Mismatched decoding for the relay channel

Mismatched Decoding for the Relay Channel

Charlotte Hucher and Parastoo Sadeghi
The Australian National University, Canberra, ACT, 0200, Australia
Email: {charlotte.hucher,parastoo.sadeghi}@anu.edu.au

***Abstract*— We consider a discrete memoryless relay channel, where both relay and destination may have an incorrect estimation of the channel. This estimation error is modeled with mismatched decoders. In this paper, we provide a lower-bound on the mismatch capacity of the relay channel. Moreover, we prove that this lower-bound is indeed the exact capacity for the degraded relay channel when random coding is used.**

I. Introduction

The decoding method that minimizes the error probability is the maximum-likelihood (ML) decoder. However, it cannot always be implemented in practice because of some channel estimation errors or hardware limitations. An alternative decoder can then be a mismatched one, based on a different metric. The theoretical performance of mismatched decoding has been studied since the 1980's when Csiszàr and Körner in [1], and Hui in [2] both provided a lower-bound on the achievable capacity in a point-to-point communication channel. In [3], the authors proved that this lower-bound is the exact capacity when random coding is used. The mismatch capacity of multiple-access channels has also been characterized in [4].

There is increasing evidence that future wireless communications will be based not on point-to-point transmission anymore, but on cooperation between the nodes in a network (see [5],[6]). The simplest model of a cooperative network is the relay channel for which capacity bounds have been derived in 1979 by Cover and El Gamal in [7].

In this paper, we consider a discrete memoryless relay channel with mismatched decoders at both receivers (the relay and the destination). We provide a lower-bound on the mismatch capacity of such a channel and prove that it is indeed the exact capacity of the mismatched degraded relay channel when random coding is used.

II. The relay channel and mismatched decoder

We consider a discrete memoryless relay channel consisting of one source, one relay and one destination. We use the same setup as in [7]: The source broadcasts a signal $x_1 \in \mathcal{X}_1$ which is received by both the relay and the destination. The relay transmits a signal $x_2 \in \mathcal{X}_2$ which is received by the destination. Received signals at relay and destination are denoted by $y_1 \in \mathcal{Y}_1$ and $y \in \mathcal{Y}$ respectively (see Figure 1).

The channel is modeled by a set of probability distributions $p(y_1, y|x_1, x_2)$. We consider three mismatched decoders using the metrics $q_{sr}(x_1, x_2, y_1)$, $q_{rd}(x_2, y)$ and $q_{sd}(x_1, x_2, y)$, where the subscripts sr, rd and sd stand for the source-relay, relay-destination and source-destination links, respectively.

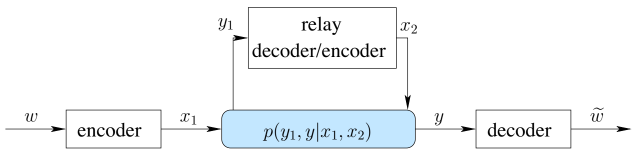

Fig. 1. Relay channel

The following setup is used to prove the achievability of the lower-bound on the mismatch capacity derived in this paper.

We consider the transmission of B blocks of length n. In each block $i \in \{1, \ldots, B\}$, a message $w_i \in \{1, \ldots, 2^{nR}\}$ is transmitted from the source. Let us partition the set $\{1, \ldots, 2^{nR}\}$ into 2^{nR_0} independent subsets denoted by $S_s, s \in \{1, \ldots, 2^{nR_0}\}$, such that any $w_i \in \{1, \ldots, 2^{nR}\}$ belongs to a unique subset S_{s_i}. The message is then coded as $\mathbf{x}_1(w_i|s_i) \in \mathcal{X}_1^n$ and $\mathbf{x}_2(s_i) \in \mathcal{X}_2^n$ at the source and relay, respectively.

Random coding: The choice of the set $\mathcal{C} = \{x_1(.|.), x_2(.)\}$ of codewords is random:

- 2^{nR_0} iid codewords in $\mathcal{X}_2^n$ are first generated according to the probability distribution $p(\mathbf{x}_2) = \prod_{i=1}^n p(x_{2i})$ and indexed by $s \in \{1, \ldots, 2^{nR_0}\}$: $\mathbf{x}_2(s)$;
- for each $\mathbf{x}_2(s)$, 2^{nR} iid codewords in $\mathcal{X}_1^n$ are generated according to the probability distribution $p(\mathbf{x}_1|\mathbf{x}_2(s))$ and indexed by $w \in \{1, \ldots, 2^{nR}\}$: $\mathbf{x}_1(w|s)$.

Two transmission steps: Let us assume that $(i-2)$ blocks have already been sent. Thus the relay has already decoded w_{i-2} and s_{i-1}, the destination has decoded w_{i-3} and s_{i-2}. In order to derive a lower bound on the capacity of the mismatched relay channel, we choose to use threshold decoders as in [2], [4]. Indeed, it can be proven that the decoding error probability of an ML mismatched decoder (which is implemented in practice) is upper-bounded by the decoding error probability of the considered threshold decoder.

In block $(i-1)$, the source and relay transmit $\mathbf{x}_1(w_{i-1}|s_{i-1})$ and $\mathbf{x}_2(s_{i-1})$, respectively; the relay and destination receive $\mathbf{y}_1(i-1)$ and $\mathbf{y}(i-1)$. The relay is able to detect w_{i-1} as the unique w such that $(\mathbf{x}_1(w|s_{i-1}), \mathbf{x}_2(s_{i-1}), \mathbf{y}_1(i-1))$ is jointly typical and $q_{sr}(\mathbf{x}_1(w|s_{i-1}), \mathbf{x}_2(s_{i-1}), \mathbf{y}_1(i-1))$ is larger than a threshold to be defined later. The relay is thus able to determine s_i such that $w_{i-1} \in S_{s_i}$. The source is obviously also aware of s_i.

In block i, the source and relay transmit $\mathbf{x}_1(w_i|s_i)$ and $\mathbf{x}_2(s_i)$, respectively; the relay and destination receive $\mathbf{y}_1(i)$ and $\mathbf{y}(i)$. The destination can detect s_i as the unique s such that $(\mathbf{x}_2(s_i), \mathbf{y}(i))$ is jointly typical and $q_{rd}(\mathbf{x}_2(s_i), \mathbf{y}(i))$ is larger than some threshold. It is then able to detect

w_{i-1} as the unique w such that $w \in S_{s_i} \cap \mathcal{L}(\mathbf{y}(i-1))$, where $\mathcal{L}(\mathbf{y}(i-1))$ is the set of all $w \in \{1, \ldots, 2^{nR}\}$ such that $(\mathbf{x}_1(w|s_{i-1}), \mathbf{x}_2(s_{i-1}), \mathbf{y}(i-1))$ is jointly typical and $q_{sd}(\mathbf{x}_1(w|s_{i-1}), \mathbf{x}_2(s_{i-1}), \mathbf{y}(i-1))$ is larger than some threshold.

Notation: Let $E_p(q(x))$ denote the expected value of $q(x)$ w.r.t. the probability distribution $p(x)$. Let $I_f(X;Y)$ denote the usual mutual information between X and Y w.r.t. the probability distribution $f(x,y)$.

For a probability distribution $p(x)$ on a finite set $\mathcal{X}$ and a constant $\delta > 0$, let $N^\delta_{p(x)}$ denote the set of all probability distributions on $\mathcal{X}$ that are within δ of $p(x)$: $N^\delta_{p(x)} = \{f \in \mathcal{P}(\mathcal{X}) : \forall x \in \mathcal{X}, |f(x) - p(x)| \leq \delta\}$. Let $T^\delta_{p(x)}$ be the set of all sequences in $\mathcal{X}^n$ whose type is in $N^\delta_{p(x)}$: $T^\delta_{p(x)} = \left\{\mathbf{x} \in \mathcal{X}^n : \mathbf{f}_x \in N^\delta_{p(x)}\right\}$, where $\mathbf{f}_x(x)$ is the number of elements of the sequence $\mathbf{x}$ that are equal to x, normalized by the sequence length n. In the following, we drop the subscript arguments when the context is clear enough and write $f(x) \in N^\delta_p$ and $(\mathbf{x}, \mathbf{y}) \in T^\delta_p$.

III. Main results

Theorem 1: (Achievability) Consider a discrete memoryless relay channel described by the probability distribution $p(y_1, y|x_1, x_2)$. The capacity C_M obtained using the mismatched decoders $q_{sr}(x_1, x_2, y_1)$, $q_{rd}(x_2, y)$ and $q_{sd}(x_1, x_2, y)$ is lower-bounded by:

$$C_{LM} = \max_{p(x_1,x_2)} I_{LM}(p(x_1, x_2)) \tag{1}$$

$$I_{LM}(p(x_1, x_2)) = \min\{I_{sr}(p(x_1, x_2)), I_{rd}(p(x_2)) + I_{sd}(p(x_1, x_2))\} \tag{2}$$

where

$$I_{sr}(p(x_1, x_2)) = \min_{f \in \mathcal{D}_{sr}} I_f(X_1; Y_1|X_2) + I_f(X_1; X_2) \tag{3}$$

$$I_{rd}(p(x_2)) = \min_{f \in \mathcal{D}_{rd}} I_f(X_2; Y) \tag{4}$$

$$I_{sd}(p(x_1, x_2)) = \min_{f \in \mathcal{D}_{sd}} I_f(X_1; Y|X_2) + I_f(X_1; X_2) \tag{5}$$

and

$$\mathcal{D}_{sr} = \{f(x_1, x_2, y_1) : f(x_1) = p(x_1), f(x_2, y_1) = p(x_2, y_1), E_f(q_{sr}(x_1, x_2, y_1)) \geq E_p(q_{sr}(x_1, x_2, y_1))\} \tag{6}$$

$$\mathcal{D}_{rd} = \{f(x_2, y) : f(x_2) = p(x_2), f(y) = p(y), E_f(q_{rd}(x_2, y)) \geq E_p(q_{rd}(x_2, y))\} \tag{7}$$

$$\mathcal{D}_{sd} = \{f(x_1, x_2, y) : f(x_1) = p(x_1), f(x_2, y) = p(x_2, y), E_f(q_{sd}(x_1, x_2, y)) \geq E_p(q_{sd}(x_1, x_2, y))\}. \tag{8}$$

Let now suppose that there exist three probability distributions $\hat{f}_i$, $i \in \{sr, rd, sd\}$ such that $E_{\hat{f}_i}(q_i) > E_p(q_i)$ with strict inequality.

Theorem 2: (Converse) With the above assumption and for a degraded relay channel, if for some input distribution $p(x_1, x_2)$, the rate $R > C_{LM}$, then the average probability of error, averaged over all random codebooks drawn according to $p(x_1, x_2)$, approaches one as the block length tends to infinity.

IV. Proof of Theorem 1

A. Upper-bounding the error probability

In order to prove the achievability of this lower-bound, we consider the four possible error events described in [7] adapted to the use of a threshold decoder. For each block i, these possible error events are:

- E_{0i}: $(\mathbf{x}_1(w_i|s_i), \mathbf{x}_2(s_i), \mathbf{y}_1(i), \mathbf{y}(i))$ is not jointly typical;
- E_{1i}: there exists $\tilde{w} \neq w_i$ such that $(\mathbf{x}_1(\tilde{w}|s_i), \mathbf{x}_2(s_i), \mathbf{y}_1(i))$ is jointly typical and $q_{sr}(\mathbf{x}_1(\tilde{w}|s_i), \mathbf{x}_2(s_i), \mathbf{y}_1(i))$ is larger than some threshold;
- E_{2i}: there exists $\tilde{s} \neq s_i$ such that $(\mathbf{x}_2(\tilde{s}), \mathbf{y}(i))$ is jointly typical and $q_{rd}(\mathbf{x}_2(\tilde{s}), \mathbf{y}(i))$ is larger than some threshold;
- $E_{3i} = E'_{3i} \cup E''_{3i}$
  - E'_{3i}: $w_{i-1} \notin S_{s_i} \cap \mathcal{L}(\mathbf{y}(i-1))$
  - E''_{3i}: there exists $\tilde{w} \neq w_{i-1}$ such that $\tilde{w} \in S_{s_i} \cap \mathcal{L}(\mathbf{y}(i-1))$.

Let F_i be the decoding error event in block i. Let us assume that no error has occurred till block $i-1$. Thus, the decoding error probability in block i is given by:

$$p_e(i) = \sum_{k=0}^{3} \Pr\left\{E_{ki} \cap F^c_{i-1} \bigcap_{l=0}^{k-1} E^c_{li}\right\} \triangleq \sum_{k=0}^{3} p_{ek}(i).$$

1) Probability of error event E_{0i}: By Sanov's theorem [8, Theorem 11.4.1], this probability is exponentially small in n. There exists $\psi > 0$ such that $p_{e0}(i) < 2^{-n\psi}$.

2) Probability of error event E_{1i}: An error occurs if there exists $\tilde{w} \neq w_i$ such that the metric $q_{sr}(\mathbf{x}_1(\tilde{w}|s_i), \mathbf{x}_2(s_i), \mathbf{y}_1(i))$ is greater than the threshold

$$\Upsilon^\delta_{sr} = \min_{\tilde{p} \in N^\delta_p} \sum_{(x_1,x_2,y_1) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}_1} \tilde{p}(x_1, x_2, y_1) q_{sr}(x_1, x_2, y_1), \tag{9}$$

where δ is a small positive number.

The probability of error event E_{1i} is thus given by

$$p_{e1}(i) = \Pr\{\exists \tilde{w} \neq w_i, \mathbf{x}_1(\tilde{w}|s_i) \in T^\delta_p, q_{sr}(\mathbf{x}_1(\tilde{w}|s_i), \mathbf{x}_2(s_i), \mathbf{y}_1(i)) \geq \Upsilon^\delta_{sr} | F^c_{i-1}\}.$$

Using Sanov's theorem, we obtain the upper-bound

$$p_{e1}(i) \leq (2^{nR} - 1)(n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}_1|} 2^{-n\tilde{R}^\delta_{sr}} \leq (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}_1|} 2^{-n(\tilde{R}^\delta_{sr} - R)},$$

where

$$\tilde{R}^\delta_{sr} = \min_{f(x_1,x_2,y) \in \mathcal{D}^\delta_{sr}} D(f(x_1, x_2, y_1) \| p(x_1) p(x_2, y_1)) \tag{10}$$

$$\mathcal{D}^\delta_{sr} = \{f(x_1, x_2, y_1) : f(x_1) \in N^\delta_p, f(x_2, y_1) \in N^\delta_p, E_f(q_{sr}(x_1, x_2, y_1)) \geq \Upsilon^\delta_{sr}\}, \tag{11}$$

with $D(.\|.)$ denoting the KL divergence [8, equation (2.26)].

Thus, if $R < \tilde{R}^\delta_{sr}$, the probability of error event E_{1i} is exponentially small in n: $p_{e1}(i) < 2^{-n\psi}$ for some $\psi > 0$.

3) Probability of error event E_{2i}: Using the threshold

$$\Upsilon_{rd}^{\delta} = \min_{\tilde{p} \in N_p^{\delta}} \sum_{(x_2,y) \in \mathcal{X}_2 \times \mathcal{Y}} \tilde{p}(x_2,y) q_{rd}(x_2,y), \tag{12}$$

we can write the probability of error event E_{2i} as

$$\begin{aligned} p_{e2}(i) &= \Pr\left\{\exists \tilde{s} \neq s_i, \mathbf{x}_2(\tilde{s}) \in T_p^{\delta}, q_{rd}(\mathbf{x}_2(\tilde{s}), \mathbf{y}(i)) \geq \Upsilon_{rd}^{\delta} | F_{i-1}^c\right\} \\ &\leq (n+1)^{|\mathcal{X}_2||\mathcal{Y}|} 2^{-n(\tilde{R}_{rd}^{\delta} - R_0)}, \end{aligned}$$

where the upper-bound is obtained using Sanov's theorem with

$$\tilde{R}_{rd}^{\delta} = \min_{f(x_2,y) \in \mathcal{D}_{rd}^{\delta}} D(f(x_2,y) \| p(x_2)p(y)) \tag{13}$$

$$\begin{aligned} \mathcal{D}_{rd}^{\delta} = \{ & f(x_2,y) : f(x_2) \in N_p^{\delta}, f(y) \in N_p^{\delta}, \\ & E_f(q_{rd}(x_2,y)) \geq \Upsilon_{rd}^{\delta} \}. \end{aligned} \tag{14}$$

If $R_0 < \tilde{R}_{rd}^{\delta}$, then $p_{e2}(i) < 2^{-n\psi}$ for some $\psi > 0$.

4) Probability of error event E_{3i}: Error event E_{3i} can be decomposed into two different events E'_{3i} and E''_{3i}.

If we assume that the previous transmission was correctly received at destination, then $w_{i-1} \in \mathcal{L}(\mathbf{y}(i-1))$. Moreover, the fact that the error event E_{2i} does not occur implies that $w_{i-1} \in S_{\hat{s}_i} = S_{s_i}$. Thus the first term of the decomposition has a probability zero and we only need to consider E''_{3i}.

$$\begin{aligned} p_{e3}(i) &= \Pr\left\{\exists \tilde{w} \neq w_{i-1}, \tilde{w} \in S_{s_i} \cap \mathcal{L}(\mathbf{y}(i-1)) | F_{i-1}^c\right\} \\ &\leq E\left\{ \sum_{\tilde{w} \neq w_{i-1}, \tilde{w} \in \mathcal{L}(\mathbf{y}(i-1))} \Pr\{\tilde{w} \in S_{s_i}\} | F_{i-1}^c \right\} \\ &\leq E\left\{ \|\mathcal{L}(\mathbf{y}(i-1))\| 2^{-nR_0} | F_{i-1}^c \right\} \end{aligned}$$

where $\|.\|$ denotes the cardinality of the considered set.

Let

$$\varphi(\tilde{w}|\mathbf{y}) = \begin{cases} 1, & \mathbf{x}_1(\tilde{w}|s_{i-1}) \in T_p^{\delta}, \\ & q_{sd}(\mathbf{x}_1(\tilde{w}|s_{i-1}), \mathbf{x}_2(s_{i-1}), \mathbf{y}(i-1)) \geq \Upsilon_{sd}^{\delta} \\ 0, & \text{otherwise.} \end{cases}$$

where the threshold is defined by

$$\Upsilon_{sd}^{\delta} = \min_{\tilde{p} \in N_p^{\delta}} \sum_{(x_1,x_2,y) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}} \tilde{p}(x_1,x_2,y) q_{sd}(x_1,x_2,y). \tag{15}$$

Using Sanov's theorem, we can upper-bound the expected cardinality of $\mathcal{L}(\mathbf{y}(i-1))$ given that $\tilde{w} \neq w_{i-1}$

$$\begin{aligned} & E\left\{ \|\mathcal{L}(\mathbf{y}(i-1)) | \tilde{w} \neq w_{i-1}\| \,|F_{i-1}^c \right\} \\ &= E\left\{ \sum_{\tilde{w} \neq w_{i-1}} \varphi(\tilde{w}|\mathbf{y}) \big| F_{i-1}^c \right\} \\ &\leq (2^{nR} - 1)(n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} 2^{-n\tilde{R}_{sd}^{\delta}} \\ &\leq (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} 2^{-n(\tilde{R}_{sd}^{\delta} - R)}, \end{aligned}$$

where

$$\tilde{R}_{sd}^{\delta} = \min_{f(x_1,x_2,y) \in \mathcal{D}_{sd}^{\delta}} D(f(x_1,x_2,y) \| p(x_1)p(x_2,y)) \tag{16}$$

$$\begin{aligned} \mathcal{D}_{sd}^{\delta} = \{ & f(x_1,x_2,y) : f(x_1) \in N_p^{\delta}, f(x_2,y) \in N_p^{\delta}, \\ & E_f(q_{sd}(x_1,x_2,y)) \geq \Upsilon_{sd}^{\delta} \}. \end{aligned} \tag{17}$$

The error probability is then upper-bounded by

$$\begin{aligned} p_{e3}(i) &\leq (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} 2^{-n(\tilde{R}_{sd}^{\delta} - R)} 2^{-nR_0} \\ &\leq (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} 2^{-n(\tilde{R}_{sd}^{\delta} + R_0 - R)}. \end{aligned}$$

Replacing R_0 by the constraint previously computed

$$p_{e3}(i) \leq (n+1)^{|\mathcal{X}_1||\mathcal{X}_2||\mathcal{Y}|} 2^{-n(\tilde{R}_{sd}^{\delta} + \tilde{R}_{rd}^{\delta} - R)}.$$

If $R < \tilde{R}_{sd}^{\delta} + \tilde{R}_{rd}^{\delta}$, then $p_{e3}(i) < 2^{-n\psi}$ for some $\psi > 0$.

B. Existence of a random code

If $R < \tilde{R}_{sr}^{\delta}$ and $R < \tilde{R}_{rd}^{\delta} + \tilde{R}_{sd}^{\delta}$, then the total error probability is exponentially small in n: $p_e < 4B \times 2^{-n\psi}$. Thus, as n tends to infinity, the probability of finding a set of codewords $\mathcal{C}$ respecting $p_e(\mathcal{C}) < 4B \times 2^{-n\psi}$ tends to one.

Let $\mathcal{C}$ be such a set of codewords of length n. Its average error probability is lower than $4B \times 2^{-n\psi}$. Throwing away the worst half of the codewords, we end up with a set of codewords $\mathcal{C}^*$ of length $\frac{n}{2}$ whose maximum error probability is lower than $2 \times 4B \times 2^{-n\psi}$, which tends to zero, and whose rate is $R - \frac{1}{n}$ which tends to R.

C. Letting δ tend to zero

We note that $\lim_{\delta \to 0} \Upsilon_{sr}^{\delta} = E_p(q_{sr}(x_1,x_2,y_1))$. Thus, the set $\mathcal{D}_{sr}^{\delta}$ becomes

$$\begin{aligned} \mathcal{D}_{sr} = \{ & f(x_1,x_2,y_1) : f(x_1) = p(x_1), f(x_2,y_1) = p(x_2,y_1), \\ & E_f(q_{sr}(x_1,x_2,y_1)) \geq E_p(q_{sr}(x_1,x_2,y_1)) \} \end{aligned}$$

and with these new constraints on the probability distribution

$$D(f(x_1,x_2,y_1) \| p(x_1)p(x_2,y_1)) = I_f(X_1;Y_1|X_2) + I_f(X_1;X_2).$$

Thus the first constraint of the rate becomes

$$I_{sr}(p(x_1,x_2)) \triangleq \min_{f \in \mathcal{D}_{sr}} I_f(X_1;Y_1|X_2) + I_f(X_1;X_2). \tag{18}$$

In the same way, we find the final expressions of $I_{rd}(p(x_1,x_2))$ and $I_{sd}(p(x_1,x_2))$.

V. Proof of Theorem 2

The proof of Theorem 2 is in essence similar to the one of [4, Theorem 3].

A. Decoding at relay

Let us assume that

$$R > \min_{f \in \mathcal{D}_{sr}} I_f(X_1;Y_1|X_2) + I_f(X_1;X_2). \tag{19}$$

Let f^* be the probability distribution that achieves (19). Let $\tilde{f} = (1-\epsilon)f^* + \epsilon \hat{f}_{sr}$. We recall that $\hat{f}_{sr}$ is a probability distribution that respects $E_{\hat{f}_{sr}}(q_{sr}) > E_p(q_{sr})$. Then, for sufficiently small ϵ,

$$R > I_{\tilde{f}}(X_1;Y_1|X_2) + I_{\tilde{f}}(X_1;X_2) \tag{20}$$

$$E_{\tilde{f}}(q_{sr}) > E_p(q_{sr}). \tag{21}$$

Using (20), (21) and the continuity of the divergence, we can find $\Delta > 0$, $\epsilon > 0$ and a neighborhood U of $\tilde{f}(x_1, x_2, y_1)$ such that for all $f \in U$ and $p'(x_2, y_1) \in N_p^\epsilon$, we have

$$R > D(f(x_1|x_2, y_1)\|p(x_1)|p'(x_2, y_1)) + \Delta \quad (22)$$

$$E_f(q_{sr}) > E_p(q_{sr}) + \Delta, \quad (23)$$

where $D(.\|.|.)$ is defined in [4].

Let V be a sufficiently small neighborhood of $p(x_1, x_2, y_1)$, such that for every $p'(x_1, x_2, y_1) \in V$, then $p'(x_2, y_1) \in N_p^\epsilon$ and

$$E_{p'}(q_{sr}) < E_p(q_{sr}) + \Delta. \quad (24)$$

Let us assume that the true message in block i is w_i and that the triple $(\mathbf{x}_1(w_i|s_i), \mathbf{x}_2(s_i), \mathbf{y}_1(i))$ has empirical type in V. If there exists another message $\tilde{w} \neq w_i$ such that the empirical type of $(\mathbf{x}_1(\tilde{w}|s_i), \mathbf{x}_2(s_i), \mathbf{y}_1(i))$ is in U, then a decoding error occurs.

Let $W(\tilde{w})$ take the value 1 if the empirical type of $(\mathbf{x}_1(\tilde{w}|s_i), \mathbf{x}_2(s_i), \mathbf{y}_1(i))$ is in U and 0 otherwise.

The expectation of $W = \sum_{\tilde{w} \neq w_i} W(\tilde{w})$ given $\mathbf{y}_1$ is

$$E(W|\mathbf{y}_1) = (2^{nR} - 1)\pi_0 \doteq 2^{nR}\pi_0, \quad (25)$$

where $\doteq$ denotes the behavior of the expression when $n \to \infty$ and $\pi_0 = E(W(\tilde{w}^*)|\mathbf{y}_1) = \Pr\{W(\tilde{w}^*) = 1|\mathbf{y}_1\}$ with $\tilde{w}^*$ being a random message different from w_i.

For two different messages $\tilde{w}$ and $\tilde{w}'$, the events $W(\tilde{w})$ and $W(\tilde{w}')$ are independent. Thus the variance of W given $\mathbf{y}_1$ is

$$\text{Var}(W|\mathbf{y}_1) = \sum_{\tilde{w} \neq w_i} \text{Var}(W(\tilde{w})|\mathbf{y}_1).$$

Since $W(\tilde{w})$ can only take the values 0 and 1, we can upper-bound $\text{Var}(W(\tilde{w})|\mathbf{y}_1) \leq E(W(\tilde{w})|\mathbf{y}_1)$. Thus

$$\text{Var}(W|\mathbf{y}_1) \leq \sum_{\tilde{w} \neq w_i} E(W(\tilde{w})|\mathbf{y}_1) \doteq 2^{nR}\pi_0. \quad (26)$$

Using (25), (26) and the fact that for any random variable X, $\Pr(X = 0) \leq \frac{\text{Var}(X)}{E(X)^2}$, we can write

$$\Pr(W = 0|\mathbf{y}_1) \dot{\leq} \frac{2^{nR}\pi_0}{(2^{nR}\pi_0)^2} = 2^{-nR}\frac{1}{\pi_0}. \quad (27)$$

Using the second part of Sanov's theorem, we obtain the asymptotic behavior $\pi_0 \doteq 2^{-n\tilde{R}_{sr}}$, where $\tilde{R}_{sr} = \min_{f \in \mathcal{U}} D(f(x_1|x_2, y_1)\|p(x_1)|p'(x_2, y_1))$. Using (22), we can then lower-bound $\pi_0 \dot{\geq} 2^{-n(R-\Delta)}$ and conclude that the probability of no decoding error tends to 0 when $n \to \infty$:

$$\Pr(W = 0|\mathbf{y}_1) \dot{\leq} 2^{-nR}2^{n(R-\Delta)} = 2^{-n\Delta}. \quad (28)$$

B. Decoding at destination

The second inequality can be dealt in two separate parts. Indeed, we have shown in the direct part that

$$R_0 \leq \min_{f \in \mathcal{D}_{rd}} I_f(X_2; Y) \quad (29)$$

$$R \leq R_0 + \min_{f \in \mathcal{D}_{sd}} I_f(X_1; Y|X_2) + I_f(X_1; X_2). \quad (30)$$

We thus have to show that if one of these inequalities is reversed, an error occurs with asymptotic probability one. This can be done using the same reasoning as in previous subsection.

VI. Matched Decoding Case

In the matched decoding case, i.e. $q_{sr}(x_1, x_2, y_1) = \log p(y_1|x_1, x_2)$, $q_{rd}(x_2, y) = \log p(y|x_2)$ and $q_{sd}(x_1, x_2, y) = \log p(y|x_1, x_2)$, the capacity coincides with the one of degraded relay computed by Cover and El Gamal in [7]. Indeed, for any distribution $f \in \mathcal{D}_{sr}$, we have

$$\begin{aligned}
&I_f(X_1; Y_1|X_2) + I_f(X_1; X_2) \\
&\geq I_f(X_1; Y_1|X_2) \\
&= H(Y_1|X_2) - H_f(Y_1|X_1, X_2) && (31) \\
&\geq H(Y_1|X_2) + \sum_{x_1, x_2, y_1} f(x_1, x_2, y_1) \log p(y_1|x_1, x_2) && (32) \\
&\geq H(Y_1|X_2) + \sum_{x_1, x_2, y_1} p(x_1, x_2, y_1) \log p(y_1|x_1, x_2) && (33) \\
&= I(X_1; Y_1|X_2),
\end{aligned}$$

where (31) holds because $f(x_2, y_1) = p(x_2, y_1)$, (32) follows from the non-negativity of the divergence and (33) is obtained using $E_f(\log p(y_1|x_1, x_2)) \geq E(\log p(y_1|x_1, x_2))$.

Moreover, by choosing $f(x_1, x_2, y_1) = p(x_1)p(x_2)p(y_1|x_1, x_2) \in \mathcal{D}_{sr}$, $I_f(X_1; X_2) = 0$ and $I_f(X_1; Y_1|X_2) + I_f(X_1; X_2) = I(X_1; Y_1|X_2)$. The bound is achievable, so $I_{sr}(p(x_1, x_2)) = I(X_1; Y_1|X_2)$.

In the same way, we can prove that $I_{rd}(p(x_1, x_2)) = I(X_2; Y)$ and $I_{sd}(p(x_1, x_2)) = I(X_1; Y|X_2)$.

Finally, in the matched case, the following rate is achievable

$$R = \min\{I(X_1; Y_1|X_2), I(X_2; Y) + I(X_1; Y|X_2)\} \quad (34)$$

$$= \min\{I(X_1; Y_1|X_2), I(X_1, X_2; Y)\}. \quad (35)$$

Acknowledgement

This work was supported under Australian Research Council's Discovery Projects funding scheme (project no. DP0984950). The authors would also like to thank Prof. Lapidoth for useful discussions.

References

[1] I. Csiszàr and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inform. Theory*, vol. 27, no. 1, pp. 5–12, Jan. 1981.
[2] J. Hui, "Fundamental issues of multiple accessing," Ph.D. dissertation, M.I.T., 1983, chap. IV.
[3] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai Shitz, "On information rates for mismatched decoders," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.
[4] A. Lapidoth, "Mismatched decoding and the multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 42, no. 5, pp. 1439–1452, Sep. 1996.
[5] A. Sendonaris, E. Erkip, and B. Aazhang, "User Cooperation Diversity. Part I and II," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1948, Nov. 2003.
[6] A. Nosratinia and A. Hedayat, "Cooperative Communication in Wireless Networks," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.
[7] T. Cover and A. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inform. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.
[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.