

Article

The Maximum Error Probability Criterion, Random Encoder, and Feedback, in Multiple Input Channels

Ning Cai

The State Key Lab. of ISN, Xidian University, No. 2 Taibai South Road, Xi'an 710071, China; E-Mail: caining@mail.xidian.edu.cn; Tel.: +86-(0)29-88204275

Received: 30 October 2013; in revised form: 10 December 2013 / Accepted: 24 December 2013 / Published: 25 February 2014

Abstract: For a multiple input channel, one may define different capacity regions, according to the criterions of error, types of codes, and presence of feedback. In this paper, we aim to draw a complete picture of relations among these different capacity regions. To this end, we first prove that the average-error-probability capacity region of a multiple input channel can be achieved by a random code under the criterion of maximum error probability. Moreover, we show that for a non-deterministic multiple input channel with feedback, the capacity regions are the same under two different error criterions. In addition, we discuss two special classes of channels to shed light on the relation of different capacity regions. In particular, to illustrate the roles of feedback, we provide a class of MAC, for which feedback may enlarge maximum-error-probability capacity regions, but not average-error-capacity regions. Besides, we present a class of MAC, as an example for which the maximum-error-probability capacity regions are strictly smaller than the average-error-probability capacity regions (first example showing this was due to G. Dueck). Differently from G. Dueck's enlightening example in which a deterministic MAC was considered, our example includes and further generalizes G. Dueck's example by taking both deterministic and non-deterministic MACs into account. Finally, we extend our results for a discrete memoryless two-input channel, to compound, arbitrarily varying MAC, and MAC with more than two inputs.

Keywords: average/maximum probability of error; random/deterministic encoder; feedback; multiple input channel; Chernoff bound

1. Introduction

The behavior of a multiple user channel is quite different from that of a point-to-point channel. For example, it is well known that the choice of the error criterion, *i.e.*, the *average* or *maximum* probability of error, makes no difference for the capacity of a point-to-point channel; whereas G. Dueck [1] showed that, maximum-error-probability capacity regions of multiple access channel (MAC), two way channel (TWC), and interference channel (IC) can be strictly smaller than their average-error-probability capacity regions. It is worth mentioning that this had been pointed out in an earlier work by R. Ahlswede [2]. Nevertheless, the two criterions may lead to the same capacity region for an MAC as well, as such an example was presented by P. Vanroose in [3].

The difference of the two capacity regions could be intuitively explained as follows. For a channel with two inputs, two senders may not always well cooperate (although in most cases they may cooperate well). If the criterion of maximum probability of error is used, the worst case has to be counted. As a direct consequence, the capacity region in this case may be strictly smaller than the one under the criterion of average error probaility (where only the average case is considered). This also explains why for a broadcast channel (BC), which is a channel with one single input, the maximum-error-probability capacity region is always equal to the average-error-probability capacity region [4] (p.293).

In network information theory, a lot of excellent work has been done on the criterion of average error probability (for example c.f. [5]). However, the progress of research on the criterion of maximum error probability seems to be relatively slow. As a comparison, for a discrete memoryless MAC, to determine its maximum-error-probability capacity region is still a challenging open problem even today; whilst its average-error-probability capacity region was completely determined in early of 70's of the last century [6,7].

Another interesting issue is *feedback*. It is well known that feedback may not increase the capacity of a point-to-point memoryless channel [8]; whilst it may enlarge the capacity region of a multiple user channel (e.g., [9–14]). Remarkably, feedback may not increase the capacity region of a physical degraded BC [15], as to a point-to-point channel. Moreover, the capacity region of a multiple-user channel with feedback also depends on the error criterion used. As shown in Dueck's example in [1], the maximum-error-probability capacity region of an MAC, even when feedback is present, may be strictly smaller than its average-error-probability region without feedback. Therefore, for a MAC with feedback, the capacity regions under two different error criterions may be different. To clear the clouds, an investigation on the difference made by error criterions in coding theorems for multiple user channels with feedback, is necessary.

Finally, let us take a look at the impact on the capacity regions made by different types of codes, *i.e., deterministic* and *random* codes. If the criterion of average probability of error is used, obviously the capacity regions of a channel for both deterministic codes and random codes are the same. Actually, the proofs of most direct coding theorems are based on this fact. So, when speaking of average-error-probability capacity region, (without or with feedback,) one does not have to distinguish which types of codes are employed. As a direct consequence, a point-to-point channel has only one capacity, no matter which criterion of error is considered, whether feedback is present or not, and which type of codes are employed. And, the same applies to a BC without feedback. However, this is not

applicable to an MAC, IC, or TWC, since their maximum- and average-error-probability regions might be different when deterministic codes are employed. Therefore, there is no reason to believe that for them, random codes and deterministic codes have the same maximum-probability-error capacity region. In fact, we shall see in this paper, that random codes may have a larger maximum-error-probability capacity region than deterministic codes.

In general, for an MAC or IC, we can define the following 6 capacity regions:

- average-error-probability capacity regions, without and with feedback;
- maximum-error-probability capacity regions of random codes, without and with feedback;
- maximum-error-probability capacity regions of deterministic codes, without and with feedback.

Similarly, when speaking of capacity region of a TWC, we must distinguish which type of codes and error criterion are used.

We observe that the relations of these capacity regions of MAC, IC, and TWC are similar, due to the fact that they all have multiple inputs. We refer to them as multiple input channels. The goal of the paper is to clarify the relation of the capacity regions of multiple input channels, but not to determine them. To simplify the discussion, we first assume that all channels have two inputs, and then extend them to the general case.

First, we show that the average-error-probability capacity region for a multiple input channel may be achieved by a random code under the criterion of maximum probability of error, no matter whether feedback is present or not. Thus, we reduce the above "6 capacity regions" to 4 of them. By definition, a random code for a two-input channel may have two random encoders, or one random encoder and one deterministic encoder. Notice that a deterministic encoder can be considered as a "special random encoder". Therefore, capacity region of random codes with two random encoders may not be smaller than that of random codes with one random encoder and one deterministic encoder, no matter maximum or average error criterion to be considered. For the same reason, no matter which criterion of error to be used, the capacity region of random codes with one random encoder and one deterministic encoder may not be smaller than the capacity region of deterministic codes. On the other hand, for random codes with two random encoders, the maximum-error-probability capacity region may not be larger than its average-error-probability capacity region, which is equal to average-error-probability capacity region of deterministic codes. We propose a coding scheme, which achieves the average-error-probability capacity region by employing a random code under the criterion of maximum probability of error. In our coding scheme, randomization at one encoder is sufficient. Therefore, the maximum-error-probability capacity regions of random codes with two random encoders, and one random encoder and one deterministic encoder must be the same.

Consider the remaining 4 capacity regions. Recall that for a MAC which employs deterministic codes, feedback may enlarge the average- and maximum-error-probability capacity regions; and the maximum-error-probability capacity region may be strictly smaller than the average-error-probability capacity region. In particular, G. Dueck [1] gave an example of a deterministic channel and showed that, the maximum-error-probability capacity region of a MAC with feedback or TWC may be strictly contained by its average-error-probability capacity region, even when the feedback is absent. So, it nature for us to ask whether maximum-error-probability capacity region for a multiple input channel with

feedback may be larger than the average-error-probability capacity region for the same channel when the feedback is absent. This motivate us to study maximum-error-probability capacity region with feedback. In contrast to Dueck's example, in this paper, we demonstrate that, *for a non-deterministic discrete memoryless or Gaussian MAC and IC with feedback, the average- and maximum-error-probability capacity regions are always the same, as well as for a TWC.* This tells us that only the deterministic channel may possibly have the properties in [1]. By combining our result and Cover-Leung bound [10], it is not difficult to find a MAC such that its maximum-error-probability capacity region with feedback is strictly larger than its average-error-probability capacity region without feedback. As a conclusion, we obtain a complete picture about the relation of the capacity regions of a multiple input channel.

- The choice of criterion of error makes no difference for random codes.
- The maximum-error-probability capacity region for deterministic codes without feedback is contained by all the other capacity regions; and, the contained relation can be strict.
- The average-error-probability capacity region with feedback contains all the other capacity regions; and, the containing relation can be strict.
- The maximum-error-probability capacity region for deterministic codes with feedback may be strictly larger than, or strictly smaller than the average-error-probability capacity region without feedback.
- For deterministic codes with feedback, if the channel is deterministic, the choice of criterion of error makes a difference; whereas if the channel is non-deterministic, it makes no difference.

When deterministic codes are employed, feedback may enlarge both the average- and maximum-error-probability capacity regions. However, the reasons may be different. To illustrate it, we provide a class of MAC, for which feedback may enlarge the maximum-error-probability capacity regions, but not the average-error-probability capacity regions. As *the contraction channel* (the channel used by G. Dueck in [1] to show that the maximum-error-probability capacity region), is deterministic, we are especially interested in having a non-deterministic example. For this purpose, *we generalize the contraction channel to a class of MAC, which contains both deterministic and non-deterministic MACs, and show that, for all channels in the class, maximum-error-probability capacity regions are strictly smaller than average-error-probability capacity regions.*

Moreover, we extend above results to compound channels, arbitrarily varying channels, and channels with more than two inputs. It turns out that for a random code under the criterion of maximum probability of error, to achieve the average-error-probability capacity region of an arbitrarily varying MAC, a randomization at one encoder is not sufficient any longer; and it is necessary to have randomization at both encoders. This leads us to the 3rd capacity region of an arbitrarily varying MAC, which we shall call *semi-average-*error-probability capacity region.

The rest of the paper is organized as follows. In the next section, we briefly describe our problems and review the contraction channel which was introduced in [1]. In Section 3, we prove that the average-error-probability capacity region of a multiple input channel can be achieved by a random code under the criterion of maximum probability of error. In Section 4, we show that for a non-deterministic discrete memoryless channel with feedback, and a Gaussian channel with feedback,

the maximum-error-probability capacity region is always equal to the average-error-probability capacity region. To better illustrate the results of random coding and coding with feedback, we provide two examples in Section 5. In Section 6, the results are extended to compound channels, arbitrarily varying channels, and channels with more than two inputs. Finally we conclude the paper by a brief discussion in Section 7.

2. Problem Description and Previous Work

2.1. Channel Codes and their Error Criterion

Consider a MAC, which has two inputs and one output. Two senders aim to send messages from their own message sets \mathcal{M}_i , i = 1, 2 to a receiver, who accesses the output of the channel. A code C for the MAC is specified by its encoding functions ϕ_i , i = 1, 2 and decoding function ψ . Note that the encoding functions ϕ_i , i = 1, 2 are two mappings from the message sets \mathcal{M}_i , i = 1, 2 to the two input alphabets of the channel, respectively; whilst the decoding function ψ is a mapping from the output space of the channel to the message set $\mathcal{M}_1 \times \mathcal{M}_2$.

Consider an IC, which has two inputs and two outputs. Two senders aim to send their own messages from sets \mathcal{M}_1 and \mathcal{M}_2 to two receivers, respectively. Similarly, a code C for the IC is specified by its the encoding functions ϕ_i , i = 1, 2 and the decoding functions ψ_j , j = 1, 2. In general case, the capacity region of an IC remains unknown; and, the best inner bound is due to T. S Han and K. Kobayashi [16].

For a MAC/IC, we say that feedback from the output $\mathbf{y}^{(j)}$ (for a MAC, the output index j = 1; whilst for an IC, $j \in \{1, 2\}$) is available at the encoder i for $i \in \{1, 2\}$, if the codeword sent by the *i*-th sender, not only depends on the message m_i , but also the output $\mathbf{y}^{(j)}$ causally. I. e., the codeword can be written as

$$\phi_i(m_i, \mathbf{y}^{(j)}) = (\phi_{i,1}(m_i), \phi_{i,2}(m_i, y_1^{(j)}), \dots, \phi_{i,n}(m_i, y_1^{(j)}, y_2^{(j)}, \dots, y_{n-1}^{(j)}))$$

where $\mathbf{y}^{(j)} = (y_1^{(j)}, y_2^{(j)}, \dots, y_n^{(j)})$; and $\phi_{i,t}(m_i, y_1^{(j)}, y_2^{(j)}, \dots, y_{t-1}^{(j)})$ is the symbol sent by the sender *i* at the time *t*.

A TWC has two inputs and two outputs, and is used by two users for exchanging messages. The first user sends messages to the first input of the channel, and receives messages from the second output of the channel; and the second user sends messages to the second input, and receives messages from the first output. The symbol sent by a user at the time t depends on not only the message that he/she wants to send, but also the output that he/she has received before the time t. A user decodes the messages according to not only the output that he/she received, but also the message sent by him/her. TWC was introduced by C. E. Shannon in his milestone paper [17]. It was actually the first model in network information theory. In the case that the users encode only according to their own messages, respectively, but not the outputs received by them, we refer the channel as a *TWC without feedback*. In other words, a TWC without feedback is an "IC", for which his/her own message of a user can be used in decoding.

To distinguish MAC, IC, and TWC from point-to-point channels and BC, we refer them as multiple input channels. Note that in general a MAC may have more than two inputs. In the following, we first consider a MAC with only two inputs; and results obtained will be extended to the general case later.

It is well known that, a discrete channel $W^n : \mathcal{X}_1^n \times \mathcal{X}_2^n \to \mathcal{Y}^n$ for a MAC (or $W^n : \mathcal{X}_1^n \times \mathcal{X}_2^n \to \mathcal{Y}_1^n \times \mathcal{Y}_2^n$ for an IC/TWC), is *memoryless* if for all $\mathbf{x}^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}) \in \mathcal{X}_1^n, \mathbf{x}^{(2)} = (x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)}) \in \mathcal{X}_2^n$, and all $\mathbf{y} = (y_1, y_2, \dots, y_n) \in \mathcal{Y}^n$ (or all $\mathbf{y}^{(1)} = (y_1^{(1)}, y_2^{(1)}, \dots, y_n^{(1)}) \in \mathcal{Y}_1^n, \mathbf{y}^{(2)} = (y_1^{(2)}, y_2^{(2)}, \dots, y_n^{(2)}) \in \mathcal{Y}_2^n$)

$$\begin{split} W^n(\mathbf{y}|\mathbf{x}^{(1)},\mathbf{x}^{(2)}) &= \prod_{t=1}^n W(y_t|x_t^{(1)},x_t^{(2)}) & \text{for a MAC} \\ (\text{ or } W^n(\mathbf{y}^{(1)},\mathbf{y}^{(2)}|\mathbf{x}^{(1)},\mathbf{x}^{(2)}) &= \prod_{t=1}^n W(y_t^{(1)},y_t^{(2)}|x_t^{(1)},x_t^{(2)}) & \text{for an IC/TWC}) \end{split}$$

where \mathcal{X}_i , i = 1, 2 are input alphabets; $\mathcal{Y}(\mathcal{Y}_j, j = 1, 2)$ is (are) output alphabet(s) of the channel. We call W the *generic* of the channel. For simplicity, we call the discrete memoryless channel with generic W, the discrete memoryless channel W in the sequel. With a slight abuse of notation, we also denote a channel not necessary to be discrete memoryless, abstractly by W.

For a multiple input channel, let $p_e(C, (m_1, m_2))$ be the *error probability* of a given code C for the message pair $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$ (*i.e.*, the probability, that the decoder of the MAC, or at least one decoder of the IC or TWC, does not correctly decode, when the messages m_1 and m_2 are sent simultaneously through the channel). Then the *average probability of error* and *maximum probability of error* of the code are defined as

$$p_a(C) := \frac{1}{M_1} \frac{1}{M_2} \sum_{(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} p_e(C, (m_1, m_2))$$
(1)

$$p_m(C) := \max_{(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} p_e(C, (m_1, m_2))$$
 (2)

respectively, where $M_i = |\mathcal{M}_i|, i = 1, 2$.

We say that an encoder *i* is *random*, if it can choose the codewords randomly, where $i \in \{1, 2\}$. That is, in the case that feedback is absent, the encoder *i* first generates a random key K_i , and then chooses a codeword according to the message to be sent, and the outcome of the random key; whilst in the case that feedback is present, the choice of codeword is additionally according to the received output of channel via the feedback causally. If the *i*-th encoder is random, accordingly we denote its encoding function by capital Greek Φ_i .

In this paper, we assume that the random key is uniformly distributed. A random code may have two random encoders, or one random encoder and one deterministic encoder. If a random code has two random encoders, we assume that the two random keys K_1 and K_2 are independent. We shall see in the next section that the number of random encoders makes no difference in the capacity regions of an (ordinary) multiple input channel. However, we shall see in Subsection 6.2, that random codes with two random encoders, and random codes with one random encoder and one deterministic encoder may have different capacity regions for an arbitrarily varying MAC. Thus, we do not need to distinguish the number of random encoders, when speaking of the capacity region, until Subsection 6.2.

For a random code C_r , the error probability $p_e(C_r, (m_1, m_2))$ for a fixed message pair (m_1, m_2) is a function of the random keys, and thus is a random variable. Therefore, it must be replaced by

its expectation in the definitions of error probabilities. That is, the *average probability of error* and *maximum probability of error* of the random code C_r are defined as

$$p_a(C_r) := \frac{1}{M_1} \frac{1}{M_2} \mathbf{E} \sum_{(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} p_e(C_r, (m_1, m_2))$$
(3)

$$p_m(C_r) := \max_{(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} \mathbf{E} p_e(C_r, (m_1, m_2))$$
(4)

respectively, where E is operator of expectation. Obviously random codes may not have a capacity region smaller than deterministic codes, because by definition, a deterministic code is a "special random code".

The capacity regions of a multiple input channel are defined in a standard way. According to the presence of feedback, criterions of probabilities of error, and types of codes (random or deterministic), formally one might have 8 capacity regions.

- As a matter of fact, every random code such that its average probability of error defined by (3) is smaller than a given number, has a realization with an average probability of error (defined in Equation (1) smaller than the same number. So, the type of codes makes no difference, if the criterion of the average error probability is employed. Therefore, we may simply speak of average-error-probability capacity region, and do not need to distinguish the type of codes.
- For the criterion of the average probability of error, we have two capacity regions for a given channel W. We denote them by $\bar{\mathcal{R}}(W)$ and $\bar{\mathcal{R}}_f(W)$, respectively, according to whether feedback is absent or present.
- For the criterion of maximum probability of error, one may define capacity regions of random codes without and with feedback, denoted by $\mathcal{R}_r(W)$ and $\mathcal{R}_{r,f}(W)$, respectively; and, capacity regions of deterministic codes without and with feedback, denoted by $\mathcal{R}_d(W)$ and $\mathcal{R}_{d,f}(W)$, respectively.
- As a special kind of random codes, deterministic codes may not have larger capacity regions than random codes; and, feedback may not reduce capacity regions. Therefore, relationship of the six capacity regions are presented as follows.

$$\begin{array}{rcl}
\mathcal{R}_{d}(W) &\subset & \mathcal{R}_{d,f}(W) \\
\cap & & \cap \\
\mathcal{R}_{r}(W) &\subset & \mathcal{R}_{r,f}(W) \\
\cap & & \cap \\
\bar{\mathcal{R}}(W) &\subset & \bar{\mathcal{R}}_{f}(W)
\end{array}$$
(5)

Remarks 2.1 We have the following observations:

- (1) By definition, a multiple input channel may have different kinds of feedback; and, its capacity regions may depend on the kinds of feedback. Nevertheless, we denote the capacity regions with different kinds of feedback by the same notation. It makes no problem in this paper, since we are interested in the relation between the capacity regions, but not the capacity regions themselves. And, the kinds of feedback are also clear by the context.
- (2) Let C_r be a random code with a random encoder Φ_1 and a deterministic encoder ϕ_2 for a multiple input channel. Then $\{\Phi_1(m_1) : m_1 \in \mathcal{M}_1\}$ is not necessary to be a set of independent random

variables. However, one can obtain a random code C'_r with the same rate, the same average and maximum probabilities of error for the same channel, from C_r , such that $\Phi'_1(m_1), m_1 \in \mathcal{M}_1$ are independent. The reasons are the followings:

- (1) For Φ'_1 , one can simply choose values of $\Phi'_1(m_1)$ for all $m_1 \in \mathcal{M}_1$, randomly and independently according to marginal distribution of $\Phi_1(m_1)$, and keep the deterministic encoder and the decoder(s) unchanged.
- (2) Then $\{p_e(C'_r, (m_1, m_2)), m_i \in \mathcal{M}_i\}$ is a set of independent random variables with the same marginal distribution as $p_e(C_r, (m_1, m_2))$, for all $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$. That is, we have that

$$\mathbf{E}p_e(C'_r, (m_1, m_2)) = \mathbf{E}p_e(C_r, (m_1, m_2))$$

(3) Thus, without loss of generality, one may assume that $\{\Phi_1(m_1), m_1 \in \mathcal{M}_1\}$ and $\{p_e(C_r, (m_1, m_2)), m_1 \in \mathcal{M}_1\}$ for $m_2 \in \mathcal{M}_2$, are sets of independent random variables.

By the same reason, for a random code with two random encoders Φ_i , i = 1, 2, we may also assume $\Phi_i(m_i), m_i \in \mathcal{M}_i, i = 1, 2$ are independent and $p_e(C_r, (m_1, m_2)), m_1 \in \mathcal{M}_1$ are conditionally independent, given $\Phi_2(m_2) = \mathbf{x}^{(2)}$.

(3) We may also assume that the decoder (decoders) of a random code accesses (access) the random key(s) generated by encoder(s), and decodes (decode) according to the outcome(s) of the key(s). In this case, the decoder(s) is (are) random. The code is often referred as a correlated random code in coding theorems of arbitrarily varying channels. One may have a random code with deterministic decoder(s), from a correlated random code with vanishing key rate(s), by sending the outcome(s) of random key(s) to decoder(s), if the capacity region of deterministic codes has non-empty topological interior.

2.2. Dueck's Contraction Channel

In [1] G. Dueck showed by an example that, for MAC and TWC, maximum-error-probability capacity regions can be strictly smaller than their average-error-probability capacity regions. Without being pointed out explicitly in [1], the same example also implies that the maximum-error-probability capacity region of an IC may be strictly contained in its average-error-probability capacity region.

Dueck's contraction channel is a discrete memoryless channel with a generic $W : \mathcal{X}_1 \times \mathcal{X}_2 \rightarrow \mathcal{Y}_1 \times \mathcal{Y}_2$, where

$$\mathcal{X}_1 := \{a(0), b(0), a(1), b(1)\}$$
$$\mathcal{X}_2 := \{0, 1\}$$
$$\mathcal{Y}_1 := \{a(0), b(0), e(0), a(1), b(1), e(1)\}$$
$$\mathcal{Y}_2 := \{0, 1\}$$

Let g be a function from $\mathcal{X}_1 \times \mathcal{X}_2$ to \mathcal{Y}_1 , such that

$$g(a(0), 0) = g(b(0), 0) = e(0)$$

 $g(a(1), 0) = a(1)$

g(b(1), 0) = b(1) g(a(0), 1) = a(0) g(b(0), 1) = b(0)g(a(1), 1) = g(b(1), 1) = e(1)

Let W_1 be a channel $W_1 : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}_1$, such that for all $x_i \in \mathcal{X}_i, i = 1, 2$ and $y_1 \in \mathcal{Y}_1$, $W_1(y_1|x_1, x_2) = 1$ if and only if $g(x_1, x_2) = y_1$; and W_2 be the identity channel on $\{0, 1\}$, *i.e.*, for all $x_2 \in \mathcal{X}_2$ and $y_2 \in \mathcal{Y}_2$, $W_2(y_2|x_2) = 1$ if and only if $y_2 = x_2$. Then the generic channel $W : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}_1 \times \mathcal{Y}_2$ can be represented as $W(y_1, y_2|x_1, x_2) = W_1(y_1|x_1, x_2)W_2(y_2|x_2)$, for $x_i \in \mathcal{X}_i, i = 1, 2, y_j \in \mathcal{Y}_j, j = 1, 2$.

Dueck [1] considered the following scenarios:

(1) **TWC**

- (1.w) *TWC without feedback*, or IC: There are two users in the communication system, say user 1 and user 2. The user 1 (user 2) accesses the first (second) message set \mathcal{M}_1 (\mathcal{M}_2) and the second (first) output of the channel with the alphabet \mathcal{Y}_2^n (\mathcal{Y}_1^n), and sends messages to the first (second) input of the channel with alphabet \mathcal{X}_1^n (\mathcal{X}_2^n). No feedback is available. It is easy to see that, for this channel, their own messages may not help them to decode. So TWC and IC have the same capacity region.
- (1.f) *TWC with feedback*: In addition to (1.w), the feedbacks from the second output to the first encoder and from the first output to the second encoder are available. That is, the channel is used as the standard TWC in Shannon's sense [17].

(2) MAC

- (2.w) MAC without feedback: Two senders send their messages to two inputs of the channel independently; and a single receiver, who wants to decode both messages, accesses both outputs. That is, by combining two outputs as one single output, we use the channel as an MAC. No feedback is available.
- (2.f) *MAC with feedback*: In addition to (1.w), the feedbacks from the output to both encoders are available.

Notice that a codeword $\mathbf{x}^{(2)}$, which is sent to the second input of the contraction channel, always can be recovered from both outputs correctly with probability 1; and, the second output is a function of the first output of the channel. Therefore, one may assume that the receiver only accesses the first output of channel in scenario (2). Moreover, a code is decodable in scenario (1) if and only if it is decodable in scenario (2). Thus, its capacity regions in scenarios (1.w) and (2.w) are the same; and, its capacity regions in scenarios (1.f) and (2.f) are the same.

Let \mathcal{R} be the average-error-probability capacity region of channels in scenarios (1) and (2). Define

$$\mathcal{R}_{out} := \{ (R_1, R_2) : R_1 \le h(\frac{1}{3}) + \frac{2}{3} - p, R_2 \le h(p), p \in [0, \frac{1}{2}] \}$$
(6)

where $h(\cdot)$ is the binary entropy function. Dueck [1] proved that the maximum-error-probability capacity regions in all the above four scenarios are contained by \mathcal{R}_{out} , which is strictly smaller than $\overline{\mathcal{R}}$.

3. Randomization at One Encoder Enhances Codes with Maximum Error Probability

In this section, we show that randomization at one encoder may enhance a code under the criterion of maximum error probability for a multiple input channel, without or with feedback, to achieve the average-error-probability capacity region. That is, for a multiple input channel W, $\mathcal{R}_r(W) = \bar{\mathcal{R}}(W)$ and $\mathcal{R}_{r,f}(W) = \bar{\mathcal{R}}_f(W)$.

First, let us recall Chernoff bound which will be applied in this paper repeatedly.

Lemma 3.1 Let $V_1, V_2, \ldots V_L$ be L independent random variables such that $\mathbf{E}V_l \leq \alpha < \beta/3$ and $0 \leq V_l \leq 1$ with probability 1, for $l = 1, 2, \ldots, L$, and positive real numbers α and β . Then

$$\Pr\{\sum_{l=1}^{L} V_l > L\beta\} \le e^{-L(\beta - 3\alpha)}$$
(7)

Proof:

$$\Pr\{\sum_{l=1}^{L} V_{l} > L\beta\} = \Pr\{e^{-L\beta + \sum_{l=1}^{L} V_{l}} > 1\}$$

$$\leq e^{-L\beta} \mathbf{E} e^{\sum_{l=1}^{L} V_{l}} \stackrel{(a)}{=} e^{-L\beta} \prod_{l=1}^{L} \mathbf{E} e^{V_{l}}$$

$$\stackrel{(b)}{\leq} e^{-L\beta} \prod_{l=1}^{L} \mathbf{E} [1 + V_{l} + \frac{e}{2} V_{l}^{2}] \stackrel{(c)}{\leq} e^{-L\beta} \prod_{l=1}^{L} \mathbf{E} [1 + (1 + \frac{e}{2}) V_{l}]$$

$$\leq e^{-L\beta} \prod_{l=1}^{L} (1 + 3\mathbf{E} V_{l}) \stackrel{(d)}{\leq} e^{-L\beta} (1 + 3\alpha)^{L}$$

$$\stackrel{(e)}{\leq} e^{-L(\beta - 3\alpha)}$$

where the equality in (a) holds because V_1, V_2, \ldots, V_L are independent; (b) follows from the inequality $e^x \leq 1 + x + \frac{e}{2}x^2$ for $x \in [0, 1]$; (c) holds because $V_l \leq 1$ with probability 1; (d) follows from the assumption that $\mathbf{E}V_l \leq \alpha$; and (e) follows from the inequality $1 + x \leq e^x$. That is (7).

Theorem 3.2 Let W be a multiple input channel. Then

$$\mathcal{R}_r(W) = \bar{\mathcal{R}}(W) \tag{8}$$

provided that $\overline{\mathcal{R}}(W)$ has a non-empty topological interior; and

$$\mathcal{R}_{r,f}(W) = \bar{\mathcal{R}}_f(W) \tag{9}$$

provided that $\overline{\mathcal{R}}_f(W)$ has a non-empty topological interior.

Moreover, $\overline{\mathcal{R}}(W)$ has a non-empty topological interior if and only if $\mathcal{R}_r(W)$ has a non-empty topological interior; and, $\overline{\mathcal{R}}_f(W)$ has a non-empty topological interior if and only if $\mathcal{R}_{r,f}(W)$ has a non-empty topological interior.

Proof: Since $\mathcal{R}_r(W) \subset \overline{\mathcal{R}}(W)$ and $\mathcal{R}_{r,f}(W) \subset \overline{\mathcal{R}}_f(W)$, it is sufficient to show that, (1) $\overline{\mathcal{R}}(W)$ is achievable by a random code with maximum error probability and without feedback, if it has a non-empty topological interior; and, (2) $\overline{\mathcal{R}}_f(W)$ is achievable by a random code with maximum error probability and with feedback, if it has a non-empty topological interior. Both are proven in the same way.

Let $\overline{\mathcal{R}}(W)$ or $\overline{\mathcal{R}}_f(W)$ have a non-empty topological interior; (R_1, R_2) be a rate pair in $\overline{\mathcal{R}}(W)$ or $\overline{\mathcal{R}}_f(W)$, and $\lambda, \epsilon > 0$. Then, for a sufficiently large n, there exists a deterministic code C of length n, with the encoding function $\phi_i, i = 1, 2$, suitable decoder(s), and rate pair (at least) $(R_1 - \frac{\epsilon}{2}, R_2 - \frac{\epsilon}{2})$, such that the average probability of error

$$\frac{1}{|\mathcal{M}_1|} \frac{1}{|\mathcal{M}_2|} \sum_{(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} p_e(C, (m_1, m_2)) < \frac{1}{12}\lambda$$
(10)

where \mathcal{M}_i , i = 1, 2, are the two message sets. Let

$$Q_e(m_2) = \frac{1}{|\mathcal{M}_1|} \sum_{m_1 \in \mathcal{M}_1} p_e(C, (m_1, m_2))$$
(11)

for $m_2 \in \mathcal{M}_2$; and,

$$\mathcal{M}_{2,0} = \{ m_2 : Q_e(m_2) < \frac{1}{6}\lambda, m_2 \in \mathcal{M}_2 \}$$
(12)

and $\mathcal{M}_{2,1} = \mathcal{M}_2 \setminus \mathcal{M}_{2,0}$. Then it follows from (10) and Markov inequality that

$$\frac{1}{12}\lambda > \frac{1}{|\mathcal{M}_2|} \sum_{m_2 \in \mathcal{M}_2} Q_e(m_2) \ge \frac{|\mathcal{M}_{2,1}|}{|\mathcal{M}_2|} \frac{\lambda}{6}$$

This gives us $|\mathcal{M}_{2,1}| < \frac{1}{2}|\mathcal{M}_2|$ or $|\mathcal{M}_{2,0}| > \frac{1}{2}|\mathcal{M}_2|$. By using the same encoding functions and decoding function(s), while deleting messages in $\mathcal{M}_{2,1}$, we have a subcode C_0 of C, such that for all $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_{2,0}$, $p_e(C_0, (m_1, m_2)) = p_e(C, (m_1, m_2))$.

Next we construct a code with a random encoder and a deterministic encoder under the criterion of maximum error probability. To this end, we let Σ be the permutation group on \mathcal{M}_1 ; and, $\tilde{\sigma}_k, k = 1, 2, \ldots n^2$, be i.i.d. random permutations uniformly distributed on Σ , where *n* is the length of code *C*. Then for all $m_1, m'_1 \in \mathcal{M}_1$, and all *k*,

$$\Pr(\tilde{\sigma}_k(m_1) = m_1') = \frac{(|\mathcal{M}_1| - 1)!}{|\mathcal{M}_1|!} = |\mathcal{M}_1|^{-1}$$
(13)

For a $\sigma \in \Sigma$, we define $\sigma(C_0)$ as a code with the same message sets as C_0 , encoding functions $\phi_{1,\sigma}(\cdot) = \phi_1(\sigma(\cdot))$ and $\phi_{2,\sigma}(\cdot) = \phi_2(\cdot)$, and a suitable modification in decoder(s) (*i.e.*, by decoding the first message to be $\sigma^{-1}(m'_1)$ if the decoding result in C_0 is m'_1). Then, we have that $p_e(\sigma(C_0), (m_1, m_2)) = p_e(C_0, (\sigma(m_1), m_2))$, for all $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_{2,0}$. Thus by Equation (13), we have that for all $k \in \{1, 2, \ldots, n^2\}, (m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_{2,0}$,

$$\mathbf{E}p_{e}(\tilde{\sigma}_{k}(C_{0}), (m_{1}, m_{2})) = \sum_{\substack{m_{1}' \in \mathcal{M}_{1}}} \Pr(\tilde{\sigma}_{k}(m_{1}) = m_{1}')p_{e}(C_{0}, (m_{1}', m_{2}))$$

$$\stackrel{(a)}{=} \sum_{\substack{m_{1}' \in \mathcal{M}_{1}}} \frac{1}{|\mathcal{M}_{1}|}p_{e}(C_{0}, (m_{1}', m_{2}))$$

$$\stackrel{(b)}{=} Q_e(m_2) \stackrel{(c)}{<} \frac{1}{6}\lambda \tag{14}$$

where (a) and (b) hold by Equations (13) and (11), respectively; and (c) follows from (12) and the fact $m_2 \in \mathcal{M}_{2,0}$. Consequently, by Chernoff bound, *i.e.*, Lemma 3.1 (taking $L = n^2, \alpha = \frac{\lambda}{6}, \beta = \frac{2\lambda}{3}$), and (14), we obtain that for all fixed $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_{2,0}$,

$$\Pr\{\frac{1}{n^2}\sum_{k=1}^{n^2} p_e(\tilde{\sigma}_k(C_0), (m_1, m_2)) \ge \frac{2}{3}\lambda\} \le e^{-\frac{n^2\lambda}{6}}$$
(15)

Thus, by union bound, we have that with a probability at least $1 - |\mathcal{M}_1 \times \mathcal{M}_{2,0}| e^{-\frac{n^2 \lambda}{6}} \to 1$ (as $n \to \infty$), $(\tilde{\sigma}_1, \tilde{\sigma}_2, \dots, \tilde{\sigma}_{n^2})$ has a realization $(\sigma_1, \sigma_2, \dots, \sigma_{n^2})$, such that for all $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_{2,0}$,

$$\frac{1}{n^2} \sum_{k=1}^{n^2} p_e(\sigma_k(C_0), (m_1, m_2)) < \frac{2}{3}\lambda$$
(16)

Now we are ready to construct a random code \tilde{C}_r with a random encoder and a deterministic encoder. Our code consists of two blocks.

We first prepare a code C fulfilling the criterion of average probability error, choose its subcode C_0 , and then find permutation groups $(\sigma_1, \sigma_2, \ldots, \sigma_{n^2})$, satisfying (16).

Encoders: The first sender uniformly at random generates a "key" K_1 from $\{1, 2, ..., n^2\}$, and then sends the outcome k_1 of K_1 in the first block, by using a code with *average* probability of error $\frac{\lambda}{6}$. At the same time, the second sender sends a suitably chosen dummy codeword to help the transmission. Then, the first sender sends the codeword $\phi_{1,\sigma_{k_1}}(m_1)$ through the channel in the second block, if he/she wants to send the message m_1 and the outcome of the key is k_1 . At the same time, to send the message m_2 , the second sender sends $\phi_{2,\sigma_{k_1}}(m_2)(=\phi_2(m_2))$ through the channel. That is, the two senders send their messages by using the code $\sigma_{k_1}(C_0)$ in the second block. Notice that the second sender does not need to know the outcome of the key, because his/her encoding function does not depend on the permutation σ_{k_1} . The ratio of the length of the first block in the whole length of the code can be arbitrarily small, when n is sufficiently large, because the key rate $\frac{\log n^2}{n}$ vanishes, as n increases. Thus, the rates of code are larger than $R_i - \epsilon$, i = 1, 2, respectively, if n is sufficiently large.

Decoder(s): At the receiver(s), first the key is decoded as \hat{k} ; then the second block is decoded by using the decoding function of the code $\sigma_{\hat{k}}(C_0)$.

Error Analysis: Let \mathcal{E}_0 be the event, of that an error occurs in the first block; and \mathcal{E}_0^c be its complement. Then, for all $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_{2,0}$,

$$\begin{aligned} \mathbf{E}p_{e}(\tilde{C}_{r},(m_{1},m_{2})) &= \mathbf{E}[\mathbf{E}(p_{e}(\tilde{C}_{r},(m_{1},m_{2}))|K_{1})] \\ &= \sum_{k_{1}=1}^{n^{2}}[\Pr(K_{1}=k_{1})\Pr(\mathcal{E}_{0}|K_{1}=k_{1})\mathbf{E}(p_{e}(\tilde{C}_{r},(m_{1},m_{2}))|K_{1}=k_{1},\mathcal{E}_{0}) \\ &+ \Pr(K_{1}=k_{1})\Pr(\mathcal{E}_{0}^{c}|K_{1}=k_{1})\mathbf{E}(p_{e}(\tilde{C}_{r},(m_{1},m_{2}))|K_{1}=k_{1},\mathcal{E}_{0}^{c})] \end{aligned}$$

$$\overset{(a)}{\leq} \sum_{k_{1}=1}^{n^{2}} \Pr(K_{1} = k_{1}) \Pr(\mathcal{E}_{0} | K_{1} = k_{1}) + \sum_{k_{1}=1}^{n^{2}} \Pr(K_{1} = k_{1}) \Pr(\mathcal{E}_{0}^{c} | K_{1} = k_{1}) \mathbf{E}(p_{e}(\tilde{C}_{r}, (m_{1}, m_{2})) | K_{1} = k_{1}, \mathcal{E}_{0}^{c})] \overset{(b)}{\leq} \frac{\lambda}{6} + \sum_{k_{1}=1}^{n^{2}} \Pr(K_{1} = k_{1}) \Pr(\mathcal{E}_{0}^{c} | K_{1} = k_{1}) \mathbf{E}(p_{e}(\tilde{C}_{r}, (m_{1}, m_{2})) | K_{1} = k_{1}, \mathcal{E}_{0}^{c})] = \frac{\lambda}{6} + \sum_{k_{1}=1}^{n^{2}} \Pr(K_{1} = k_{1}) \Pr(\mathcal{E}_{0}^{c} | K_{1} = k_{1}) p_{e}(\sigma_{k_{1}}(C_{0}), (m_{1}, m_{2}))] \overset{(c)}{\leq} \frac{\lambda}{6} + \frac{1}{n^{2}} \sum_{k_{1}=1}^{n^{2}} p_{e}(\sigma_{k_{1}}(C_{0}), (m_{1}, m_{2})) \overset{(d)}{\leq} \frac{5}{6}\lambda < \lambda$$

$$(17)$$

where (a) holds since $p_e(\tilde{C}_r(m_1, m_2)) \leq 1$ with probability 1; (b) holds since the average probability of error in the first block is no larger than $\frac{\lambda}{6}$; (c) follows from the code construction; (d) follows from (16). Thus we complete the proof of Theorem 3.2.

The part of the proof will be used in the next section and Section 6, for readers' convenience, we summary them as the following lemma:

Lemma 3.3 Let C be a deterministic code of length n for a multiple input channel, without or with feedback, with average probability of error $\lambda_1 = \frac{\lambda}{12}$, and a message set $\mathcal{M}_1 \times \mathcal{M}_2$.

- (1) Then, there is a subcode C_0 of C, with a message set $\mathcal{M}_1 \times \mathcal{M}_{2,0}$, such that $\mathcal{M}_{2,0} \subset \mathcal{M}_2$, $|\mathcal{M}_{2,0}| \geq \frac{|\mathcal{M}_2|}{2}$, and $Q_e(m_2) < 2\lambda_1$, for $Q_e(m_2)$ as defined in Equation (11), and $m_2 \in \mathcal{M}_{2,0}$.
- (2) Let $\tilde{\sigma}_k, k = 1, 2, ..., n^2$, be i.i.d. random permutations uniformly distributed on Σ , then Equation (15) holds for C_0 as described in 1).
- (3) Therefore, there are n^2 permutations $\sigma_k, k = 1, 2, ..., n^2$ on \mathcal{M}_1 such that (16) holds.

Next, let us look into the reason *why the average- and maximum-error-probability capacity regions of deterministic codes for a multiple input channel may be different.* The two senders of a multiple input channel may cooperate only in the choice of codebook, since they must choose codewords independently due to the independence of messages. The combinations of the codeword pairs sent by them may be "good" or "bad". Intuitively, comparing to the whole codebooks, the bad combinations are very few. Thus their contributions to the average probability of error can be ignored, when the codebooks are sufficiently large. On the other hand, if the criterion of maximum probability of error is used, one must consider the probability of error for "worst combination of codeword pairs". To guarantee the worst combinations of codeword pairs to be correctly decoded with a high probability, the capacity region for codes satisfying the criterion of maximum probability of error often must be reduced. The randomization at an encoder plays a similar role to probability algorithms in theory of computation, or mixed strategies in game theory, mixing the good and bad combinations. As a consequence, the average-error-probability capacity regions are achievable.

The proof of the theorem tells us that adding randomness with a vanishing rate at one single encoder is sufficient for a code fulfilling the criterion of maximum error probability to achieve the average-error-probability capacity region. That is, the cost for the randomization is very low. This suggests us to use random coding in the network communication, because in any sense a small maximum probability of error is better than a small average probability of error, although in most cases the latter is acceptable.

By Theorem 3.2, the 6 capacity regions in (5) can be actually reduced to 4 capacity regions. Namely, (5) can be represented by

$$\begin{array}{rcl}
\mathcal{R}_d(W) &\subset & \mathcal{R}_{d,f}(W) \\
\cap & & \cap \\
\bar{\mathcal{R}}(W)/\mathcal{R}_r(W) &\subset & \bar{\mathcal{R}}_f(W)/\mathcal{R}_{r,f}(W)
\end{array}$$
(18)

So far we have known that there is a MAC W, such that $\overline{\mathcal{R}}(W) \subsetneq \overline{\mathcal{R}}_f(W)$; and, there are MAC, IC, and TWC W, such that $\mathcal{R}_d(W) \subset \mathcal{R}_{d,f}(W) \subsetneq \overline{\mathcal{R}}(W)$. In the next section, we shall apply Lemma 3.3 to non-deterministic and Gaussian multiple input channels with feedback, and show that for all non-deterministic discrete memoryless and Gaussian multiple input channels W,

$$\mathcal{R}_{d,f}(W) = \bar{\mathcal{R}}_f(W) \tag{19}$$

By applying Cover-Leung inner bound [10], it is not hard to find a non-deterministic discrete memoryless MAC W, such that $\overline{\mathcal{R}}(W) \subsetneq \overline{\mathcal{R}}_f(W)$, which with (19) yields that $\overline{\mathcal{R}}(W) \subsetneq \mathcal{R}_{d,f}(W)$. Thus, we have a complete picture of the relation of the capacity regions of multiple input channels. That is, all contained relations in (18) may possibly be strict; and, $\mathcal{R}_{d,f}(W)$ may be strictly contained by, and strictly contain $\overline{\mathcal{R}}(W)$. Therefore, the relations in (18) may not be simplified further.

4. An Application to Feedback

In this section, we apply Lemma 3.3 to non-deterministic discrete memoryless and Gaussian multiple input channels with feedback.

An output of a multiple input channel is *deterministic*, if the output is a function of inputs. That is, for a deterministic output, no matter what input letters are input to the channel, at the output the channel always outputs a letter with probability 1. We say that the output is *non-deterministic*, if it is not deterministic. A multiple input channel is *non-deterministic*, if it has at least one non-deterministic output. Throughout this section, we assume that feedback from a non-deterministic output is available at least at one encoder, say the first encoder, if a multiple input channel is non-deterministic. Obviously, the contraction channel is deterministic. Here, by Gaussian channels we mean *discrete-time additive Gaussian noise channels* in the sense of [5] (93,137–138,438) or [18] ([Chapter 15]). It is easy to see from the following that, the result in Corollary 4.1 is true for all multiple input channels with the property (*) below, although for simplicity, we assume that the channel is non-deterministic discrete memoryless channels or Gaussian.

(*) There is a positive α such that for all $\beta > 0, \gamma > 0$ and a sufficiently large integer n_0 , there exist two codewords $\mathbf{x}^{(i)}, i = 1, 2$ of length n_0 , and a subset A in an output space, whose output is sent to the first encoder as a feedback, such that

• with probability at least $1 - \beta$, the output sequence falls in A;

Typically, we require that $\beta \to 0$ and $\gamma \to 0$ as $n_0 \to \infty$.

Obviously, for a non-deterministic discrete memoryless or a Gaussian multiple input channel, one may always choose a pair of codewords $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ of length n_0 , such that the random output of the channel has positive entropy $n_0\alpha$, when $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ are sent to the channel. It is easy to see that, by taking the set of typical sequences as A, by AEP, one can partition A into (nearly) $2^{n_0\alpha}$ parts with nearly equal probabilities. For instance, for a non-deterministic discrete memoryless channel, there are about $2^{n_0(\alpha+o(1))}$ output typical sequences with nearly equal probabilities, if the output entropy is $n_0\alpha$. Thus, one may partition the typical sequences into $a = 2^{n_0(\alpha+o(1))}$ parts, such that each part contains nearly equal number of typical sequences, and therefore has nearly equal probability, similarly for a Gaussian channel. That is, (*) follows.

Feedback in memoryless channel usually may play the following 3 roles:

- (1) *Reducing Size of the Decoding List:* The idea was introduced by C. E. Shannon in his pioneering work [8]. In short, the receiver lists all messages, which are possibly to be sent to him/her. Initially, the list contains all messages in the message sets. At each time *t*, the receiver reduces the list according to the output symbols, that he/she has received so far. Finally, the receiver decodes the message, when only one single message remains in the list. The senders learn the list via feedback, and thus can effectually cooperate with the receiver(s) to work on the list (instead of the whole sets of the messages).
- (2) Shifting the Private Messages to the Common Messages: Perhaps, it is the most common way to use feedback to enlarge capacity regions (e.g., [9,10,19] and etc). One difficulty for the senders of an multiple input channel (without feedback) to cooperate is that, their (private) messages are independent. When feedback is present, a sender can obtain certain information about the message sent by the other sender(s), via his own inputs and previous outputs, which he has received via feedback. Thus, he/she can shift the information from the private messages (of the other encoder) to the common messages, and cooperate with the other encoder(s) to resend them.
- (3) *Extracting Common Randomness:* Finally, receiver(s) may extract a common randomness, or "a random key" from the output(s) and sends (send) it to the sender(s) via feedback. Thus, they may use the common randomness to help transmission (e.g., [20,21]).

The feedback will play the 3rd role in the following corollary. That is, by the property (*), we shall use a block of length n_0 to generate a random key of size a. Then by using the random key, we shall "mix the good and bad combinations" of input codeword pairs.

Note that

$$\mathcal{R}_{d,f}(W) \subset \bar{\mathcal{R}}_f(W) \tag{20}$$

Therefore, if the topological interior of $\bar{\mathcal{R}}_f(W)$ is empty, then the topological interior of $\mathcal{R}_{d,f}(W)$ is empty too. For simplicity of the discussion and ignoring this trivial case, we assume that the topological interior of $\bar{\mathcal{R}}_f(W)$ is not empty.

Corollary 4.1 For a non-deterministic discrete memoryless or Gaussian multiple input channel W, such that the topological interior of $\overline{\mathcal{R}}_f(W)$ is not empty, we have that

$$\mathcal{R}_{d,f}(W) = \bar{\mathcal{R}}_f(W) \tag{21}$$

Proof: By (20), it is sufficient for us to show that, all $(R_1, R_2) \in \mathcal{R}_f(W)$ is achievable by (deterministic) codes under the criterion of maximum probability of error and with feedback. For any given $\lambda, \epsilon > 0$ and a sufficiently large n, we first find a code C of length n with feedback, such that its average probability of error is at most $\frac{\lambda}{12}$, and $\frac{1}{n} \log |\mathcal{M}_i| \ge R_i - \frac{\epsilon}{2}, i = 1, 2$, where $\mathcal{M}_1 \times \mathcal{M}_2$ is the message set of C. Then we find a subcode C_0 of C and permutation groups $(\sigma_1, \sigma_2, \ldots, \sigma_{n^2})$ as described in Lemma 3.3. Next, we choose $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)})$ with the smallest possible n_0 in the property (*) such that

$$2^{n_0\alpha} \approx a = n^2, \quad \beta < \frac{\lambda}{12}, \quad \gamma < \frac{\lambda}{12}$$

Obviously, $\frac{n_0}{n}$ can be arbitrarily small, by choosing a sufficiently large n, since an integer n_0 no less than $\frac{\log n^2}{\alpha}$ is sufficient, and $\frac{\log n^2}{n} \to 0$, as $n \to \infty$.

We assume that the channel is a MAC. Our code consists of two blocks. In the first block, two senders send $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ (in the property (*)) respectively, no matter what messages they want to send. After the first block, both the first sender and the receiver learn the output \mathbf{y} of the first block. In the case that \mathbf{y} does not fall into A, the receiver declares an error, and we denote this error event by \mathcal{E}_0 . By assumption, we have that

$$\Pr(\mathcal{E}_0) < \frac{\lambda}{12} \tag{22}$$

In the case that \mathcal{E}_0 does not occur, we assume that y falls into the k-th subset in the partition of A. Then in the second block, two senders send message m_1 and m_2 , respectively, by using the code $\sigma_k(C_0)$ as described in Lemma 3.3. Recall (16). We have that for any fixed (m_1, m_2) , the probability of error in the second block is upper bounded by

$$\frac{1}{a}(1+\gamma)\sum_{k=1}^{n^2} p_e(\sigma_k(C_0), (m_1, m_2)) < \frac{1}{n^2}(1+\frac{\lambda}{12})\sum_{k=1}^{n^2} p_e(\sigma_k(C_0), (m_1, m_2)) < (1+\frac{\lambda}{12})\frac{2\lambda}{3} < \frac{3\lambda}{4}$$
(23)

which with (22) together implies that, the total probability of error of the code for the MAC is upper bounded by $\frac{\lambda}{12} + \frac{3\lambda}{4} = \frac{5\lambda}{6}$. Moreover, by choosing sufficiently small $\frac{n_0}{n}$, the rates of code may be larger than $R_i - \epsilon$, i = 1, 2.

Notice that the second sender does not need to know k although he/she may know it. That is, the statement of the corollary still holds in the case that the feedback is available at one encoder only.

Differently from an MAC, a TWC or IC has two outputs. Let A be in any of outputs. Then a receiver accessing the other output, may not know the outcome of the random key. Thus, we need an additional block, for the first encoder to inform him/her the outcome of the key. This can be done by using a code with feedback and with average error probability at most $\frac{\lambda}{6}$. Thus as in the proof of Theorem 3.2, the total maximum probability of error may not exceed λ . Again, because the rate of random key $\frac{\log n^2}{n} \to 0$ as $n \to \infty$, a vanishing rate is sufficient for the block. Thus, the total rates of code are not smaller than $R_i - \epsilon, i = 1, 2$. This completes our proof.

Remark 4.2 Now, among the capacity regions of Gaussian MAC, only the maximum-error-probability capacity region without feedback is unknown, because the average-error-probability capacity region of a Gaussian MAC with feedback was determined by H. Ozarow [11]; and, by Corollary 4.1, it is equal to the maximum-error-probability capacity region of the same channel. We conjecture that the maximum-error-probability capacity region of a Gaussian MAC would be strictly smaller than the average-error-probability capacity region, when the feedback is absent. However, a proof or disproof has not been discovered so far. It would be very appreciated, if one might point out its existence. Otherwise, it would be a good problem for future research.

Corollary 4.1 tells us that, for a non-deterministic channel with feedback, the error criterion makes no difference on its capacity region. But, it is not always true in the general case, because we have learnt from the contraction channel [1], which is a deterministic channel, that its maximum-error-probability capacity region of deterministic codes with feedback is strictly smaller than its average-error-probability capacity region without feedback. Now we know from Corollary 4.1, that one can find neither example of non-deterministic channel nor example of Gaussian channel, which leads us to a similar result as in [1]. So, Corollary 4.1 actually provides a counterpart of Dueck's case [1]. By comparing these two cases, we see an essential difference between deterministic and non-deterministic multiple user channels. The reason behind the difference is that: for a non-deterministic channel, one can build a common randomness through the feedback and use it to mix the "good cases" and "bad cases"; but for a deterministic channel, it is impossible. In this sense, noise may help transmission when feedback is present. We have seen this phenomenon from arbitrarily varying channel [20]. We will illustrate it in more details by examples in the next section.

5. Examples

Motivated by the following, we present two subclasses of MAC, as examples, in this section.

- We have seen that one can apply feedback, not only for shifting the private messages to the common messages, but also for extracting common randomness, to help the transmission under the criterion of the maximum probability of error. On the other hand, the common randomness may not help the transmission under the criterion of the average probability of error, as random codes and deterministic codes have the same average-error-probability capacity region. So, it is expected that there exists a multiple input channel, for which feedback may enlarge the maximum-error-probability capacity region, but not the average-error-probability capacity region.
- We wonder if there is a non-deterministic MAC, whose average- and maximum-error-probability capacity regions are different, since Dueck's contraction channel is special for being deterministic.

First let us recall an inner bound on the average-error-probability capacity regions of discrete memoryless MAC W with feedback, due to T. M. Cover and C. S. K. Leung [10], which will be used in this section:

$$\mathcal{R}_{cl}(W) = \{ (R_1, R_2) : R_1 \le I(X_1; Y | X_2, U), R_2 \le I(X_2; Y | X_1, U) \\ R_1 + R_2 \le I(X_1, X_2; Y), (U, X_1, X_2, Y) \in \mathcal{Q}(W) \}$$
(24)

where $\mathcal{Q}(W)$ is the set of quadruples (U, X_1, X_2, Y) of random variables, with a joint distribution

$$P_{UX_1X_2Y}(u, x_1, x_2, y) = P_U(u)P_{X_1|U}(x_1|u)P_{X_2|U}(x_2|u)W(y|x_1, x_2)$$

for all $(u, x_1, x_2, y) \in \mathcal{U} \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$. Here $\mathcal{X}_i, i = 1, 2$ are input alphabets of the MAC; \mathcal{Y} is the output alphabet of the MAC; and \mathcal{U} is a finite set with $|\mathcal{U}| \leq \min(|\mathcal{X}_1||\mathcal{X}_2|+1, |\mathcal{Y}|+2)$.

It was shown by F. M. J. Willems [22] that, the inner bound is tight, when the second input of the MAC is uniquely determined by the first input and the output of the MAC. We denote by \mathcal{K}_w this special class of MAC. In this section, we will discuss a subclass of \mathcal{K}_w to illustrate the role of feedback.

5.1. A Class of MAC, for which Feedback Does not Enlarge Average-Error-Probability Capacity Region

We say that a MAC W is in the class \mathcal{K}^* , if there is a function $g : \mathcal{Y} \to \mathcal{X}_2$, such that $x_2 = g(y)$, whenever there is an $x_1 \in \mathcal{X}_1$ with $W(y|x_1, x_2) > 0$. Obviously, $\mathcal{K}^* \subset \mathcal{K}_w$. Furthermore, we partition \mathcal{K}^* into two subclasses:

$$\mathcal{K}_d^* = \{ W : W \in \mathcal{K}^* \text{ and } W \text{ is deterministic} \}$$

$$\mathcal{K}_n^* = \{ W : W \in \mathcal{K}^* \text{ and } W \text{ is non-deterministic} \}$$

The maximum-error-probability capacity region of channels in \mathcal{K}_d^* with feedback was obtained by R. Ahlswede and N. Cai [23], where the channels in \mathcal{K}_d^* were referred as *semi-noisy deterministic MAC*, based on the 1st role of feedback: reducing size of the decoding list. It is easy to see that both the contraction channel in [1] and noiseless binary switching MAC in [3] belong to this subclass \mathcal{K}_d^* .

As $\mathcal{K}^* \subset \mathcal{K}_w$, by F. M. J. Willems' theorem in [22], the average-error-probability capacity region of a channel $W \in \mathcal{K}^*$ is equal to Cover-Leung inner bound $\mathcal{R}_{cl}(W)$ in (24), when feedback is present. By applying Corollary 4.1 to non-deterministic members in \mathcal{K}^* , *i.e.*,the channels in \mathcal{K}^*_n , we have $\mathcal{R}_{d,f}(W) = \overline{\mathcal{R}}_f(W)$ for $W \in \mathcal{K}^*_n$. In summary, we have

$$\mathcal{R}_{d,f}(W) = \bar{\mathcal{R}}_f(W) = \mathcal{R}_{cl}(W), \quad \text{for } W \in \mathcal{K}_n^*$$

In fact, for all channels in \mathcal{K}^* , feedback does not enlarge the average-error-probability capacity region.

Proposition 5.1 (*i*) For all $W \in \mathcal{K}^*$,

$$\bar{\mathcal{R}}(W) = \bar{\mathcal{R}}_f(W) \tag{25}$$

(ii) For all $W \in \mathcal{K}_n^*$,

$$\mathcal{R}_{d,f}(W) = \bar{\mathcal{R}}(W) = \bar{\mathcal{R}}_f(W) \tag{26}$$

Proof: Since (ii) immediately follows from (i) and Corollary 4.1, it is sufficient for us to show (i).

To show (i), we only need to verify that, the average-error-probability capacity region of any channel in \mathcal{K}^* without feedback, is equal to Cover-Leung inner bound. Recall that the average-error-probability capacity region of an MAC W is equal to the convex hull of

$$\mathcal{R}'(W; X_1, X_2) = \{ (R_1, R_2) : R_1 \le I(X_1; Y | X_2), R_2 \le I(X_2; Y | X_1), R_1 + R_2 \le I(X_1, X_2; Y) \}$$
(27)

where (X_1, X_2, Y) are a triple of random variables with a joint distribution $P_{X_1X_2Y}(x_1, x_2, y) = P_{X_1}(x_1)P_{X_2}(x_2)W(y|x_1, x_2)$, for all $(x_1, x_2, y) \in \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}$, where $\mathcal{X}_i, i = 1, 2$, are input alphabets of the MAC, and \mathcal{Y} is the output alphabet of the MAC. Now we only have to show that for all $W \in \mathcal{K}^*$, Cover-Leung bound $\mathcal{R}_{cl}(W)$ in (24) is contained in the convex hull of (27), since $\overline{\mathcal{R}}(W) \subset \mathcal{R}_{cl}(W)$.

To this end, we fix a $W \in \mathcal{K}^*$. First we observe that, for any pair of *independent* input random variables (X_1, X_2) and the corresponding random output Y, we have that

$$H(X_2|Y) = H(X_2|X_1, Y) = 0$$

due to the fact that by definition of \mathcal{K}^* , the random input X_2 is uniquely determined by the random output Y with probability 1. Therefore,

$$I(X_2;Y) = I(X_2;Y|X_1) = H(X_2)$$
(28)

as X_1 and X_2 are independent. Then the bounds on R_2 and $R_1 + R_2$ in (27) can be rewritten as

$$\begin{aligned} R_2 &\leq H(X_2), \\ R_1 + R_2 &\leq I(X_1; Y | X_2) + I(X_2; Y) = I(X_1; Y | X_2) + H(X_2) \end{aligned}$$

respectively. In addition, we notice that $R_1 \leq I(X_1; Y | X_2)$ and $R_2 \leq H(X_2)$ imply that $R_1 + R_2 \leq I(X_1; Y | X_2) + H(X_2)$. Thus for a $W \in \mathcal{K}^*$, one can simplify (27) to the following:

$$\mathcal{R}'(W; X_1, X_2) = \{ (R_1, R_2) : R_1 \le I(X_1; Y | X_2), R_2 \le H(X_2) \}$$
(29)

On the other hand, since for $(U, X_1, X_2, Y) \in \mathcal{Q}(W)$, $I(X_2; Y|X_1, U) \leq H(X_2|U)$ (actually the equality holds for $W \in \mathcal{K}^*$), additionally omitting the bound on $R_1 + R_2$ in (24), we obtain that

$$\mathcal{R}_{cl}(W) \subset \{(R_1, R_2) : R_1 \le I(X_1; Y | X_2, U), R_2 \le H(X_2 | U), (U, X_1, X_2, Y) \in \mathcal{Q}(W)\}$$

That is, $\mathcal{R}_{cl}(W)$ is contained in the convex hull of $\mathcal{R}'(W; X_1, X_2)$. This completes our proof.

By Proposition 5.1, feedback may not enlarge the average-error-probability capacity region of a MAC in \mathcal{K}^* . In the next subsection, we shall present a subset of MACs in K_n^* , for which the maximum-error-probability capacity regions are strictly smaller than the average-error-probability capacity regions. This serves as an example, where feedback does enlarge the maximum-error-probability capacity region (to the average-error-probability capacity region), but does not enlarge the average-error-probability capacity region. Intuitively, this can be explained by the roles of feedback as follows.

Feedback may not enlarge the average-error-probability capacity region of a MAC in \mathcal{K}^* : Cover-Leung bound was obtained by the technique of *shifting the private messages to the common messages*. That is, a sender may obtain some information about the message sent by the other sender, via his/her own inputs and previous outputs of the channel, which he/she learnt from the feedback. Then he/she may shift the information from the private messages of the other sender to common messages, and cooperate with the other sender to resend it. The key idea is that the "common message" is more easily to be sent, under the cooperation of two senders. A premise is that the common message should be unknown by the receiver, because otherwise it is not necessary to be resent. For a MAC in \mathcal{K}^* , the receiver is able to decode the second input directly, and so shifting it to common message is unnecessary at all. Thus, in this case the feedback may not be used for shifting message from private to common, to enlarge average-error-probability capacity region.

Feedback can enlarge the maximum-error-probability capacity region of a MAC in K_n^* : However, one can apply the 3rd role of feedback *i.e.*, *extracting common randomness* from feedback for mixing the "good cases" and the "bad case". This may enlarge the maximum-error-probability capacity region, when the worst case is considered.

5.2. Generalized Contraction MAC

In the previous subsection, we have that for a channel W in \mathcal{K}_n^* , $\mathcal{R}_{d,f}(W) = \overline{R}(W) = \overline{R}_f(W)$. To have an example in \mathcal{K}_n^* , for which feedback enlarges the maximum-error-probability capacity region, but does not enlarge the average-error-probability capacity region, we need an example in \mathcal{K}_n^* , for which $\mathcal{R}_d(W) \subsetneq \overline{R}(W)$. To this end, we extend the contraction channel to generalized contraction MACs as follows.

Let \tilde{W} be a (point-to-point) channel with input alphabet $\tilde{\mathcal{X}}$ and output alphabet $\tilde{\mathcal{Y}}$. Denote its capacity by γ_* . Let $\tilde{W}_j, j = 0, 1$, be two copies of \tilde{W} . That is, for j = 0, 1, \tilde{W}_j is a channel with an input alphabet $\tilde{\mathcal{X}}(j) = \{\tilde{x}(j) : \tilde{x} \in \tilde{\mathcal{X}}\}$, and output alphabet $\tilde{\mathcal{Y}}(j) = \{\tilde{y}(j) : \tilde{y} \in \tilde{\mathcal{Y}}\}$, such that $\tilde{W}_j(\tilde{y}(j)|\tilde{x}(j)) =$ $\tilde{W}(\tilde{y}|\tilde{x})$, for all $\tilde{x}(j) \in \tilde{\mathcal{X}}(j)$, and $\tilde{y}(j) \in \mathcal{Y}(j)$. (Here, $\tilde{\mathcal{X}}(j)$ and $\tilde{\mathcal{Y}}(j)$ can be understood as copies of alphabets $\tilde{\mathcal{X}}$ and $\tilde{\mathcal{Y}}$ respectively; $\tilde{x}(j)$ and $\tilde{y}(j)$ as copies of letters $\tilde{x} \in \tilde{\mathcal{X}}$ and $\tilde{y} \in \tilde{\mathcal{Y}}$, respectively; and j = 0, 1, are indexes of the copies.) Let $\mathcal{X}_1 = \tilde{\mathcal{X}}(0) \cup \tilde{\mathcal{X}}(1)$, $\mathcal{X}_2 = \{0, 1\}$ and $\mathcal{Y} = \tilde{\mathcal{Y}}(0) \cup \tilde{\mathcal{Y}}(1) \cup$ $\{e(0), e(1)\}$. The generic of the generalized contraction MAC originating from \tilde{W} is defined as

$$W(y|x_1, x_2) = \begin{cases} 1 & \text{if } y = e(j), x_1 \in \tilde{\mathcal{X}}(j) \text{ with } j = x_2\\ \tilde{W}_j(y|x_1) & \text{if } x_1 \in \tilde{\mathcal{X}}(j), y \in \tilde{\mathcal{Y}}(j) \text{ with } j \neq x_2\\ 0 & \text{else} \end{cases}$$
(30)

for all $x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2$, and $y \in \mathcal{Y}$. For a generalized contraction channel, the decoder may decode the second input from the output with probability 1. The message carried by the first input is "erased", if $x_1 \in \tilde{\mathcal{X}}(j)$ and $x_2 = j$ are sent to a generalized contraction channel; otherwise, the generalized channel transmits the message as to \tilde{W}_j . Obviously, a generalized contraction MAC is in the class \mathcal{K}^* and is non-deterministic if and only if the original channel \tilde{W} is non-deterministic. By taking the identity channel on $\{a, b\}$ as \tilde{W} , we have that the contraction MAC is a special generalized contraction MAC.

Since the generalized contraction channels are in \mathcal{K}^* , their average-error-probability capacity regions are convex hulls of $\mathcal{R}'(W; X_1, X_2)$ as given in (29). Let J be the binary random variable, such that J = j if the first random input $X_1 \in \tilde{\mathcal{X}}(j)$. Then the first inequality at the right hand side of (29) is

$$R_{1} \leq I(X_{1};Y|X_{2}) \stackrel{(a)}{=} I(X_{1},J;Y|X_{2})$$

= $I(J;Y|X_{2}) + I(X_{1};Y|J,X_{2}) = H(J|X_{2}) - H(J|X_{2},Y) + I(X_{1};Y|J,X_{2})$
 $\stackrel{(b)}{=} H(J|X_{2}) + I(X_{1};Y|J,X_{2}) \stackrel{(c)}{=} H(J) + I(X_{1};Y|J,X_{2})$

where (a) holds since J is a function of X_1 ; (b) holds since J is a function of Y; (c) holds since J and X_2 are independent. Next we let $P_{X_2}(1) = p$ and Pr(J = 1) = q. Continue with the above inequality,

$$R_{1} \leq h(q) + p(1-q)I(X_{1};Y|J=0, X_{2}=1) + (1-p)qI(X_{1};Y|J=1, X_{2}=0)$$

$$\stackrel{(d)}{\leq} h(q) + [(1-p)q + p(1-q)]\gamma_{*}$$

where h(q) is the binary entropy of q. Note that in (d) the equality holds if and only if $I(X_1; Y|J = 0, X_2 = 1) = I(X_1; Y|J = 1, X_2 = 0) = \gamma_*$; or, equivalently both $P_{X_1|J}(\cdot|0)$ and $P_{X_1|J}(\cdot|1)$ are optimal input of \tilde{W} . By the symmetry, we assume that $0 \le p \le \frac{1}{2}$. Consequently, the average-error-probability capacity region of a generalized contraction MAC W is the convex hull of rectangles

$$\mathcal{R}''(p) = \{ (R_1, R_2); R_1 \le \max_{q \in [0,1]} [h(q) + [(1-p)q + p(1-q)]\gamma_*], R_2 \le h(p) \}, \text{ for } p \in [0, \frac{1}{2}]$$
(31)

To show that the maximum-error-probability capacity region of a generalized contraction MAC is strictly smaller than its average-error-probability capacity region, we need an outer bound on the maximum-error-probability capacity region. G. Dueck [1] elegantly applied *vertex isoperimetric theorem in the binary Hamming space*, to derive an outer bound of the maximum-error-probability capacity region of the contraction channel. Along his way, we will derive our outer bound.

The isoperimetric problem is a basic problem in combinatorics, which asks how large at least, the boundary of a subset with a given cardinality in a discrete metric space has to be. Its asymptotic version is known as "Blowing Up Lemma", by people in Information Theory (e.g., c.f. [4,24]). The vertex isoperimetric theorem in the binary Hamming space first was discovered by K. H. Harper [25]. Its different proofs were then given by G.O.H. Katona [26], P. Frankl and Z. Füredi [27]. The theorem says that, the optimal configurations have a nested structure. In this paper, we only need its simplified version as given below. More about the isoperimetric theorem can be found in [26] or [27].

For a subset $\mathcal{A} \subset \{0,1\}^n$ and a positive integer l, the l-boundary of \mathcal{A} is defined as

$$\Gamma^{l}(\mathcal{A}) = \{\mathbf{b} : \mathbf{b} \in \{0, 1\}^{n} \text{ and there exists an } \mathbf{a} \in \mathcal{A} \text{ with } d_{H}(\mathbf{a}, \mathbf{b}) \leq l\}$$

where d_H is the Hamming distance. The following lemma is a simplified version of isoperimetric theorem in the binary Hamming space.

Lemma 5.2 [25] Let $\mathcal{A} \subset \{0,1\}^n$. Then if $|\mathcal{A}| \geq \sum_{i=0}^k \binom{n}{i}$, we have that for $l \leq n-k$,

$$|\Gamma^{l}(\mathcal{A})| \ge \sum_{i=0}^{k+l} \binom{n}{i}$$
(32)

Let C be a code with maximum probability of error λ and block length n, with codebooks \mathcal{U}_1 and \mathcal{U}_2 . Let k be the largest integer such that $|\mathcal{U}_2| \geq \sum_{i=0}^k \binom{n}{i}$, *i.e.*,

$$\sum_{i=0}^{k} \binom{n}{i} \le |\mathcal{U}_2| < \sum_{i=0}^{k+1} \binom{n}{i}$$
(33)

without loss of generality, we may assume $k \leq \frac{n}{2}$, since we are interested in the asymptotic rates of the codes. For $\mathbf{j} = (j_1, j_2, \dots, j_n) \in \{0, 1\}^n$, we denote $\tilde{\mathcal{X}}^n(\mathbf{j}) = \prod_{i=1}^n \tilde{\mathcal{X}}(j_i)$ and partition \mathcal{U}_1 into 2^n

Entropy 2014, 16

subsets $\mathcal{U}_1 \cap \tilde{\mathcal{X}}^n(\mathbf{j}), \mathbf{j} \in \{0, 1\}^n$. Suppose that a codeword $\mathbf{x}^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}) \in \mathcal{U}_1 \cap \tilde{\mathcal{X}}^n(\mathbf{j})$ for a $\mathbf{j} \in \{0, 1\}^n$, and a codeword $\mathbf{x}^{(2)} = (x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)}) \in \mathcal{U}_2$, are sent by the two senders, respectively. Then, with probability 1, the *t*-th output of the channel is $e(j_t)$ if $j_t = x_t^{(2)}$. Thus the receiver may distinguish two codewords in $\mathcal{U}_1 \cap \tilde{\mathcal{X}}^n(\mathbf{j})$, only by the $d_H(\mathbf{j}, \mathbf{x}^{(2)})$ symbols at the coordinates *t* with $j_t \neq x_t^{(2)}$, when a codeword $\mathbf{x}^{(2)}$ is sent to the second input of the channel. In other words, $\mathcal{U}_1 \cap \tilde{\mathcal{X}}^n(\mathbf{j})$ must form a decodable codebook (with small probability of error) for the channel \tilde{W} (by neglect of the indexes of the copies), at the coordinates *t* with $j_t \neq x_t^{(2)}$, for all $\mathbf{x}^{(2)}$. Thus by converse coding theorem for the point-to-point channels, we have that for all $\epsilon > 0$ and a sufficiently large *n*,

$$|\mathcal{U}_1 \cap \tilde{\mathcal{X}}^n(\mathbf{j})| \le 2^{d_H(\mathbf{j}, \, \mathcal{U}_2)(\gamma_* + \epsilon)} \tag{34}$$

where $d_H(\mathbf{j}, \mathcal{U}_2) = \min\{d_H(\mathbf{j}, \mathbf{x}^{(2)}), \mathbf{x}^{(2)} \in \mathcal{U}_2\}$. Thus we have that

$$\begin{aligned} |\mathcal{U}_{1}| &= \sum_{\mathbf{j}\in\{0,1\}^{n}} |\mathcal{U}_{1}\cap\mathcal{X}^{n}(\mathbf{j})| \\ &\leq \sum_{\mathbf{j}\in\{0,1\}^{n}} 2^{d_{H}(\mathbf{j},\mathcal{U}_{2})(\gamma_{*}+\epsilon)} = \sum_{l=0}^{n} (|\Gamma^{l}(\mathcal{U}_{2})| - |\Gamma^{l-1}(\mathcal{U}_{2})|) 2^{l(\gamma_{*}+\epsilon)} \end{aligned} \tag{35}$$

Now we have to maximize the right hand side of (35). By Lemma 5.2 and (33), we have that $\Gamma^{l}(\mathcal{U}_{2}) = 2^{n}$ for all $l \ge n - k$. Thus, the right hand side of (35) can be rewritten as

$$\sum_{l=0}^{n-k} (|\Gamma^{l}(\mathcal{U}_{2})| - |\Gamma^{l-1}(\mathcal{U}_{2})|) 2^{l(\gamma_{*}+\epsilon)} = -\sum_{l=0}^{n-k-1} |\Gamma^{l}(\mathcal{U}_{2})| (2^{(l+1)(\gamma_{*}+\epsilon)} - 2^{l(\gamma_{*}+\epsilon)}) + 2^{n} 2^{(n-k)(\gamma_{*}+\epsilon)}$$

which is maximized by $|\Gamma^l(\mathcal{U}_2)| = \sum_{i=0}^{k+l} \binom{n}{i}$ by Lemma 5.2. That is,

$$\mathcal{U}_1| \le \sum_{i=0}^n \binom{n}{i} 2^{|i-k|^+(\gamma_*+\epsilon)} \le n \max_{i\ge k} \binom{n}{i} 2^{(i-k)(\gamma_*+\epsilon)}$$
(36)

where $|z|^+ = \max\{0, z\}$. Thus by s (33) and (36) for $\frac{k}{n} = p$, we have an outer bound of the maximumerror-probability capacity region of W:

$$\mathcal{R}_{out}(W) = \{ (R_1, R_2) : R_1 \le \max_{q' \in [p, 1]} [h(q') + (q' - p)\gamma_*], R_2 \le h(p), p \in [0, \frac{1}{2}] \}$$

Because for a fixed p, the function $h(q') + (q'-p)\gamma_*$ achieves the maximum value at

$$q_0 = \frac{2^{\gamma_*}}{2^{\gamma_*} + 1} \tag{37}$$

and $q_0 \in [\frac{1}{2}, 1)$ for $\gamma_* \ge 0$, and $q_0 = \frac{1}{2}$ if and only if $\gamma_* = 0$, the outer bound can be rewritten as

$$\mathcal{R}_{out}(W) = \{ (R_1, R_2) : R_1 \le h(q_0) + (q_0 - p)\gamma_*, R_2 \le h(p), p \in [0, \frac{1}{2}] \}$$
(38)

In particular, when \tilde{W} is the identity channel on $\{a, b\}$, $\gamma_* = 1$ and $q_0 = \frac{2}{3}$, W becomes the contraction channel, and the outer bound (38) becomes Dueck's outer bound (6). It is easy to verify that, if $\gamma_* > 0$,

the outer bound in (38) is always strictly smaller than the average-error-probability capacity region, the convex hull of $\mathcal{R}''(p), p \in [0, \frac{1}{2}]$ in (31), since for any fixed $p \in [0, \frac{1}{2}]$,

$$\max_{q \in [0,1]} [h(q) + [(1-p)q + p(1-q)]\gamma_*] - [h(q_0) + (q_0 - p)\gamma_*]$$

$$\geq [h(q_0) + [(1-p)q_0 + p(1-q_0)]\gamma_*] - [h(q_0) + (q_0 - p)\gamma_*]$$

$$= 2p(1-q_0)\gamma_* > 0$$
(39)

if $\gamma_* > 0$. That is, for any (nontrivial) generalized contraction MAC, the maximum-probability-error capacity region is always strictly smaller than the average-error-probability capacity region. In particular, by Corollary 4.1 and Proposition 5.1, we have that for all non-deterministic generalized contraction MAC W,

$$\mathcal{R}_d(W) \subsetneq \overline{\mathcal{R}}(W) = \overline{\mathcal{R}}_f(W) = \mathcal{R}_{d,f}(W)$$

6. Extensions

In this section, we extend our results to compound channels, arbitrarily varying channels, and channels with more than two inputs. For the sake of simplicity, we will focus on the discrete memoryless MAC. As one can see, it is not that difficult to extend the results in Subsections 6.1 and 6.3 to Gaussian MAC and other multiple input channels. However, it is not very easy to extend Theorem 3.2 to arbitrarily varying Gaussian channel defined in the sense of [28], due to the following reason. By its definition, a Gaussian arbitrarily varying MAC with average power constraint Λ to the state sequences, is possibly governed by any state sequence with average power not exceeding Λ . In our coding scheme, we need to send the outcome of random key by a block with a vanishing rate. This would not work, when a state sequence, such that powers of its components in the "small block" are much larger than Λ , and powers in the rest part are slightly smaller than Λ , governs the channel.

6.1. Extension to Compound MAC

A compound MAC \mathfrak{W} is specified by a set of MACs $\{W(\cdot|\cdot, \cdot, s) : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}, s \in S\}$, where S is an index set, whose members are often called *states*. Without loss of generality, one may assume that S is finite, since an infinite state set can be approached by a finite state set (e.g., [4] [pp. 219–220]). A compound channel outputs a sequence $\mathbf{y} = (y_1, y_2, \dots, y_n)$ with probability

$$W(\mathbf{y}|\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, s) = \prod_{t=1}^{n} W(y_t|x_t^{(1)}, x_t^{(2)}, s)$$

if the sequences $\mathbf{x}^{(j)} = (x_1^{(j)}, x_2^{(j)}, \dots, x_n^{(j)}), j = 1, 2$ are sent to the channel, and the channel is governed by the state s. Similarly, for deterministic a code C (random code C_r) with message set $\mathcal{M}_1 \times \mathcal{M}_2$, we denote by $p_e(C; (m_1, m_2); s)$ ($p_e(C_r; (m_1, m_2); s)$) for $m_j \in \mathcal{M}_j, j = 1, 2$ and $s \in S$, the probability of error for the code $C(C_r)$, when a message pair (m_1, m_2) is sent and the state of the channel is s. We assume that the state governing the channel is unknown by the senders and the receiver. Accordingly, for a deterministic code C, the average and maximum probabilities of error are defined as

$$p_a(C) = \max_{s \in S} \frac{1}{M_1} \frac{1}{M_2} \sum_{(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} p_e(C; (m_1, m_2); s)$$
(40)

$$p_m(C) = \max_{s \in \mathcal{S}} \max_{(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} p_e(C; (m_1, m_2); s)$$
(41)

respectively; and, for a random code C_r ,

$$p_a(C_r) = \max_{s \in S} \frac{1}{M_1} \frac{1}{M_2} \mathbf{E} \{ \sum_{(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} p_e(C_r; (m_1, m_2); s) \}$$
(42)

$$p_m(C_r) = \max_{s \in \mathcal{S}} \max_{(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2} \mathbf{E} p_e(C_r; (m_1, m_2); s)$$

$$(43)$$

respectively. In the following, we extend the results in Sections 3, 4 to compound channels.

First we extend Theorem 3.2 by showing that

$$\mathcal{R}_r(\mathfrak{W}) = \overline{\mathcal{R}}(\mathfrak{W}) \text{ and } \mathcal{R}_{r,f}(\mathfrak{W}) = \overline{\mathfrak{R}}_f(\mathfrak{W}),$$
(44)

for all compound channels \mathfrak{W} . The proof is done by modifying the proof of Theorem 3.2. Suppose that we are given a deterministic code C of length n, with average probability of error $\frac{\lambda}{12|S|}$. Similarly, for all $s \in S$, we split the message set \mathcal{M}_2 into two parts: $\mathcal{M}_{2,0}(s)$ and $\mathcal{M}_{2,1}(s)$, such that

$$\frac{1}{M_1} \sum_{m_1 \in \mathcal{M}_1} p_e(C; (m_1, m_2); s) \ge \frac{\lambda}{6}$$

if and only if $m_2 \in \mathcal{M}_{2,1}(s)$. Then, by Markov inequality, we have that $|\mathcal{M}_{2,1}(s)| \leq \frac{|\mathcal{M}_2|}{2|\mathcal{S}|}$, for all $s \in \mathcal{S}$. Thus, we have that $|\bigcup_{s\in\mathcal{S}}\mathcal{M}_{2,1}(s)| \leq \frac{|\mathcal{M}_2|}{2}$; or equivalently, $|\mathcal{M}_{2,0}| \geq \frac{|\mathcal{M}_2|}{2}$ if we take $\mathcal{M}_{2,0} = \bigcap_{s\in\mathcal{S}}\mathcal{M}_{2,0}(s)$ as the second message set of "the new subcode" C_0 .

By applying Lemma 3.3-(2) to each channel in the set of \mathfrak{W} respectively, we have that for all $s \in S$,

$$\Pr\{\frac{1}{n^2}\sum_{k=1}^{n^2} p_e(\tilde{\sigma}_k(C_0), (m_1, m_2); s) \ge \frac{2}{3}\lambda\} < e^{-\frac{n^2\lambda}{6}}$$

Then by the union bound, with probability at least $1 - |\mathcal{S}||\mathcal{M}_1 \times \mathcal{M}_{2,0}|e^{-\frac{n^2\lambda}{6}}$, there is a realization of $(\tilde{\sigma}_1, \tilde{\sigma}_2, \dots, \tilde{\sigma}_{n^2})$, $(\sigma_1, \sigma_2, \dots, \sigma_{n^2})$, such that for all $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_{2,0}$ and all $s \in \mathcal{S}$

$$\frac{1}{n^2} \sum_{k=1}^{n^2} p_e(\sigma_k(C_0), (m_1, m_2); s) < \frac{2}{3}\lambda$$
(45)

We omit the rest parts of the proof of (44), since they follow exactly the proof of Theorem 3.2.

The extension of Corollary 4.1 to a compound MAC is straightforward. The proof follows from the argument in Section 4. Note that the definition of the non-deterministic compound channel slightly makes a difference here. If we define a non-deterministic compound MAC as a compound MAC, such that there is a pair of input letter (x_1, x_2) , for which no y and s are with $W(y|x_1, x_2, s) = 1$, then the same as in Section 4, we can spend a block to build the common randomness. However, it seems more natural to define a non-deterministic compound MAC as an compound MAC, such that for all $s \in S$, there exists at least a pair of input letter (x_1, x_2) for which no y is with $W(y|x_1, x_2, s) = 1$. In this case, the common randomness can be built by $|\mathcal{X}_1||\mathcal{X}_2|$ blocks. That is, every pair of input letter (x_1, x_2) takes a block. Surely, the sender and the receiver know from which block(s) they may extract randomness, because a "deterministic block" always produces the same output letter. The number of the blocks makes no difference, since their rates will vanish as the length of the code increases. Thus, the extension of Corollary 4.1 follows.

6.2. Extension to Arbitrarily Varying MAC

Similar to a compound MAC, an *arbitrarily varying MAC (AVMAC)* \mathcal{W} is specified by a finite set of MACs $\{W(\cdot|\cdot, \cdot, s) : \mathcal{X}_1 \times \mathcal{X}_2 \rightarrow \mathcal{Y}, s \in S\}$. Differently from a compound MAC, AVMAC is non-stationary, and governed by a state sequence in S^n . That is, an AVMAC outputs a sequence $\mathbf{y} = (y_1, y_2, \dots, y_n)$ with probability

$$W(\mathbf{y}|\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{s}) = \prod_{t=1}^{n} W(y_t|x_t^{(1)}, x_t^{(2)}, s_t)$$

if the sequences $\mathbf{x}^{(j)} = (x_1^{(j)}, x_2^{(j)}, \dots, x_n^{(j)}), j = 1, 2$, are sent to the channel, and the channel is governed by the state sequence $\mathbf{s} = (s_1, s_2, \dots, s_n)$. For a given deterministic code C (random code C_r) with message set $\mathcal{M}_1 \times \mathcal{M}_2$, we denote by $p_e(C; (m_1, m_2); \mathbf{s})$ ($p_e(C_r; (m_1, m_2); \mathbf{s})$) for $m_j \in \mathcal{M}_j, j = 1, 2$, and $\mathbf{s} \in S^n$, the probability of error of the code $C(C_r)$, when a message pair (m_1, m_2) is sent to the channel, and the channel is governed by a state sequence \mathbf{s} . Similarly, for a deterministic code C, the average and maximum probabilities of error are defined as

$$p_a(C) = \max_{\mathbf{s}\in\mathcal{S}^n} \frac{1}{M_1} \frac{1}{M_2} \sum_{(m_1,m_2)\in\mathcal{M}_1\times\mathcal{M}_2} p_e(C; (m_1,m_2); \mathbf{s})$$
(46)

$$p_m(C) = \max_{\mathbf{s}\in\mathcal{S}^n} \max_{(m_1,m_2)\in\mathcal{M}_1\times\mathcal{M}_2} p_e(C;(m_1,m_2);\mathbf{s})$$
(47)

respectively; and, for a random code C_r ,

$$p_a(C_r) = \max_{\mathbf{s}\in\mathcal{S}^n} \frac{1}{M_1} \frac{1}{M_2} \mathbf{E} \{ \sum_{(m_1,m_2)\in\mathcal{M}_1\times\mathcal{M}_2} p_e(C_r; (m_1,m_2); \mathbf{s}) \}$$
(48)

$$p_m(C_r) = \max_{\mathbf{s}\in\mathcal{S}^n} \max_{(m_1,m_2)\in\mathcal{M}_1\times\mathcal{M}_2} \mathbf{E}p_e(C_r; (m_1,m_2); \mathbf{s})$$
(49)

respectively.

According to our knowledge, most known works on AVMAC focused on the average-error-probability capacity regions. By elimination technique, J. H. Jahn [29] proved that, the average-error-probability capacity regions of random correlated codes and deterministic codes are the same, provided that the average-error-probability capacity region has a non-empty topological interior. Thereafter, it became a key problem to find conditions for which the average-error-probability capacity region of an AVMAC has a non-empty the topological interior. In [30], J. A. Gubner found a necessary condition and conjectured that it is also sufficient. Eventually, R. Ahlswede and N. Cai [31] proved that his conjecture is true. Since random codes and deterministic codes for an AVMAC in general may have different average-error-probability capacity regions, we add the subscript "d" to the capacity region of deterministic codes. However, we assume in the following that the average-error-probability capacity regions of random and deterministic codes are the same [29].

Actually, one can define the 3rd kind of probabilities of error, which we call *semi-average* probability of error:

$$p_s(C) = \max_{\mathbf{s}\in\mathcal{S}^n} \max_{m_2\in\mathcal{M}_2} \frac{1}{M_1} \sum_{m_1\in\mathcal{M}_1} p_e(C; (m_1, m_2); \mathbf{s}) \quad \text{for a deterministic code } C$$
(50)

$$p_s(C_r) = \max_{\mathbf{s}\in\mathcal{S}^n} \max_{m_2\in\mathcal{M}_2} \frac{1}{M_1} \mathbf{E} \sum_{m_1\in\mathcal{M}_1} p_e(C_r; (m_1, m_2); \mathbf{s}) \quad \text{for a random code } C_r$$
(51)

Accordingly, we denote the semi-average error probability capacity regions of deterministic codes, without and with feedback by $\tilde{\mathcal{R}}_d(\mathcal{W})$ and $\tilde{\mathcal{R}}_{d,f}(\mathcal{W})$, respectively.

The difference among the criterions of the average, semi-average, and maximum error probability regions is obvious. At first, we have that

$$\mathcal{R}_{d}(\mathcal{W}) \subset \tilde{\mathcal{R}}_{d}(\mathcal{W}) \subset \bar{\mathcal{R}}_{d}(\mathcal{W}) \text{ and } \mathcal{R}_{d,f}(\mathcal{W}) \subset \tilde{\mathcal{R}}_{d,f}(\mathcal{W}) \subset \bar{\mathcal{R}}_{f,d}(\mathcal{W})$$
 (52)

Secondly, we observe that $\tilde{\mathcal{R}}_d(\mathcal{W})$ may strictly contain $\mathcal{R}_d(\mathcal{W})$, and be strictly contained by $\bar{\mathcal{R}}_d(\mathcal{W})$. To see it, let us recall that it was proven in [32] by examples, that the maximum-error-probability capacity of a (point-to-point) AVC may be strictly smaller than its average-error-probability capacity. Then we let \mathcal{W}_i , i = 1, 2, be two (point-to-point) AVCs with average-error-probability capacity $\gamma_a(\mathcal{W}_i)$, and maximum-error-probability capacity $\gamma_m(\mathcal{W}_i)$ (of deterministic codes), input alphabets \mathcal{X}_i , output alphabets \mathcal{Y}_i , and set of state \mathcal{S}_i , respectively, such that $\gamma_m(\mathcal{W}_i) < \gamma_a(\mathcal{W}_i)$, for i = 1, 2. Let \mathcal{W} be the AVMAC having input alphabets \mathcal{X}_i , i = 1, 2, output alphabet $\mathcal{Y} = \mathcal{Y}_1 \times \mathcal{Y}_2$, and set of states $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$, and specified by the set of MACs $\{W(\cdot|\cdot, \cdot, (s_1, s_2)) : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{Y}, (s_1, s_2) \in \mathcal{S}\}$, such that

$$W((y_1, y_2)|x_1, x_2, (s_1, s_2)) = W_1(y_1|x_1, s_1)W_2(y_2|x_2, s_2)$$

for all $x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, (y_1, y_2) \in \mathcal{Y}$, and $(s_1, s_2) \in \mathcal{S}$. Then $\mathcal{R}_d(\mathcal{W}), \tilde{\mathcal{R}}_d(\mathcal{W})$, and $\bar{\mathcal{R}}_d(\mathcal{W})$ are given by

$$\{(R_1, R_2) : R_1 \leq \gamma_m(\mathcal{W}_1), R_2 \leq \gamma_m(\mathcal{W}_2)\}$$
$$\{(R_1, R_2) : R_1 \leq \gamma_a(\mathcal{W}_1), R_2 \leq \gamma_m(\mathcal{W}_2)\}$$
$$\{(R_1, R_2) : R_1 \leq \gamma_a(\mathcal{W}_1), R_2 \leq \gamma_a(\mathcal{W}_2)\}$$

respectively. Obviously we have that

$$\mathcal{R}_d(\mathcal{W}) \subsetneq \tilde{\mathcal{R}}_d(\mathcal{W}) \subsetneq \bar{\mathcal{R}}_d(\mathcal{W})$$

i.e., indeed, the three capacity regions are different. Similarly we can see the difference, when feedback is present.

We have discussed in Section 3 that, randomization at one single encoder is sufficient for a random code under the criterion of maximum probability of error to achieve the average-error-probability capacity region, and therefore the number of random encoders makes no difference for an (ordinary) MAC. Now, we shall see that it does make a difference for an AVMAC. Let $\mathcal{R}_{r,d}(\mathcal{W})$ and $\mathcal{R}_{r,r}(\mathcal{W})$ be the maximum-error-probability capacity regions of random codes with a random encoder and a deterministic encoder, and with two random encoders, respectively, when feedback is absent.

Theorem 6.1 For an AVMAC W, whose average-error-probability capacity region has a non-empty topological interior,

$$\mathcal{R}_{r,d}(\mathcal{W}) = \mathcal{R}_d(\mathcal{W}) \tag{53}$$

$$\mathcal{R}_{r,r}(\mathcal{W}) = \bar{\mathcal{R}}_d(\mathcal{W}) \tag{54}$$

Proof: To show (53), we first show that

$$\mathcal{R}_{r,d}(\mathcal{W}) \supset \mathcal{R}_d(\mathcal{W}) \tag{55}$$

This can be done by simply modifying the proof of Theorem 3.2, similar to the extension to compound channels in last subsection. For a $\lambda > 0$ and any pair of rates $(R_1, R_2) \in \tilde{\mathcal{R}}_d(\mathcal{W})$, we have a code deterministic code C for an AVMAC of length n with rates larger than $R_i - \frac{\epsilon}{2}$, i = 1, 2, and the semi-average probability of error $\frac{\lambda}{6}$, *i.e.*,

$$\frac{1}{M_1} \sum_{m_1 \in \mathcal{M}_1} p_e(C; (m_1, m_2); \mathbf{s}) \le \frac{\lambda}{6}$$
(56)

for all $m_2 \in \mathcal{M}_2$ and $\mathbf{s} \in \mathcal{S}^n$. Now, we have to construct a random code $\tilde{C}_{r,d}$ with one random encoder and one deterministic encode and show (49) for $\tilde{C}_{r,d}$, *i.e.*, for all $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$, $\mathbf{s} \in \mathcal{S}^n$,

$$\mathbf{E}p_e(\tilde{C}_{r,d}; (m_1, m_2); \mathbf{s}) < \lambda$$

By (56), the code C already has the property of C_0 in Lemma 3.3, which is needed by us. Thus, it is sufficient to show that, (by replacing C_0 in Lemma 3.3-(2) with C,) the set of random permutations in Lemma 3.3 has a realization $(\sigma_1, \sigma_2, \ldots, \sigma_{n^2})$, such that for all $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$, $\mathbf{s} \in S^n$,

$$\frac{1}{n^2} \sum_{k=1}^{n^2} p_e(\sigma_k(C), (m_1, m_2); \mathbf{s}) < \frac{2}{3}\lambda$$
(57)

since the rest part of the proof will directly follow from the proof of Theorem 3.2, similar to proof of extension to compound channel in the last subsection. Applying Lemma 3.3 to all $s \in S^n$, we have that for all $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$, $s \in S^n$,

$$\Pr\{\frac{1}{n^2}\sum_{k=1}^{n^2} p_e(\tilde{\sigma}_k(C), (m_1, m_2); \mathbf{s}) \ge \frac{2}{3}\lambda\} < e^{-\frac{n^2\lambda}{6}}$$
(58)

That is, with a probability no less than $1 - |\mathcal{M}_1 \times \mathcal{M}_2| |\mathcal{S}^n| e^{-\frac{n^2 \lambda}{6}}$, a realization satisfying (57) exists.

Next, we show the opposite relation

$$\mathcal{R}_{r,d}(\mathcal{W}) \subset \mathcal{R}_d(\mathcal{W}) \tag{59}$$

To do that, we have to prove that, given $C_{r,d}$ having a random encoder Φ_1 and a deterministic encoder ϕ_2 , with the maximum probability of error $\frac{\lambda}{4}$, one can always construct a deterministic code C' with the same rate and semi-average probability of error λ . It is done by choosing a realization of $C_{r,d}$. Note that by Remark 2.1-(2), we may for a fixed $m_2 \in \mathcal{M}_2$ and $\mathbf{s} \in S^n$, assume that $\{p_e(C_{r,d}; (m_1, m_2); \mathbf{s}) : m_1 \in \mathcal{M}_1\}$ are independent. Then, applying Chernoff bound, *i.e.*, Lemma 3.1 for $\beta = 1$ and $\alpha = \frac{1}{4}$, we have that for fixed $m_2 \in \mathcal{M}_2$, $\mathbf{s} \in S^n$,

$$\Pr\{\frac{1}{M_1}\sum_{m_1\in\mathcal{M}_1}p_e(C_{r,d};(m_1,m_2);\mathbf{s})>\lambda\}\leq e^{-\frac{M_1\lambda}{4}}$$

Consequently, applying the union bound, we have

$$\Pr\{\max_{\mathbf{s}\in\mathcal{S}^n}\max_{m_2\in\mathcal{M}_2}\frac{1}{M_1}\sum_{m_1\in\mathcal{M}_1}p_e(C_{r,d};(m_1,m_2);\mathbf{s})>\lambda\}<|\mathcal{S}^n||\mathcal{M}_2|e^{-\frac{M_1}{4}\lambda}$$

which can be arbitrarily close to 0, for a sufficiently large n, since M_1 exponentially increases as n increases. This implies that $C_{r,d}$ has a realization with semi-average probability of error at most λ , *i.e.*, (59).

Now we proceed to the proof of (54). Notice that under our assumption, $\overline{\mathcal{R}}_d(\mathcal{W})$ is equal to the average-error-probability capacity region of random correlated codes, which obviously contains $\mathcal{R}_{r,r}(\mathcal{W})$. So it is sufficient for us to show

$$\mathcal{R}_{r,r}(\mathcal{W}) \supset \bar{\mathcal{R}}_d(\mathcal{W}) \tag{60}$$

Similarly to the proof of Theorem 3.2, for a deterministic code C for \mathcal{W} , with average probability of error $\frac{\lambda}{6}$, and permutations $\sigma^{(i)}$ on \mathcal{M}_i , i = 1, 2, we define a code $(\sigma^{(1)}, \sigma^{(2)})(C)$, as a code having encoders $\phi_i(\sigma_i(m_i))$, for $m_i \in \mathcal{M}_i$, i = 1, 2, and decoder $((\sigma^{(1)})^{-1}(\psi_1(\mathbf{y})), ((\sigma^{(2)})^{-1}(\psi_2(\mathbf{y})))$, for $\mathbf{y} \in \mathcal{Y}^n$, where $\mathcal{M}_i, \phi_i, i = 1, 2$, and $\psi = (\psi_1, \psi_2)$ are message sets, encoders, and decoder of C, respectively.

Next we randomly, independently and uniformly generate n^2 random permutations $\tilde{\sigma}_k^{(i)}$, $k = 1, 2, \ldots, n^2$ on \mathcal{M}_i for i = 1, 2, respectively. Then, similarly to the proof of Theorem 3.2, we have that $p_e((\tilde{\sigma}_{k_1}^{(1)}, \tilde{\sigma}_{k_2}^{(2)})(C); (m_1, m_2); \mathbf{s}), k_1 = 1, 2, \ldots, n^2, k_2 = 1, 2, \ldots, n^2$, are n^4 independent random variables with expectations

$$\mathbf{E}p_e((\tilde{\sigma}_{k_1}^{(1)}, \tilde{\sigma}_{k_2}^{(2)})(C); (m_1, m_2); \mathbf{s}) = \frac{1}{M_1} \frac{1}{M_2} \sum_{(m_1', m_2') \in \mathcal{M}_1 \times \mathcal{M}_2} p_e(C; (m_1', m_2'); \mathbf{s}) \le \frac{\lambda}{6}$$

for all fixed $(m_1, m_2) \in \mathcal{M}_1 \times \mathcal{M}_2$ and $\mathbf{s} \in \mathcal{S}^n$. Thus, it follows from Chernoff bound that

$$\Pr\{\frac{1}{n^2}\sum_{k_1=1}^{n^2}\sum_{k_2=1}^{n^2}p_e((\tilde{\sigma}_{k_1}^{(1)},\tilde{\sigma}_{k_2}^{(2)})(C);(m_1,m_2);\mathbf{s}) > \frac{2\lambda}{3}\} \le e^{-\frac{n^4\lambda}{6}}$$

As a direct consequence, by the union bound, with a probability at least $1 - |\mathcal{M}_1 \times \mathcal{M}_2||\mathcal{S}^n|e^{-\frac{n^4\lambda}{6}}$, $\{\tilde{\sigma}_{k_i}^{(i)}, k_i = 1, 2, \dots, n^2, i = 1, 2\}$ has a realization $\{\sigma_{k_i}^{(i)}, k_i = 1, 2, \dots, n^2, i = 1, 2\}$, such that for all (m_1, m_2) , s, we have

$$\frac{1}{n^4} \sum_{k_1=1}^{n^2} \sum_{k_2=1}^{n^2} p_e((\sigma_{k_1}^{(1)}, \sigma_{k_2}^{(2)})(C); (m_1, m_2); \mathbf{s}) \le \frac{2\lambda}{3}$$

Now we construct a two-block random code by using the code C and the realization. That is, the two encoders randomly generate two keys K_i , i = 1, 2 from $\{1, 2, ..., n^2\}$, and send them in the first block independently, by using a code with average error probability at most $\frac{\lambda}{3}$. In the second block, they use the code $(\sigma_{k_1}^{(1)}, \sigma_{k_2}^{(2)})(C)$ to sent their messages, if the outcomes of K_i are k_i , i = 1, 2. Then the total maximum probability of error may not be larger than λ , if the average probability of error in the first block is no larger than $\frac{\lambda}{3}$. That is, (60) holds, which completes the proof of the theorem.

Similarly to in Section 3, here we have the following corollary.

Corollary 6.2 For an AVMAC \mathcal{W} , if there exist $x_i \in \mathcal{X}_i$, i = 1, 2 such that for all $s \in S$, there is no $y \in \mathcal{Y}$ with $W(y|x_1, x_2, s) = 1$, then

$$\mathcal{R}_{d,f}(\mathcal{W}) = ar{\mathcal{R}}_f(\mathcal{W})$$

- Remarks 6.3 (1) The random codes in Theorem 6.1 seems to be slightly different from the random codes in [29]. In Theorem 6.1, we assume that the two encoders choose codewords according to the outcomes of two random keys, respectively; and, the decoder does not know the outcomes of the keys (initially). In [29] J. H. Jahn assumed that the decoder knows the keys. But in the case that the capacity region has a non-empty topological interior, they are equivalent. This is because (as we did,) in this case the encoders may send the outcomes of keys to the decoder and by elimination technique, a block with a vanishing rate is sufficient for the purpose.
 - (2) The assumption in Corollary 6.2 greatly simplified the problem. The problem would become very complicated when the assumption is violated (i.e., in the case that for all $x_i \in X_i$, i = 1, 2, there exist an $s \in S$ and a $y \in Y$ with $W(y|x_1, x_2, s) = 1$). In this case, the conclusion in the corollary likely does not hold any more. For a (point-to-point) AVC, the maximum-error-probability capacity with feedback may be positive, but strictly smaller than its average-error-probability capacity, if for all input letter x, there are a state s and an output letter y such that W(y|x, s) = 1 [20].

6.3. Extension to Multiple Input MAC

The results in Sections 3 and 4 can be easily extended to an MAC with I inputs for I > 2. As the proof of the extension is very similar to the proof in Sections 3 and 4, we just give a brief outline as follows. Suppose that we are given a deterministic code C with average probability of error $\frac{\lambda}{12}$. Let $\phi_i, i = 1, 2, ..., I$, be its encoding functions.

• To find a "good" subcode: It follows from Markov inequality that, C contains a subcode C_0 with message sets $\mathcal{M}_i, i = 1, 2, ..., I - 1, \mathcal{M}_{I,0} \subset \mathcal{M}_I, |\mathcal{M}_{I,0}| \geq \frac{1}{2} |\mathcal{M}_I|$, and

$$\left(\prod_{i=1}^{I-1} \frac{1}{|\mathcal{M}_i|}\right) \sum_{i=1}^{I-1} \sum_{m_i \in \mathcal{M}_i} p_e(C; (m_1, m_2, \dots, m_{I-1}, m_I)) < \frac{\lambda}{6}$$

for all $m_I \in \mathcal{M}_{I,0}$, where $\mathcal{M}_i, i = 1, 2, \ldots, I$, are message sets of C.

Existence of permutations for construction of a random code: Similarly to the proofs in Section 3 and Subsection 6.2, define a code (σ⁽¹⁾, σ⁽²⁾,..., σ^(I-1))(C₀), for a set of permutations σ⁽ⁱ⁾ on M_i, i = 1, 2, ..., I − 1 *i.e.*, by taking φ_i(σ⁽ⁱ⁾(·)) for i = 1, 2, ..., I − 1, φ_I(·), as encoding function, and properly choosing a decoding function. Randomly and independently generate n² permutations σ⁽ⁱ⁾_{k_i}, k_i = 1, 2, ..., n² on M_i, for i = 1, 2, ..., I − 1. Apply Chernoff bound to show that there exists a realization of the random permutations, {σ⁽ⁱ⁾_{k_i}, k_i = 1, 2, ..., n², i = 1, 2, ..., I − 1} such that

$$\frac{1}{n^{2(I-1)}} \sum_{(k_1,k_2,\dots,k_{I-1})} p_e((\sigma_{k_1}^{(1)},\sigma_{k_2}^{(2)},\dots,\sigma_{k_{I-1}}^{(I-1)})(C_0);(m_1,m_2,\dots,m_{I-1},m_I)) < \frac{2\lambda}{3}$$

for all $(m_1, m_2, \ldots, m_{I-1}, m_I) \in \mathcal{M}_1 \times \mathcal{M}_2 \times \ldots \times \mathcal{M}_{I-1} \times \mathcal{M}_{I,0}$.

- Constructing a random code with maximum probability of error: Construct a random code under the criterion of maximum probability of error, with I - 1 random encoders and a deterministic encoder, in two blocks such that the outcomes of random keys are sent in the first block, and messages are sent in the second block.
- Constructing a deterministic code with feedback for non-deterministic MAC: Construct deterministic code with feedback under the criterion of maximum probability of error in two blocks, such that common randomness between the *i*-th encoders and the decoder, for *i* = 1, 2, ... *I* − 1, are generated in the first block; and, messages are sent in the second block. Here *C* is understood as a code with feedback.

7. Discussions

In [32], R. Ahlswede defined 6 capacities for AVC and discussed their relation. Readers who are familiar with AVC, may recognize that a multiple input channel has properties similar to AVC, especially when the criterion of maximum probability of error is considered; and, our results are similar to those in [32]. A (point-to-point) AVC can be regarded as an MAC or IC, where an input is controlled by a malicious user. The similarity is caused by the fact that for both AVC and multiple input channels under the criterion of maximum error probability, one must consider the "bad combination" of input codewords. For multiple input channels, this is due to the fact that the senders may not always cooperate properly. A user might be considered to be "malicious" by other users, if they do not cooperate well. This can be reflected by the following fact. For a random code under the criterion of maximum error probability, to achieve the average-error-probability capacity region of an I input channel, one needs I - 1 random encoders; and, for arbitrarily varying I input channel, one needs one more random encoder. A difference is that, for an MAC the receiver has to decode the messages from both input; and, for an IC, a receiver may decode the message not to be sent to him/her, as a side information; whereas it is impossible for a receiver of an AVC to decode the state sequence.

We have seen that randomization at encoders serves as an efficient way to solve the problem caused by the bad combination of input codewords. As in many cases, a perfect cooperation in network communication is impossible, this suggests us to employ random codes.

So far we have known very little about the maximum-error-probability capacity regions. One reason is that in most cases a small average probability of error is acceptable. Perhaps, another reason is that, to determine the maximum-error-probability capacity regions is much harder than to determine the average-error-probability capacity regions. Nevertheless, from a theoretical point of view, to well understand the multiple user channels, exploring their maximum-error-probability capacity regions is necessary. Likely, the study on the maximum-error-probability capacity regions is closely related to the study on AVCs.

Acknowledgements

The author would like to thank Yanling Chen for many discussions and helpful suggestions in details for revising the paper. This work was partially supported by grant from the National Natural Science Foundation of China (Ref. No. 61271174).

Conflicts of Interest

The authors declare no conflict of interest.

References

- 1. Dueck, G. Maximal error capacity regions are smaller than average error capacity regions for multi-user channels. *Probl. Contr. Inform. Theor.* **1978**, *7*, 11–19.
- Ahlswede, R. On two-way communication channels and a problem. In Proceedings of the Zarankiewiez, Sixth Prague Conference on Information Thery, Statistical Decision Functions and Random Processes, Prague, Czech, 19–25 September 1971.
- 3. Vanroose, P. Code construction for the noiseless binary switching multiple-access channel. *IEEE Trans. Inform. Theor.* **1988**, *34*, 1100–1106.
- 4. Csiszár, I.; Körner, J. *Information Theory: Coding Theorem for Discrete Memoryless Sytems*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2011.
- 5. Gamal, A.E.; Kim, Y.H. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2011.
- Ahlswede, R. Multi-way communication channels. In Proceedings of the 2nd International Symposium on Information Theory, Tsahkadsor, Armenian USSR, 2–8 September 1971; Publishing House of the Hungarian Academy of Sciences: Budapest, Hungary, 1973; pp. 23–52.
- 7. Liao, H. A coding theorem for multiple access communications. In Proceedings of the International Symposium on Information Theory, Asilomar, USA, 1972.
- 8. Shannon, C.E. The zero-error capacity of a noisy channel. *IRE Trans. Inform. Theor.* **1956**, *2*, 8–19.
- 9. Gaarder, N.T.; Wolf, J.K. The capacity region of a multiple-access discrete memoryless channel can increase with feedback. *IEEE Trans. Inform. Theor.* **1975**, *21*, 100–102.
- 10. Cover, T.M.; Leung, C.S.K. An achievable rate region for the multiple-access channel with feedback. *IEEE Trans. Inform. Theor.* **1981**, *27*, 292–298.
- 11. Ozarow, L.H. The capacity of the white gaussian multiple access channel with feedback. *IEEE Trans. Inform. Theor.* **1984**, *30*, 623–629.
- 12. Ozarow, L.H.; Leung, C.S.K. An achievable region and outer bound for the Gaussian broadcast channel with feedback. *IEEE Trans. Inform. Theor.* **1984**, *30*, 667–671.
- 13. Kramer, G. Capacity results for the discrete memoryless network. *IEEE Trans. Inform. Theor.* **2003**, *49*, 4–21.
- 14. Bross, S,I.; Lapidoth, A. An improved achievable region for the discrete memoryless two-user multiple-access channel with noiseless feedback. *IEEE Trans. Inform. Theor.* **2005**, *51*, 811–833.
- 15. El Gamal, A. The feedback capacity of degraded broadcast channels. *IEEE Trans. Inform. Theor.* **1978**, *24*, 379–381.
- 16. Han, T.S.; Kobayashi, K. A new achievable rate region for the interference channel. *IEEE Trans. Inform. Theor.* **1981**, *27*, 49–60.
- Shannon, C.E. Two-way communication channels. Berkeley, CA, USA, 20 June 30 July 1960; University of California Press: Berkeley, CA, USA, 1961; Volume I, pp. 611–644.

- 18. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Weinheim, Germany, 2006.
- 19. Dueck, G. The zero error feedback capacity region of a certain class of multiple-access channels. *Probl. Control Inform. Theory* **1985**, *14*, 89–103.
- Ahlswede, R.; Cai, N. The AVC with noiseless feedback and maximal error probability: A capacity formula with a trichotomy. In *Numbers, Information, and Complexity (Festschrift for Rudolf Ahlswede)*; Althöfer, I., Cai, N., Dueck, G., Khachatian, G., Pinsker, M.S., Sarkozy, A., Wegener, I., Zhang, Z. Eds; Kluwer: Dordrecht, The Netherlands, 2000; pp. 151–176.
- Ahlswede, R.; Cai, N. Transmission, Identification and Common Randomness Capacities for Wire-tape Channels with Secure Feedback from the Decoder. In *General Theory of Information Transfer and Combinatorics*, Lecture Notes in Computer Science; Springer: Berlin, Germany, 2006; Volume 4123, pp. 257–274.
- 22. Willems, F.M.J. The feedback capacity region of a class of discrete memoryless multiple access channels. *IEEE Trans. Inform. Theor.* **1982**, *28*, 93–95.
- 23. Ahlswede, R.; Cai, N. Seminoisy deterministic multiple-access channels: Coding theorems for list codes and codes with feedback. *IEEE Trans. Inform. Theor.* **2002**, *48*, 2153–2162.
- 24. Alhswede, R.; Gács, P.; Körner, J. Bounds on Conditional Probabilities with Application in Muli-User Communication. *Probab. Theor. Relat. Field.* **1976**, *34*, 157–177.
- 25. Harper, K.H. Optimal numberings and insperimetric problems on graphs. J. Comb. Theory **1966**, 1, 385–393.
- 26. Katona, G.O.H. The Hamming-sphere Has Minimum Boundary. *Studia Sci. Math. Hugar.* **1975**, *10*, 131–140.
- 27. Frankl, P.; Füredi, Z. A short proof for a theorem of harper about hamming-spheres. *Discrete Math.* **1981**, *34*, 311–313.
- 28. Csiszár, I.; Narayan, P. Capacity of the gaussian arbitrarily varying channel. *IEEE Trans. Inform. Theor.* **1991**, *31*, 18–26.
- 29. Jahn, J.H. Coding of arbitrarily varying multiuser channels. *IEEE Trans. Inform. Theor.* **1981**, 27, 212–226.
- 30. Gubner, J.A. On the deterministic-code capacity of the multiple-access arbitrarily varying channel. *IEEE Trans. Inform. Theor.* **1990**, *36*, 262–275.
- 31. Ahlswede, R.; Cai, N. Arbitrarily varying multiple-access channels part I, Ericson's symmetrizability is adequate, Gubner's conjecture is true. *IEEE Trans. Inform. Theor.* **1999**, 45, 742–749.
- 32. Ahlswede, R. Elimination of correlation in random codes for arbitrarily varying channels. *Probab. Theor. Relat. Field.* **1978**, *44*, 159–175.

© 2014 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).