On Coding Schemes for Channels with Mismatched Decoding

Anelia Somekh-Baruch
Faculty of Engineering, Bar-Ilan University
Ramat-Gan, Israel
Email: anelia.somekhbaruch@gmail.com

***Abstract*—The problem of mismatched decoding for discrete memoryless channels is addressed. A mismatched cognitive multiple-access channel is introduced, and an inner bound on its capacity region is derived using two alternative encoding methods: superposition coding and random binning. The inner bounds are derived by analyzing the average error probability of the code ensemble for both methods and by a tight characterization of the resulting error exponents. Random coding converse theorems are also derived. A comparison of the achievable regions shows that in the matched case, random binning performs as well as superposition coding, i.e., the region achievable by random binning is equal to the capacity region. The achievability results are further specialized by investigating a cognitive multiple access channel whose achievable sum-rate serves as a lower bound on the single-user channel's capacity. In certain cases, for given auxiliary random variables this bound strictly improves on the achievable rate derived by Lapidoth.**

I. Introduction

The mismatch capacity is the highest achievable rate using a given decoding rule. The mismatch capacity reflects a situation in which due to inaccurate knowledge of the channel, or other practical limitations, the receiver is constrained to use a possibly suboptimal decoder. This paper focuses on mismatched decoders that are defined by a mapping q, which for convenience will be referred to as "metric" from the product of the channel input and output alphabets to the reals. The decoding rule maximizes, among all the codewords, the accumulated sum of metrics between the channel output sequence and the codeword.

Mismatched decoding has been studied extensively for discrete memoryless channels (DMCs). A random coding lower bound on the mismatched capacity was derived by Csiszár and Körner and by Hui [1], [2]. Csiszár and Narayan [3] showed that the random coding bound is not tight and that the positivity of the random coding lower-bound is a necessary condition for a positive mismatched capacity. A converse theorem for the mismatched binary-input DMC was proved in [4], but in general, the problem of determining the mismatch capacity of the DMC remains open.

Lapidoth [5] introduced the mismatched multiple access channel (MAC) and derived an inner bound on its capacity region. The study of the MAC case led to an improved lower bound on the mismatch capacity of the single-user DMC by considering the maximal sum-rate of an appropriately chosen mismatched MAC whose codebook is obtained by expurgating codewords from the product of the codebooks of the two users.

In [1], an error exponent for random coding with fixed composition codes and mismatched decoding was established using a graph decomposition theorem. In a recent work, Scarlett and Guillén i Fàbregas [6] characterized the achievable error exponents obtained by a constant-composition random coding scheme for the mismatched MAC. For other related works and extensions see [7]–[13] and references therein.

This paper introduces the *cognitive* mismatched two-user MAC. Encoder 1 shares the message it wishes to transmit with (the cognitive) Encoder 2, and the latter transmits an additional message to the same receiver. Two achievability schemes with a mismatched decoder are presented. The first scheme is based on superposition coding and the second uses random binning. The achievable regions are compared, and an example is shown in which the achievable region obtained by random binning is strictly larger than the rate-region achieved by superposition coding. In general it seems that neither achievable region dominates the other, and conditions are shown under which each scheme is guaranteed to perform at least as well as the other. As a special case it is shown that in the matched case, where superposition coding is known to be capacity-achieving, so is binning. It is also shown that the region achievable by binning contains the mismatched non-cognitive MAC achievable region [5]. In certain cases, for fixed auxiliary random variables cardinalities, the suggested schemes serve to derive an improved achievable rate for the mismatched single-user channel.

The outline of this paper is as follows. Section II provides some notation and background. Section III introduces the mismatched cognitive MAC and presents the achievable regions and error exponents. The concluding section IV discusses the single-user mismatched DMC. We note that due to space limitations, the proofs are omitted and can be found in [14].

II. Notation and Preliminaries

Throughout this paper, scalar random variables are denoted by capital letters, and their sample values and alphabets are denoted by the respective lower case letters and calligraphic letters, respectively, e.g. X, x, and $\mathcal{X}$. A similar convention applies to random n-vectors which are denoted similarly with boldface font. To emphasize the dependence of an information-theoretic quantity on a certain underlying probability distribution, say μ, it is subscripted by μ, i.e., with notations such as $H_\mu(X)$, $H_\mu(X|Y)$, $I_\mu(X;Y)$, etc. The divergence between two probability measures μ and p is denoted by $D(\mu\|p)$. The expectation operator w.r.t. the distribution μ is denoted by $\boldsymbol{E}_\mu\{\cdot\}$. The cardinality of a finite set A is denoted by $|A|$. The indicator function of an event $\mathcal{E}$ is denoted by $1\{\mathcal{E}\}$.

Let $\mathcal{P}(\mathcal{X})$ denote the set of all probability measures on $\mathcal{X}$. For a given sequence $\boldsymbol{y} \in \mathcal{Y}^n$, $\mathcal{Y}$ being a finite alphabet, $\hat{P}_{\boldsymbol{y}}$ is the vector $\{\hat{P}_{\boldsymbol{y}}(y), y \in \mathcal{Y}\}$, where $\hat{P}_{\boldsymbol{y}}(y)$ is the relative frequency of the symbol y in the vector $\boldsymbol{y}$. The type-class of $\boldsymbol{x}$ is the set of $\boldsymbol{x}' \in \mathcal{X}^n$ such that $\hat{P}_{\boldsymbol{x}'} = \hat{P}_{\boldsymbol{x}}$, which is denoted $T(\hat{P}_{\boldsymbol{x}})$. The conditional type-class of $\boldsymbol{y}$ given $\boldsymbol{x}$ is the set of $\tilde{\boldsymbol{y}}$'s such that $\hat{P}_{\boldsymbol{x},\tilde{\boldsymbol{y}}} = \hat{P}_{\boldsymbol{x},\boldsymbol{y}}$, which is denoted $T(\hat{P}_{\boldsymbol{x},\boldsymbol{y}}|\boldsymbol{x})$ with a little abuse of notation. The set of empirical measures of order n on alphabet $\mathcal{X}$ is denoted $\mathcal{P}_n(\mathcal{X})$. For two sequences of positive numbers, $\{a_n\}$ and $\{b_n\}$, the notation $a_n \doteq b_n$ means that $\frac{1}{n}\ln\frac{a_n}{b_n} \to 0$ as $n \to \infty$. Another notation is that for a real number x, $|x|^+ = \max\{0, x\}$. Throughout this paper logarithms are taken to base 2.

Consider a DMC with a finite input alphabet $\mathcal{X}$ and finite output alphabet $\mathcal{Y}$, which is governed by the conditional p.m.f. $P_{Y|X}$, i.e., $P(y_i|x_1,...,x_i,y_1,...,y_{i-1}) = P_{Y|X}(y_i|x_i)$, $i = 1,2,...,n,$. A rate-R block-code of length n consists of 2^{nR} n-vectors $\boldsymbol{x}(m)$, $m = 1,2,...,2^{nR}$, which represent 2^{nR} different messages, i.e., it is defined by the encoding function

$$f_n : \{1,...,2^{nR}\} \to \mathcal{X}^n. \tag{1}$$

It is assumed that all possible messages are equiprobable.

A mismatched decoder for the channel is defined by a mapping $q_n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathbb{R}$, where the decoder declares that message i was transmitted iff

$$q_n(\boldsymbol{x}(i), \boldsymbol{y}) > q_n(\boldsymbol{x}(j), \boldsymbol{y}), \forall j \neq i, \tag{2}$$

and if no such i exists, an error is declared. The results in this paper refer to additive decoding functions, i.e., $q_n(x^n, y^n) = \frac{1}{n}\sum_{i=1}^n q(x_i, y_i)$, where q is a mapping from $\mathcal{X} \times \mathcal{Y}$ to $\mathbb{R}$.

A rate R is said to be achievable for the channel $P_{Y|X}$ with a decoding metric q if there exists a sequence of codebooks $\mathcal{C}_n$, $n \geq 1$ of rate R such that the average probability of error incurred by the decoder q_n applied to the codebook $\mathcal{C}_n$ and $\boldsymbol{Y}$ vanishes as n tends to infinity. The capacity of the channel with decoding metric q is the supremum of all achievable rates.

The notion of mismatched decoding can be extended to a MAC $P_{Y|X_1,X_2}$ with codebooks $\mathcal{C}_{n,1} = \{\boldsymbol{x}_1(i)\}, i = 1,....,2^{nR_1}$, $\mathcal{C}_{n,2} = \{\boldsymbol{x}_2(j)\}, j = 1,....,2^{nR_2}$. A mismatched decoder for a MAC is defined by the mapping $q_n : \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}^n \to \mathbb{R}$, where similar to the single-user's case, the decoder outputs the messages (i,j) iff for all $(i',j') \neq (i,j)$

$$q_n(\boldsymbol{x}_1(i), \boldsymbol{x}_2(j), \boldsymbol{y}) > q_n(\boldsymbol{x}_1(i'), \boldsymbol{x}_2(j'), \boldsymbol{y}). \tag{3}$$

The focus here is on additive decoding functions, i.e.,

$$q_n(x_1^n, x_2^n, y^n) = \frac{1}{n}\sum_{i=1}^n q(x_{1,i}, x_{2,i}, y_i), \tag{4}$$

where q is a mapping from $\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$ to $\mathbb{R}$. The achievable rate-region of the MAC $P_{Y|X_1,X_2}$ with decoding metric q is the closure of the set of rate-pairs (R_1, R_2) for which there exists a sequence of codebooks $\mathcal{C}_{n,1}, \mathcal{C}_{n,2}$, $n \geq 1$ of rates R_1 and R_2, respectively, such that the average probability of error that is incurred by the decoder q_n when applied to the codebooks $\mathcal{C}_{n,1}, \mathcal{C}_{n,2}$ and $\boldsymbol{Y}$ vanishes as n tends to infinity.

The best known inner bound on the capacity region of the mismatched (non-cognitive) MAC [5] is given by $\mathcal{R}_{LM}$ where

$$\begin{aligned}
\mathcal{R}_{LM} = &\text{closure of the CH of} \bigcup_{P_{X_1},P_{X_2}} \{(R_1,R_2) :\\
&R_1 < \min_{f \in \mathcal{D}_{(1)}} I_f(X_1;Y|X_2) + I_f(X_1;X_2)\\
&R_2 < \min_{f \in \mathcal{D}_{(2)}} I_f(X_2;Y|X_1) + I_f(X_1;X_2)\\
&R_1 + R_2 < \min_{f \in \mathcal{D}_{(0)}} I_f(X_1,X_2;Y) + I_f(X_1;X_2)\},
\end{aligned} \tag{5}$$

where CH stands for convex hull,

$$\begin{aligned}
\mathcal{D}_{(1)} = \{&f \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_1 \times \mathcal{Y}) : f_{X_1} = P_{X_1},\\
&f_{X_2,Y} = P_{X_2,Y}, \boldsymbol{E}_f(q) \geq \boldsymbol{E}_P(q)\}\\
\mathcal{D}_{(2)} = \{&f \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_1 \times \mathcal{Y}) : f_{X_2} = P_{X_2},\\
&f_{X_1,Y} = P_{X_1,Y}, \boldsymbol{E}_f(q) \geq \boldsymbol{E}_P(q)\}\\
\mathcal{D}_{(0)} = \{&f \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_1 \times \mathcal{Y}) : f_{X_1} = P_{X_1},\\
&f_{X_2} = P_{X_2}, f_Y = P_Y, \boldsymbol{E}_f(q) \geq \boldsymbol{E}_P(q),\\
&I_f(X_1;Y) \leq R_1, I_f(X_2;Y) \leq R_2\}.
\end{aligned} \tag{6}$$

and where $P_{X_1,X_2,Y} = P_{X_1} \times P_{X_2} \times P_{Y|X_1,X_2}$.

III. The Mismatched Cognitive MAC

The two-user discrete memoryless cognitive MAC is defined by the input alphabets $\mathcal{X}_1$, $\mathcal{X}_2$, output alphabet $\mathcal{Y}$ and conditional transition probability $P_{Y|X_1,X_2}$. A block-code of length n for the channel is defined by the two encoding mappings

$$\begin{aligned}
&f_{1,n} : \{1,...,2^{nR_1}\} \to \mathcal{X}_1^n\\
&f_{2,n} : \{1,...,2^{nR_1}\} \times \{1,...,2^{nR_2}\} \to \mathcal{X}_2^n,
\end{aligned} \tag{7}$$

resulting in two codebooks $\{\boldsymbol{x}_\ell(i)\}, i = 1,....,2^{nR_\ell}$, $\ell = 1,2$. A mismatched decoder for the cognitive MAC is defined by a mapping of the form (4) where the decoder outputs the message (i,j) iff for all $(i',j') \neq (i,j)$

$$q_n(\boldsymbol{x}_1(i), \boldsymbol{x}_2(i,j), \boldsymbol{y}) > q_n(\boldsymbol{x}_1(i'), \boldsymbol{x}_2(i',j'), \boldsymbol{y}). \tag{8}$$

The capacity region of the cognitive mismatched MAC is defined similarly to that of the mismatched MAC.

Denote by W_1, W_2 the random messages, and the corresponding outputs of the decoder by $\hat{W}_1, \hat{W}_2$. It is said that $E \geq 0$ is an achievable error exponent for the MAC if there exists a sequence of codebooks $\mathcal{C}_{n,1}, \mathcal{C}_{n,2}$, $n \geq 1$ of rates R_1 and R_2, respectively, such that the average probability of error, $\bar{P}_{e,n} = \Pr\{(\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)\}$, that is incurred by the decoder q_n when applied to codebooks $\mathcal{C}_{n,1}, \mathcal{C}_{n,2}$ and the channel output satisfies $\liminf_{n\to\infty} -\frac{1}{n}\log \bar{P}_{e,n} \geq E$.

Two achievability schemes tailored for the mismatched cognitive MAC are presented next. The first encoding scheme is based on constant composition superposition coding, and the second on constant composition random binning.

Codebook Generation of User 1: (for both superposition coding and random binning) Fix a distribution $P_{X_1,X_2} \in \mathcal{P}_n(\mathcal{X}_1 \times \mathcal{X}_2)$. User 1 draws 2^{nR_1} codewords $\{\boldsymbol{x}_1(i)\}, i = 1, ..., 2^{nR_1}$ independently, each uniformly over $T(P_{X_1})$.

Codebook Generation of User 2 and Encoding:

• **Superposition coding**: For each $\boldsymbol{x}_1(i)$, user 2 draws 2^{nR_2} codewords $\boldsymbol{x}_2(i,j), j = 1, ..., 2^{nR_2}$ conditionally independent given $\boldsymbol{x}_1(i)$ uniformly over $T(P_{X_1,X_2}|\boldsymbol{x}_1(i))$. To transmit the messages (m_1, m_2), encoder 1 transmits $\boldsymbol{x}_1(m_1)$ and encoder 2 transmits $\boldsymbol{x}_2(m_1, m_2)$.

• **Random binning:** Let $\gamma = I_P(X_1; X_2) + \epsilon$ for an arbitrarily small $\epsilon > 0$. User 2 draws $2^{n(R_2+\gamma)}$ codewords independently, each uniformly over $T(P_{X_2})$ and partitions them into 2^{nR_2} bins, i.e., $\{\boldsymbol{x}_2[k,j]\}$, $k = 1, ..., 2^{n\gamma}$, $j = 1, ..., 2^{nR_2}$. To transmit message m_1, encoder 1 transmits $\boldsymbol{x}_1(m_1)$. To transmit message m_2, encoder 2 looks for k such that $(\boldsymbol{x}_1(m_1), \boldsymbol{x}_2[k, m_2]) \in T(P_{X_1,X_2})$. If more than one such k exists, the encoder chooses one of them arbitrarily, otherwise an error is declared. Thus, the encoding of user 2 defines a mapping, which is denoted by $\boldsymbol{x}_2(m_1, m_2)$, in parentheses, as opposed to the square brackets of $\boldsymbol{x}_2[k, m_2]$.

Let $\bar{P}_{e,1}^{sup} = \Pr\{\hat{W}_1 \neq W_1\}$ and $\bar{P}_{e,2}^{sup} = \Pr\{\hat{W}_1 = W_1, \hat{W}_2 \neq W_2\}$ when superposition coding is employed.

Let $Q \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y})$ be given. Define the following sets of p.m.f.'s that will be useful in what follows:

$$\begin{aligned}
&\mathcal{K}(Q) \triangleq \{f \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : \ f_{X_1,X_2} = Q_{X_1,X_2}\} \\
&\mathcal{G}_q(Q) \triangleq \{f \in \mathcal{K}(Q) : \\
&\qquad \boldsymbol{E}_f\{q(X_1, X_2, Y)\} \geq \boldsymbol{E}_Q\{q(X_1, X_2, Y)\}\} \\
&\mathcal{L}_1(Q) \triangleq \{f \in \mathcal{G}_q(Q) : \ f_{X_2,Y} = Q_{X_2,Y}\} \\
&\mathcal{L}_2(Q) \triangleq \{f \in \mathcal{G}_q(Q) : \ f_{X_1,Y} = Q_{X_1,Y}\} \\
&\mathcal{L}_0(Q) \triangleq \{f \in \mathcal{G}_q(Q) : \ f_Y = Q_Y\}.
\end{aligned} \tag{9}$$

Theorem 1. *Let $P = P_{X_1,X_2}P_{Y|X_1,X_2}$, then*

$$\bar{P}_{e,2}^{sup} \doteq 2^{-nE_2(P,R_2)}, \ \bar{P}_{e,1}^{sup} \doteq 2^{-nE_1(P,R_1,R_2)} \tag{10}$$

where $E_2(P, R_2) =$

$$\begin{aligned}
&\min_{P' \in \mathcal{K}(P)} \left[D(P' \| P) + \min_{\tilde{P} \in \mathcal{L}_2(P')} |I_{\tilde{P}}(X_2; Y|X_1) - R_2|^+ \right], \\
&E_1(P, R_1, R_2) = \min_{P' \in \mathcal{K}(P)} \left[D(P' \| P) \right. \\
&\left. + \min_{\tilde{P} \in \mathcal{L}_0(P')} \left| I_{\tilde{P}}(X_1; Y) + |I_{\tilde{P}}(X_2; Y|X_1) - R_2|^+ - R_1 \right|^+ \right]
\end{aligned} \tag{11}$$

We note that Theorem 1 implies that $E_{sup}(P, R_1, R_2) = \min\{E_2(P, R_2), E_1(P, R_1, R_2)\}$ is the error exponent induced by the superposition coding scheme. Define the following functions

$$\begin{aligned}
R_1'(P) &\triangleq \min_{\tilde{P} \in \mathcal{L}_1(P)} I_{\tilde{P}}(X_1; Y, X_2) \\
R_2'(P) &\triangleq \min_{\tilde{P} \in \mathcal{L}_2(P)} I_{\tilde{P}}(X_2; Y|X_1) \\
R_1''(P, R_2) &\triangleq \min_{\tilde{P} \in \mathcal{L}_0(P)} I_{\tilde{P}}(X_1; Y) + |I_{\tilde{P}}(X_2; Y|X_1) - R_2|^+ \\
R_2''(P, R_1) &\triangleq \min_{\tilde{P} \in \mathcal{L}_0(P)} \{I_{\tilde{P}}(X_2; Y) - I_{\tilde{P}}(X_2; X_1) \\
&\qquad + |I_{\tilde{P}}(X_1; Y, X_2) - R_1|^+\}.
\end{aligned} \tag{12}$$

Note that for $E_2(P, R_2), E_1(P, R_1, R_2)$ to be zero one must have $P' = P$. Theorem 1 therefore implies achievability of

$$\mathcal{R}_{cog}^{sup}(P) = \left\{ (R_1, R_2) : \begin{array}{l} R_2 \leq R_2'(P), \\ R_1 \leq R_1''(P, R_2) \end{array} \right\}. \tag{13}$$

Consider the following region $\tilde{\mathcal{R}}_{cog}^{sup}(P) =$

$$\left\{ (R_1, R_2) : \begin{array}{r} R_2 \leq \min_{\tilde{P} \in \mathcal{L}_2(P)} I_{\tilde{P}}(X_2; Y|X_1), \\ R_1 + R_2 \leq \min_{\tilde{P} \in \mathcal{L}_0^{sup}(P)} I(X_1, X_2; Y) \end{array} \right\} \tag{14}$$

where $\mathcal{L}_0^{sup}(P) = \{\tilde{P} \in \mathcal{L}_0(P) : \ I_{\tilde{P}}(X_1; Y) \leq R_1\}$. Since by definition the capacity region is a closed convex set, and $\tilde{\mathcal{R}}_{cog}^{sup}(P) \subseteq \mathcal{R}_{cog}^{sup}(P)$ this yields the following theorem.

Theorem 2. *The capacity region of the finite alphabet cognitive MAC $P_{Y|X_1,X_2}$ with decoding metric $q(x_1, x_2, y)$ contains the set of rate-pairs*

$$\mathcal{R}_{cog}^{sup} = \text{ closure of CH of } \bigcup_P \tilde{\mathcal{R}}_{cog}^{sup}(P) \tag{15}$$

where the union is over all $P \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y})$ with conditional $P_{Y|X_1,X_2}$ given by the channel.

The following theorem provides a random coding converse using superposition coding.

Theorem 3. *If $(R_1, R_2) \notin \tilde{\mathcal{R}}_{cog}^{sup}(P_{X_1,X_2,Y})$ then the average probability of error, averaged over the ensemble of random codebooks drawn according to P_{X_1,X_2} using superposition coding, approaches one as the blocklength tends to infinity.*

We note that the proof follows similarly to the proof of Theorem 3 of [5]. The following corollary follows.

Corollary 1. *$\mathcal{R}_{cog}^{sup}(P) = \tilde{\mathcal{R}}_{cog}^{sup}(P)$.*

The error exponents achievable by random binning are presented next. Let $\bar{P}_{e,2}^{bin}, \bar{P}_{e,1}^{bin}$ be defined as follows

$$\bar{P}_{e,1}^{bin} = \Pr\{\hat{W}_1 \neq W_1\}, \ \bar{P}_{e,2}^{bin} = \Pr\{\hat{W}_1 = W_1, \hat{W}_2 \neq W_2\}$$

when random binning is employed. Recall (11).

Theorem 4. *Let $P = P_{X_1,X_2}P_{Y|X_1,X_2}$, then*

$$\bar{P}_{e,2}^{bin} \doteq 2^{-nE_2(P,R_2)}, \ \bar{P}_{e,1}^{bin} \doteq 2^{-n\min\{E_0(P,R_1,R_2), E_1(P,R_1)\}}$$

where $E_1(P, R_1) =$

$$\begin{aligned}
&\min_{P' \in \mathcal{K}(P)} \left[D(P' \| P) + \min_{\tilde{P} \in \mathcal{L}_1(P')} |I_{\tilde{P}}(X_1; Y, X_2) - R_1|^+ \right] \\
&E_0(P, R_1, R_2) = \max\{E_1(P, R_1, R_2), E_{0,b}(P, R_1, R_2)\}
\end{aligned}$$

and where

$$\begin{aligned}
&E_{0,b}(P, R_1, R_2) \\
&= \min_{P' \in \mathcal{K}(P)} \left[D(P' \| P) + \min_{\tilde{P} \in \mathcal{L}_0(P')} \left| I_{\tilde{P}}(X_2; Y) - I_P(X_1; X_2) \right. \right. \\
&\qquad \left. \left. + |I_{\tilde{P}}(X_1; Y, X_2) - R_1|^+ - R_2 \right|^+ \right].
\end{aligned} \tag{16}$$

The derivation of the exponent associated with $\bar{P}_{e,1}^{bin}$ makes use of [6, Lemma 3]. Note that Theorem 4 implies that $E_{bin}(P, R_1, R_2) = \min\{E_2(P, R_2), E_0(P, R_1, R_2), E_1(P, R_1)\}$ is the error exponent induced by the random binning scheme. Theorem 4 also implies that for fixed $P = P_{X_1,X_2}P_{Y|X_1,X_2}$, the region

$$\mathcal{R}_{cog}^{bin}(P) = \left\{ \begin{array}{l} (R_1, R_2) : \\ R_1 \leq R_1'(P), \\ R_2 \leq R_2'(P), \\ R_1 \leq R_1''(P, R_2) \text{ or } R_2 \leq R_2''(P, R_1) \end{array} \right\}, \tag{17}$$

is achievable, where $R_1'(P), R_2'(P), R_1''(P, R_2), R_2''(P, R_1)$ are defined in (12). Note that $R_{cog}^{bin}(P)$ can be potentially enlarged as follows:

Lemma 1. *Let $(R_1, R_2) \in \mathcal{R}_{cog}^{bin}(P)$ then $(R_1 + R_2, 0)$ is also achievable by random binning.*

The lemma follows since the cognitive encoder can assign some of the information it transmits to the non-cognitive user. The resulting region of rates achievable by random binning is described in the following theorem.

Theorem 5. *The capacity region of the finite alphabet cognitive MAC $P_{Y|X_1,X_2}$ with decoding metric $q(x_1, x_2, y)$ contains the set of rate-pairs*

$$\mathcal{R}_{cog}^{bin} = \text{ closure of CH of } \bigcup_{P} \mathcal{R}_{cog}^{bin,*}(P) \tag{18}$$

where the union is over all $P \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y})$ with conditional $P_{Y|X_1,X_2}$ given by the channel.

Consider the rate-region:

$$\tilde{\mathcal{R}}_{cog}^{bin}(P) = \left\{ \begin{array}{l} (R_1, R_2) : \\ R_1 \leq \min_{\tilde{P} \in \mathcal{L}_1(P)} I_{\tilde{P}}(X_1; Y, X_2), \\ R_2 \leq \min_{\tilde{P} \in \mathcal{L}_2(P)} I_{\tilde{P}}(X_2; Y|X_1), \\ R_1 + R_2 \leq \min_{\tilde{P} \in \mathcal{L}_0^{bin}(P)} I_{\tilde{P}}(X_1, X_2; Y) \end{array} \right\}$$

where $\mathcal{L}_0^{bin}(P) = \{\tilde{P} \in \mathcal{L}_0(P) : I_{\tilde{P}}(X_1; Y) \leq R_1, I_{\tilde{P}}(X_2; Y) - I_P(X_2; X_1) \leq R_2\}$. The following theorem provides a random coding converse for random binning.

Theorem 6. *If $(R_1, R_2) \notin \tilde{\mathcal{R}}_{cog}^{bin}(P_{X_1,X_2,Y})$ then the average probability of error, averaged over the ensemble of random codebooks drawn according to P_{X_1,X_2} using binning, approaches one as the blocklength tends to infinity.*

The proof of Theorem 6 adheres closely to that of Theorem 3. The theorem implies the following corollary.

Corollary 2. $\mathcal{R}_{cog}^{bin}(P) = \tilde{\mathcal{R}}_{cog}^{bin}(P)$.

A few comments are in order:

• It is easy to show, as one may expect, that the achievable region of the cognitive mismatched MAC, $\mathcal{R}_{cog}^{bin}$, contains the achievable region of the mismatched MAC, $\mathcal{R}_{LM}$ (5):

Lemma 2. $\mathcal{R}_{LM} \subseteq \mathcal{R}_{cog}^{bin}$.

• The following proposition specializes the results to the matched case.

Proposition 1. *In the matched case where $q(x_1, x_2, y) = \log p(y|x_1, x_2)$, $\mathcal{R}_{cog}^{sup} = \mathcal{R}_{cog}^{bin} = \mathcal{R}_{cog}^{match}$ where*

$$\mathcal{R}_{cog}^{match} = \bigcup_{P_{X_1,X_2}} \{(R_1, R_2) : R_2 \leq I_P(X_2; Y|X_1), R_1 + R_2 \leq I_P(X_1, X_2; Y)\},$$

and P abbreviates $P_{X_1,X_2}P_{Y|X_1,X_2}$.

Theorem 1 is clearly an example for which $\mathcal{R}_{LM} \subseteq \mathcal{R}_{cog}^{sup} = \mathcal{R}_{cog}^{bin}$ with obvious cases in which the inclusion is strict.

• In principle, neither region $\mathcal{R}_{cog}^{sup}(P)$ (17), nor $\mathcal{R}_{cog}^{bin,*}(P)$ (see Lemma 1) dominates the other, as the second inequality in (13) is stricter than the third inequality in (17) and the first inequality in (17) does not appear in (13). It is easily verified that unless $R_1''(P, R_2'(P)) > R_1'(P)$ and $R_2''(P, R_1'(P)) > R_2'(P)$ we have $\mathcal{R}_{cog}^{sup}(P) \subseteq \mathcal{R}_{cog}^{bin,*}(P)$, otherwise, the opposite inclusion $\mathcal{R}_{cog}^{bin,*}(P) \subseteq \mathcal{R}_{cog}^{sup}(P)$ may occur.

• Consider the following parallel MAC which is a special case of the channel that was studied in [5, Section IV, Example 2]. Let X_1, X_2 be binary $\{0, 1\}$. The output of the channel is given by $Y = (Y_1, Y_2) = (X_1, X_2 \oplus Z)$ where $\oplus$ denotes modulo-2 addition, and $Z \sim$ Bernoulli(p''), with the decoding metric $q(x_1, x_2, (y_1, y_2)) = -\frac{1}{2}(x_1 \oplus y_1 + x_2 \oplus y_2)$. In [14] we obtain that $\mathcal{R}_{cog}^{bin} = \left\{(R_1, R_2) : \begin{array}{l} R_2 \leq 1 - h_2(p'') \\ R_1 + R_1 \leq 2 - h_2(p'') \end{array}\right\}$, which is also the capacity region of the matched cognitive MAC. It is shown [14] that the vertex point $(R_1, R_2) = (1, 1 - h_2(p''))$ is not achievable by superposition coding.

• The following proposition generalizes the fact that binning performs as well as superposition coding in the matched case.

Proposition 2. *If $R_2'(P) \geq I_P(X_2; Y) - I_P(X_1; X_2)$ then $\mathcal{R}_{cog}^{sup}(P) \subseteq \mathcal{R}_{cog}^{bin}(P)$.*

• It is noted that in certain cases, binning is more advantageous than superposition coding in terms of memory requirements: superposition coding requires the cognitive user to use a separate codebook for every possible codeword of the non-cognitive user, i.e., a collection of $2^{n(R_1+R_2)}$ codewords. Binning on the other hand allows encoder 2 to decrease memory requirements to $2^{nR_1} + 2^{n(R_2+I_P(X_1;X_2))}$ codewords at the cost of increased encoding complexity[1].

IV. The Mismatched Single-User Channel

This section shows that achievable rates for the mismatched single-user DMC can be derived from the maximal sum-rate of an appropriately chosen mismatched cognitive MAC.

Similar to [5], consider the single-user mismatched DMC $P_{Y|X}$ with input alphabet $\mathcal{X}$ and decoding metric $q(x, y)$. Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be finite alphabets and let ϕ be a given mapping $\phi : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{X}$. Consider the mismatched cognitive MAC with input alphabets $\mathcal{X}_1, \mathcal{X}_2$ and output alphabet $\mathcal{Y}$, whose input-output relation is given by

$$P_{Y|X_1,X_2}(y|x_1, x_2) = P_{Y|X}(y|\phi(x_1, x_2)). \tag{19}$$

[1]In this case, the choice of $\boldsymbol{x}_2(i, j)$ should not be arbitrary among the vectors that are jointly typical with $\boldsymbol{x}_1(i)$ in the j-th bin, but rather, a deterministic rule, e.g., pick the jointly typical $\boldsymbol{x}_2[k, j]$ with the lowest k.

The decoding metric $q(x_1,x_2,y)$ of the mismatched cognitive MAC is defined in terms of that of the single-user channel:

$$q(x_1,x_2,y) = q(\phi(x_1,x_2),y). \tag{20}$$

The resulting mismatched cognitive MAC will be referred to as the cognitive MAC induced by the single-user channel. Note that, in fact, X_1 and X_2 can be regarded as auxiliary random variables for the original single-user channel $P_{Y|X}$.

In [5, Theorem 4], it is shown that the mismatch capacity of the single-user channel is lower-bounded by R_1+R_2 for any pair (R_1,R_2) that, for some mapping ϕ and for some distributions P_{X_1} and P_{X_2} satisfy

$$\begin{aligned} &R_1 < \min_{f\in\mathcal{D}'_{(1)}} I_f(X_1;Y|X_2),\ R_2 < \min_{f\in\mathcal{D}'_{(2)}} I_f(X_2;Y|X_1) \\ &R_1+R_2 < \min_{f\in\mathcal{D}'_{(0)}} I_f(X_1,X_2;Y), \end{aligned} \tag{21}$$

where $\mathcal{D}'_{(i)} = \{f \in \mathcal{D}_{(i)} : f_{X_1,X_2} = f_{X_1}f_{X_2}\}$, $i=0,1,2$, the $\mathcal{D}_{(i)}$'s are defined in (6), and $P_{X_1,X_2,Y} = P_{X_1}P_{X_2}P_{Y|\phi(X_1,X_2)}$. The proof of [5, Theorem 4] is based on expurgating the product codebook of the MAC (19) containing $2^{n(R_1+R_2)}$ codewords $\boldsymbol{v}(i,j) = (\boldsymbol{x}_1(i),\boldsymbol{x}_2(j))$ by keeping only the $\boldsymbol{v}(i,j)$'s whose components, $\boldsymbol{x}_1(i),\boldsymbol{x}_2(j)$, are jointly ϵ-typical w.r.t. the product p.m.f. $P_{X_1}P_{X_2}$. This expurgation causes a negligible rate loss and therefore makes is possible to consider minimization over product measures in (21).

It is easy to realize that in the cognitive MAC case as well, if (R_1,R_2) is an achievable rate-pair for the induced cognitive MAC, R_1+R_2 is an achievable rate for the inducing single-user channel. While the users of the non-cognitive induced MAC of [5] exercise a limited degree of cooperation (by expurgation), the induced cognitive MAC introduced here enables a much higher degree of cooperation between users. There is no need to expurgate codewords for cases of either superposition coding or random binning, since the codebook generation guarantees that for all (i,j), $(\boldsymbol{x}_1(i),\boldsymbol{x}_2(i,j))$ lies in the desired joint type-class $T(P_{X_1,X_2})$.

Let $\mathcal{R}_{cog}(P_{X_1,X_2,Y},q(x_1,x_2,y)) = \mathcal{R}_{cog}^{sup}(P) \cup \mathcal{R}_{cog}^{bin,*}(P)$.

Theorem 7. *For all finite $\mathcal{X}_1,\mathcal{X}_2$, $P_{X_1,X_2} \in \mathcal{P}(\mathcal{X}_1,\mathcal{X}_2)$ and $\phi : \mathcal{X}_1\times\mathcal{X}_2 \to \mathcal{X}$, the capacity of the single-user mismatched DMC $P_{Y|X}$ with decoding metric $q(x,y)$ is lower bounded by the sum-rate resulting from $\mathcal{R}_{cog}(P_{X_1,X_2}P_{Y|\phi(X_1,X_2)},q(\phi(x_1,x_2),y))$.*

While Theorem 7 may not improve the rate (21) in optimizing over all $(\mathcal{X}_1,\mathcal{X}_2,\phi,P_{X_1},P_{X_2})$, it can certainly improve the achieved rates for given (ϕ,P_{X_1},P_{X_2}) as demonstrated in Section III, and thereby may reduce the computational complexity required to find a good code.

Next, it is demonstrated how superposition coding can be used to achieve the rate of Theorem 7. Consider the region of rate-pairs (R_1,R_2) which satisfy

$$\begin{aligned} R_1 &\le r_1(P) \triangleq \min_{\tilde{P}\in\mathcal{L}_1(P)} I_{\tilde{P}}(X_1;Y|X_2) \\ R_2 &\le r_2(P,R_1) \\ &\triangleq \min_{\tilde{P}\in\mathcal{L}_0(P)} I_{\tilde{P}}(X_2;Y) + |I_{\tilde{P}}(X_1;Y|X_2) - R_1|^+. \end{aligned} \tag{22}$$

This region is obtained by reversing the roles of the users in (13), i.e., setting user 1 as the cognitive one.

It can be shown [14] that the sum-rate resulting from $\mathcal{R}_{cog}^{bin}(P)$ (17) is upper-bounded by the sum-rate resulting from the union of the regions (22) and (13) that are achievable by superposition coding. This yields the following corollary.

Corollary 3. *For all finite $\mathcal{X}_1,\mathcal{X}_2$, $P_{X_1,X_2} \in \mathcal{P}(\mathcal{X}_1,\mathcal{X}_2)$ and $\phi : \mathcal{X}_1\times\mathcal{X}_2 \to \mathcal{X}$, the capacity of the single-user mismatched DMC $P_{Y|X}$ with decoding metric $q(x,y)$ is lower bounded by the rate $\max\{R_2'(P)+R_1''(P,R_2'(P)), r_1(P)+r_2(P,r_1(P))\}$, which is achievable by superposition coding, where $P = P_{X_1,X_2}\times P_{Y|\phi(X_1,X_2)}$.*

V. Acknowledgement

The author would like to thank the anonymous reviewers of ISIT 2013 for their very helpful comments and suggestions, which helped improving this paper.

References

[1] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *infoth*, vol. 27, no. 1, pp. 5–12, 1981.

[2] J. Hui, "Fundamental issues of multiple accessing," *PhD dissertation, MIT*, 1983.

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

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

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

[6] J. Scarlett and A. Guillén i Fàbregas, "An achievable error exponent for the mismatched multiple-access channel," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, October 2012, pp. 1975–1982.

[7] V. Balakirsky, "Coding theorem for discrete memoryless channels with given decision rule," in *Algebraic Coding*, ser. Lecture Notes in Computer Science, G. Cohen, A. Lobstein, G. Zémor, and S. Litsyn, Eds. Springer Berlin Heidelberg, 1992, vol. 573, pp. 142–150. [Online]. Available: http://dx.doi.org/10.1007/BFb0034351

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

[9] Y.-S. Liu and B. Hughes, "A new universal random coding bound for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 376–386, Match 1996.

[10] A. Lapidoth, "Nearest neighbor decoding for additive non-gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, Sept. 1996.

[11] A. Ganti, A. Lapidoth, and I. Telatar, "Mismatched decoding revisited: general alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[12] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched decoding: finite-length bounds, error exponents and approximations," *arXiv:1303.6166 [cs.IT]*, March 2013.

[13] ——, "Ensemble-tight error exponents for mismatched decoders," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, 2012, pp. 1951–1958.

[14] A. Somekh-Baruch, "On achievable rates for channels with mismatched decoding," May 2013, arXiv 1305.0547 [cs.IT].