A Sphere-Packing Error Exponent for Mismatched Decoding

Ehsan Asadi Kangarshahi[ID] and Albert Guillén i Fàbregas[ID], *Fellow, IEEE*

Abstract—We derive a sphere-packing error exponent for coded transmission over discrete memoryless channels with a fixed decoding metric. By studying the error probability of the code over an auxiliary channel, we find a lower bound to the probability of error of mismatched decoding. The bound is shown to decay exponentially for coding rates smaller than a new upper bound to the mismatch capacity which is established in this paper. For rates higher than the new upper bound, the error probability is shown to be bounded away from zero. The new upper bound is shown to improve over previous upper bounds to the mismatch capacity.

Index Terms—Mismatched decoding, mismatch, Shannon capacity, channel capacity, decoding, error exponent, sphere packing.

I. INTRODUCTION

COMMUNICATION problems where the receiver needs to employ a suboptimal decoder are typically cast within the mismatched decoding framework [1]. These situations arise when optimal maximum-likelihood decoding cannot be used: i) the channel transition is unknown and imperfectly estimated or, ii) when, for complexity reasons, the channel likelihood is too difficult to compute and an alternative decoding metric is needed. In addition, some important problems in information theory like the zero-error or zero-undetected error capacities can be cast as instances of mismatched decoding [2]. In the mismatched decoding problem, the optimal maximum-likelihood decoder is replaced by a maximum metric decoder, in which the metric is not necessarily the channel likelihood. For a fixed channel W and decoding metric q, finding a single-letter expression for the mismatch capacity $C_q(W)$ remains an open problem and only bounds are known.

A number of single-letter lower bounds have been derived in the literature [2], [3], [4], [5] (see also [1] for a recent survey). A number of lower bounds based on multiuser coding

techniques have been derived [6], [7], [8], some yielding improvements over standard single-user coding. Most of these lower bounds have been derived via random coding which in turn yield single-letter lower bounds on the error exponent. As suggested by [2], multiletter versions of achievable rates can yield strict improvements over their single-letter counterparts.

Instead, up until recently, not much progress had been made on upper bounds. Balakirsky [9] claimed that for binary-input discrete memoryless channels (DMC), the mismatch capacity coincided with the lower bound in [3] and [4]. Reference [10] provided a binary-input ternary-output counterexample to this converse invalidating its claim. In particular, it was shown that the order-2 multiletter version of the multiuser coding rate in [7] and [8] is strictly higher than the bound derived by in [3] and [4]. In [11] (see also [12]), we proposed a single-letter upper bound to the mismatch capacity based on transforming the channel in such a way that errors on the auxiliary channel imply a mismatched decoding error in the original channel. Reference [13] cast the bound in [11] as multicast transmission over a broadcast channel, significantly simplifying the proof. The bounds in [13] improved over that in [11] in several directions. In addition, [13] also provided conditions that a pair of channel and decoding metric must fulfil for the bound to be tight and thus give the mismatch capacity. Recently, further improvements were presented in [14] and [15]. Reference [14] builds on the idea of multicast transmission allowing the possibility that when an error is made in the auxiliary channel, a mismatched decoding error in the original channel is made with a certain probability, instead of deterministically as in [11] and [13]. The bound in [15] is a preliminary part of this work and also relaxes this condition in a different way and will be discussed in detail in this paper. All bounds [11], [13], [14], [15] belong to the same family of constrained minimizations of the mutual information of an auxiliary channel, and can be expressed as

$$C_q(W) \leq \max_{P_X} \min_{\substack{P_{Y\hat{Y}|X} \in \mathcal{M} \\ P_{Y|X}=W}} I(P_X, P_{\hat{Y}|X}), \tag{1}$$

where the set $\mathcal{M}$ quantifies the statistical relationships among the channel input X, output Y and auxiliary channel output $\hat{Y}$, ensuring, either deterministically or probabilistically, that errors in the auxiliary channel induce mismatched decoding errors in the true channel. The set $\mathcal{M}$ may depend on the input distribution P_X. Therefore, it is of interest to enlarge the set $\mathcal{M}$ of joint conditional distributions $P_{Y\hat{Y}|X}$, or broadcast

channels, such that the aforementioned error condition is fulfilled.

Not many single-letter upper bounds on the error exponent of mismatched decoding are available, other than the trivial upper-bounds to the standard channel coding problem. In a recent paper [16], the authors proved that the expurgated error exponent derived in [17] at rate zero is tight for a wide class of channels and decoding metrics. In this paper, we derive a sphere-packing upper bound to the error exponent of mismatched decoding. We also show that the rate where the sphere-packing upperbound becomes equal to zero is a new upper bound on the mismatch capacity.

This paper is organized as follows. Section II introduces the notation and preliminary concepts. Section III introduces the main results of the paper, the new upper bound to the mismatch capacity, the sphere-packing bound to the error exponent and a comparison of the new bound and previously proposed bounds. Section IV discusses an optimization interpretation of the design of the set $\mathcal{M}$. Proofs of the main results can be found in Section V. Proofs of auxiliary results can be found in the appendices.

II. PRELIMINARIES

We consider reliable communication over a DMC W defined over input and output alphabets $\mathcal{X} = \{1, 2, \ldots, J\}$ and $\mathcal{Y} = \{1, 2, \ldots, K\}$. We denote the channel transition probability by $W(k|j)$. A codebook $\mathcal{C}_n$ is defined as a set of M sequences $\mathcal{C}_n = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$, where $\boldsymbol{x}_m = (x_{1,m}, \ldots, x_{n,m}) \in \mathcal{X}^n$, for $m \in \{1, \ldots, M\}$. A message $m \in \{1, \ldots, M\}$ is chosen equiprobably and $\boldsymbol{x}_m$ is sent over the channel. The channel produces a noisy observation $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathcal{Y}^n$ according to $W^n(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} W(y_i|x_i)$.

Upon observing $\boldsymbol{y} \in \mathcal{Y}^n$ the decoder produces an estimate of the transmitted message $\hat{m} \in \{1, \ldots, M\}$. The average and maximal error probabilities are respectively defined as

$$P_e(\mathcal{C}_n) = \mathbb{P}[\hat{m} \neq m] \tag{2}$$

$$P_{e,\max}(\mathcal{C}_n) = \max_{m \in \{1,\ldots,M\}} \mathbb{P}[\hat{m} \neq m | m \text{ is sent}]. \tag{3}$$

The decoder that minimizes the error probability is the maximum-likelihood (ML) decoder, that produces the message estimate $\hat{m}$ according to

$$\hat{m} = \arg\max_{m \in \{1,\ldots,M\}} W^n(\boldsymbol{y}|\boldsymbol{x}_m). \tag{4}$$

Rate $R > 0$ is said to be achievable if for any $\epsilon > 0$ there exists a sequence of length-n codebooks $\{\mathcal{C}_n\}_{n=1}^{\infty}$ such that $|\mathcal{C}_n| \geq 2^{n(R-\epsilon)}$, and $\liminf_{n \to \infty} P_e(\mathcal{C}_n) = 0$. The capacity of W, denoted by $C(W)$, is defined as the largest achievable rate.

In situations with channel uncertainty, it is not possible to use ML decoding and instead, the decoder produces the message estimate $\hat{m}$ as

$$\hat{m} = \arg\max_{m \in \{1,\ldots,M\}} q^n(\boldsymbol{x}_m, \boldsymbol{y}), \tag{5}$$

where $q^n(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{n} q(x_i, y_i)$ and $q : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is the decoding metric. We refer to this decoder as q-decoder.

When $q(x, y) = \log W(y|x)$, the decoder is ML, otherwise, the decoder is said to be mismatched [1], [2], [3], [4], [5]. The average and maximal error probabilities of codebook $\mathcal{C}_n$ under q-decoding are respectively denoted by $P_e^q(\mathcal{C}_n, W)$ and $P_{e,\max}^q(\mathcal{C}_n, W)$. The mismatch capacity $C_q(W)$ is defined as supremum of all achievable rates with q-decoding.

The method of types [18, Ch. 2] will be used extensively in this paper. We recall some of the basic definitions and introduce some notation. The type of a sequence $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ is a column vector representing its empirical distribution, i.e., $\hat{\boldsymbol{p}}_{\boldsymbol{x}}(j) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{x_i = j\}$. The set of all types of $\mathcal{X}^n$ is denoted by $\mathcal{P}_n(\mathcal{X})$. For $\boldsymbol{p}_X \in \mathcal{P}_n(\mathcal{X})$, the type class $\mathcal{T}(\boldsymbol{p}_X)$ is set of all sequences in $\mathcal{X}^n$ with type $\boldsymbol{p}_X$, $\mathcal{T}(\boldsymbol{p}_X) = \{\boldsymbol{x} \in \mathcal{X}^n | \hat{\boldsymbol{p}}_{\boldsymbol{x}} = \boldsymbol{p}_X\}$. The joint type of sequences $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n) \in \mathcal{Y}^n$ is defined as a matrix representing their empirical distribution

$$\hat{\boldsymbol{p}}_{\boldsymbol{x}\boldsymbol{y}}(j, k) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{x_i = j, y_i = k\}. \tag{6}$$

The set of joint types on $\mathcal{X}^n, \mathcal{Y}^n$ is given by $\mathcal{P}_n(\mathcal{X}\mathcal{Y})$. The conditional type of $\boldsymbol{y}$ given $\boldsymbol{x}$ is the matrix

$$\hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}}(k|j) = \begin{cases} \frac{\hat{\boldsymbol{p}}_{\boldsymbol{x}\boldsymbol{y}}(j,k)}{\hat{\boldsymbol{p}}_{\boldsymbol{x}}(j)} & \hat{\boldsymbol{p}}_{\boldsymbol{x}}(j) > 0 \\ \frac{1}{|\mathcal{Y}|} & \text{otherwise.} \end{cases} \tag{7}$$

The set of conditional types on $\mathcal{Y}^n$ given $\mathcal{X}^n$ is denoted by $\mathcal{P}_n(\mathcal{Y}|\mathcal{X})$. For $\boldsymbol{p}_{Y|X} \in \mathcal{P}_n(\mathcal{Y}|\mathcal{X})$ and sequence $\boldsymbol{x} \in \mathcal{T}(\boldsymbol{p}_X)$, the conditional type class $\mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})$ is defined as $\mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X}) = \{\boldsymbol{y} \in \mathcal{Y}^n | \hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}} = \boldsymbol{p}_{Y|X}\}$.

Similarly, we can define the joint type of $\boldsymbol{x}, \boldsymbol{y}, \hat{\boldsymbol{y}}$, as the empirical distribution of the triplet. For $j \in \mathcal{X}$ and $k_1, k_2 \in \mathcal{Y}$,

$$\hat{\boldsymbol{p}}_{\boldsymbol{x}\boldsymbol{y}\hat{\boldsymbol{y}}}(j, k_1, k_2) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{x_i = j, y_i = k_1, \hat{y}_i = k_2\}. \tag{8}$$

We define the joint conditional type of $\boldsymbol{y}, \hat{\boldsymbol{y}}$ given $\boldsymbol{x} \in \mathcal{T}(\boldsymbol{p}_X)$ as

$$\hat{\boldsymbol{p}}_{\boldsymbol{y}\hat{\boldsymbol{y}}|\boldsymbol{x}}(k_1, k_2|j) = \begin{cases} \frac{\hat{\boldsymbol{p}}_{\boldsymbol{x}\boldsymbol{y}\hat{\boldsymbol{y}}}(j,k_1,k_2)}{\hat{\boldsymbol{p}}_{\boldsymbol{x}}(j)} & \hat{\boldsymbol{p}}_{\boldsymbol{x}}(j) > 0 \\ \frac{1}{|\mathcal{Y}|} \mathbb{1}\{k_1 = k_2\} & \text{otherwise.} \end{cases} \tag{9}$$

The set of all joint conditional types is denoted by $\mathcal{P}_n(\mathcal{Y}\hat{\mathcal{Y}}|\mathcal{X})$. Additionally, for $\boldsymbol{p}_{Y\hat{Y}|X} \in \mathcal{P}_n(\mathcal{Y}\hat{\mathcal{Y}}|\mathcal{X})$ we define:

$$\mathcal{T}_{\boldsymbol{y}\boldsymbol{x}}(\boldsymbol{p}_{Y\hat{Y}|X}) = \{\hat{\boldsymbol{y}} \in \mathcal{Y}^n | \hat{\boldsymbol{p}}_{\boldsymbol{y}\hat{\boldsymbol{y}}|\boldsymbol{x}} = \boldsymbol{p}_{Y\hat{Y}|X}\}. \tag{10}$$

Throughout the paper use the notation $\boldsymbol{p}_Y \boldsymbol{p}_{X|Y}$ to denote the distribution P_{XY} defined by

$$P_{XY}(j, k) = \boldsymbol{p}_Y(k) \boldsymbol{p}_{X|Y}(j|k) \tag{11}$$

Note that the former multiplication for two generic types is not necessarily a type, therefore we denote the result with the probability distribution notation.

The mutual information and conditional relative entropy are respectively defined as

$$I(P_X, P_{Y|X}) \triangleq \mathbb{E}\left[\log \frac{P_{Y|X}(Y|X)}{\sum_{x'} P_X(x') P_{Y|X}(Y|x')}\right],$$
(12)

$$D(P_{Y'|X} \| P_{Y|X} | P_X) \triangleq \sum_{x \in \mathcal{X}} P_X(x) \cdot D(P_{Y'|X=x} \| P_{Y|X=x}).$$
(13)

Definition 1: A random variable X is called sub-Gaussian with parameter (sub-Gaussian norm) θ if for any $\xi > 0$ we have

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq \xi] \leq \Gamma e^{\frac{-\theta^2 \xi^2}{2}},$$
(14)

where Γ is a constant. Throughout the paper we use $\Gamma = 2$ for simplicity of notation. Additionally, $\Gamma = 2$ is sufficient for the relevant random variables to our proofs. Therefore, we use the following definition instead of the previous one

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq \xi] \leq 2e^{\frac{-\theta^2 \xi^2}{2}}.$$
(15)

Definition 2: Let P, Q be probability distributions defined on alphabet $\mathcal{X}$. Then, the infinity norm between P, Q is defined as

$$|P - Q|_\infty = \max_{x \in \mathcal{X}} |P(x) - Q(x)|.$$
(16)

Throughout the paper and proofs, whenever we use $|P - Q|_\infty$ we will implicitly assume that P is absolutely continuous with respect to Q and vice versa.

Definition 3: Let $\mathcal{C}_n = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}$ be a codebook and $\boldsymbol{p}_{Y|X}$ be a conditional type. The maximum type-conflict error probability is defined as

$$P_{\text{tce}}^{\max}(\mathcal{C}_n, \boldsymbol{p}_{Y|X})$$
$$\triangleq \max_{m \in \{1, \ldots, M\}} \mathbb{P}\left[\bigcup_{\bar{m} \neq m} \{\hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}_m} = \hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}_{\bar{m}}} = \boldsymbol{p}_{Y|X}\}\right], \quad (17)$$

where the probability is with respect to the uniform distribution over the type class $\mathcal{T}_{\boldsymbol{x}_m}(\boldsymbol{p}_{Y|X})$.

Similarly to [11], the main idea of this paper is to relate the type-conflict error performance of a given codebook over an auxiliary channel V with the q-decoding performance of the same code over channel W. The main reason for studying type-conflict errors is that an equation of the form $\hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}_2} = \hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}_1}$ provides more information about the properties of the error than ML errors, where we simply have a scalar likelihood inequality. In addition, it can be shown that for rates $R > C(V)$, then the probability of type-conflict errors bounded away from zero.

We proceed by introducing a few definitions. Recall the definition of maximal set from [11]. Consider the set

$$\mathcal{S}_q(k_1, k_2) \triangleq \left\{j \in \mathcal{X} \,\middle|\, j = \arg\max_{j' \in \mathcal{X}} q(j', k_2) - q(j', k_1)\right\}.$$
(18)

A joint conditional distribution $P_{Y\hat{Y}|X}$ is said to be maximal if for all $(j, k_1, k_2) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Y}$,

$$P_{Y\hat{Y}|X}(k_1, k_2|j) = 0 \quad \text{if} \quad j \notin \mathcal{S}_q(k_1, k_2).$$
(19)

The set of all maximal joint conditional distributions was defined to be $\mathcal{M}_{\max}(q)$. In this work, for a given distribution P_{X_1}, we define the set of maximal joint conditional distributions as follows.

Definition 4: $\mathcal{M}_{\max}(q, P_{X_1})$ is the set of all joint conditional distributions $P_{Y\hat{Y}|X_1}$ such that

$$\min_{\substack{P_{X_2|X_1\hat{Y}}: \\ X_2 - X_1\hat{Y} - Y \\ P_{\hat{Y}X_2} = P_{\hat{Y}X_1}}} \mathbb{E}[q(X_2, Y)] \geq \mathbb{E}[q(X_1, Y)], \quad (20)$$

where the notation $X_2 - X_1\hat{Y} - Y$ denotes that $X_2, (X_1\hat{Y})$ and Y form a Markov chain.

We close this section by showing that that $\mathcal{M}_{\max}(q) \subset \mathcal{M}_{\max}(q, P_{X_1})$ for any input distribution P_{X_1}. Assume that $P_{Y\hat{Y}|X_1} \in \mathcal{M}_{\max}(q)$. Then from [11, Lemma 3] we have for any X_2 such that $P_{\hat{Y}X_1} = P_{\hat{Y}X_2}$

$$\mathbb{E}[q(X_2, Y)] \geq \mathbb{E}[q(X_1, Y)].$$
(21)

This implies that $P_{Y\hat{Y}|X_1}$ satisfies (20) and as a result, $P_{Y\hat{Y}|X_1} \in \mathcal{M}_{\max}(q, P_{X_1})$. As we will show, this enlarged set of maximal distributions yields an improved upper bound on the mismatch capacity. Throughout the paper, we have taken the convention that X_1 represents the sent codeword and X_2 represents an auxiliary codeword.

III. MAIN RESULTS

In this section, we introduce over the main results of this paper. We first introduce an upper bound to the mismatch capacity.

Theorem 1: Let W, q be channel and decoding metric, respectively. Then,

$$C_q(W) \leq \bar{R}(W, q).$$
(22)

where

$$\bar{R}(W, q) \triangleq \max_{P_X} \min_{\substack{P_{Y\hat{Y}|X} \in \mathcal{M}_{\max}(q, P_X) \\ P_{Y|X} = W}} I(P_X, P_{\hat{Y}|X}), \quad (23)$$

Proof: See section V-A for the proof of this theorem. ∎

Corollary 1: If some joint conditional distribution $P_{Y\hat{Y}|X} \in \mathcal{M}_{\max}(q, P_X)$ for all input distributions P_X, then

$$C_q(W) \leq C(P_{\hat{Y}|X}).$$
(24)

The next result introduces a sphere-packing upper bound to the error exponent of mismatched decoding.

Theorem 2: Consider a fixed composition codebook $\mathcal{C}_n$ with length n, rate R and composition $\boldsymbol{p}_X$. The error probability of $\mathcal{C}_n$ with q-decoding over channel W satisfies

$$-\frac{1}{n} \log P_{e,\max}^q(\mathcal{C}_n, W) \leq E_{\text{sp}}^q(\boldsymbol{p}_X, R + \zeta_n) - \delta_n, \quad (25)$$

where

$$E_{\text{sp}}^q(P_X, R) = \min_{\substack{P_{Y'\hat{Y}|X} \in \mathcal{M}_{\max}(q, P_X) \\ I(P_X, P_{\hat{Y}|X}) \leq R}} D(P_{Y'|X} \| P_{Y|X} | P_X) \quad (26)$$

TABLE I
NONZERO ENTRIES OF $P_{Y\hat{Y}|X}$ FOR EXAMPLE 1

(j, k_1, k_2)	$P_{Y\hat{Y}|X}$	(j, k_1, k_2)	$P_{Y\hat{Y}|X}$
$(1,1,1)$	0.3778	$(2,1,1)$	0.1000
$(1,1,2)$	0.5922	$(2,2,2)$	0.0911
$(1,2,2)$	0.0300	$(2,3,3)$	0.6956
		$(2,3,2)$	0.1133

and

$$\zeta_n = (JK - 1)\frac{\log(n+1)}{n} + \frac{\log 2}{n} \tag{27}$$

$$\delta_n = \mathcal{O}\left(\frac{\log n}{n}\right) \tag{28}$$

Proof: See section V-B for the proof of this theorem. ∎

Next we introduce the analogous version of Theorem 2 for a family of type dependent metrics. Firstly we define the analogous version of $\mathcal{M}_{\max}$ for type dependent metrics. With a slight abuse of notation we use $q(\boldsymbol{p}_{XY})$ to denote a type-dependent metric q computed for type $\boldsymbol{p}_{XY}$.

Definition 5: Set $\mathcal{M}_{\max}^{\mathrm{td}}(q, P_{X_1})$ is defined as follows

$$
\mathcal{M}_{\max}^{\mathrm{td}}(q, P_{X_1})
$$
$$
\triangleq \left\{ P_{Y\hat{Y}|X_1} \middle| \min_{\substack{P_{X_2|X_1\hat{Y}}: \\ X_2 - \hat{Y}X_1 - Y \\ P_{\hat{Y}X_1} = P_{\hat{Y}X_2}}} q(P_{X_2Y}) \geq q(P_{X_1Y}) \right\} \tag{29}
$$

Consider type-dependant metrics $q(P_{XY})$ where q is convex in $P_{Y|X}$ when P_X is fixed. This is an important family since important metrics such as maximum mutual information (MMI) metric defined as $q(P_{XY}) = I(P_{XY})$ have this property. For this family of metrics, we have exactly the same statement as that of Theorem 2, but replacing $\mathcal{M}_{\max}(q, P_{X_1})$ by $\mathcal{M}_{\max}^{\mathrm{td}}(q, P_{X_1})$. See Section V-C for the proof of this statement.

A. Example

In this part we show the application of our bound to the counterexample in [10], where the channel and metric are

$$W = \begin{bmatrix} 0.97 & 0.03 & 0 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} \tag{30}$$

$$q = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \log(0.5) & \log(1.36) \end{bmatrix}. \tag{31}$$

For this example $C(W) = 0.7133$ bits/use, the rate achievable by 2-letter superposition coding from [10] is $R_{\mathrm{sc}}^{(2)}(W, q) = 0.1991$ bits/use and our previous converse [11] stated that $C_q(W) \leq \bar{R}_q(W) = 0.6182$ bits/use. By numerically solving the optimization problem in (20) we observe the joint conditional distribution given in Table I is maximal for all input distributions P_X.

Marginalizing the above $P_{Y\hat{Y}|X}$ over Y we obtain

$$P_{\hat{Y}|X} = \begin{bmatrix} 0.3756 & 0.6244 & 0 \\ 0.1 & 0.2044 & 0.6956 \end{bmatrix}. \tag{32}$$

Therefore, by using Corollary 1 we have

$$C_q(W) \leq C(P_{\hat{Y}|X}) \tag{33}$$
$$= 0.4999 \text{ bits/use.} \tag{34}$$

Observe that the above result can be further improved by solving the optimization problem in (23). In terms of computation, unlike the bound proposed in [11], optimizing (23) is not a simple task. This observation stems from the fact that the maximal set $\mathcal{M}_{\max}(q, P_X)$ in (23) depends on P_X, unlike the maximal set $\mathcal{M}_{\max}(q)$ in [11]. In addition, the set $\mathcal{M}_{\max}(q, P_X)$ is itself defined as an optimization problem over distributions $P_{X_2|X\hat{Y}}$ and this makes the problem more difficult than [11]. As illustrated next, the advantages of the bound in (23) are potentially significant even under the conditions of Corollary 1.

B. Comparison With Other Bounds

In this section, we compare the new bound to the mismatch capacity given in (23) with some of the recent bounds that have appeared in the recent literature. Recall that all bounds have the same form

$$C_q(W) \leq \max_{P_X} \min_{\substack{P_{Y\hat{Y}|X} \in \mathcal{M} \\ P_{Y|X} = W}} I(P_X, P_{\hat{Y}|X}), \tag{35}$$

where $\mathcal{M}$ is a set of joint conditional distributions. In the following, we compare the set

$$
\mathcal{M}_{\max}(q, P_{X_1})
$$
$$
\triangleq \left\{ P_{Y\hat{Y}|X_1} : \min_{\substack{P_{X_2|X_1\hat{Y}}: \\ X_2 - \hat{Y}X_1 - Y \\ P_{\hat{Y}X_1} = P_{\hat{Y}X_2}}} \mathbb{E}[q(X_2, Y)] \geq \mathbb{E}[q(X_1, Y)] \right\}
$$
$$\tag{36}$$

defined in Definition 4 with those from previously proposed bounds. In particular, we show that all previously proposed bounds are such that $\mathcal{M} \subset \mathcal{M}_{\max}(q, P_X)$.

To begin with, we compare our recent upper bound derived in [11]. The expression of the set $\mathcal{M}$ is

$$
\mathcal{M}_{\max}(q)
$$
$$
= \left\{ P_{Y\hat{Y}|X_1} : P_{Y\hat{Y}|X_1}(k_1, k_2|j) = 0 \text{ if } j \notin \mathcal{S}_q(k_1, k_2) \right\}, \tag{37}
$$

where

$$
\mathcal{S}_q(k_1, k_2) = \left\{ j \in \mathcal{X} \middle| j = \arg\max_{j' \in \mathcal{X}} q(j', k_2) - q(j', k_1) \right\}. \tag{38}
$$

From [11, Lemma 3], we know that if $P_{Y\hat{Y}|X_1} \in \mathcal{M}_{\max}(q)$ then

$$\forall X_1 : P_{\hat{Y}X_1} = P_{\hat{Y}X_2} \Rightarrow \mathbb{E}[q(X_2, Y)] \geq \mathbb{E}[q(X_1, Y)]. \tag{39}$$

However, $P_{Y\hat{Y}|X_1} \in \mathcal{M}_{\max}(q)$ is not a necessary condition for (39) to hold. Instead, for $P_{Y\hat{Y}|X_1} \in \mathcal{M}_{\max}(q, P_{X_1})$ we have

$$\forall X_2 : X_2 - X_1\hat{Y} - Y, P_{\hat{Y}X_1} = P_{\hat{Y}X_2}$$
$$\Rightarrow \mathbb{E}[q(X_2, Y)] \geq \mathbb{E}[q(X_1, Y)]. \tag{40}$$

We observe that $P_{Y\hat{Y}|X_1} \in \mathcal{M}_{\max}(q, P_{X_1})$ is both a necessary and sufficient condition for (40) being true. Therefore, we see thanks to the additional Markov chain constraint on $\tilde{X}$, $\mathcal{M}_{\max}(q) \subset \mathcal{M}_{\max}(q, P_{X_1})$. Indeed, the more constraints are added to (36), the more joint conditional distributions are able to satisfy the inequality, thus making the set larger.

Two improved upper bounds of the same family were presented in [13]. The first bound is expressed as

$$C_q(W) \leq \min_{\substack{P_{Y\hat{Y}|X} \in \Gamma(\rho, q) \\ P_{Y|X} = W}} C_\rho(P_{\hat{Y}|X}), \qquad (41)$$

where

$$\Gamma(\rho, q) = \{P_{Y\hat{Y}|X} | P_{Y\hat{Y}|X}(y, \hat{y}|x) = 0 \ \text{if} \ x \notin \mathcal{S}_{\rho,q}(y, \hat{y})\}, \qquad (42)$$

$$\mathcal{S}_{\rho,q}(y, \hat{y}) = \left\{ x' \in \mathcal{X} | x' \notin \arg\max_{x \in \mathcal{X}} \rho(x, \hat{y}) - q(x, y) \right\}. \qquad (43)$$

The expression of the second bound, which is also valid for type-dependent metrics, is given by

$$C_q(W) \leq \max_{P_X} \min_{\substack{P_{Y\hat{Y}|X} \in \Gamma^*(q, P_X) \\ P_{Y|X} = W}} I(P_X, P_{\hat{Y}|X}), \qquad (44)$$

where

$$\Gamma^*(q, P_X) = \{P_{Y\hat{Y}|X} \,|\, \forall V_{Y\hat{Y}X\tilde{X}} : V_{Y\hat{Y}X} \ll P_X \times P_{Y\hat{Y}|X},$$
$$V_{\hat{Y}X} = V_{\hat{Y}\tilde{X}} \Rightarrow \mathbb{E}[q(\tilde{X}, Y)] \geq \mathbb{E}[q(X, Y)]\}, \quad (45)$$

and $V_{Y\hat{Y}X} \ll P_X \times P_{Y\hat{Y}|X}$ denotes $P_X \times P_{Y\hat{Y}|X}$ being absolutely continuous with respect to $V_{Y\hat{Y}X}$. The second bound was shown to be stronger than the first one [13], and we therefore focus on the comparison with the second. By expressing the set $\mathcal{M}_{\max}(q, P_X)$ with a similar notation we get

$$\mathcal{M}_{\max}(q, P_X) = \{P_{Y\hat{Y}|X} | \forall P_{Y\hat{Y}X\tilde{X}} : P_{\hat{Y}X} = P_{\hat{Y}\tilde{X}},$$
$$\tilde{X} - X\hat{Y} - Y \Rightarrow \mathbb{E}[q(\tilde{X}, Y)] \geq \mathbb{E}[q(X, Y)]\}. \qquad (46)$$

Observe that the constraint $\mathbb{E}[q(\tilde{X}, Y)] \geq \mathbb{E}[q(X, Y)]$ in the set $\Gamma^*(q, P_X)$ should hold for all $V_{Y\hat{Y}X\tilde{X}}$ with $V_{Y\hat{Y}X} \ll P_X \times P_{Y\hat{Y}|X}, V_{\hat{Y}X} = V_{\hat{Y}\tilde{X}}$. Instead, the constraint $\mathbb{E}[q(\tilde{X}, Y)] \geq \mathbb{E}[q(X, Y)]$ in the set $\mathcal{M}_{\max}(q, P_X)$ must hold for all $P_{Y\hat{Y}X\tilde{X}}$ such that distribution of $Y\hat{Y}X$ is equal to $P_{Y\hat{Y}X}$ but $\tilde{X}$ is further constrained by the Markov chain property $\tilde{X} - X\hat{Y} - Y$. Therefore, similarly to the previous comparison, we find that

$$\Gamma^*(q, P_X) \subset \mathcal{M}_{\max}(q, P_X). \qquad (47)$$

More recently, a further improvement was reported in [14]. The main bound in [14] is expressed as

$$C_q(W) \leq \max_{P_X} \min_{\substack{P_{Y\hat{Y}|X} \in \Theta^*(q, P_X) \\ P_{Y|X} = W}} I(P_X, P_{\hat{Y}|X}) \qquad (48)$$

where Θ^* is defined as

$$\Theta^*(q, P_X) = \{P_{Y\hat{Y}|X} \,|\, \forall P_{Y\hat{Y}X\tilde{X}} : P_{\hat{Y}X} = P_{\hat{Y}\tilde{X}}$$
$$\Rightarrow \mathbb{E}[q(X, Y)] \leq \mathbb{E}[q(\tilde{X}, Y)]\} \quad (49)$$

By comparing the set $\Theta^*(q, P_X)$ to $\mathcal{M}_{\max}(q, P_X)$, we find that the constraint $\mathbb{E}[q(\tilde{X}, Y)] \geq \mathbb{E}[q(X, Y)]$ in the set $\Theta^*(q, P_X)$ must hold for all $P_{Y\hat{Y}X\tilde{X}}$ such that distribution of $Y\hat{Y}X$ is equal to $P_{Y\hat{Y}X}$, but a further constraint on $\tilde{X}$ is missing. Since $\mathcal{M}_{\max}(q, P_X)$ has an additional Markov chain constraint on $\tilde{X}$, we have that

$$\Theta^*(q, P_X) \subset \mathcal{M}_{\max}(q, P_X). \qquad (50)$$

IV. ALTERNATIVE INTERPRETATION OF MAXIMAL SETS

So far, every joint conditional distribution $P_{Y\hat{Y}|X}$ that belongs to the corresponding maximal set

$$\mathcal{M}_{\max}(q, P_{X_1})$$
$$\triangleq \left\{ P_{Y\hat{Y}|X_1} : \min_{\substack{P_{X_2|X_1\hat{Y}}: \\ X_2 - \hat{Y}X_1 - Y \\ P_{\hat{Y}X_1} = P_{\hat{Y}X_2}}} \mathbb{E}[q(X_2, Y)] \geq \mathbb{E}[q(X_1, Y)] \right\}$$
$$(51)$$

from Definition 4 yields a valid upper bound to the mismatch capacity. The joint conditional distribution that minimizes the mutual information of the auxiliary channel yields the best bound. This is the case because maximal joint conditional distributions are such that if an error is made over the auxiliary channel $P_{\hat{Y}|X}$, then a mismatched decoding error is made on the original channel $P_{Y|X}$, constrained to be $P_{Y|X} = W$. This latter statement also holds for a significant fraction of the errors, not necessarily all.

In this section, we discuss a different approach to the construction of the maximal set. Specifically, we first fix the auxiliary channel $V = P_{\hat{Y}|X}$, and then optimize the resulting joint conditional distribution to fulfill the maximality constraint, i.e., if an error is made over the auxiliary channel $V = P_{\hat{Y}|X}$, then a mismatched decoding error is made on the original channel $W = P_{Y|X}$. This naturally gives maximal set of auxiliary channels. Not fixing to the joint conditional distribution between V, W offers the possibility to derive a potentially stronger upper bound. Specifically, we first consider the type $p^\star_{\hat{Y}X_1X_2}$ from Lemma 2 such that for most type conflict errors on channel V, the empirical type $\hat{p}_{\hat{y}x_1x_2}$ is equal to $p^\star_{\hat{Y}X_1X_2}$. Then given this type, we can optimize the joint conditional distribution $P_{Y\hat{Y}|X}$ to fulfill the maximality condition for type $p^\star_{\hat{Y}X_1X_2}$. This is in contrast to only knowing the type conflict error condition over the auxiliary channel, i.e., $\hat{p}_{\hat{y}x_1} = \hat{p}_{\hat{y}x_2}$ for every joint type $\hat{p}_{\hat{y}x_1x_2}$.

The above interpretation of the construction of the maximal set, suggests to define it as the following set of all auxiliary channels V

$$\mathcal{V}_{\max}(q, P_{X_1})$$
$$\triangleq \left\{ V : \max_{\substack{P_{Y\hat{Y}|X_1}: \\ P_{\hat{Y}|X_1} = W \\ P_{\hat{Y}|X_2} = V}} \min_{\substack{P_{X_2|X_1\hat{Y}}: \\ X_2 - \hat{Y}X_1 - Y \\ P_{\hat{Y}X_1} = P_{\hat{Y}X_2}}} \mathbb{E}[q(X_2, Y)] \geq \mathbb{E}[q(X_1, Y)] \right\},$$
$$(52)$$

where the inner minimization corresponds to the choice of type $p^\star_{\hat{Y}X_1X_2}$ and the outer maximization refers to the choice of the joint conditional distribution with marginals V, W. Observe that $\mathbb{E}[q(X_1, Y)]$ is constant for any given P_{X_1}, channel W and metric q. The following lemma implies that this alternative definition gives the same bounds to the mismatch capacity and error exponent as those described in previous sections for additive decoding metrics.

Lemma 1: The optimization order in (52) can be exchanged. More precisely,

$$
\max_{\substack{P_{Y\hat{Y}|X_1}: \\ P_{Y|X_1}=W \\ P_{\hat{Y}|X_1}=V}} \min_{\substack{X_2 \\ X_2-\hat{Y}X_1-Y \\ P_{\hat{Y}X_1}=P_{\hat{Y}X_2}}} \mathbb{E}[q(X_2, Y)]
$$
$$
= \min_{\substack{P_{X_2|X_1\hat{Y}}: \\ X_2-\hat{Y}X_1-Y \\ P_{\hat{Y}X_1}=P_{\hat{Y}X_2}}} \max_{\substack{P_{Y\hat{Y}|X_1}: \\ P_{Y|X_1}=W \\ P_{\hat{Y}|X_1}=V}} \mathbb{E}[q(X_2, Y)]. \quad (53)
$$

Proof: We show the following 3 facts in order to prove the lemma.

- The set of $P_{Y\hat{Y}|X_1}$ where $P_{Y|X_1} = W, P_{\hat{Y}|X_2} = V$ is convex. This is evident, since marginalizing a probability distribution is a linear operation.
- The set of $P_{X_2|X_1\hat{Y}}$ such that $X_2 - \hat{Y}X_1 - Y$ and $P_{\hat{Y}X_1} = P_{\hat{Y}X_2}$ is a convex set. To prove this statement we fix $P_{Y\hat{Y}|X_1}$ and consider the set of joint probability distributions $P_{Y\hat{Y}X_1X_2}$

$$
\{P_{Y\hat{Y}X_1X_2}|P_{\hat{Y}X_1} = P_{\hat{Y}X_2},
$$
$$
P_{Y\hat{Y}X_1X_2} = P_{X_2|\hat{Y}X_1}P_{Y\hat{Y}X_1}\} \quad (54)
$$

Therefore, if two random variables $\tilde{X}_2, \bar{X}_2$ both have joint probability distributions in the above set (54), then, any new random variable $\hat{X}_2$ drawn according to $P_{\hat{X}_2|\hat{Y}X_1} = \alpha P_{\tilde{X}_2|\hat{Y}X_1} + (1-\alpha) P_{\bar{X}_2|\hat{Y}X_1}$ is in set (54).
- Finally, we need to show $\mathbb{E}[q(X_2, Y)]$ is linear in both $P_{Y\hat{Y}|X_1}$ and $P_{X_2|\hat{Y}X_1}$ when fixing either of them. This is proven by expanding $\mathbb{E}[q(X_2, Y)]$

$$
\mathbb{E}[q(X_2, Y)] = \sum_{x_1,x_2,y,\hat{y}} q(x_2, y) P_{X_1}(x_1)
$$
$$
\times P_{Y\hat{Y}|X_1}(y, \hat{y}|x_1) P_{X_2|\hat{Y}X_1}(x_2|\hat{y}, x_1), \quad (55)
$$

which is linear both in $P_{Y\hat{Y}|X_1}$ and $P_{X_2|\hat{Y}X_1}$ when we fix either of the two.

As a result, we have a convex-concave optimization problem, and therefore, by the minimax theorem [19], the order of optimization can be exchanged. ∎

Observe that, as a consequence of the above lemma, every joint conditional distribution $P_{Y\hat{Y}|X_1} \in \mathcal{M}_{\max}(q, P_{X_1})$, then the corresponding $P_{\hat{Y}|X_1} \in \mathcal{V}_{\max}(q, P_{X_1})$. Conversely, for every $P_{\hat{Y}|X_1} \in \mathcal{V}_{\max}(q, P_{X_1})$, there exists a joint conditional distribution $P_{Y\hat{Y}|X_1} \in \mathcal{M}_{\max}(q, P_{X_1})$. Therefore, the optimization problems involving $\mathcal{M}_{\max}(q, P_{X_1})$ or $\mathcal{V}_{\max}(q, P_{X_1})$ in the calculation of the upper bound to the mismatch capacity and error exponent give the same result.

We next illustrate how this argument continues to hold for the optimization of the error exponent for type-dependent metrics, but not necessarily for the upper bound to the mismatch capacity. In particular, for type-dependent metrics, consider the following set of auxiliary channels

$$
\mathcal{V}^{\text{td}}_{\max}(q, P_{X_1}, W)
$$
$$
\triangleq \left\{ V : \max_{\substack{P_{Y\hat{Y}|X_1}: \\ P_{\hat{Y}|X_1}=W \\ P_{\hat{Y}|X_2}=V}} \min_{\substack{P_{X_2|X_1\hat{Y}}: \\ X_2-\hat{Y}X_1-Y \\ P_{\hat{Y}X_1}=P_{\hat{Y}X_2}}} q(P_{X_2Y}) \geq q(P_{X_1Y}) \right\}.
$$
$$
(56)
$$

We have the following result for the error exponent.

Theorem 3: Consider a fixed composition codebook $\mathcal{C}_n$ with length n, rate R and composition p_X. The error probability of $\mathcal{C}_n$ with a type-dependent metric decoder using q over channel W satisfies

$$
-\frac{1}{n} \log P_e^q(\mathcal{C}_n, W) \leq E_{\text{sp}}^q(p_X, R + \zeta_n) - \delta_n, \quad (57)
$$

where we have

$$
E_{\text{sp}}^q(P_X, R)
$$
$$
= \min_{P_{Y'|X}} \min_{\substack{V \in \mathcal{V}^{\text{td}}_{\max}(q, P_X, P_{Y'|X}) \\ I(P_X, V) \leq R}} D(P_{Y'|X} \| P_{Y|X} | P_X) \quad (58)
$$

and ζ_n, δ_n are defined in (27), (28), respectively.

Proof: See Section V-C. ∎

The rate where the the exponent becomes equal to zero is the following

$$
\max_{P_X} \min_{V \in \mathcal{V}^{\text{td}}_{\max}(q, P_X, W)} I(P_X, V). \quad (59)
$$

Unfortunately, the analysis of Section V for this expression fails to work. The main reason is that the error probability may in principle decay subexponentially for rates above (59), and the techniques to prove the mismatch capacity upper bound of Section V are not sufficient.

V. PROOFS OF THE MAIN RESULTS

A. Proof of Theorem 1

We will use the following results proved in Appendix A.

Theorem 4: Let $\mathcal{C}_n = \{x_1, \ldots, x_M\}$ be a constant composition codebook of composition p_X and length n. Assume that $P_{Y\hat{Y}|X} \in \mathcal{M}_{\max}(q, p_X)$ is a maximal joint conditional distribution. Then, there exists a joint conditional distribution $\bar{P}_{Y\hat{Y}|X} \in \mathcal{M}_{\max}(q, p_X)$ satisfying

$$
\bar{P}_{\hat{Y}|X} \times p_X \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}) \quad (60)
$$
$$
|\bar{P}_{\hat{Y}|X} \times p_X - P_{\hat{Y}|X} \times p_X|_\infty \leq \frac{1}{n} \quad (61)
$$
$$
|\bar{P}_{Y|X} \times p_X - P_{Y|X} \times p_X|_\infty \leq \frac{K}{n}, \quad (62)
$$

and a constant $\gamma > 0$ that depends only on $P_{Y|X}$ and q such that

$$
P_{e,\max}^q(\mathcal{C}_n, \bar{P}_{Y\hat{Y}|X}) \geq \gamma P_{\text{tce}}^{\max}(\mathcal{C}_n, \bar{P}_{Y\hat{Y}|X}). \quad (63)
$$

The next result from [11] lower bounds the type-conflict error probability.

Theorem 5: Under the assumptions of Theorem 4, for every type $\boldsymbol{p}_X$, there exist $n_0, \bar{E}(R) > 0$ such that if $n > n_0$ and $\frac{1}{n} \log |\mathcal{C}_n| > I(P_X, P_{\hat{Y}|X})$

$$P_{\text{tce}}^{\max}(\mathcal{C}_n, P_{\hat{Y}|X}) \geq 1 - 2^{-n\bar{E}(R)}. \tag{64}$$

We show that for any $R > \bar{R}(W, q)$ there exist $n_0 > 0$, $0 < \gamma < 1$ and $\delta > 0$ such that for any codebook $\mathcal{C}_n, n > n_0$ with $\frac{1}{n} \log |\mathcal{C}_n| \geq R$, we have

$$P_{e,\max}^q(\mathcal{C}_n, W) \geq \gamma e^{-\delta}(1 - 2^{-n\bar{E}(R)}). \tag{65}$$

We set $R = \bar{R}(W, q) + 2\varepsilon$. We know that for any codebook $\mathcal{C}_n$ of length n and rate R, there exists a constant composition sub-codebook $\mathcal{C}'_n \subset \mathcal{C}_n$ with length n satisfying, rate $R' > R - \frac{J-1}{n} \log(n+1)$, and composition $\boldsymbol{p}_X$ such that

$$P_{e,\max}^q(\mathcal{C}_n, W) \geq P_{e,\max}^q(\mathcal{C}'_n, W). \tag{66}$$

Additionally, from [11, Lemma 5] for any $\varepsilon > 0$ there exists a $\nu > 0$ such that there exists a codebook $\tilde{\mathcal{C}}_{\tilde{n}}$ with the following properties

$$\min_{\tilde{\boldsymbol{p}}_X(j)>0} \tilde{\boldsymbol{p}}_X(j) \geq \nu \tag{67}$$

$$\tilde{n} \geq n\big(1 - (|\mathcal{X}| - 1)\nu\big) \tag{68}$$

$$P_{e,\max}^q(\mathcal{C}'_n, W) \geq P_{e,\max}^q(\tilde{\mathcal{C}}_{\tilde{n}}, W) \tag{69}$$

$$\frac{1}{\tilde{n}} \log(|\tilde{\mathcal{C}}_{\tilde{n}}|) \geq \frac{1}{n} \log(|\mathcal{C}'_n|) - \varepsilon + \mathcal{O}\Big(\frac{\log n}{n}\Big), \tag{70}$$

where $\tilde{\mathcal{C}}_{\tilde{n}}$ is of composition $\tilde{\boldsymbol{p}}_X$. Now, let

$$P_{Y\hat{Y}|X}^\star = \mathop{\arg\min}_{\substack{P_{Y\hat{Y}|X} \in \mathcal{M}_{\max}(q, \tilde{\boldsymbol{p}}_X) \\ P_{Y|X}=W}} I(\tilde{\boldsymbol{p}}_X, P_{\hat{Y}|X}) \tag{71}$$

be the best joint conditional distribution for constant composition codes of composition $\tilde{\boldsymbol{p}}_X$. Then, by applying Theorem 4 to $P_{Y\hat{Y}|X}^\star$ we have that there exists a distribution $\bar{P}_{Y\hat{Y}|X}$ that fulfills the following conditions

$$P_{e,\max}^q(\tilde{\mathcal{C}}_{\tilde{n}}, \bar{P}_{\hat{Y}|X}) \geq \gamma P_{\text{tce}}^{\max}(\tilde{\mathcal{C}}_{\tilde{n}}, \bar{P}_{\hat{Y}|X}). \tag{72}$$

$$|\bar{P}_{\hat{Y}|X} \times \tilde{\boldsymbol{p}}_X - P_{\hat{Y}|X}^\star \times \tilde{\boldsymbol{p}}_X|_\infty \leq \frac{1}{\tilde{n}} \tag{73}$$

$$|\bar{P}_{Y|X} \times \tilde{\boldsymbol{p}}_X - P_{Y|X}^\star \times \tilde{\boldsymbol{p}}_X|_\infty \leq \frac{K}{\tilde{n}} \tag{74}$$

$$\bar{P}_{\hat{Y}|X} \times \tilde{\boldsymbol{p}}_X \in \mathcal{P}_{\tilde{n}}(\mathcal{X} \times \mathcal{Y}) \tag{75}$$

On the other hand, by using Lemma 11 for $P_{XY}^\star = W \times \boldsymbol{p}_X$, $\bar{P}_{XY} = \bar{P}_{Y|X} \times \boldsymbol{p}_X$ we have

$$P_{e,\max}^q(\tilde{\mathcal{C}}_{\tilde{n}}, P_{Y|X}) \geq e^{-\delta} P_{e,\max}^q(\tilde{\mathcal{C}}_{\tilde{n}}, \bar{P}_{Y|X}), \tag{76}$$

where $\delta = \frac{2K}{\min_{P_{XY}(j,k)>0} P_{XY}(j,k)}$. To provide an upper bound on δ note that, we have

$$P_{XY}^\star(j, k) = W(k|j)\tilde{\boldsymbol{p}}_X(j). \tag{77}$$

We observe that since $\tilde{\mathcal{C}}_{\tilde{n}}$ is a ν-reduction of $\mathcal{C}'_n$, from (67) we have that the right hand side of (77) is either equal to zero or bigger than or equal to $\nu W(k|j)$. Therefore,

$$\min_{P_{XY}(j,k)>0} P_{XY}(j, k) \geq \nu \cdot \min_{W(k|j)>0} W(k|j) \tag{78}$$

As a result we have,

$$0 \leq \delta \leq \frac{2K}{\nu \cdot \min_{W(k|j)>0} W(k|j)}, \tag{79}$$

which only depends on the channel, and ε, since ν depends on ε.

Finally, we apply Theorem 5 to codebook $\tilde{\mathcal{C}}_{\tilde{n}}$. Therefore, we have that there exists n_0 such that for $n > n_0$ if

$$\frac{1}{\tilde{n}} \log |\tilde{\mathcal{C}}_{\tilde{n}}| > \max_{P_X} \min_{\substack{P_{Y\hat{Y}|X} \in \mathcal{M}_{\max}(q, P_X) \\ P_{Y|X}=W}} I(P_X, P_{\hat{Y}|X}) \tag{80}$$

$$\geq \min_{\substack{P_{Y\hat{Y}|X} \in \mathcal{M}_{\max}(q, \tilde{\boldsymbol{p}}_X) \\ P_{Y|X}=W}} I(\tilde{\boldsymbol{p}}_X, P_{\hat{Y}|X}), \tag{81}$$

we have that

$$P_{\text{tce}}^{\max}(\tilde{\mathcal{C}}_{\tilde{n}}, P_{\hat{Y}|X}) \geq 1 - 2^{-\tilde{n}\bar{E}(R)}, \tag{82}$$

where in (81) we have chosen $\tilde{\boldsymbol{p}}_X$ as input distribution instead of the maximizing one.

Finally, by combining (66), (72), (76) and (82) we get

$$P_{e,\max}^q(\mathcal{C}_n, P_{Y|X}) \geq P_{e,\max}^q(\mathcal{C}'_n, P_{Y|X}) \tag{83}$$

$$\geq P_{e,\max}^q(\tilde{\mathcal{C}}_{\tilde{n}}, P_{Y|X}) \tag{84}$$

$$\geq e^{-\delta} P_{e,\max}^q(\tilde{\mathcal{C}}_{\tilde{n}}, \bar{P}_{Y|X}) \tag{85}$$

$$\geq e^{-\delta} \gamma P_{\text{tce}}^{\max}(\tilde{\mathcal{C}}_{\tilde{n}}, \bar{P}_{\hat{Y}|X}) \tag{86}$$

$$\geq \gamma e^{-\delta}(1 - 2^{-\tilde{n}\bar{E}(R)}) \tag{87}$$

$$\geq \gamma e^{-\delta}(1 - 2^{-n(1-(|\mathcal{X}|-1)\nu)\bar{E}(R)}). \tag{88}$$

where (88) is bounded away from zero as n tends to infinity.

B. Proof of Theorem 2

The proof is based on three lemmas. Lemma 2, shows a lower bound to the type-conflict error probability of code $\mathcal{C}_n$ over an auxiliary channel. Lemma 3, shows that if the outputs of W and those of the auxiliary channel and connected by an appropriately constructed graph, then a type-conflict error in the auxiliary channel yields a q-decoding error in W. Lemma 4, shows that if the joint conditional distribution that defines W and the auxiliary channel is maximal according to (20), then, the error probability of the q-decoder over channel W is lower-bounded by the type-conflict error probability over the auxiliary channel multiplied by a constant.

Lemma 2: Assume codebook $\mathcal{C}_n$ consists of M codewords of composition $\boldsymbol{p}_X$ used over a DMC $P_{\hat{Y}|X}$. Assume that the conditional type $\boldsymbol{p}_{\hat{Y}|X_1}$ is such that $M|\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})| \geq 2|\mathcal{T}(\boldsymbol{p}_{\hat{Y}})|$. Then, there exists a joint type $\boldsymbol{p}_{\hat{Y}X_1X_2}$ such that $\boldsymbol{p}_{\hat{Y}X_1} = \boldsymbol{p}_{\hat{Y}X_2}$ and

$$\mathbb{P}\big[\exists \boldsymbol{x}_2 \in \mathcal{C}_n \setminus \{\boldsymbol{x}_1\} \text{ s.t. } \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \boldsymbol{p}_{\hat{Y}X_1X_2} | \boldsymbol{x}_1\big]$$
$$\geq \frac{1}{2(n+1)^{J^2K-1}} \mathbb{P}\big[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1}) | \boldsymbol{x}_1\big], \tag{89}$$

where the probabilities are computed w.r.t. n uses of channel $P_{\hat{Y}|X}$.

Proof: Before proving this lemma we explain its main application. This lemma implies that at least a polynomial fraction of elements $\hat{\boldsymbol{y}}$ of $\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{\boldsymbol{y}}|\boldsymbol{x}_1})$ cause a type conflict

error with some codeword $\boldsymbol{x}_2$, when $\boldsymbol{x}_1$ is sent and $\hat{\boldsymbol{y}}$ received as the output of the auxiliary channel, for a fixed joint type $\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \boldsymbol{p}_{\hat{Y}X_1X_2}$.

From [20, Lemma 4] we have there exist a codeword $\boldsymbol{x}_1 \in \mathcal{C}_n$ such that

$$\mathbb{P}\big[\exists \boldsymbol{x}_2 \in \mathcal{C}_n \setminus \{\boldsymbol{x}_1\} \text{ s.t. } \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1} = \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_2} = \boldsymbol{p}_{\hat{Y}X_1}|\boldsymbol{x}_1\big]$$
$$\geq \frac{1}{2}\mathbb{P}\big[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})|\boldsymbol{x}_1\big], \quad (90)$$

where the probabilities are computed w.r.t. n uses of channel $P_{\hat{Y}|X}$. This implies that, assuming $\boldsymbol{x}_1 \in \mathcal{C}_n$ was transmitted, for at least half of the $\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ we can find a codeword $\boldsymbol{x}_2 \neq \boldsymbol{x}_1$ such that $\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}|\boldsymbol{x}_1} = \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}|\boldsymbol{x}_2}$. Observe that there are at most $(n+1)^{J^2K-1}$ joint types $\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2}$. Consider an arbitrary joint type $\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2}$ and define the subset

$$\mathcal{E}_{\boldsymbol{x}_1}(\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1})$$
$$= \big\{\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1}) \,|\, \exists \boldsymbol{x}_2 \in \mathcal{C}_n \setminus \{\boldsymbol{x}_1\},$$
$$\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2}, \tilde{\boldsymbol{p}}_{\hat{Y}X_1} = \tilde{\boldsymbol{p}}_{\hat{Y}X_2} = \boldsymbol{p}_{\hat{Y}X_1}\big\}. \quad (91)$$

In other words, the set $\mathcal{E}_{\boldsymbol{x}_1}(\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1})$ is the set of outputs $\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ such that the joint type of $\boldsymbol{y}, \boldsymbol{x}_1, \boldsymbol{x}_2$ is equal to $\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2}$ and the $\hat{Y}X_1$ and $\hat{Y}X_2$ marginal types are equal to the given $\boldsymbol{p}_{\hat{Y}X_1}$. We now define the joint type $\boldsymbol{p}_{\hat{Y}X_1X_2}^{\star}$ that satisfies the following

$$\boldsymbol{p}_{\hat{Y}X_1X_2}^{\star} = \argmax_{\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2} \in \mathcal{P}_n(\mathcal{Y} \times \mathcal{X}^2)} |\mathcal{E}_{\boldsymbol{x}_1}(\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1})|, \quad (92)$$

i.e., the joint type $\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2}$ that induces the largest subset $\mathcal{E}_{\boldsymbol{x}_1}(\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1})$ for any given $\boldsymbol{p}_{\hat{Y}X_1}$. In other words, out of all joint types $\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2}$, $\boldsymbol{p}_{\hat{Y}X_1X_2}^{\star}$ is the one that contains the maximum number of outputs $\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ that yield a type-conflict error.

Observe that the left hand side of (90) can be bounded as

$$\mathbb{P}\big[\exists \boldsymbol{x}_2 \in \mathcal{C}_n \setminus \{\boldsymbol{x}_1\} \text{ s.t. } \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1} = \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_2} = \boldsymbol{p}_{\hat{Y}X_1}|\boldsymbol{x}_1\big]$$
$$\leq \sum_{\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2} \in \mathcal{P}_n(\mathcal{Y} \times \mathcal{X}^2)} \mathbb{P}[\mathcal{E}_{\boldsymbol{x}_1}(\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1})|\boldsymbol{x}_1] \quad (93)$$
$$\leq (n+1)^{J^2K-1}\mathbb{P}[\mathcal{E}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}X_1X_2}^{\star}, \boldsymbol{p}_{\hat{Y}X_1})|\boldsymbol{x}_1], \quad (94)$$

where the probability is computed with respect to n uses of channel $P_{\hat{Y}|X}$, and thus, from (90), we get

$$\mathbb{P}\big[\exists \boldsymbol{x}_2 \in \mathcal{C}_n \setminus \{\boldsymbol{x}_1\} \text{ s.t. } \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \boldsymbol{p}_{\hat{Y}X_1X_2}^{\star}|\boldsymbol{x}_1\big]$$
$$\geq \frac{1}{2(n+1)^{J^2K-1}}\mathbb{P}\big[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})|\boldsymbol{x}_1\big] \quad (95)$$

which completes the proof. The joint type $\boldsymbol{p}_{\hat{Y}X_1X_2}^{\star}$ is the type $\boldsymbol{p}_{\hat{Y}X_1X_2}$ whose existence is stated in the lemma. ∎

In the rest of this section whenever $\boldsymbol{p}_{\hat{Y}X_1X_2}^{\star}$ is used we refer to the type defined in (92).

Corollary 2: The above statement implies that

$$\frac{|\mathcal{E}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}X_1X_2}^{\star}, \boldsymbol{p}_{\hat{Y}X_1})|}{|\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})|} \geq \frac{1}{2(n+1)^{J^2K-1}}. \quad (96)$$

Proof: We have that

$$\frac{|\mathcal{E}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}X_1X_2}^{\star}, \boldsymbol{p}_{\hat{Y}X_1})|}{|\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})|} = \frac{|\mathcal{E}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}X_1X_2}^{\star}, \boldsymbol{p}_{\hat{Y}X_1})| \cdot \mathbb{P}[\hat{\boldsymbol{y}}|\boldsymbol{x}_1]}{|\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})| \cdot \mathbb{P}[\hat{\boldsymbol{y}}|\boldsymbol{x}_1]}$$
$$\quad (97)$$
$$= \frac{\mathbb{P}\big[\mathcal{E}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}X_1X_2}^{\star}, \boldsymbol{p}_{\hat{Y}X_1})|\boldsymbol{x}_1\big]}{\mathbb{P}\big[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})|\boldsymbol{x}_1\big]}$$
$$\quad (98)$$
$$\geq \frac{1}{2(n+1)^{J^2K-1}}, \quad (99)$$

where $\boldsymbol{p}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1} = \boldsymbol{p}_{\hat{Y}X_1}$, (98) follows from the fact that all elements of $\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ are equiprobable when $\boldsymbol{x}_1$ is sent and (99) is equivalent to (89). ∎

Note that in the next lemmas' proof we will employ Corollary 2 rather than Lemma 2.

Similarly to [11], we construct a bipartite graph $\mathcal{G}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ in the following way (see [11] for details). Vertices of this graph are the elements of $\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})$ and $\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$. Moreover, $\boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})$ and $\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ are connected if $\hat{\boldsymbol{p}}_{\boldsymbol{y}'\hat{\boldsymbol{y}}\boldsymbol{x}_1} = \boldsymbol{p}_{Y'\hat{Y}X_1}$. The graph is regular and we denote the left degree by d_1. Ideally, we need the graph $\mathcal{G}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ to satisfy the following property: if $\boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})$ is connected to $\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ in this graph then for some $\boldsymbol{x}_2 \in \mathcal{C}_n/\{\boldsymbol{x}_1\}$

$$q^n(\boldsymbol{x}_2, \boldsymbol{y}') \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}'). \quad (100)$$

However, in contrast to [11] this is not always the case here. The next lemma proves a lower bound to the fraction of the edges in $\mathcal{G}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ that satisfy the aforementioned desired property.

Lemma 3: Consider a conditional maximal joint type $\boldsymbol{p}_{Y'\hat{Y}|X} \in \hat{\mathcal{M}}_{\max}(q, \boldsymbol{p}_X)$, for some composition $\boldsymbol{p}_X$, and construct a graph $\mathcal{G}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ between the type classes $\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ and $\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})$ as described above. Then, for every $\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ such that $\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \boldsymbol{p}_{\hat{Y}X_1X_2}^{\star}$ there are at least $e^{-n\Lambda_n}d_1$ of its neighbours $\boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})$ such that for some $\boldsymbol{x}_2 \in \mathcal{C}\setminus\{\boldsymbol{x}_1\}$ we have a q-decoding error when $\boldsymbol{x}_1$ is sent, i.e.,

$$q^n(\boldsymbol{x}_2, \boldsymbol{y}') \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}'), \quad (101)$$

where $\Lambda_n = \mathcal{O}\big(\frac{\log n}{n}\big)$.

Proof: Consider $\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$. By construction, all $\boldsymbol{y}'$ that are connected to $\hat{\boldsymbol{y}}$ in graph $\mathcal{G}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ satisfy

$$\hat{\boldsymbol{p}}_{\boldsymbol{y}'\hat{\boldsymbol{y}}|\boldsymbol{x}_1} = \boldsymbol{p}_{Y'\hat{Y}|X}. \quad (102)$$

As a result, by using Lemma 5 with $T = Y', S = X_2, Z = (\hat{Y}, X_1)$ and $f(T, S) = q(X_2, Y')$ we have

$$\mathbb{E}_U[q^n(\boldsymbol{x}_2, \boldsymbol{y}')] = n\mathbb{E}_{\boldsymbol{p}_{Y'|\hat{Y}X_1} \times \boldsymbol{p}_{\hat{Y}X_1X_2}^{\star}}[q(X_2, Y')] \quad (103)$$
$$\geq n\mathbb{E}[q(X_1, Y')] \quad (104)$$
$$= q^n(\boldsymbol{x}_1, \boldsymbol{y}'), \quad (105)$$

where U is a equiprobable random variable over all sequences $\boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})$ that satisfy $\hat{\boldsymbol{p}}_{\boldsymbol{y}'\hat{\boldsymbol{y}}\boldsymbol{x}_1} = \boldsymbol{p}_{Y'\hat{Y}X_1}$ and $\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \boldsymbol{p}_{\hat{Y}X_1X_2}^{\star}$, (104) follows from $\boldsymbol{p}_{Y'\hat{Y}|X_1}$ being maximal and (105) is derived from the additivity of the metric q.

The above equation has an important implication: the expected metric computed on all sequences $\boldsymbol{y}'$ that satisfy $\hat{\boldsymbol{p}}_{\boldsymbol{y}'\hat{\boldsymbol{y}}\boldsymbol{x}_1} = \boldsymbol{p}_{Y'\hat{Y}X_1}$ is larger than or equal to $q^n(\boldsymbol{x}_1, \boldsymbol{y}')$. In the rest of this section, we derive a subset of all such $\boldsymbol{y}'$ that satisfy the mismatch pairwise error condition $q^n(\boldsymbol{x}_2, \boldsymbol{y}') \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}')$.

The main difficulty in deducing finding such a set directly from (105) is that $\boldsymbol{p}_{Y'|\hat{Y}X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}$ might not be a type. As a result, there might not be any $\boldsymbol{y}'$ with the type $\boldsymbol{p}_{Y'|\hat{Y}X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}$ which satisfies the desired inequality (105). Therefore, we attempt to express this distribution $\boldsymbol{p}_{Y'|\hat{Y}X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}$ as a linear combination of types that are in a neighborhood of $\boldsymbol{p}_{Y'|\hat{Y}X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}$, and then prove the desired property for one such type.

By using Lemma 8 with $Z = Y', S = (\hat{Y}, X_1), U = X_2$, respectively, we can express the distribution $\boldsymbol{p}_{Y'|\hat{Y}X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}$ as a convex combination of joint types $\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2}$ with marginals $\boldsymbol{p}_{Y'\hat{Y}X_1}$ and $\boldsymbol{p}^\star_{\hat{Y}X_1X_2}$ for which $|\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2} - \boldsymbol{p}_{Y'|\hat{Y}X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}|_\infty \leq \frac{1}{n}$. More precisely, we have

$$\boldsymbol{p}_{Y'|\hat{Y}X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}$$
$$= \sum_{\substack{\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2} \\ |\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2} - \boldsymbol{p}_{Y'|\hat{Y}X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}|_\infty \leq \frac{1}{n} \\ \tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1} = \boldsymbol{p}_{Y'\hat{Y}X_1}, \tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2} = \boldsymbol{p}^\star_{\hat{Y}X_1X_2}}} \alpha(\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2})\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2},$$

(106)

where

$$\sum_{\substack{\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2} \\ |\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2} - \boldsymbol{p}_{Y'|\hat{Y}X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}|_\infty \leq \frac{1}{n} \\ \tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1} = \boldsymbol{p}_{Y'\hat{Y}X_1}, \tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2} = \boldsymbol{p}^\star_{\hat{Y}X_1X_2}}} \alpha(\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2}) = 1 \quad (107)$$

and $\alpha(\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2}) \geq 0$.

Therefore, from (103) and (106) we have

$$\mathbb{E}_U[q(\boldsymbol{x}_2, \boldsymbol{y}')]$$
$$= n\mathbb{E}_{\boldsymbol{p}_{Y'|\hat{Y}X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}}[q(X_2, Y')] \quad (108)$$
$$= n \sum_{\substack{\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2} \\ |\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2} - \boldsymbol{p}_{Y'|\hat{Y}X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}|_\infty \leq \frac{1}{n} \\ \tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1} = \boldsymbol{p}_{Y'\hat{Y}X_1}, \tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2} = \boldsymbol{p}^\star_{\hat{Y}X_1X_2}}} \alpha(\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2})$$

$$\times \mathbb{E}_{\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2}}[q(\boldsymbol{x}_2, \boldsymbol{y}')].$$

(109)

Moreover, from (105) we know $\mathbb{E}_U[q^n(\boldsymbol{x}_2, \boldsymbol{y}')] \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}')$, and therefore, there exists a joint type $\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2}$ such that

$$n\mathbb{E}_{\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2}}[q(X_2, Y')] = q^n(\boldsymbol{x}_2, \boldsymbol{y}') \quad (110)$$
$$\geq q^n(\boldsymbol{x}_1, \boldsymbol{y}'). \quad (111)$$

As a result, by using $|\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2} - \boldsymbol{p}_{Y'|\hat{Y}X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}|_\infty \leq \frac{1}{n}$ and Lemma 9 we obtain a lower bound on the number of $\boldsymbol{y}'$

with the above type $\hat{\boldsymbol{p}}_{\boldsymbol{y}'\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2}$

$$|\{\boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1\hat{\boldsymbol{y}}}(\boldsymbol{p}_{Y'|\hat{Y}X_1})|\exists \boldsymbol{x}_2 \in \mathcal{C}_n\backslash\{\boldsymbol{x}_1\},$$
$$\hat{\boldsymbol{p}}_{\boldsymbol{y}'\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2}\}| = e^{nH(Y'|\hat{Y},X_1) - n\omega_n}.$$

(112)

where the entropy is computed using probability distribution $\boldsymbol{p}_{Y'|\hat{Y}X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}$.

On the other hand, since d_1 is defined as degree of every $\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ in graph $\mathcal{G}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ we have that $d_1 = e^{nH(Y'|\hat{Y},X_1) - n\kappa_n}$ where $\kappa_n = \mathcal{O}\left(\frac{\log n}{n}\right)$. This follows from the type counting lemma from Gallager's notes [20] and it can be derived by noting that degree d_1 is equal to the number of sequences $\boldsymbol{y}'$ such that $\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{y}'\boldsymbol{x}_1} = \boldsymbol{p}_{\hat{Y}Y'X_1}$ when $\boldsymbol{x}_1, \hat{\boldsymbol{y}}$ are fixed, more precisely $d_1 = |\mathcal{T}_{\boldsymbol{x}_1\hat{\boldsymbol{y}}}(P_{Y'|\hat{Y}X_1})|$.

As a result, by combining (112) and the fact that $d_1 = e^{nH(Y'|\hat{Y},X_1) - n\kappa_n}$ we have

$$|\{\boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1\hat{\boldsymbol{y}}}(\boldsymbol{p}_{Y'\hat{Y}|X_1})|\exists \boldsymbol{x}_2 \in \mathcal{C}_n\backslash\{\boldsymbol{x}_1\},$$
$$\hat{\boldsymbol{p}}_{\boldsymbol{y}'\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \boldsymbol{p}^\star_{Y'\hat{Y}X_1X_2}\}| = e^{-n\omega_n + n\kappa_n} \cdot d_1$$

(113)

Also, from (111), for every $\boldsymbol{y}'$ in the above set (113), we have

$$q^n(\boldsymbol{x}_2, \boldsymbol{y}') \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}'). \quad (114)$$

By setting $\Lambda_n = \omega_n - \kappa_n$ we get the desired result. ∎

Now we construct a new graph $\tilde{\mathcal{G}}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ using Lemma 3. We construct this graph by starting from $\mathcal{G}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ and for each $\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ only keeping the edges that are connected to $\boldsymbol{y}'$ that for some $\boldsymbol{x}_2 \in \mathcal{C}\backslash\{\boldsymbol{x}_1\}$ we have a q-decoding error, more precisely

$$q^n(\boldsymbol{x}_2, \boldsymbol{y}') \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}'). \quad (115)$$

As described in [11], the graph $\mathcal{G}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ is regular: for every $\boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})$ the number of $\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ such that $\hat{\boldsymbol{p}}_{\boldsymbol{y}'\hat{\boldsymbol{y}}\boldsymbol{x}_1} = \boldsymbol{p}_{Y'\hat{Y}X_1}$ is the same; similarly, for every $\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ the number of $\boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})$ such that $\hat{\boldsymbol{p}}_{\boldsymbol{y}'\hat{\boldsymbol{y}}\boldsymbol{x}_1} = \boldsymbol{p}_{Y'\hat{Y}X_1}$ is the same.

The graph $\tilde{\mathcal{G}}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ is no longer regular. The previous lemma shows that the degree of any vertex in $\tilde{\mathcal{G}}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ is at least $e^{-n\Lambda_n}d_1$ and $\Lambda_n = \mathcal{O}\left(\frac{\log n}{n}\right)$. Now we can use this fact to prove the next lemma which relates the q-decoding error probability in channel $P_{Y'|X}$ with the type-conflict error probability in channel $P_{\hat{Y}|X}$.

Lemma 4: Let $\boldsymbol{p}_{Y'\hat{Y}|X_1} \in \hat{\mathcal{M}}_{\max}(q, \boldsymbol{p}_{X_1})$ be a maximal joint conditional type and $\boldsymbol{x}_1 \in \mathcal{T}(\boldsymbol{p}_{X_1})$ be the transmitted codeword. Then

$$P^q_{e,\max}(\mathcal{C}_n, W) \geq e^{-n\sigma_n}\mathbb{P}[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|\boldsymbol{x}_1], \quad (116)$$

where $\sigma_n = \mathcal{O}\left(\frac{\log n}{n}\right)$ and both probabilities are computed with respect to n uses of channel W.

Proof: Consider the bipartite graph $\tilde{\mathcal{G}}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ obtained by connecting elements $\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})$ with $\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ as described above. For any $\mathcal{B} \subset \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ we

define $\Psi(\mathcal{B})$ as

$$
\Psi(\mathcal{B}) = \left\{ \boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1}) \mid \boldsymbol{y}' \text{ is connected} \right.
$$
$$
\left. \text{to some } \hat{\boldsymbol{y}} \in \mathcal{B} \text{ in graph } \tilde{\mathcal{G}}_{\boldsymbol{x}}(\boldsymbol{p}_{Y'\hat{Y}|X_1}) \right\}
$$
(117)

We apply Lemma 10 to graph $\tilde{\mathcal{G}}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ and we obtain that for any $\mathcal{B} \subset \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$

$$
\frac{|\Psi(\mathcal{B})|}{|\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|} \geq e^{-n\Lambda_n} \frac{|\mathcal{B}|}{|\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})|}.
$$
(118)

Now, let $\mathcal{B}$ be the set of all $\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ such that there exist a type-conflict error with another codeword $\boldsymbol{x}_2$ such that $\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \boldsymbol{p}^\star_{\hat{Y}X_1X_2}$ from Lemma 2 Eq. (92), i.e.,

$$
\mathcal{B} = \mathcal{E}_{\boldsymbol{x}_1}(\boldsymbol{p}^\star_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1}).
$$
(119)

Therefore, from Lemma 3 we have for any $\boldsymbol{y}' \in \Psi(\mathcal{B})$ there exists a codeword $\boldsymbol{x}_2 \neq \boldsymbol{x}_1$ such that

$$
q^n(\boldsymbol{x}_2, \boldsymbol{y}') \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}').
$$
(120)

We bound the probability of error as follows

$$
P^q_{e,\max}(\mathcal{C}_n, W)
$$
$$
= \mathbb{P}[\exists \boldsymbol{x}_2 \in \mathcal{C}_n \backslash \{\boldsymbol{x}_1\}, q^n(\boldsymbol{x}_2, \boldsymbol{y}') \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}')|\boldsymbol{x}_1] \quad (121)
$$
$$
\geq \mathbb{P}[\exists \boldsymbol{x}_2 \in \mathcal{C}_n \backslash \{\boldsymbol{x}_1\}, q^n(\boldsymbol{x}_2, \boldsymbol{y}') \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}'),
$$
$$
\boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|\boldsymbol{x}_1] \quad (122)
$$
$$
= \mathbb{P}[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|\boldsymbol{x}_1] \cdot \mathbb{P}[\exists \boldsymbol{x}_2 \in \mathcal{C}_n \backslash \{\boldsymbol{x}_1\},
$$
$$
q^n(\boldsymbol{x}_2, \boldsymbol{y}') \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}')|\boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1}), \boldsymbol{x}_1] \quad (123)
$$
$$
= \mathbb{P}[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|\boldsymbol{x}_1] \cdot \quad (124)
$$
$$
\frac{\left|\left\{\boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|\exists \boldsymbol{x}_2 \in \mathcal{C}_n \backslash \{\boldsymbol{x}_1\}, q^n(\boldsymbol{x}_2, \boldsymbol{y}') \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}')\right\}\right|}{|\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|}
$$
(125)
$$
\geq \mathbb{P}[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|\boldsymbol{x}_1] \cdot \frac{\left|\Psi\left(\mathcal{E}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1})\right)\right|}{|\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|} \quad (126)
$$
$$
\geq \mathbb{P}[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|\boldsymbol{x}_1] \cdot e^{-n\Lambda_n} \cdot \frac{|\mathcal{E}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1})|}{|\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})|}
$$
(127)
$$
\geq \mathbb{P}[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|\boldsymbol{x}_1] \cdot e^{-n\Lambda_n} \frac{1}{2(n+1)^{J^2K-1}}, \quad (128)
$$

where all of probabilities are computed with respect to n uses of channel W, (126) follows from all elements of $\Psi(\mathcal{B})$ satisfying (120), (127) follows from (118) and (128) follows from (96). By setting $\delta_n = \Lambda_n + (J^2K-1)\frac{\log(n+1)}{n}$ we get the desired result. ∎

Using a standard property of conditional types we have that

$$
\mathbb{P}[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|\boldsymbol{x}_1] \geq e^{-n\left(D(P_{Y'|X_1} \| P_{Y|X_1}|\boldsymbol{p}_{X_1}) + \delta_n\right)} \quad (129)
$$

with $\delta_n = \mathcal{O}\left(\frac{\log n}{n}\right)$. From standard arguments of the method of types we obtain (25), where we have set $\boldsymbol{p}_X = \boldsymbol{p}_{X_1}$.

Again using standard arguments (see e.g. [20, Th. 2]) the result of Theorem 2 is applicable to any code, and not only constant composition codes. This is due to the fact that every

codebook $\mathcal{C}_n$ of rate R has a constant composition sub-codebook $\mathcal{C}'_n \subseteq \mathcal{C}_n$ with rate $R' > R - \frac{J-1}{n}\log(n+1)$ with

$$
P^q_{e,\max}(\mathcal{C}_n, W) \geq P^q_{e,\max}(\mathcal{C}'_n, W). \quad (130)
$$

Additionally, a similar analysis would give an identical upper bound to the error exponent using the maximal sets $\hat{\mathcal{M}}_{\max}(q)$ from [11].

As is well known, the exponent from Theorem 2 is decreasing in R and $E^q_{\mathrm{sp}}(\boldsymbol{p}_X, R) = 0$ by choosing $Y' = Y$ in (26) at a rate equal to

$$
\bar{R}_q(W, \boldsymbol{p}_X) \triangleq \min_{\substack{P_{Y\hat{Y}|X} \in \mathcal{M}_{\max}(q, \boldsymbol{p}_X) \\ P_{Y|X} = W}} I(\boldsymbol{p}_X, P_{\hat{Y}|X}) \quad (131)
$$

We have shown that for rates $R < \bar{R}_q(W, \boldsymbol{p}_X)$, the error probability decays at most exponentially. The proof of Theorem 1 in Section V-A shows that for rates $R > \max_{P_X} \bar{R}_q(W, P_X)$ the error probability cannot decay sub-exponentially and is bounded away from zero as n tends to infinity. In the next subsection, we extend our error exponent analysis to type-dependent metrics.

C. Type-Dependent Metrics

In this part we show the previous analysis holds for an important family of type-dependent metrics as well. Namely, we show the analysis holds for type-dependant metric $q(P_{XY})$ where q is convex in $P_{Y|X}$ when P_X is fixed. This is an important family since important metrics such as maximum mutual information (MMI) metric defined as $q(P_{XY}) = I(P_{XY})$ have this property. With a slight abuse of notation we use $q(\boldsymbol{p}_{XY})$ to denote a type-dependent metric q computed for type $\boldsymbol{p}_{XY}$. Recall definition of $\mathcal{M}^{\mathrm{td}}_{\max}(q, P_X)$

$$
\mathcal{M}^{\mathrm{td}}_{\max}(q, P_{X_1})
$$
$$
\triangleq \left\{ P_{Y\hat{Y}|X_1} \middle| \min_{\substack{P_{X_2|X_1\hat{Y}}: \\ X_2 - \hat{Y}X_1 - Y \\ P_{\hat{Y}X_1} = P_{\hat{Y}X_2}}} q(P_{X_2Y}) \geq q(P_{X_1Y}) \right\}. \quad (132)
$$

For this family of metrics, we have exactly the same statement as that of Theorem 2, but replacing $\mathcal{M}_{\max}(q, P_{X_1})$ by $\mathcal{M}^{\mathrm{td}}_{\max}(q, P_{X_1})$.

Here we only discuss the parts of the proof that are different from that of Theorem 2. To begin with, let $\boldsymbol{p}_{Y'\hat{Y}|X} \in \hat{\mathcal{M}}_{\max}(q, \boldsymbol{p}_X)$. Lemma 2 remains valid since the result and its proof do not depend on the decoding metric nor its form. We now adapt Lemma 3 to type-dependent metrics. Assume, we have $\boldsymbol{p}^\star_{\hat{Y}X_1X_2}$ as explained in the proof of the Lemma 2. Moreover, the graph $\mathcal{G}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ is constructed similarly. We now want to construct a graph $\tilde{\mathcal{G}}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ analogously to the proof of Lemma 3. To this end, by using Lemma 8 with $Z, S, U = Y', (\hat{Y}, X_1), X_2$, respectively, we can express $\boldsymbol{p}_{Y'|\hat{Y},X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}$ as a convex combination of types that have marginals equal to $\boldsymbol{p}_{Y'\hat{Y},X_1}, \boldsymbol{p}^\star_{\hat{Y}X_1X_2}$ and satisfying $|\tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1X_2} - \boldsymbol{p}_{Y'|\hat{Y},X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}|_\infty \leq \frac{1}{n}$. More

precisely, we have

$$\boldsymbol{p}_{Y'|\hat{Y},X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}$$
$$= \sum_{\substack{\tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1X_2} \\ |\tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1X_2} - \boldsymbol{p}_{Y'|\hat{Y},X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}|_\infty \le \frac{1}{n} \\ \tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1} = \boldsymbol{p}_{Y'\hat{Y},X_1}, \tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2} = \boldsymbol{p}^\star_{\hat{Y}X_1X_2}}} \alpha(\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2})\tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1X_2},$$

(133)

where

$$\sum_{\substack{\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2} \\ |\tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1X_2} - \boldsymbol{p}_{Y'|\hat{Y},X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}|_\infty \le \frac{1}{n} \\ \tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1} = \boldsymbol{p}_{Y'\hat{Y},X_1}, \tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2} = \boldsymbol{p}^\star_{\hat{Y}X_1X_2}}} \alpha(\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2}) = 1 \quad (134)$$

and $\alpha(\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2}) \ge 0$. We will now show that there exists a type $\boldsymbol{p}^\star_{Y'\hat{Y}X_1X_2}$ such that $|\tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1X_2} - \boldsymbol{p}_{Y'|\hat{Y},X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}|_\infty \le \frac{1}{n}$ and the mismatched decoder makes an error, i.e.,

$$q(\boldsymbol{p}_{YX_2}) \ge q(\boldsymbol{p}_{YX_1}). \quad (135)$$

This can be seen by the fact that $q(P_{XY})$ is convex in $P_{Y|X}$ and using (133). More precisely, if we define $f(P_{Y\hat{Y}X_1X_2}) = q(P_{X_2Y})$, then, $f(P_{Y\hat{Y}X_1X_2})$ is convex in $P_{Y\hat{Y}X_1X_2}$ when P_{X_2} is fixed. As a result, we have

$$q(P_{Y'X_2})$$
$$= f(\boldsymbol{p}_{Y'|\hat{Y},X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}) \quad (136)$$
$$= f\left(\sum_{\substack{\tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1X_2} \\ |\tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1X_2} - \boldsymbol{p}_{Y'|\hat{Y},X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}|_\infty \le \frac{1}{n} \\ \tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1} = \boldsymbol{p}_{Y'\hat{Y},X_1}, \tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2} = \boldsymbol{p}^\star_{\hat{Y}X_1X_2}}} \alpha(\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2})\tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1X_2} \right)$$

(137)

$$\le \sum_{\substack{\tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1X_2} \\ |\tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1X_2} - \boldsymbol{p}_{Y'|\hat{Y},X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}|_\infty \le \frac{1}{n} \\ \tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1} = \boldsymbol{p}_{Y'\hat{Y},X_1}, \tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2} = \boldsymbol{p}^\star_{\hat{Y}X_1X_2}}} \alpha(\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2})f(\tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1X_2})$$

(138)

$$= \sum_{\substack{\tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1X_2} \\ |\tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1X_2} - \boldsymbol{p}_{Y'|\hat{Y},X_1} \times \boldsymbol{p}^\star_{\hat{Y}X_1X_2}|_\infty \le \frac{1}{n} \\ \tilde{\boldsymbol{p}}_{Y'\hat{Y},X_1} = \boldsymbol{p}_{Y'\hat{Y},X_1}, \tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2} = \boldsymbol{p}^\star_{\hat{Y}X_1X_2}}} \alpha(\tilde{\boldsymbol{p}}_{Y'\hat{Y}X_1X_2})q(\tilde{\boldsymbol{p}}_{Y'X_2}),$$

(139)

where (137) follows by substituting (133), and (138) follows by the convexity of f. Throughout this derivation $\boldsymbol{p}_{X_2}$ is fixed due to the codebook being of constant composition. Therefore, we obtain the desired result. We now use this type $\boldsymbol{p}^\star_{Y'\hat{Y}X_1X_2}$ to construct the graph $\tilde{\mathcal{G}}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$. The proof proceeds as that of Theorem 2 with the new graph $\tilde{\mathcal{G}}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$.

Proof of Theorem 5: The proof is almost identical to the proof of Theorem 2; we point out the steps that are different. The main difference with the proof of Theorem 2 is the order of the choice of the type $\boldsymbol{p}^\star_{\hat{Y}X_1X_2}$ and choice of the maximal joint conditional type as mentioned above. Suppose that we

fix $\boldsymbol{p}_{Y'|X}$ and $\boldsymbol{p}_{\hat{Y}|X} \in \hat{\mathcal{V}}^{\text{td}}_{\max}(q, P_{X_1}, \boldsymbol{p}_{Y'|X_1})$ and consider the type classes $\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1}), \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})$. Note that we cannot yet construct a graph between these two type classes, because we have not specified a joint conditional type. Yet, Lemma 2 still holds, since the maximality condition of the underlying joint conditional type is not used. Consider the joint type $\boldsymbol{p}^\star_{\hat{Y}X_1X_2}$ from Lemma 2. We can now use the definition of $\hat{\mathcal{V}}^{\text{td}}_{\max}(q, P_X, \boldsymbol{p}_{Y'|X})$ for which

$$\max_{\substack{P_{Y\hat{Y}|X_1}: \\ P_{YX_1} = \boldsymbol{p}_{Y'X_1} \\ P_{\hat{Y}X_1} = \boldsymbol{p}_{\hat{Y}X_1}}} \quad \min_{\substack{P_{X_2|X_1\hat{Y}}: \\ X_2 - \hat{Y}X_1 - Y \\ P_{\hat{Y}X_1} = P_{\hat{Y}X_2}}} q(P_{X_2Y}) \ge q(P_{X_1Y}). \quad (140)$$

In other words, there exists a joint conditional distribution $P_{\hat{Y}Y'|X_1}$ which is the maximizer in (140) such that the marginals satisfy

$$P_{Y'X_1} = \boldsymbol{p}_{Y'X_1} \quad (141)$$
$$P_{\hat{Y}X_1} = \boldsymbol{p}_{\hat{Y}X_1}, \quad (142)$$

and additionally,

$$q(P_{X_2Y}) \ge q(P_{X_1Y}), \quad (143)$$

where $X_1, X_2, Y', \hat{Y} \sim P_{Y'|\hat{Y}X_1}\boldsymbol{p}^\star_{\hat{Y}X_1X_2}$. The proof proceeds as in Section V-C.

APPENDIX A
PROOF OF THEOREM 4

In this section, we prove Theorem 4. The proof uses several results stated and proved in Appendices B and C.

We first show existence of a joint conditional distribution $\bar{P}_{Y\hat{Y}|X} \in \mathcal{M}_{\max}(q, \boldsymbol{p}_X)$ with properties (60), (61) and (62). To this end let the joint distribution be $P_{Y\hat{Y}X} = P_{Y\hat{Y}|X} \times \boldsymbol{p}_X$. Then, we can use Lemma 7 to express $P_{X\hat{Y}}$ as follows

$$P_{X\hat{Y}} = \sum_{\substack{P'_{\hat{Y}X} \in \mathcal{P}_n(\mathcal{X}\mathcal{Y}): \\ |P'_{\hat{Y}X} - P_{\hat{Y}X}|_\infty \le \frac{1}{n}}} \alpha(P'_{X\hat{Y}})P'_{X\hat{Y}}, \quad (144)$$

where the coefficients $\alpha(\cdot)$ are non-negative and are such that $\sum_{\substack{P'_{\hat{Y}X} \in \mathcal{P}_n(\mathcal{X}\mathcal{Y}): \\ |P'_{\hat{Y}X} - P_{\hat{Y}X}|_\infty \le \frac{1}{n}}} \alpha(P'_{X\hat{Y}}) = 1$. As a result, by multiplying both sides of (144) by $P_{Y|\hat{Y}X_1}$ we have that

$$P_{Y\hat{Y}X} = \sum_{\substack{P'_{\hat{Y}X} \in \mathcal{P}_n(\mathcal{X}\mathcal{Y}): \\ |P'_{\hat{Y}X} - P_{\hat{Y}X}|_\infty \le \frac{1}{n}}} \alpha(P'_{\hat{Y}X})P_{Y|\hat{Y}X}P'_{\hat{Y}X}. \quad (145)$$

Define the joint conditional distributions

$$P'_{Y\hat{Y}X} = P_{Y|\hat{Y}X}P'_{\hat{Y}X} \quad (146)$$

in the sum (145).

The theorem statement assumes that $P_{Y\hat{Y}|X}$ is maximal. We now claim that at least one of the joint conditional distributions $P'_{Y\hat{Y}X} = P_{Y|\hat{Y}X}P'_{\hat{Y}X}$ in the sum (145) is maximal. To see this, assume by contradiction none of the joint conditional distributions $P'_{Y\hat{Y}X} = P_{Y|\hat{Y}X}P'_{\hat{Y}X}$ in the sum (145) are maximal. This this implies that for each distribution $P'_{Y\hat{Y}X_1}$

there exists a distribution $P^\star_{X_2|X_1\hat{Y}}$ such that the optimization problem in the definition of the maximal set gives

$$P^\star_{X_2|X_1\hat{Y}} = \underset{\substack{P_{X_2|X_1\hat{Y}}: \\ X_2 - X_1\hat{Y} - Y \\ P'_{\hat{Y}X_2} = P'_{\hat{Y}X_1}}}{\arg\min} \mathbb{E}[q(X_2, Y)] < \mathbb{E}[q(X_1, Y)], \quad (147)$$

where the expectations in (147) are computed over joint distributions $P'_{Y\hat{Y}X_1X_2} = P^\star_{X_2|X_1\hat{Y}}P'_{Y\hat{Y}X_1}$.

Define

$$P_{Y\hat{Y}X_1X_2} = \sum_{\substack{P'_{\hat{Y}X} \in \mathcal{P}_n(\mathcal{X}\mathcal{Y}): \\ |P'_{\hat{Y}X} - P_{\hat{Y}X}|_\infty \le \frac{1}{n}}} \alpha(P'_{\hat{Y}X})P'_{Y\hat{Y}X_1X_2} \quad (148)$$

with the same coefficients as in (144). We have that

$$P_{\hat{Y}X_1} = P_{\hat{Y}X_2} \quad (149)$$
$$P_{Y\hat{Y}X_1X_2} = P_{Y|\hat{Y}X_1}P_{\hat{Y}X_1X_2}, \quad (150)$$

where (149) follows from the fact that all $P'_{\hat{Y}X_1X_2}$ in the sum of equation (148) are such that $P'_{\hat{Y}X_1} = P'_{\hat{Y}X_2}$ by construction and (150) follows from the definition of $P'_{Y\hat{Y}X}$ in (146).

We write the expectation condition in (147) as

$$\mathbb{E}_{P_{Y\hat{Y}X_1X_2}}[q(X_2, Y) - q(X_1, Y)]$$
$$= \sum_{\substack{P'_{\hat{Y}X} \in \mathcal{P}_n(\mathcal{X}\mathcal{Y}): \\ |P'_{\hat{Y}X} - P_{\hat{Y}X}|_\infty \le \frac{1}{n}}} \alpha(P'_{\hat{Y}X})\mathbb{E}_{P'_{Y\hat{Y}X_1X_2}}[q(X_2, Y) - q(X_1, Y)]$$
$$\quad (151)$$
$$< 0, \quad (152)$$

where (151) follows from (148) and (152) follows from (147). The above inequality contradicts the maximality assumption of $P_{Y\hat{Y}X_1}$. Therefore, there must exist at least one $P'_{Y\hat{Y}X}$ in the sum (145) which is maximal. We call this maximal joint conditional distribution $\bar{P}_{Y\hat{Y}X}$. The distribution $\bar{P}_{Y\hat{Y}X}$ is such that

$$\bar{P}_{X\hat{Y}} \in \mathcal{P}_n(\mathcal{X}\mathcal{Y}) \quad (153)$$
$$|\bar{P}_{\hat{Y}X} - P_{\hat{Y}X}|_\infty \le \frac{1}{n} \quad (154)$$
$$\bar{P}_{Y\hat{Y}X} = P_{Y|\hat{Y}X}\bar{P}_{\hat{Y}X} \quad (155)$$

fulfilling properties (60) and (61). In addition we have that

$$|\bar{P}_{XY}(j, k) - P_{XY}(j, k)|$$
$$= |\sum_{k'} P_{Y|\hat{Y}X}(k|j, k')\bar{P}_{X\hat{Y}}(j, k') \quad (156)$$
$$\quad - P_{Y|\hat{Y}X}(k|j, k')P_{X\hat{Y}}(j, k')| \quad (157)$$
$$\le \sum_{k'} |\bar{P}_{X\hat{Y}}(j, k') - P_{X\hat{Y}}(j, k')| \quad (158)$$
$$\le \frac{K}{n}, \quad (159)$$

where (158) follows from the triangle inequality and $P_{Y|\hat{Y}X}(k|j, k') \le 1$ and (159) follows from (154), proving property (62).

Now we have found a $\bar{P}_{Y\hat{Y}X}$ with properties (60)–(62). We need to show that for this $\bar{P}_{Y\hat{Y}X}$, we have that

$$P^q_{e,\max}(\mathcal{C}_n, \bar{P}_{Y|X}) \ge \gamma P^{\max}_{\text{tce}}(\mathcal{C}_n, \bar{P}_{\hat{Y}|X}). \quad (160)$$

In the following, we prove (160). Without loss of generality assume that x_1 is the codeword with maximum type conflict error on channel $\bar{P}_{\hat{Y}|X}$. For every message $\ell = 2\ldots, M$, define the sets

$$\mathcal{A}_\ell = \{y \mid q^n(x_\ell, y) \ge q^n(x_1, y)\} \quad (161)$$
$$\mathcal{B}_\ell = \{\hat{y} \mid \hat{p}_{\hat{y}|x_\ell} = \hat{p}_{\hat{y}|x_1} = \bar{P}_{\hat{Y}|X}\}. \quad (162)$$

The sets $\mathcal{A}_\ell, \mathcal{B}_\ell$ are the sets of outputs that result in a pairwise mismatched decoding error or type-conflict error, respectively. Using these definitions we write the probability of mismatched decoding error over channel $\bar{P}_{Y|X}$ and the type-conflict error probability over channel $\bar{P}_{\hat{Y}|X}$ as

$$P^q_{e,\max}(\mathcal{C}_n, \bar{P}_{Y|X}) = \mathbb{P}\left[\bigcup_{m'=2}^M \mathcal{A}_{m'}\right] \quad (163)$$

$$P^{\max}_{\text{tce}}(\mathcal{C}_n, \bar{P}_{\hat{Y}|X}) = \mathbb{P}\left[\bigcup_{\ell=2}^M \mathcal{B}_\ell,\right] \quad (164)$$

where both probabilities in (163) and (164) are computed with respect to $\bar{P}^n_{Y|\hat{Y}X} \times P_U$, where P_U denotes the equiprobable distribution over the type class $\mathcal{T}_{x_1}(\bar{P}_{\hat{Y}|X})$. Also define

$$\mathcal{D}_\ell = \mathcal{B}_\ell \setminus \cup_{i=1}^{\ell-1} \mathcal{B}_i \quad (165)$$

with $\mathcal{B}_0 = \emptyset$. Observe that while $\mathcal{B}_\ell$ are not necessarily disjoint, the newly constructed sets $\mathcal{D}_\ell$ are, and thus $\bigcup_{\ell=2}^M \mathcal{B}_\ell = \bigcup_{\ell=2}^M \mathcal{D}_\ell$. Then, we have

$$P^q_{e,\max}(\mathcal{C}_n, \bar{P}_{Y|X}) = \mathbb{P}\left[\bigcup_{m'=2}^M \mathcal{A}_{m'}\right] \quad (166)$$

$$\ge \mathbb{P}\left[\bigcup_{m'=2}^M \mathcal{A}_{m'} \bigcap \bigcup_{\ell=2}^M \mathcal{B}_\ell\right] \quad (167)$$

$$= \mathbb{P}\left[\bigcup_{m'=2}^M \mathcal{A}_{m'} \bigcap \bigcup_{\ell=2}^M \mathcal{D}_\ell\right] \quad (168)$$

$$= \sum_{\ell=2}^M \mathbb{P}\left[\bigcup_{m'=2}^M \mathcal{A}_{m'}|\mathcal{D}_\ell\right]\mathbb{P}[\mathcal{D}_\ell] \quad (169)$$

$$\ge \sum_{\ell=2}^M \mathbb{P}[\mathcal{A}_\ell|\mathcal{D}_\ell]\mathbb{P}[\mathcal{D}_\ell], \quad (170)$$

where (168) follows from $\bigcup_{\ell=2}^M \mathcal{B}_\ell = \bigcup_{\ell=2}^M \mathcal{D}_\ell$, (169) follows from the fact that the sets $\mathcal{D}_\ell$ are disjoint and (170) is follows from lower bounding $\mathbb{P}\left[\bigcup_{m'=2}^M \mathcal{A}_{m'}|\mathcal{D}_\ell\right]$ by $\mathbb{P}[\mathcal{A}_\ell|\mathcal{D}_\ell]$. Although, inequality (170) has removed many error events, it does not weaken our bound since a type conflict error in the auxiliary channel induces a q-decoding error in the original channel for the same codewords.

We now proceed to lower-bounding $\mathbb{P}[\mathcal{A}_\ell|\mathcal{D}_\ell]$. We first rewrite $\mathbb{P}[\mathcal{A}_\ell|\mathcal{D}_\ell]$ as follows

$$\mathbb{P}[\mathcal{A}_\ell|\mathcal{D}_\ell] = \mathbb{P}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_1, \boldsymbol{y})|\mathcal{D}_\ell] \qquad (171)$$

$$= \sum_{\hat{\boldsymbol{y}} \in \mathcal{D}_\ell} \mathbb{P}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_1, \boldsymbol{y})|\hat{\boldsymbol{y}}]\frac{\mathbb{P}[\hat{\boldsymbol{y}}]}{\mathbb{P}[\mathcal{D}_\ell]}, \quad (172)$$

where the probability in (172) is over output sequences $\boldsymbol{y}$. A consequence of Lemma 12 is that, given that $\boldsymbol{x}_1, \boldsymbol{x}_\ell$ are fixed, $\mathbb{P}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_1, \boldsymbol{y})|\hat{\boldsymbol{y}}]$ depends on $\hat{\boldsymbol{y}}$ only through their joint type, i.e.,

$$\mathbb{P}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_1, \boldsymbol{y})|\hat{\boldsymbol{y}}]$$
$$= \mathbb{P}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_1, \boldsymbol{y})|\boldsymbol{p}_{\hat{Y}X_1X_\ell}], \qquad (173)$$

where the joint type $\boldsymbol{p}_{\hat{Y}X_1X_\ell} = \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_\ell}$. We now proceed to lower bound the right hand side of (173) by using Lemma 6. In order to apply the lemma to obtain a lower bound on $\mathbb{P}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) - q^n(\boldsymbol{x}_1, \boldsymbol{y}) \geq 0|\boldsymbol{p}_{\hat{Y}X_1X_\ell}]$ we proceed with the following steps:

1) We derive a single-letter expression of the expectation $\mathbb{E}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) - q^n(\boldsymbol{x}_1, \boldsymbol{y})|\boldsymbol{p}_{\hat{Y}X_1X_\ell}]$. To this end, we use Lemma 13 for $Z_i = (\boldsymbol{x}_1(i), \boldsymbol{x}_\ell(i), \hat{\boldsymbol{y}}(i))$ and $S_i = Y_i$ and $f(Z_i, S_i) = q(\boldsymbol{x}_\ell(i), Y_i) - q(\boldsymbol{x}_1(i), Y_i)$. Then, by using Lemma 13 we obtain

$$\mathbb{E}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) - q^n(\boldsymbol{x}_1, \boldsymbol{y})|\boldsymbol{p}_{\hat{Y}X_1X_\ell}]$$
$$= n\mathbb{E}_{P_{Y|X_1,\hat{Y}} \times \boldsymbol{p}_{\hat{Y}X_1X_\ell}}[q(X_\ell, Y) - q(X_1, Y)] \quad (174)$$

As a result, since $\bar{P}_{Y\hat{Y}|X_1} \in \mathcal{M}_{\max}(q, \boldsymbol{p}_X)$ is maximal, then

$$\mathbb{E}_{P_{Y|X_1,\hat{Y}} \times \boldsymbol{p}_{\hat{Y}X_1X_\ell}}[q(X_\ell, Y) - q(X_1, Y)] \geq 0. \quad (175)$$

2) We use Corollary 4 to write the conditional variance $\text{Var}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) - q^n(\boldsymbol{x}_1, \boldsymbol{y})|\boldsymbol{p}_{\hat{Y}X_1X_\ell}]$ as

$$\text{Var}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) - q^n(\boldsymbol{x}_1, \boldsymbol{y})|\boldsymbol{p}_{\hat{Y}X_1X_\ell}]$$
$$= n\mathbb{E}_{\boldsymbol{p}_{\hat{Y}X_1X_\ell}}\left[\text{Var}_{P_{Y|\hat{Y}X_1}}[q(X_\ell, Y) - q(X_1, Y)]\right].$$
$$(176)$$

3) From Lemma 17, we have that $q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) - q^n(\boldsymbol{x}_1, \boldsymbol{y})$ given $\boldsymbol{p}_{\hat{Y}X_1X_\ell}$ is sub-Gaussian, i.e.,

$$\mathbb{P}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) - q^n(\boldsymbol{x}_1, \boldsymbol{y}) \geq \xi|\boldsymbol{p}_{\hat{Y}X_1X_\ell}] \leq e^{\frac{-\xi^2}{n(b-a)^2}},$$
$$(177)$$

where $a = 2\min_{x,y} q(x, y), b = 2\max_{x,y} q(x, y)$.

4) We apply Corollary 3 to the random variable $\frac{q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) - q^n(\boldsymbol{x}_1, \boldsymbol{y})}{\sqrt{n}}$ and setting $\theta = \frac{1}{|a-b|}$ we obtain

$$\mathbb{P}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) - q^n(\boldsymbol{x}_1, \boldsymbol{y}) \geq 0|\boldsymbol{p}_{\hat{Y}X_1X_\ell}]$$
$$\geq \frac{\mathbb{E}_{\boldsymbol{p}_{\hat{Y}X_1X_\ell}}\left[\text{Var}_{P_{Y|\hat{Y}X_1}}[q(X_\ell, Y) - q(X_1, Y)]\right]}{2\kappa^2(a-b)^2}$$
$$- |a - b|e^{\frac{-\kappa^2}{2}}\left(1 + \sqrt{2} + \frac{\sqrt{2\pi}}{\kappa} + \frac{1}{\kappa^2}\right).$$
$$(178)$$

The expected conditional variance in the right hand side of (178) can potentially be very small. This can happen for types

$\boldsymbol{p}_{\hat{Y}X_1X_\ell}$ that have substantial mass in the entries where the conditional variance $\text{Var}_{P_{Y|\hat{Y}X_1}}[q(X_\ell, Y) - q(X_1, Y)]$ is zero. This implies that conditioning on this type does not allow us to lower bound the probability by a constant, independent of n, as we would like.

To overcome this problem, we shorten the code and received sequences by discarding the entries where the above conditional variance is zero. Then, we use again Corollary 3. More precisely, we define a new type $\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_\ell}$ which places zero mass in the entries where the conditional variance

$$\sigma^2_{j_1,j_2,k} \triangleq \text{Var}_{P_{Y|\hat{Y}X_1}}[q(X_\ell, Y) - q(X_1, Y)$$
$$|X_1 = j_1, X_\ell = j_2, \hat{Y} = k] \quad (179)$$

is zero

$$\boldsymbol{p}^*_{\hat{Y}X_1X_\ell}(k, j_1, j_2) = \begin{cases} 0 & \sigma^2_{j_1,j_2,k} = 0 \\ \frac{\boldsymbol{p}_{\hat{Y}X_1X_\ell}(k,j_1,j_2)}{n^*} & \text{otherewise} \end{cases}, \quad (180)$$

where $\tilde{n} \leq n$ is the length of the sequences after removing the zero-variance entries is defined as follows

$$n^* = n \sum_{j_1,j_2,k} \boldsymbol{p}_{\hat{Y}X_1X_\ell}(k, j_1, j_2)\mathbb{1}\{\sigma^2_{j_1,j_2,k} \neq 0\}. \quad (181)$$

This type consists of only the k, j_1, j_2 for which the conditional variance $\sigma^2_{j_1,j_2,k}$ in (179) is $\sigma^2_{j_1,j_2,k} \neq 0$. We redefine the auxiliary channel output and the two codewords accordingly, by eliminating the entries with zero variance. More precisely, $\boldsymbol{x}^*_1, \boldsymbol{x}^*_\ell, \hat{\boldsymbol{y}}^*$ are defined by eliminating indices $0 \leq i \leq n$ from $\boldsymbol{x}_1, \boldsymbol{x}_2, \hat{\boldsymbol{y}}$ when $\boldsymbol{x}_1(i) = j_1, \boldsymbol{x}_\ell(i) = j_2, \hat{\boldsymbol{y}}(i) = k$ and $\sigma^2_{j_1,j_2,k} = 0$.

We define $\boldsymbol{y}^*$ as the corresponding shortened length-n^* channel output sequence. Then, we notice that

$$\mathbb{P}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_1, \boldsymbol{y})|\boldsymbol{p}_{\hat{Y}X_1X_\ell}]$$
$$= \mathbb{P}[q^{n^*}(\boldsymbol{x}^*_\ell, \boldsymbol{y}^*) \geq q^{n^*}(\boldsymbol{x}^*_1, \boldsymbol{y}^*) + \mu^*|\boldsymbol{p}^*_{\hat{Y}X_1X_\ell}], \quad (182)$$

where we replace the zero-variance entries by

$$\mu^* = \mathbb{E}[q^{n^*}(\boldsymbol{x}^*_\ell, \boldsymbol{y}^*) - q^{n^*}(\boldsymbol{x}^*_1, \boldsymbol{y}^*)|\boldsymbol{p}^*_{\hat{Y}X_1X_\ell}]$$
$$- \mathbb{E}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) - q^n(\boldsymbol{x}_1, \boldsymbol{y})|\boldsymbol{p}_{\hat{Y}X_1X_\ell}], \quad (183)$$

where μ^* is the overall change in metric difference.

Therefore, (182) follows from the fact that eliminating zero-variance entries at the positions as described in (180) corresponds to cases where the metric difference was a constant in that position.

Notice that with the previous procedure we have

$$\text{Var}[q^{n^*}(\boldsymbol{x}^*_\ell, \boldsymbol{y}^*) - q^{n^*}(\boldsymbol{x}^*_1, \boldsymbol{y}^*)|\boldsymbol{p}^*_{\hat{Y}X_1X_\ell}]$$
$$= n^*\mathbb{E}_{\boldsymbol{p}^*_{\hat{Y}X_1X_\ell}}[\text{Var}_{P_{Y|\hat{Y}X_1}}[q(X_\ell, Y)] - q(X_1, Y)]] \quad (184)$$
$$\geq n^*\sigma^2, \quad (185)$$

where

$$\sigma^2 = \min_{\substack{j_1,j_2,k: \\ \sigma^2_{j_1,j_2,k}>0}} \sigma^2_{j_1,j_2,k} \quad (186)$$
$$> 0 \quad (187)$$

where $\sigma^2_{j_1,j_2,k}$ has been defined in (179).

We now proceed to repeat steps 3) and 4) of the above procedure. We first use Lemma 17 and obtain that

$$\mathbb{E}[q^{n^*}(\boldsymbol{x}_\ell^*, \boldsymbol{y}^*) - q^{n^*}(\boldsymbol{x}_1^*, \boldsymbol{y}^*) | \boldsymbol{p}_{\hat{Y}X_1X_\ell}^*] = \mu \geq \mu^*, \quad (188)$$

$$\mathbb{P}[|q^{n^*}(\boldsymbol{x}_\ell^*, \boldsymbol{y}^*) - q^{n^*}(\boldsymbol{x}_1^*, \boldsymbol{y}^*) - \mu| \geq \xi | \boldsymbol{p}_{\hat{Y}X_1X_\ell}^*]$$
$$\leq 2e^{\frac{-\xi^2}{n^*(b-a)^2}}. \quad (189)$$

We now apply Corollary 3 as in step 4), and get that

$$\mathbb{P}[q^{n^*}(\boldsymbol{x}_\ell^*, \boldsymbol{y}^*) - q^{n^*}(\boldsymbol{x}_1^*, \boldsymbol{y}^*) \geq \mu^* | \boldsymbol{p}_{\hat{Y}X_1X_\ell}^*]$$
$$= \mathbb{P}[q^{n^*}(\boldsymbol{x}_\ell^*, \boldsymbol{y}^*) - q^{n^*}(\boldsymbol{x}_1^*, \boldsymbol{y}^*) \geq \mu | \boldsymbol{p}_{\hat{Y}X_1X_\ell}^*] \quad (190)$$
$$\geq \frac{\sigma^2}{2\kappa^2(a-b)^2} - |a-b|e^{\frac{-\kappa^2}{2}}\Big(1 + \sqrt{2} + \frac{\sqrt{2\pi}}{\kappa} + \frac{1}{\kappa^2}\Big), \quad (191)$$

where $a = 2\min_{x,y} q(x,y), b = 2\max_{x,y} q(x,y)$.

By setting κ large enough we get a uniform bound for all $n^* > 0$. Let $\gamma > 0$ denote such a bound, i.e.,

$$\mathbb{P}[q^{n^*}(\boldsymbol{x}_\ell^*, \boldsymbol{y}^*) - q^{n^*}(\boldsymbol{x}_1^*, \boldsymbol{y}^*) \geq \mu^* | \boldsymbol{p}_{\hat{Y}X_1X_\ell}^*] \geq \gamma \quad (192)$$

for all $n^* > 0$. In case $n^* = 0$ the expression in left hand side of (192) equals to 1 and the rest of the proof holds. Therefore, from (171) we get

$$\mathbb{P}[\mathcal{A}_\ell | \mathcal{D}_\ell] = \sum_{\hat{\boldsymbol{y}} \in \mathcal{D}_\ell} \mathbb{P}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}) | \hat{\boldsymbol{y}}] \frac{\mathbb{P}[\hat{\boldsymbol{y}}]}{\mathbb{P}[\mathcal{D}_\ell]} \quad (193)$$

$$\geq \gamma \sum_{\hat{\boldsymbol{y}} \in \mathcal{D}_\ell} \frac{\mathbb{P}[\hat{\boldsymbol{y}}]}{\mathbb{P}[\mathcal{D}_\ell]} \quad (194)$$

$$= \gamma, \quad (195)$$

where (194) follows from (192) and (195) follows from the fact that $\mathbb{P}[\mathcal{D}_\ell] = \sum_{\hat{\boldsymbol{y}} \in \mathcal{D}_\ell} \mathbb{P}[\hat{\boldsymbol{y}}]$.

Therefore, combining the above inequality with (170) we get

$$P_{e,\max}^q(\mathcal{C}_n, \bar{P}_{Y|X}) \geq \sum_{\ell=2}^M \mathbb{P}[\mathcal{A}_\ell | \mathcal{D}_\ell]\mathbb{P}[\mathcal{D}_\ell] \quad (196)$$

$$\geq \sum_{\ell=2}^M \gamma \mathbb{P}[\mathcal{D}_\ell] \quad (197)$$

$$= \gamma \mathbb{P}\Big[\bigcup_{\ell=2}^M \mathcal{D}_\ell\Big] \quad (198)$$

$$= \gamma \mathbb{P}\Big[\bigcup_{\ell=2}^M \mathcal{B}_\ell\Big] \quad (199)$$

$$= \gamma P_{\text{tce}}^{\max}(\mathcal{C}_n, \bar{P}_{\hat{Y}|X}), \quad (200)$$

where (197) is deduced from (195), (198) is resulted from the fact that the sets $\mathcal{D}_\ell$ are disjoint, (199) follows from $\bigcup_{\ell=2}^M \mathcal{B}_\ell = \bigcup_{\ell=2}^M \mathcal{D}_\ell$ and (200) follows from (164). This concludes the proof.

Unfortunately, the techniques introduced in the proof of Theorem 4 do not seem to naturally extend to type-dependent metrics. This implies that the rate at which the error exponent derived in Section V-C becomes zero might not be the best possible bound to the mismatch capacity and might potentially be further improved, since there might be smaller rates where the error probability decays sub-exponentially.

APPENDIX B
AUXILIARY LEMMAS

In this appendix we study expected values of functions under the equiprobable distribution over a type class. Let $\boldsymbol{s}, \boldsymbol{z}$ be sequences of length n from alphabets $\mathcal{S}, \mathcal{Z}$ respectively with joint type $\hat{\boldsymbol{p}}_{\boldsymbol{zs}} = \boldsymbol{p}_{ZS}$. Moreover, let $f : \mathcal{T} \times \mathcal{S} \to \mathbb{R}$ be an arbitrary additive function, i.e.,

$$f(\boldsymbol{t}, \boldsymbol{s}) = \sum_{i=1}^n f(t_i, s_i), \quad (201)$$

where with a slight abuse of notation we have used the same f for sequences and their entries.

Let P_U be the equiprobable distribution over all sequences $\boldsymbol{t}$ such that $\hat{\boldsymbol{p}}_{\boldsymbol{tz}} = \boldsymbol{p}_{TZ}$. In other words, P_U denotes the equiprobable distribution over elements of the conditional type class $\mathcal{T}_{\boldsymbol{z}}(\boldsymbol{p}_{T|Z})$, where $\boldsymbol{p}_{TZ}$ is a given type. The lemma below provides a single-letter expression for $\mathbb{E}_U[f(\boldsymbol{t}, \boldsymbol{s})]$.

Lemma 5: With the above assumptions we have

$$\mathbb{E}_U[f(\boldsymbol{t}, \boldsymbol{s})] = n\mathbb{E}_{\boldsymbol{p}_{T|Z} \times \boldsymbol{p}_{ZS}}[f(T, S)]. \quad (202)$$

Proof: We have

$$\mathbb{E}_U[f(\boldsymbol{t}, \boldsymbol{s})] = \mathbb{E}_U\Big[\sum_{i=1}^n f(t_i, s_i)\Big] \quad (203)$$

$$= \sum_{i=1}^n \mathbb{E}_U[f(t_i, s_i)] \quad (204)$$

$$= \sum_{i=1}^n \mathbb{E}_{\boldsymbol{p}_{T|Z} \times \boldsymbol{p}_{Z|S=s_i}}[f(T, S)|S = s_i] \quad (205)$$

$$= \sum_s n\boldsymbol{p}_S(s)\mathbb{E}_{\boldsymbol{p}_{T|Z} \times \boldsymbol{p}_{Z|S=s}}[f(T, S)|S = s] \quad (206)$$

$$= n\mathbb{E}_{\boldsymbol{p}_{T|Z} \times \boldsymbol{p}_{ZS}}[f(T, S)], \quad (207)$$

where (204) follows from linearity of expectation and (205) is deduced from $\hat{\boldsymbol{p}}_{\boldsymbol{tz}} = \boldsymbol{p}_{TZ}$. ∎

Lemma 6: Let Y be a zero-mean sub-Gaussian random variable with parameter θ i.e. $\mathbb{P}[|Y| \geq \xi] \leq 2e^{\frac{-\xi^2\theta^2}{2}}$ for all $\xi \geq 0$, then for any $a > 0$ we have

$$\mathbb{P}[Y \geq 0] \geq \frac{\text{Var}[Y]}{2a^2} - 2e^{\frac{-a^2\theta^2}{2}}\Big(1 + \sqrt{2} + \frac{\sqrt{2\pi}}{a\theta} + \frac{1}{a^2\theta^2}\Big). \quad (208)$$

Proof: To begin with, we have that

$$\mathbb{1}\{Y \geq 0\} \geq \frac{Y(Y+a)}{2a^2}\mathbb{1}\{-a \leq Y \leq a\}. \quad (209)$$

For simplicity of notation let $\mathcal{I} = [-a, a]$. Therefore, by taking expectations from both sides of (209) we get

$$\mathbb{P}\{Y \geq 0\} \geq \mathbb{E}\Big[\frac{Y(Y+a)\mathbb{1}\{Y \in \mathcal{I}\}}{2a^2}\Big] \quad (210)$$

$$= \mathbb{E}\Big[\frac{Y^2\mathbb{1}\{Y \in \mathcal{I}\}}{2a^2}\Big] + \mathbb{E}\Big[\frac{Y\mathbb{1}\{Y \in \mathcal{I}\}}{2a}\Big] \quad (211)$$

$$= \mathbb{E}\left[\frac{Y^2}{2a^2}\right] + \mathbb{E}\left[\frac{Y}{2a}\right] - \mathbb{E}\left[\frac{Y^2}{2a^2}\mathbb{1}\{Y \notin \mathcal{I}\}\right]$$
$$- \mathbb{E}\left[\frac{Y}{2a}\mathbb{1}\{Y \notin \mathcal{I}\}\right]. \quad (212)$$

Now by substituting $\mathbb{E}[Y] = 0$ and $\mathbb{E}[Y^2] = \text{Var}[Y]$ we have

$$\mathbb{P}\{Y \geq 0\} \geq \frac{\text{Var}[Y]}{2a^2} - \mathbb{E}\left[\frac{Y^2}{2a^2}\mathbb{1}\{Y \notin \mathcal{I}\}\right]$$
$$- \mathbb{E}\left[\frac{Y}{2a}\mathbb{1}\{Y \notin \mathcal{I}\}\right] \quad (213)$$

We now evaluate both expectations in (213). We have that

$$\mathbb{E}\left[\frac{Y^2}{2a^2}\mathbb{1}\{Y \notin \mathcal{I}\}\right] = \int_0^\infty \mathbb{P}\left(\frac{Y^2}{2a^2}\mathbb{1}\{Y \notin \mathcal{I}\} > t\right)dt \quad (214)$$
$$= \int_0^\infty \mathbb{P}\left(|Y| \geq \max\{\sqrt{2t}a, a\}\right)dt \quad (215)$$
$$\leq 2\int_0^\infty e^{-\max\{2ta^2, a^2\}\theta^2}dt \quad (216)$$
$$= 2\int_0^{\frac{1}{2}} e^{-a^2\theta^2}dt + 2\int_{\frac{1}{2}}^\infty e^{-2ta^2\theta^2}dt \quad (217)$$
$$= e^{-a^2\theta^2} + \frac{1}{a^2\theta^2}e^{-a^2\theta^2}, \quad (218)$$

where (214) follows from rewriting the expectation, (216) is followed from the sub-Gaussianity of Y. Similarly, we have

$$\mathbb{E}\left[\frac{Y}{2a}\mathbb{1}\{Y \notin \mathcal{I}\}\right]$$
$$\leq \mathbb{E}\left[\frac{|Y|}{2a}\mathbb{1}\{Y \notin \mathcal{I}\}\right] \quad (219)$$
$$\leq \int_0^\infty \mathbb{P}\left(\frac{|Y|}{2a}\mathbb{1}\{Y \notin \mathcal{I}\} > t\right)dt \quad (220)$$
$$= \int_0^\infty \mathbb{P}\left(|Y| \geq \max\{2ta, a\}\right)dt \quad (221)$$
$$\leq 2\int_0^\infty e^{-\max\{2t^2a^2, a^2\}\theta^2}dt \quad (222)$$
$$= 2\int_0^{\sqrt{\frac{1}{2}}} e^{-a^2\theta^2}dt + 2\int_{\sqrt{\frac{1}{2}}}^\infty e^{-2t^2a^2\theta^2}dt \quad (223)$$
$$= \sqrt{2}e^{-a^2\theta^2} + 2\sqrt{2\pi}\frac{1}{2a\theta}Q(\sqrt{2}a\theta) \quad (224)$$
$$\leq \sqrt{2}e^{-a^2\theta^2} + \frac{\sqrt{2\pi}}{a\theta}e^{-a^2\theta^2}, \quad (225)$$

where Q is the Gaussian Q-function, (220) follows from the sub-Gaussianity of Y and (224) follows from the change of variable $u = 2a\theta t$. Moreover, (225) is resulted from using the Chernoff bound on the Q-function. by substituting (218) and (225) in (213) we get the desired result. ∎

Corollary 3: For a sub-Gaussian random variable Z with parameter θ i.e. $\mathbb{P}[|Z - \mathbb{E}[Z]| \geq \xi] \leq 2e^{\frac{-\xi^2\theta^2}{2}}$ for all $\xi \geq 0$, for any $\kappa > 0$ we get

$$\mathbb{P}[Z \geq \mathbb{E}[Z]] \geq \frac{\theta^2\text{Var}[Z]}{2\kappa^2} - 2e^{\frac{-\kappa^2}{2}}\left(1 + \sqrt{2} + \frac{\sqrt{2\pi}}{\kappa} + \frac{1}{\kappa^2}\right). \quad (226)$$

Proof: By setting $a = \frac{\kappa}{\theta}$ and substituting $Y = Z - \mathbb{E}[Z]$ in the above lemma we get the above inequality. ∎

We will use this form of the inequality throughout the paper.

The next lemma compares the size of the type class $\mathcal{T}(\boldsymbol{p}_{ZSU})$ for joint type $\boldsymbol{p}_{ZSU}$, with number of sequences whose marginal types are equal to $\boldsymbol{p}_{ZS}$ and $\boldsymbol{p}_{SU}$.

Lemma 7: Consider type $\boldsymbol{p}_{ZSU}$ which is the multiplication of two types $\boldsymbol{p}_{ZSU} = \boldsymbol{p}_{Z|S}\boldsymbol{p}_{SU}$. We have the following inequality

$$\frac{\left|\{(\boldsymbol{z}, \boldsymbol{s}, \boldsymbol{u}) \in \mathcal{Z} \times \mathcal{S} \times \mathcal{U}|\hat{\boldsymbol{p}}_{\boldsymbol{zsu}} = \boldsymbol{p}_{ZSU}\}\right|}{\left|\{(\boldsymbol{z}, \boldsymbol{s}, \boldsymbol{u}) \in \mathcal{Z} \times \mathcal{S} \times \mathcal{U}|\hat{\boldsymbol{p}}_{\boldsymbol{zs}} = \boldsymbol{p}_{ZS}, \hat{\boldsymbol{p}}_{\boldsymbol{su}} = \boldsymbol{p}_{SU}\}\right|} \geq 2^{-n\omega_n}, \quad (227)$$

where $\omega_n = \mathcal{O}(\frac{\log n}{n})$.

Proof: From method of types properties (see e.g. [18], [20]) we have

$$\left|\{(\boldsymbol{z}, \boldsymbol{s}, \boldsymbol{u}) \in \mathcal{Z} \times \mathcal{S} \times \mathcal{U}|\hat{\boldsymbol{p}}_{\boldsymbol{zsu}} = \boldsymbol{p}_{ZSU}\}\right|$$
$$= 2^{n(H(Z,S,U) + \zeta_n)} \quad (228)$$
$$= 2^{n(H(Z|S,U) + H(S,U) + \zeta_n)} \quad (229)$$
$$= 2^{n(H(Z|S) + H(S,U) + \zeta_n)}, \quad (230)$$

where $\zeta_n = \mathcal{O}(\frac{\log n}{n})$, the entropies in the above expressions are computed with respect to probability distribution $\boldsymbol{p}_{ZSU}$, (228) follows from counting elements of a type class [20], (229) is derived by using the chain rule of entropy and (230) is deduced by using $\boldsymbol{p}_{ZSU} = \boldsymbol{p}_{Z|S}\boldsymbol{p}_{SU}$ implying that Z is independent of U given S.

On the other hand

$$\left|\{(\boldsymbol{z}, \boldsymbol{s}, \boldsymbol{u}) \in \mathcal{Z} \times \mathcal{S} \times \mathcal{U}|\hat{\boldsymbol{p}}_{\boldsymbol{zs}} = \boldsymbol{p}_{ZS}, \hat{\boldsymbol{p}}_{\boldsymbol{su}} = \boldsymbol{p}_{SU}\}\right| \quad (231)$$
$$= \sum_{\substack{\tilde{\boldsymbol{p}}_{ZSU} \\ \tilde{\boldsymbol{p}}_{\boldsymbol{zs}} = \tilde{\boldsymbol{p}}_{ZS}, \\ \tilde{\boldsymbol{p}}_{\boldsymbol{su}} = \tilde{\boldsymbol{p}}_{SU}}} \left|\{(\boldsymbol{z}, \boldsymbol{s}, \boldsymbol{u}) \in \mathcal{Z} \times \mathcal{S} \times \mathcal{U}|\hat{\boldsymbol{p}}_{\boldsymbol{zsu}} = \tilde{\boldsymbol{p}}_{ZSU}\}\right| \quad (232)$$
$$\leq (n+1)^{|\mathcal{Z}||\mathcal{S}||\mathcal{U}|}$$
$$\max_{\substack{\tilde{\boldsymbol{p}}_{ZSU} \\ \hat{\boldsymbol{p}}_{\boldsymbol{zs}} = \tilde{\boldsymbol{p}}_{ZS}, \\ \hat{\boldsymbol{p}}_{\boldsymbol{su}} = \tilde{\boldsymbol{p}}_{SU}}} \left|\{(\boldsymbol{z}, \boldsymbol{s}, \boldsymbol{u}) \in \mathcal{Z} \times \mathcal{S} \times \mathcal{U}|\hat{\boldsymbol{p}}_{\boldsymbol{zsu}} = \tilde{\boldsymbol{p}}_{ZSU}\}\right| \quad (233)$$
$$= 2^{n(H(Z,S,U) + \theta_n)} \quad (234)$$
$$= 2^{n(H(Z|S,U) + H(S,U) + \theta_n)} \quad (235)$$
$$\leq 2^{n(H(Z|S) + H(S,U) + \theta_n)}, \quad (236)$$

where $\theta_n = \mathcal{O}(\frac{\log n}{n})$, (232) is derived by considering all types with marginals $\boldsymbol{p}_{ZS}$ and $\boldsymbol{p}_{SU}$, (233) follows by upper bounding the number of types with $(n+1)^{|\mathcal{Z}||\mathcal{S}||\mathcal{U}|}$ and number of elements of each type class with the number of elements of the largest one, and (234) follows by counting the elements of the type class that maximises the expression

$$\max_{\substack{\tilde{\boldsymbol{p}}_{ZSU} \\ \hat{\boldsymbol{p}}_{\boldsymbol{zs}} = \tilde{\boldsymbol{p}}_{ZS}, \\ \hat{\boldsymbol{p}}_{\boldsymbol{su}} = \tilde{\boldsymbol{p}}_{SU}}} \left|\{(\boldsymbol{z}, \boldsymbol{s}, \boldsymbol{u}) \in \mathcal{Z} \times \mathcal{S} \times \mathcal{U}|\hat{\boldsymbol{p}}_{\boldsymbol{zsu}} = \tilde{\boldsymbol{p}}_{ZSU}\}\right|. \quad (237)$$

In other words, the entropy in (234) is computed with respect to this maximising type $\tilde{\boldsymbol{p}}_{ZSU}$ in the previous expression. In the proceeding expressions, (234), (235) and (236) the

same type and its corresponding marginals are used. Eq. (235) is derived using chain rule of entropy and (236) follows from $H(Z|S,U) \leq H(Z|S)$. Since the marginals of both types $\boldsymbol{p}_{ZSU}$ and $\tilde{\boldsymbol{p}}_{ZSU}$ are the same, $H(Z|S), H(S,U)$ are the same in (236) and (230). Therefore, we have that

$$\frac{\left|\{(\boldsymbol{z},\boldsymbol{s},\boldsymbol{u}) \in \mathcal{Z} \times \mathcal{S} \times \mathcal{U}|\hat{\boldsymbol{p}}_{\boldsymbol{zsu}} = \boldsymbol{p}_{ZSU}\}\right|}{\left|\{(\boldsymbol{z},\boldsymbol{s},\boldsymbol{u}) \in \mathcal{Z} \times \mathcal{S} \times \mathcal{U}|\hat{\boldsymbol{p}}_{\boldsymbol{zs}} = \boldsymbol{p}_{ZS}, \hat{\boldsymbol{p}}_{\boldsymbol{su}} = \boldsymbol{p}_{SU}\}\right|} \geq 2^{n(\zeta_n - \theta_n)}. \quad (238)$$

By setting $\omega_n = \zeta_n - \theta_n$ we get the desired result. ■

Before stating the next lemma we need to define the convex hull of a set of vectors.

Definition 6: Let $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_\ell \in \mathbb{R}^d$ be vectors in a vector space. Then convex hull of these vectors denoted by $\text{CVH}(\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_\ell\})$ is defined as the following set

$$\begin{aligned}&\text{CVH}(\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_\ell\}) \\ &= \left\{\boldsymbol{v} \in \mathbb{R}^d \mid \boldsymbol{v} = \sum_{i=1}^{\ell} \alpha_i \boldsymbol{v}_i, \alpha_i \in [0,1], \sum_{i=1}^{\ell} \alpha_i = 1\right\}.\end{aligned}$$
$$(239)$$

Lemma 8: Let $\boldsymbol{p}_{ZS}, \boldsymbol{p}_{SU}$ be two joint types. Define the distribution $P_{ZSU}^\star = \boldsymbol{p}_{Z|S}\boldsymbol{p}_{SU}$ and let

$$\begin{aligned}\mathcal{A} = \Big\{&\boldsymbol{p}_{Z_1 S_1 U_1} \in \mathcal{P}_n(\mathcal{Z} \times \mathcal{S} \times \mathcal{U})|\boldsymbol{p}_{Z_1 S_1} = \boldsymbol{p}_{ZS}, \\ &\boldsymbol{p}_{S_1 U_1} = \boldsymbol{p}_{SU}, |\boldsymbol{p}_{Z_1 S_1 U_1} - P_{ZSU}^\star|_\infty \leq \frac{1}{n}\Big\}. \quad (240)\end{aligned}$$

Then $P_{ZSU}^\star \in \text{CVH}(\mathcal{A})$.

The implication of the above lemma is that while $P_{ZSU}^\star$ is not necessarily a type it can be expressed as a convex combination of types that are in the neighborhood of $P_{ZSU}^\star$ and also has marginals equal to the marginals of $P_{ZSU}^\star$, which are types by definition. *Proof:* We prove a stronger result than the one mentioned in the theorem statement. We have used a simpler version of the lemma's statement since it will be what we need in the proofs of the main results of this paper. We will use induction to prove the the following statement. For every probability distribution $P_{ZSU}^\star$ and set $\mathcal{J} \subseteq \mathcal{Z} \times \mathcal{S} \times \mathcal{U}$, we have $P_{ZSU}^\star \in \text{CVH}(\mathcal{A}_\mathcal{J})$, where

$$\begin{aligned}\mathcal{A}_\mathcal{J} = \Big\{&\boldsymbol{p}_{Z_1 S_1 U_1} \in \mathcal{P}_n(\mathcal{Z} \times \mathcal{S} \times \mathcal{U}) \mid \boldsymbol{p}_{Z_1 S_1} = \boldsymbol{p}_{ZS}, \\ &\boldsymbol{p}_{S_1 U_1} = \boldsymbol{p}_{SU}, |\boldsymbol{p}_{Z_1 S_1 U_1} - P_{ZSU}^\star|_\infty \leq \frac{1}{n}, \\ &\forall(i,j,k) \in \mathcal{J}, \boldsymbol{p}_{Z_1 S_1 U_1}(i,j,k) = P_{ZSU}^\star(i,j,k)\Big\}. \\ &\quad (241)\end{aligned}$$

Define the hyperplane $\mathcal{H}$ as the set

$$\mathcal{H} = \{P_{Z_1 S_1 U_1}|P_{Z_1 S_1} = \boldsymbol{p}_{ZS}, P_{S_1 U_1} = \boldsymbol{p}_{SU}\} \quad (242)$$

where in the definition of $\mathcal{H}$, the quantities $P_{Z_1 S_1 U_1}$ are not assumed to satisfy $P_{Z_1 S_1 U_1}(i,j,k) \geq 0$, but they satisfy $\sum_{z,s,u} P_{Z_1 S_1 U_1}(z,s,u) = 1$. This makes the above set a

hyperplane. Therefore, $P_{ZSU}^\star \in \mathcal{H}$, but since it is a distribution, it satisfies that $P_{ZSU}^\star(i,j,k) \geq 0$. Define also the set

$$\begin{aligned}\mathcal{H}_\mathcal{J} = \{&P_{Z_1 S_1 U_1}|\forall(i,j,k) \in \mathcal{J}, P_{Z_1 S_1}(i,j) = \boldsymbol{p}_{ZS}(i,j), \\ &P_{S_1 U_1}(j,k) = \boldsymbol{p}_{SU}(j,k)\} \subseteq \mathcal{H}. \quad (243)\end{aligned}$$

We perform the induction on the dimension, or number of degrees of freedom, of the set $\mathcal{H}_\mathcal{J}$. Recall that $P_{ZSU}^\star$ is not necessarily a type but its marginals $P_{ZS}^\star = \boldsymbol{p}_{ZS}, P_{SU}^\star = \boldsymbol{p}_{SU}$ are types.

Additionally, define the set $\mathcal{B}$

$$\begin{aligned}\mathcal{B} = \Big\{&\boldsymbol{p}_{Z_1 S_1 U_1} \mid \forall(i,j,k) \; \boldsymbol{p}_{Z_1 S_1 U_1}(i,j,k) \geq 0, \\ &|\boldsymbol{p}_{Z_1 S_1 U_1} - P_{ZSU}^\star|_\infty \leq \frac{1}{n} \\ &\forall(i,j,k) \in \mathcal{J}, \boldsymbol{p}_{Z_1 S_1 U_1}(i,j,k) = P_{ZSU}^\star(i,j,k)\Big\}, \\ &\quad (244)\end{aligned}$$

where in the definition of the set $\mathcal{B}$, the quantities $\boldsymbol{p}_{Z_1 S_1 U_1}$ are not assumed to sum to 1, but instead, they are assumed to satisfy $\boldsymbol{p}_{Z_1 S_1 U_1}(i,j,k) \geq 0$ for all (i,j,k).

Then, we deduce that $P_{ZSU}^\star \in \text{CVH}(\mathcal{B}) \cap \mathcal{H}$ because $P_{ZSU}^\star$ belongs to both $\text{CVH}(\mathcal{B})$ and $\mathcal{H}$. Moreover, the intersection of $\text{CVH}(\mathcal{B})$ and $\mathcal{H}$ is a convex set since the intersection of any convex set and a hyperplane is a convex set.

For any $\tilde{\mathcal{J}} \supset \mathcal{J}$ and $P_{ZSU}^\dagger$ where $\forall(i,j,k) \in \mathcal{J}$ we have $P_{ZSU}^\dagger(i,j,k) = P_{ZSU}^\star(i,j,k)$ we define a side of set $\text{CVH}(\mathcal{B})$ as the set $\text{CVH}(\mathcal{B}_{\tilde{\mathcal{J}}})$, where

$$\begin{aligned}\mathcal{B}_{\tilde{\mathcal{J}}} = \Big\{&\boldsymbol{p}_{Z_1 S_1 U_1} \mid \forall(i,j,k), \; \boldsymbol{p}_{Z_1 S_1 U_1}(i,j,k) \geq 0, \\ &|\boldsymbol{p}_{Z_1 S_1 U_1} - P_{ZSU}^\star|_\infty \leq \frac{1}{n} \\ &\forall(i,j,k) \in \tilde{\mathcal{J}}, \boldsymbol{p}_{Z_1 S_1 U_1}(i,j,k) = P_{ZSU}^\dagger(i,j,k)\Big\} \\ &\quad (245)\end{aligned}$$

We claim that the intersection of any side of $\text{CVH}(\mathcal{B})$ with set $\mathcal{H}_\mathcal{J}$, i.e., $\text{CVH}(\mathcal{B}_{\tilde{\mathcal{J}}}) \cap \mathcal{H}_\mathcal{J}$ is the convex hull of all types in $\text{CVH}(\mathcal{B}_{\tilde{\mathcal{J}}}) \cap \mathcal{H}_\mathcal{J}$.

Observe that if we prove this, then the induction step is proved. This is true since if the previous claim is proven, we would deduce that $\text{CVH}(\mathcal{B}) \cap \mathcal{H}_\mathcal{J}$ is itself the convex hull of types in $\text{CVH}(\mathcal{B}) \cap \mathcal{H}_\mathcal{J}$. As a result, any element of $\text{CVH}(\mathcal{B}) \cap \mathcal{H}_\mathcal{J}$ including $P_{ZSU}^\star$ can be expressed as a convex combination of types in $\text{CVH}(\mathcal{B}) \cap \mathcal{H}_\mathcal{J}$. Additionally, observe that from their definition, the set of types that belong to $\text{CVH}(\mathcal{B}) \cap \mathcal{H}_\mathcal{J}$ is equal to $\mathcal{A}_\mathcal{J}$.

To prove our claim we notice that for any $\mathcal{J} \subset \tilde{\mathcal{J}}$ we have that

$$\text{CVH}(\mathcal{B}_{\tilde{\mathcal{J}}}) \cap \mathcal{H}_\mathcal{J} = \text{CVH}(\mathcal{B}_{\tilde{\mathcal{J}}}) \cap \mathcal{H}_{\tilde{\mathcal{J}}}. \quad (246)$$

In addition, observe that

$$\text{CVH}(\mathcal{B}_{\tilde{\mathcal{J}}}) \cap \mathcal{H}_{\tilde{\mathcal{J}}} = \mathcal{A}_{\tilde{\mathcal{J}}}. \quad (247)$$

Therefore, from the induction step we deduce that any distribution $\tilde{P}_{ZSU} \in \text{CVH}(\mathcal{B}_{\tilde{\mathcal{J}}}) \cap \mathcal{H}_{\tilde{\mathcal{J}}}$ can be expressed as a convex combination of types in this side $\text{CVH}(\mathcal{B}_{\tilde{\mathcal{J}}}) \cap \mathcal{H}_{\tilde{\mathcal{J}}}$.

Therefore, the desired induction step is proven. Since we perform induction over the dimension of $\mathcal{H}_{\mathcal{J}}$, when this dimension is 1, this is a trivial statement. Therefore, the proof by induction is complete. ∎

Lemma 9: Let $\boldsymbol{p}_X$ be a type and P_X be a distribution such that $|\boldsymbol{p}_X - P_X|_\infty \leq \frac{1}{n}$. Then we have

$$|\mathcal{T}(\boldsymbol{p}_X)| = 2^{nH(X)-\omega_n} \qquad (248)$$

$$|\omega_n| \leq \frac{|\mathcal{X}|\log n}{n}, \qquad (249)$$

where the entropy is computed with respect to distribution P_X.

Proof: This lemma has been proven in [20]. ∎

The following lemma is eventually used for connecting type conflict errors of the auxiliary channel $P_{\hat{Y}|X}$ and mismatch decoding errors of $P_{Y|X}$. Since for such connections we use a bipartite graph and not all of the edges of the bipartite graph are useful we need the following lemma as a lower bound to the number of erroneous sequences under mismatched decoding.

Lemma 10: Let $\mathcal{G}_{\boldsymbol{x}}(\boldsymbol{p}_{Y\hat{Y}|X})$ be a regular bipartite graph between type classes $\mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})$ and $\mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X})$ with right degree r_1 and left degree r_2. Construct a graph $\tilde{\mathcal{G}}_{\boldsymbol{x}}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ by removing some edges connected to every $\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X})$ in such a way that at least αr_1 of these edges for $0 < \alpha < 1$ remain. Then, for every set $\mathcal{B} \subset \mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X})$ we have

$$\frac{|\Psi(\mathcal{B})|}{|\mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})|} \geq \alpha \frac{|\mathcal{B}|}{|\mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X})|}, \qquad (250)$$

where $\Psi(\mathcal{B})$ is defined as

$$\Psi(\mathcal{B}) = \big\{ \boldsymbol{y} \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y|X_1}) \mid \boldsymbol{y} \text{ is connected}$$
$$\text{to some } \hat{\boldsymbol{y}} \in \mathcal{B} \text{ in graph } \tilde{\mathcal{G}}_{\boldsymbol{x}}(\boldsymbol{p}_{Y\hat{Y}|X_1})\big\}. \quad (251)$$

Proof: Observe that when we eliminate some edges, the degree of every $\boldsymbol{y} \in \mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})$ is at most r_2 and degree of every element in $\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X})$ is at least αr_1. Therefore, if we count the number of edges between $\mathcal{B}$ and $\Psi(\mathcal{B})$ and call it e, we have

$$\alpha|\mathcal{B}|r_1 \leq e \leq r_2|\Psi(\mathcal{B})|. \qquad (252)$$

The above inequality holds since degree of every vertex in $\mathcal{B} \subset \mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X})$ is at least αr_1 therefore, e is at least $|\mathcal{B}|\alpha r_1$. On the other hand, degree of every element in $\Psi(\mathcal{B})$ is at most r_2. As a result, e is at most $|\Psi(\mathcal{B})|r_2$. In addition, observe that

$$|\mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X})|r_1 = |\mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})|r_2. \qquad (253)$$

which follows by counting the number all edges in graph $\mathcal{G}_{\boldsymbol{x}}(\boldsymbol{p}_{Y\hat{Y}|X})$. As a result, by substituting $\frac{r_1}{r_2}$ from (253) in (252) we get

$$\frac{|\Psi(\mathcal{B})|}{|\mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})|} \geq \alpha \frac{|\mathcal{B}|}{|\mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X})|}. \qquad (254)$$

∎

Lemma 11: Let $P_{YX}, \bar{P}_{YX}$ be two joint distributions such that $|\bar{P}_{XY} - P_{XY}|_\infty \leq \frac{K}{n}$. Then, there exists an N_0 such

that for any $n > N_0$ and for any pair of sequences $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ with joint type $\hat{\boldsymbol{p}}_{\boldsymbol{xy}}$ we have

$$e^{-\delta} \leq \prod_{i=1}^{n} \frac{P_{XY}(x_i, y_i)}{\bar{P}_{XY}(x_i, y_i)} \leq e^{\delta}, \qquad (255)$$

where $\delta = \frac{2K}{\min_{P_{XY}(j,k)>0} P_{XY}(j,k)}$.

In addition, for any codebook $\mathcal{C}_n$

$$P_{e,\max}^q(\mathcal{C}_n, P_{Y|X}) \geq e^{-\delta} P_{e,\max}^q(\mathcal{C}_n, \bar{P}_{Y|X}). \qquad (256)$$

Proof: We have the following

$$\prod_{i=1}^{n} \frac{P_{XY}(x_i, y_i)}{\bar{P}_{XY}(x_i, y_i)}$$

$$= \prod_{j,k} \left[\frac{P_{XY}(j,k)}{\bar{P}_{XY}(j,k)} \right]^{n\hat{\boldsymbol{p}}_{\boldsymbol{xy}}(j,k)} \qquad (257)$$

$$= \prod_{j,k} \left[\frac{\bar{P}_{XY}(j,k) + P_{X,Y}(j,k) - \bar{P}_{X,Y}(j,k)}{\bar{P}_{XY}(j,k)} \right]^{n\hat{\boldsymbol{p}}_{\boldsymbol{xy}}(j,k)} \qquad (258)$$

$$\leq \prod_{j,k} \left[\frac{\bar{P}_{XY}(j,k) + |P_{XY}(j,k) - \bar{P}_{XY}(j,k)|}{\bar{P}_{Y|X}(j,k)} \right]^{n\hat{\boldsymbol{p}}_{\boldsymbol{xy}}(j,k)} \qquad (259)$$

$$\leq \prod_{j,k} \left(1 + \frac{\delta}{n} \right)^{n\hat{\boldsymbol{p}}_{\boldsymbol{xy}}(j,k)} \qquad (260)$$

$$= \left(1 + \frac{\delta}{n} \right)^n \qquad (261)$$

$$\leq e^{\delta}, \qquad (262)$$

where (260) follows from $|\bar{P}_{XY} - P_{XY}|_\infty \leq \frac{K}{n}$ and the definition of δ. Moreover, there exists an N_0, such that for $n > N_0$ we have that $\bar{P}_{Y|X}(j,k) \geq \frac{P_{Y|X}(j,k)}{2}$. The other inequality is derived similarly.

As a result, without loss of generality assume $\boldsymbol{x}_1$ is the codeword with maximum probability of error and $\mathcal{B}$ be the set of all output sequences such that cause a q-decoding error when $\boldsymbol{x}_1$ is sent. Therefore,

$$P_{e,\max}^q(\mathcal{C}_n, P_{Y|X}) = \mathbb{P}[\mathcal{B}|\boldsymbol{x}_1] \qquad (263)$$

$$\leq e^{\delta} \bar{\mathbb{P}}[\mathcal{B}|\boldsymbol{x}_1] \qquad (264)$$

$$= e^{\delta} P_{e,\max}^q(\mathcal{C}_n, \bar{P}_{Y|X}), \qquad (265)$$

where the probabilities in (263) and (264) are computed with respect to P_{XY} and $\bar{P}_{XY}$, respectively. This concludes the proof. ∎

APPENDIX C
CONDITIONING ON THE TYPE OF A SEQUENCE

In this section, we study the effect of conditioning on the type of a sequence when computing some statistical properties of functions of random sequences.

Lemma 12: Let $f : \mathcal{Z} \times \mathcal{S} \to \mathbb{R}$ be an arbitrary function and $(Z_i, S_i), i = 1, 2, \ldots, n$ be random variables taking values

on alphabets $\mathcal{Z}, \mathcal{S}$, respectively. Further assume that

$$\mathbb{P}\Big[S = \boldsymbol{s} \Big| Z = \boldsymbol{z}\Big] = \prod_{i=1}^{n} \mathbb{P}[S_i = s_i | Z_i = z_i] \qquad (266)$$

and $P_{Z_i S_i}$ does not depend on index i. Let $\hat{\boldsymbol{p}}_{\boldsymbol{z}}, \hat{\boldsymbol{p}}_{\boldsymbol{s}}$ denote the types of $\boldsymbol{z} = (z_1, \ldots, z_n), \boldsymbol{s} = (s_1, \ldots, s_n)$, respectively. Then, for any function g the expectation

$$\mathbb{E}\Big[g\Big(\sum_{i=1}^{n} f(Z_i, S_i)\Big)\Big| Z = \boldsymbol{z}\Big] \qquad (267)$$

only depends on $\hat{\boldsymbol{p}}_{\boldsymbol{z}}$.

Proof: It is sufficient to show that for any $\boldsymbol{z}_1, \boldsymbol{z}_2$ that $\hat{\boldsymbol{p}}_{\boldsymbol{z}_1} = \hat{\boldsymbol{p}}_{\boldsymbol{z}_2}$ we have

$$\mathbb{E}\Big[g\Big(\sum_{i=1}^{n} f(Z_i, S_i)\Big)\Big| Z = \boldsymbol{z}_1\Big]$$
$$= \mathbb{E}\Big[g\Big(\sum_{i=1}^{n} f(Z_i, S_i)\Big)\Big| Z = \boldsymbol{z}_2\Big]. \qquad (268)$$

We have

$$\mathbb{E}\Big[g\Big(\sum_{i=1}^{n} f(Z_i, S_i)\Big)\Big| Z = \boldsymbol{z}\Big]$$
$$= \int_{\mathcal{S}^n} g\Big(\sum_{i=1}^{n} f(z_i, s_i)\Big)\mathbb{P}\Big[S_1 = s_1, \cdots, S_n = s_n\Big| Z = \boldsymbol{z}\Big] dS \qquad (269)$$
$$= \int_{\mathcal{S}^n} g\Big(\sum_{i=1}^{n} f(z_i, s_i)\Big) \prod_{i=1}^{n} \mathbb{P}[S_i = s_i | Z_i = z_i] dS \qquad (270)$$

where (270) follows from (266). Now notice that with a permutation of indices we can turn $\boldsymbol{z}_1$ into $\boldsymbol{z}_2$. Moreover, the expression in (270) is invariant under permutation of indices because $P_{Z_i S_i}$ does not depend on index i. Therefore, the expression in (270) is equal for $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$. This finishes the proof. ∎

Having the above lemma in mind, we study the problem of conditioning on types in the following results.

Lemma 13: Under the assumptions of Lemma 12, we have that

$$\mathbb{E}\Big[\sum_{i=1}^{n} f(Z_i, S_i)\Big| \hat{\boldsymbol{p}}_{\boldsymbol{z}}\Big] = n\mathbb{E}_{P_{S|Z} \times \hat{\boldsymbol{p}}_{\boldsymbol{z}}}[f(\tilde{Z}, S)] \qquad (271)$$

where $\tilde{Z}$ is a random variable with distribution $\hat{\boldsymbol{p}}_{\boldsymbol{z}}$.

Proof: We have that

$$\mathbb{E}\Big[\sum_{i=1}^{n} f(Z_i, S_i)\Big| \hat{\boldsymbol{p}}_{\boldsymbol{z}}\Big] = \mathbb{E}\Big[\sum_{i=1}^{n} \sum_{z} f(z, S_i)\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z)\Big] \qquad (272)$$
$$= \sum_{i=1}^{n} \sum_{z} \mathbb{E}[f(z, S_i)]\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z) \qquad (273)$$
$$= \sum_{i=1}^{n} \sum_{z} \mathbb{E}_{P_{S|\tilde{Z}=z}}[f(z, S)]\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z) \qquad (274)$$

$$= n\mathbb{E}_{\hat{\boldsymbol{p}}_{\boldsymbol{z}}}\Big[\mathbb{E}_{P_{S|\tilde{Z}}}\Big[f(\tilde{Z}, S)\Big|\tilde{Z}\Big]\Big] \qquad (275)$$
$$= n\mathbb{E}_{P_{S|\tilde{Z}} \times \hat{\boldsymbol{p}}_{\boldsymbol{z}}(z)}[f(\tilde{Z}, S)], \qquad (276)$$

where (272) is derived using the fact that $P_{Z_i S_i}$ does not depend on index i, (273) is derived from the linearity of expectation, (274) follows by replacing random variables S_i by S which does not affect the expectation and (276) follows from the tower rule of conditional expectation. ∎

Lemma 14: Under the assumptions of Lemma 13 we have

$$\mathbb{E}\Big[\Big(\sum_{i=1}^{n} f(Z_i, S_i)\Big)^2\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\Big] = n^2\mathbb{E}_{P_{S|Z} \times \hat{\boldsymbol{p}}_{\boldsymbol{z}}}[f(\tilde{Z}, S)]^2$$
$$+ n\mathbb{E}_{P_{S|Z} \times \hat{\boldsymbol{p}}_{\boldsymbol{z}}}[f(\tilde{Z}, S)^2] - n\mathbb{E}_{\hat{\boldsymbol{p}}_{\boldsymbol{z}}}\Big[\mathbb{E}_{P_{S|Z}}[f(\tilde{Z}, S)|\tilde{Z}]^2\Big], \qquad (277)$$

where $\tilde{Z}$ is a random variable with distribution $\hat{\boldsymbol{p}}_{\boldsymbol{z}}$.

Proof: By expanding the term in the expectation we have

$$\mathbb{E}\Big[\Big(\sum_{i=1}^{n} f(Z_i, S_i)\Big)^2\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\Big]$$
$$= \mathbb{E}\Big[\sum_{i \neq k} f(Z_i, S_i)f(Z_k, S_k)\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\Big] + \mathbb{E}\Big[\sum_{i=1}^{n} f(Z_i, S_i)^2\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\Big] \qquad (278)$$

Then for the first term of the right hand side of (278) we can use Lemma 13

$$\mathbb{E}\Big[\sum_{i=1}^{n} f(Z_i, S_i)^2\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\Big] = n\mathbb{E}_{P_{Z|S} \times \hat{\boldsymbol{p}}_{\boldsymbol{z}}}[f(\tilde{Z}, S)^2], \qquad (279)$$

where $\tilde{Z}$ is a random variable with distribution $\hat{\boldsymbol{p}}_{\boldsymbol{z}}$. Moreover, for the second term of right hand side of (278) we have

$$\mathbb{E}\Big[\sum_{i \neq k} f(Z_i, S_i)f(Z_k, S_k)\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\Big]$$
$$= \mathbb{E}\Big[\sum_{z_1 \neq z_2} \sum_{i \neq k} f(z_1, S_i)f(z_2, S_k)\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_1)\frac{n\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_2)}{n-1}\Big]$$
$$+ \mathbb{E}\Big[\sum_{z} \sum_{i \neq k} f(z, S_i)f(z, S_k)\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z)\frac{n\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z)-1}{n-1}\Big] \qquad (280)$$
$$= \frac{n}{n-1}\mathbb{E}\Big[\sum_{i \neq k} \sum_{z_1, z_2} f(z_1, S_i)\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_1)f(z_2, S_k)\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_2)\Big]$$
$$- \frac{1}{n-1}\mathbb{E}\Big[\sum_{i \neq k} \sum_{z} f(z, S_i)f(z, S_k)\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z)^2\Big] \qquad (281)$$
$$= \frac{n}{n-1}\sum_{i \neq k} \mathbb{E}_{P_{S|Z} \times \hat{\boldsymbol{p}}_{\boldsymbol{z}}}\Big[f(\tilde{Z}, S_i)\Big]\mathbb{E}_{P_{S|Z} \times \hat{\boldsymbol{p}}_{\boldsymbol{z}}}\Big[f(\tilde{Z}, S_k)\Big]$$
$$- \frac{1}{n-1}\sum_{i \neq k} \mathbb{E}_{\hat{\boldsymbol{p}}_{\boldsymbol{z}}}\Big[\mathbb{E}_{P_{S|Z}}[f(\tilde{Z}, S_i)|\tilde{Z}]\mathbb{E}_{P_{S|Z}}[f(\tilde{Z}, S_k)|\tilde{Z}]\Big] \qquad (282)$$
$$= 2\binom{n}{2}\Big(\frac{n}{n-1}\mathbb{E}_{P_{S|Z} \times \hat{\boldsymbol{p}}_{\boldsymbol{z}}}\Big[f(\tilde{Z}, S)\Big]^2$$
$$- \frac{1}{n-1}\mathbb{E}_{\hat{\boldsymbol{p}}_{\boldsymbol{z}}}\Big[\mathbb{E}_{P_{S|Z}}[f(\tilde{Z}, S)|\tilde{Z}]^2\Big]\Big), \qquad (283)$$

where (280) follows from expanding the expectation when the type of the sequence is known and $P_{Z_i S_i}$ does not depend on index i. Observe that (280) is divided into two parts because it addresses z_1, z_2 being equal or not in the expression $f(z_1, S_i) f(z_2, S_k)$. Observe that there are two terms separating all cases depending on whether z_1, z_2 are equal or not. When they are not equal, the number of such possibilities is $n\hat{p}_z(z_1) n\hat{p}_z(z_2)$ while the number of choices is $n(n-1)$, yielding a probability equal to $\frac{n}{n-1}\hat{p}_z(z_1)\hat{p}_z(z_2)$. Similarly, when $z_1 = z_2 = z$, the number of such possibilities is $n\hat{p}_z(z_1)(n\hat{p}_z(z_2) - 1)$, while the number of choices remains $n(n-1)$, yielding a probability equal to $\frac{1}{n-1}\hat{p}_z(z_1)(n\hat{p}_z(z_2) - 1)$. Eq. (281) follows by rearranging the terms. Additionally, (282) follows by taking the expectation inside using Lemma 13. Combining (279) and (283) with (278) we get the result. ∎

Corollary 4: Under the assumptions of Lemma 13 we have

$$\mathrm{Var}\left[\sum_{i=1}^n f(Z_i, S_i)\Big|\hat{p}_z\right] = n\mathbb{E}_{\hat{p}_z}[\mathrm{Var}_{P_{S|Z}}[f(\tilde{Z}, S)|\tilde{Z}]],$$
(284)

where $\tilde{Z}$ is a random variable with distribution $\hat{p}_z$.

Proof:

$$\mathrm{Var}\left[\sum_{i=1}^n f(Z_i, S_i)\Big|\hat{p}_z\right]$$

$$= \mathbb{E}\left[\left(\sum_{i=1}^n f(Z_i, S_i)\right)^2\Big|\hat{p}_z - \mathbb{E}\left[\sum_{i=1}^n f(Z_i, S_i)\Big|\hat{p}_z\right]^2\right]$$
(285)

$$= n^2 \mathbb{E}_{P_{S|Z}\times\hat{p}_z}[f(\tilde{Z}, S)]^2 + n\mathbb{E}_{P_{S|Z}\times\hat{p}_z}[f(\tilde{Z}, S)^2]$$
$$- n\mathbb{E}_{\hat{p}_z}\left[\mathbb{E}_{P_{S|Z}}[f(\tilde{Z}, S)|\tilde{Z}]^2\right] - n^2\mathbb{E}_{P_{S|Z}\times\hat{p}_z}[f(\tilde{Z}, S)]^2$$
(286)

$$= n\mathbb{E}_{P_{S|Z}\times\hat{p}_z}[f(\tilde{Z}, S)^2] - n\mathbb{E}_{\hat{p}_z}\left[\mathbb{E}_{P_{S|Z}}[f(\tilde{Z}, S)|\tilde{Z}]^2\right]$$
(287)

$$= n\mathbb{E}_{\hat{p}_z}[\mathrm{Var}_{P_{S|Z}}[f(\tilde{Z}, S)|\tilde{Z}]]$$
(288)

where (285) follows from the definition of variance, and (286) follows by directly using Lemmas 13 and 14. ∎

Lemma 15: Let $(Z_i, S_i), i = 1, 2, \ldots, n$ be i.i.d random variables, $z = (Z_1, Z_2, \ldots, Z_n)$ and $\mathcal{A} \subset \mathcal{P}_z^n$ then

$$\mathbb{E}\left[\sum_{i=1}^n f(Z_i, S_i)\Big|\mathcal{A}\right] \geq n\min_{\hat{p}_z\in\mathcal{A}}\mathbb{E}_{P_{S|Z}\times\hat{p}_z}\left[f(\tilde{Z}, S)\right]$$
(289)

Proof: We have

$$\mathbb{E}\left[\sum_{i=1}^n f(Z_i, S_i)\Big|\mathcal{A}\right]$$

$$= \frac{1}{\mathbb{P}(\mathcal{A})}\mathbb{E}\left[\left(\sum_{i=1}^n f(Z_i, S_i)\right)\mathbb{1}\{\hat{p}_z \in \mathcal{A}\}\right]$$
(290)

$$\geq \min_{\hat{p}_z\in\mathcal{A}}\mathbb{E}\left[\left(\sum_{i=1}^n f(Z_i, S_i)\right)\Big|\hat{p}_z\right]$$
(291)

$$= n\min_{\hat{p}_z\in\mathcal{A}}\mathbb{E}_{P_{S|Z}\times\hat{p}_z}\left[f(\tilde{Z}, S)\right],$$
(292)

where (290) follows from the definition of conditional expectation and (292) follows from using Lemma 13. ∎

The next result introduces a lower bound on the variance.

Lemma 16: With the above assumptions we have

$$\mathrm{Var}\left[\sum_{i=1}^n f(Z_i, S_i)\Big|\mathcal{A}\right] \geq \min_{\tilde{p}_Z\in\mathcal{A}}\mathrm{Var}\left[\sum_{i=1}^n f(Z_i, S_i)\Big|\tilde{p}_Z\right]$$
(293)

$$= n\min_{\tilde{P}_Z\in\mathcal{A}}\mathbb{E}_{\tilde{P}_Z}[\mathrm{Var}_{P_{S|Z}}[f(\tilde{Z}, S)|\tilde{Z}]]$$
(294)

Proof: To show this we use the law of total variance which is stated below as a reminder. For any two random variable X, Y we have

$$\mathrm{Var}[X] = \mathbb{E}_Y[\mathrm{Var}[X|Y]] + \mathrm{Var}_Y[\mathbb{E}[X|Y]]$$
(295)

As a result, by setting $Y = \hat{p}_Z$ meaning the random variable that denoted the type of the random variable Z. Then, we have

$$\mathrm{Var}\left[\sum_{i=1}^n f(Z_i, S_i)\Big|\mathcal{A}\right]$$

$$= \mathbb{E}_Y\left[\mathrm{Var}\left[\sum_{i=1}^n f(Z_i, S_i)\Big|\mathcal{A}, Y\right]\right]$$
$$+ \mathrm{Var}_Y\left[\mathbb{E}\left[\sum_{i=1}^n f(Z_i, S_i)\Big|\mathcal{A}, Y\right]\right]$$
(296)

$$\geq \mathbb{E}_Y\left[\mathrm{Var}\left[\sum_{i=1}^n f(Z_i, S_i)\Big|\mathcal{A}, Y\right]\right]$$
(297)

$$\geq \min_{\tilde{p}_Z\in\mathcal{A}}\mathrm{Var}\left[\sum_{i=1}^n f(Z_i, S_i)\Big|\tilde{p}_Z\right]$$
(298)

$$= n\min_{\tilde{P}_Z\in\mathcal{A}}\mathbb{E}_{\tilde{P}_Z}[\mathrm{Var}_{P_{S|Z}}[f(\tilde{Z}, S)|\tilde{Z}]].$$
(299)

∎

In the next lemma we prove a concentration inequality for the same setting. In particular, we show that by conditioning on the type of a sequence we get a sub-Gaussian random variable. We prove the tail bound for the sum $\sum_{i=1}^n f(Z_i, S_i)$ when conditioned on the type $\hat{p}_z$.

The following lemma is an application of the Hoeffding's lemma but we prove it for completeness.

Lemma 17: Let f, Z_i, S_i be defined the same as Lemma 13. Further assume for all z we have $a \leq f(z, S) - \mathbb{E}[f(z, S)] \leq b$. Then,

$$\mathbb{P}\left[\Big|\sum_{i=1}^n f(Z_i, S_i) - \mu\Big| \geq \xi \Big|\hat{p}_z\right] \leq 2e^{\frac{-\xi^2}{n(b-a)^2}},$$
(300)

where $\mu = \mathbb{E}\left[\sum_{i=1}^n f(Z_i, S_i)|\hat{p}_z\right]$. *Proof:* Assume $\mu_z = \mathbb{E}[f(z, S)]$, then for any $\lambda > 0$

$$\mathbb{P}\left[\sum_{i=1}^n f(Z_i, S_i) - \mu \geq \xi\Big|\hat{p}_z\right]$$

$$= \mathbb{P}\left[e^{\lambda\left(\sum_{i=1}^n f(Z_i, S_i) - \mu\right)} \geq e^{\lambda\xi}\Big|\hat{p}_z\right]$$
(301)

$$\leq \frac{\mathbb{E}\left[e^{\lambda\left(\sum_{i=1}^n f(Z_i, S_i) - \mu\right)}\Big|\hat{p}_z\right]}{e^{\lambda\xi}}$$
(302)

$$= \frac{\prod_z \mathbb{E}\left[e^{\lambda(f(z,S)-\mu_z)}\right]^{n\hat{p}_z(z)}}{e^{\lambda\xi}} \tag{303}$$

$$\leq \frac{\prod_z e^{\frac{1}{8}\lambda^2(a-b)^2 n\hat{p}_z(z)}}{e^{\lambda\xi}} \tag{304}$$

$$= \frac{e^{\frac{1}{8}\lambda^2(a-b)^2}}{e^{\lambda\xi}} \tag{305}$$

$$\leq e^{\frac{-\xi^2}{n(b-a)^2}}, \tag{306}$$

where (302) is derived for Markov's inequality. Additionally, (303) follows from noticing that frequency of $f(z, S_i)$ appearing in the expression $\sum_{i=1}^n f(Z_i, S_i)$ for some i is exactly $n\hat{p}_z(z)$ and because S_is are i.i.d. the index of the appearance does not impact the moment generating function. Moreover, (305) is followed from setting $\lambda = \frac{4\xi}{n(b-a)^2}$. Therefore, (300) follows from

$$\mathbb{P}\left[\left|\sum_{i=1}^n f(Z_i, S_i) - \mu\right| \geq \xi \Big| \hat{p}_z\right]$$

$$= \mathbb{P}\left[\sum_{i=1}^n f(Z_i, S_i) - \mu \geq \xi \Big| \hat{p}_z\right]$$

$$+ \mathbb{P}\left[\sum_{i=1}^n f(Z_i, S_i) - \mu \leq -\xi \Big| \hat{p}_z\right] \tag{307}$$

and writing the same steps (301)–(306) for $\mathbb{P}\left[\sum_{i=1}^n f(Z_i, S_i) - \mu \leq -\xi \Big| \hat{p}_z\right]$. ∎

REFERENCES

[1] J. Scarlett, A. Guillén i Fàbregas, A. Somekh-Baruch, and A. Martinez, *Information-Theoretic Foundations of Mismatched Decoding* (Foundations and Trends in Communications and Information Theory). Norwell, MA, USA: Now Publishers, 2020.

[2] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 35–43, Jan. 1995.

[3] J. Y. N. Hui, "Fundamental issues of multiple accessing," Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 1983.

[4] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 5–12, Jan. 1981.

[5] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.

[6] A. Lapidoth, "Mismatched decoding and the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1439–1452, Sep. 1996.

[7] A. Somekh-Baruch, "On achievable rates and error exponents for channels with mismatched decoding," *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 727–740, Feb. 2015.

[8] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Multiuser random coding techniques for mismatched decoding," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3950–3970, Jul. 2016.

[9] V. B. Balakirsky, "A converse coding theorem for mismatched decoding at the output of binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1889–1902, Nov. 1995.

[10] J. Scarlett, A. Somekh-Baruch, A. Martinez, and A. Guillén i Fàbregas, "A counter-example to the mismatched decoding converse for binary-input discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 61, no. 10, pp. 5387–5395, Oct. 2015.

[11] E. A. Kangarshahi and A. Guillén i Fàbregas, "A single-letter upper bound to the mismatch capacity," *IEEE Trans. Inf. Theory*, vol. 67, no. 4, pp. 2013–2033, Apr. 2021.

[12] E. A. Kangarshahi and A. Guillén i Fàbregas, "An upper bound to the mismatch capacity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2019, pp. 287–2877.

[13] A. Somekh-Baruch, "A single-letter upper bound on the mismatch capacity via multicast transmission," *IEEE Trans. Inf. Theory*, vol. 68, no. 5, pp. 2801–2812, May 2022.

[14] A. Somekh-Baruch, "Robust multicasting and an upper bound on the mismatch capacity of the DMC," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 1130–1135.

[15] E. A. Kangarshahi and A. Guillén i Fàbregas, "A sphere-packing exponent for mismatched decoding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2021, pp. 1136–1141.

[16] M. Bondaschi, A. Guillén i Fàbregas, and M. Dalai, "Mismatched decoding reliability function at zero rate," *IEEE Trans. Inf. Theory*, vol. 68, no. 3, pp. 1482–1495, Mar. 2021.

[17] J. Scarlett, L. Peng, N. Merhav, A. Martinez, and A. G. I. Fàbregas, "Expurgated random-coding ensembles: Exponents, refinements, and connections," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4449–4462, Aug. 2014.

[18] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[19] J. von Neumann, "Zur theorie der gesellschaftsspiele," *Math. Annalen.*, vol. 100, pp. 295–320, Aug. 1928.

[20] R. G. Gallager. (1994). *Fixed Composition Arguments and Lower Bounds to the Error Probability*. Lecture Notes MIT 6.441. [Online]. Available: https://www.rle.mit.edu/rgallager/documents/notes5.pdf and https://tinyurl.com/22akp3ba

Ehsan Asadi Kangarshahi received the dual bachelor's degree in mathematics and electrical engineering from the Sharif University of Technology in 2017, the Ph.D. degree from the Department of Engineering, University of Cambridge, in October 2021, and the Ph.D. degree from the University of Cambridge in 2022.

From October 2017 to December 2017, he was a Research Intern at the École Polytechnique Fédérale de Lausanne (EPFL). He was also a member of Trinity Hall. His research interests are in the areas of information theory, communication theory, and statistics.

Albert Guillén i Fàbregas (Fellow, IEEE) received the degree in telecommunications engineering from Universitat Politècnica de Catalunya, the degree in electronics engineering from the Politecnico di Torino, in 1999, and the Ph.D. degree in communications systems from École Polytechnique Fédérale de Lausanne (EPFL), in 2004.

In 2020, he returned to a full-time faculty position at the Department of Engineering, University of Cambridge, where he had been a full-time Faculty Member and a fellow of Trinity Hall from 2007 to 2012. Since 2011, he has been an ICREA Research Professor at Universitat Pompeu Fabra (currently on leave). He has held appointments at the New Jersey Institute of Technology, Telecom Italia, European Space Agency (ESA), Institut Eurécom, University of South Australia, Universitat Pompeu Fabra, and University of Cambridge, and visiting appointments at EPFL, École Nationale des Télécommunications, Paris, Universitat Pompeu Fabra, University of South Australia, Centrum Wiskunde & Informatica, and Texas A&M University at Qatar. His specific research interests are in the areas of information theory, communication theory, coding theory, and statistical inference.

Dr. Guillén i Fàbregas is a fellow of the Institute for Mathematics and its Applications (IMA) and a member of the Young Academy of Europe. He received Starting and Consolidator Grants from the European Research Council, the Young Authors Award of the 2004 European Signal Processing Conference, the 2004 Best Doctoral Thesis Award from the Spanish Institution of Telecommunications Engineers, and a Research Fellowship of the Spanish Government to join ESA. Since 2013, he has been an Editor of the *Foundations and Trends in Communications and Information Theory* and Now Publishers and was an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY (2013–2020) and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2007–2011).