# Interference channel capacity region for randomized fixed-composition codes

Cheng Chang

HP Laboratories
HPL-2008-194

**Keyword(s):**
interference channels, randomized coding, capacity region

**Abstract:**
The random coding error exponents are studied [5], [6] for the finite alphabet interference channel (IFC) with two transmitter receiver pairs. The code words are uniform on a fixed-composition set and the decoding is optimum, as opposed to decoding based on interference cancellation, and decoding that considers the interference as additional noises. In this paper we further study the error exponents of randomized fixed-composition coding, some simple lower bounds are derived for universal decoding rules. Furthermore, we give a complete characterization of the capacity region of this coding scheme that is first proposed in [5] and [6]. It is shown that even with a sophisticated time-sharing scheme among randomized fixed-composition codes, the capacity region of the randomized fixed-composition coding is not bigger than the known Han-Kobayashi capacity region first appeared in [12]. This suggests that the average behavior of random codes are not sufficient to get new capacity regions.

External Posting Date: November 21, 2008 [Fulltext] - Approved for External Publication

Internal Posting Date: November 21, 2008 [Fulltext]

To be published in IEEE Transactions on Information Theory & IEEE International Symposium on Information Theory

# Interference channel capacity region for randomized fixed-composition codes

Cheng Chang

***Abstract*—The random coding error exponents are studied [5], [6] for the finite alphabet interference channel (IFC) with two transmitter receiver pairs. The code words are uniform on a fixed-composition set and the decoding is optimum, as opposed to decoding based on interference cancellation, and decoding that considers the interference as additional noises. In this paper we further study the error exponents of randomized fixed-composition coding, some simple lower bounds are derived for universal decoding rules. Furthermore, we give a complete characterization of the capacity region of this coding scheme that is first proposed in [5] and [6]. It is shown that even with a sophisticated time-sharing scheme among randomized fixed-composition codes, the capacity region of the randomized fixed-composition coding is not bigger than the known Han-Kobayashi capacity region first appeared in [12]. This suggests that the average behavior of random codes are not sufficient to get new capacity regions.**

***Index Terms*—interference channels, randomized coding, capacity region**

## I. Introduction and Problem Setup

In [12], the capacity region of interference channel is studied for both discrete and Gaussian cases. In this paper we study the discrete interference channels with 2 pairs of encoders and decoders as shown in Figure 1. The two channel inputs are $x^n \in \mathcal{X}^n$ and $y^n \in \mathcal{Y}^n$, outputs are $z^n \in \mathcal{Z}^n$ and $\tilde{z}^n \in \tilde{\mathcal{Z}}^n$ respectively, where $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ and $\tilde{\mathcal{Z}}$ are finite sets. We study a simpler channel model where each encoder only has a private message to the correspondent decoder while in [12], a more complicated case is studied while allowing public messages.

The capacity regions for general interference channels are generally unknown. We focus on the capacity region for a specific coding scheme– randomized fixed-composition codes while the error probability is defined as the average error over all code book with a certain composition. Fixed-composition coding is a useful technique in proving classical information theoretical problems, especially error exponent results [7]. In [5] and [6], randomized fixed composition coding is used to drive a lower bound on the error exponent for discrete interference channels. Some numerical results on this error exponents are presented, this is the first attempt in investigating the error exponents for interference channels as we know. In this paper, we derive the interference channel capacity region for randomized fixed-composition codes. We show that the fixed composition coding scheme does not achieve new capacity regions known as the Han-Kobayashi region first appeared in [12].

The outline of the paper is as follows, we first formally define randomized fixed-composition codes and its capacity region. In Section II we present the main result of this paper: the interference channel capacity region for randomized fixed composition code in Theorem 1. The proof is later shown in Section III with some details in the appendix.

### A. Randomized fixed-composition codes

In this section, we introduce the randomized fixed composition code. This is a subset of all coding schemes that can be used by the coding system. However, the understanding of this particular kind of codes may help us to understand the nature of the interference channels. First we introduce the notion of type set [1].

A type set $\mathcal{T}^n(P)$ is a set of all the strings $x^n \in \mathcal{X}^n$ with the same type $P$ where $P$ is a probability distribution [1]. A sequence of type sets $\mathcal{T}^n \subseteq \mathcal{X}^n$ has composition $P_X$ if the types of $\mathcal{T}^n$ converges to $P_X$, i.e. $\lim\limits_{n\to\infty} \frac{N(a|\mathcal{T}^n)}{n} = P_X(a)$ for all $a \in \mathcal{X}$ and $P_X(a) > 0$ and $N(a|\mathcal{T}^n) = 0$ for all $a \in \mathcal{X}$ and $P_X(a) = 0$, where $N(a|\mathcal{T}^n)$ is the number of occurrence of $a$ in type $\mathcal{T}^n$.

We ignore the nuisance in the integer effect and assume that $nP_X(a)$ is an integer for all $a \in \mathcal{X}$, this is indeed a reasonable assumption since we study long block length $n$ and all the information theoretic quantities studied in this paper are continuous on the distributions $P_X$. We will simply denote by $\mathcal{T}^n(P_X)$ the length-$n$ type set which has "asymptotic" type $P_X$, later in the appendix we abuse the notations by simply writing $x^n \in P_X$ instead of $x^n \in \mathcal{T}^n(P_X)$. Similarly, we assume that $nR_x$ and $nR_y$ are integers.

In this paper, we are concerned with the randomized fixed-composition codes, where a code word for message $i$ is uniformly i.i.d distributed on the type set $\mathcal{T}^n(P_X)$, formally defined as follows.

*Definition 1:* Randomized fixed-composition codes: for a probability distribution $P_X$ on $\mathcal{X}$, a rate $R_x$ randomized fixed-composition-$P_X$ encoder picks a code book with the following probability, for any fixed-composition-$P_X$ code book $\theta^n = (\theta^n(1), \theta^n(2), ..., \theta(2^{nR_x}))$, where $\theta^n(i) \in \mathcal{T}^n(P_X)$, $i = 1, 2, ..., 2^{nR_x}$, and $\theta^n(i)$ and $\theta^n(j)$ may not be different for $i \neq j$, the code book $\theta_n$ is chosen, i.e. $x^n(i) = \theta^n(i), \quad i = 1, 2, ..., 2^{nR_x}$ , with probability

$$\left(\frac{1}{|\mathcal{T}^n(P_X)|}\right)^{2^{nR_x}}$$

C. Chang is with the Hewlett-Packard Laboratories, Palo Alto, CA
Email: cheng.chang@hp.com

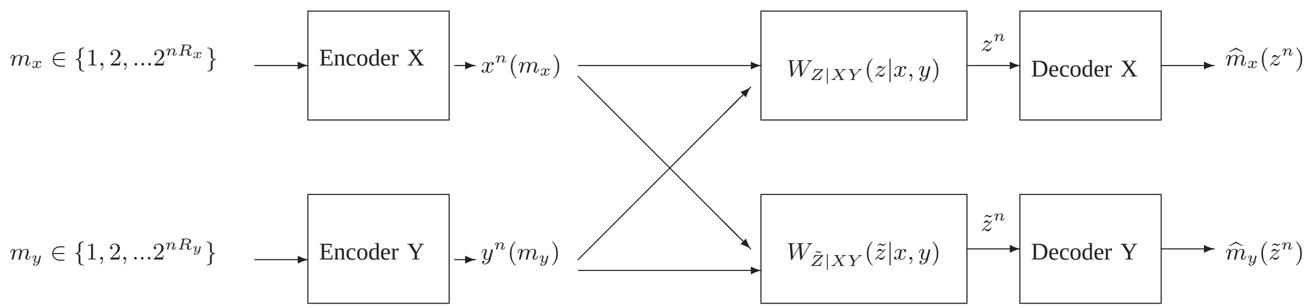


Fig. 1. A discrete memoryless interference channel of two users

In other words, the choice of the code book is a random variable $c_X$ uniformly distributed on the index set of all the possible code books with fixed-composition $P_X$: $\{1, 2, 3, ..., |\mathcal{T}^n(P_X)|^{2^{nR_x}}\}$, while $c_X$ is shared between the encoder $X$ and the decoders $X$ and $Y$.

The key property of the randomized fixed-composition code is that for any message subset $\{i_1, i_2, ...i_l\} \subseteq \{1, 2, ..., 2^{nR_x}\}$, the code words for these messages are identical independently distributed on the type set of $P_X$.

For randomized fixed-composition codes, the average error probability $P^n_{e(x)}(R_x, R_y, P_X, P_Y)$ for X is the expectation of decoding error over all code books and all channel behaviors.

$$\begin{aligned}
&P^n_{e(x)}(R_x, R_y, P_X, P_Y) \\
&= \left(\frac{1}{|\mathcal{T}^n(P_X)|}\right)^{2^{nR_x}} \left(\frac{1}{|\mathcal{T}^n(P_Y)|}\right)^{2^{nR_y}} \\
&\quad \sum_{c_X}\sum_{c_Y} \frac{1}{2^{nR_x}} \sum_{m_x} \frac{1}{2^{nR_y}} \sum_{m_y}\sum_{z^n} \\
&\quad W_{Z|XY}(z^n|x^n(m_x), y^n(m_y))1(\widehat{m}_x(z^n) \neq m_x)
\end{aligned} \tag{1}$$

where $x^n(m_x)$ is the code word of message $m_x$ in code book $c_X$, similarly for $y^n(m_y)$, $\widehat{m}_x(z^n)$ is the decision made by the decoder knowing the code books $c_X$ and $c_Y$.

### *B. Randomized fixed-composition coding capacity for interference channels*

Given the definitions of randomized fixed-composition coding and the average error probability in (1) for such codes, we can formally define the capacity region for such codes.

*Definition 2:* Capacity region for randomized fixed-composition codes: for a fixed composition $P_X$ and $P_Y$, a rate pair $(R_x, R_y)$ is said to be achievable for $X$, if for all $\delta > 0$, there exists $N_\delta < \infty$, s.t. for all $n > N_\delta$,

$$P^n_{e(x)}(R_x, R_y, P_X, P_Y) < \delta \tag{2}$$

We denote by $\mathcal{R}_x(P_X, P_Y)$ the closure of the union of the all achievable rate pairs. Similarly we can define the achievable region for $Y$, $\mathcal{R}_y(P_X, P_Y)$ and $(X, Y)$, $\mathcal{R}_{xy}(P_X, P_Y)$ where both decoding errors are small and obviously

$$\mathcal{R}_{xy}(P_X, P_Y) = \mathcal{R}_x(P_X, P_Y) \bigcap \mathcal{R}_y(P_X, P_Y). \tag{3}$$

We only need to focus our investigation on $\mathcal{R}_x(P_X, P_Y)$.

## II. Capacity region for fixed-composition code

The main result of this paper is the complete characterization of the capacity region for randomized fixed-composition codes: $\mathcal{R}_x(P_X, P_Y)$. The region is illustrated in Figure 2. $\mathcal{R}_x(P_X, P_Y)$ is the union of Region $I$ and $II$.

*Theorem 1:* Interference channel capacity region $\mathcal{R}_x(P_X, P_Y)$ for randomized fixed-composition codes with composition $P_X$ and $P_Y$:

$$\begin{aligned}
&\mathcal{R}_x(P_X, P_Y) \\
&= \{(R_x, R_y) : 0 \le R_x < I(X; Z), 0 \le R_y\} \bigcup \\
&\quad \{(R_x, R_y) : 0 \le R_x < I(X; Z|Y), \\
&\qquad R_x + R_y < I(X, Y; Z)\}
\end{aligned} \tag{4}$$

where the random variables in (4), $(X, Y, Z) \sim P_X P_Y W_{Z|X,Y}$.

The achievable part of the theorem states that: for a rate pair $(R_x, R_y) \in \mathcal{R}_x(P_X, P_Y)$, the union of Region $I$ and $II$ in Figure 2, for all $\delta > 0$, there exists $N_\delta < \infty$, s.t. for all $n > N_\delta$, the average error probability (1) for the randomized code from compositions $P_X$ and $P_Y$ is smaller than $\delta$ for $X$:

$$P^n_{e(x)}(R_x, R_y, P_X, P_Y) < \delta$$

for some decoding rule. Region $II$ is also the multiple-access capacity region for fixed composition codes $(P_X, P_Y)$ for channel $W_{Z|XY}$.

The converse of the theorem states that for any rate pair in the interior of the complement of $\mathcal{R}_x(P_X, P_Y)$, region $III$, $IV$ and $IV$ in Figure 2, there exists $\delta > 0$, such that for all $n$,

$$P^n_{e(x)}(R_x, R_y, P_X, P_Y) > \delta$$

no matter what decoding rule is used.

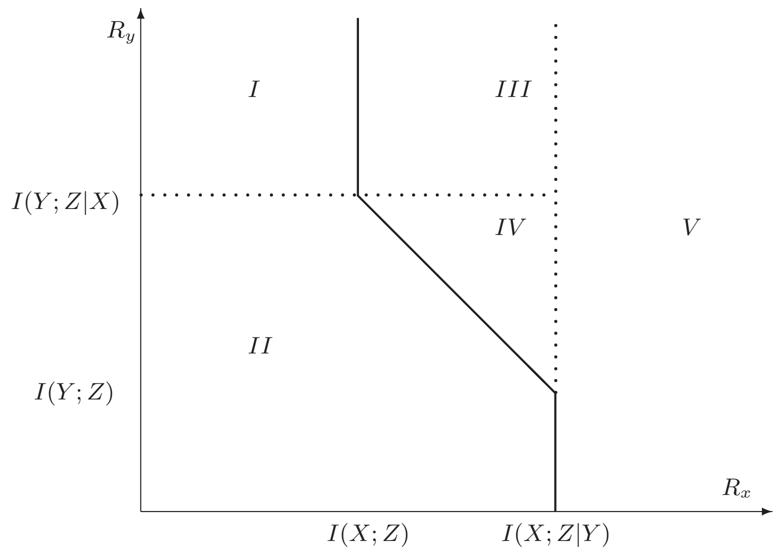


Fig. 2. Randomized fixed-composition capacity region $\mathcal{R}_x(P_X,P_Y)$ for $X$, the achievable region is the union of Region $I$ and $II$.

The proof of Theorem 1 is in Section III.

In the achievability part of Theorem 1, we prove that the average error probability for decoder $X$ is small for a randomized fixed-composition code if the rate pair $(R_x,R_y)$ is inside the capacity region $\mathcal{R}_x(P_X,P_Y)$. For interference channels, it is desired to have error probabilities small for both user $X$ and $Y$. It is obvious that the rate region for this problem is

$$\mathcal{R}_{xy}(P_X,P_Y)=\mathcal{R}_x(P_X,P_Y)\cap\mathcal{R}_y(P_X,P_Y),$$

where $\mathcal{R}_y(P_X,P_Y)$ is defined in the same manner as $\mathcal{R}_x(P_X,P_Y)$ but the channel is $W_{\tilde{Z}|XY}$ instead of $W_{Z|XY}$ as shown in Figure 1. A typical capacity region $\mathcal{R}_{xy}(P_X,P_Y)$ is shown in Figure 3. It is not necessarily convex.

However, by time sharing between different rate pairs for the same composition, we can convexify the capacity region. Then the convex hull of the union of all such capacity regions for different compositions gives the biggest known achievable capacity region from our coding scheme. i.e. the capacity region of the interference channel is a superset of

$$CONVEX\left(\bigcup_{P_X,P_Y}\mathcal{R}_{xy}(P_X,P_Y)\right).$$

A more thorough discussion of time sharing and the achievable error exponent and capacity region is detailed in Section IV.

### A. Existence of a good code for the interference channel

In this paper we are concerned with the average error probability over the code book ensemble with the same composition. This is in contrast to the case where only the existence of a code book is needed. The achievability of the randomized coding implies the existence of *a* good code book. This can be proved as a simple corollary of Theorem 1. Similar to the error probability for $X$ defined in (1), we define the average error

$$\begin{aligned}
&P^n_{e(xy)}(R_x,R_y,P_X,P_Y)\\
=\;&\left(\frac{1}{|\mathcal{T}^n(P_X)|}\right)^{2^{nR_x}}\left(\frac{1}{|\mathcal{T}^n(P_Y)|}\right)^{2^{nR_y}}\\
&\sum_{c_X}\sum_{c_Y}\frac{1}{2^{nR_x}}\sum_{m_x}\frac{1}{2^{nR_y}}\sum_{m_y} \qquad (5)\\
&\Big\{\sum_{z^n}W_{Z|XY}(z^n|x^n(m_x),y^n(m_y))1(\widehat{m}_x(z^n)\neq m_x)\\
&+\sum_{\tilde{z}^n}W_{\tilde{Z}|XY}(\tilde{z}^n|x^n(m_x),y^n(m_y))1(\widehat{m}_y(\tilde{z}^n)\neq m_y)\Big\}
\end{aligned}$$

For a rate pair $(R_x,R_y)\in\mathcal{R}_{xy}(P_X,P_Y)=\mathcal{R}_x(P_X,P_Y)\bigcap\mathcal{R}_y(P_X,P_Y)$. We know that for all $\delta>0$, there exists $N_\delta<\infty$, s.t. for all $n>N_\delta$, the average error probability is smaller than $\delta$ for user $X$ and user $Y$: $P^n_{e(x)}(R_x,R_y,P_X,P_Y)<\delta$ and $P^n_{e(y)}(R_x,R_y,P_X,P_Y)<\delta$. It is easy to see that the average error probability for user $X$ and $Y$ can be bounded by:

$$\begin{aligned}
&P^n_{e(xy)}(R_x,R_y,P_X,P_Y)\\
&\quad=P^n_{e(x)}(R_x,R_y,P_X,P_Y)+P^n_{e(y)}(R_x,R_y,P_X,P_Y)\\
&\quad\leq 2\delta \qquad (6)
\end{aligned}$$

From (5), we know that $P^n_{e(xy)}(R_x,R_y,P_X,P_Y)$ is the average error probability of *all* fixed-composition codes. With (6), we know that there exists at least *a* codebook such that the error probability is no bigger than $2\delta$.

The existence of a code book that achieves the error exponents can also be shown. The proof is similar to that in [9] and Exercise 30 (b) on page 198 [3]. This is quite obvious.

The converse of the randomized coding does not guarantee that there is not a single good fixed-composition code book.

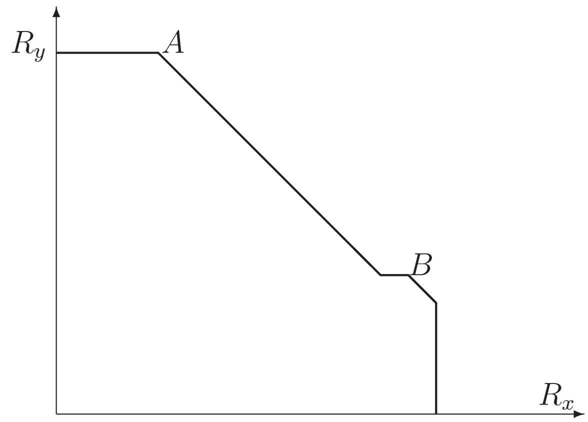


Fig. 3. A typical randomized fixed-composition capacity region $\mathcal{R}_{xy}(P_X, P_Y) = \mathcal{R}_x(P_X, P_Y) \cap \mathcal{R}_y(P_X, P_Y)$, this capacity region is not necessarily convex.

Our converse does not claim that no such code book exists. Instead, we claim that, the average decoding error probability does not converge to $0$ if the rate pair is outside the capacity region in Theorem 1.

## III. Proof of Theorem 1

There are two parts of the theorem, achievability and converse. The achievability part is proved by applying the classical techniques in point to point channel coding and multiple access channel coding for randomized fixed-composition code. The converse is proved by contradiction and using a technique first developed in [4].

### A. Achievability

We show that in the interior of the capacity region, i.e. the union of Region $I$ and $II$ in Figure 2, a positive error exponent is achieved by applying the randomized fixed-composition coding defined in Definition 1. We present the error exponent results in Lemmas 1 and 2 that covers Region $I$ and $II$ respectively. Then in Lemma 3, we show that these error exponents are positive in the interior of the capacity region $\mathcal{R}_x(P_X, P_Y)$ and hence conclude the proof of the achievability part in Theorem 1.

*1) Region $II$:* In Region $II$, we show that decoder $X$ can decode both message $m_x$ and $m_y$ with small error probabilities. This is essentially a multiple-access channel coding problem. We use the technique developed in [3] to derive the positive error exponents that parallel to those in [11]. The decoder is a simple maximum mutual information[1] decoder [3], this decoding rule is universal in the sense that the decoder does not need to know the multiple access channel $W_{Z|XY}$. We describe the decoding rule here, the estimate of the joint message is the message pair such that the input to the channel $W_{Z|XY}$ and the output of the channel have the maximal empirical mutual information. i.e.:

$$(\widehat{m}_x(z^n), \widehat{m}_y(z^n)) = \arg\max_{i \in \{1,2,\ldots,2^{nR_x}\}, j \in \{1,2,\ldots,2^{nR_y}\}} I(z^n; x^n(i), y^n(j)) \quad (7)$$

where $z^n$ is the channel output and $x^n(i)$ and $y^n(j)$ are the channel inputs for message $i$ and $j$ respectively. $I(z^n; x^n, y^n)$ is the empirical mutual information between $z^n$ and $(x^n, y^n)$, the point to point maximal mutual mutual information decoding is studied in [3].

If there is a tie, the decoder can choose an arbitrary winner or simply declare error. In Lemma 1, we show that by using the randomized fixed composition encoding and the maximal mutual information decoding, a non-negative error exponent is achieved in Region $II$.

*2) Region $I$:* In Region $I$, decoder $X$ only estimates $m_x$ by treating the input of encoder $Y$ as a source of random noise. This is a point to point channel coding problem. The channel itself has memory since the input of encoder $Y$ is not memoryless. Similar to the multiple access channel coding problem studied in Region $II$, we use a maximal mutual information decoding rule:

$$\widehat{m}_x(z^n) = \arg\max_{i \in \{1,2,\ldots,2^{nR_x}\}} I(z^n; x^n(i)) \quad (8)$$

In Lemma 2, we show that the by using the randomized fixed composition encoding and the maximal mutual information decoding, a non-negative error exponent is achieved in Region $I$.

We use the method of types [2] in both proofs.

*Lemma 1:* (Region $II$) Multiple-access channel error exponents (joint error probability). For the randomized coding scheme described in Definition 1, and the decoding rule described in (7), the decoding error probability averaged over all messages, code books and channel behaviors is upper bounded by an exponential term:

$$\begin{aligned}
&\Pr((m_x, m_y) \neq (\widehat{m}_x, \widehat{m}_y)) \\
&= \left(\frac{1}{|\mathcal{T}^n(P_X)|}\right)^{2^{nR_x}} \left(\frac{1}{|\mathcal{T}^n(P_Y)|}\right)^{2^{nR_y}} \\
&\quad \sum_{c_X}\sum_{c_Y} \frac{1}{2^{nR_x}} \sum_{m_x} \frac{1}{2^{nR_y}} \sum_{m_y}\sum_{z^n} \qquad (9)\\
&\quad W_{Z|XY}(z^n|x^n(m_x), y^n(m_y)) \\
&\quad 1\left((\widehat{m}_x(z^n), \widehat{m}_y(z^n)) \neq (m_x, m_y)\right) \\
&\leq 2^{-n(E-\epsilon_n)}. \qquad (10)
\end{aligned}$$

$\epsilon_n$ converges to zero as $n$ goes to infinity, and

$$E = \min\{E_{xy}, E_{x|y}, E_{y|x}\}, \text{ where}$$

[1] A more sophisticated decoding rule based on minimum conditional entropy decoding for multiple-access channel is developed in [10], it is shown that this decoding rule achieves a bigger error exponent in low rate regime. The goal of this paper is, however, not to derive the tightest lower bound on the error exponent. We only need a coding scheme to achieve positive error exponent in the capacity region in Theorem 1. Hence we use the simpler decoding scheme in [11].

$$\begin{aligned} E_{xy} &= \min_{Q_{XYZ}:Q_X=P_X,Q_Y=P_Y} D(Q_{Z|XY}\|W|Q_{XY}) \\ &\quad +D(Q_{XY}\|P_X\times P_Y) \\ &\quad +|I_Q(X,Y;Z)-R_x-R_y|^+ \\ E_{x|y} &= \min_{Q_{XYZ}:Q_X=P_X,Q_Y=P_Y} D(Q_{Z|XY}\|W|Q_{XY}) \\ &\quad +D(Q_{XY}\|P_X\times P_Y)+|I_Q(X;Z|Y)-R_x|^+ \\ E_{y|x} &= \min_{Q_{XYZ}:Q_X=P_X,Q_Y=P_Y} D(Q_{Z|XY}\|W|Q_{XY}) \\ &\quad +D(Q_{XY}\|P_X\times P_Y)+|I_Q(Y;Z|X)-R_y|^+ \end{aligned}$$

where $|t|^+ = \max\{0,t\}$ and the random variables $(X,Y,Z) \sim Q_{XYZ}$ in $I_Q(X;Z|Y), I_Q(Y;Z|X)$ and $I_Q(X,Y;Z)$.

*Remark: it is easy to verify that $D(Q_{Z|XY}\|W|Q_{XY}) + D(Q_{XY}\|P_X\times P_Y) = D(Q_{XYZ}\|P_X\times P_Y\times W)$, so the expressions for the error exponents can be further simplified. We use the expressions similar to those in [11] because they are more intuitive.*

*Remark: The proof parallels to that in [11] which is in turn an extension to the point to point channel coding problem studied in [3]. The method of types is the main tool for the proofs. The difference is that we need to show the lower bound to the average error probability instead of showing the existence of a good code book in [11]. Without giving details, we follow Gallager's proof in [9] and claim the existence of a good code with the same error exponent as that in [11]. This is a simple corollary of Lemma 1.*

*Proof:* First we have an obvious upper bound on the error probability

$$\begin{aligned} &\Pr((m_x,m_y)\neq(\widehat{m}_x,\widehat{m}_y)) \\ &= \Pr(m_x\neq\widehat{m}_x, m_y\neq\widehat{m}_y)+\Pr(m_x\neq\widehat{m}_x, m_y=\widehat{m}_y) \\ &\quad +\Pr(m_x=\widehat{m}_x, m_y\neq\widehat{m}_y) \\ &\leq \Pr(m_x\neq\widehat{m}_x, m_y\neq\widehat{m}_y)+\Pr(m_x\neq\widehat{m}_x|m_y=\widehat{m}_y) \\ &\quad +\Pr(m_y\neq\widehat{m}_y|m_x=\widehat{m}_x)) \end{aligned} \tag{11}$$

The inequality is true because $P(A,B)=P(A|B)P(B)\leq P(A|B)$. Now we upper bound each individual error probability in (11) respectively by exponentials. By symmetry, we only need to show that

$$\Pr(m_x\neq\widehat{m}_x, m_y\neq\widehat{m}_y)\leq 2^{-n(E_{xy}-\epsilon_n)} \tag{12}$$

and

$$\Pr(m_x\neq\widehat{m}_x|m_y=\widehat{m}_y)\leq 2^{-n(E_{x|y}-\epsilon_n)}. \tag{13}$$

We leave the proof of (12) and (13) to Appendix A, where a standard method of type argument is used. □

*Lemma 2:* (Region $I$) point to point channel coding error exponent (decoding $X$ only). For the randomized coding scheme described in Definition 1, and the decoding rule described in (8), the decoding error probability averaged over all messages, code books and channel behaviors is upper bounded by an exponential term:

$$\begin{aligned} &\Pr(m_x\neq\widehat{m}_x) \\ &= \left(\frac{1}{|\mathcal{T}^n(P_X)|}\right)^{2^{nR_x}}\left(\frac{1}{|\mathcal{T}^n(P_Y)|}\right)^{2^{nR_y}} \\ &\quad \sum_{c_X}\sum_{c_Y}\frac{1}{2^{nR_x}}\sum_{m_x}\frac{1}{2^{nR_y}}\sum_{m_y}\sum_{z^n} \\ &\quad W_{Z|XY}(z^n|x^n(m_x),y^n(m_y)) \\ &\quad 1\left(\widehat{m}_x(z^n)\neq m_x\right) \\ &\leq 2^{-n(E_x-\epsilon_n)}. \end{aligned} \tag{14}$$

$\epsilon_n$ converges to zero as $n$ goes to infinity, and

$$\begin{aligned} E_x &= \min_{Q_{XYZ}:Q_X=P_X,Q_Y=P_Y} D(Q_{Z|XY}\|W|Q_{XY}) \\ &\quad +D(Q_{XY}\|P_X\times P_Y)+|I_Q(X;Z)-R_x|^+ \end{aligned}$$

*Proof:* We give a unified proof for (12), (13) and (14) in Appendix A. □

With Lemma 1 and Lemma 2, we know that some non-negative error exponents can be achieved as long as the rate pair $(R_x,R_y)\in\mathcal{R}_x(P_X,P_Y)$ for randomized fixed-composition code. This is because both Kullback-Leibler divergence and $|\cdot|^+$ are always non-negative. Now we only need to show the positiveness of those error exponents when the rate pair is in the interior of $\mathcal{R}_x(P_X,P_Y)$.

*Lemma 3:* Positiveness of the error exponents, for rate pairs $(R_x,R_y)$ in the interior of $\mathcal{R}_x(P_X,P_Y)$ defined in Theorem 1, we have:

$$\max\{\min\{E_{xy},E_{x|y},E_{y|x}\},E_x\}>0.$$

More specifically, we will show two things. First, if $R_x<I(X,Z)$, where $(X,Z)\sim P_X\times P_Y\times W_{Z|XY}$, then $E_x>0$. This covers Region $I$. Second, if $R_x<I(X,Z|Y)$, $R_y<I(Y,Z|X)$ and $R_x+R_y<I(X,Y;Z)$, where $(X,Y,Z)\sim P_X\times P_Y\times W_{Z|XY}$, then $\min\{E_{xy},E_{x|y},E_{y|x}\}>0$, this covers Region $II$.

*Proof:* First, suppose that for some $R_x<I(X,Z)$, $E_x\leq 0$. Since both Kullback-Leibler divergence and $|\cdot|^+$ are non-negative functions, we must have $E_x=0$ and there exists a distribution $Q_{XYZ}$, s.t. $Q_X=P_X$, $Q_Y=P_Y$ and all the individual non-negative functions are zero:

$$\begin{aligned} D(Q_{XY}\|P_X\times P_Y) &= 0 \\ D(Q_{Z|XY}\|W|Q_{XY}) &= 0 \\ |I_Q(X;Z)-R_x|^+ &= 0 \end{aligned}$$

The first equation tells us that $Q_{XY}=P_X\times P_Y$. Then the second equation becomes $D(Q_{Z|XY}\|W|P_X\times P_Y)=0$, this means that $Q_{Z|XY}\times P_X\times P_Y=W\times P_X\times P_Y$, so $I_Q(X;Z)=I(X;Z)$ where the random variables $(X,Y,Z)\sim P_X\times P_Y\times W_{Z|XY}$ in $I(X;Z)$. Now the third equation becomes $|I(X;Z)-R_x|^+=0$ which is equivalent to $I(X;Z)\leq R_x$, this is a contradiction to the fact that $R_x<I(X,Z)$.

Secondly, suppose that for some rate pair $(R_x, R_y)$ in Region $II$, i.e. $R_x < I(X,Z|Y)$, $R_y < I(Y,Z|X)$ and $R_x + R_y < I(X,Y;Z)$ and $\min\{E_{xy}, E_{x|y}, E_{y|x}\} \leq 0$, then $\min\{E_{xy} = 0$ or $E_{x|y} = 0$ or $E_{y|x}\} = 0$. Following exactly the same argument as that in the first part of the proof of Lemma 3, we can get contradictions with the fact that the rate pair $(R_x, R_y)$ is in the interior of capacity region $\mathcal{R}_x(P_X, P_Y)$. □

From the above three lemmas, we know that the error probability for decoding message $X$ is upper bounded by $2^{-n(E-\epsilon_n)}$ for all $(R_x, R_y) \in \mathcal{R}_x(P_X, P_Y)$, where $E > 0$ and $\lim_{n\to\infty} \epsilon_n = 0$. Hence the error probability goes to zero exponentially fast for large $n$. This concludes the achievability part of the proof for Theorem 1.

### B. Converse

We show that the average decoding error of Decoder $X$ does not converge to 0 with increasing $n$ if the rate pair $(R_x, R_y)$ is outside the capacity region $\mathcal{R}_x(P_X, P_Y)$ shown in Figure 2. There are two parts of the proof.

First, we show that in Region $V$ the average error probability does not go to zero as block length goes to infinity. This is proved by using a modified version of the reliability function for rate higher than the channel capacity [4].

*Lemma 4:* Region $V$, the average error probability for $X$ does not converge to 0 with block length $n$ if $R_x > I(X;Z|Y)$, where $(X,Y,Z) \sim P_X \times P_Y \times W_{Z|XY}$.

*Proof:* It is enough to show the case where there is only one message for $Y$ and encoder $Y$ sends a code word $y^n$ with composition $P_Y$. The code book for encoder $X$ is still uniformly generated among all the fixed-composition-$P_X$ code books. In the rest of the proof, we investigate the typical behavior of the codewords $x^n$ and modify the Lemma 3 and Lemma 5 from [4] to show that

$$\Pr(\widehat{m}_x \neq m_x) = P^n_{e(x)}(R_x, R_y, P_X, P_Y) > \frac{1}{2} \qquad (15)$$

for large $n$. The details of the proof are in Appendix I-B. □

The more complicated case is in Region $IV$. We show that the decoding error probability for user $X$ does not converge to zero with block length $n$ by constructing a decoder that decodes both message $m_x$ and message $m_y$ correctly with high probability. Then again by using the reliability function for rate higher than channel capacity [4], we get a contradiction.

*Lemma 5:* Region $IV$, the average error probability for $X$ does not converge to 0 with block length $n$ if $R_x < I(X;Z|Y)$, $R_y < I(Y;Z|X)$ and $R_x + R_y > I(X,Y;Z)$ where $(X,Y,Z) \sim P_X \times P_Y \times W_{Z|XY}$.

*Proof:* Suppose that

$$\Pr(\widehat{m}_x \neq m_x) = P^n_{e(x)}(R_x, R_y, P_X, P_Y) \leq \delta_n$$

where $\delta_n$ goes to zero with $n$. Now let decoder $X$ decode $m_y$ by the same decoding rule devised in (7):

$$\widehat{m}_y(z^n) = \underset{j \in \{1,2,\ldots,2^{nR_y}\}}{\arg\max} I(z^n; x^n(\widehat{m}_x(z^n)), y^n(j)). \qquad (16)$$

The decoding error for either message at decoder $X$ is now:
$\Pr((\widehat{m}_x, \widehat{m}_y) \neq (m_x, m_y))$

$$\begin{aligned} &= \Pr(\widehat{m}_x \neq m_x) + \Pr(\widehat{m}_x = m_x, \widehat{m}_y \neq m_y) \\ &\leq \Pr(\widehat{m}_x \neq m_x) + \Pr(\widehat{m}_y \neq m_y | \widehat{m}_x = m_x) \qquad (17) \end{aligned}$$

Given $\widehat{m}_x = m_x$, (18) becomes

$$\widehat{m}_y(z^n) = \underset{j \in \{1,2,\ldots,2^{nR_y}\}}{\arg\max} I(z^n; x^n(m_x), y^n(j)). \qquad (18)$$

So the second term in the RHS of (17) can be bounded by *exactly* the same way as that in the proof of (13). From (17) we know:

$$\Pr((\widehat{m}_x, \widehat{m}_y) \neq (m_x, m_y)) \leq \delta_n + 2^{-n(E_{y|x} - \epsilon_n)} \qquad (19)$$

This upper bound goes to zero as $n$ goes to infinity. However in Appendix I-B, we show that

$$\begin{aligned} &P^n_{e(xy)}(R_x, R_y, P_X, P_Y) \\ &\qquad = \Pr((\widehat{m}_x, \widehat{m}_y) \neq (m_x, m_y)) > \frac{1}{2} \qquad (20) \end{aligned}$$

This is contradicted to (19). □

As a corollary of Lemma 5, the decoding error for encoder $X$ does not converge to 0 with $n$ if the rate pair $(R_x, R_y)$ is in Region $III$. For a $(R_x, R_y)$ decoder, we can construct a new decoder for $(R_x, R'_y)$ where $R'_y < R_y$, by revealing a random selection of a $(R_x, R_y)$ code book that is the superset of the $(R_x, R'_y)$ code book to the $(R_x, R_y)$ decoder and accept the estimate of the $(R_x, R_y)$ decoder as the estimate for the $(R_x, R'_y)$ decoder. If the average error probability is small for the $(R_x, R_y)$ code books, the average error probability is small for this particular $(R_x, R'_y)$ decoder as well, this is a contradiction to Lemma 5.

This concludes the converse part of the proof for Theorem 1.

## IV. Beyond Fixed-composition

There are two parts in this section. First we investigate the error exponent performance for simple time sharing among different rate pairs for the same randomized fixed-composition codes, and time sharing among different compositions. In the second part, we study the necessity of more sophisticated time-sharing coding schemes.

### A. Simple time sharing exponents

The simple idea of time sharing is well developed and understood for multi-user information theory, especially in the capacity region results for multiple-access channel coding and broadcast channel coding.

*Definition 3:* Time-sharing randomized code: for a positive vector $(\xi_1, \ldots \xi_L)$ and $\sum_i \xi_i = 1$, a $\big((\xi_1, R_x^{(1)}, R_y^{(1)}, P_X^{(1)}, P_Y^{(1)}), \ldots, (\xi_L, R_x^{(L)}, R_y^{(L)}, P_X^{(L)}, P_Y^{(L)})\big)$ randomized code is such that a map for source $X$, $\mathcal{E}_x$:

$$\{1, 2, \ldots, 2^{nR_x}\} \to \{1, 2, \ldots, 2^{n\xi_1 R_x^{(1)}}\} \times \ldots \times \{1, 2, \ldots, 2^{n\xi_L R_x^{(L)}}\}$$

$$\mathcal{E}_x(m_x) = (m_x^{(1)}, \ldots m_x^{(L)})$$

where $m_x^{(1)}$ is the first $n\xi_1 R_x^{(1)}$ bits of $m_x$ and so on. Then for each $m_x^{(i)}$, we use the randomized fixed-composition code with composition $P_X^{(i)}$ defined in Definition 1. The rate for this encoder is $R_x$ where $R_x = \sum_{i=1}^{L} \xi_i R_x^{(i)}$.

Similarly for $\mathcal{E}_y$.

The decoder simply decodes $(m_x^{(i)}, m_y^{(i)})$ individually for $i = 1, ..., L$. And put them back to form the whole estimation $\widehat{m}_x, \widehat{m}_y$.

The decoding error probability is upper bounded by the union bound of the error probability of individual messages.

$$\Pr(\widehat{m}_x \neq m_x) \leq L \max_{i \in \{1,2,...,L\}: R_x^{(i)} > 0} \Pr(\widehat{m}_x^{(i)} \neq m_x^{(i)})$$

$L$ is a fixed number, so let $n$ goes to infinity, the overall error exponents is

$$E^x = \min_{i \in \{1,2,...,L\}: R_x^{(i)} > 0} \xi_i E_x(R_x^{(i)}, R_y^{(i)}, P_X^{(i)}, P_Y^{(i)}) \quad (21)$$

where $E_x(R_x^{(i)}, R_y^{(i)}, P_X^{(i)}, P_Y^{(i)})$ is the randomized fixed-composition error exponent for rate $(R_x^{(i)}, R_y^{(i)})$ with fixed compositions $(P_X^{(i)}, P_Y^{(i)})$ defined in Lemma 3. Similarly:

$$E^y = \min_{i \in \{1,2,...,L\}: R_y^{(i)} > 0} \xi_i E_y(R_x^{(i)}, R_y^{(i)}, P_X^{(i)}, P_Y^{(i)}) \quad (22)$$

With (21) and (22), we know that a positive error exponent pair can be achieved by properly time-share among different fixed-composition codes $(R_x, R_y, P_X, P_Y)$ for rate pairs in the interior of

$$CONVEX\left(\bigcup_{P_X,P_Y} \mathcal{R}_{xy}(P_X, P_Y)\right) =$$

$$CONVEX\left(\bigcup_{P_X,P_Y} \mathcal{R}_x(P_X, P_Y) \bigcap R_y(P_X, P_Y)\right). \quad (23)$$

The last equality is by the definition of $R_{xy}(P_X, P_Y)$.

The order of the union and intersection operators cannot be changed in (23). We further our discussion on the achievable capacity region (23). Fix the composition of the two encoders at $P_X$ and $P_Y$, the region where a positive error exponent can be achieved by $X$ is defined as $\mathcal{R}_x(P_X, P_Y)$. The convex hull (time sharing region) of all such regions over all possible input compositions is

$$\begin{aligned} & CONVEX\left(\bigcup_{P_X,P_Y} \mathcal{R}_x(P_X, P_Y)\right) \\ = \quad & MAC(W_{Z|XY}) \bigcup P2P_X(W_{Z|XY}) \end{aligned}$$

where $P2P_X(W_{Z|XY}) =$

$$\{(R_x, R_y) : R_x \leq \max_{(X,Z) \sim Q_{XY} W_{Z|XY}} I(X;Z)\}$$

is the point-to-point capacity region for user $X$.

$MAC(W_{Z|XY})$ is the multiple access channel capacity region for $W$. In the analysis of the capacity region for multiple-access channels in [1], the authors give two interpretations of the same capacity region. First

$$\begin{aligned} CONVEX( \bigcup_{P_X \times P_Y \times W_{Z|XY}} & \{(R_x, R_y) : R_x \leq I(X;Z|Y), \\ & R_y \leq I(Y;Z|X), R_x + R_y \leq I(X,Y;Z)\}) \end{aligned}$$

Equivalently:

$$\begin{aligned} & CLOSURE( \bigcup_{P_U \times P_{X|U} \times P_{Y|U} \times W_{Z|XY}} \{(R_x, R_y) : \\ & R_x \leq I(X;Z|Y,U), R_y \leq I(Y;Z|X,U), \\ & R_x + R_y \leq I(X,Y;Z|U)\}) \end{aligned}$$

where $U$ is the time-sharing auxiliary random variable and $|U| = 4$.

All the above discussions in this subsection is for $X$ only. For interference channels, $Y$ has to be reliably communicated simultaneously, hence the intersection operator in (23). The order of the intersection operator and the time sharing operator cannot be changed due to this.

*Remark 1: It is interesting that we not only need to time-share between different compositions but also need to time-share between different rate pairs even for the same fixed-composition! This is due to the non-convexity of the capacity region $\mathcal{R}_x(P_X, P_Y)$ for randomized fixed-composition code shown in Figure 3. Clearly $E^x$ is positive as long as each $E_x(R_x^{(i)}, R_y^{(i)}, P_X^{(i)}, P_Y^{(i)})$ is positive. From $X$'s point of view, the best chance to get the message decoded correctly is to spread the codewords uniformly in the code space $\{x^n : x^n \in P_X\}$. But this might have a negative impact on $Y$. By the simple time-sharing technique among different rate pairs with the same composition, we randomly sample codewords from a subset of the fixed-composition set which is clearly less optimal than sampling over the whole fixed-composition set for $X$. This however helps the decoding of the other user as now the capacity region (with positive error exponents) is the convex hull of those not necessarily convex regions $\mathcal{R}_{xy}(P_X, P_Y)$.*

### B. Beyond simple time-sharing

The error exponent achieved by simple time-sharing is determined by two factors, the fraction of time a particular randomized code $(R_x^{(i)}, R_y^{(i)}, P_X^{(i)}, P_Y^{(i)})$ is used and the error exponents achieved by such code. This simple time-sharing technique is also used in achieving the multiple access channel capacity region [1]. After all, part of the achievablity is proved by reducing the interference channel coding problem into a multiple access channel problem. However, the difference between interference and multiple-access channels is the number of receivers. Now there are two receivers, the achievable region for a particular fixed-composition is the *intersection* of two non-convex regions.

In this section we give a time sharing coding scheme that was first developed by Gallager [8] and later further studied for universal decoding by Pokorny and Wallmeier [11]. This type of randomized time-sharing schemes not only achieves better error exponents, more importantly, we show that this might achieve **bigger** capacity region than the simple time-sharing scheme does! Unlike the multiple-access channels where the simple time-sharing achieves the whole capacity region, this is unique to the interference channels, due to the fact that the capacity region is the convex hull of the intersections of pairs of non-convex regions (convex or not is not the issue here, the real difference is the intersection operation).

The organization of this section parallel to that for the fixed-composition. We first introduce the randomized time-sharing coding scheme, then give the achievable error exponents and lastly drive the achievable rate region for such coding schemes. The proofs are omitted since they are extremely similar to those for the randomized fixed-composition codes.

*Definition 4:* Randomized time-sharing codes: for a probability distribution $P_U$ on $\mathcal{U}$, where $\mathcal{U} = \{u_1, u_2, ..., u_K\}$ with $\sum_{i=1}^K P_U(u_i) = 1$, and a pair of conditional independent distributions $P_{X|U}, P_{Y|U}$. We define the two codeword sets[2] as $X_c(n) = \{x^n : x_1^{nP_U(u1)} \in P_{X|u_1}, x_{nP_U(u_1)+1}^{n(P_U(u_1)+P_U(u_2))} \in P_{X|u_2}, ..., x^n_{n(1-P_U(u_1))} \in P_{X|u_L}\}$ i.e. the $i$'th chunk of the codeword $x^n$ with length $nP_U(u_i)$ has composition $P_{X|u_i}$, and similarly $Y_c(n) = \{y^n : y_1^{nP_U(u1)} \in P_{Y|u_1}, y_{nP_U(u_1)+1}^{n(P_U(u_1)+P_U(u_2))} \in P_{Y|u_2}, ..., y^n_{n(1-P_U(u_1))} \in P_{Y|u_L}\}$. A randomized time sharing code $(R_x, R_y, P_U P_{X|U} P_{Y|U})$ encoder picks a code book with the following probability: for any message $m_x \in \{1, 2, ..., 2^{nR_x}\}$, the code word $x^n(m_x)$ is uniformly distributed in $X_c(n)$, similarly for encoder Y.

After the code book is randomly generated and revealed to the decoder, the decoder uses a maximum mutual information decoding rule. Similar to the fixed-composition coding, the decoder needs to either decode both message $X$ and $Y$ jointly or simply treats $Y$ as noise and decode $X$ only, depending on where the rate pairs are in Region $I$ or $II$, as shown in Figure 4. The error probability we investigate is again the average error probability over all messages and codebooks.

*Theorem 2:* Interference channel capacity region $\mathcal{R}_x(P_U P_{X|U} P_{Y|U})$ for randomized time-sharing codes with composition $P_U P_{X|U} P_{Y|U}$:

$$\begin{aligned}
&\mathcal{R}_x(P_U P_{X|U} P_{Y|U}) \\
&= \{(R_x, R_y) : 0 \le R_x < I(X;Z|U), 0 \le R_y\} \bigcup \\
&\quad \{(R_x, R_y) : 0 \le R_x < I(X;Z|Y,U), \\
&\quad\quad R_x + R_y < I(X,Y;Z|U)\} \qquad (24)
\end{aligned}$$

where the random variables in (24), $(U, X, Y, Z) \sim P_U P_{X|U} P_{Y|U} W_{Z|X,Y}$.

The rate region defined in (24) itself does not give any new $X$-capacity regions for $X$, since this region is a subset of the convex hull(time-sharing) of the capacity regions in (4). But for the interference channel capacity, we will argue in next section that this coding scheme might give a strictly bigger capacity region than that given by the simple time-sharing of fixed composition codes in (23).

The proof of Theorem 2 is extremely similar to that of Theorem 1. We omit the details here. We only point out that the achievability part is proved by deriving a positive error exponent for rate pair in the interior of the capacity region defined in Theorem 2. As shown in [11] and also detailed in this paper for the randomized coding, the error exponents in Region $II$ of in Figure 4 is:

$$E = \min\{E_{xy}, E_{x|y}, E_{y|x}\}, \text{ where}$$

$$\begin{aligned}
E_{xy} &= \min_{Q_{XYZ|U}: Q_{X|U}=P_{X|U}, Q_{Y|U}=P_{Y|U}} D(Q_{Z|XY} \| W | Q_{XYU}) \\
&\quad + D(Q_{XY|U} \| P_{X|U} \times P_{Y|U} | U) \\
&\quad + |I_Q(X, Y; Z) - R_x - R_y|^+ \\
E_{x|y} &= \min_{Q_{XYZ|U}: Q_{X|U}=P_{X|U}, Q_{Y|U}=P_{Y|U}} D(Q_{Z|XY} \| W | Q_{XYU}) \\
&\quad + D(Q_{XY|U} \| P_{X|U} \times P_{Y|U} | U) + |I_Q(X; Z|Y) - R_x|^+ \\
E_{y|x} &= \min_{Q_{XYZ|U}: Q_{X|U}=P_{X|U}, Q_{Y|U}=P_{Y|U}} D(Q_{Z|XY} \| W | Q_{XYU}) \\
&\quad + D(Q_{XY|U} \| P_{X|U} \times P_{Y|U} | U) + |I_Q(Y; Z|X, U) - R_y|^+
\end{aligned}$$

This is the error exponents in Lemma 1 with a conditional auxiliary random variable $U$.

The error exponent in Region $I$ is

$$\begin{aligned}
E_x &= \\
&\min_{Q_{XYZ|U}: Q_{X|U}=P_{X|U}, Q_{Y|U}=P_{Y|U}} D(Q_{Z|XY} \| W | Q_{XYU}) \\
&+ D(Q_{XY|U} \| P_{X|U} \times P_{Y|U} | U) + |I_Q(X; Z|U) - R_x|^+
\end{aligned}$$

### C. *Why the coding scheme in Theorem 2 is useful in studying the capacity regions*

It is obvious that the time-sharing fixed-composition coding gives a bigger error exponent than the simple time sharing coding does. More interestingly, we argue that it might gives a bigger interference channel capacity region. First we write down the capacity region for the time-sharing fixed-composition coding:

$$\begin{aligned}
CONVEX \quad &\Big( \bigcup_{P_{X|U} P_{Y|U} P_U} \qquad (25) \\
&[\mathcal{R}_x(P_{X|U} P_{Y|U} P_U) \bigcap \mathcal{R}_y(P_{X|U} P_{Y|U} P_U)] \Big).
\end{aligned}$$

$U$ is a time sharing auxiliary random variable. Unlike the MAC coding problem, where simple time sharing of fixed composition codes achieve the full capacity region, it is not guaranteed for interference channels. The reason is the intersection operator in the achievable capacity regions (4) and (24). In the following example, we illustrate why (24) **might** be bigger than (4).

Suppose we have a symmetric interference channel, i.e. $\mathcal{R}_x(P_X, P_Y) = \mathcal{R}_y^T(P_Y, P_X)$ where $^T$ is the transpose operation. The comparison of simple timesharing capacity region and the more sophisticated time-sharing fixed composition capacity region are illustrated by a toy example in Figure 5.

[2]Again, we ignore the nuisance of the non-integers here.

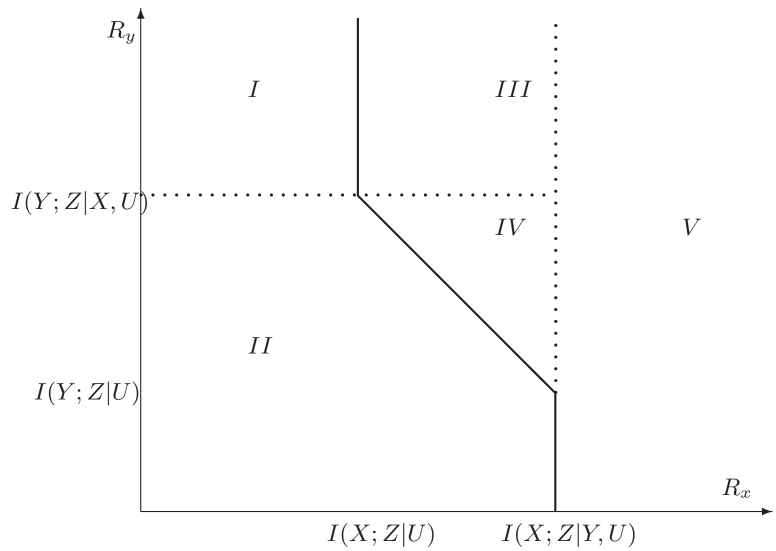


Fig. 4. Randomized time-sharing capacity region $\mathcal{R}_x(P_U P_{X|U} P_{Y_U})$ for $X$, the achievable region is the union of Region $I$ and $II$. This region is very similar to that for fixed-composition coding shown in Figure 2, only difference is now there is an auxiliary time-sharing random variable $U$.

For a distribution $(P_X, P_Y)$, the achievable region for the fixed composition code is illustrated in Figure 5, $\mathcal{R}_x(P_X, P_Y)$ and $\mathcal{R}_y(P_X, P_Y)$ respectively, these are bounded by the red dotted lines and red dash-dotted lines respectively, so the interference capacity region $\mathcal{R}_x y(P_X, P_Y)$ is bounded by the pentagon $ABEFO$. By symmetry, $\mathcal{R}_x(P_Y, P_X)$ and $\mathcal{R}_y(P_X, P_Y)$ are bounded by the blue dotted lines and blue dash-dotted lines respectively, the capacity region $\mathcal{R}_x y(P_Y, P_X)$ is bounded by the pentagon $HGCDO$. So the convex hull of these two regions is $ABCDO$.

Now consider the following timesharing fixed-composition coding $P_{X|U} P_{Y|U} P_U$ where $\mathcal{U} = \{0, 1\}$, $P_U(0) = P_U(1) = 0.5$ and $P_{X|0} = P_{Y|1} = P_X$, $P_{X|1} = P_{Y|0} = P_Y$. The interference capacity region is obviously bounded by the black pentagon in Figure 5. This toy example shows why (25) might be bigger than (23).

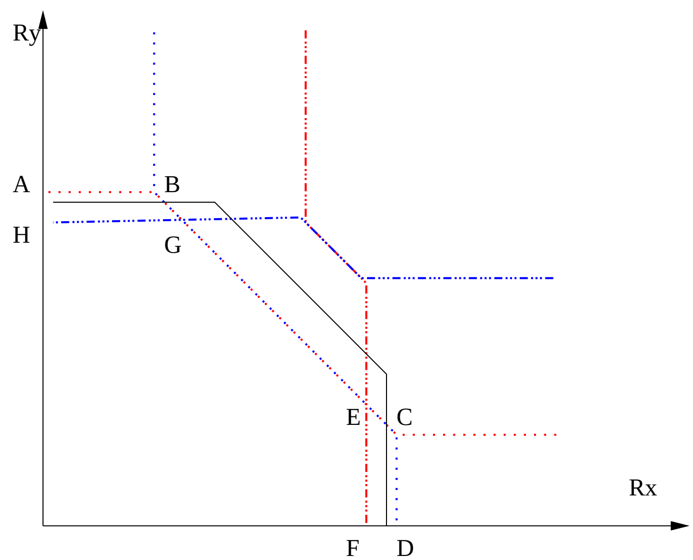


Fig. 5. Simple timesharing of fixed composition capacity $ABCDO$ VS time-sharing fixed composition capacity(0.5) ( the black pentagon)

## V. Discussions

In this paper we investigate the randomized fixed-composition coding error exponents for interference channels. We derive the standard random coding error exponents for interference channels. A better error exponent can be achievable by using more sophisticated coding schemes, as in the multiple access channel coding problem [10]. The capacity regions for such randomized coding are completely characterized. It is clear that this region is a subset of the well-known Han-Kobayashi region [12]. As a simple corollary, the existence of *a* good code is also proved. By time sharing between different rates and code compositions, a convex achievable rate region is derived for interference channels. The relation of this achievable capacity region and the Han-Kobayashi region [12] is unknown. An interesting future direction is to incorporate an auxiliary random variable into the code book generation. This coding scheme will certainly gives new error exponent results and possibly new achievable capacity region results.

## Acknowledgments

The author thanks Raul Etkin, Neri Merhav and Erik Ordentlich for introducing the problem and helpful discussions along the way.

## References

[1] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory, 2nd Edition.* John Wiley and Sons Inc., New York, 2006.
[2] Imre Csiszár. The method of types. *IEEE Transactions on Information Theory*, 44:2505–2523, 1998.
[3] Imre Csiszár and János Körner. *Information Theory*. Akadémiai Kiadó, Budapest, 1986.

[4] G Dueck and J Korner. Reliability function of a discrete memoryless channel at rates above capacity. *IEEE Transactions on Information Theory*, 25:82–85, 1979.

[5] Raul Etkin, Neri Merhav, and Erik Ordentlich. Error exponents of optimum decoding for the interference channel. *ISIT*, 2008.

[6] Raul Etkin, Neri Merhav, and Erik Ordentlich. Error exponents of optimum decoding for the interference chann. *IEEE Transactions on Information Theory*, In preparation.

[7] Robert Gallager. Fixed composition arguments and lower bounds to error probability. `http://web.mit.edu/gallager/www/notes/notes5.pdf`.

[8] Robert Gallager. A perspective on multiaccess channels. *IEEE Transactions on Information Theory*, 31, 1985.

[9] Robert G. Gallager. *Information Theory and Reliable Communication*. John Wiley, New York, NY, 1971.

[10] Yu-Sun Liu and Brian L. Hughes. A new universal random coding bound for the multiple access channel. *IEEE Transactions on Information Theory*, 42:376–386, 1996.

[11] Jutta Pokorny and Hans-Martin Wallmeier. Random coding bound and codes produced by permutations for the multiple-access channel. *IEEE Transactions on Information Theory*, 31:741 – 750, 1985.

[12] Te sun Han and Kingo Kobayashi. A new achievable rate region for the interference channel. *IEEE Transactions on Information Theory*, 27:49–60, 1981.

## APPENDIX

### A. Proof of (12), (13) and (14)

The expectation of the error probabilities in (12), (13) and (14) are taken over all messages, code books and channel behaviors. Because of the symmetry of the code book selection, we can fix the message pair $(m_x, m_y) = (1, 1)$.

We examine the object function to be minimized in (12), (13) and (14). First, the *common* part of the three error exponents $E_{xy}$, $E_{x|y}$ and $E_x$: $D(Q_{Z|XY}\|W|Q_{XY}) + D(Q_{XY}\|P_X \times P_Y)$. $D(Q_{XY}\|P_X \times P_Y)$ is the logarithm of the inverse of the probability that type $Q_{XY}$ is the empirical distribution of the code pair $x^n(1), y^n(1)$ individually generated from fixed-compositions $P_X$ and $P_Y$. $D(Q_{Z|XY}\|W|Q_{XY})$ is logarithm of the inverse of the conditional probability that the input to the channel $W$ is $Q_{XY}$, while the empirical type of the input/output is $Q_{XYZ} = Q_{XY} \times Q_{Z|XY}$. For the individual part of the error exponents in (12), (13) and (14): $|I_Q(X,Y;Z) - R_x - R_y|^+$, $|I_Q(X;Z|Y) - R_x|^+$ and $|I_Q(X;Z) - R_x|^+$ respectively, each one is the logarithm of the inverse of an upper bound on the probability that there exists another message (pair) with higher mutual information with the channel output, while the channel inputs/ouput has type $Q_{XYZ}$. This is derived by a union bound argument.

First we write the error probability (12) in the following way:

$$\begin{aligned}
&\Pr(m_x \neq \widehat{m}_x, m_y \neq \widehat{m}_y)\\
&= \left(\frac{1}{|\mathcal{T}^n(P_X)|}\right)^{2^{nR_x}} \left(\frac{1}{|\mathcal{T}^n(P_Y)|}\right)^{2^{nR_y}} \sum_{c_X}\sum_{c_Y} \qquad (26)\\
&\qquad \frac{1}{2^{nR_x}}\sum_{m_x}\frac{1}{2^{nR_y}}\sum_{m_y}\sum_{z^n} W_{Z|XY}(z^n|x^n(m_x), y^n(m_y))1(\widehat{m}_x(z^n)\neq m_x, \widehat{m}_y(z^n)\neq m_y)\\
&= \left(\frac{1}{|\mathcal{T}^n(P_X)|}\right)^{2^{nR_x}} \left(\frac{1}{|\mathcal{T}^n(P_Y)|}\right)^{2^{nR_y}} \sum_{c_X}\sum_{c_Y}\sum_{z^n} W_{Z|XY}(z^n|x^n(1), y^n(1))1(\widehat{m}_x(z^n)\neq 1, \widehat{m}_y(z^n)\neq 1)\\
&= \sum_{Q_{XY}:Q_X=P_X, Q_Y=P_Y} \Big\{ \Pr\big((x^n(1), y^n(1)) \in Q_{XY}\big) \sum_{Q_{Z|XY}} \Pr(z^n|(x^n(1), y^n(1)) \in Q_{Z|XY})\\
&\qquad \Pr(\widehat{m}_x(z^n)\neq 1, \widehat{m}_y(z^n)\neq 1)\Big\} \qquad (27)\\
&\leq \sum_{Q_{XY}:Q_X=P_X, Q_Y=P_Y} \Big\{ \Pr\big((x^n(1), y^n(1)) \in Q_{XY}\big) \sum_{Q_{Z|XY}} \Pr(z^n|(x^n(1), y^n(1)) \in Q_{Z|XY})\\
&\qquad \min\{1, \sum_{i=2}^{2^{nR_x}}\sum_{j=2}^{2^{nR_y}} \Pr\left(I(z^n; x^n(1), y^n(1)) \leq I(z^n; x^n(i), y^n(j)) | (x^n(1), y^n(1), z^n) \in Q_{XYZ}\right)\}\Big\}\\
&\leq |\mathcal{T}^n_{XYZ}| \max_{Q_{XYZ}:Q_X=P_X, Q_Y=P_Y} \Pr\big((x^n(1), y^n(1)) \in Q_{XY}\big) \Pr(z^n|(x^n(1), y^n(1)) \in Q_{Z|XY})\\
&\qquad \min\{1, \sum_{i=2}^{2^{nR_x}}\sum_{j=2}^{2^{nR_y}} \Pr\left(I(z^n; x^n(1), y^n(1)) \leq I(z^n; x^n(i), y^n(j)) | (x^n(1), y^n(1), z^n) \in Q_{XYZ}\right)\} \qquad (28)
\end{aligned}$$

(26) and (27) are two different interpretations of the same error probability. In (26), we first randomly pick a fixed composition codebook pair $c_X$ and $c_Y$, then sum over the all probabilities that the output of the channel causes a decoding error for the chosen codebook pair. (27) is an equivalent interpretation of the above error probability because the codewords for each message is independently generated. We interpret (27) as follows, we first randomly pick a codeword pair for message 1 in $X$ and message 1 in $Y$, then the codeword pair is transmitted to through the channel. Then we randomly generate the rest of the codebook and investigate the probability that other message pairs maximize the mutual information with the channel output.

We upper bound the four components in (28) as follows. The number of type sets of length $n$:

$$|\mathcal{T}^n_{XYZ}| \leq (n+1)^{|\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}|} = 2^{n(\frac{\log(n+1)}{n}|\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}|)} = 2^{na_n}. \qquad (29)$$

For any $Q_{XY}$, s.t. $Q_X = P_X$ and $Q_Y = P_Y$, from the method of types [1] and [2], we know that $2^{n(H(P_Y) - \frac{\log n}{n}|\mathcal{Y}|)} \leq |\mathcal{P}_Y| \leq 2^{nH(P_Y)}$, similar bounds applies to $|\mathcal{P}_X|$. And for a fixed $X$-sequence, $x^n(1) \in P_X = Q_X$, we have $2^{n(H(Q_{Y|X}) - \frac{\log n}{n}|\mathcal{X}\mathcal{Y}|)} \leq |\{y^n \in \mathcal{Y}^n : (x^n(1), y^n) \in Q_{XY}\}| \leq 2^{nH(Q_{Y|X})}$. $x^n(1)$ and $y^n(1)$ are independently distributed in type set $P_X$ and $P_Y$. Hence,

$$\Pr\big((x^n(1), y^n(1)) \in Q_{XY}\big) = \frac{|\{y^n \in \mathcal{Y}^n : (x^n(1), y^n) \in Q_{XY}\}|}{|P_Y|} \leq 2^{n(H(Q_{Y|X}) - H(Q_Y) + \frac{\log n}{n}|\mathcal{X}|)}$$

Notice that $H(Q_{Y|X}) - H(Q_Y) = -D(Q_{XY}\|Q_X \times Q_Y) = -D(Q_{XY}\|P_X \times P_Y)$ and let $b_n = \frac{\log n}{n}|\mathcal{X}|$, we have:

$$\Pr\left((x^n(1), y^n(1)) \in Q_{XY}\right) \leq 2^{-n(D(Q_{XY}\|P_X \times P_Y) - b_n)} \tag{30}$$

For $(x^n(1), y^n(1)) \in Q_{XY}$, for any empirical channel behavior $Q_{Z|XY}$:

$$\begin{aligned}
\Pr(z^n|(x^n(1), y^n(1)) \in Q_{Z|XY}) &= |\{z^n : (x^n(1), y^n(1), z^n) \in Q_{XYZ}\}| W_{Z|XY}(Q_{Z|XY}) \\
&\leq 2^{nH(Q_{Z|XY})} \times 2^{n(-D(Q_{Z|XY}\|W|Q_{XY}) - H(Q_{Z|XY}))} \\
&= 2^{-nD(Q_{Z|XY}\|W|Q_{XY})}
\end{aligned} \tag{31}$$

Finally, for $(x^n(1), y^n(1), z^n) \in Q_{XYZ}$, we investigate the probability that there exists $(i,j)$, $i \neq 1, j \neq 1$, s.t. the mutual information between $(x^n(i), y^n(j))$ and $z^n$ is at least as much as the mutual information between $(x^n(1), y^n(1))$ and $z^n$. For all $i \neq 1$, the codeword $x^n(i)$ is uniformly distributed on the fixed-composition set $P_X$, same for $Y$. Given $(x^n(1), y^n(1), z^n) \in Q_{XYZ}$, we have $I(z^n; x^n(1), y^n(1)) = I_Q(Z; X, Y)$, so:

$$\begin{aligned}
&\min\{1, \sum_{i=2}^{2^{nR_x}} \sum_{j=2}^{2^{nR_y}} \Pr\left(I(z^n; x^n(1), y^n(1)) \leq I(z^n; x^n(i), y^n(j)) | (x^n(1), y^n(1), z^n) \in Q_{XYZ}\right)\} \\
&\leq \min\{1, 2^{n(R_x+R_y)} \sum_{V_{XYZ}: V_X = Q_X, V_Y = Q_Y, V_Z = Q_Z, I_Q(Z;X,Y) \leq I_V(Z;X,Y)} \Pr(((x^n(i), y^n(j), z^n) \in V_{XYZ} | z^n \in Q_Z)\} \\
&= \min\{1, 2^{n(R_x+R_y)} \sum_{V_{XYZ}: V_X = Q_X, V_Y = Q_Y, V_Z = Q_Z, I_Q(Z;X,Y) \leq I_V(Z;X,Y)} \frac{|\{(x^n, y^n) \in P_X \times P_Y : (x^n, y^n, z^n) \in V_{XYZ}\}|}{|\{x^n : x^n \in P_X\}||\{y^n : y^n \in P_Y\}|}\} \\
&\leq \min\{1, 2^{n(R_x+R_y)} \sum_{V_{XYZ}: V_X = Q_X, V_Y = Q_Y, V_Z = Q_Z, I_Q(Z;X,Y) \leq I_V(Z;X,Y)} 2^{n(H_V(X,Y|Z) - H_V(X) - H_V(Y) + \frac{\log n(|\mathcal{X}|+|\mathcal{Y}|)}{n})}\} \\
&\leq \min\{1, 2^{n(R_x+R_y)} \sum_{V_{XYZ}: V_X = Q_X, V_Y = Q_Y, V_Z = Q_Z, I_Q(Z;X,Y) \leq I_V(Z;X,Y)} 2^{n(H_V(X,Y|Z) - H_V(X,Y) + \frac{\log n(|\mathcal{X}|+|\mathcal{Y}|)}{n})}\} \\
&= \min\{1, 2^{n(R_x+R_y)} \sum_{V_{XYZ}: V_X = Q_X, V_Y = Q_Y, V_Z = Q_Z, I_Q(Z;X,Y) \leq I_V(Z;X,Y)} 2^{n(-I_V(X,Y;Z) + \frac{\log n(|\mathcal{X}|+|\mathcal{Y}|)}{n})}\} \\
&\leq \min\{1, 2^{n(R_x+R_y)} n^{|\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}|} 2^{n(-I_Q(X,Y;Z) + \frac{\log n(|\mathcal{X}|+|\mathcal{Y}|)}{n})}\} \\
&= 2^{-n(|I_Q(X,Y;Z) - R_x - R_y|^+ - c_n)}
\end{aligned} \tag{32}$$

Combining (29), (30), (31) and (32), and noticing that $a_n$ $b_n$ and $c_n$ converges to zero when $n$ goes to infinity, we have just proved (12).
(13) and (14) can be proved by following the same argument. Similar to the way we upper bound the LHS of (12) in (28), we bound the LHS of (13) as follows:

$$\begin{aligned}
&\Pr(m_x \neq \widehat{m}_x | m_y = \widehat{m}_y) \\
&\leq |\mathcal{T}^n_{XYZ}| \max_{Q_{XYZ}: Q_X = P_X, Q_Y = P_Y} \Pr\left((x^n(1), y^n(1)) \in Q_{XY}\right) \Pr(z^n|(x^n(1), y^n(1)) \in Q_{Z|XY}) \\
&\quad \min\{1, \sum_{i=2}^{2^{nR_x}} \Pr\left(I(z^n; x^n(1), y^n(1)) \leq I(z^n; x^n(i), y^n(1)) | (x^n(1), y^n(1), z^n) \in Q_{XYZ})\right)\}.
\end{aligned} \tag{33}$$

Similarly, we upper bound the LHS of (14) by

$$\begin{aligned}
&\Pr(m_x \neq \widehat{m}_x) \\
&\leq |\mathcal{T}^n_{XYZ}| \max_{Q_{XYZ}: Q_X = P_X, Q_Y = P_Y} \Pr\left((x^n(1), y^n(1)) \in Q_{XY}\right) \Pr(z^n|(x^n(1), y^n(1)) \in Q_{Z|XY}) \\
&\quad \min\{1, \sum_{i=2}^{2^{nR_x}} \Pr\left(I(z^n; x^n(1)) \leq I(z^n; x^n(i)) | (x^n(1), y^n(1), z^n) \in Q_{XYZ})\right)\}.
\end{aligned} \tag{34}$$

The common parts (first line) in (28), (33) and (34) are proved in (29) (30) and (31). The individual part of (28) is proved in (31). The proof for the individual part of (33) and (34) follow similar argument. We omit the details here.

### B. *Proof of (15) and (20)*

To prove (15), we give an upper bound of the *correct probability* $\Pr(\widehat{m}_x = m_x)$.

$$\begin{aligned}\Pr(\widehat{m}_x = m_x) &= P^n_{e(x)}(R_x, R_y, P_X, P_Y)\\ &= \left(\frac{1}{|\mathcal{T}^n(P_X)|}\right)^{2^{nR_x}} \sum_{c_X} \frac{1}{2^{nR_x}} \sum_{m_x}\sum_{z^n} W_{Z|XY}(z^n|x^n(m_x), y^n) 1(\widehat{m}_x(z^n) = m_x)\end{aligned}$$

The codewords $x^n(m_x)$ is uniformly distributed on the type set $P_X$, so the probability that the joint type of $(x^n(m_x), y^n)$ is close to $P_X \times P_Y$ with high probability [1], i.e. for all $\sigma > 0$, for large $n$,

$$\Pr(D((x^n(m_x), y^n)\|P_X \times P_Y)) > \sigma) < \sigma. \tag{35}$$

We denote by $T_\sigma(y^n) = \{x^n : D((x^n, y^n)\|P_X \times P_Y)) \leq \sigma\}$, the typical set given $y^n$. Now we look at individual codebooks $c_X$, we say a codebook $c_X$ is good if

$$|c_X \bigcap T^C_\sigma(y^n)| \leq \frac{|c_X|}{4} \tag{36}$$

where $|c_X| = 2^{nR_x}$The set of all good codebooks is denoted by $G$, at most $4\sigma$ of the codebooks are not in $G$ because of (35). For a good codebook $c_X$, we use the technique from [4] to upper bound the correct probability for the good codebook $c_X$.

$$\begin{aligned}\Pr(\widehat{m}_x = m_x) &\leq \frac{|c_X \bigcap T^C_\sigma(y^n)|}{|c_X|} + \frac{1}{|c_X|} \sum_{i:x^n(i)\in T_\sigma(y^n)} \Pr(i = \widehat{m}_x(z^n))\\ &\leq \frac{1}{4} + \frac{1}{|c_X|} \sum_{i:x^n(i)\in T_\sigma(y^n)} \Pr(i = \widehat{m}_x(z^n))\\ &\leq \frac{1}{4} + 2^{-n(E-\epsilon_n)}\end{aligned}$$

The last inequality is proved by Lemma 6 and 7 which are extensions of Lemma 3 and Lemma 5 in [4] from memoryless to conditional on $y^n$, where $\epsilon_n$ goes to zero with $n$, and

$$E = \min_{Q_{XYZ}:D(Q_{XY}\|P_X\times P_Y)<\sigma} D(Q_{Z|XY}\|W_{Z|XY}|Q_{XY}) + |R_x - I_Q(X;Z|Y)|^+$$

Following the argument in Lemma 3, it is easy to see that $E > 0$ for $R_x > I(X;Z|Y)$ and small $\sigma$, where $(X, Y, Z) \sim W_{Z|XY} \times P_X \times P_Y$. Now we have

$$\begin{aligned}\Pr(\widehat{m}_x = m_x) &= \left(\frac{1}{|\mathcal{T}^n(P_X)|}\right)^{2^{nR_x}} (\sum_{c_X\in G} \Pr(\widehat{m}_x = m_x) + \sum_{c_X\in G^C} \Pr(\widehat{m}_x = m_x))\\ &\leq \frac{1}{4} + 2^{-n(E-\epsilon_n)} + 4\sigma\end{aligned} \tag{37}$$

Let $\sigma$ be small enough and let $n$ goes to infinity, (15) is proved. □

(20) is proved in the same way. We only need to introduce the notion of good codebook pair, $(c_X, c_Y)$ is good if

$$|c_X \times c_Y \bigcap T^C_\sigma| \leq \frac{|c_X||c_Y|}{4} \tag{38}$$

where the typical set $T_\sigma = \{(x^n, y^n) : D((x^n, y^n)\|P_X \times P_Y) < \sigma\}$. The rest of the proof are similar to that in the proof for (15). We conclude that

$$\Pr(\widehat{m}_x = m_x) \leq \frac{1}{4} + 2^{-n(E-\epsilon_n)} + 4\sigma \tag{39}$$

where $E = \min_{Q_{XYZ}:D(Q_{XY}\|P_X\times P_Y)<\sigma} D(Q_{Z|XY}\|W_{Z|XY}|Q_{XY}) + |R - I_Q(X, Y; Z)|^+ > 0$, for $R_x + R_y > I(X, Y; Z)$.

Again, we need to use an modified version of Lemma 3 and 5 from [4], the proof are extremely similar to those in Lemma 6 and 7. We omit the details here. This concludes the proof. □

The following two lemmas are simple extensions of Lemma 3 and 5 in [4]. Instead studying the upper bound on the probability of correct decoding for memoryless channels, we give an upper bound on the correct probability for a multiple access channel with one encoder with fixed output and the joint composition of the two encoders is fixed. That is, we fix $y^n$ and this $y^n$ is known to both encoders and the decoder. The multiple access channel is $W_{Z|XY}$ and the input $x^n$ from $X$ is such that $(x^n(i), y^n) \in Q_{XY}$, $i = 1, 2, ...2^{nR}$.

*Lemma 6:* Extension of Lemma 3 in [4] from memoryless to condition on $y^n$, for any $R \geq R_x > 0$, for any coding system $X(y^n)$ with joint input distribution $(x^n(i), y^n) \in Q_{XY}$, $i = 1, 2, ...2^{nR_x}$, and decoding rule $\phi : \mathcal{Z}^n \to \{1, 2, ..., 2^{nR_x}\}$, let $Q_{Z|XY}(x^n(i), y^n) = \{z^n : (x^n(i), y^n, z^n) \in Q_{XYZ}\}$ (this is the V-shell notation $T_V$ used in [4]), we have:

$$\frac{1}{2^{nR}} \sum_{i=1}^{2^{nR_x}} \frac{|Q_{Z|XY}(x^n(i), y^n) \bigcap \phi^{-1}(i)|}{|Q_{Z|XY}(x^n(i), y^n)|} \leq 2^{-n|R - I_Q(X;Z|Y) - \epsilon_n|^+} \tag{40}$$

where $\epsilon_n = \epsilon(n, |\mathcal{X}|, |\mathcal{Y}|, |\mathcal{Z}|)$ goes to zero as $n$ goes to infinity.

*Proof:* Write $Q_{Z|Y}(y^n) = \{z^n : (y^n, z^n) \in Q_{ZY}\}$. By the method of types [2], we know that

$$(n+1)^{-|\mathcal{Z}|} 2^{nH_Q(Z|XY)} \leq |Q_{Z|XY}(x^n(i), y^n)| \leq 2^{nH_Q(Z|XY)}$$

$$\text{and } (n+1)^{-|\mathcal{Z}|} 2^{nH_Q(Z|Y)} \leq |Q_{Z|Y}(y^n)| \leq 2^{nH_Q(Z|Y)}.$$

So the LHS of (40) is upper bounded by

$$\begin{aligned}
\frac{1}{2^{nR}} \sum_{i=1}^{2^{nR_x}} \frac{|Q_{Z|XY}(x^n(i), y^n) \bigcap \phi^{-1}(i)|}{|Q_{Z|XY}(x^n(i), y^n)|} &\leq (n+1)^{|\mathcal{Z}|} 2^{-nH_Q(Z|XY)} 2^{-nR} \sum_{i=1}^{2^{nR_x}} |Q_{Z|XY}(x^n(i), y^n) \bigcap \phi^{-1}(i)| \\
&\leq (n+1)^{|\mathcal{Z}|} 2^{-nH_Q(Z|XY)} 2^{-nR} |Q_{Z|Y}(y^n)| \qquad (41) \\
&\leq (n+1)^{|\mathcal{Z}|} 2^{-nH_Q(Z|XY)} 2^{-nR} (n+1)^{|\mathcal{Z}|} 2^{nH_Q(Z|Y)} \\
&= 2^{-n(R - I_Q(X;Z|Y) - \epsilon_n)} \qquad (42)
\end{aligned}$$

(41) is true because $Q_{Z|XY}(x^n(i), y^n) \bigcap \phi^{-1}(i)$, $i = 1, 2, ..., 2^{nR_x}$ are disjoint and $\bigcup_i Q_{Z|XY}(x^n(i), y^n) \subseteq Q_{Z|Y}(y^n)$. Now notice that the LHS of (40) is at most $2^{n(R_x - R)} \leq 1$, hence the LHS of (40) is no bigger than 1. This together with (42), we just proved Lemma 6. $\square$

Now we are ready to prove Lemma 7.

*Lemma 7:* Extension of Lemma 5 in [4] from memoryless to condition on $y^n$, for a good codebook $c_X \in G$ defined in (36). Recall that $|c_X \bigcap T_\sigma(y^n)| \geq \frac{3|c_X|}{4} = \frac{3}{4} \times 2^{nR_x}$, then for any decoding rule (previously known as $\widehat{m}_x$) $\phi : \mathcal{Z}^n \to \{1, 2, ..., 2^{nR_x}\}$,

$$\frac{1}{|c_X|} \sum_{i : x^n(i) \in T_\sigma(y^n)} \Pr(i = \phi(z^n)) \leq 2^{-n(E - \epsilon_n)} \tag{43}$$

$$\text{where } E = \min_{Q_{XYZ} : D(Q_{XY} \| P_X \times P_Y) < \sigma} D(Q_{Z|XY} \| W_{Z|XY} | Q_{XY}) + |R_x - I_Q(X;Z|Y)|^+$$

and $\epsilon_n = \epsilon(|\mathcal{X}|, |\mathcal{Y}|, |\mathcal{Z}|, n)$ which converges to zero as $n$ goes to infinity.

*Proof:* We write $M = \{i \in \{1, 2, ..., 2^{nR_x}\} : x^n(i) \in T_\sigma(y^n)\}$ then we know that from the definition of a good codebook: $\frac{3}{4} \times 2^{nR_x} \leq |M| \leq 2^{nR_x} = |c_X|$. Notice that

$$\Pr(i = \phi(z^n)) = \sum_{z^n \in \phi^{-1}(i)} W_{Z|XY}(z^n | x^n(i), y^n) = W_{Z|XY}(\phi^{-1}(i) | x^n(i), y^n) \tag{44}$$

We rewrite the LHS of (43):

$$\begin{aligned}
&= 2^{-nR_x} \sum_{i : x^n(i) \in T_\sigma(y^n)} W_{Z|XY}(\phi^{-1}(i) | x^n(i), y^n) \\
&= 2^{-nR_x} \sum_{Q_{XY} : D(Q_{XY} \| P_X \times P_Y) < \sigma} \left( \sum_{i : (x^n(i), y^n) \in Q_{XY}} W_{Z|XY}(\phi^{-1}(i) | x^n(i), y^n) \right) \\
&\leq (n+1)^{|\mathcal{X}||\mathcal{Y}|} \max_{Q_{XY} : D(Q_{XY} \| P_X \times P_Y) < \sigma} \left( 2^{-nR_x} \sum_{i : (x^n(i), y^n) \in Q_{XY}} W_{Z|XY}(\phi^{-1}(i) | x^n(i), y^n) \right) \\
&= (n+1)^{|\mathcal{X}||\mathcal{Y}|} \max_{Q_{XY} : D(Q_{XY} \| P_X \times P_Y) < \sigma} \left( 2^{-nR_x} \sum_{i : (x^n(i), y^n) \in Q_{XY}} \sum_{Q_{Z|XY}} W_{Z|XY}(\phi^{-1}(i) \bigcap Q_{Z|XY}(x^n(i), y^n) | x^n(i), y^n) \right) \\
&\leq (n+1)^{|\mathcal{X}||\mathcal{Y}| + |\mathcal{X}||\mathcal{Y}||\mathcal{Z}|} \max_{Q_{XYZ} : D(Q_{XY} \| P_X \times P_Y) < \sigma} \left( 2^{-nR_x} \sum_{i : (x^n(i), y^n) \in Q_{XY}} W_{Z|XY}(\phi^{-1}(i) \bigcap Q_{Z|XY}(x^n(i), y^n) | x^n(i), y^n) \right)
\end{aligned}$$

$$
\begin{aligned}
&\le 2^{n\epsilon_n(1)} \max_{Q_{XYZ}:D(Q_{XY}\|P_X\times P_Y)<\sigma} \big(2^{-nR_x} \sum_{i:(x^n(i),y^n)\in Q_{XY}} W_{Z|XY}(Q_{Z|XY}(x^n(i),y^n)|x^n(i),y^n) \\
&\qquad \frac{|Q_{Z|XY}(x^n(i),y^n)\bigcap \phi^{-1}(i)|}{|Q_{Z|XY}(x^n(i),y^n)|}\big) \\
&\le 2^{n\epsilon_n(1)} \max_{Q_{XYZ}:D(Q_{XY}\|P_X\times P_Y)<\sigma} \big(2^{-nD(Q_{Z|XY}\|W_{Z|XY}|Q_{XY})}2^{-nR_x} \sum_{i:(x^n(i),y^n)\in Q_{XY}} \frac{|Q_{Z|XY}(x^n(i),y^n)\bigcap \phi^{-1}(i)|}{|Q_{Z|XY}(x^n(i),y^n)|}\big) \\
&\le 2^{n\epsilon_n(1)} \max_{Q_{XYZ}:D(Q_{XY}\|P_X\times P_Y)<\sigma} \big(2^{-nD(Q_{Z|XY}\|W_{Z|XY}|Q_{XY})}2^{-n|R-I_Q(X;Z|Y)-\epsilon_n(2)|^+}\big) \qquad (45) \\
&= 2^{-n(E-\epsilon_n)} \qquad (46)
\end{aligned}
$$

where (45) is true by Lemma 6. The rest are obvious by the method of types. □

The proof here for Lemma 6 and 7 are almost identical to that for Lemma 3 and 5 in [4]. The difference is here we need to deal with the other channel inputs $y^n$ and the new challenges coming from the typicalities of the codebook.