An Upper Bound to the Mismatch Capacity

Ehsan Asadi Kangarshahi
University of Cambridge
ea460@cam.ac.uk

Albert Guillén i Fàbregas
ICREA & Universitat Pompeu Fabra
University of Cambridge
guillen@ieee.org

***Abstract*—We derive a single-letter upper bound to the mismatched-decoding capacity for discrete memoryless channels. The bound is expressed as the mutual information of a transformation of the channel, such that a maximum-likelihood decoding error on the translated channel implies a mismatched-decoding error in the original channel. We show this bound recovers the binary-input binary-output mismatch capacity which is known to either be the channel capacity or zero. In addition, a strong converse is shown for this upper bound: if the rate exceeds the upper-bound, the probability of error tends to 1 exponentially when the block-length tends to infinity.**

I. Introduction and Preliminaries

We consider reliable communication over a discrete memoryless channel (DMC) W defined over input and output alphabets $\mathcal{X} = \{1, 2, \dots, J\}$ and $\mathcal{Y} = \{1, 2, \dots, K\}$. We denote the channel transition probability by $W(k|j)$ and define $\boldsymbol{W} \in \mathbb{R}^{J\times K}$ as the matrix defined by the channel $\boldsymbol{W}(j, k) = W(k|j)$. A codebook $\mathcal{C}_n$ is defined as a set of M sequences $\mathcal{C}_n = \big\{\boldsymbol{x}(1), \boldsymbol{x}(2), \dots, \boldsymbol{x}(M)\big\}$, where $\boldsymbol{x}(m) = \big(x_1(m), x_2(m), \dots, x_n(m)\big) \in \mathcal{X}^n$, for $m \in \{1, 2, \dots, M\}$. A message $m \in \{1, 2, \dots, M\}$ is chosen equiprobably and $\boldsymbol{x}(m)$ is sent over the channel. The channel produces a noisy observation $\boldsymbol{y} = (y_1, y_2, \dots, y_n) \in \mathcal{Y}^n$ 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 $\hat{m} \in \{1, 2, \dots, M\}$. The average and maximal error probabilities are respectively defined as $P_e(\mathcal{C}_n) = \mathbb{P}[\hat{m} \neq m]$ and $P_{e,\max}(\mathcal{C}_n) = \max_{i\in\{1,2,\dots,M\}} \mathbb{P}[\hat{m} \neq m | m = i]$. The decoder that minimizes the error probability is the maximum-likelihood (ML) decoder, that produces the message estimate $\hat{m}$ according to

$$\hat{m} = \operatorname*{arg\,max}_{i\in\{1,2,\dots,M\}} W^n\big(\boldsymbol{y}|\boldsymbol{x}(i)\big). \tag{1}$$

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 $C(W)$ or $C(\boldsymbol{W})$, is defined as the largest achievable rate.

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

$$\hat{m} = \operatorname*{arg\,max}_{i\in\{1,2,\dots,M\}} d\big(\boldsymbol{x}(i), \boldsymbol{y}\big), \tag{2}$$

This work was supported in part by the European Research Council under Grant 725411, and by the Spanish Ministry of Economy and Competitiveness under Grant TEC2016-78434-C3-1-R.

where $d\big(\boldsymbol{x}(i), \boldsymbol{y}\big) = \sum_{\ell=1}^n d\big(x_\ell(i), y_\ell\big)$ and $d : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is the decoding metric. We will refer to this decoder as d-decoder. When $d(x, y) = \log W(y|x)$, the decoder is ML, otherwise, for a general decoding metric d the decoder is said to be mismatched [1]–[4]. We define the metric matrix $\boldsymbol{D} \in \mathbb{R}^{J\times K}$ with entries $\boldsymbol{D}(j, k) = d(j, k)$. The average and maximal error probabilities of codebook $\mathcal{C}_n$ under d-decoding are respectively denoted by $P_e^d(\mathcal{C}_n)$ and $P_{e,\max}^d(\mathcal{C}_n)$. The mismatch capacity $C_d(W)$ or $C_d(\boldsymbol{W})$ is defined as supremum of all achievable rates with d-decoding.

Deriving a single-letter expression for the mismatch capacity is a long-standing open problem. Multiple achievability results have been reported in the literature [1]–[4] (see also [5] for an account of more recent progress). The only single-letter converse was [6], where it was claimed that for binary-input DMCs, the mismatch capacity was the achievable rate derived in [1], [3]. Reference [7] provided a counterexample to this converse invalidating its claim. Multiletter converse results were proposed in [8]. In particular, for DMCs, [8] shows that for rational decoding metrics, the probability of error cannot decay faster than $O(n^{-1})$ for rates above the achievable rate in [1], [3].

In this paper, we introduce a new single-letter upper bound to the mismatch capacity based on transforming the channel in such a way that a ML error on the transformed channel implies a mismatched-decoding error in the original channel.

A. Notation

The method of types [9] will be used extensively. We recall some of the basic definitions and introduce some notation. The type of a sequence $\boldsymbol{x} = (x_1, x_2, \dots, x_n) \in \mathcal{X}^n$ 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 $\mathcal{T}^n(\boldsymbol{p}_X)$ is set of all sequences in $\mathcal{X}^n$ with type $\boldsymbol{p}_X$, $\mathcal{T}^n(\boldsymbol{p}_X) = \{\boldsymbol{x} \in \mathcal{X}^n \,|\, \hat{\boldsymbol{p}}_{\boldsymbol{x}} = \boldsymbol{p}_X\}$.

The joint type of sequences $\boldsymbol{x} = (x_1, x_2, \dots, x_n) \in \mathcal{X}^n$ and $\boldsymbol{y} = (y_1, y_2, \dots, y_n) \in \mathcal{Y}^n$ is defined as a matrix representing their empirical distribution $\hat{\boldsymbol{p}}_{\boldsymbol{x}\boldsymbol{y}}(j, k) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{x_i = j, y_i = k\}$. The conditional type of $\boldsymbol{y}$ given $\boldsymbol{x}$ is the matrix

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

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

$\boldsymbol{x} \in \mathcal{T}^n(\boldsymbol{p}_X)$, the conditional type class $\mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})$ is defined as $\mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X}) = \{\boldsymbol{y} \in \mathcal{Y}^n \,|\, \hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}} = \boldsymbol{p}_{Y|X}\}$.

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 $k_1, k_2 \in \mathcal{Y}$,

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

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

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

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:

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

The l dimensional simplex is defined as $\Delta^l = \{\boldsymbol{x} \in \mathbb{R}^l \,|\, x_i \geq 0, \sum_{i=1}^{l} x_i = 1\}$. The mutual information $I(\boldsymbol{V}; \boldsymbol{p})$ between two random variables defined by a conditional probability mass matrix $\boldsymbol{V} \in \mathbb{R}^{J\times K}$ and a marginal $\boldsymbol{p} \in \Delta^J$ is defined as

$$I(\boldsymbol{V}; \boldsymbol{p}) = \sum_{j=1}^{J}\sum_{k=1}^{K} \boldsymbol{p}(j)\boldsymbol{V}(j,k) \log \frac{\boldsymbol{V}(j,k)}{\sum_{j'=1}^{J} \boldsymbol{p}(j')\boldsymbol{V}(j',k)}. \quad (7)$$

Definition 1: Let $P_{Y\hat{Y}|X}$ be a joint conditional distribution and define the set $\mathcal{S}(k_1, k_2) \triangleq \{i \in \mathcal{X} | i = \arg\max_{i' \in \mathcal{X}} \boldsymbol{D}(i', k_2) - \boldsymbol{D}(i', k_1)\}$. We say that $P_{Y\hat{Y}|X}$ is a maximal joint conditional distribution if for all $(j, k_1, k_2) \in \mathcal{X}\times\mathcal{Y}\times\mathcal{Y}$, $P_{Y\hat{Y}|X}(j, k_1, k_2) = 0$ if $j \notin \mathcal{S}(k_1, k_2)$. Moreover, if $\hat{\boldsymbol{p}}_{Y\hat{Y}|X} \in \mathcal{P}_n(\mathcal{Y}\hat{\mathcal{Y}}|\mathcal{X})$ satisfies the same condition, we call it a maximal joint conditional type.

For a given decoding metric matrix $\boldsymbol{D}$, we define the set of maximal joint distributions to be $\mathcal{M}_{\max}(\boldsymbol{D})$.

The above definition will become helpful when relating decoding errors in channel $P_{Y|X} = W$ under d-decoding to errors in channel $P_{\hat{Y}|X}$ under ML decoding.

Definition 2: Let $\mathcal{C}_n = \{\boldsymbol{x}(1), \cdots, \boldsymbol{x}(M)\}$ and m be the transmitted message. We say that the decoder makes a *type conflict error* for a given $\boldsymbol{y} \in \mathcal{Y}^n$ if there is at least one codeword $\boldsymbol{x}(i) \neq \boldsymbol{x}(m)$ such that that $\hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}(i)} = \hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}(m)}$.

If there is a type conflict error, every α-decoder makes an error, including ML and d-decoding; the converse is not true. With the same method developed in the paper, it can be shown that the type conflict error probability over channel W goes to 1 exponentially for $R > C(W)$; even with a genie-aided ML decoder knowing the exact conditional type $\hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}(m)}$, the error probability would still tend to 1 exponentially above capacity.

II. Main Result

In this section, we introduce the main result and discuss some of its properties.

Theorem 1: Let $\boldsymbol{W}, \boldsymbol{D}$ be channel and decoding metric matrices, respectively. We define $\bar{R}_d(\boldsymbol{W})$ as follows

$$\bar{R}_d(\boldsymbol{W}) = \max_{\boldsymbol{p}\in\Delta^J} \min_{\substack{P_{Y\hat{Y}|X}\in\mathcal{M}_{\max}(\boldsymbol{D}) \\ P_{Y|X}=W}} I(P_{\hat{Y}|X}; \boldsymbol{p}). \quad (8)$$

If $R > \bar{R}_d(\boldsymbol{W})$, $\exists N_0 \in \mathbb{N}$ and $E^d_{\text{sc}}(R) > 0$ such that for $n > N_0$, the error probability of codebook $\mathcal{C}_n$ of length n and $M \geq 2^{nR}$ codewords satisfies $P^d_{e,\max}(\mathcal{C}_n) \geq 1 - 2^{-nE^d_{\text{sc}}(R)}$.

Proof: The proof is developed over the next sections of the paper. The complete details can be found in [10]. The main idea behind proof of Theorem 1 is that of lower-bounding the error probability of a codebook $\mathcal{C}_n$ with d-decoding over channel W by that of the same codebook over a different channel V with ML decoding, with $V = P_{\hat{Y}|X}$ as per the theorem statement. We will be able to construct a graph $\mathcal{G}$ in the output space such that if ML decoding over V makes a type conflict error for some $\boldsymbol{y} \in \mathcal{Y}^n$, then, the d-decoder makes an error for some $\hat{\boldsymbol{y}} \in \mathcal{Y}^n$ connected to $\boldsymbol{y}$ in $\mathcal{G}$. ■

Theorem 1 implies that $C_d(W) \leq \bar{R}_d(\boldsymbol{W})$. Setting Y such that $P_{Y|X} = W$ and $\hat{Y} = Y$ makes $P_{YY|X}$ a maximal joint conditional distribution (Def. 1). As a result, $C_d(W) \leq C(P_{Y|X}) = C(W)$. In addition, it is implied in Theorem 1 that for any $P_{Y\hat{Y}|X} \in M_{\max}(\boldsymbol{D})$ such that $P_{Y|X} = W$,

$$\bar{R}_d(\boldsymbol{W}) \leq C(P_{\hat{Y}|X}). \quad (9)$$

This result is derived by using the min-max inequality:

$$\bar{R}_d(\boldsymbol{W}) = \max_{\boldsymbol{p}\in\Delta^J} \min_{\substack{P_{Y\hat{Y}|X}\in\mathcal{M}_{\max}(\boldsymbol{D}) \\ P_{Y|X}=W}} I(P_{\hat{Y}|X}; \boldsymbol{p}) \quad (10)$$

$$\leq \min_{\substack{P_{Y\hat{Y}|X}\in\mathcal{M}_{\max}(\boldsymbol{D}) \\ P_{Y|X}=W}} \max_{\boldsymbol{p}\in\Delta^J} I(P_{\hat{Y}|X}; \boldsymbol{p}) \quad (11)$$

$$= \min_{\substack{P_{Y\hat{Y}|X}\in\mathcal{M}_{\max}(\boldsymbol{D}) \\ P_{Y|X}=W}} C(P_{\hat{Y}|X}). \quad (12)$$

We thus get $C_d(W) \leq C(P_{\hat{Y}|X})$. Theorem 1 characterizes a family of bounds to the mismatch capacity, not only the minimum in (8)). The above property is helpful to construct bounds without necessarily performing the optimization. In addition, it can be shown that the optimization problems in (10) and (12) are convex (see [10] for details).

In the following, we discuss the applicability of our upper bound to two relevant cases. First, we show that our bound recovers known results on binary-input binary-output channels. Next, we show that our bound makes a non-trivial improvement over the channel-metric combination used in [7] to state the counterexample to Balakirsky's result [6].

Example 1 (Binary-input binary-output channels): Suppose that the channel and decoding metric matrices of binary-input binary-output channels are given by

$$\boldsymbol{W} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{and} \quad \boldsymbol{D} = \begin{bmatrix} \hat{a} & \hat{b} \\ \hat{c} & \hat{d} \end{bmatrix}. \quad (13)$$

Without loss of generality we assume $a + d \geq b + c$. We show the following known result [2]: if $\hat{a}+\hat{d} < \hat{b}+\hat{c}$ then $\bar{R}_d(W) = 0$. On the other hand, if $\hat{a}+\hat{d} \geq \hat{b}+\hat{c}$, then $\bar{R}_d(W) = C(W)$.

<u>*Case 1*</u>: $\hat{a} + \hat{d} < \hat{b} + \hat{c}$

We choose the joint conditional distribution in Table I.

It can be checked that indeed it is a valid joint conditional distribution for $0 \leq r_1 \leq a$ and $0 \leq r_2 \leq d$, and that

TABLE I
JOINT CONDITIONAL DISTRIBUTION $P_{Y\hat{Y}|X}$ FOR EXAMPLE 1

(j,k_1,k_2)	$P_{Y\hat{Y}|X}$	(j,k_1,k_2)	$P_{Y\hat{Y}|X}$
$(1,1,1)$	$a-r_1$	$(2,2,2)$	$d-r_2$
$(1,1,2)$	r_1	$(2,2,1)$	r_2
$(1,2,2)$	b	$(2,1,1)$	c
$(1,2,1)$	0	$(2,1,2)$	0

$\sum_{k_2} P_{Y\hat{Y}|X}(j,k_1,k_2) = P_{Y|X}(j,k_1) = W(k_1|j)$. In order to check its maximality, we first notice that for $k_1 = k_2$ we always have that $\boldsymbol{D}(i,k_2) - \boldsymbol{D}(i,k_1) = 0$ for all $i \in \mathcal{X}$, implying that $\mathcal{S}(k_1,k_2) = \{1,2\}$. Thus, since every $j \in \mathcal{X}$ is such that $j \in \mathcal{S}(k_1,k_2)$, the corresponding four entries can be nonzero. As for entry $(1,1,2)$ (resp. $(2,2,1)$), using the assumption $\hat{a}+\hat{d} < \hat{b}+\hat{c}$ we have that $\mathcal{S}(k_1,k_2) = \{1\}$ (resp . $\mathcal{S}(k_1,k_2) = \{2\}$), and thus they both can be nonzero. Since by assumption $\hat{a}+\hat{d} < \hat{b}+\hat{c}$, it can be checked that for entry $(2,1,2)$, $\mathcal{S}(k_1,k_2) = \{1\}$, and thus we must have $P_{Y\hat{Y}|X}(j,k_1,k_2) = 0$. Similarly for entry $(1,2,1)$, $\mathcal{S}(k_1,k_2) = \{2\}$. Marginalizing the above over Y gives

$$P_{\hat{Y}|X} = \begin{bmatrix} a-r_1 & b+r_1 \\ c+r_2 & d-r_2 \end{bmatrix}. \tag{14}$$

Without loss of generality assume that a is the largest element of $\boldsymbol{W}$. By setting $r_1 = r_2 = \frac{a-c}{2} = \frac{d-b}{2}$ we obtain

$$P_{\hat{Y}|X} = \begin{bmatrix} \frac{a+c}{2} & \frac{b+d}{2} \\ \frac{a+c}{2} & \frac{b+d}{2} \end{bmatrix}. \tag{15}$$

Since $C(P_{\hat{Y}|X}) = 0$, we have that $C_d(W) \le 0$.

Case 2: $\hat{a}+\hat{e} \ge \hat{b}+\hat{c}$

In [3] it is shown that the LM achievable rate is equal to $C(W)$. Therefore, our upper-bound also matches the achievable rate (see discussion before (9).)

Example 2: In this example we consider the channel and metric studied in [7] to show a counterexample to [6]

$$\boldsymbol{W} = \begin{bmatrix} 0.97 & 0.03 & 0 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} \text{ and } \boldsymbol{D} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0.5 & 1.36 \end{bmatrix}. \tag{16}$$

We choose the maximal $P_{Y\hat{Y}|X}$ in Table II such that $P_{Y|X} = W$. By marginalizing over Y we find that

$$P_{\hat{Y}|X} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}. \tag{17}$$

Using the above we have that $C(P_{\hat{Y}|X}) = 0.61$ bits/use, while the rate achieved in [7] is approximately 0.1991 bits/use; the capacity is $C(W) = 0.71$ bits/use.

TABLE II
NONZERO ENTRIES OF $P_{Y\hat{Y}|X}$ FOR EXAMPLE 2

(j,k_1,k_2)	$P_{Y\hat{Y}|X}$	(j,k_1,k_2)	$P_{Y\hat{Y}|X}$
$(1,1,1)$	0.5	$(2,1,1)$	0.1
$(1,1,2)$	0.47	$(2,2,2)$	0.1
$(1,2,2)$	0.03	$(2,3,3)$	0.8

In the above example, if we change $d(2,2)$ from 0.5 to 1, the same $P_{Y\hat{Y}|X}$ in Table II remains maximal and gives $C(P_{\hat{Y}|X}) = 0.61$ bits/use, matching the LM rate [3].

III. Graph Construction

In this section, we outline how to construct a graph between different conditional joint types.

Definition 3: Let $\mathcal{G} = \{\mathcal{V}_1, \mathcal{V}_2, \mathcal{E}\}$ be a regular bipartite graph with vertex sets $\mathcal{V}_1$ and $\mathcal{V}_2$, edge set $\mathcal{E}$ and degrees r_1 on vertex set $\mathcal{V}_1$ and r_2 on vertex set $\mathcal{V}_2$. For $\mathcal{B} \in \mathcal{V}_2$ we define the set of vertices in $\mathcal{V}_1$ connected to $\mathcal{B}$ as

$$\Psi_{21}(\mathcal{B}) = \big\{v \in \mathcal{V}_1 \,|\, \exists b \in \mathcal{B}; (b,v) \in \mathcal{E}\big\}. \tag{18}$$

Lemma 1: Suppose $\mathcal{G} = \{\mathcal{V}_1, \mathcal{V}_2, E\}$ is a regular bipartite graph with degrees $r_1 > 0, r_2 > 0$. Then, for any $\mathcal{B} \subset \mathcal{V}_2$ we have that

$$\frac{|\Psi_{21}(\mathcal{B})|}{|\mathcal{V}_1|} \ge \frac{|\mathcal{B}|}{|\mathcal{V}_2|}. \tag{19}$$

Our aim is to construct a graph between different two conditional type classes, in order to be able to relate type conflict errors of codebook $\mathcal{C}_n$ over channel V and errors of $\mathcal{C}_n$ over channel W under d-decoding. Suppose $\boldsymbol{p}_{Y\hat{Y}|X} \in \mathcal{P}_n(\mathcal{Y}\hat{\mathcal{Y}}|\mathcal{X})$ is an arbitrary joint conditional type. In this section we construct a graph between $\mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})$ and $\mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X})$.

Definition 4: The graph

$$\mathcal{G}_{\boldsymbol{x}}(\boldsymbol{p}_{Y\hat{Y}|X}) = \big\{\mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X}), \mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X}), \mathcal{E}\big\} \tag{20}$$

has the following edge set:

$$\mathcal{E} = \big\{(\boldsymbol{y},\hat{\boldsymbol{y}}) \,|\, \hat{\boldsymbol{p}}_{\boldsymbol{y}\hat{\boldsymbol{y}}|\boldsymbol{x}} = \boldsymbol{p}_{Y\hat{Y}|X}\big\}. \tag{21}$$

Lemma 2: The graph $\mathcal{G}_{\boldsymbol{x}}(\boldsymbol{p}_{Y\hat{Y}|X})$ is regular.

Proof: For a given $\boldsymbol{x} \in \mathcal{T}^n(\boldsymbol{p}_X)$, $|\mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})|$ is independent of the chosen $\boldsymbol{x} \in \mathcal{T}^n(\boldsymbol{p}_X)$, but dependent on $\boldsymbol{p}_X$. Similarly, for a given $\boldsymbol{y} \in \mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})$, $|\mathcal{T}^n_{\boldsymbol{yx}}(\boldsymbol{p}_{Y\hat{Y}|X})|$ is independent of the chosen $\boldsymbol{x}, \boldsymbol{y}$, but dependent on the joint type $\boldsymbol{p}_{XY}$. Therefore, the total number of edges that are connected to any given $\boldsymbol{y} \in \mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})$ is equal to $|\mathcal{T}^n_{\boldsymbol{yx}}(\boldsymbol{p}_{Y\hat{Y}|X})|$ (see (6)). This proves the left-regularity, i.e., for vertex set $\mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})$. The same argument holds for $\hat{\boldsymbol{y}} \in \mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X})$ and therefore the graph is regular. ■

As we show next, the combination of Lemmas 1 and 2 will prove to be helpful. Assume for a codeword $\boldsymbol{x}$ we find a set of type conflict errors $\mathcal{B} \subset \mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X})$. Then, the probability of an element $\hat{\boldsymbol{y}} \in \mathcal{B}$ being the output of the channel when conditional type $\boldsymbol{p}_{\hat{Y}|X}$ happens, is

$$\mathbb{P}\big[\hat{\boldsymbol{y}} \in \mathcal{B} \,|\, \hat{\boldsymbol{y}} \in \mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X}), \boldsymbol{x} \text{ is sent}\big] = \frac{|\mathcal{B}|}{|\mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X})|} \tag{22}$$

where equality holds because all members of $\mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X})$ are equally likely to appear at the output when $\boldsymbol{x}$ is sent. Now if the graph $\mathcal{G}_{\boldsymbol{x}}(\boldsymbol{p}_{Y\hat{Y}|X})$ is connecting a type conflict error to a d-decoder error, by Lemma 1 we show the existence of a set $\Psi_{21}(\mathcal{B}) \subset \mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})$ satisfying

$$\frac{|\Psi_{21}(\mathcal{B})|}{|\mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})|} \ge \frac{|\mathcal{B}|}{|\mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X})|}. \tag{23}$$

Now by the same argument as in (22) we have

$$\mathbb{P}\big[\boldsymbol{y} \in \Psi_{21}(\mathcal{B}) \,|\, \boldsymbol{y} \in \mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X}), \boldsymbol{x} \text{ is sent}\big] = \frac{|\Psi_{21}(\mathcal{B})|}{|\mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})|}. \quad (24)$$

Combining (24) and (23) we get

$$\mathbb{P}[\boldsymbol{y} \in \Psi_{21}(\mathcal{B}) \,|\, \boldsymbol{y} \in \mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X}), \boldsymbol{x} \text{ is sent}] \geq \mathbb{P}[\hat{\boldsymbol{y}} \in \mathcal{B} \,|\, \hat{\boldsymbol{y}} \in \mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X}), \boldsymbol{x} \text{ is sent}]. \quad (25)$$

As a result, we get a lower bound on the probability of error of d-decoder. In the next section, we prove that a graph constructed based on a maximal joint conditional type has the property of connecting type conflict errors to d-decoder errors.

IV. Connecting Errors

We next introduce a property of maximal joint conditional types and use it to relate type conflict and d-decoding errors.

Lemma 3: Let $\boldsymbol{p}_X \in \mathcal{P}_n(\mathcal{X})$, $\boldsymbol{x}, \hat{\boldsymbol{x}} \in \mathcal{T}^n(\boldsymbol{p}_X)$, and $\boldsymbol{p}_{Y\hat{Y}|X}$ be a maximal joint conditional type. If $\hat{\boldsymbol{y}} \in \mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X}) \cap \mathcal{T}^n_{\hat{\boldsymbol{x}}}(\boldsymbol{p}_{\hat{Y}|X})$ is connected to $\boldsymbol{y} \in \mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{Y|X})$ in $\mathcal{G}_{\boldsymbol{x}}(\boldsymbol{p}_{Y\hat{Y}|X})$ then,

$$d(\boldsymbol{x}, \boldsymbol{y}) \leq d(\hat{\boldsymbol{x}}, \boldsymbol{y}). \quad (26)$$

Proof: From the definition of type, for any $\bar{\boldsymbol{x}} \in \mathcal{X}^n$,

$$\hat{\boldsymbol{p}}_{\boldsymbol{y}\hat{\boldsymbol{y}}}(k_1, k_2) = \sum_j \hat{\boldsymbol{p}}_{\bar{\boldsymbol{x}}\boldsymbol{y}\hat{\boldsymbol{y}}}(j, k_1, k_2). \quad (27)$$

Now, if we use the above equation once by setting $\bar{\boldsymbol{x}} = \boldsymbol{x}$ and once by setting $\bar{\boldsymbol{x}} = \hat{\boldsymbol{x}}$ we get

$$\sum_j \hat{\boldsymbol{p}}_{\boldsymbol{x}\boldsymbol{y}\hat{\boldsymbol{y}}}(j, k_1, k_2) = \sum_j \hat{\boldsymbol{p}}_{\hat{\boldsymbol{x}}\boldsymbol{y}\hat{\boldsymbol{y}}}(j, k_1, k_2). \quad (28)$$

We continue by bounding $d(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) - d(\hat{\boldsymbol{x}}, \boldsymbol{y})$ as

$$\begin{aligned}
&d(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) - d(\hat{\boldsymbol{x}}, \boldsymbol{y}) \\
&= n \sum_{j,k_1,k_2} \hat{\boldsymbol{p}}_{\hat{\boldsymbol{x}}\boldsymbol{y}\hat{\boldsymbol{y}}}(j, k_1, k_2)\big(\boldsymbol{D}(j, k_2) - \boldsymbol{D}(j, k_1)\big) && (29)\\
&\leq n \sum_{k_1,k_2} \Big(\sum_j \hat{\boldsymbol{p}}_{\hat{\boldsymbol{x}}\boldsymbol{y}\hat{\boldsymbol{y}}}(j, k_1, k_2)\Big) \max_j \big(\boldsymbol{D}(j, k_2) - \boldsymbol{D}(j, k_1)\big) && (30)\\
&= n \sum_{k_1,k_2} \Big(\sum_j \hat{\boldsymbol{p}}_{\boldsymbol{x}\boldsymbol{y}\hat{\boldsymbol{y}}}(j, k_1, k_2)\Big) \max_j \big(\boldsymbol{D}(j, k_2) - \boldsymbol{D}(j, k_1)\big) && (31)\\
&= n \sum_{k_1,k_2} \sum_j \hat{\boldsymbol{p}}_{\boldsymbol{x}\boldsymbol{y}\hat{\boldsymbol{y}}}(j, k_1, k_2)\big(\boldsymbol{D}(j, k_2) - \boldsymbol{D}(j, k_1)\big) && (32)\\
&= d(\boldsymbol{x}, \hat{\boldsymbol{y}}) - d(\boldsymbol{x}, \boldsymbol{y}) && (33)
\end{aligned}$$

where (29) follows from the definition of metric and type, since for a joint type $\hat{\boldsymbol{p}}_{\boldsymbol{x}\boldsymbol{y}}$ we have that $d(\boldsymbol{x}, \boldsymbol{y}) = n \sum_{j,k} \hat{\boldsymbol{p}}_{\boldsymbol{x}\boldsymbol{y}}(j, k)\boldsymbol{D}(j, k)$, (30) follows from upper-bounding $(\boldsymbol{D}(j, k_2) - \boldsymbol{D}(j, k_1))$ by $\max_j(\boldsymbol{D}(j, k_2) - \boldsymbol{D}(j, k_1))$, (31) follows from (28), (32) follows from the maximality of $\boldsymbol{p}_{Y\hat{Y}|X}$ (see Definition (1)) and graph construction $\mathcal{G}_{\boldsymbol{x}}(\boldsymbol{p}_{Y\hat{Y}|X})$ (see Definition (4)) and (33) follows from the metric definition.

Now, using the fact that $\hat{\boldsymbol{y}} \in \mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X}) \cap \mathcal{T}^n_{\hat{\boldsymbol{x}}}(\boldsymbol{p}_{\hat{Y}|X})$ we get a type conflict error, i.e., $\hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}|\boldsymbol{x}} = \hat{\boldsymbol{p}}_{\hat{\boldsymbol{y}}|\hat{\boldsymbol{x}}}$. Thus, $d(\boldsymbol{x}, \hat{\boldsymbol{y}}) = d(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})$. Finally, combining with (33) we get the desired result $d(\boldsymbol{x}, \boldsymbol{y}) \leq d(\hat{\boldsymbol{x}}, \boldsymbol{y})$, i.e., a d-decoding error. ■

The condition of the Lemma says that if $\hat{\boldsymbol{y}} \in \mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{p}_{\hat{Y}|X}) \cap \mathcal{T}^n_{\hat{\boldsymbol{x}}}(\boldsymbol{p}_{\hat{Y}|X})$ and if $\boldsymbol{x}, \hat{\boldsymbol{x}} \in \mathcal{C}_n$, by observing $\hat{\boldsymbol{y}}$ when $\boldsymbol{x}$ is sent, there would be a type conflict error. Moreover, if such a $\hat{\boldsymbol{y}}$ is connected to $\boldsymbol{y}$ in $\mathcal{G}_{\boldsymbol{x}}(\boldsymbol{p}_{Y\hat{Y}|X})$, then, based on (26), by observing $\boldsymbol{y}$ when $\boldsymbol{x}$ is sent, the d-decoder makes an error.

In the next theorem, we show that if $P_{Y\hat{Y}|X}$ is a maximal joint conditional distribution and M is large enough, then we will find many type conflict errors over conditional types close to $V = P_{\hat{Y}|X}$. These are then linked to d-decoding errors over channel $W = P_{Y|X}$. The corresponding error probability is lower bounded using (25).

Definition 5: Let $\boldsymbol{W}$ be a channel and $\boldsymbol{p}_X$ an input type. We define the channel type neighborhood as the set of conditional types that are close to $\boldsymbol{W}$,

$$\mathcal{N}_{\epsilon,\boldsymbol{p}_X}(\boldsymbol{W}) = \big\{\boldsymbol{p}_{Y|X} \in \mathcal{P}_{\mathcal{Y}|\mathcal{X}} \,|\, |\boldsymbol{W} - \boldsymbol{p}_{Y|X}|_\infty \leq \epsilon\big\}. \quad (34)$$

Theorem 2: Let $\mathcal{C}_n$ be a codebook with M codewords and composition $\boldsymbol{p}_X$ with $\boldsymbol{p}_{\min} \triangleq \min_{j,\boldsymbol{p}_X(j)>0} \boldsymbol{p}_X(j)$. Let $P_{Y\hat{Y}|X}$ be a maximal joint conditional distribution such that and $P_{Y|X} = W, P_{\hat{Y}|X} = V$. Let $\epsilon \geq \frac{2|\mathcal{Y}|}{n\boldsymbol{p}_{\min}}$ and suppose $\mathcal{N}_{\epsilon,\boldsymbol{p}_X}(\boldsymbol{V}) = \{\boldsymbol{V}^1, \boldsymbol{V}^2 \cdots, \boldsymbol{V}^t\}$. Define $\boldsymbol{q}^i = \boldsymbol{p}_X^T \boldsymbol{V}^i$. If for some integer $a \geq 2$, for every $\boldsymbol{x} \in \mathcal{T}^n(\boldsymbol{p}_X)$ and for all $i \in \{1, \ldots, t\}$ we have that

$$M|\mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{V}^i)| \geq a^2(n+1)^{JK-1} \max_{1\leq i'\leq t} |\mathcal{T}^n(\boldsymbol{q}^{i'})|, \quad (35)$$

then, there exists a codeword $\boldsymbol{x}(m) \in \mathcal{C}_n$ such that

$$\mathbb{P}\Big[\hat{m} \neq m \,\big|\, \hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}(m)} \in \mathcal{N}_{\frac{\epsilon}{2},\boldsymbol{p}_X}(\boldsymbol{W}), \boldsymbol{x}(m) \text{ is sent}\Big] > 1 - \frac{1}{a}. \quad (36)$$

The above theorem (see [10] for details of the proof) gives us a sphere-packing type of bound. From the method of types approximations we know that $|\mathcal{T}^n_{\boldsymbol{x}}(\boldsymbol{V}^i)| \approx 2^{nH(\boldsymbol{V}^i|\boldsymbol{p}_X)} \approx 2^{nH(V|\boldsymbol{p}_X)}$ and that $|\mathcal{T}^n(\boldsymbol{q}^i)| \approx 2^{nH(\boldsymbol{q}^i)} \approx 2^{nH(\boldsymbol{p}_X^T V)}$. Therefore, the inequality (35) approximately implies that

$$2^{nR} 2^{nH(V|\boldsymbol{p}_X)} > 2^{nH(\boldsymbol{p}_X^T \boldsymbol{V})}, \quad (37)$$

or equivalently, $R > I(V; \boldsymbol{p}_X)$. The result of this theorem states the error probability of one of the messages is large under d-decoding.

The following result enables us to show $\hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}}$ is close to channel $\boldsymbol{W}$ for large block-lengths. This is necessary because the kind of result we have in (36).

Lemma 4: Let $\boldsymbol{x} \in \mathcal{T}^n(\boldsymbol{p}_X)$ be a codeword and $\boldsymbol{y}$ the output when $\boldsymbol{x}$ is sent. Then, $\forall \gamma > 0$ we have:

$$\mathbb{P}\big[|\boldsymbol{W} - \hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}}|_\infty \leq \gamma \,|\, \boldsymbol{x} \;\text{ is sent}\big] > 1 - |\mathcal{X}||\mathcal{Y}|e^{-2n\boldsymbol{p}_{\min}\gamma^2}. \quad (38)$$

Definition 6: Let $\mathcal{C}_n$ be a codebook. We say that $\hat{\mathcal{C}}_{\hat{n}}$ is a δ-reduction of $\mathcal{C}_n$ if there exist a sub-codebook $\hat{\mathcal{C}}_n$ of $\mathcal{C}_n$ of composition $\boldsymbol{p}_X$ that $\hat{\mathcal{C}}_{\hat{n}}$ is obtained by eliminating all symbols in the set $\mathcal{I} = \{j \in \mathcal{X} \,|\, \boldsymbol{p}_X(j) < \delta\}$ from $\hat{\mathcal{C}}_n$.

Lemma 5: Let $R > 0$ be a rate, then for any $\varepsilon > 0$ there exists a $\delta > 0$ independent of n such that for any codebook $\mathcal{C}_n$ of rate R there exists a δ-reduction constant composition codebook $\hat{\mathcal{C}}_{\hat{n}}$ with the following properties:

$$\hat{n} \geq \big(1 - (|\mathcal{X}| - 1)\delta\big)n \tag{39}$$
$$P^d_{e,\max}(\hat{\mathcal{C}}_{\hat{n}}) \leq P^d_{e,\max}(\mathcal{C}_n) \tag{40}$$
$$\frac{1}{\hat{n}}\log(|\hat{\mathcal{C}}_{\hat{n}}|) \geq \frac{1}{n}\log(|\mathcal{C}_n|) - \varepsilon + O\left(\frac{\log(n)}{n}\right). \tag{41}$$

V. Proof of the main theorem

In this part we prove the final part of Theorem 1 using the material developed in the previous sections. Assume $R = \bar{R}_d(\boldsymbol{W}) + \sigma$ for some $\sigma > 0$. Now, choose $\epsilon > 0$ small enough such that if $|\overline{\boldsymbol{V}} - \boldsymbol{V}|_\infty \leq \epsilon$ for conditional distribution $\boldsymbol{V}, \overline{\boldsymbol{V}}$, then for any distribution $\boldsymbol{p}$ on $\mathcal{X}$ we have that

$$|H(\overline{\boldsymbol{V}}|\boldsymbol{p}) - H(\boldsymbol{V}|\boldsymbol{p})| < \frac{\sigma}{4} \tag{42}$$
$$|H(\boldsymbol{p}^T\overline{\boldsymbol{V}}) - H(\boldsymbol{p}^T\boldsymbol{V})| < \frac{\sigma}{4}. \tag{43}$$

From Lemma 5 with $\varepsilon = \frac{\sigma}{4}$, for any codebook $\mathcal{C}_n$ with $M \geq 2^{nR}$ codewords, there exists a δ-reduction constant composition codebook $\hat{\mathcal{C}}_{\hat{n}}$ of length $\hat{n}$ and type $\hat{\boldsymbol{p}}_{\hat{n}}$ such that (39)–(41) are satisfied. Since the required δ to satisfy the above inequalities is independent of n, then choose N_0 large enough such that $\epsilon \geq \frac{2|\mathcal{Y}|}{N_0(1-(|\mathcal{X}|-1)\delta)\delta}$. Set $n > N_0$. Now choose a maximal joint conditional distribution $P_{Y\hat{Y}|X}$ such that $I(P_{\hat{Y}|X}; \hat{\boldsymbol{p}}_{\hat{n}}) \leq \bar{R}_d(\boldsymbol{W})$ and let $V = P_{\hat{Y}|X}$. Such a minimizing $P_{Y\hat{Y}|X}$ always exists because the domain of the minimization in (8) $\mathcal{M}_{\max}(\boldsymbol{D}) \cap \{P_{Y\hat{Y}|X} | P_{Y|X} = W\}$ is a compact set and the function $I(\boldsymbol{V}; \boldsymbol{p})$ is continuous. Now, for any conditional distributions $\widehat{\boldsymbol{V}}$ such that $|\widehat{\boldsymbol{V}} - \boldsymbol{V}|_\infty \leq \epsilon$

$$\Big|\max_{\overline{\boldsymbol{V}} \in \mathcal{N}_{\epsilon, \hat{\boldsymbol{p}}_n}(\boldsymbol{V})} H(\hat{\boldsymbol{p}}^T_{\hat{n}}\overline{\boldsymbol{V}}) - H(\widehat{\boldsymbol{V}}|\hat{\boldsymbol{p}}_{\hat{n}})\Big| \tag{44}$$
$$\leq |H(\hat{\boldsymbol{p}}^T_{\hat{n}}\boldsymbol{V}) - H(\boldsymbol{V}|\hat{\boldsymbol{p}}_{\hat{n}})| + \frac{\sigma}{2} \tag{45}$$
$$= I(\boldsymbol{V}; \hat{\boldsymbol{p}}_{\hat{n}}) + \frac{\sigma}{2} \tag{46}$$

where (45) follows from (42) and (43).

Now suppose $\mathcal{N}_{\epsilon, \hat{\boldsymbol{p}}_{\hat{n}}}(\boldsymbol{V}) = \{\boldsymbol{V}^1, \boldsymbol{V}^2, \cdots, \boldsymbol{V}^t\}$ and $\boldsymbol{q}^i = \hat{\boldsymbol{p}}^T_{\hat{n}}\boldsymbol{V}^i$. Now for any $1 \leq i \leq t$

$$\frac{1}{\hat{n}}\log\frac{\max_{1\leq s\leq t}|\mathcal{T}^{\hat{n}}(\boldsymbol{q}^s)|}{|\mathcal{T}^{\hat{n}}_{\boldsymbol{x}}(\boldsymbol{V}^i)|} = \frac{1}{\hat{n}}\log\frac{2^{\hat{n}\left(H(\boldsymbol{q}^{i'}) + O\left(\frac{\log(\hat{n})}{\hat{n}}\right)\right)}}{2^{\hat{n}\left(H(V^i|\hat{\boldsymbol{p}}_{\hat{n}}) + O\left(\frac{\log(\hat{n})}{\hat{n}}\right)\right)}} \tag{47}$$
$$\leq I(\boldsymbol{V}; \hat{\boldsymbol{p}}_{\hat{n}}) + \frac{\sigma}{2} + O\left(\frac{\log(\hat{n})}{\hat{n}}\right) \tag{48}$$

where $i' = \arg\max_{1\leq s\leq t}|\mathcal{T}^{\hat{n}}(\boldsymbol{q}^s)|$ and (48) follows form (46). Now, for $n > N_0$ we have from (41) with $\varepsilon = \frac{\sigma}{4}$, (48) and the condition $I(P_{\hat{Y}|X}; \hat{\boldsymbol{p}}_{\hat{n}}) \leq \bar{R}_d(\boldsymbol{W})$ that

$$|\hat{\mathcal{C}}_{\hat{n}}|\frac{|\mathcal{T}^{\hat{n}}_{\boldsymbol{x}}(\boldsymbol{V}^i)|}{\max_{1\leq s\leq t}|\mathcal{T}^{\hat{n}}(\boldsymbol{q}^s)|} \geq 2^{\hat{n}\left(R - \frac{\sigma}{4} - I(\boldsymbol{V}; \hat{\boldsymbol{p}}_{\hat{n}}) - \frac{\sigma}{2} + O\left(\frac{\log(\hat{n})}{\hat{n}}\right)\right)}. \tag{49}$$

As a result,

$$|\hat{\mathcal{C}}_{\hat{n}}||\mathcal{T}^{\hat{n}}_{\boldsymbol{x}}(\boldsymbol{V}^i)| \geq 2^{\hat{n}\left(\frac{\sigma}{4} + O\left(\frac{\log(\hat{n})}{\hat{n}}\right)\right)}\max_{1\leq s\leq t}|\mathcal{T}^{\hat{n}}(\boldsymbol{q}^s)|. \tag{50}$$

Setting $a = \left\lfloor \frac{2^{\frac{1}{2}\hat{n}\left(\frac{\sigma}{4} + O\left(\frac{\log(\hat{n})}{\hat{n}}\right)\right)}}{(\hat{n}+1)^{\frac{1}{2}(JK-1)}} \right\rfloor$ validates the conditions of Theorem 2. As a result, there exists $\boldsymbol{x}(m) \in \hat{\mathcal{C}}_{\hat{n}}$ such that

$$\mathbb{P}\Big[\hat{m} \neq m | \hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}(m)} \in \mathcal{N}_{\frac{\epsilon}{2}, \hat{\boldsymbol{p}}_{\hat{n}}}, \boldsymbol{x}(m) \text{ is sent}\Big] > 1 - \frac{1}{a}. \tag{51}$$

Choosing N_1 such that if $n > N_1$ is large enough, we bound

$$a > \frac{1}{2} \cdot \frac{2^{\frac{1}{2}\hat{n}\left(\frac{\sigma}{4} + O\left(\frac{\log(\hat{n})}{\hat{n}}\right)\right)}}{(\hat{n}+1)^{\frac{1}{2}(JK-1)}} \tag{52}$$
$$\geq 2^{\frac{1}{2}\hat{n}\left(\frac{\sigma}{4} + O\left(\frac{\log(\hat{n})}{\hat{n}}\right) - (JK-1)\frac{\log(\hat{n}+1)}{\hat{n}} - \frac{\log(2)}{\hat{n}}\right)} \tag{53}$$

Finally, we write,

$$P^d_{e,\max}(\mathcal{C}_n) \geq P^d_{e,\max}(\hat{\mathcal{C}}_{\hat{n}}) \tag{54}$$
$$\geq \mathbb{P}[\hat{m} \neq m \,|\, \boldsymbol{x}(m) \text{ is sent}] \tag{55}$$
$$\geq \mathbb{P}[\hat{m} \neq m \,|\, \hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}(m)} \in \mathcal{N}_{\frac{\epsilon}{2}, \hat{\boldsymbol{p}}_{\hat{n}}}(\boldsymbol{W}), \boldsymbol{x}(m) \text{ is sent}]$$
$$\cdot\, \mathbb{P}[\hat{\boldsymbol{p}}_{\boldsymbol{y}|\boldsymbol{x}(m)} \in \mathcal{N}_{\frac{\epsilon}{2}, \hat{\boldsymbol{p}}_{\hat{n}}}(\boldsymbol{W}), \boldsymbol{x}(m) \text{ is sent}] \tag{56}$$
$$\geq \left(1 - \frac{1}{a}\right)\left(1 - |\mathcal{X}||\mathcal{Y}|2^{-2\hat{n}\delta\frac{\epsilon^2}{4}}\right) \tag{57}$$
$$\geq 1 - 2^{-\hat{n}E^d_{\mathrm{sc}}(R)} \tag{58}$$

where $E^d_{\mathrm{sc}}(R) \triangleq \min\left\{\frac{\delta\epsilon^2}{2} - \log\frac{|\mathcal{X}||\mathcal{Y}|}{\hat{n}}, \frac{1}{2}\left(\frac{\sigma}{4} + O\left(\frac{\log(\hat{n})}{\hat{n}}\right)\right)\right\}$. Setting n larger than $\max\{N_0, N_1\}$ we get the desired result.

References

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

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

[4] 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.

[5] J. Scarlett, "Reliable communication under mismatched decoding," Ph.D. dissertation, Ph. D. dissertation, University of Cambridge, 2014, [Online: http://itc.upf.edu/biblio/1061, 2014.

[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] A. Somekh-Baruch, "Converse theorems for the dmc with mismatched decoding," *IEEE Trans. Inf. Theory*, vol. 64, pp. 6196–6207, Sept. 2018.

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

[10] E. Asadi Kangarshahi and A. Guillén i Fàbregas, "An upper bound to the mismatch capacity," in prepration.