A Sphere-Packing Exponent for Mismatched Decoding

Ehsan Asadi Kangarshahi and Albert Guillén i Fàbregas

Abstract

We derive a sphere-packing error exponent for mismatched decoding over discrete memoryless channels. We find a lower bound to the probability of error of mismatched decoding that decays exponentially for coding rates smaller than a new upper bound to the mismatch capacity. 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.

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: a) the channel transition is unknown and imperfectly estimated or, b) when, for complexity reasons, the channel likelihood is too complex 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].

Finding a single-letter expression for the mismatch capacity remains an open problem. A number of single-letter lower bounds have been derived in the literature [2]–[5] (see also [1] for a recent survey, including multiuser coding achievable rates). Instead, up until recently, not much progress had been made on upper bounds. Balakirsky [6] claimed that for binary-input discrete memoryless channels (DMC), the mismatch capacity coincided with the lower bound in [3], [4]. Reference [7] provided a counterexample to this converse invalidating its claim. In [8], we proposed a single-letter upper bound to the mismatch capacity based on transforming the channel in such a way that errors on the transformed channel with high probability imply a mismatched-decoding error in the original channel. Reference [9] derived a new single-letter upper bound based on a multicast approach that improves over [8] in zero-error problems and remains valid for continuous channels.

In this paper, we derive a sphere-packing upper bound to the error exponent that yields a new upper bound on the mismatch capacity. The new bound improves over known bounds, subsumes that in [8], and provides significant gains.

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 $P_e(\mathcal{C}_n) = \mathbb{P}[\hat{m} \neq m]$ and $P_{e,\max}(\mathcal{C}_n) = \max_{m\in\{1,\ldots,M\}} \mathbb{P}[\hat{m} \neq m|m \text{ is sent}]$. The decoder that minimizes the error probability is the maximum-likelihood (ML) decoder, that produces the message estimate $\hat{m}$ according to

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

Rate $R > 0$ is 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

$$\hat{m} = \underset{m\in\{1,\ldots,M\}}{\arg\max}\, q^n(\boldsymbol{x}_m, \boldsymbol{y}), \tag{2}$$

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]–[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 [10, 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 $\hat{\boldsymbol{p}}_{\boldsymbol{xy}}(j,k) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{x_i = j, y_i = k\}$. The conditional type of $\boldsymbol{y}$ given $\boldsymbol{x}$ is the matrix

$$\hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}}(k|j) = \begin{cases} \frac{\hat{\boldsymbol{p}}_{\boldsymbol{xy}}(j,k)}{\hat{\boldsymbol{p}}_{\boldsymbol{x}}(j)} & \hat{\boldsymbol{p}}_{\boldsymbol{x}}(j) > 0 \\ \frac{1}{|\mathcal{Y}|} & \text{otherwise.} \end{cases} \tag{3}$$

The set of all 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 $k_1, k_2 \in \mathcal{Y}$,

$$\hat{\boldsymbol{p}}_{\boldsymbol{xy}\hat{\boldsymbol{y}}}(j, k_1, k_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{x_i = j, y_i = k_1, \hat{y}_i = k_2\}. \tag{4}$$

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

$$\hat{\boldsymbol{p}}_{\boldsymbol{y}\hat{\boldsymbol{y}}|\boldsymbol{x}}(k_1, k_2|j) = \begin{cases} \frac{\hat{\boldsymbol{p}}_{\boldsymbol{xy}\hat{\boldsymbol{y}}}(j,k_1,k_2)}{\hat{\boldsymbol{p}}_{\boldsymbol{x}}(j)} & \hat{\boldsymbol{p}}_{\boldsymbol{x}}(j) > 0 \\ \frac{1}{|\mathcal{Y}|}\mathbb{1}\{k_1 = k_2\} & \text{otherwise.} \end{cases} \tag{5}$$

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:

$$\mathcal{T}_{\boldsymbol{yx}}(\boldsymbol{p}_{Y\hat{Y}|X}) = \{\hat{\boldsymbol{y}} \in \mathcal{Y}^n \,|\, \hat{\boldsymbol{p}}_{\boldsymbol{y}\hat{\boldsymbol{y}}|\boldsymbol{x}} = \boldsymbol{p}_{Y\hat{Y}|X}\}. \tag{6}$$

The mutual information and conditional relative entropy are respectively defined as

$$I(P_X, P_{Y|X}) \triangleq \mathbb{E}\Big[\log \frac{P_{Y|X}(Y|X)}{\sum_{x'} P_X(x') P_{Y|X}(Y|x')}\Big], \tag{7}$$

$$D(P_{Y'|X}\|P_{Y|X}|P_X) \triangleq \sum_{x\in\mathcal{X}} P_X(x)\cdot D(P_{Y'|X=x}\|P_{Y|X=x}). \tag{8}$$

Definition 1: Let $\mathcal{C}_n = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}$ be a codebook and W be a channel. The type-conflict error probability is defined as

$$\begin{aligned} &P_{\text{tce}}^{\max}(\mathcal{C}_n, W) \\ &\triangleq \max_{m\in\{1,\ldots,M\}} \mathbb{P}\big[\cup_{\bar{m}\neq m}\{\hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}_m} = \hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}_{\bar{m}}}\}|\boldsymbol{x}_m \text{ sent}\big] \end{aligned} \tag{9}$$

where the probability is over $W^n(\boldsymbol{y}|\boldsymbol{x}_m)$.

Definition 2: Let $\mathcal{C}_n = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}$ be a codebook and W be a channel. Then, for $\varepsilon \geq 0$, we define

$$\begin{aligned} &P_{e,\max}^{q}(\mathcal{C}_n, W, \varepsilon) \\ &\triangleq \max_{m\in\{1,\ldots,M\}} \mathbb{P}\big[\cup_{\bar{m}\neq m}\{q^n(\boldsymbol{x}_{\bar{m}}, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_m, \boldsymbol{y}) + \varepsilon\}|\boldsymbol{x}_m \text{ sent}\big] \end{aligned} \tag{10}$$

where the probability is over $W^n(\boldsymbol{y}|\boldsymbol{x}_m)$, and $P_{e,\max}^{q}(\mathcal{C}_n, W) = P_{e,\max}^{q}(\mathcal{C}_n, W, \varepsilon = 0)$.

Then, $P_{e,\max}^{q}(\mathcal{C}_n, W, \varepsilon)$ is a generalization of the probability of error of codebook $\mathcal{C}_n$ under mismatch decoding, as it allows for some margin ε.

Similarly to [8], 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 an ML error, where we simply have a likelihood inequality. In addition, it can be shown that for rates $R > C(V)$, then the probability of type-conflict errors tends to one exponentially.

We proceed by introducing a few definitions. Recall the definition of maximal set from [8]. Consider the set

$$\mathcal{S}_q(k_1, k_2) \triangleq \big\{j \in \mathcal{X} | j = \arg\max_{j'\in\mathcal{X}} q(j', k_2) - q(j', k_1)\big\}. \tag{11}$$

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

$$P_{Y\hat{Y}|X}(k_1, k_2|j) = 0 \quad \text{if} \quad j \notin \mathcal{S}_q(k_1, k_2). \tag{12}$$

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

$$\min_{\substack{P_{X_2|X_1\hat{Y}}:\\ X_2-X_1\hat{Y}-Y\\ P_{\hat{Y}X_2}=P_{\hat{Y}X_1}}} \mathbb{E}[q(X_2,Y)] \geq \mathbb{E}[q(X_1,Y)] \tag{13}$$

where the notation $X_2 - X_1\hat{Y} - Y$ denotes that $X_2, (X_1\hat{Y})$ and Y form a Markov chain. In addition, define $\mathcal{M}^{\delta}_{\max}(q, P_{X_1})$ as the set of all distributions satisfying

$$\min_{\substack{P_{X_2|X_1\hat{Y}}:\\ X_2-X_1,\hat{Y}-Y\\ P_{\hat{Y}X_2}=P_{\hat{Y}X_1}}} \mathbb{E}[q(X_2,Y)] \geq \mathbb{E}[q(X_1,Y)] + \delta \tag{14}$$

so that $\mathcal{M}^{\delta}_{\max}(q, P_{X_1})$ is an approximation of $\mathcal{M}_{\max}(q, P_{X_1})$. For types, $\hat{\mathcal{M}}_{\max}$ and $\hat{\mathcal{M}}^{\delta}_{\max}$ are similarly defined.

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 [8, Lemma 3] we have for any X_2 such that $P_{\hat{Y}X_1} = P_{\hat{Y}X_2}$

$$\mathbb{E}[q(X_2,Y)] \geq \mathbb{E}[q(X_1,Y)] \tag{15}$$

Therefore $P_{Y\hat{Y}|X_1}$ satisfies (13) and as a result $P_{Y\hat{Y}|X_1} \in \mathcal{M}_{\max}(q, P_{X_1})$. This enlarged set of maximal distributions enables a better upper bound on the mismatch capacity.

III. Sphere-Packing Exponent

In this section, we derive a sphere-packing exponent for mismatched decoding using the method developed in [8].

Theorem 1: 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

$$-\frac{1}{n}\log P_e^q(\mathcal{C}_n, W) \leq E^q_{\text{sp}}(\boldsymbol{p}_X, R+\zeta_n) - \delta_n, \tag{16}$$

where

$$E^q_{\text{sp}}(P_X, R) = \min_{\substack{P_{Y'\hat{Y}|X}\in\mathcal{M}_{\max}(q,P_X)\\ I(P_X,P_{\hat{Y}|X})\leq R}} D(P_{Y'|X}\|P_{Y|X}|P_X) \tag{17}$$

and

$$\zeta_n = (JK-1)\frac{\log(n+1)}{n} + \frac{\log 2}{n} \tag{18}$$

$$\delta_n = \big((JK-1)+(J^2K-1)\big)\frac{\log(n+1)}{n} + \frac{\log 2}{n}. \tag{19}$$

The derivation of the above exponent follows similar footsteps as that in Gallager's lecture notes on fixed composition codes [11]. The proof is based on three lemmas. The first lemma, shows a lower bound to the type-conflict error probability of code $\mathcal{C}_n$ over an auxiliary channel. The second lemma 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. The third lemma shows that if the joint conditional distribution that defines W and the auxiliary channels is maximal according to (13), then, the error probability of the q-decoder over channel W is lower-bounded by the type-conflict error probability over the auxiliary channel.

Lemma 1: Assume codebook $\mathcal{C}_n$ consists of M codewords of composition $\boldsymbol{p}_X$ used over a DMC $P_{\hat{Y}|X}$. Assume that noise composition $\boldsymbol{p}_{\hat{Y}|X_1}$ is such that $M|\mathcal{T}_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X_1})| \geq 2|\mathcal{T}(\boldsymbol{p}_{\hat{Y}})|$. 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

$$\begin{aligned}\mathbb{P}\big[\exists \boldsymbol{x}_2 \in \mathcal{C}_n\backslash\{\boldsymbol{x}_1\} \text{ s.t. } \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \boldsymbol{p}_{\hat{Y}X_1X_2}|\boldsymbol{x}_1\big]&\\ \geq \frac{1}{2(n+1)^{J^2K-1}}\mathbb{P}\big[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})|\boldsymbol{x}_1\big]&\end{aligned} \tag{20}$$

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

Proof: From Gallager's lecture notes on fixed composition codes [11, Lemma 4] we have there exist a codeword $\boldsymbol{x}_1 \in \mathcal{C}_n$ such that

$$\begin{aligned}\mathbb{P}\big[\exists \boldsymbol{x}_2 \in \mathcal{C}_n \setminus \{\boldsymbol{x}_1\} \text{ s.t. } \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1} = \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_2} = \boldsymbol{p}_{\hat{Y}X_1}|\boldsymbol{x}_1\big]&\\ \geq \frac{1}{2}\mathbb{P}\big[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})|\boldsymbol{x}_1\big]&\end{aligned} \tag{21}$$

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 $\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}|\boldsymbol{x}_1} = \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}|\boldsymbol{x}_2}$. We now construct a joint type 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

$$\begin{aligned}&\mathcal{E}_{\boldsymbol{x}_1}(\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1})\\ &= \big\{\hat{\boldsymbol{y}} \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1}) \,|\, \exists \boldsymbol{x}_2 \in \mathcal{C}_n \setminus \{\boldsymbol{x}_1\},\\ &\qquad \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2}, \tilde{\boldsymbol{p}}_{\hat{Y}X_1} = \tilde{\boldsymbol{p}}_{\hat{Y}X_2} = \boldsymbol{p}_{\hat{Y}X_1}\big\}. \end{aligned} \tag{22}$$

In 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 $\boldsymbol{p}^\star_{\hat{Y}X_1X_2}$ that satisfies the following

$$\boldsymbol{p}^\star_{\hat{Y}X_1X_2} = \operatorname*{arg\,max}_{\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2} \in \mathcal{P}_n(\mathcal{Y}\times\mathcal{X}^2)} |\mathcal{E}_{\boldsymbol{x}_1}(\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1})|, \tag{23}$$

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 $\boldsymbol{p}_{\hat{Y}X_1}$. Out of all joint types $\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2}$, $\boldsymbol{p}^\star_{\hat{Y}X_1X_2}$ is the one that contains the maximum number of outputs $\hat{\boldsymbol{y}}$ that yield a type-conflict error.

Observe that the left hand side of (21) can be bounded as

$$\begin{aligned}&\mathbb{P}\big[\exists \boldsymbol{x}_2 \in \mathcal{C}_n \setminus \{\boldsymbol{x}_1\} \text{ s.t. } \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1} = \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_2} = \boldsymbol{p}_{\hat{Y}X_1} | \boldsymbol{x}_1\big]\\ &\quad= \sum_{\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2} \in \mathcal{P}_n(\mathcal{Y}\times\mathcal{X}^2)} \mathbb{P}[\mathcal{E}_{\boldsymbol{x}_1}(\tilde{\boldsymbol{p}}_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1}) | \boldsymbol{x}_1] && (24)\\ &\quad\le (n+1)^{J^2K-1} \mathbb{P}[\mathcal{E}_{\boldsymbol{x}_1}(\boldsymbol{p}^\star_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1})] && (25)\end{aligned}$$

and thus, from (21), we get

$$\begin{aligned}&\mathbb{P}\big[\exists \boldsymbol{x}_2 \in \mathcal{C}_n \setminus \{\boldsymbol{x}_1\} \text{ s.t. } \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \boldsymbol{p}^\star_{\hat{Y}X_1X_2}\big]\\ &\qquad \ge \frac{1}{2(n+1)^{J^2K-1}} \mathbb{P}\big[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1}) | \boldsymbol{x}_1\big] \end{aligned} \tag{26}$$

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

Observe that the above statement implies that

$$\begin{aligned}\frac{|\mathcal{E}_{\boldsymbol{x}_1}(\boldsymbol{p}^\star_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1})|}{|\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})|} &= \frac{\mathbb{P}\big[\mathcal{E}_{\boldsymbol{x}_1}(\boldsymbol{p}^\star_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1})\big]}{\mathbb{P}\big[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1}) | \boldsymbol{x}_1\big]} && (27)\\ &\ge \frac{1}{2(n+1)^{J^2K-1}} && (28)\end{aligned}$$

where (27) is true because all elements of $\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ are equiprobable when $\boldsymbol{x}_1$ is sent.

Similarly to [8], we construct a bipartite graph $\mathcal{G}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ in the following way (see [8] for details). Vertices of this graph consists of 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 $\hat{\boldsymbol{p}}_{\boldsymbol{y}'\hat{\boldsymbol{y}}\boldsymbol{x}_1} = \boldsymbol{p}_{Y'\hat{Y}X_1}$.

Lemma 2: Consider a conditional 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. 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 graph $\mathcal{G}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$, then, for every $\boldsymbol{x}_2$ such that

$$\hat{\boldsymbol{p}}_{\boldsymbol{y}'\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \hat{\boldsymbol{p}}_{\boldsymbol{y}'|\hat{\boldsymbol{y}}\boldsymbol{x}_1} \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2}, \tag{29}$$

$$\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_2} = \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1} \tag{30}$$

we have a q-decoding error

$$q^n(\boldsymbol{x}_2, \boldsymbol{y}') \ge q^n(\boldsymbol{x}_1, \boldsymbol{y}'). \tag{31}$$

Proof: This is proven due to the fact that we know if $\hat{\boldsymbol{p}}_{\boldsymbol{y}'\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \boldsymbol{p}_{Y'\hat{Y}X_1X_2}$ we can write the metric difference as

$$q^n(\boldsymbol{x}_2, \boldsymbol{y}') - q^n(\boldsymbol{x}_1, \boldsymbol{y}') = \mathbb{E}[q(X_2, Y') - q(X_1, Y')] \tag{32}$$

where the expectation is taken with respect to type $\boldsymbol{p}_{Y'\hat{Y}X_1X_2}$. Since $\boldsymbol{p}_{Y'\hat{Y}|X_1} \in \hat{\mathcal{M}}_{\max}(q, \boldsymbol{p}_X)$, and from (29) and (30) we have that $\boldsymbol{p}_{\hat{Y}X_1} = \boldsymbol{p}_{\hat{Y}X_2}$ and $\boldsymbol{p}_{Y'\hat{Y}X_1X_2} = \boldsymbol{p}_{Y'|\hat{Y}X_1}\boldsymbol{p}_{\hat{Y}X_1X_2}$, i.e, $X_2 - X_1\hat{Y} - Y'$ form a Markov chain, based on definition of $\hat{\mathcal{M}}_{\max}(q, \boldsymbol{p}_{X_1})$ we have

$$\mathbb{E}[q(X_2, Y') - q(X_1, Y')] \geq 0 \tag{33}$$

and thus, from (32), we get the desired result. ■

The next lemma relates the q-decoding error probability in channel $P_{Y'|X}$ with the type-conflict error probability in channel $P_{\hat{Y}|X}$ by using the fact that in the conditions of the maximal set we have included that $X_2 - X_1\hat{Y} - Y'$ is a Markov chain.

Lemma 3: Let $\boldsymbol{p}_{Y'\hat{Y}|X_1} \in \hat{\mathcal{M}}_{\max}(q, \boldsymbol{p}_{X_1})$ and $\boldsymbol{x}_1 \in \mathcal{T}(\boldsymbol{p}_{X_1})$, then

$$P_e^q(\mathcal{C}_n, W) \geq \frac{1}{2(n+1)^{J^2K-1}}\mathbb{P}[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})|\boldsymbol{x}_1] \tag{34}$$

Where the probability of error is computed w.r.t channel W and the second probability is computed w.r.t the channel $P_{\hat{Y}|X}$.

Proof: Consider the bipartite graph $\mathcal{G}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'\hat{Y}|X_1})$ connecting elements of $\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})$ and $\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$. As described in [8], the graph 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. For any $\mathcal{B} \subset \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$ we define $\Psi(\mathcal{B})$ as

$$\begin{aligned}\Psi(\mathcal{B}) = \{&\boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1}) \mid \boldsymbol{y}' \text{ is connected to some}\\ &\hat{\boldsymbol{y}} \in \mathcal{B} \text{ in graph } \mathcal{G}_{\boldsymbol{x}}(\boldsymbol{p}_{Y'\hat{Y}|X_1})\}\end{aligned} \tag{35}$$

As a result using the result stated in [8] we get that for any $\mathcal{B} \subset \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})$

$$\frac{|\Psi(\mathcal{B})|}{|\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|} \geq \frac{|\mathcal{B}|}{|\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})|} \tag{36}$$

Now we 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 $\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_2} = \boldsymbol{p}_{\hat{Y}X_1X_2}$, i.e.,

$$\mathcal{B} = \mathcal{E}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1}). \tag{37}$$

Therefore, from Lemma 2 we have for any $\boldsymbol{y}' \in \Psi(\mathcal{B})$ there exists a codeword $\boldsymbol{x}_2 \neq \boldsymbol{x}_1$ such that

$$q^n(\boldsymbol{x}_2, \boldsymbol{y}') \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}') \tag{38}$$

and we bound the probability of error as follows

$$\begin{aligned}
&P_e^q(\mathcal{C}_n, W)\\
&= \mathbb{P}[\exists \boldsymbol{x}_2 \in \mathcal{C}_n\backslash\{\boldsymbol{x}_1\}, q^n(\boldsymbol{x}_2, \boldsymbol{y}') \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}')] && (39)\\
&\geq \mathbb{P}[\exists \boldsymbol{x}_2 \in \mathcal{C}_n\backslash\{\boldsymbol{x}_1\}, q^n(\boldsymbol{x}_2, \boldsymbol{y}') \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}'), \boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})] && (40)\\
&\geq \mathbb{P}[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|\boldsymbol{x}_1]\\
&\cdot \mathbb{P}[\exists \boldsymbol{x}_2 \in \mathcal{C}_n\backslash\{\boldsymbol{x}_1\}, q^n(\boldsymbol{x}_2, \boldsymbol{y}') \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}')|\boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})] && (41)\\
&= \mathbb{P}[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|\boldsymbol{x}_1]\cdot\\
&\cdot \frac{|\{\boldsymbol{y}' \in \mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1}) \mid \exists \boldsymbol{x}_2 \in \mathcal{C}_n\backslash\{\boldsymbol{x}_1\}, q^n(\boldsymbol{x}_2, \boldsymbol{y}') \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}')\}|}{|\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|} && (42)\\
&\geq \mathbb{P}[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|\boldsymbol{x}_1]\frac{\left|\Psi\left(\mathcal{E}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1})\right)\right|}{|\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|} && (43)\\
&\geq \mathbb{P}[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|\boldsymbol{x}_1]\frac{|\mathcal{E}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}X_1X_2}, \boldsymbol{p}_{\hat{Y}X_1})|}{|\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{\hat{Y}|X_1})|} && (44)\\
&\geq \mathbb{P}[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|\boldsymbol{x}_1] \cdot \frac{1}{2(n+1)^{J^2K-1}} && (45)
\end{aligned}$$

where all of probabilities are computed with respect to channel W^n, (43) follows from all elements of $\Psi(\mathcal{B})$ satisfying (38), (44) follows from (36) and (45) follows from (28). ■

Using a standard property of noise types we have that

$$\mathbb{P}[\mathcal{T}_{\boldsymbol{x}_1}(\boldsymbol{p}_{Y'|X_1})|\boldsymbol{x}_1] \geq e^{-n\left(D(P_{Y'|X_1}\|P_{Y|X_1}|\boldsymbol{p}_{X_1})+\zeta_n\right)} \tag{46}$$

with $\zeta_n = \frac{JK-1}{n}\log(n+1)$. From standard arguments of the method of types we obtain (16), where we have set $\boldsymbol{p}_X = \boldsymbol{p}_{X_1}$.

Again using standard arguments (see e.g. [11, Th. 2]) the result of Theorem 1 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

$$P^q_{e,\max}(\mathcal{C}_n, W) \geq P^q_{e,\max}(\mathcal{C}'_n, W). \tag{47}$$

Additionally, a similar analysis would give an identical upperbound to the error exponent using the maximal sets $\hat{\mathcal{M}}_{\max}(q)$ from [8].

As is well known, the exponent from Theorem 1 is decreasing in R and $E^q_{\rm sp}(\boldsymbol{p}_X, R) = 0$ by choosing $Y' = Y$ in (17) at a rate equal to

$$\bar{R}_q(W, \boldsymbol{p}_X) \triangleq \min_{\substack{P_{Y\hat{Y}|X} \in \mathcal{M}_{\max}(q, \boldsymbol{p}_X) \\ P_{Y|X} = W}} I(\boldsymbol{p}_X, P_{\hat{Y}|X}) \tag{48}$$

We have shown that for rates $R < \bar{R}_q(W, \boldsymbol{p}_X)$, the error probability decays at most exponentially. In the next section, we show that for rates $R > \bar{R}_q(W, \boldsymbol{p}_X)$ the error probability cannot decay sub-exponentially and is bounded away from zero as n tends to infinity.

IV. Converse

In this section, we show that for coding rates R

$$R > \bar{R}_q(W, P_X) = \min_{\substack{P_{Y\hat{Y}|X} \in \mathcal{M}_{\max}(q, P_X) \\ P_{Y|X} = W}} I(P_X, P_{\hat{Y}|X}) \tag{49}$$

for a fixed input distribution P_X, the error probability is bounded away from zero as n tends to infinity. Proofs are not included due to space limitations and can be found in B.

Theorem 2: Let $\mathcal{C}_n = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ be a constant composition codebook of composition $\boldsymbol{p}_X$ and length n. Assume $P_{Y\hat{Y}|X} \in \mathcal{M}^\delta_{\max}(q, \boldsymbol{p}_X)$ and $P_{Y|X} = W$. Then for any $\epsilon > 0$, there exists a constant $\gamma_n(\epsilon)$ that depends on n, W and q, such that $0 < \gamma_n(\epsilon) < 1$ for every n, W and q, such that

$$P^q_{e,\max}(\mathcal{C}_n, W, n\epsilon) \geq \big(1 - \gamma_n(\epsilon)\big) P^{\max}_{\rm tce}(\mathcal{C}_n, P_{\hat{Y}|X}). \tag{50}$$

The next result from [8] lower bounds the type-conflict error probability.

Theorem 3: With the assumptions of Theorem 2, for every P_X, there exist $n_0, \bar{E}(R) > 0$ such that if $n > n_0$ and $\frac{1}{n}\log|\mathcal{C}_n| > I(P_X, P_{\hat{Y}|X})$

$$P^{\max}_{\rm tce}(\mathcal{C}_n, P_{\hat{Y}|X}) \geq 1 - 2^{-n\bar{E}(R)}. \tag{51}$$

The following result, also from [8] allows to establish a connection between codes of arbitrary distributions and constant composition codes.

Theorem 4: Let W, q be channel and decoding metric, respectively. Define, for any input distribution P_X,

$$\bar{R}_q(W, P_X) = \min_{\substack{P_{Y\hat{Y}|X} \in \mathcal{M}_{\max}(q, P_X) \\ P_{Y|X} = W}} I(P_X, P_{\hat{Y}|X}) \tag{52}$$

If $R > \bar{R}_q(W, P_X)$, $\exists n_0 \in \mathbb{N}$, $0 < \gamma < 1$ and $\bar{E}(R) > 0$ such that for $n > n_0$, the error probability of any codebook $\mathcal{C}_n$ of length n, $M \geq 2^{nR}$ codewords satisfies

$$P^q_{e,\max}(\mathcal{C}_n, W) \geq (1 - \gamma)(1 - 2^{-n\bar{E}(R)}). \tag{53}$$

Proof: For any distribution P_X, set the code rate to be $R > \bar{R}_q(W, P_X)$. Similarly to the previous section, we know that for any code $\mathcal{C}_n$ of length n and rate R, there exists a constant composition subcode $\mathcal{C}'_n \subset \mathcal{C}_n$ with length n satisfying, rate $R' > R - \frac{J-1}{n}\log(n+1)$, and composition $\boldsymbol{p}_X$ such that

$$P^q_{e,\max}(\mathcal{C}_n, W) \geq P^q_{e,\max}(\mathcal{C}'_n, W). \tag{54}$$

Applying Theorems 2 and 3 to code $\mathcal{C}'_n$, we get that for any $\delta > 0$, if

$$R > \bar{R}_q(W, P_X) > \min_{\substack{P_{Y\hat{Y}|X} \in \mathcal{M}^\delta_{\max}(q, \boldsymbol{p}_X) \\ P_{Y|X} = W}} I(\boldsymbol{p}_X, P_{\hat{Y}|X}) \tag{55}$$

we have that

$$P^q_{e,\max}(\mathcal{C}'_n, W, n\epsilon) \geq \big(1 - \gamma_n(\epsilon)\big) P^{\max}_{\rm tce}(\mathcal{C}'_n, P_{\hat{Y}|X}) \tag{56}$$

$$\geq \big(1 - \gamma_n(\epsilon)\big)(1 - 2^{-n\bar{E}(R)}) \tag{57}$$

where (57) is bounded away from zero as n tends to infinity. Now since the above inequality holds for any $\delta > 0$ we get the desired result.

■

Corollary 1: We have

$$C_q(W) \le \max_{P_X} \min_{\substack{P_{Y\hat{Y}|X} \in \mathcal{M}_{\max}(q, P_X) \\ P_{Y|X} = W}} I(P_X, P_{\hat{Y}|X}) \tag{58}$$

In terms of computation, unlike the bound proposed in [8], optimizing (58) is not a simple task. This observation stems from the fact that the maximal set $\mathcal{M}_{\max}(q, P_X)$ in (58) depends on P_X, unlike the maximal set $\mathcal{M}_{\max}(q)$ in [8]. 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 [8]. As illustrated next, the advantages of the new bound are potentially significant.

A. Example

In this part we show the application of our bound to the counterexample in [7], where the channel and metric are

$$W = \begin{bmatrix} 0.97 & 0.03 & 0 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}, q = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \log(0.5) & \log(1.36) \end{bmatrix}. \tag{59}$$

For this example $C(W) = 0.7133$ bits/use, the rate achievable by 2-letter superposition coding from [7] is $R_{\rm sc}^{(2)}(W, q) = 0.1991$ bits/use and our previous converse [8] stated that $C_q(W) \le \bar{R}_q(W) = 0.6182$ bits/use. Due to the intricate nature of the optimization problem (58) (see above discussion), we have so far only been able to compute the bound for the fixed input distribution $P_X = [0.75597 \;\; 0.24403]$, which is the maximizing input distribution for the LM rate [3], [4]. The joint conditional distribution given in I is maximal for the above P_X.

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)	$P_{Y\hat{Y}\|X}$
$(1,1,1)$	0.37	$(2,1,1)$	0.1
$(1,1,2)$	0.6	$(2,2,2)$	0.1
$(1,2,2)$	0.03	$(2,3,3)$	0.62
		$(2,3,2)$	0.18

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

$$P_{\hat{Y}|X} = \begin{bmatrix} 0.37 & 0.63 & 0 \\ 0.1 & 0.28 & 0.62 \end{bmatrix}. \tag{60}$$

which upper bounds the rates achievable with distribution $P_X = [0.75597 \;\; 0.24403]$ by

$$I(P_X; P_{\hat{Y}|X}) = 0.3824 \text{ bits/use}. \tag{61}$$

APPENDIX A
CONDITIONING ON THE TYPE OF A SEQUENCE

In this section we study the effect of conditioning on type of a sequence in its statistical properties.

Lemma 4: 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 i.i.d. random variables taking values on alphabets $\mathcal{Z}, \mathcal{S}$, respectively. Moreover, let $\hat{\boldsymbol{p}}_{\boldsymbol{z}}$ denote the type of $\boldsymbol{z} = (z_1, z_2, \ldots, z_n)$. Analogously, $\hat{\boldsymbol{p}}_{\boldsymbol{s}}$ denotes the type of $\boldsymbol{s} = (s_1, s_2, \ldots, s_n)$. Then, we have

$$\mathbb{E}\left[\sum_{i=1}^{n} f(Z_i, S_i) \middle| \hat{\boldsymbol{p}}_{\boldsymbol{z}}\right] = n\mathbb{E}_{P_{S|Z} \times \hat{\boldsymbol{p}}_{\boldsymbol{z}}}[f(\tilde{Z}, S)] \tag{62}$$

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

Proof:

$$\mathbb{E}\left[\sum_{i=1}^{n} f(Z_i,S_i)\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\right] = \mathbb{E}\left[\sum_{i=1}^{n}\sum_{z} f(z,S_i)\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z)\right] \tag{63}$$

$$= \sum_{i=1}^{n}\sum_{z} \mathbb{E}\left[f(z,S_i)\right]\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z) \tag{64}$$

$$= \sum_{i=1}^{n}\sum_{z} \mathbb{E}_{P_{S|\tilde{Z}=z}}[f(z,S)]\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z) \tag{65}$$

$$= n\mathbb{E}_{\hat{\boldsymbol{p}}_{\boldsymbol{z}}}\left[\mathbb{E}_{P_{S|\tilde{Z}}}\left[f(\tilde{Z},S)\Big|\tilde{Z}\right]\right] \tag{66}$$

$$= n\mathbb{E}_{P_{S|\tilde{Z}}\times\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z)}[f(\tilde{Z},S)] \tag{67}$$

Where (63) and (64) are derived from the definition of type and conditional expectation, (65) follows by replacing random variables S_i by S which does not effect the expectation and (67) follows from the tower rule of conditional expectation. ■

Lemma 5: With the assumptions of Lemma 4 we have

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} f(Z_i,S_i)\right)^2\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\right] = n^2\mathbb{E}_{P_{S|Z}\times\hat{P}_{\boldsymbol{z}}}[f(\tilde{Z},S)]^2 + n\mathbb{E}_{P_{S|Z}\times\hat{P}_{\boldsymbol{z}}}[f(\tilde{Z},S)^2] - n\mathbb{E}_{\hat{P}_{\boldsymbol{z}}}\left[\mathbb{E}_{P_{S|Z}}[f(\tilde{Z},S)|\tilde{Z}]^2\right] \tag{68}$$

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

Proof: By expanding the term in the expectation we have

$$\mathbb{E}\left[\left(\sum_{i=1}^{n} f(Z_i,S_i)\right)^2\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\right] = \mathbb{E}\left[\sum_{i\neq k} f(Z_i,S_i)f(Z_k,S_k)\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\right] + \mathbb{E}\left[\sum_{i=1}^{n} f(Z_i,S_i)^2\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\right] \tag{69}$$

Then for the first term of the right hand side of (69) we can use Lemma 4

$$\mathbb{E}\left[\sum_{i=1}^{n} f(Z_i,S_i)^2\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\right] = n\mathbb{E}_{P_{Z|S}\times\hat{\boldsymbol{p}}_{\boldsymbol{z}}}[f(\tilde{Z},S)^2] \tag{70}$$

Where $\tilde{Z}$ is a random variable with distribution $\hat{\boldsymbol{p}}_{\boldsymbol{z}}$. Moreover, for the second term of right hand side of (69) we have

$$\begin{aligned}
\mathbb{E}\left[\sum_{i\neq k} f(Z_i,S_i)f(Z_k,S_k)\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\right] &= \mathbb{E}\left[\sum_{z_1\neq z_2}\sum_{i\neq k} f(z_1,S_i)f(z_2,S_k)\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_1)\frac{n\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_2)}{n-1}\right] \\
&\quad + \mathbb{E}\left[\sum_{z}\sum_{i\neq k} f(z,S_i)f(z,S_k)\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z)\frac{n\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z)-1}{n-1}\right] &(71)\\
&= \frac{n}{n-1}\mathbb{E}\left[\sum_{i\neq k}\sum_{z_1,z_2} f(z_1,S_i)\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_1)f(z_2,S_k)\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_2)\right] \\
&\quad - \frac{1}{n-1}\mathbb{E}\left[\sum_{i\neq k}\sum_{z} f(z,S_i)f(z,S_k)\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z)^2\right] &(72)\\
&= \frac{n}{n-1}\sum_{i\neq k}\mathbb{E}_{P_{S|Z}\times\hat{\boldsymbol{p}}_{\boldsymbol{z}}}\left[f(\tilde{Z},S_i)\right]\mathbb{E}_{P_{S|Z}\times\hat{\boldsymbol{p}}_{\boldsymbol{z}}}\left[f(\tilde{Z},S_k)\right] \\
&\quad - \frac{1}{n-1}\sum_{i\neq k}\mathbb{E}_{\hat{\boldsymbol{p}}_{\boldsymbol{z}}}\left[\mathbb{E}_{P_{S|Z}}[f(\tilde{Z},S_i)|\tilde{Z}]\mathbb{E}_{P_{S|Z}}[f(\tilde{Z},S_k)|\tilde{Z}]\right] &(73)\\
&= 2\binom{n}{2}\left(\frac{n}{n-1}\mathbb{E}_{P_{S|Z}\times\hat{\boldsymbol{p}}_{\boldsymbol{z}}}\left[f(\tilde{Z},S)\right]^2 - \frac{1}{n-1}\mathbb{E}_{\hat{\boldsymbol{p}}_{\boldsymbol{z}}}\left[\mathbb{E}_{P_{S|Z}}[f(\tilde{Z},S)|\tilde{Z}]^2\right]\right) &(74)
\end{aligned}$$

where (71) follows from expanding the expectation when the type of the sequence is known. 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 $n\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_1)n\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_2)$ while the number of choices is $n(n-1)$, yielding a probability equal to $\frac{n}{n-1}\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_1)\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_2)$. Similarly, when $z_1 = z_2 = z$, the number of such possibilities is $n\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_1)(n\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_2)-1)$, while the number of choices remains $n(n-1)$, yielding a probability equal to $\frac{1}{n-1}\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_1)(n\hat{\boldsymbol{p}}_{\boldsymbol{z}}(z_2)-1)$. Eq. (72) follows by rearranging the terms. Additionally, (73) follows by taking the expectation inside using Lemma 4. Combining (70) and (74) with (69) we get the result. ■

Corollary 2: With the assumptions of Lemma 4 we have

$$\mathrm{Var}\left[\sum_{i=1}^{n} f(Z_i,S_i)\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\right] = n\mathbb{E}_{\hat{\boldsymbol{p}}_{\boldsymbol{z}}}\left[\mathrm{Var}_{P_{S|Z}}[f(\tilde{Z},S)|\tilde{Z}]\right] \tag{75}$$

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

Proof:

$$\mathrm{Var}\left[\sum_{i=1}^{n} f(Z_i,S_i)\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\right] = \mathbb{E}\left[\left(\sum_{i=1}^{n} f(Z_i,S_i)\right)^2\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\right] - \mathbb{E}\left[\sum_{i=1}^{n} f(Z_i,S_i)\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\right]^2 \tag{76}$$

$$= n^2\mathbb{E}_{P_{S|Z}\times\hat{P}_{\boldsymbol{z}}}[f(\tilde{Z},S)]^2 + n\mathbb{E}_{P_{S|Z}\times\hat{P}_{\boldsymbol{z}}}[f(\tilde{Z},S)^2] - n\mathbb{E}_{\hat{P}_{\boldsymbol{z}}}\left[\mathbb{E}_{P_{S|Z}}[f(\tilde{Z},S)|\tilde{Z}]^2\right] - n^2\mathbb{E}_{P_{S|Z}\times\hat{P}_{\boldsymbol{z}}}[f(\tilde{Z},S)]^2 \tag{77}$$

$$= n\mathbb{E}_{P_{S|Z}\times\hat{P}_{\boldsymbol{z}}}[f(\tilde{Z},S)^2] - n\mathbb{E}_{\hat{P}_{\boldsymbol{z}}}\left[\mathbb{E}_{P_{S|Z}}[f(\tilde{Z},S)|\tilde{Z}]^2\right] \tag{78}$$

$$= n\mathbb{E}_{\hat{P}_{\boldsymbol{z}}}[\mathrm{Var}_{P_{S|Z}}[f(\tilde{Z},S)|\tilde{Z}]] \tag{79}$$

where (76) follows from the definition of variance, and (77) follows by directly using Lemmas 4 and 5. ∎

Lemma 6: Let $(Z_i,S_i), i=1,2,\ldots,n$ be i.i.d random variables, $\boldsymbol{z}=(Z_1,Z_2,\ldots,Z_n)$ and $\mathcal{A}\subset\mathcal{P}_{\mathcal{Z}}^n$ then

$$\mathbb{E}\left[\sum_{i=1}^{n} f(Z_i,S_i)\Big|\mathcal{A}\right] \geq n\min_{\hat{\boldsymbol{p}}_{\boldsymbol{z}}\in\mathcal{A}} \mathbb{E}_{P_{S|Z}\times\hat{\boldsymbol{p}}_{\boldsymbol{z}}}\left[f(\tilde{Z},S)\right] \tag{80}$$

Proof: We have

$$\mathbb{E}\left[\sum_{i=1}^{n} f(Z_i,S_i)\Big|\mathcal{A}\right] = \frac{1}{\mathbb{P}(\mathcal{A})}\mathbb{E}\left[\left(\sum_{i=1}^{n} f(Z_i,S_i)\right)\mathbb{1}\{\hat{\boldsymbol{p}}_{\boldsymbol{z}}\in\mathcal{A}\}\right] \tag{81}$$

$$\geq \min_{\hat{\boldsymbol{p}}_{\boldsymbol{z}}\in\mathcal{A}}\mathbb{E}\left[\left(\sum_{i=1}^{n} f(Z_i,S_i)\right)\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\right] \tag{82}$$

$$= n\min_{\hat{\boldsymbol{p}}_{\boldsymbol{z}}\in\mathcal{A}} \mathbb{E}_{P_{S|Z}\times\hat{\boldsymbol{p}}_{\boldsymbol{z}}}\left[f(\tilde{Z},S)\right] \tag{83}$$

Where (81) is by definition of conditional expectation, (83) is by using the Lemma 4. ∎

The analogous of the above lemma does not hold for the variance and we need an extra assumption on the boundedness of $\max_{z\in\mathcal{Z}}\left|\mathbb{E}_{P_{S|Z=z}}\left[f(z,S)\right]\right|$.

Lemma 7: Let $(Z_i,S_i), i=1,2,\ldots,n$ be i.i.d. random variables, $\boldsymbol{z}=(z_1,z_2,\ldots,z_n)$ and $\mathcal{A}\subset\mathcal{P}_{\mathcal{Z}}^n$. Then, we have

$$\mathrm{Var}\left[\sum_{i=1}^{n} f(Z_i,S_i)\Big|\mathcal{A}\right] \leq n\max_{\tilde{P}_Z\in\mathcal{A}}\mathbb{E}_{\tilde{P}_Z}\left[\mathrm{Var}_{P_{S|Z}}[f(\tilde{Z},S)|\tilde{Z}]\right] + 4n^2G_1G_2 \tag{84}$$

where $\tilde{Z}$ is a random variable with distribution $\tilde{P}_Z$. Moreover, G_1, G_2 are defined as

$$G_1 = \max_{z\in\mathcal{Z}}\left|\mathbb{E}_{P_{S|Z=z}}\left[f(z,S)\right]\right| \tag{85}$$

$$G_2 = \max_{\tilde{P}_Z,\bar{P}_Z\in\mathcal{A}}\Big(\mathbb{E}_{P_{S|Z}\times\tilde{P}_Z}[f(\tilde{Z},S)] - \mathbb{E}_{P_{S|Z}\times\bar{P}_Z}[f(\bar{Z},S)]\Big). \tag{86}$$

Proof:

$$\mathrm{Var}\left[\sum_{i=1}^{n} f(Z_i,S_i)\Big|\mathcal{A}\right] = \mathbb{E}\left[\left(\sum_{i=1}^{n} f(Z_i,S_i)\right)^2\Big|\mathcal{A}\right] - \mathbb{E}\left[\sum_{i=1}^{n} f(Z_i,S_i)\Big|\mathcal{A}\right]^2 \tag{87}$$

$$\leq \max_{\hat{\boldsymbol{p}}_{\boldsymbol{z}}\in\mathcal{A}}\mathbb{E}\left[\left(\sum_{i=1}^{n} f(Z_i,S_i)\right)^2\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\right] - \min_{\hat{\boldsymbol{p}}_{\boldsymbol{z}}\in\mathcal{A}}\mathbb{E}\left[\left(\sum_{i=1}^{n} f(Z_i,S_i)\right)\Big|\hat{\boldsymbol{p}}_{\boldsymbol{z}}\right]^2 \tag{88}$$

$$\leq n^2\mathbb{E}_{P_{S|Z}\times\tilde{P}_Z}[f(\tilde{Z},S)]^2 + n\mathbb{E}_{P_{S|Z}\times\tilde{P}_Z}[f(\tilde{Z},S)^2] - n\mathbb{E}_{\bar{P}_Z}\left[\mathbb{E}_{P_{S|Z}}[f(\bar{Z},S)|\bar{Z}]^2\right] - \\ - n^2\mathbb{E}_{P_{S|Z}\times\bar{P}_Z}[f(\bar{Z},S)]^2 \tag{89}$$

$$\leq n\mathbb{E}_{\tilde{P}_Z}\left[\mathrm{Var}_{P_{S|Z}}[f(\tilde{Z},S)|\tilde{Z}]\right] + n^2\Big(\mathbb{E}_{P_{S|Z}\times\tilde{P}_Z}[f(\tilde{Z},S)]^2 - \mathbb{E}_{P_{S|Z}\times\bar{P}_Z}[f(\bar{Z},S)]^2\Big) \tag{90}$$

where $\tilde{Z}$ and $\bar{Z}$ are random variables with distribution $\tilde{P}_Z$ and $\bar{P}_Z$ that correspond to the maximizing and minimizing types in (88), respectively. Additionally, for the second term of the right hand side of (90) we have

$$n^2\Big(\mathbb{E}_{P_{S|Z}\times\tilde{P}_Z}[f(\tilde{Z},S)]^2 - \mathbb{E}_{P_{S|Z}\times\bar{P}_Z}[f(\bar{Z},S)]^2\Big) \tag{91}$$
$$= n^2\Big(\mathbb{E}_{P_{S|Z}\times\tilde{P}_Z}[f(\tilde{Z},S)] - \mathbb{E}_{P_{S|Z}\times\bar{P}_Z}[f(\bar{Z},S)]\Big)\Big(\mathbb{E}_{P_{S|Z}\times\tilde{P}_Z}[f(\tilde{Z},S)] + \mathbb{E}_{P_{S|Z}\times\bar{P}_Z}[f(\bar{Z},S)]\Big) \tag{92}$$
$$\leq n^2(2G_1)(2G_2) \tag{93}$$
$$= 4n^2G_1G_2. \tag{94}$$

■

Appendix B
Proof of Theorem 2

In this section, we prove Theorem 2, i.e.,

$$P^q_{e,\max}(\mathcal{C}_n, W, n\epsilon) \geq (1-\gamma_n(\epsilon))P^{\max}_{\text{tce}}(\mathcal{C}_n, P_{\hat{Y}|X}). \tag{95}$$

where

$$\gamma_n(\epsilon) = 1 - \frac{1}{N(\varepsilon)}\Big(1 - \frac{2n\sigma_0^2 + 4n^2\kappa_1\kappa_2}{n^2(\delta-\epsilon)^2}\Big) \tag{96}$$
$$\sigma_0^2 = \max_{x\in\mathcal{X}} \text{Var}[q(x,Y)] \tag{97}$$
$$\kappa_1 = \max_{x_1,x_2\in\mathcal{X}} \big|\mathbb{E}_{P_{Y\hat{Y}X_1=x,X_2=x_2}}[q(x_1,Y) - q(x_2,Y)]\,\big| \tag{98}$$
$$\kappa_2 = \max_{\tilde{P}_{\hat{Y}X_1X_2},\bar{P}_{\hat{Y}X_1X_2}\in\mathcal{F}} \Big(\mathbb{E}_{P_{Y|\hat{Y}X_1}\times\tilde{P}_{\hat{Y}X_1X_2}}[q(x_1,Y) - q(x_2,Y)] - \mathbb{E}_{P_{Y|\hat{Y}X_1}\times\bar{P}_{\hat{Y}X_1X_2}}[q(x_1,Y) - q(x_2,Y)]\Big) \tag{99}$$

and the constant $N(\varepsilon)$ and set $\mathcal{F}$ are to be specified in the proof.

Proof: Without loss of generality assume that $\boldsymbol{x}_1$ is the codeword with maximum type conflict error on channel $P_{\hat{Y}|X}$. For every message $\ell = 2\ldots, M$, define the sets

$$\mathcal{A}_\ell = \{\boldsymbol{y} \,|\, q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}) + n\epsilon\} \tag{100}$$
$$\mathcal{B}_\ell = \{\boldsymbol{y} \,|\, \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}|\boldsymbol{x}_\ell} = \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}|\boldsymbol{x}_1}, \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_\ell\boldsymbol{x}_1} \in \mathcal{F}\} \tag{101}$$

Firstly we prove two lemmas which would be helpful in choosing of $N(\varepsilon)$ and set $\mathcal{F}$.

Lemma 8: We have

$$\kappa_2 \leq \max_{\tilde{P}_{\hat{Y}X_1X_2},\bar{P}_{\hat{Y}X_1X_2}\in\mathcal{F}} 4\cdot q_{\max}\cdot \|\tilde{P}_{\hat{Y}X_1X_2} - \bar{P}_{\hat{Y}X_1X_2}\|_1 \tag{102}$$

Where $q_{\max} = \max_{x\in\mathcal{X},y\in\mathcal{Y}} q(x,y)$

Proof: The proof directly follows from definition of κ_2 in (99) and triangle inequality. ■

Lemma 9: For every $\varepsilon > 0$ there are $N(\varepsilon)$ different probability distributions $\{P^1_{\hat{Y}X_1X_2}, P^2_{\hat{Y}X_1X_2}, \ldots, P^{N(\varepsilon)}_{YX_1X_2}\} \subset \mathcal{P}_{\mathcal{YXX}}$ such that every other distribution is in ε ball of one these distributions, i.e., for every $P_{YX_1X_2}$ there exists an index $1 \leq s \leq N(\varepsilon)$ such that

$$\|P_{YX_1X_2} - P^s_{YX_1X_2}\|_1 \leq \varepsilon \tag{103}$$

Proof: Since $\mathcal{P}_{\mathcal{YXX}}$ is a compact set under the L_1 distance, the result follows. ■

Then, we have

$$P^q_{e,\max}(\mathcal{C}_n, W, n\epsilon) = \mathbb{P}\Big[\bigcup_{m'=2}^{M} \mathcal{A}_{m'}\Big] \tag{104}$$
$$= \sum_{\ell=2}^{M} \mathbb{P}\Big[\bigcup_{m'=2}^{M} \mathcal{A}_{m'} | \mathcal{B}_\ell\Big]\mathbb{P}[\mathcal{B}_\ell] \tag{105}$$
$$\geq \sum_{\ell=2}^{M} \mathbb{P}[\mathcal{A}_\ell|\mathcal{B}_\ell]\mathbb{P}[\mathcal{B}_\ell] \tag{106}$$

where (104) follows from the definition or error probability, (105) is obtained by conditioning on $\mathcal{B}_\ell, \ell = 2, \dots, M$ and (106) is by using the following inequality

$$\mathbb{P}\bigg[\bigcup_{m'=2}^{M} \mathcal{A}_{m'}|\mathcal{B}_\ell\bigg] \geq \mathbb{P}[\mathcal{A}_\ell|\mathcal{B}_\ell]. \tag{107}$$

We proceed by lower-bounding $\mathbb{P}[\mathcal{A}_\ell|\mathcal{B}_\ell]$ as follows

$$\begin{aligned}
\mathbb{P}[\mathcal{A}_\ell|\mathcal{B}_\ell] &= \mathbb{P}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) \geq q^n(\boldsymbol{x}_1, \boldsymbol{y}) + n\epsilon|\mathcal{B}_\ell] && (108)\\
&= 1 - \mathbb{P}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) - q^n(\boldsymbol{x}_1, \boldsymbol{y}) < n\epsilon|\mathcal{B}_\ell] && (109)\\
&\geq 1 - \mathbb{P}\big[|q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) - q^n(\boldsymbol{x}_1, \boldsymbol{y}) - \mu| > \mu - n\epsilon|\mathcal{B}_\ell\big] && (110)\\
&\geq 1 - \frac{\sigma^2}{(\mu - n\epsilon)^2} && (111)
\end{aligned}$$

where $\mu = \mathbb{E}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) - q^n(\boldsymbol{x}_1, \boldsymbol{y})|\mathcal{B}_\ell]$ and $\sigma^2 = \mathrm{Var}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) - q^n(\boldsymbol{x}_1, \boldsymbol{y})|\mathcal{B}_\ell]$ and (111) is derived by from Chebychev's inequality.

To compute μ, σ in (111) we use Lemma 6. We choose $(Z_i, S_i), i = 1, 2, \dots, n$ in Lemma 6 to be $((\boldsymbol{x}_1(i), \boldsymbol{x}_\ell(i), \hat{Y}_i), Y_i), i = 1, 2, \dots, n$ where Z_i corresponds to triplet of $(\boldsymbol{x}_1(i), \boldsymbol{x}_\ell(i), \hat{Y}_i)$ and S_i corresponds to Y_i. Moreover, if we define $f(Z_i, S_i) = f(\boldsymbol{x}_1(i), \boldsymbol{x}_\ell(i), \hat{Y}_i, Y_i) = q^n(\boldsymbol{x}_\ell(i), Y_i) - q^n(\boldsymbol{x}_1(i), Y_i)$ we have

$$\begin{aligned}
q^n(\boldsymbol{x}_\ell, Y^n) - q^n(\boldsymbol{x}_1, Y^n) &= \sum_{i=1}^{n} q(\boldsymbol{x}_\ell(i), Y_i) - q(\boldsymbol{x}_1(i), Y_i) && (112)\\
&= \sum_{i=1}^{n} f(\boldsymbol{x}_1(i), \boldsymbol{x}_\ell(i), \hat{Y}_i, Y_i). && (113)
\end{aligned}$$

Therefore, from Lemma 6 we have

$$\begin{aligned}
\mathbb{E}[q^n(\boldsymbol{x}_\ell, \boldsymbol{y}) - q^n(\boldsymbol{x}_1, \boldsymbol{y})|\mathcal{B}_\ell] &= \mathbb{E}\bigg[\sum_{i=1}^{n} q(\boldsymbol{x}_\ell(i), \boldsymbol{y}(i)) - q(\boldsymbol{x}_1(i), \boldsymbol{y}(i))|\mathcal{B}_\ell\bigg] && (114)\\
&= \mathbb{E}\bigg[\sum_{i=1}^{n} f(\boldsymbol{x}_1(i), \boldsymbol{x}_\ell(i), \boldsymbol{y}(i), \hat{\boldsymbol{y}}(i))|\mathcal{B}_\ell\bigg] && (115)\\
&\geq n \min_{\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_\ell} \in \mathcal{B}_\ell} \mathbb{E}_{P_{Y|\hat{Y}\boldsymbol{x}_1\boldsymbol{x}_\ell} \times \hat{p}_{\hat{Y}\boldsymbol{x}_1\boldsymbol{x}_\ell}}[q(\tilde{X}_\ell, Y) - q(\tilde{X}_1, Y)] && (116)\\
&= n \min_{\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_\ell} \in \mathcal{B}_\ell} \mathbb{E}_{P_{Y|\hat{Y}\boldsymbol{x}_1} \times \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_\ell}}[q(\tilde{X}_\ell, Y) - q(\tilde{X}_1, Y)] && (117)\\
&\geq n \min_{X_2 : P_{\hat{Y}X_2} = P_{\hat{Y}X_1}} \mathbb{E}[q(X_2, Y) - q(X_1, Y)] && (118)\\
&\geq n\delta && (119)
\end{aligned}$$

where (116) follows from Lemma 6, (117) is by Y being independent from $\boldsymbol{x}_\ell$ given $\hat{Y}, \boldsymbol{x}_1$ and (118) follows from the definition of events $\mathcal{B}_\ell$, and (119) follows from the definition of set $\mathcal{M}^{\delta}_{\max}(d, P_X)$. As for σ^2, we use Corollary 2 in the same way. We have

$$\begin{aligned}
\mathrm{Var}[q(\boldsymbol{x}_\ell, \boldsymbol{y}) - q(\boldsymbol{x}_1, \boldsymbol{y})|\mathcal{B}_\ell] &= \mathrm{Var}\bigg[\sum_{i=1}^{n} q(\boldsymbol{x}_\ell(i), \boldsymbol{y}(i)) - q(\boldsymbol{x}_1(i), \boldsymbol{y}(i))|\mathcal{B}_\ell\bigg] && (120)\\
&\leq n \max_{\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_\ell} \in \mathcal{B}_\ell} \mathbb{E}_{\hat{\boldsymbol{p}}_{\hat{Y}\boldsymbol{x}_1\boldsymbol{x}_\ell}}\big[\mathrm{Var}_{P_{Y|\hat{Y}\boldsymbol{x}_1\boldsymbol{x}_\ell}}[q(\tilde{X}_\ell, Y) - q(\tilde{X}_1, Y)]\big] + 4n^2\kappa_1\kappa_2 && (121)\\
&= n \max_{\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_\ell} \in \mathcal{B}_\ell} \mathbb{E}_{\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}\boldsymbol{x}_1\boldsymbol{x}_\ell}}\big[\mathrm{Var}_{P_{Y|\hat{Y}\boldsymbol{x}_1}}[q(\tilde{X}_\ell, Y) - q(\tilde{X}_1, Y)]\big] + 4n^2\kappa_1\kappa_2 && (122)\\
&\leq 2n\sigma_0^2 + 4n^2\kappa_1\kappa_2 && (123)
\end{aligned}$$

Therefore, combining (119) and (123) we get

$$\mathbb{P}[\mathcal{A}_\ell|\mathcal{B}_\ell] \geq 1 - \frac{2n\sigma_0^2 + 4n^2\kappa_1\kappa_2}{n^2(\delta - \epsilon)^2} \tag{124}$$

Now recall that $\boldsymbol{x}_1$ is the codeword which has the maximum type conflict error over channel $P_{\hat{Y}|X}$. Moreover, by combining (124) and (106) we get the following result

$$P^q_{e,\max}(\mathcal{C}_n, W, n\epsilon) \geq \sum_{\ell=2}^{M} \mathbb{P}[\mathcal{A}_\ell|\mathcal{B}_\ell]\mathbb{P}[\mathcal{B}_\ell] \tag{125}$$

$$\geq \left(1 - \frac{2n\sigma_0^2 + 4n^2\kappa_1\kappa_2}{n^2(\delta-\epsilon)^2}\right)\sum_{\ell=2}^{M}\mathbb{P}[B_\ell] \tag{126}$$

The above analysis is valid for an arbitrary set $\mathcal{F}$. Now we need to choose the set $\mathcal{F}$ in such a way that $\mathbb{P}[\cup_{\ell=2}^M B_\ell] \geq \frac{1}{N(\varepsilon)} P^{\max}_{\rm tce}(\mathcal{C}_n, P_{\hat{Y}|X})$. If we choose $\mathcal{F} = \mathcal{P}_{\mathcal{Y}\mathcal{X}\mathcal{X}}$, then, the variance in (120) can be too large but $\mathbb{P}[\cup_{\ell=2}^M B_\ell] = P^{\max}_{\rm tce}$. In order to control the variance, instead, since the union of the $N(\varepsilon)$ ε-neighbourhoods of the distributions $P^s_{YX_1X_2}$ for $s = 1, \dots, N(\varepsilon)$, completely covers the space of joint distributions $\mathcal{P}_{\mathcal{Y}\mathcal{X}\mathcal{X}}$, we choose $\mathcal{F}$ to be the ε-neighbourhood of some distribution $P^{\bar{s}}_{YX_1X_2}$. We choose the distribution $P^{\bar{s}}_{YX_1X_2}$ in such a way that most of the joint types that yield a type-conflict error in the ε-neighbourhood of $P^{\bar{s}}_{YX_1X_2}$. This way, we can guarantee, that for such ε-neighbourhood of $P^{\bar{s}}_{YX_1X_2}$,

$$\sum_{\ell=2}^{M}\mathbb{P}[B_\ell] \geq \mathbb{P}[\cup_{\ell=2}^M B_\ell] \tag{127}$$

$$\geq \frac{1}{N(\varepsilon)} P^{\max}_{\rm tce}(\mathcal{C}_n, P_{\hat{Y}|X}) \tag{128}$$

As a result from (128) and (126) we get

$$P^q_{e,\max}(\mathcal{C}_n, W, n\epsilon) \geq \frac{1}{N(\varepsilon)}\left(1 - \frac{2n\sigma_0^2 + 4n^2\kappa_1\kappa_2}{n^2(\delta-\epsilon)^2}\right) P^{\max}_{\rm tce}(\mathcal{C}_n, P_{\hat{Y}|X}) \tag{129}$$

Now note that based on Lemma 8 we can choose ε small enough such that κ_2 is as small as we want, since the L_1 norm in (102) is always smaller than ε by Lemma 9. Then $N(\varepsilon)$ is the constant defined in lemma 9. Therefore $1 - \frac{2n\sigma_0^2 + 4n^2\kappa_1\kappa_2}{n^2(\delta-\epsilon)^2} \in (0,1)$. When ε is small, then $N(\varepsilon)$ increases, which is not a problem since $N(\varepsilon)\big(1 - \frac{2n\sigma_0^2 + 4n^2\kappa_1\kappa_2}{n^2(\delta-\epsilon)^2}\big)$ still remains in the interval $(0,1)$. ■

References

[1] J. Scarlett, A. Guillén i Fàbregas, A. Somekh-Baruch, and A. Martinez, *Information-theoretic foundations of mismatched decoding*. Found. Trends in Commun. and Inf. Theory, Now Publishers, 2020.

[2] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 41, pp. 35–43, Jan. 1995.

[3] J. Y. N. Hui, "Fundamental issues of multiple accessing," Ph.D. dissertation, Massachusetts Institute of Technology, 1983.

[4] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. 27, pp. 5–12, Jan. 1981.

[5] N. Merhav, G. Kaplan, A. Lapidoth, and S. S. Shitz, "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, pp. 1953–1967, Nov. 1994.

[6] 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, 1995.

[7] 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, pp. 5387–5395, Oct. 2015.

[8] E. Asadi Kangarshahi and A. Guillén i Fàbregas, "A single-letter upper bound to the mismatch capacity," *https://arxiv.org/abs/2004.01785, to appear IEEE Trans. Inf. Theory*, 2020.

[9] A. Somekh-Baruch, "A single-letter upper bound on the mismatch capacity via a multicasting approach," in *Proc. 2020 IEEE Inf. Theory Workshop, Riva del Garda, Italy (see also https://arxiv.org/abs/2007.14322)*, 2020.

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

[11] 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,tinyurl.com/22akp3ba