mathcal{F}_{M_X,M_Y}] \\ &\supset (\mathcal{E}[\varphi_{M_X,M_Y}] \cup \mathcal{E}[\bar{\varphi}_{M_X,M_Y}]) \,. \end{aligned} \tag{34}$$

This bound can be tight under certain circumstances, as we will illustrate later.

In addition, the following result illustrates that, when the communication constraint on one node is much stronger than that on the other node, the performance of DHT is dominated by this stronger constraint. A proof is provided in Appendix H.

Proposition 2: For given $M_Y \geq 1$, $M_X > 2^{M_Y}$, and $R_X \in [0, \infty)$, we have

$$\mathcal{E}(R_X, 0_{M_Y}) = \mathcal{E}(0_{M_X}, 0_{M_Y}) = \mathcal{E}(0_{2^{M_Y}}, 0_{M_Y}).$$

Therefore, without loss of generality we may assume that $M_Y \leq M_X \leq 2^{M_Y}$.

C. Construction of Sufficient Decoder Families

We then demonstrate a construction of sufficient decoder families. As a first step, we partition the collection $\mathcal{F}_{M_X,M_Y}$ of $M_X \times M_Y$ decoders as

$$\mathcal{F}_{M_X,M_Y} = \Omega_{M_X,M_Y} \cup \bar{\Omega}_{M_X,M_Y}, \tag{35}$$

where we have defined

$$\begin{aligned}\Omega_{M_X,M_Y} &\triangleq \{\phi \in \mathcal{F}_{M_X,M_Y} : \phi \text{ is completely reducible}\}, \\ \bar{\Omega}_{M_X,M_Y} &\triangleq \mathcal{F}_{M_X,M_Y} \setminus \Omega_{M_X,M_Y}. \end{aligned} \tag{36}$$

Then, we discuss the decoders in Ω_{M_X,M_Y} and $\bar{\Omega}_{M_X,M_Y}$ separately.

For the decoders in Ω_{M_X,M_Y}, we have the following useful characterization, a proof of which is provided in Appendix I.

Proposition 3: Let ϕ denote an $M_X \times M_Y$ decoder with $M_X, M_Y \geq 2$. Then, the following statements are equivalent:

S1) ϕ is completely reducible;
S2) each 2×2 subdecoder of ϕ is reducible;
S3) there exists a monotonic decoder ϕ' such that $\phi \simeq \phi'$.

As an immediate consequence of Proposition 3, note that the threshold decoders are monotonic decoders, and thus are completely reducible. Furthermore, we have the following consequence of Proposition 3, which demonstrates the sufficiency of threshold decoders.

Lemma 5: For given $M_X \geq M_Y \geq 1$, we have $\Omega_{M_X,M_Y} \preceq \Phi_{M_X,M_Y}$, where we have defined

$$\Phi_{M_X,M_Y} \triangleq \{\varphi_{M_X,M_Y}, \bar{\varphi}_{M_X,M_Y}\}. \tag{37}$$

A proof of Lemma 5 is provided in Appendix J, and makes use of the following simple fact.

Fact 5: If $\phi \simeq \phi'$, then $\mathcal{E}[\phi] = \mathcal{E}[\phi']$. If ϕ' is a subdecoder of ϕ, then $\{\phi'\} \preceq \{\phi\}$.

Remark 4: If $M_X > M_Y$, we have $\varphi_{M_X,M_Y} \simeq \bar{\varphi}_{M_X,M_Y}$, and it follows from Fact 5 that

$$\Phi_{M_X,M_Y} = \{\varphi_{M_X,M_Y}, \bar{\varphi}_{M_X,M_Y}\} \preceq \{\varphi_{M_X,M_Y}\}.$$

Therefore, the statement of Lemma 5 can be further refined to $\Omega_{M_X,M_Y} \preceq \{\varphi_{M_X,M_Y}\}$ for the case $M_X > M_Y$.

Moreover, we have the following characterization for decoders in $\bar{\Omega}_{M_X,M_Y}$, a proof of which is provided in Appendix K.

Proposition 4: For each decoder ϕ that is not completely reducible, there exists a unique irreducible decoder ϕ' that can be reduced from ϕ.

Specifically, for each ϕ that is not completely reducible, we refer to the ϕ' given by Proposition 4 as the *reduced form* of ϕ, denoted by $\omega^*(\phi)$. Then, we can further partition $\bar{\Omega}_{M_X,M_Y}$ as

$$\bar{\Omega}_{M_X,M_Y} = \bar{\Omega}^{(0)}_{M_X,M_Y} \cup \bar{\Omega}^{(1)}_{M_X,M_Y}, \tag{38}$$

where

$$\begin{aligned}\bar{\Omega}^{(0)}_{M_X,M_Y} &\triangleq \{\phi \in \bar{\Omega}_{M_X,M_Y} : \omega^*(\phi) \text{ is indecomposable}\}, \\ \bar{\Omega}^{(1)}_{M_X,M_Y} &\triangleq \{\phi \in \bar{\Omega}_{M_X,M_Y} : \omega^*(\phi) \text{ is decomposable}\}. \end{aligned} \tag{39}$$

In addition, we have the following lemma, a proof of which is provided in Appendix L.

Lemma 6: Given $M_X \geq M_Y \geq 1$, we have $\bar{\Omega}^{(1)}_{M_X,M_Y} \preceq (\Omega_{M_X,M_Y} \cup \bar{\Omega}^{(0)}_{M_X,M_Y})$, where $\bar{\Omega}^{(1)}$ is as defined in (37), and where Ω and $\bar{\Omega}^{(0)}$ are as defined in (39).

By applying Lemma 5 and Lemma 6, the following theorem provides a construction of sufficient decoder families.

Theorem 4: Given $M_X \geq M_Y \geq 1$, the decoder family $\Phi_{M_X,M_Y} \cup \bar{\Omega}^{(0)}_{M_X,M_Y}$ is sufficient for $\mathcal{F}_{M_X,M_Y}$, where Φ and $\bar{\Omega}^{(0)}$ are as defined in (37) and (39), respectively.

Proof: From (35) and (38), we have

$$\mathcal{F}_{M_X,M_Y} = \Omega_{M_X,M_Y} \cup \bar{\Omega}^{(0)}_{M_X,M_Y} \cup \bar{\Omega}^{(1)}_{M_X,M_Y} \tag{40}$$
$$\preceq \Omega_{M_X,M_Y} \cup \bar{\Omega}^{(0)}_{M_X,M_Y} \tag{41}$$
$$\preceq \Phi_{M_X,M_Y} \cup \bar{\Omega}^{(0)}_{M_X,M_Y} \tag{42}$$

where to obtain (41) we have used Lemma 5 and Fact 2, and where to obtain (42) we have used Lemma 6. ∎

D. Error Exponent Regions

We then provide several conditions where the inner bound (34) is tight, i.e., the threshold decoders Φ_{M_X,M_Y} are sufficient for $\mathcal{F}_{M_X,M_Y}$. To this end, we first introduce the following lemma, a proof of which is provided in Appendix M.

Lemma 7: Given $M_X \geq M_Y \geq 1$ with $(M_X - 2)(M_Y - 2) < 2$, we have

$$\mathcal{E}(0_{M_X}, 0_{M_Y}) = \mathcal{E}[\varphi_{M_X,M_Y}] \cup \mathcal{E}[\bar{\varphi}_{M_X,M_Y}]. \tag{43}$$

From Lemma 7, it suffices to consider threshold decoders for the one-bit compression settings with $M_X \geq M_Y = 2$ and two-sided one-trit compression ($M_X = M_Y = 3$). As a straightforward corollary, we first revisit the *two-sided one-bit compression* setting, i.e., $M_X = M_Y = 2$, with the following characterization of corresponding exponent region $\mathcal{E}(0_2, 0_2)$.

Corollary 1 ([3, Theorem 5], [4, Theorem 5.6]): With one-bit compression for both nodes, we have

$$\mathcal{E}(0_2, 0_2) = \mathcal{E}[\varphi_{2,2}] \cup \mathcal{E}[\bar{\varphi}_{2,2}],$$

where $\mathcal{E}[\varphi_{2,2}]$ and $\mathcal{E}[\bar{\varphi}_{2,2}]$ are as given by Theorem 3, and can be represented as

$$\mathcal{E}[\varphi_{2,2}] = \{(E_0, E_1)\colon \mathcal{D}_0(E_0) \cap \mathcal{B}_1(E_1) = \varnothing\},$$
$$\mathcal{E}[\bar{\varphi}_{2,2}] = \{(E_0, E_1)\colon \mathcal{B}_0(E_0) \cap \mathcal{D}_1(E_1) = \varnothing\},$$

where for $i \in \{0, 1\}$ and $t \geq 0$, we have defined

$$\mathcal{B}_i(t) \triangleq \{(Q_X, Q_Y)\colon D(Q_X \| P^{(i)}_X) < t, D(Q_Y \| P^{(i)}_Y) < t\}.$$

Remark 5: It has been shown in [3] that the same result can be established when we relax the strict positive assumption (21) to $D(P^{(0)}_{XY} \| P^{(1)}_{XY}) < \infty$.

Similarly, the error exponent region with *two-sided one-trit compression* is as follows.

Corollary 2: The exponent region of $M_X = M_Y = 3$ is

$$\mathcal{E}(0_3, 0_3) = \mathcal{E}[\varphi_{3,3}] \cup \mathcal{E}[\bar{\varphi}_{3,3}],$$

where $\mathcal{E}[\varphi_{3,3}]$ and $\mathcal{E}[\bar{\varphi}_{3,3}]$ are as given by Theorem 3.

In addition, by combining Proposition 2 and Lemma 7, we can establish the error exponent region for *one-sided one-bit compression*.

Corollary 3: For all $M \geq 3$ and $R \in [0, \infty)$, we have

$$\mathcal{E}(R, 0_2) = \mathcal{E}(0_M, 0_2) = \mathcal{E}(0_3, 0_2) = \mathcal{E}[\varphi_{3,2}].$$

Remark 6: It is worth noting that in general we have $\mathcal{E}(0_2, 0_2) \subsetneq \mathcal{E}(0_3, 0_2) = \mathcal{E}(R, 0_2)$. Therefore, when one distributed node can only transmit one bit, to obtain the optimal performance, the other node is required to transmit at least a one-trit message. This situation differs from the one appeared in investigating the optimal type-II error exponent E_1 with type-I error π_0 constrained by a constant (cf. [2, Corollary 7]), where it requires only a one-bit message sent from the other node to obtain the optimal performance.

Finally, when the observations at both nodes are conditionally independent given $\mathsf{H} = 0$ or $\mathsf{H} = 1$, the inner bound (34) is tight for all $M_X \geq M_Y \geq 1$, illustrated as follows. A proof is provide in Appendix N.

Theorem 5: Suppose $P^{(i)}_{XY} = P^{(i)}_X P^{(i)}_Y$ for some $i \in \{0, 1\}$, then we have

$$\mathcal{E}(0_{M_X}, 0_{M_Y}) = \mathcal{E}[\varphi_{M_X,M_Y}] \cup \mathcal{E}[\bar{\varphi}_{M_X,M_Y}],$$

for all $M_X \geq M_Y \geq 1$, where $\mathcal{E}[\varphi_{M_X,M_Y}]$ and $\mathcal{E}[\bar{\varphi}_{M_X,M_Y}]$ are as given by Theorem 3.

Appendix A
Proof of Fact 1

Given $\alpha \in (0, 1)$ and $i \in \{0, 1\}$, for all (Q_X, Q_Y) and $(R_X, R_Y) \in \mathcal{P}_\star$, we have

$$\begin{aligned} &D^*_i(\alpha Q_X + (1-\alpha) R_X, \alpha Q_Y + (1-\alpha) Q_Y) \\ &\leq D(\alpha \tilde{Q}_{XY} + (1-\alpha) \tilde{R}_{XY} \| P^{(i)}_{XY}) \\ &\leq \alpha D(\tilde{Q}_{XY} \| P^{(i)}_{XY}) + (1-\alpha) D(\tilde{R}_{XY} \| P^{(i)}_{XY}) \\ &= \alpha D^*_i(Q_X, Q_Y) + (1-\alpha) D^*_i(R_X, R_Y), \end{aligned}$$

where we have chosen $\tilde{Q}_{XY}$ and $\tilde{R}_{XY}$ such that

$$[\tilde{Q}_{XY}]_X = Q_X, [\tilde{Q}_{XY}]_Y = Q_Y, \tag{44}$$
$$[\tilde{R}_{XY}]_X = R_X, [\tilde{R}_{XY}]_Y = R_Y, \tag{45}$$

and

$$D(\tilde{Q}_{XY} \| P^{(i)}_{XY}) = D^*_i(Q_X, Q_Y),$$
$$D(\tilde{R}_{XY} \| P^{(i)}_{XY}) = D^*_i(R_X, R_Y).$$

∎

Appendix B
Proof of Fact 3

To begin, for each given marginal type pair $(Q_X, Q_Y) \in \hat{\mathcal{P}}^{\mathcal{X}}_n \times \hat{\mathcal{P}}^{\mathcal{Y}}_n$, let us define

$$\mathcal{R} \triangleq \{Q_{XY} \in \mathcal{P}^{\mathcal{X}\times\mathcal{Y}}\colon [Q_{XY}]_X = Q_X, [Q_{XY}]_Y = Q_Y\}$$
$$\hat{\mathcal{R}}_n \triangleq \{Q_{XY} \in \hat{\mathcal{P}}^{\mathcal{X}\times\mathcal{Y}}_n\colon [Q_{XY}]_X = Q_X, [Q_{XY}]_Y = Q_Y\}.$$

and

$$\tilde{Q}_{XY}^{(i)} \triangleq \arg\min_{Q_{XY}\in\mathcal{R}} D(Q_{XY}\|P_{XY}^{(i)}), \quad \text{for } i \in \{0,1\}.$$

Then, under the hypothesis $\mathsf{H} = i \in \{0,1\}$, the probability of observing a sequence (X^n, Y^n) with marginal types (Q_X, Q_Y) can be represented as

$$\begin{aligned}
&\mathbb{P}\left\{X^n \in \mathcal{T}_{Q_X}^n, Y^n \in \mathcal{T}_{Q_Y}^n \middle| \mathsf{H} = i\right\} \\
&\quad= \sum_{Q_{XY}\in\hat{\mathcal{R}}_n} \mathbb{P}\left\{(X^n, Y^n) \in \mathcal{T}_{Q_{XY}}^n \middle| \mathsf{H} = i\right\} \\
&\quad\le \sum_{Q_{XY}\in\hat{\mathcal{R}}_n} \exp(-nD(Q_{XY}\|P_{XY}^{(i)})) &(46)\\
&\quad\le \left|\hat{\mathcal{R}}_n\right| \cdot \exp(-nD_i^*(Q_X, Q_Y)) &(47)\\
&\quad\le \left|\hat{\mathcal{P}}_n^{\mathcal{X}\times\mathcal{Y}}\right| \cdot \exp(-nD_i^*(Q_X, Q_Y)) \\
&\quad\le (n+1)^{|\mathcal{X}||\mathcal{Y}|} \exp(-nD_i^*(Q_X, Q_Y)) &(48)\\
&\quad= \exp(-n(D_i^*(Q_X, Q_Y) - \epsilon_n)), &(49)
\end{aligned}$$

where (46) follows from (19), where (47) follows from the definition of D_i^* [cf. (9)], where (48) follows from (18), and where we have defined

$$\epsilon_n \triangleq \frac{|\mathcal{X}||\mathcal{Y}|\log(n+1)}{n} = o(1). \tag{50}$$

In addition, for each $i \in \{0,1\}$, since $\hat{\mathcal{R}}_n$ is dense in $\mathcal{R}$, there exists $\hat{Q}_{XY}^{(n)} \in \hat{\mathcal{R}}_n$ satisfying

$$d_{\max}(\hat{Q}_{XY}^{(n)}, \tilde{Q}_{XY}^{(i)}) = o(1),$$

and it follows from the uniform continuity of KL divergence that

$$\begin{aligned}
&|D(\hat{Q}_{XY}^{(n)}\|P_{XY}^{(i)}) - D_i^*(Q_X, Q_Y)| \\
&\quad= |D(\hat{Q}_{XY}^{(n)}\|P_{XY}^{(i)}) - D(\tilde{Q}_{XY}^{(i)}\|P_{XY}^{(i)})| < \nu_n &(51)
\end{aligned}$$

for some $\nu_n = o(1)$.

Therefore, we obtain

$$\begin{aligned}
&\mathbb{P}\left\{X^n \in \mathcal{T}_{Q_X}^n, Y^n \in \mathcal{T}_{Q_Y}^n \middle| \mathsf{H} = i\right\} \\
&\quad= \sum_{Q_{XY}\in\hat{\mathcal{R}}_n} \mathbb{P}\left\{(X^n, Y^n) \in \mathcal{T}_{Q_{XY}}^n \middle| \mathsf{H} = i\right\} \\
&\quad\ge \mathbb{P}\left\{(X^n, Y^n) \in \mathcal{T}_{\hat{Q}_{XY}^{(n)}}^n \middle| \mathsf{H} = i\right\} \\
&\quad\ge (n+1)^{-|\mathcal{X}||\mathcal{Y}|} \cdot \exp(-nD(\hat{Q}_{XY}^{(n)}\|P_{XY}^{(i)})) &(52)\\
&\quad\ge (n+1)^{-|\mathcal{X}||\mathcal{Y}|} \cdot \exp(-n(D_i^*(Q_X, Q_Y) + \nu_n)) &(53)\\
&\quad= \exp(-n(D_i^*(Q_X, Q_Y) + \nu_n + \epsilon_n), &(54)
\end{aligned}$$

where to obtain (52) we have again used (19), to obtain (53) we have used (51), and where ϵ_n is as defined in (50).

Finally, (22) is obtained via combining (49) and (54). ■

APPENDIX C
PROOF OF LEMMA 4

For a given pair of marginal distributions $(Q_X, Q_Y) \in \hat{\mathcal{P}}_n^{\mathcal{X}} \times \hat{\mathcal{P}}_n^{\mathcal{Y}}$, we first define

$$\begin{aligned}
\mathcal{S}_X &\triangleq \{x^n \in \mathcal{X}^n : f_n(x^n) = \theta_X(Q_X)\}, \\
\mathcal{S}_Y &\triangleq \{y^n \in \mathcal{Y}^n : g_n(y^n) = \theta_Y(Q_Y)\}
\end{aligned}$$

and $\mathcal{S}_{XY} \triangleq \mathcal{S}_X \times \mathcal{S}_Y$, where for given f_n and g_n, we have defined $\theta_X : \mathcal{P}^{\mathcal{X}} \to \mathcal{M}_X^{(n)}$ and $\theta_Y : \mathcal{P}^{\mathcal{Y}} \to \mathcal{M}_Y^{(n)}$ such that for all $P_X \in \mathcal{P}^{\mathcal{X}}$ and $P_Y \in \mathcal{P}^{\mathcal{Y}}$,

$$\theta_X(P_X) \triangleq \arg\max_{m_X\in\mathcal{M}_X^{(n)}} \mathbb{P}\left\{f_n(X^n) = m_X \middle| X^n \sim P_X^{\otimes n}\right\}, \tag{55}$$

$$\theta_Y(P_Y) \triangleq \arg\max_{m_Y\in\mathcal{M}_Y^{(n)}} \mathbb{P}\left\{g_n(Y^n) = m_Y \middle| Y^n \sim P_Y^{\otimes n}\right\}. \tag{56}$$

By symmetry, it suffices to establish (23) for $i = 0$. To this end, let (X^n, Y^n) be i.i.d. generated from $P_{XY}^{(0)}$, and we define $Q_{XY} \in \mathcal{P}^{\mathcal{X}\times\mathcal{Y}}$ such that it satisfies $[Q_{XY}]_X = Q_X, [Q_{XY}]_Y = Q_Y$ and $D_0^*(Q_X, Q_Y) = D(Q_{XY}\|P_{XY}^{(0)})$. Then, we can equivalently express (23) as

$$\begin{aligned}
&\mathbb{P}\{(X^n, Y^n) \in \mathcal{S}_{XY}\} \\
&\quad\ge \exp(-n \cdot (D(Q_{XY}\|P_{XY}^{(0)}) + \epsilon_n)) &(57)
\end{aligned}$$

with $\epsilon_n = o(1)$.

We then illustrate that (57) holds, if there exists a sequence of positive integers $\{l_n\}_{n\ge1}$ with $l_n = o(n)$, such that for n sufficiently large, we have

$$\max_{\tilde{Q}_{XY}\in\mathcal{Q}_n} \gamma_n(\tilde{Q}_{XY}) \ge \frac{1}{2}, \tag{58}$$

where for each $n \ge 1$ and $\tilde{Q}_{XY} \in \hat{\mathcal{P}}_n^{\mathcal{X}\times\mathcal{Y}}$, we have defined

$$\gamma_n(\tilde{Q}_{XY}) \triangleq \frac{\left|\mathcal{T}_{\tilde{Q}_{XY}}^n \cap \mathcal{N}_{\mathrm{H}}^{l_n}(\mathcal{S}_{XY})\right|}{\left|\mathcal{T}_{\tilde{Q}_{XY}}^n\right|} \tag{59}$$

with $\mathcal{N}_{\mathrm{H}}^d(\cdot)$ denoting the Hamming d-neighborhood (cf. Definition 2), $\eta_n \triangleq n^{-\frac{1}{3}}$, and

$$\mathcal{Q}_n \triangleq \left\{\tilde{Q}_{XY} \in \hat{\mathcal{P}}_n^{\mathcal{X}\times\mathcal{Y}} : d_{\max}(\tilde{Q}_{XY}, Q_{XY}) \le \eta_n\right\}. \tag{60}$$

To see this, first note that from Lemma 2, we have

$$\begin{aligned}
&\mathbb{P}\{(X^n, Y^n) \in \mathcal{S}_{XY}\} \\
&\quad\ge \mathbb{P}\left\{(X^n, Y^n) \in \mathcal{N}_{\mathrm{H}}^{l_n}(\mathcal{S}_{XY})\right\} \cdot \exp(-n\epsilon_n') &(61)
\end{aligned}$$

for some $\epsilon_n' = o(1)$.

Moreover, from (58), for sufficiently large n, there exists $Q_{XY}' \in \mathcal{Q}_n$, such that $\gamma_n(Q_{XY}') \ge \frac{1}{2}$. As a result, we have

$$\begin{aligned}
&\mathbb{P}\left\{(X^n, Y^n) \in \mathcal{N}_{\mathrm{H}}^{l_n}(\mathcal{S}_{XY})\right\} \\
&\quad\ge \mathbb{P}\left\{(X^n, Y^n) \in \mathcal{T}_{Q_{XY}'}^n \cap \mathcal{N}_{\mathrm{H}}^{l_n}(\mathcal{S}_{XY})\right\} \\
&\quad= \mathbb{P}\left\{(X^n, Y^n) \in \mathcal{T}_{Q_{XY}'}^n\right\} \cdot \gamma_n(Q_{XY}') \\
&\quad\ge \frac{1}{2} \cdot \mathbb{P}\left\{(X^n, Y^n) \in \mathcal{T}_{Q_{XY}'}^n\right\} &(62)
\end{aligned}$$

where the equality follows from the fact that different sequences within a type class are equiprobable.

In addition, it follows from the definition of $\mathcal{Q}_n$ [cf. (60)] that $d_{\max}(Q'_{XY}, Q_{XY}) \leq \eta_n$. Hence, from the uniform continuity of KL divergence, there exists $\epsilon''_n = o(1)$ such that

$$\left| D(Q'_{XY} \| P^{(0)}_{XY}) - D(Q_{XY} \| P^{(0)}_{XY}) \right| < \epsilon''_n.$$

This implies that

$$\begin{aligned} &\mathbb{P}\left\{(X^n, Y^n) \in \mathcal{T}_{Q'_{XY}}\right\} \\ &\qquad \geq (n+1)^{-|\mathcal{X}||\mathcal{Y}|} \exp(-nD(Q'_{XY} \| P^{(0)}_{XY})) \\ &\qquad \geq (n+1)^{-|\mathcal{X}||\mathcal{Y}|} \exp(-n\epsilon''_n) \cdot \exp(-nD(Q_{XY} \| P^{(0)}_{XY})), \end{aligned} \tag{63}$$

where the first inequality follows from the lower bound in (19) for probability of a type class, see, e.g., [17, Theorem 11.1.4] or [13, Lemma 2.6].

Then, it can then be verified from (61), (62) and (63) that (57) holds with

$$\epsilon_n = \epsilon'_n + \epsilon''_n + \frac{1}{n} \log 2 + \frac{|\mathcal{X}||\mathcal{Y}|}{n} \log(n+1) = o(1).$$

Hence, it remains to establish (58). To this end, we turn to consider probabilities under the measure Q_{XY}, and let $(\tilde{X}^n, \tilde{Y}^n)$ be i.i.d. generated from Q_{XY}. Then, it follows from Lemma 1 that

$$\mathbb{P}\left\{(\tilde{X}^n, \tilde{Y}^n) \in \mathcal{T}^n_{Q_{XY};\eta_n}\right\} \geq 1 - \frac{|\mathcal{X}||\mathcal{Y}|}{4n\eta_n^2} = 1 - \frac{|\mathcal{X}||\mathcal{Y}|}{4n^{\frac{1}{3}}}. \tag{64}$$

Moreover, from (55) we have

$$\begin{aligned} \mathbb{P}\left\{\tilde{X}^n \in \mathcal{S}_X\right\} &= \mathbb{P}\left\{f_n(\tilde{X}^n) = \theta_X(Q_X)\right\} \\ &\geq \frac{1}{\|f_n\|} = \exp\left(-n \cdot \frac{\log \|f_n\|}{n}\right), \end{aligned} \tag{65}$$

and, similarly, from (56) we have

$$\mathbb{P}\left\{\tilde{Y}^n \in \mathcal{S}_Y\right\} \geq \exp\left(-n \cdot \frac{\log \|g_n\|}{n}\right). \tag{66}$$

Then, since f_n and g_n are with zero-rates, both $\frac{1}{n} \log \|f_n\|$ and $\frac{1}{n} \log \|g_n\|$ vanish as n tends to infinity. Therefore, it follows from Lemma 3 that there exist $d_n = o(n)$ and $\nu_n = o(1)$, such that

$$\mathbb{P}\left\{\tilde{X}^n \in \mathcal{N}^{d_n}_{\mathrm{H}}(\mathcal{S}_X)\right\} \geq 1 - \nu_n, \tag{67}$$

$$\mathbb{P}\left\{\tilde{Y}^n \in \mathcal{N}^{d_n}_{\mathrm{H}}(\mathcal{S}_Y)\right\} \geq 1 - \nu_n, \tag{68}$$

where $\mathcal{N}^d_{\mathrm{H}}(\cdot)$ denotes the d-Hamming neighborhood, as defined in Definition 2.

Let $l_n \triangleq 2d_n = o(n)$, and it follows from the fact $\mathcal{N}^{d_n}_{\mathrm{H}}(\mathcal{S}_X) \times \mathcal{N}^{d_n}_{\mathrm{H}}(\mathcal{S}_Y) \subset \mathcal{N}^{2d_n}_{\mathrm{H}}(\mathcal{S}_X \times \mathcal{S}_Y) = \mathcal{N}^{l_n}_{\mathrm{H}}(\mathcal{S}_{XY})$ that

$$\begin{aligned} &\mathbb{P}\left\{(\tilde{X}^n, \tilde{Y}^n) \in \mathcal{N}^{l_n}_{\mathrm{H}}(\mathcal{S}_{XY})\right\} \\ &\geq \mathbb{P}\left\{(\tilde{X}^n, \tilde{Y}^n) \in \mathcal{N}^{d_n}_{\mathrm{H}}(\mathcal{S}_X) \times \mathcal{N}^{d_n}_{\mathrm{H}}(\mathcal{S}_Y)\right\} \\ &\geq \mathbb{P}\left\{\tilde{X}^n \in \mathcal{N}^{d_n}_{\mathrm{H}}(\mathcal{S}_X)\right\} + \mathbb{P}\left\{\tilde{Y}^n \in \mathcal{N}^{d_n}_{\mathrm{H}}(\mathcal{S}_Y)\right\} - 1 \\ &\geq 1 - 2\nu_n \\ &= 1 - o(1), \end{aligned} \tag{69}$$

where the second inequality follows from the elementary fact that, for two events E_1 and E_2,

$$\begin{aligned} \mathbb{P}\{E_1 \cap E_2\} &= \mathbb{P}\{E_1\} + \mathbb{P}\{E_2\} - \mathbb{P}\{E_1 \cup E_2\} \\ &\geq \mathbb{P}\{E_1\} + \mathbb{P}\{E_2\} - 1. \end{aligned} \tag{70}$$

As a result, for sufficiently large n, we can obtain

$$\begin{aligned} &\mathbb{P}\left\{(\tilde{X}^n, \tilde{Y}^n) \in \mathcal{T}^n_{Q_{XY};\eta_n} \cap \mathcal{N}^{l_n}_{\mathrm{H}}(\mathcal{S}_{XY})\right\} \\ &\geq \mathbb{P}\left\{(\tilde{X}^n, \tilde{Y}^n) \in \mathcal{T}^n_{Q_{XY};\eta_n}\right\} \\ &\qquad + \mathbb{P}\left\{(\tilde{X}^n, \tilde{Y}^n) \in \mathcal{N}^{l_n}_{\mathrm{H}}(\mathcal{S}_{XY})\right\} - 1 \\ &\geq \frac{1}{2}. \end{aligned} \tag{71}$$

Therefore, with $\mathcal{Q}_n$ as defined in (60), we obtain

$$\begin{aligned} &\max_{\tilde{Q}_{XY} \in \mathcal{Q}_n} \gamma_n(\tilde{Q}_{XY}) \\ &\geq \sum_{\tilde{Q}_{XY} \in \mathcal{Q}_n} \gamma_n(\tilde{Q}_{XY}) \cdot \mathbb{P}\left\{\hat{P}_{\tilde{X}^n\tilde{Y}^n} = \tilde{Q}_{XY}\right\} \\ &= \sum_{\tilde{Q}_{XY} \in \mathcal{Q}_n} \gamma_n(\tilde{Q}_{XY}) \cdot \mathbb{P}\left\{(\tilde{X}^n, \tilde{Y}^n) \in \mathcal{T}^n_{\tilde{Q}_{XY}}\right\} \\ &= \sum_{\tilde{Q}_{XY} \in \mathcal{Q}_n} \mathbb{P}\left\{(\tilde{X}^n, \tilde{Y}^n) \in \mathcal{T}^n_{\tilde{Q}_{XY}} \cap \mathcal{N}^{l_n}_{\mathrm{H}}(\mathcal{S}_{XY})\right\} \\ &= \mathbb{P}\left\{(\tilde{X}^n, \tilde{Y}^n) \in \mathcal{T}^n_{Q_{XY};\eta_n} \cap \mathcal{N}^{l_n}_{\mathrm{H}}(\mathcal{S}_{XY})\right\} \\ &\geq \frac{1}{2}, \end{aligned} \tag{72}$$

where to obtain the second equality we have used the fact that

$$\gamma_n(\tilde{Q}_{XY}) = \mathbb{P}\left\{(\tilde{X}^n, \tilde{Y}^n) \in \mathcal{N}^{l_n}_{\mathrm{H}}(\mathcal{S}_{XY}) \middle| (\tilde{X}^n, \tilde{Y}^n) \in \mathcal{T}^n_{\tilde{Q}_{XY}}\right\},$$

and where the last equality follows from that

$$\mathcal{T}^n_{Q_{XY};\eta_n} = \bigcup_{\tilde{Q}_{XY} \in \mathcal{Q}_n} \mathcal{T}^n_{\tilde{Q}_{XY}}.$$

■

Appendix D
Proof of Fact 4

It suffices to prove the first equality, since the second equality follows immediately from the definitions of $\mathcal{E}[\cdot]$ and sufficient decoder families.

To this end, suppose $(E_0, E_1) \in \mathcal{E}(0_{M_X}, 0_{M_Y})$, then for each $\epsilon > 0$, there exists a sequence of coding scheme $\{\mathcal{C}_n\}_{n\ge1}$, such that [cf. (4)]

$$-\lim_{n\to\infty} \frac{1}{n} \log \pi_i(\mathcal{C}_n) = E_i - \epsilon, \quad i = 0, 1, \tag{73}$$

where each coding scheme $\mathcal{C}_n$ is equipped with some decoder in $\mathcal{F}_{M_X,M_Y}$.

Note that since the set $\mathcal{F}_{M_X,M_Y}$ is finite, there exists a decoder $\phi \in \mathcal{F}_{M_X,M_Y}$ and an infinite subsequence $\{m_k\}_{k\ge1}$ of positive integers, such that for each $k \ge 1$, the corresponding coding scheme $\mathcal{C}_{m_k}$ is equipped with ϕ.

Moreover, we define a new sequence of coding scheme $\mathcal{C}'_n \triangleq \mathcal{C}_{m_{\hat{k}}}$ where $\hat{k} = \hat{k}(n) \triangleq \max\{k : m_k \le n\}$. It can be verified that

$$\begin{aligned} -\lim_{n\to\infty} \frac{1}{n} \log \pi_i(\mathcal{C}'_n) &= -\lim_{k\to\infty} \frac{1}{n} \log \pi_i(\mathcal{C}_{n_k}) \\ &= E_i - \epsilon, \quad \text{for } i = 0, 1, \end{aligned} \tag{74}$$

which implies that $(E_0, E_1) \in \mathcal{E}[\phi]$.

Therefore, we obtain

$$\mathcal{E}(0_{M_X}, 0_{M_Y}) \subset \bigcup_{\phi\in\mathcal{F}_{M_X,M_Y}} \mathcal{E}[\phi] = \mathcal{E}[\mathcal{F}_{M_X,M_Y}]. \tag{75}$$

In addition, note that for each decoder $\phi \in \mathcal{F}_{M_X,M_Y}$, we have $\mathcal{E}[\phi] \subset \mathcal{E}(0_{M_X}, 0_{M_Y})$, which implies the reverse inclusion

$$\mathcal{E}[\mathcal{F}_{M_X,M_Y}] \subset \mathcal{E}(0_{M_X}, 0_{M_Y}). \tag{76}$$

From (75) and (76), we obtain $\mathcal{E}(0_{M_X}, 0_{M_Y}) = \mathcal{E}[\mathcal{F}_{M_X,M_Y}]$ as desired. ■

Appendix E
Proof of Theorem E

We first demonstrate that $(E_0, E_1) \in \mathcal{E}[\phi]$ if $(\mathcal{D}_0(E_0), \mathcal{D}_1(E_1))$ is separable under ϕ. To this end, we consider the error exponents associated with the coding schemes $\{\mathcal{C}_n\}_{n\ge1}$ with $\mathcal{C}_n \triangleq (f_n, g_n, \phi)$, where $f_n(x^n) \triangleq \theta_X(\hat{P}_{x^n}), g_n(y^n) \triangleq \theta_Y(\hat{P}_{y^n})$, and θ_X and θ_Y are the corresponding functions as defined in Definition 9 to separate $(\mathcal{D}_0(E_0), \mathcal{D}_1(E_1))$.

To begin, first note that from Fact 3, there exists some $\epsilon_n = o(1)$, such that for each $i \in \{0, 1\}$, we have

$$\begin{aligned} &\mathbb{P}\left\{X^n \in \mathcal{T}^n_{Q_X}, Y^n \in \mathcal{T}^n_{Q_Y} | \mathsf{H} = i\right\} \\ &\quad\le \exp(-n(D^*_i(Q_X, Q_Y) - \epsilon_n)). \end{aligned} \tag{77}$$

In addition, for each $i = 0, 1$ and $n \ge 1$, let us define

$$\Gamma^n_i \triangleq \{(Q_X, Q_Y) \in \hat{\mathcal{P}}^{\mathcal{X}}_n \times \hat{\mathcal{P}}^{\mathcal{Y}}_n : \phi(\theta_X(Q_X), \theta_Y(Q_Y)) \ne i\},$$

and it can be verified from Definition 9 that

$$D^*_i(Q_X, Q_Y) \ge E_i \quad \text{for all} \quad (Q_X, Q_Y) \in \Gamma^n_i. \tag{78}$$

Therefore, the type-I error π_0 and type-II error π_1 can be represented as

$$\begin{aligned} \pi_i(\mathcal{C}_n) &= \mathbb{P}\left\{\phi(\theta_X(\hat{P}_{X^n}), \theta_Y(\hat{P}_{Y^n})) \ne i \middle| \mathsf{H} = i\right\} \\ &= \sum_{(Q_X,Q_Y)\in\Gamma^n_i} \mathbb{P}\left\{X^n \in \mathcal{T}^n_{Q_X}, Y^n \in \mathcal{T}^n_{Q_Y} \middle| \mathsf{H} = i\right\} \\ &\le \sum_{(Q_X,Q_Y)\in\Gamma^n_i} \exp(-n \cdot (D^*_i(Q_X, Q_Y) - \epsilon_n)) &(79)\\ &\le \sum_{(Q_X,Q_Y)\in\Gamma^n_i} \exp(-n(E_i - \epsilon_n)) &(80)\\ &\le |\Gamma^n_i| \cdot \exp(-n(E_i - \epsilon_n)) &(81)\\ &\le (n + 1)^{|\mathcal{X}|+|\mathcal{Y}|} \exp(-n(E_i - \epsilon_n)). &(82)\\ &\le \exp(-n(E_i - \epsilon'_n)), &(83) \end{aligned}$$

where (79) follows from (77), (80) follows from (78), (83) follows from (29), and where ϵ'_n is as defined in (30).

Note that since $\epsilon'_n = o(1)$, we obtain $(E_0, E_1) \in \mathcal{E}[\phi]$.

In addition, we illustrate that for each $(E_0, E_1) \in \mathcal{E}[\phi]$, $(\mathcal{D}_0(E_0), \mathcal{D}_1(E_1))$ is separable under ϕ. To this end, first note that from Theorem 1, it suffices to consider coding schemes $\tilde{\mathcal{C}}_n = (\tilde{f}_n, \tilde{g}_n, \phi)$ with type-based encoders $\tilde{f}_n : x^n \mapsto \hat{\theta}^{(n)}_X(\hat{P}_{x^n})$ and $\tilde{g}_n : y^n \mapsto \hat{\theta}^{(n)}_Y(\hat{P}_{y^n})$, where $\hat{\theta}^{(n)}_X : \hat{\mathcal{P}}^{\mathcal{X}}_n \to \lceil M_X \rfloor$ and $\hat{\theta}^{(n)}_Y : \hat{\mathcal{P}}^{\mathcal{Y}}_n \to \lceil M_Y \rfloor$ are the encoders for marginal types.

Then, it can be verified that for n sufficiently large, the $\hat{\theta}^{(n)}_X$ and $\hat{\theta}^{(n)}_Y$ satisfy that, for both $i = 0, 1$, and each $(Q_X, Q_Y) \in \mathcal{D}_i(E_i) \cap (\hat{\mathcal{P}}^{\mathcal{X}}_n \times \hat{\mathcal{P}}^{\mathcal{Y}}_n)$,

$$\phi(\hat{\theta}^{(n)}_X(Q_X), \hat{\theta}^{(n)}_Y(Q_Y)) = i. \tag{84}$$

By symmetry, it suffices to establish (84) for the case $i = 0$, which can be shown by contraction. Indeed, suppose that there exists some $(Q_X, Q_Y) \in \mathcal{D}_0(E_0) \cap (\hat{\mathcal{P}}^{\mathcal{X}}_n \times \hat{\mathcal{P}}^{\mathcal{Y}}_n)$ such that $\phi(\hat{\theta}^{(n)}_X(Q_X), \hat{\theta}^{(n)}_Y(Q_Y)) = 1$, then from Fact 3, there exists some $\nu_n = o(1)$, such that the type-I error $\pi_0(\tilde{\mathcal{C}}_n)$ satisfies

$$\begin{aligned} \pi_0(\tilde{\mathcal{C}}_n) &\ge \mathbb{P}\left\{X^n \in \mathcal{T}^n_{Q_X}, Y^n \in \mathcal{T}^n_{Q_Y} | \mathsf{H} = 0\right\} \\ &\ge \exp(-n(D^*_0(Q_X, Q_Y) + \nu_n)) \end{aligned}$$

Therefore, the type-I error exponent is at most $D^*_0(Q_X, Q_Y)$, which is strictly less than E_0, since $(Q_X, Q_Y) \in \mathcal{D}(E_0)$. This contradicts the assumption $(E_0, E_1) \in \mathcal{E}[\phi]$.

Furthermore, let us define functions $\tilde{\theta}^{(n)}_X : \mathcal{P}^{\mathcal{X}} \to \lceil M_X \rfloor$ and $\tilde{\theta}^{(n)}_Y : \hat{\mathcal{P}}^{\mathcal{Y}}_n \to \lceil M_Y \rfloor$ such that

$$\tilde{\theta}^{(n)}_X(Q_X) \triangleq \hat{\theta}^{(n)}_X(\hat{Q}^{(n)}_X) \quad \text{and} \quad \tilde{\theta}^{(n)}_Y(Q_Y) \triangleq \hat{\theta}^{(n)}_X(\hat{Q}^{(n)}_Y)$$

for all $Q_X \in \mathcal{P}^{\mathcal{X}}$ and $Q_Y \in \mathcal{P}^{\mathcal{Y}}$, where

$$\begin{aligned} \hat{Q}^{(n)}_X &\triangleq \mathop{\arg\min}_{Q'_X\in\hat{\mathcal{P}}^{\mathcal{X}}_n} d_{\max}(Q'_X, Q_X), \\ \hat{Q}^{(n)}_Y &\triangleq \mathop{\arg\min}_{Q'_Y\in\hat{\mathcal{P}}^{\mathcal{Y}}_n} d_{\max}(Q'_Y, Q_Y). \end{aligned} \tag{85}$$

Note that for each $(Q_X, Q_Y) \in \mathcal{D}_0(E_0)$, we have

$$D^*_0(Q_X, Q_Y) < E_0.$$

Note that from (85), we have

$$d_{\max}(\hat{Q}_X^{(n)}, Q_X) \le \frac{1}{n} \quad \text{and} \quad d_{\max}(\hat{Q}_Y^{(n)}, Q_Y) \le \frac{1}{n},$$

and it follows from the uniform continuity of D_0^* that,

$$D_0^*(\hat{Q}_X^{(n)}, \hat{Q}_Y^{(n)}) < E_0$$

for n sufficiently large.

This implies that $(\hat{Q}_X^{(n)}, \hat{Q}_X^{(n)}) \in \mathcal{D}_0(E_0) \cap (\hat{\mathcal{P}}_n^{\mathcal{X}} \times \hat{\mathcal{P}}_n^{\mathcal{Y}})$. Hence, from (84) we obtain

$$\phi(\tilde{\theta}_X^{(n)}(Q_X), \tilde{\theta}_Y^{(n)}(Q_Y)) = 0. \tag{86}$$

Similarly, we have

$$\phi(\tilde{\theta}_X^{(n)}(Q_X), \tilde{\theta}_Y^{(n)}(Q_Y)) = 1 \tag{87}$$

for each $(Q_X, Q_Y) \in \mathcal{D}_1(E_1)$. From (86) and (87), $\mathcal{D}_0(E_0)$ and $\mathcal{D}_1(E_1)$ is separable under ϕ, which completes the proof. ∎

APPENDIX F
PROOF OF PROPOSITION 1

It suffices to consider the first statement for $i = 0$, and other cases can be similarly established. To this end, let $\mathbf{A} \leftrightarrow \phi$ and $\phi' \triangleq \omega_X^{(0)}(\phi) \leftrightarrow \mathbf{A}_X^{(0)}$ denote the associated decision matrix of ϕ and the reduced decoder, as defined in Definition 6, respectively. We also define

$$\mathcal{A}_0' \triangleq \mathcal{A}_0 \overset{X}{\triangleright} \mathcal{A}_1, \quad \text{and} \quad \mathcal{A}_1' \triangleq \mathcal{A}_1 \overset{X}{\triangleright} \mathcal{A}_1 = \mathcal{A}_1. \tag{88}$$

Without loss of generality, suppose the 0-dominated columns of $\mathbf{A}$ are its last d columns, i.e., we have

$$\phi(m_X, m_Y) = 0, \tag{89}$$

for each $m_X = M_X - d, \dots, M_X - 1$ and $m_Y \in \lceil M_Y \rfloor$.

Moreover, it can be verified that ϕ' is the restriction of ϕ to $\lceil M_X - d \rfloor \times \lceil M_Y \rfloor$, and we have

$$\phi'(m_X, m_Y) = \phi(m_X, m_Y) \tag{90}$$

for each $(m_X, m_Y) \in \lceil M_X - d \rfloor \times \lceil M_Y \rfloor$.

To prove the "only if" part of the claim, suppose $(\mathcal{A}_0, \mathcal{A}_1)$ is separable under ϕ. Then, from Definition 9, there exist mappings $\theta_X \colon \mathcal{P}^{\mathcal{X}} \to \lceil M_X \rfloor$ and $\theta_Y \colon \mathcal{P}^{\mathcal{Y}} \to \lceil M_Y \rfloor$, such that for both $i \in \{0, 1\}$, we have

$$\phi(\theta_X(Q_X), \theta_Y(Q_Y)) = i, \quad \text{for all } (Q_X, Q_Y) \in \mathcal{A}_i. \tag{91}$$

For each $Q_X \in \Pi_X(\mathcal{A}_1)$, it can be verified that $\theta_X(Q_X) \in \lceil M_X - d \rfloor$. Otherwise, there exists $Q_Y' \in \mathcal{P}^{\mathcal{Y}}$ with $(Q_X, Q_Y') \in \mathcal{A}_1$, and it follows from (89) that $\phi(\theta_X(Q_X), \theta_Y(Q_Y')) = 0$, which contradicts the claim (91).

Then, we define $\theta' \colon \mathcal{P}^{\mathcal{X}} \to \lceil M_X - d \rfloor$ such that

$$\theta'(Q_X) = \begin{cases} \theta'(Q_X) & \text{if } Q_X \in \Pi_X(\mathcal{A}_1), \\ 0 & \text{otherwise,} \end{cases} \tag{92}$$

and it follows from (90) that, for each $Q_X \in \Pi_X(\mathcal{A}_1)$ and $Q_Y \in \mathcal{P}^{\mathcal{Y}}$, we have

$$\phi(\theta_X(Q_X), \theta_Y(Q_Y)) \equiv \phi'(\theta_X'(Q_X), \theta_Y(Q_Y)).$$

Moreover, from (91) we have, for both $i \in \{0, 1\}$,

$$\phi'(\theta_X'(Q_X), \theta_Y(Q_Y)) = i, \quad \text{for all } (Q_X, Q_Y) \in \mathcal{A}_i', \tag{93}$$

which implies that $(\mathcal{A}_0', \mathcal{A}_1')$ is separable under ϕ'.

For the "if" part of the claim, suppose $(\mathcal{A}_0', \mathcal{A}_1')$ is separable under ϕ', then there exist functions $\hat{\theta}_X \colon \mathcal{P}^{\mathcal{X}} \to \lceil M_X - d \rfloor$ and $\hat{\theta}_Y \colon \mathcal{P}^{\mathcal{Y}} \to \lceil M_X - d \rfloor$, such that for both $i \in \{0, 1\}$, we have

$$\phi'(\hat{\theta}_X(Q_X), \hat{\theta}_Y(Q_Y)) = i, \quad \text{for all } (Q_X, Q_Y) \in \mathcal{A}_i'. \tag{94}$$

Then, let us define $\hat{\theta}' \colon \mathcal{P}^{\mathcal{X}} \to \lceil M_X \rfloor$ such that

$$\hat{\theta}'(Q_X) = \begin{cases} \hat{\theta}(Q_X) & \text{if } Q_X \in \Pi_X(\mathcal{A}_1), \\ M_X - d & \text{otherwise.} \end{cases} \tag{95}$$

From (89), for both $i \in \{0, 1\}$, we have

$$\phi(\hat{\theta}_X'(Q_X), \hat{\theta}_Y(Q_Y)) = i, \quad \text{for all } (Q_X, Q_Y) \in \mathcal{A}_i, \tag{96}$$

which implies that $(\mathcal{A}_0, \mathcal{A}_1)$ is separable under ϕ. ∎

APPENDIX G
PROOF OF THEOREM 3

We first introduce several useful facts on the separability, which can be readily verified from Definition 9.

Fact 6: Given $\mathcal{A}, \mathcal{A}' \subset \mathcal{P}_\star$ and decoders $\phi \simeq \phi'$, $(\mathcal{A}, \mathcal{A}')$ is separable under ϕ if and only if it is separable under ϕ'.

Fact 7: Given $\mathcal{A}, \mathcal{A}' \subset \mathcal{P}_\star$, $(\mathcal{A}, \mathcal{A}')$ is separable under $\varphi_{1,1}$ if and only if $\mathcal{A}' = \varnothing$.

Fact 8: For any given $\mathcal{A}, \mathcal{A}' \subset \mathcal{P}_\star$ and $\phi \in \mathcal{F}$, $(\mathcal{A}, \mathcal{A}')$ is separable under ϕ if and only if $(\mathcal{A}', \mathcal{A})$ is separable under its complement $\bar{\phi}$.

The following corollary of Proposition 1 would also be useful.

Corollary 4: Suppose $\mathcal{A}$ and $\mathcal{A}'$ are two disjoint subsets of $\mathcal{P}_\star$. Given $M \ge 2$, the following statements are equivalent:

S1) $(\mathcal{A}, \mathcal{A}')$ is separable under $\varphi_{M,M}$;
S2) $(\mathcal{A}', \mathcal{A} \triangleright \mathcal{A}')$ is separable under $\varphi_{M-1,M-1}$;
S3) $\mathcal{A} \triangleright_M \mathcal{A}' = \varnothing$.

In addition, for given $M_X > M_Y \ge 1$, $(\mathcal{A}, \mathcal{A}')$ is separable under φ_{M_X,M_Y} if and only if $(\mathcal{A}, \mathcal{A}' \overset{X}{\triangleright} \mathcal{A})$ is separable under φ_{M_Y,M_Y}.

Proof of Corollary 4: First, note that

$$\bar{\varphi}_{M-1,M-1} = \omega_Y^{(0)} \left(\omega_X^{(0)}(\varphi_{M,M}) \right).$$

Therefore, from Proposition 1, we have

$$\begin{aligned} &\text{S1 } (\mathcal{A}, \mathcal{A}') \text{ is separable under } \varphi_{M,M} \\ &\iff (\mathcal{A} \overset{X}{\triangleright} \mathcal{A}', \mathcal{A}') \text{ is separable under } \omega_X^{(0)}(\varphi_{M,M}) \\ &\iff ((\mathcal{A} \overset{X}{\triangleright} \mathcal{A}') \overset{Y}{\triangleright} \mathcal{A}', \mathcal{A}') \text{ is separable under } \bar{\varphi}_{M-1,M-1} \\ &\iff (\mathcal{A} \triangleright \mathcal{A}', \mathcal{A}') \text{ is separable under } \bar{\varphi}_{M-1,M-1} \\ &\iff \text{S2 } (\mathcal{A}', \mathcal{A} \triangleright \mathcal{A}') \text{ is separable under } \varphi_{M-1,M-1}, \end{aligned}$$

where the third "$\iff$" follows from $(\mathcal{A} \overset{X}{\triangleright} \mathcal{A}') \overset{Y}{\triangleright} \mathcal{A}' = \mathcal{A} \triangleright \mathcal{A}'$. To obtain the last "$\iff$", we have used Fact 8.

Then, by repeatedly applying the equivalence "S1 $\iff$ S2" $(M-1)$ times, we know that statements S1 and S2 are further equivalent to

$$\begin{aligned}&(\mathcal{A} \triangleright_{M-1} \mathcal{A}', \mathcal{A} \triangleright_M \mathcal{A}') \text{ is separable under } \varphi_{1,1}\\ &\quad\iff \text{ S3 } \mathcal{A} \triangleright_M \mathcal{A}' = \varnothing,\end{aligned}$$

where we have used Fact 7.

Similarly, we can establish the second statement of the claim, by noting that

$$\varphi_{M_Y,M_Y} = \omega_X^{(1)}(\varphi_{M_X,M_Y}), \quad \text{for all } M_X > M_Y \geq 1.$$

■

Proceeding to the proof of Theorem 3, we first consider the case $M_X = M_Y$. From Theorem 2 we have

$$\begin{aligned}&(E_0, E_1) \in \mathcal{E}[\varphi_{M_Y,M_Y}]\\ &\iff (\mathcal{D}_0(E_0), \mathcal{D}_1(E_1)) \text{ is separable under } \varphi_{M_Y,M_Y}\\ &\iff \mathcal{D}_0(E_0) \triangleright_M \mathcal{D}_1(E_1) = \varnothing,\end{aligned}$$

where the last "$\iff$" follows from Corollary 4. Then, it follows from Fact 8 that

$$(E_0, E_1) \in \mathcal{E}[\varphi_{M_Y,M_Y}] \iff \mathcal{D}_1(E_1) \triangleright_M \mathcal{D}_0(E_0) = \varnothing.$$

For the case $M_X > M_Y$, it can be verified that

$$\begin{aligned}\mathcal{E}[\varphi_{M_X,M_Y}] &= \mathcal{E}[\varphi_{M_Y+1,M_Y}]\\ &= \mathcal{E}[\bar{\varphi}_{M_Y+1,M_Y}] = \mathcal{E}[\bar{\varphi}_{M_X,M_Y}],\end{aligned}$$

where the second equality follows from Fact 6 and that $\varphi_{M_Y+1,M_Y} \simeq \bar{\varphi}_{M_Y+1,M_Y}$. To obtain the first equality, note that the decision matrix associated with φ_{M_X,M_Y} and that associated with φ_{M_Y,M_Y} differ only in duplicated columns. The last equality follows from symmetry considerations.

Then, from Theorem 2 and Corollary 4 we can obtain

$$\begin{aligned}&(E_0, E_1) \in \mathcal{E}[\varphi_{M_X,M_Y}]\\ &\iff (\mathcal{D}_0(E_0), \mathcal{D}_1(E_1)) \text{ is separable under } \varphi_{M_X,M_Y}\\ &\iff (\mathcal{D}_0(E_0), \mathcal{D}_1(E_1) \overset{X}{\triangleright} \mathcal{D}_0(E_0))\\ &\qquad\quad \text{is separable under } \varphi_{M_Y,M_Y}\\ &\iff \mathcal{D}_0(E_0) \triangleright_{M_Y} (\mathcal{D}_1(E_1) \overset{X}{\triangleright} \mathcal{D}_0(E_0)) = \varnothing,\end{aligned}$$

which completes the proof. ■

Appendix H
Proof of Proposition 2

First, we define $R_{\max} \triangleq \max\{H(P_X^{(0)}), H(P_X^{(1)})\}$ with $H(\cdot)$ representing the entropy. Then, due to the inclusion chain

$$\begin{aligned}\mathcal{E}(0_{2^{M_Y}}, 0_{M_Y}) &\subset \mathcal{E}(0_{M_X}, 0_{M_Y})\\ &\subset \mathcal{E}(R_X, 0_{M_Y}) \subset \mathcal{E}(R_{\max}, 0_{M_Y}), \qquad (97)\end{aligned}$$

it suffices to demonstrate $\mathcal{E}(0_{2^{M_Y}}, 0_{M_Y}) = \mathcal{E}(R_{\max}, 0_{M_Y})$.

Specifically, note that under the constraints $(R_{\max}, 0_{M_Y})$, the decoder can obtain the full side information of the X sequence. Then, for each $n \geq 1$, the corresponding coding scheme can be characterized as a encoder g_n that encodes Y^n, and a central decoder $\phi_n \colon \mathcal{X}^n \times \lceil M_Y \rfloor \to \{0, 1\}$. When nodes N_X and N_Y observe sequences $X^n = x^n$ and $Y^n = y^n$, respectively, the decision at the center can be represented as $\hat{\mathsf{H}} = \phi_n(x^n, g_n(y^n))$.

Then, we introduce a new encoder $f_n \colon \mathcal{X}^n \to \lceil 2^{M_X} \rfloor$ for encoding X^n, such that

$$f_n(x^n) \triangleq \sum_{j \in \lceil M_Y \rfloor} \phi_n(x^n, j) \cdot 2^j, \quad \text{for all } x^n \in \mathcal{X}^n.$$

We also define decoder $\phi' \colon \lceil 2^{M_Y} \rfloor \times \lceil M_Y \rfloor \to \{0, 1\}$ as

$$\phi'(m_X, m_Y) \triangleq b_{m_Y}, \quad (m_X, m_Y) \in \lceil 2^{M_Y} \rfloor \times \lceil M_Y \rfloor,$$

where for each $j \in \lceil M_Y \rfloor$, $b_j \in \{0, 1\}$ denotes the $(j+1)$-th digit of the binary representation of m_X, such that

$$m_X = (b_{M_Y-1} \cdots b_1 b_0)_2 \triangleq \sum_{j \in \lceil M_Y \rfloor} b_j \cdot 2^j.$$

It can be verified that for each $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$, the decision $\hat{\mathsf{H}}'$ associated with the coding scheme (f_n, g_n, ϕ') is

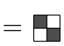
$$\hat{\mathsf{H}}' = \phi'(f_n(x^n), g_n(y^n)) \equiv \phi_n(x^n, g_n(y^n)) = \hat{\mathsf{H}}.$$

Therefore, for each coding scheme under the rate constraints $(R_{\max}, 0_{M_Y})$, there exists a coding scheme satisfying constraints $(0_{2^{M_Y}}, 0_{M_Y})$ which obtains the same decision result. Hence, we have $\mathcal{E}(R_{\max}, 0_{M_Y}) \subset \mathcal{E}(0_{2^{M_Y}}, 0_{M_Y})$, and it follows from (97) that $\mathcal{E}(0_{2^{M_Y}}, 0_{M_Y}) = \mathcal{E}(R_{\max}, 0_{M_Y})$. ■

Appendix I
Proof of Proposition 3

We would show the equivalences by separately establishing "S1 $\implies$ S2", "S2 $\implies$ S3", and "S3 $\implies$ S1".

First, for the claim "S1 $\implies$ S2", note that there are two irreducible 2×2 decoders, which we can denote by

$\phi_0 \leftrightarrow \mathbf{A}_0 =$  and $\phi_1 \leftrightarrow \mathbf{A}_1 =$ . (98)

We then prove the claim by contradiction. Specifically, we assume that $\phi \leftrightarrow \mathbf{A}$ has a irreducible subdecoder ϕ_0. Without loss of generality, suppose $\mathbf{A}_0$ is the submatrix of $\mathbf{A}$ composed of first two rows and first two columns of $\mathbf{A}$. Then, it suffices to show that ϕ is not completely reducible, which is trivially true if ϕ is irreducible.

We now consider the case where ϕ is reducible. Then, there exists an elementary reduction operator ω, such that $\omega(\phi)$ exists. Since the first two rows and first two columns of $\mathbf{A}$ cannot be dominated, $\mathbf{A}_0$ is also a submatrix of $\mathbf{A}' \leftrightarrow \omega(\phi)$, and thus ϕ_0 is also a subdecoder of $\omega(\phi)$. As a consequence, for all ϕ' that can be reduced from ϕ, ϕ_0 is a subdecoder of ϕ', which implies that ϕ is not completely reducible. Similarly, ϕ is not completely reducible if ϕ_1 is a subdecoder of ϕ.

Then, to prove "S2 $\implies$ S3", note that for each decoder ϕ, we can construct its equivalent decoder $\phi' \simeq \phi$ such that the functions $\sigma_X^{(\phi)}(\cdot)$ and $\sigma_Y^{(\phi)}(\cdot)$ are both non-decreasing, where for each $\phi \in \mathcal{F}_{M_X,M_Y}$, we have defined

$$\sigma_X^{(\phi)}(m_X) \triangleq \sum_{m_Y \in \lceil M_Y \rfloor} \phi(m_X, m_Y), \ \forall\, m_X \in \lceil M_X \rfloor, \quad (99a)$$

$$\sigma_Y^{(\phi)}(m_Y) \triangleq \sum_{m_X \in \lceil M_X \rfloor} \phi(m_X, m_Y), \ \forall\, m_Y \in \lceil M_Y \rfloor. \quad (99b)$$

We then establish that ϕ' is monotonic if ϕ satisfies the statement S2. To see this, first note that for all $0 \le m_X < m'_X < M_X$, we have $\sigma_X(m_X) \le \sigma_X(m'_X)$, which implies

$$\sum_{m_Y \in \lceil M_Y \rfloor} [\phi'(m_X, m_Y) - \phi'(m'_X, m_Y)] \le 0. \tag{100}$$

Now, suppose $\phi'(m_X, m_Y) - \phi'(m'_X, m_Y) > 0$ for some $m_Y \in \lceil M_Y \rfloor$. Since the summation (100) is non-negative, there exists $m'_Y \in \lceil M_Y \rfloor$ with $\phi'(m_X, m'_Y) - \phi'(m'_X, m'_Y) < 0$. Therefore,

$$\begin{aligned} \phi'(m_X, m_Y) &= 1, \quad \phi'(m'_X, m_Y) = 0, \\ \phi'(m_X, m'_Y) &= 0, \quad \phi'(m'_X, m'_Y) = 1, \end{aligned}$$

which implies that ϕ' has an irreducible 2×2 subdecoder. Thus, ϕ also has an irreducible 2×2 subdecoder, which contradicts the statement S2.

As a consequence, we obtain

$$\phi'(m_X, m_Y) - \phi'(m'_X, m_Y) \le 0$$

for all $m_Y \in \lceil M_Y \rfloor$ and $0 \le m_X < m'_X < M_X$, and, similarly,

$$\phi'(m_X, m_Y) - \phi'(m_X, m'_Y) \le 0$$

for all $m_X \in \lceil M_X \rfloor$ and $0 \le m_Y < m'_Y < M_Y$. This demonstrates the statement S3.

Finally, to establish "S3 $\implies$ S1", note that for equivalent decoders $\phi \simeq \phi'$, ϕ is completely reducible if and only if ϕ' is completely reducible. Therefore, it suffices to show that monotonic decoders are completely reducible. To this end, we first show that the monotonic decoders are reducible. Indeed, given a monotonic decoder $\phi \in \mathcal{F}_{M_X,M_Y}$, it can be verified from the definition that

- if $\phi(M_X - 1, 0) = 0$, then $\phi(m_X, 0) = 0$ for all $m_X \in \lceil M_X \rfloor$;
- if $\phi(M_X - 1, 0) = 1$, then $\phi(M_X - 1, m_Y) \equiv 1$ for all $m_Y \in \lceil M_Y \rfloor$.

Therefore, ϕ is reducible.

Moreover, if ϕ is non-trivial, then there exists an elementary reduction operator ω, such that $\omega(\phi)$ exists. Then, it can be verified that $\omega(\phi)$ is also monotonic, and we can similarly apply reduction operations on $\omega(\phi)$ until obtaining trivial decoders. This establishes the statement S1. ∎

Appendix J
Proof of Lemma 5

First, for all given M_X and M_Y, we define

$$\mathcal{F}^{\mathrm{m}}_{M_X,M_Y} \triangleq \{\phi \in \mathcal{F}_{M_X,M_Y} : \phi \text{ is monotonic}\}. \tag{101}$$

Then, from Fact 5 and the equivalence of statements S1 and S3 in Proposition 3, we obtain

$$\mathcal{E}[\Omega_{M_X,M_Y}] = \mathcal{E}[\mathcal{F}^{\mathrm{m}}_{M_X,M_Y}]. \tag{102}$$

Therefore, it suffices to establish $\{\phi\} \preceq \Phi_{M_X,M_Y}$ for each $\phi \in \mathcal{F}^{\mathrm{m}}_{M_X,M_Y}$.

To this end, we first establish a useful expression of monotonic decoders via using the functions $\sigma_X^{(\phi)}(\cdot)$ and $\sigma_Y^{(\phi)}(\cdot)$ as defined in (99). In particular, for each $\phi \in \mathcal{F}^{\mathrm{m}}_{M_X,M_Y}$, from the definition of monotonicity we have, for all $(m_X, m_Y) \in \lceil M_X \rfloor \times \lceil M_Y \rfloor$,

$$\begin{aligned} \phi(m_X, m_Y) &= \mathbb{1}_{\{m_X + \sigma_Y^{(\phi)}(m_Y) \ge M_X\}} && (103) \\ &= \mathbb{1}_{\{\sigma_X^{(\phi)}(m_X) + m_Y \ge M_Y\}}. && (104) \end{aligned}$$

If $M_X > M_Y$, for each $m_X \in \lceil M_X \rfloor$, we have $\sigma_X^{(\phi)}(m_X) \in \lceil M_X \rfloor$. Then, it follows from (104) that, for all $m_Y \in \lceil M_Y \rfloor$,

$$\begin{aligned} \phi(m_X, m_Y) &= \mathbb{1}_{\{\sigma_X^{(\phi)}(m_X) + m_Y \ge M_Y\}} \\ &= \varphi_{M_X,M_Y}(\sigma_X^{(\phi)}(m_X), m_Y), && (105) \end{aligned}$$

which implies that ϕ is a subdecoder of φ_{M_X,M_Y}. Therefore, from Fact 5 we obtain

$$\{\phi\} \preceq \{\varphi_{M_X,M_Y}\} \preceq \Phi_{M_X,M_Y}. \tag{106}$$

For the case $M_X = M_Y$, let $M \triangleq M_X$, then $\sigma_Y^{(\phi)}(\cdot)$ is a non-decreasing function on $\lceil M \rfloor$. If $\sigma_Y^{(\phi)}(\cdot)$ is not strictly increasing, then there exists $m'_Y \in \lceil M - 1 \rfloor$, such that $\sigma_Y^{(\phi)}(m'_Y) = \sigma_Y^{(\phi)}(m'_Y + 1)$, and from (103) we obtain

$$\phi(m_X, m'_Y) = \phi(m_X, m'_Y + 1), \quad \text{for all } m_X \in \lceil M \rfloor.$$

This implies that the m'_Y-th and $(m'_Y + 1)$-th rows of the associated decision matrix $\mathbf{A} \leftrightarrow \phi$ are the same. Let $\mathbf{A}'$ denote the submatrix of $\mathbf{A}$ obtained by deleting its $(m'_Y + 1)$-th row. Then, it can be verified that, the decoder $\phi' \leftrightarrow \mathbf{A}'$ is an $M \times (M - 1)$ monotonic decoder with $\mathcal{E}[\phi] = \mathcal{E}[\phi']$.

Therefore, we obtain

$$\{\phi\} \preceq \mathcal{F}^{\mathrm{m}}_{M,M-1} \preceq \{\varphi_{M,M-1}\} \preceq \{\varphi_{M,M}\} \preceq \Phi_{M,M},$$

where the second "$\preceq$" follows from (106), and where the third "$\preceq$" follows from Fact 5 and that $\varphi_{M,M-1}$ is a subdecoder of $\varphi_{M,M}$.

It remains to establish the claim for the case where $M_X = M_Y = M$ and $\sigma_Y^{(\phi)}(\cdot)$ is strictly increasing on $\lceil M \rfloor$. To this end, first note that if $\sigma_Y^{(\phi)}(0) = 0$, for each $m_X \in \lceil M \rfloor$ we have $\phi(m_X, 0) = 0$. Therefore, we have $\sigma_X^{(\phi)}(m_X) \in \lceil M \rfloor$, and it follows from (105)–(106) that $\{\phi\} \preceq \mathcal{F}^{\mathrm{m}}_{M,M}$. Moreover, if $\sigma_Y^{(\phi)}(\cdot)$ is strictly increasing and $\sigma_Y^{(\phi)}(0) \ne 0$, we have

$$\sigma_Y^{(\phi)}(m_Y) = m_Y + 1, \quad \text{for all } m_Y \in \lceil M \rfloor.$$

Hence, from (103) we have, for all $(m_X, m_Y) \in \lceil M_X \rfloor \times \lceil M_Y \rfloor$,

$$\begin{aligned} \phi(m_X, m_Y) &= \mathbb{1}_{\{m_X + m_Y \ge M - 1\}} \\ &= \mathbb{1}_{\{(M-1-m_X) + (M-1-m_Y) \le M-1\}} \\ &= \mathbb{1}_{\{(M-1-m_X) + (M-1-m_Y) < M\}} \\ &= \bar{\varphi}_{M,M}(M - 1 - m_X, M - 1 - m_Y), \end{aligned}$$

which implies that $\phi \simeq \bar{\varphi}_{M,M}$. As a result, we obtain

$$\{\phi\} \preceq \{\bar{\varphi}_{M,M}\} \preceq \Phi_{M,M},$$

which completes the proof. ∎

APPENDIX K
PROOF OF PROPOSITION 4

To begin, we consider a decoder ϕ that is not completely reducible. If ϕ is irreducible, it suffices to let $\phi' = \phi$. Otherwise, since ϕ cannot be reduced to trivial decoders, each decoder reduced from ϕ is either an irreducible decoder, or a non-trivial reducible decoder. Therefore, we can apply a series of elementary reduction operators on ϕ, until obtaining some irreducible decoder.

It remains only to demonstrate the uniqueness of obtained irreducible decoders. To see this, suppose both ϕ'' and $\tilde{\phi}''$ are the irreducible decoders obtained from the above procedures.

Note that since ϕ'' is an irreducible subdecoder of ϕ, its associated rows and columns in the decision matrix $\mathbf{A} \leftrightarrow \phi$ cannot be dominated during the above reduction procedures. Therefore, it is also a subdecoder of all decoders reduced from ϕ.

As a result, ϕ'' is a subdecoder of $\tilde{\phi}''$, and, similarly, $\tilde{\phi}''$ is a subdecoder of ϕ''. Hence, we have $\phi'' = \tilde{\phi}''$, corresponding to the unique decoder ϕ' reduced from ϕ. ∎

APPENDIX L
PROOF OF LEMMA 6

Our proof makes use of the notion of open sets in $\mathcal{P}_\star$, together with discussions on the separability (cf. Definition 9) under reducible and decomposable decoders.

As a first step, we define the open sets in $\mathcal{P}_\star$ as follows. With slight abuse of notation, we use $Q_X Q_Y$ to represent $(Q_X, Q_Y) \in \mathcal{P}_\star$. Then, we introduce the metric $d_\star$ on $\mathcal{P}_\star$, such that for all given $Q_X Q_Y, Q'_X Q'_Y \in \mathcal{P}_\star$,

$$\begin{aligned} &d_\star(Q_X Q_Y, Q'_X Q'_Y) \\ &\triangleq \max\{d_{\max}(Q_X, Q'_X), d_{\max}(Q_Y, Q'_Y)\}. \end{aligned}$$

Moreover, $\mathcal{A} \subset \mathcal{P}_\star$ is open, if for each $Q_X Q_Y \in \mathcal{A}$, there exists $\eta > 0$, such that for all $Q'_X Q'_Y \in \mathcal{P}_\star$ satisfying $d_\star(Q_X Q_Y, Q'_X Q'_Y) < \eta$, we have $Q'_X Q'_Y \in \mathcal{A}$.

Specifically, with assumption (21), the functions $D_0^*(\cdot)$ and $D_1^*(\cdot)$ as defined in (9) are uniformly continuous, from which we can obtain the following useful fact.

Fact 9: Suppose the assumption (21) holds. Then, for all $t \geq 0$ and $i \in \{0, 1\}$, $\mathcal{D}_i(t)$ is open.

To better illustrate the separability under reducible decoders, we introduce notations as follows.

For given $\mathcal{A}_0, \mathcal{A}_1 \subset \mathcal{P}_\star$ and a reduction operator ω, we define the sets $\tau_i(\mathcal{A}_0, \mathcal{A}_1; \omega)$ for $i = 0, 1$, such that for $j \in \{0, 1\}$ and $\bar{j} \triangleq 1 - j$,

$$\tau_i(\mathcal{A}_0, \mathcal{A}_1; \omega_X^{(j)}) \triangleq \mathcal{A}_i \overset{X}{\triangleright} \mathcal{A}_{\bar{j}}, \tag{107a}$$

$$\tau_i(\mathcal{A}_0, \mathcal{A}_1; \omega_Y^{(j)}) \triangleq \mathcal{A}_i \overset{Y}{\triangleright} \mathcal{A}_{\bar{j}}, \tag{107b}$$

and, for each composite reduction operator $\omega \circ \omega'$,

$$\tau_i(\mathcal{A}_0, \mathcal{A}_1; \omega \circ \omega') \triangleq \tau_i(\mathcal{A}'_0, \mathcal{A}'_1; \omega), \tag{108}$$

where $\mathcal{A}'_j \triangleq \tau_j(\mathcal{A}_0, \mathcal{A}_1; \omega')$ for $j \in \{0, 1\}$.

Then, we have the following useful fact, which can be verified by definition.

Fact 10: If $\mathcal{A}_0, \mathcal{A}_1 \subset \mathcal{P}_\star$ are open and convex, then for all reduction operator ω and $i \in \{0, 1\}$, $\tau_i(\mathcal{A}_0, \mathcal{A}_1; \omega)$ is open and convex.

The following fact, as an immediate consequences of Proposition 1, is also useful.

Fact 11: Suppose ϕ is a reducible decoder and can be reduced to $\psi \triangleq \omega(\phi)$ by some reduction operator ω. Then, for all $\mathcal{A}_0, \mathcal{A}_1 \subset \mathcal{P}_\star$, $(\mathcal{A}_0, \mathcal{A}_1)$ is separable under ϕ if and only if $(\tau_0(\mathcal{A}_0, \mathcal{A}_1; \omega), \tau_1(\mathcal{A}_0, \mathcal{A}_1; \omega))$ is separable under ψ.

In addition, our proof will make use of the following result.

Lemma 8: Suppose ϕ is a decomposable decoder with the decomposition [cf. (14)]

$$\phi = \phi_0 \oplus \phi_1 \oplus \bar{\imath} \tag{109}$$

for some $i \in \{0, 1\}$, where ϕ_0 and ϕ_1 satisfy (15). Suppose $\mathcal{A}_0$ and $\mathcal{A}_1$ are open convex subsets of $\mathcal{P}_\star$. If $(\mathcal{A}_0, \mathcal{A}_1)$ is separable under ϕ, then $(\mathcal{A}_0, \mathcal{A}_1)$ is separable under ϕ_j for some $j \in \{0, 1\}$.

Proof of Lemma 8: By symmetry, it suffices to consider the case where (109) holds for $i = 1$, i.e.,

$$\phi = \phi_0 \oplus \phi_1. \tag{110}$$

Since $(\mathcal{A}_0, \mathcal{A}_1)$ is separable under ϕ, from Definition 9, there exists $\theta_X : \mathcal{P}^{\mathcal{X}} \to \lceil M_X \rfloor$ and $\theta_Y : \mathcal{P}^{\mathcal{Y}} \to \lceil M_Y \rfloor$, such that, we have

$$\phi(\theta_X(Q_X), \theta_Y(Q_Y)) = 0, \text{ for all } (Q_X, Q_Y) \in \mathcal{A}_0, \tag{111}$$

and

$$\phi(\theta_X(Q_X), \theta_Y(Q_Y)) = 1, \text{ for all } (Q_X, Q_Y) \in \mathcal{A}_1. \tag{112}$$

From (15) and (110), we have, for all $(m_X, m_Y) \in \lceil M_X \rfloor \times \lceil M_Y \rfloor$,

$$\phi(m_X, m_Y) = \max\{\phi_0(m_X, m_Y), \phi_1(m_X, m_Y)\}, \tag{113}$$

$$\phi_0(m_X, m_Y) \cdot \phi_1(m_X, m_Y) = 0. \tag{114}$$

Therefore, we obtain, for each $(Q_X, Q_Y) \in \mathcal{A}_0$,

$$\phi_0(\theta_X(Q_X), \theta_Y(Q_Y)) = \phi_1(\theta_X(Q_X), \theta_Y(Q_Y)) = 0, \tag{115}$$

and, for each $(Q_X, Q_Y) \in \mathcal{A}_1$,

$$\phi_0(\theta_X(Q_X), \theta_Y(Q_Y)) + \phi_1(\theta_X(Q_X), \theta_Y(Q_Y)) = 1.$$

Furthermore, we can demonstrate that, for either $i = 0$ or $i = 1$,

$$\phi_i(\theta_X(Q_X), \theta_Y(Q_Y)) \equiv 1, \quad \text{for all } (Q_X, Q_Y) \in \mathcal{A}_1. \tag{116}$$

To see this, we define, for $i \in \{0, 1\}$,

$$\begin{aligned} \mathcal{A}_1^{(i)} \triangleq \{&(Q_X, Q_Y) \in \mathcal{A}_1 : \\ &\phi_i(\theta_X(Q_X), \theta_Y(Q_Y)) = 1\}, \end{aligned} \tag{117}$$

from which we obtain the partition $\mathcal{A}_1 = \mathcal{A}_1^{(0)} \cup \mathcal{A}_1^{(1)}$ with $\mathcal{A}_1^{(0)} \cap \mathcal{A}_1^{(1)} = \varnothing$. Then, it suffices to show that $\mathcal{A}_1^{(i)} = \varnothing$ for $i = 0$ or $i = 1$, which we will establish by contradiction.

To begin, suppose we have $(Q_X, Q_Y) \in \mathcal{A}_1^{(0)}$ and $(\tilde{Q}_X, \tilde{Q}_Y) \in \mathcal{A}_1^{(1)}$. Then, let us define sequences $\{(Q_X^{(n)}, Q_Y^{(n)})\}_{n\geq 0}$ and $\{(\tilde{Q}_X^{(n)}, \tilde{Q}_Y^{(n)})\}_{n\geq 0}$ such that

$$(Q_X^{(0)}, Q_Y^{(0)}) = (Q_X, Q_Y), \quad (\tilde{Q}_X^{(0)}, \tilde{Q}_Y^{(0)}) = (\tilde{Q}_X, \tilde{Q}_Y).$$

Moreover, for each $n \geq 0$, we define

$$(Q_X^{(n+1)}, Q_Y^{(n+1)}) \triangleq \begin{cases} (\hat{Q}_X^{(n)}, \hat{Q}_Y^{(n)}) & \text{if } (\hat{Q}_X^{(n)}, \hat{Q}_Y^{(n)}) \in \mathcal{A}_1^{(0)}, \\ (Q_X^{(n)}, Q_Y^{(n)}) & \text{otherwise,} \end{cases}$$

and

$$(\tilde{Q}_X^{(n+1)}, \tilde{Q}_Y^{(n+1)}) \triangleq \begin{cases} (\tilde{Q}_X^{(n)}, \tilde{Q}_Y^{(n)}), & \text{if } (\hat{Q}_X^{(n)}, \hat{Q}_Y^{(n)}) \in \mathcal{A}_1^{(0)} \\ (\hat{Q}_X^{(n)}, \hat{Q}_Y^{(n)}), & \text{otherwise,} \end{cases}$$

where we have defined

$$\hat{Q}_X^{(n)} \triangleq \frac{1}{2}(Q_X^{(n)} + \tilde{Q}_X^{(n)}), \quad \hat{Q}_Y^{(n)} \triangleq \frac{1}{2}(Q_Y^{(n)} + \tilde{Q}_Y^{(n)}),$$

and we have $(\hat{Q}_X^{(n)}, \hat{Q}_Y^{(n)}) \in \mathcal{A}_1$ due to the convexity of $\mathcal{A}_1$.

Then, for each $n \geq 0$, it can be verified that

$$(Q_X^{(n)}, Q_Y^{(n)}) \in \mathcal{A}_1^{(0)}, \quad (\tilde{Q}_X^{(n)}, \tilde{Q}_Y^{(n)}) \in \mathcal{A}_1^{(1)}, \tag{118}$$

and

$$\begin{aligned} d_\star &\left(Q_X^{(n)} Q_Y^{(n)}, \tilde{Q}_X^{(n)} \tilde{Q}_Y^{(n)}\right) \\ &= \frac{1}{2} \cdot d_\star \left(Q_X^{(n-1)} Q_Y^{(n-1)}, \tilde{Q}_X^{(n-1)} \tilde{Q}_Y^{(n-1)}\right) \\ &= \frac{1}{2^n} \cdot d_\star \left(Q_X^{(0)} Q_Y^{(0)}, \tilde{Q}_X^{(0)} \tilde{Q}_Y^{(0)}\right) \\ &= \frac{1}{2^n} \cdot d_\star \left(Q_X Q_Y, \tilde{Q}_X \tilde{Q}_Y\right). \end{aligned}$$

As a result, we obtain

$$\begin{aligned} d_\star &(Q_X^{(n)} \tilde{Q}_Y^{(n)}, Q_X^{(n)} Q_Y^{(n)}) \\ &= d_{\max}(\tilde{Q}_Y^{(n)}, Q_Y^{(n)}) \\ &\leq d_\star \left(Q_X^{(n)} Q_Y^{(n)}, \tilde{Q}_X^{(n)} \tilde{Q}_Y^{(n)}\right) \\ &\leq \frac{1}{2^n} \cdot d_\star \left(Q_X Q_Y, \tilde{Q}_X \tilde{Q}_Y\right) = o(1). \end{aligned}$$

Since $\mathcal{A}_1$ is open, for sufficiently large n we have $Q_X^{(n)} \tilde{Q}_Y^{(n)} \in \mathcal{A}_1$. Thus, it follows from (112) that

$$\phi(\theta_X(Q_X^{(n)}), \theta_Y(\tilde{Q}_Y^{(n)})) = 1. \tag{119}$$

In addition, from (117) and (118), we have

$$\theta_X(Q_X^{(n)}) \in \mathcal{I}_X^{(1)}(\phi_0) \quad \text{and} \quad \theta_Y(\tilde{Q}_Y^{(n)}) \in \mathcal{I}_Y^{(1)}(\phi_1), \tag{120}$$

where $\mathcal{I}_X^{(1)}(\cdot)$ and $\mathcal{I}_Y^{(1)}(\cdot)$ are as defined in (16). This implies that (cf. Definition 7)

$$\phi(\theta_X(Q_X^{(n)}), \theta_Y(\tilde{Q}_Y^{(n)})) = 0,$$

which contradicts (119).

Hence, we obtain (116) as desired. Finally, it follows from (115) that $(\mathcal{A}_0, \mathcal{A}_1)$ is separable under ϕ_j for some $j \in \{0, 1\}$. ∎

Our proof of Lemma 6 proceeds as follows. First, we define a mapping $\kappa\colon \mathcal{F} \to \mathbb{N}$ to indicate the reducibility of decoders. Specifically, if ϕ is completely reducible, we let $\kappa(\phi) \triangleq 0$; otherwise, suppose $\omega^*(\phi) \in \mathcal{F}_{L_X, L_Y}$ for some $L_X, L_Y \geq 2$, then we define $\kappa(\phi) \triangleq \min\{L_X, L_Y\}$, where $\omega^*(\phi)$ denotes the reduced form of ϕ as defined in Proposition 4.

If $M_Y = 1$, we have $\bar{\Omega}_{M_X, M_Y}^{(1)} \subset \bar{\Omega}_{M_X, M_Y} = \varnothing$, and Lemma 6 is trivially true. Thus, it suffices to consider the case $M_X, M_Y \geq 2$. In particular, we will show that, for each $\phi \in \bar{\Omega}_{M_X, M_Y}^{(1)}$, if $(E_0, E_1) \in \mathcal{E}[\phi]$, then there exists $\phi' \in \mathcal{F}_{M_X, M_Y}$, such that

$$(E_0, E_1) \in \mathcal{E}[\phi'] \quad \text{and} \quad \kappa(\phi') < \kappa(\phi). \tag{121}$$

It can be shown that Lemma 6 can be readily obtained from (121). Indeed, note that from (121), for each $\phi \in \bar{\Omega}_{M_X, M_Y}^{(1)}$ and $(E_0, E_1) \in \mathcal{E}[\phi]$, we can obtain some ϕ' satisfying (121). Similarly, if $\phi' \in \bar{\Omega}_{M_X, M_Y}^{(1)}$, we can again apply (121) to obtain an $M_X \times M_Y$ decoder ϕ'' with $\kappa(\phi'') < \kappa(\phi')$ and $(E_0, E_1) \in \mathcal{E}[\phi'']$. Since $\kappa(\cdot)$ is non-negative, for each $\phi \in \bar{\Omega}_{M_X, M_Y}^{(1)}$ and error exponent pair $(E_0, E_1) \in \mathcal{E}[\phi]$, we can repeatedly apply these procedures to obtain

$$\tilde{\phi} \in \mathcal{F}_{M_X, M_Y} \setminus \bar{\Omega}_{M_X, M_Y}^{(1)} = \Omega_{M_X, M_Y} \cup \bar{\Omega}_{M_X, M_Y}^{(0)},$$

such that $(E_0, E_1) \in \mathcal{E}[\tilde{\phi}]$, which demonstrates Lemma 6.

It remains only to establish (121). To this end, suppose we have a decoder $\phi \in \bar{\Omega}_{M_X, M_Y}^{(1)}$ for some $M_X, M_Y \geq 2$, and an error exponent pair $(E_0, E_1) \in \mathcal{E}[\phi]$. Let $\psi \triangleq \omega^*(\phi)$ denote the reduced form of ϕ, as defined in Proposition 4. Furthermore, suppose ψ can be reduced from ϕ by a reduction operator ω, i.e., $\psi = \omega(\phi)$. Suppose $\psi \in \mathcal{F}_{L_X, L_Y}$ for some $L_X \leq M_X$ and $L_Y \leq M_Y$. Without loss of generality, we assume that, for all $(m_X, m_Y) \in \lceil L_X \rfloor \times \lceil L_Y \rfloor$,

$$\psi(m_X, m_Y) = \phi(m_X, m_Y).$$

Then, for $i \in \{0, 1\}$, we define $\mathcal{A}_i \triangleq \mathcal{D}_i(E_i)$ and $\mathcal{A}'_i \triangleq \tau_i(\mathcal{A}_0, \mathcal{A}_1; \omega)$, with τ_i as defined in (107)–(108). Since $(E_0, E_1) \in \mathcal{E}[\phi]$, it follows from Theorem 2 that $(\mathcal{A}_0, \mathcal{A}_1)$ is separable under ϕ. From Fact 11, $(\mathcal{A}'_0, \mathcal{A}'_1)$ is separable under ψ.

Note that from Fact 1 and Fact 9, both $\mathcal{A}_0$ and $\mathcal{A}_1$ are convex and open. Hence, it follows from Fact 10 that, $\mathcal{A}'_0$ and $\mathcal{A}'_1$ are also convex and open. Then, from the definition of $\bar{\Omega}_{M_X, M_Y}^{(1)}$ [cf. (39)], ψ is decomposable and has the decomposition

$$\psi = \psi_0 \oplus \psi_1 \oplus \bar{\imath} \tag{122}$$

for some $i \in \{0, 1\}$, where $\psi_0, \psi_1 \in \mathcal{F}_{L_X, L_Y}$ satisfy (15).

Hence, it follows from Lemma 8 that $(\mathcal{A}'_0, \mathcal{A}'_1)$ is separable under ψ_0 or ψ_1. Let us define $\phi_0, \phi_1 \in \mathcal{F}_{M_X, M_Y}$ such that, for each $j \in \{0, 1\}$,

$$\begin{aligned} &\phi_j(m_X, m_Y) \\ &\triangleq \begin{cases} \psi_j(m_X, m_Y) & \text{if } (m_X, m_Y) \in \lceil L_X \rfloor \times \lceil L_Y \rfloor, \\ \phi(m_X, m_Y) & \text{otherwise.} \end{cases} \end{aligned}$$

Then, it can be verified that $(\mathcal{A}_0, \mathcal{A}_1)$ is separable under ϕ_0 or ϕ_1, which implies that $(E_0, E_1) \in \mathcal{E}[\phi_j]$ for some $j \in \{0, 1\}$.

Finally, from the definition of $\kappa(\cdot)$, for both $j \in \{0, 1\}$,

$$\kappa(\phi_j) = \kappa(\psi_j) \tag{123}$$
$$\leq \min\{|\mathcal{I}_X^{(i)}(\psi_j)|, |\mathcal{I}_Y^{(i)}(\psi_j)|\} \tag{124}$$
$$< \min\{L_X, L_Y\} \tag{125}$$
$$= \kappa(\psi) = \kappa(\phi), \tag{126}$$

where $\mathcal{I}_X^{(1)}(\cdot)$ and $\mathcal{I}_Y^{(1)}(\cdot)$ are as defined in (16), and where to obtain the inequalities (124) and (125), we have used (15). Hence, we obtain (121) as desired. ■

Appendix M
Proof of Lemma 7

The following proposition is useful in our proof.

Proposition 5: If $(M_X - 2)(M_Y - 2) < 2$, there exists no $M_X \times M_Y$ decoder that is both indecomposable and irreducible.

Proof of Proposition 5: To begin, for each $\phi \in \mathcal{F}_{M_X,M_Y}$, we define the bipartite graph $G_\phi = (U, V, E_\phi)$ with the vertex sets

$$U \triangleq \{\mathsf{u}_{m_X} : m_X \in \lceil M_X \rfloor\}, \quad V \triangleq \{\mathsf{v}_{m_Y} : m_Y \in \lceil M_Y \rfloor\}$$

and the edge sets

$$E_\phi \triangleq \{(\mathsf{u}_{m_X}, \mathsf{v}_{m_Y}) \colon \phi(m_X, m_Y) = 1\},$$

where $(\mathsf{u}_{m_X}, \mathsf{v}_{m_Y})$ represents the undirected edge connecting u_{m_X} and v_{m_Y}. This establishes the one-to-one correspondence between decoders and bipartite graphs, and it can be verified that, the decision matrix $\mathbf{A}$ associated with ϕ corresponds to the biadjacency matrix of G_ϕ.

We then illustrate that if ϕ is indecomposable and irreducible, then both G_ϕ and $G_{\bar{\phi}}$ are connected. To this end, first note that since ϕ is irreducible, there exists no isolated vertex in G_ϕ.

Now, suppose G_ϕ is disconnected and can be divided into bipartite graphs $G^{(0)} = (U_0, V_0, E^{(0)})$ and $G^{(1)} = (U_1, V_1, E^{(1)})$, with non-empty vertex sets U_0, U_1, V_0, V_1 satisfying

$$U = U_0 \cup U_1, \quad U_0 \cap U_1 = \varnothing,$$
$$V = V_0 \cup V_1, \quad V_0 \cap V_1 = \varnothing.$$

Let ϕ_0 and ϕ_1 be the decoders associated with $G^{(0)}$ and $G^{(1)}$, respectively. Then, it can be verified that ϕ satisfies (14) with $i = 0$, and thus is decomposable, which contradicts our assumption. Therefore, G_ϕ is connected. Via a symmetry argument, we can show that $G_{\bar{\phi}}$ is also connected.

Therefore, we obtain

$$|E_\phi| \geq |U| + |V| - 1 \tag{127a}$$
$$|E_{\bar{\phi}}| \geq |U| + |V| - 1, \tag{127b}$$

where we have used the simple fact that each connected graph with k vertices has at least $k - 1$ edges.

From (127), we obtain

$$M_X M_Y = |E_\phi| + |E_{\bar{\phi}}|$$
$$\geq 2(|U| + |V| - 1) = 2(M_X + M_Y - 1),$$

which is equivalent to

$$(M_X - 2)(M_Y - 2) \geq 2. \tag{128}$$

As a result, if $(M_X - 2)(M_Y - 2) < 2$, no $M_X \times M_Y$ decoder is both indecomposable and irreducible. ■

Our proof of Lemma 7 proceeds as follows. First, for each $\phi \in \bar{\Omega}^{(0)}_{M_X,M_Y}$, let $\psi \triangleq \omega^*(\phi)$. Then, we have $\psi \in \mathcal{F}_{L_X,L_Y}$ for some $L_X \leq M_X, L_Y \leq M_Y$, and ψ is both irreducible and indecomposable.

Note that if $(M_X - 2)(M_Y - 2) < 2$, we have $(L_X - 2)(L_Y - 2) < 2$. Then, it follows from Proposition 5 that such ψ does not exist. As a result, we have $\bar{\Omega}^{(0)}_{M_X,M_Y} = \varnothing$.

Hence, from Theorem 4, Φ_{M_X,M_Y} is sufficient for $\mathcal{F}_{M_X,M_Y}$, and thus

$$\mathcal{E}(0_{M_X}, 0_{M_Y}) = \mathcal{E}[\mathcal{F}_{M_X,M_Y}] = \mathcal{E}[\Phi_{M_X,M_Y}]$$
$$= \mathcal{E}[\varphi_{M_X,M_Y}] \cup \mathcal{E}[\bar{\varphi}_{M_X,M_Y}],$$

where the first equality follows from Fact 4. ■

Appendix N
Proof of Theorem 5

For given M_X, M_Y, note that if $\bar{\Omega}_{M_X,M_Y}$ and Ω_{M_X,M_Y} as defined in (36) satisfy

$$\bar{\Omega}_{M_X,M_Y} \preceq \Omega_{M_X,M_Y}, \tag{129}$$

from Fact 2 we have

$$\mathcal{F}_{M_X,M_Y} = \bar{\Omega}_{M_X,M_Y} \cup \Omega_{M_X,M_Y} \preceq \Omega_{M_X,M_Y},$$

and thus

$$\mathcal{E}(0_{M_X}, 0_{M_Y}) = \mathcal{E}[\mathcal{F}_{M_X,M_Y}] = \mathcal{E}[\Omega_{M_X,M_Y}] \tag{130}$$
$$= \mathcal{E}[\Phi_{M_X,M_Y}], \tag{131}$$

where the first equality follows from Fact 4, where the second equality follows from (130) and Definition 8, and where the last equality follows from Lemma 5.

Therefore, it suffices to establish (129). Note that if $M_Y = 1$, then $\bar{\Omega}_{M_X,M_Y} = \varnothing$, and (130) is trivially true. We then establish (129) for $M_X \geq M_Y \geq 2$. To this end, we show that for each $\phi \in \bar{\Omega}_{M_X,M_Y}$, there exists $\phi' \in \Omega_{M_X,M_Y}$, such that $\mathcal{E}[\phi] \subset \mathcal{E}[\phi']$.

To begin, note that from statement S2 of Proposition 3, ϕ has at least one irreducible 2×2 subdecoder [cf. (98)]. Without loss of generality, we assume

$$\phi(0, 0) = \phi(1, 1) = 0,$$
$$\phi(1, 0) = \phi(1, 0) = 1.$$

By symmetry, it suffices to consider the case

$$P^{(0)}_{XY} = P^{(0)}_X P^{(0)}_Y. \tag{132}$$

Let $\phi^{(0)} \triangleq \phi$, and suppose $f_n \colon \mathcal{X}^n \to \lceil M_X \rfloor$ and $g_n \colon \mathcal{Y}^n \to \lceil M_Y \rfloor$ are some given encoders. Then, we define $\phi^{(1)}$ as

$$\phi^{(1)}(m_X, m_Y) \triangleq \begin{cases} 0 & \text{if } (m_X, m_Y) = (j_X, \bar{\jmath}_X), \\ \phi^{(0)}(m_X, m_Y) & \text{otherwise,} \end{cases} \tag{133}$$

where we have defined

$$j_X \triangleq \arg\min_{j\in\{0,1\}} \mathbb{P}\left\{f_n(X^n) = j | \mathsf{H} = 0\right\} \tag{134}$$

and $\bar{\jmath}_X \triangleq 1 - j_X$.

For $k = 0, 1$, let $\mathcal{C}_n^{(k)} \triangleq (f_n, g_n, \phi^{(k)})$ denote the corresponding coding schemes. Then, it can be verified that the type-I and type-II errors for $\mathcal{C}_n^{(1)}$ satisfy

$$\pi_0(\mathcal{C}_n^{(1)}) \leq 2 \cdot \pi_0(\mathcal{C}_n^{(0)}), \tag{135a}$$

$$\pi_1(\mathcal{C}_n^{(1)}) \leq \pi_1(\mathcal{C}_n^{(0)}). \tag{135b}$$

To establish (135a), note that

$$\begin{aligned}
&\pi_0(\mathcal{C}_n^{(1)}) - \pi_0(\mathcal{C}_n^{(0)}) \\
&= \mathbb{P}\left\{(f_n(X^n), g_n(Y^n)) = (j_X, \bar{\jmath}_X) | \mathsf{H} = 0\right\} && (136)\\
&= \mathbb{P}\left\{f_n(X^n) = j_X | \mathsf{H} = 0\right\} \mathbb{P}\left\{g_n(Y^n) = \bar{\jmath}_X | \mathsf{H} = 0\right\} && (137)\\
&\leq \mathbb{P}\left\{f_n(X^n) = \bar{\jmath}_X | \mathsf{H} = 0\right\} \mathbb{P}\left\{g_n(Y^n) = \bar{\jmath}_X | \mathsf{H} = 0\right\} && (138)\\
&= \mathbb{P}\left\{(f_n(X^n), g_n(Y^n)) = (\bar{\jmath}_X, \bar{\jmath}_X) | \mathsf{H} = 0\right\} && (139)\\
&\leq \pi_0(\mathcal{C}_n^{(0)}), && (140)
\end{aligned}$$

where (137) and (139) follow from (132), and where (138) follows from (134).

Moreover, (141b) follows from the simple fact that, for all $(m_X, m_Y) \in \lceil M_X \rfloor \times \lceil M_Y \rfloor$,

$$\phi^{(1)}(m_X, m_Y) = 1 \quad \text{implies} \quad \phi^{(0)}(m_X, m_Y) = 1.$$

Furthermore, if $\phi^{(1)} \notin \Omega_{M_X,M_Y}$, we can define $\phi^{(2)}$ similar to (133). Similarly, for each $k \geq 0$, we define $\phi^{(k+1)}$ if $\phi^{(k)} \notin \Omega_{M_X,M_Y}$. Then, there exists $k' \leq M_X M_Y - 1$, such that $\phi^{(k')} \in \Omega_{M_X,M_Y}$.

To see this, it suffices to note that, for all $k \geq 0$, we have

$$0 \leq \sigma_{XY}(\phi^{(k)}) = \sigma_{XY}(\phi^{(0)}) - k \leq M_X M_Y - 1 - k,$$

where we have defined, for each $\phi \in \mathcal{F}_{M_X,M_Y}$,

$$\sigma_{XY}(\phi) \triangleq \sum_{m_X \in \lceil M_X \rfloor} \sum_{m_Y \in \lceil M_Y \rfloor} \phi(m_X, m_Y).$$

In addition, similar to (133), for each k we have

$$\pi_0(\mathcal{C}_n^{(k)}) \leq (k+1) \cdot \pi_0(\mathcal{C}_n^{(0)}), \tag{141a}$$

$$\pi_1(\mathcal{C}_n^{(k)}) \leq \pi_1(\mathcal{C}_n^{(0)}). \tag{141b}$$

This implies that

$$\pi_0(\mathcal{C}_n^{(k')}) \leq (k'+1)\pi_0(\mathcal{C}_n^{(0)}) \leq M_X M_Y \cdot \pi_0(\mathcal{C}_n^{(0)}), \tag{142a}$$

$$\pi_1(\mathcal{C}_n^{(k')}) \leq \pi_1(\mathcal{C}_n^{(0)}). \tag{142b}$$

Let $\phi' \triangleq \phi^{(k')} \in \Omega_{M_X,M_Y}$. Then, since the encoders f_n and g_n can be arbitrarily chosen, from (142) we obtain $\mathcal{E}[\phi'] \subset \mathcal{E}[\phi]$, which completes the proof. ■

Acknowledgment

The work of Shao-Lun Huang was supported in part by the National Natural Science Foundation of China under Grant 61807021, in part by the Shenzhen Science and Technology Program under Grant KQTD20170810150821146, and in part by the Innovation and Entrepreneurship Project for Overseas High-Level Talents of Shenzhen under Grant KQJSCX20180327144037831.

References

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial Intelligence and Statistics*. PMLR, 2017, pp. 1273–1282.

[2] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE transactions on information theory*, vol. 33, no. 6, pp. 759–772, 1987.

[3] T. S. Han and K. Kobayashi, "Exponential-type error probabilities for multiterminal hypothesis testing," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 2–14, 1989.

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

[5] S. Watanabe, "Neyman–pearson test for zero-rate multiterminal hypothesis testing," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 4923–4939, 2017.

[6] T. Berger, "Decentralized estimation and decision theory," in *IEEE Seven Springs Workshop on Information Theory, Mt. Kisco, NY*, 1979.

[7] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE transactions on information theory*, vol. 32, no. 4, pp. 533–542, 1986.

[8] S.-I. Amari and T. S. Han, "Statistical inference under multiterminal rate restrictions: a differential geometric approach," *IEEE Transactions on Information Theory*, vol. 35, no. 2, pp. 217–227, 1989.

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

[10] S. Sreekumar and D. Gunduz, "Distributed hypothesis testing over a noisy channel: Error-exponents trade-off," *arXiv preprint arXiv:1908.07521*, 2019.

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

[12] R. Ahlswede, P. Gács, and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 34, no. 2, pp. 157–177, 1976.

[13] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer, 1998. [Online]. Available: https://doi.org/10.1007/978-1-4612-5320-4

[14] J. Neyman and E. Pearson, "On the problems of the most efficient tests of statistical hypotheses." *Philosophical Transactions of the Royal Society of London*, 1933.

[15] W. Hoeffding *et al.*, "On probabilities of large deviations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. The Regents of the University of California, 1967.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience, 2006.