Multiuser Channels with Statistical CSI at the Transmitter: Fading Channel Alignments and Stochastic Orders, an Overview

Pin-Hsun Lin * and **Eduard A. Jorswieck** [ID]

Department of Electrical Engineering and Information Technology, Technische Universität Dresden,
Zellescher Weg 18, 01069 Dresden, Germany; eduard.jorswieck@tu-dresden.de
* Correspondence: pin-hsun.lin@tu-dresden.de; Tel.: +49-351-463-33127

Abstract: In this overview paper, we introduce an application of stochastic orders in wireless communications. In particular, we show how to use stochastic orders to investigate the ergodic capacity results for fast fading Gaussian memoryless multiuser channels when only the statistics of the channel state information are known at the transmitters (CSIT). In general, the characterization of the capacity region of multiuser channels with only statistical CSIT is open. To attain our goal, in this work we resort to classifying the random channels through their probability distributions by which we can obtain the capacity results. To be more precise, we derive sufficient conditions to attain some information theoretic channel orders such as degraded and very strong interference by exploiting the usual stochastic order and exploiting the same marginal property. After that, we apply the developed scheme to Gaussian interference channels and Gaussian broadcast channels. We also extend the framework to channels with multiple antennas. Possible scenarios for channel enhancement under statistical CSIT are also discussed. Several practical examples such as Rayleigh fading and Nakagami-m fading, etc., illustrate the application of the derived results.

Keywords: multi-user channels; fast fading; stochastic orders; same marginal property; capacity region

1. Introduction

Ordering plays a fundamental role in mathematics, science, engineering, finance, etc. The most well known and commonly used one is the trichotomy order, in which we compare real values. In the applications of wireless communications, when there is perfect channel state information at the transmitter (CSIT), it is known that for some Gaussian multiuser (MU) channels the capacity results can be attained due to the capability of ordering the quality of channels among different users. For example, capacity regions/secrecy capacity of degraded broadcast channels (BC) and wiretap channels (WTC) (even for multiple-antenna cases) [1], and also the sum capacity of low-interference regime [2] and capacity regions for some cases of IC such as strong IC [3–5] and very strong IC [4] are derived. When fading effects of wireless channels are taken into account, if there is perfect CSIT, some of the above capacity results still hold with an additional operation of taking an average over the capacity (region) with respect to fading channels. For example, in [6], the ergodic secrecy capacity of Gaussian WTC is derived; in [7], the ergodic capacity regions are derived for ergodic very strong and uniformly strong (each realization of the fading process is a strong interference channel) of the Gaussian IC. Notably, the orders in the above scenarios are all trichotomy orders.

Due to several practical limitations, for example, the finite bandwidth of feedback channels, a delay caused by channel estimation, etc., the transmitter may not be able to track channel realizations precisely instantaneously if they vary rapidly. Thus, for fast fading channels, it is more practical to consider the case with only partial CSIT of the legitimate channel, where statistical CSIT is one of

the most commonly considered one [8,9]. However, when there is only statistical CSIT (In this paper, statistical CSIT and no CSIT will be used interchangeably.), there are only few known capacity results, such as the layered BC [10], the binary fading interference channel [11], the one-sided layered IC [12], Gaussian WTC [13,14], and layered WTC [14], etc. This is because, when the transmitter has imperfect CSIT, e.g., only the statistics of the channels, the comparison of channels cannot be attained directly by trichotomy law due to the random characteristic of the fading channels. Then the optimal strategy for the transmission, including the design of the codebook, channel input distribution, resource allocation, etc., can not be easily done. In particular, those channel orders commonly used in information theory include degraded, less noisy, and more capable [9] in BC and WTC or the strong and very strong in IC [9] depends on the knowledge of CSIT. Note that by these information theoretic orders, we can highly simplify the functional optimization with respect to the channel input distribution and/or channel prefixing.

To consider an MU-channel in which the transmitters only know the distributions of the channels but not the realizations, we may ask the following questions: is it possible to compare the channel qualities just according to their distributions? If yes, how to do it? How to derive the capacity region under such comparison of channel qualities? In this work we resort to partly solve these problems by classifying the random channels into stochastically orderable and its complement. In particular, an MU channel with orderable random channels means that there exists an equivalent channel in which we can reorder the channel realizations among different transmitter-receiver pairs in the desired manner. More specifically, we resort to finding a subset of all fading channel tuples, namely, $\mathcal{A}$, which should possess the following properties:

- It allows the existence of a corresponding set $\mathcal{B}$ in which the channels follow a certain order, e.g., trichotomy order.
- It encompasses a constructive way (or easy) to find a transformation $f : a \mapsto b$, $a \in \mathcal{A}$, $b \in \mathcal{B}$.
- Capacity results are attainable.

Taking the BC as an example, an orderable two-user BC means that under the same noise distributions at the two receivers, in the equivalent BC, one channel strength is always stronger or weaker than the other for all fading states. The main tool we use for this channel classification and ordering is stochastic orders [15] from probability theory, combined with the same marginal property [16] from information theory. The stochastic orders have been widely used in the last several decades in diverse areas of probability and statistics such as reliability theory, queueing theory, and operations research, etc., see [15] and references therein. Different stochastic orders such as the usual stochastic order, the convex order, and the increasing convex order can help us to identify the location, dispersion, or both location and dispersion of random variables, respectively. Smartly choosing a proper stochastic order to compare the channels with statistical CSIT allows us to form an equivalent channel in which realizations of channel gains are ordered in a desired manner. Then we are able to derive the capacity regions of the equivalent MU channel, which is simpler than directly considering the original channel. Note that in [17] the authors also consider stochastic orders for fading channels. However, there is no such alignment concept by constructing an equivalent channel in [17] and hence the relation between the same marginal property and stochastic orders is completely not investigated there. In contrast, they discuss the stochastic dominance between fading channels by Shannon transform order. Stochastic orders are also used in stochastic geometry to analyze the performance of a random network [18]. Similarly, the inter-relation between the information theoretic channel orders and probabilistic orders are not addressed in [18].

The main issues discussed by this overview paper are summarized as follows.

- We classify fading MU channels such that we can characterize the capacity results of some memoryless Gaussian MU channels under statistical CSIT. To achieve it, we combine the concept of usual stochastic order and also the same marginal property, i.e., an intrinsic property of some MU channels. Intuitively, by doing so we can align the realizations of the fading channel gains between different users in an identical trichotomy order over time in an equivalent channel.

- We then apply the proposed scheme to characterize the capacity regions of Gaussian IC and BC, which is novel in the literature.
- We further extend the framework to channels with multiple antennas. Applicable scenarios for channel enhancement scheme under statistical CSIT, which is originally for channels with perfect CSIT, are also discussed.
- Several examples with practical channel distributions are illustrated to show the usage scenarios of the developed framework.

Notation: Upper case normal/bold letters denote random variables/random vectors (or matrices), which will be defined when they are first mentioned; lower case bold letters denote vectors. The statistical expectation is denoted by $\mathbb{E}[.]$. The mutual information between two random variables X and Y is denoted by $I(X; Y)$. The complementary cumulative density function (CCDF) is denoted by $\bar{F}_X(x) = 1 - F_X(x)$, where $F_X(x)$ is the CDF of X. In addition, we denote the probability mass function (PMF) by p and the probability density function (PDF) by f. $X \sim F$ denotes that the random variable X follows the distribution F. Markov chain relation between X, Y, and Z is described by $X - Y - Z$. $\text{Unif}(a, b)$ denotes the uniform distribution between a and b. $\mathbb{Z}^+ = \{0, \mathbb{N}\}$. The indicator function is denoted by $\mathbb{1}_{\{.\}}$. The supports of a function f and a random variable X are respectively denoted by $\text{supp}(f)$ and $\text{supp}(X)$. The logarithms used in the paper are all of base 2. We denote $C(P) = \log(1 + P)$. We denote the equality in distribution by $=_d$.

The remainder of the paper is organized as follows. In Section 2, we introduce the background knowledge and preliminaries. In Section 3, we formulate a problem and propose a framework to solve it. In Section 4 we apply the tools developed in Section 3 to fast fading Gaussian interference channels and broadcast channels with statistical CSIT. In Section 5 we consider channels with multiple antennas. In Section 6, an application of Laplace transform order on solving a power allocation problem is reviewed. Finally, Section 7 concludes the paper.

2. Preliminaries

In this section, some important properties and definitions for deriving the main results of this work will be introduced, including the same marginal properties, degradedness, and the usual stochastic orders.

2.1. Same Marginal Property

The same marginal property plays a crucial role in the proposed channel classification to obtain the capacity results. This is because it provides us the degree of freedom to construct an equivalent channel in the sense that the marginal distributions are the same as the original one, but not the joint distribution. By such a relaxation of considering an equivalent channel, we are able to reorder all channel gains under some conditions.

Two versions of the same marginal property are introduced as follows:

Theorem 1 (**Same Marginal Property for One-Transmitter (Theorem 13.9 in [19])**). *The capacity region of a multiuser channel with one transmitter and two non-cooperative receivers depends only on the conditional marginal distributions $P_{Y_1|X}$ and $P_{Y_2|X}$ and not on the jointly conditional distribution $P_{Y_1,Y_2|X}$.*

Theorem 2 (**Same Marginal Property for Two-Transmitter (Theorem 16.6 in [19])**). *The capacity region of a multiuser channel with two-transmitter and two non-cooperative receivers depends only on the conditional marginal distributions $P_{Y_1|X_1,X_2}$ and $P_{Y_2|X_1,X_2}$ and not on the jointly conditional distribution $P_{Y_1,Y_2|X_1,X_2}$.*

Remark 1. By the union bound, the error probability of a channel with multiple-receiver can be upper bounded by the sum of individual error probability. Therefore, the overall error probability approaches zero if the individual error probabilities approach zero, respectively. This fact results in the consequence that only the marginal transition probabilities affect the capacity result, but not the joint one.

Remark 2. Note that since the capacity region of a multiple access channel is determined by $p_{Y|X_1, X_2}$, the technique of reordering random fading channels developed in this paper can be useful to simplify the proof for the GMAC with statistical CSIT.

For channels with a single transmitter and multiple receivers, e.g., BC or WTC, we can get Theorem 1 from Theorem 2 by removing X_1 or X_2.

2.2. Information Theoretical Orders for Memoryless Channels and Stochastic Orders

The main task in this paper is on ordering channels. Here we introduce several important definitions describing the relation of reception qualities among different receivers from an information theoretic to the probabilistic point of view.

Definition 1. *A channel with two non-cooperative receivers and one transmitter is physically degraded if the transition distribution satisfies* $P_{Y_1 Y_2|X}(\cdot|\cdot) = P_{Y_1|X}(\cdot|\cdot)P_{Y_2|Y_1}(\cdot|\cdot)$, *i.e., X, Y_1, and Y_2 form a Markov chain $X - Y_1 - Y_2$. The channel is stochastically degraded if its conditional marginal distribution is the same as that of a physically degraded channel, i.e.,*

$$\text{there exists a } \tilde{P}_{Y_2|Y_1}(\cdot|\cdot), \text{ such that } P_{Y_2|X}(y_2|x) = \sum_{y_1} P_{Y_1|X}(y_1|x)\tilde{P}_{Y_2|Y_1}(y_2|y_1). \tag{1}$$

Denote the fading channel gains in AWGN channels from the transmitter to the first and second receivers by H_1 and H_2, respectively. Define a set of tuples of random channels $\mathcal{H}_{10} = \{(H_1, H_2)\}$ and also a set $\mathcal{H}_1 = \{(H_1, H_2) : \mathbb{1}_{\{(1)\}} = 1\}$.

Recall that $\mathbb{1}_{\{(1)\}} = 1$ means (1) is true. In the following, we call a stochastically degraded channel simply a degraded channel due to the same marginal property. Note that discussions on the relation between degradedness and other information theoretic channel orders can be referred to [19–21].

Definition 2. *A discrete memoryless-IC is said to have very strong interference if*

$$I(X_1; Y_1|X_2, H_{11}, H_{12}) \le I(X_1; Y_2|H_{21}, H_{22}), \tag{2}$$

$$I(X_2; Y_2|X_1, H_{21}, H_{22}) \le I(X_2; Y_1|H_{11}, H_{12}), \tag{3}$$

for all $f_{X_1} \cdot f_{X_2}$. Define the set $\mathcal{H}_3 = \{(H_{11}, H_{12}, H_{21}, H_{22}) : \mathbb{1}_{\{(2)\}} \cap \mathbb{1}_{\{(3)\}} = 1\}$.

After information theoretic orders, we introduce some important definitions of stochastic orders, which are the underlying tools in this paper.

Definition 3 ([15]). *For given random variables X and Y, the usual stochastic order (st), the convex order (cx), the concave order (cv), the increasing convex order (icx), the increasing concave order (icv), and the Laplace transform order (Lt) are respectively defined as follows:*

$$X \le_{st} Y, \text{ if } \mathbb{E}[f(X)] \le \mathbb{E}[f(Y)] \text{ for all increasing } f,$$

$$X \le_{cx\ (cv)} Y, \text{ if } \mathbb{E}[f(X)] \le \mathbb{E}[f(Y)] \text{ for all convex (concave) } f,$$

$$X \le_{icx\ (icv)} Y, \text{ if } \mathbb{E}[f(X)] \le \mathbb{E}[f(Y)] \text{ for all increasing convex (concave) } f,$$

$$X \le_{Lt} Y, \text{ if } \mathbb{E}[\exp(-sX)] \ge \mathbb{E}[\exp(-sY)] \text{ for all } s > 0.$$

Note that the stochastic orders in Definition 3 can be further represented by the following relations, which are more easily evaluated.

Theorem 3 ([15]). *For random variables X and Y, $X \leq_{st} Y$ if and only if $\bar{F}_X(x) \leq \bar{F}_Y(x)$ for all x, and $X \leq_{icx} Y$ if and only if*

$$\int_t^\infty \bar{F}_X(h)dh \leq \int_t^\infty \bar{F}_Y(h)dh \tag{4}$$

for all t. Moreover, $X \leq_{cx} Y$ if and only if (4) is valid for all t and $\mathbb{E}[X] = \mathbb{E}[Y]$. Similarly, $X \leq_{icv} Y$ if and only if

$$\int_{-\infty}^t \bar{F}_X(h)dh \leq \int_{-\infty}^t \bar{F}_Y(h)dh \tag{5}$$

for all t. Finally, $X \leq_{cv} Y$ if and only if (5) is valid for all t and $\mathbb{E}[X] = \mathbb{E}[Y]$. $X \leq_{Lt} Y$ if an only if

$$\int_0^\infty e^{-sh}\bar{F}_X(h)dh \leq \int_0^\infty e^{-sh}\bar{F}_Y(h)dh, \text{ for all } s > 0. \tag{6}$$

Note that when X and Y are nonnegative, the condition $\mathbb{E}[X] = \mathbb{E}[Y]$ can be further expressed as $\int_0^\infty \bar{F}_X(h)dh = \int_0^\infty \bar{F}_Y(h)dh$ [22].

Compared with the original expectation, the integral form of the CCDFs, which unifies the expression of the considered stochastic orders as functions of the CCDFs only, highly simplifies the following derivations. The relation between the aforementioned stochastic orders can be seen from the Venn diagram in Figure 1. By Definition 3, the constraint to be fulfilled by the Laplace transform order is the least restrict, so pairs of random variables belong to the concave order, increasing concave order, or the usual stochastic order, must also belong to the Laplace transform order. Note that the intersection between the concave order and the usual stochastic order happens only when the distributions of the two random variables are identical, which can be easily seen from the constraint $\mathbb{E}[X] = \mathbb{E}[Y]$ due to the concave order. In the following sections, we will develop our results mainly based on the usual stochastic order and also the Laplace transform order. Due to the indirect relation to wireless channels, we do not discuss stochastic orders such as convex/concave and increasing convex/concave orders here. Some few discussions can be referred to [14].
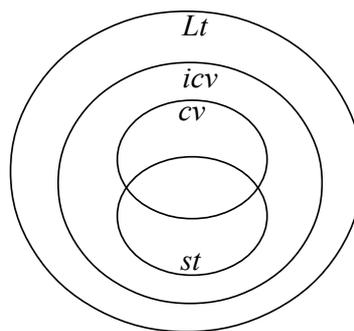

Figure 1. Venn diagram of different stochastic orders including Laplace transform order, increasing concave order, concave order, and usual stochastic order.

3. Main Results

In this section, we first formulate a problem to classify fading channels under which we can characterize the capacity results when only the statistical CSIT is available. Then, we develop a general framework in order to partly solve the formulated problem. After that, we exploit the framework to analyze the performances of several important multi-user additive white Gaussian noises (AWGN) channels including the interference channel, broadcast channel and an extension to channels with multiple antennas.

3.1. Problem Formulation

In the following, we use two simple examples to show the difficulty of the comparison in the considered scenario. In Figure 2a, the supports of the distributions of two channels are non-overlapping. Therefore, even though there is only statistical CSIT, we know that channel 2 (with PDF f_2) is always stronger than channel 1 (with PDF f_1). In contrast, in Figure 2b, the supports of the two channel distributions overlap. Intuitively, the transmitter is not able to distinguish the stronger channel just based on the channel statistics. This is because, due to the overlapping part of the PDF's, the order of the channel realizations $H_1 = h_1$ and $H_2 = h_2$ may alter for each realization. For example, in the current sample we may have $h_1 = 3.1 > h_2 = 1.7$ but in the next sample we may have $h_1 = 2.3 < h_2 = 4.9$. Then for samples within a codeword length, there is no fixed order between the two channels.
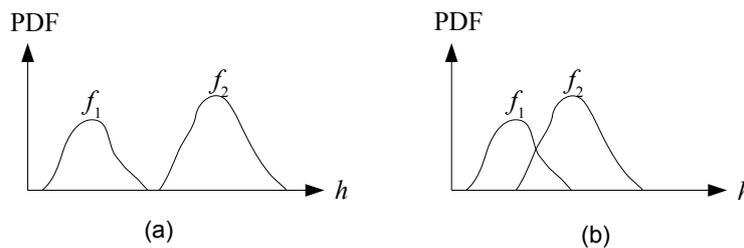

Figure 2. Two examples of relations between two fading channels: (**a**) H_2 is always stronger than H_1; (**b**) H_2 is not always stronger than H_1.

We formulate the problem as finding the tuple $(\mathcal{A}, f, \mathcal{B})$ such that the capacity results in the subset $\mathcal{B} \triangleq \{\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3\}$ are solvable, where $\mathcal{A}$ is a subset of the tuples of all random channels with a nice property, i.e., there exists a mapping f from $a \in \mathcal{A}$ to $b \in \mathcal{B}$. Figure 3 illustrates the problem formulation. Note that for each realization of b, the capacity region is provable.
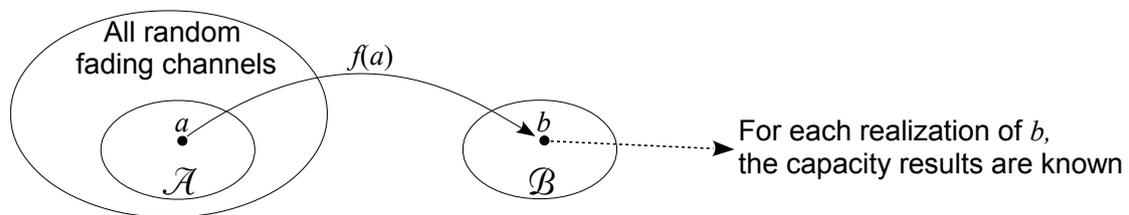

Figure 3. The proposed scheme identifies the ergodic capacity regions under statistical CSIT.

3.2. The Proposed Framework

In the following, we summarize two schemes to compare channel qualities for the case in Figure 2b, when the transmitter has only statistical CSIT. In fact, these schemes are to construct a new joint distribution which allows us to align/re-order the realizations of different fading channels in a fixed order, while the marginal distributions are not changed.

3.2.1. Coupling

In this method, we resort to constructing an explicit coupling such that each realization of a channel pair follows the same trichotomy order. To proceed, we first introduce the necessary Definition from [23].

Definition 4 ([23])**.** *The pair $(\tilde{X}, \tilde{Y})$ is a coupling of the random variables (X, Y) if $\tilde{X} =_d X$ and $\tilde{Y} =_d Y$.*

Theorem 4. *For a single-transmitter two-receiver AWGN channel, if $\mathcal{A} = \{(H_1, H_2) : H_1 \leq_{st} H_2\}$, then there exists an equivalent channel formed by $(\tilde{H}_1, \tilde{H}_2)$, where $\tilde{H}_1 \leq \tilde{H}_2$ with probability 1.*

Proof. From coupling theorem [24] we know that, $H_1 \leq_{st} H_2$ if and only if there exist random variables $\tilde{H}_1 =_d H_1$ and $\tilde{H}_2 =_d H_2$ such that $\tilde{H}_1 \leq \tilde{H}_2$ with probability 1, where the equivalent channels $\tilde{H}_1$ and $\tilde{H}_2$ can be constructed by $\tilde{H}_1 = F_{H_1}^{-1}(U)$ and $\tilde{H}_2 = F_{H_2}^{-1}(U)$, respectively, where $U \sim \text{Unif}(0,1)$. Therefore, by construction we know that distributions of $\tilde{H}_1$ and $\tilde{H}_2$ fulfill the same marginal property. In addition, the trichotomy relation $\tilde{H}_1 \leq \tilde{H}_2$ fulfills $(\tilde{H}_1, \tilde{H}_2) \in \mathcal{B}$, by construction of coupling. $\square$

Remark 3. *Originally, even though we know $H_1 \leq_{st} H_2$, the order of the channel realizations $H_1 = h_1$ and $H_2 = h_2$ may vary over time, then we may not able to claim that one channel is stronger than the other from the scope of a codeword length. However, from Theorem 4 we know that there exists an equivalent channel by which we can explicitly align all the channel realizations within a codeword length such that each channel gain realization of H_1 is no worse than that of H_2, if $H_1 \leq_{st} H_2$.*

3.2.2. Constructing a New Joint Distribution

In addition to the coupling scheme, we can also directly construct a joint distribution between the fading channels. By this way, we still can align each realization of channel pair in the same trichotomy order for all channel realizations within a codeword.

More specifically, we can design the function f by constructing a joint complementary CDF (CCDF) as follows [10,14]:

$$\bar{F}_{\tilde{H}_1, \tilde{H}_2}(\tilde{h}_1, \tilde{h}_2) = \min\{\bar{F}_{H_1}(\tilde{h}_1), \bar{F}_{H_2}(\tilde{h}_2)\}, \tag{7}$$

where $\bar{F}_{X,Y}(x,y) \triangleq P(X \geq x, Y \geq y)$, from which it is clear that the marginal distributions are unchanged, i.e., $\bar{F}_{\tilde{H}_1}(\tilde{h}_1) = \bar{F}_{\tilde{H}_1, \tilde{H}_2}(\tilde{h}_1, 0) = \bar{F}_{H_1}(\tilde{h}_1)$ and $\bar{F}_{\tilde{H}_2}(\tilde{h}_1) = \bar{F}_{\tilde{H}_1, \tilde{H}_2}(0, \tilde{h}_2) = \bar{F}_{\tilde{H}_2}(\tilde{h}_2)$. With the selection $\mathcal{A} = \{(H_1, H_2) : H_1 \leq_{st} H_2\}$, by the Definition of joint probability, we can prove that $\tilde{h}_1 \leq \tilde{h}_2, \forall (H_1, H_2) \in \mathcal{B} = f(\mathcal{A})$. In particular, assume $\tilde{h}_1 > \tilde{h}_2 + \epsilon, \epsilon > 0$, we can prove that

$$\begin{aligned}
&P(\tilde{h}_1 \leq \tilde{H}_1, \tilde{h}_2 \leq \tilde{H}_2 \leq \tilde{h}_2 + \epsilon) \\
=&P(\tilde{h}_1 \leq \tilde{H}_1, \tilde{h}_2 \leq \tilde{H}_2) - P(\tilde{h}_1 \leq \tilde{H}_1, \tilde{h}_2 + \epsilon \leq \tilde{H}_2) \\
\overset{(a)}{=}&\bar{F}_{\tilde{H}_1, \tilde{H}_2}(\tilde{h}_1, \tilde{h}_2) - \bar{F}_{\tilde{H}_1, \tilde{H}_2}(\tilde{h}_1, \tilde{h}_2 + \epsilon) \\
\overset{(b)}{=}&\bar{F}_{\tilde{H}_1}(\tilde{h}_1) - \bar{F}_{\tilde{H}_1}(\tilde{h}_1) \\
=&0, \tag{8}
\end{aligned}$$

where (a) follows from the Definition of the joint CCDF and (b) follows from (7) with the given property $H_1 \leq_{st} H_2$, which implies that $\bar{F}_{\tilde{H}_2}(\tilde{h}_2) \geq \bar{F}_{\tilde{H}_2}(\tilde{h}_2 + \epsilon) \geq \bar{F}_{\tilde{H}_1}(\tilde{h}_2 + \epsilon) \geq \bar{F}_{\tilde{H}_1}(\tilde{h}_1)$. To ensure that $\tilde{H}_1 \leq \tilde{H}_2$ for all random samples, we let $\epsilon \to 0$. Thus, as long as $H_1 \leq_{st} H_2$, we can also form an equivalent channel that has the same marginal distribution as the original one, such that the capacity is unchanged. Further discussion can be referred to [10,14].

Remark 4. *Note that the selection of $\mathcal{A}$ in Theorem 4 or in Section 3.2.2 can be easily extended to cases with K receivers by the concatenation $H_1 \leq_{st} H_2 \leq_{st} \cdots \leq_{st} H_K$.*

More discussions on the relation between the above two schemes and also the relation to copulas [25], please refer to [26]. In the following, we will use the coupling scheme introduced in Theorem 4 for MU channels due to its intuitive characteristics.

4. Applications on Gaussian MU Channels with a Single Antenna at Each Node

In this section, we will use interference channels and broadcast channels as examples to show how to apply the developed scheme. All channels in this section are assumed memoryless. All nodes are equipped with a single antenna.

4.1. Fading Gaussian Interference Channels with No CSIT

When there is only statistical CSI at the transmitter and full CSI at the receiver, the ergodic capacity region is unknown in general. In this section, we identify the sufficient condition to attain the capacity region of Gaussian interference channel with very strong interferences. Practical examples illustrate the results.

The considered received signals of a two-user fast fading Gaussian interference channel can be stated as

$$Y_1 = \sqrt{H_{11}}e^{j\Phi_{11}}X_1 + \sqrt{H_{12}}e^{j\Phi_{12}}X_2 + Z_1 \triangleq \tilde{H}_{11}X_1 + \tilde{H}_{12}X_2 + Z_1, \tag{9}$$

$$Y_2 = \sqrt{H_{21}}e^{j\Phi_{21}}X_1 + \sqrt{H_{22}}e^{j\Phi_{22}}X_2 + Z_2 \triangleq \tilde{H}_{21}X_1 + \tilde{H}_{22}X_2 + Z_2, \tag{10}$$

where H_{kj} and Φ_{kj} are real-valued non-negative independent random variables denoting the absolute square and the phase of the fading channel between the j-th transmitter to the k-th receiver, respectively, where $k, j \in \{1, 2\}$. The CCDF of H_{kj} is denoted by $\bar{F}_{H_{kj}}$. The channel inputs at the transmitters 1 and 2 are denoted by X_1 and X_2, respectively. We consider the channel input power constraint as $\mathbb{E}[|X_1|^2] \leq P_1$ and $\mathbb{E}[|X_2|^2] \leq P_2$, respectively. Noises Z_1 and Z_2 at the corresponding receivers are independent circularly symmetric AWGN with zero mean and unit variance. We assume that the transmitters only know the statistics but not the instantaneous realizations of $\{H_{kj}\}$. We also assume that each receiver knows the two channels to itself. Since we consider the case with only statistical CSIT, the channel input signals are not functions of the channel realizations. Thus, without loss of generality, we assume $\{H_{kj}\}$, $\{\Phi_{kj}\}$, $\{X_j\}$, and $\{Z_k\}$ are jointly independent.

In the following derivation, we do not exploit the commonly used standard form of GIC. This is because that the normalization operation in the standard form results in a ratio of random variables, whose distribution may not be easy to derive and thus hinders us on identifying all channels easily.

The ergodic capacity region of a very strong interference channel is as follows.

Theorem 5. *If*

$$\mathcal{A} = \left\{ (H_{11}, H_{12}, H_{21}, H_{22}) : \frac{H_{21}}{1 + P_2 H_{22}} \geq_{st} H_{11} \cap \frac{H_{12}}{1 + P_1 H_{11}} \geq_{st} H_{22} \right\}, \tag{11}$$

then the following gives the ergodic capacity region of a very strong GIC with no CSIT

$$\mathcal{C}(P_1, P_2) = \{(R_1, R_2) : R_1 \leq \mathbb{E}[C(H_{11}P_1)], R_2 \leq \mathbb{E}[C(H_{11}P_1)]\}. \tag{12}$$

Proof sketch: We first derive the ergodic capacity region of GIC with very strong interference with no CSIT and then we derive the sufficient condition to achieve the derived capacity region.

For the first part, we can solve the optimal input distributions of GIC with no CSIT by considering

$$\underset{\substack{f_{X_1}, f_{X_2}: \\ \mathbb{E}[|X_1|^2] \leq P_1, \mathbb{E}[|X_2|^2] \leq P_2}}{\arg\max} I(X_1; Y_1 | X_2, H_{11}, H_{12}) + \mu I(X_2; Y_2 | X_1, H_{22}, H_{21}), \tag{13}$$

where $\mu \in \mathbb{R}^+$. Note that (13) can be further expressed as

$$\underset{\substack{f_{X_1}, f_{X_2}: \\ \mathbb{E}[|X_1|^2] \leq P_1, \mathbb{E}[|X_2|^2] \leq P_2}}{\arg\max} \ I(X_1; H_{11}X_1 + Z_1 | H_{11}) + \mu I(X_2; H_{22}X_1 + Z_2 | H_{22}). \tag{14}$$

It is clear that for each μ, (14) can be maximized by Gaussian inputs, i.e., $X_1 \sim \mathcal{CN}(0, P_1)$ and $X_2 \sim \mathcal{CN}(0, P_2)$. Then the capacity region can be delineated by (12).

In the second part, from (2) we can derive

$$I(X_1; Y_2 | H_{21} = h_{21}, H_{22} = h_{22}) = \log\left(1 + \frac{h_{21}P_1}{1 + h_{22}P_2}\right). \tag{15}$$

After comparing (15) and $I(X_1; Y_1 | X_2, H_{11} = h_{11}, H_{12} = h_{12}) = \log(1 + h_{11}P_1)$, we can find the condition

$$h_{11} \leq \frac{h_{21}}{1 + P_2 h_{22}}. \tag{16}$$

Similarly, from (3) we can derive

$$h_{22} \leq \frac{h_{12}}{1 + P_1 h_{11}}. \tag{17}$$

By treating $H_{21}/(1 + P_2 H_{22})$ and $H_{12}/(1 + P_1 H_{11})$ as new random variables, respectively, we can extend (16) and (17) to the stochastic case as

$$\frac{H_{21}}{1 + P_2 H_{22}} \geq_{st} H_{11}, \ \text{and} \ \frac{H_{12}}{1 + P_1 H_{11}} \geq_{st} H_{22}. \tag{18}$$

Note that we can easily check that the marginal distributions are intact as shown in Theorem 2 after applying the construction by Theorem 4, which completes the proof.

Remark 5. By the considered scenarios of very strong interference channels, we can decouple the effect of the other user to single user capacity constraints, such that we are able to prove that Gaussian input can optimize the capacity region.

Examples

In this subsection, we provide examples to show the scenarios that the sufficient condition in Theorem 5 is feasible and leads to a situation which occurs in wireless communications.

Example 1: In this case, to proceed, we find the distributions of the two ratios of random variables in (11). We can first rearrange $H_{21}/(1 + P_2 H_{22})$ in the ratio of quadratic form as

$$\frac{H_{21}}{1 + P_2 H_{22}} = \frac{\boldsymbol{H}^H \boldsymbol{B_1} \boldsymbol{H}}{1 + P_2 \boldsymbol{H}^H \boldsymbol{B_2} \boldsymbol{H}} \triangleq Z, \tag{19}$$

where $\boldsymbol{H} \triangleq [\sqrt{H_{w,21}} \ \sqrt{H_{w,22}}]^T$, $H_{21} \triangleq \sigma_{21}^2 H_{w,21}$, $H_{22} \triangleq \sigma_{22}^2 H_{w,22}$, $H_{w,21} \sim N(0,1)$, $H_{w,22} \sim N(0,1)$, $H_{w,21}$ is independent to $H_{w,22}$, $\boldsymbol{B_1} \triangleq \text{diag}\{[\sigma_{21}^2, \ 0]\}$ and $\boldsymbol{B_2} \triangleq \text{diag}\{[0, \ \sigma_{21}^2]\}$. From [27] we know that the CDF of the RHS of (19) can be calculated by the following

$$\begin{aligned} F_Z(t) &\overset{(a)}{=} \text{u}(t) - \Sigma_{i=1}^2 \frac{\lambda_i^2}{\Pi_{l \neq i}(\lambda_i - \lambda_l)} \frac{1}{|\lambda_i|} e^{\frac{-t}{\lambda_i}} \text{u}\left(\frac{t}{\lambda_i}\right) \\ &\overset{(b)}{=} \text{u}(t) - \frac{1}{\sigma_{21}^2 + t P_2 \sigma_{22}^2}\left(e^{\frac{-t}{\sigma_{21}^2}} \text{u}\left(\frac{t}{\sigma_{21}^2}\right) + e^{\frac{1}{P_2 \sigma_{22}^2}} \text{u}\left(\frac{-1}{P_2 \sigma_{22}^2}\right)\right), \end{aligned} \tag{20}$$

where in (a), $\mathsf{u}(t)$ is the unit step function, $\{\lambda_i\}$ are the eigenvalues of $\boldsymbol{B_1} - tP_2\boldsymbol{B_2}$, i.e., $\{\lambda_i\} = \{\sigma_{21}^2, -tP_2\sigma_{22}^2\}$; in (b) we substitute the eigenvalues into the RHS of (a). Then we evaluate the first constraint in (11) by checking the difference of CCDFs of Z and H_{11}, i.e., $1 - F_Z(t) - e^{\frac{-t}{\sigma_{11}^2}}$ numerically. In the comparison, we fix the variances of the cross channels as $c \triangleq \sigma_{12}^2 = \sigma_{21}^2 = 1$ and the transmit powers $P \triangleq P_1 = P_2 = 1$, and scan the variances of the dedicated channels by $a \triangleq \sigma_{11}^2 = \sigma_{22}^2 = 0.1, 0.3, 0.5$ and 0.7. Since the conditions in (11) are symmetric and the considered settings are symmetric, once the first condition in (11) is valid, the second one will be automatically valid. The results are shown in Figure 4, from which we can observe that when a increases, the difference of the CCDFs decreases. This is because the support of the CCDF of H_{11} increases with increasing a and in the considered case when $a = 0.7$, the distribution of H_{11} concentrates on larger channel values than $H_{21}/(1 + P_2 H_{22})$, which reflects that the CCDFs that, H_{11} has higher CCDF than $H_{21}/(1 + P_2 H_{22})$ at lower channel values, which violates (11). In contrast, the values $a = 0.1, 0.3$ and 0.5 satisfy (11).
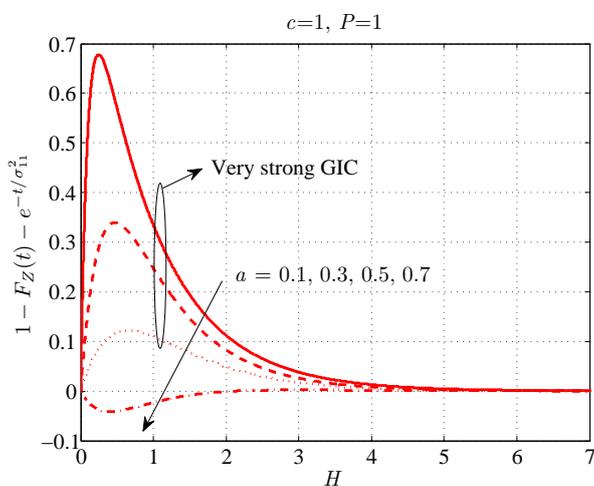

Figure 4. Identification of (11) under different variances of the dedicated channels with $c = 1$ and $P_1 = P_2 = 1$.

4.2. Fading Gaussian Broadcast Channels with No CSIT

4.2.1. Preliminaries and Results

If a 2-receiver BC is degraded, we know that the capacity region [9] is the union, over all U, X satisfying the Markov chain $U - X - Y_1 Y_2$ of rate pairs (R_1, R_2) such that

$$\begin{cases} R_1 & \leq I(X; Y_1 | U), \\ R_2 & \leq I(U; Y_2), \end{cases}$$

for some $p(u, x)$, where the cardinality of the auxiliary random variable U satisfies $|\mathcal{U}| \leq \min\{|\mathcal{X}|, |\mathcal{Y}_1|, |\mathcal{Y}_2|\} + 1$. In contrast, for non-degraded BC, only the inner [28] and outer bounds [29] are known. Therefore it is much easier to characterize the performance of a BC if we can identify its degradedness.

The capacity region of a Gaussian BC (GBC) is known for both cases with fixed and fading channels when there are perfect CSIT and CSIR. For multiple antennas GBC with perfect CSIT and CSIR, Ref. [30] invented the *channel enhancement* technique to form the degradedness and it is proved that Gaussian input is optimal. Immense endeavors have been made to find the optimal input covariance matrix of the Gaussian input, e.g., [31–33]. However, the problem is open in general when there is imperfect CSIT and only limited cases are known [10]. The fading BC with only perfect CSIR

but imperfect CSIT lacks the degraded structure in general for arbitrary fading distributions, which makes it a challenging problem. In particular, the order of channel realizations to different receivers in fast fading broadcast channels vary within a codeword length. Therefore, intuitively we are not able to compare the channels as in full CSIT cases to identify the degradedness.

We assume that there is full CSIR such that the receivers can compensate the phase rotation of their own channels, respectively, without changing the capacity to form real channels. Therefore, the signal of receiver k of the considered L-receiver fast fading Gaussian broadcast channel can be equivalently stated as

$$Y_k = \sqrt{H_k}X + Z_k, \; k = 1 \cdots L, \tag{21}$$

where X is the channel input, H_k is a real-valued non-negative independent random variable denoting the square of receiver k's fading channel with CCDF $\bar{F}_{H_k}$. We consider the channel input power constraint as $\mathbb{E}[X^2] \leq P_T$. The noises $\{Z_k\}$ at the corresponding receivers are independent AWGN with zero mean and unit variance. We assume that the transmitter only knows the statistics but not the instantaneous realizations of $\{H_k\}$.

The following result can be specialized from Remark 4.

Corollary 1. *If $\mathcal{A} = \{H_1 \geq_{st} H_2 \geq_{st} \cdots \geq_{st} H_L\}$, then it is a degraded Gaussian broadcast channel.*

Remark 6. *An important note is that degradedness does not guarantee the optimality of fading GBC's with no CSIT. In [34] it is shown that with statistical CSIT, Gaussian is not optimal in general, by local perturbation of Gaussian channel input.*

Now we can further generalize Corollary 1 to the case in which fading channels are formed by clusters of scatterers. In particular, we consider the case in which we are only provided the information of each cluster but not the superimposed result as $\sqrt{H_k}$ in (21). Therefore, phases of channels of each cluster should be taken into account, i.e., we consider the k-th clusters of the first and second users as $\tilde{H}_{1k} = \tilde{H}_{1k,Re} + i \cdot \tilde{H}_{1k,Im} = \sqrt{H_{1k}}e^{-i\phi_{1k}}$ and $\tilde{H}_{2k} = \tilde{H}_{2k,Re} + i \cdot \tilde{H}_{2k,Im} = \sqrt{H_{2k}}e^{-i\phi_{2k}}$, respectively.

The received signals at receivers 1 and 2 for M clusters are respectively expressed as

$$Y_1 = \sum_{k=1}^{M} \tilde{H}_{1,k}X + Z_1,$$

$$Y_2 = \sum_{k=1}^{M} \tilde{H}_{2,k}X + Z_2.$$

Corollary 2. *Let M be the number of clusters of scatterers for both users 1 and 2, and denote the channels of users 1's and 2's kth clusters as $\tilde{H}_{1k}$ and $\tilde{H}_{2k}$, respectively, where $k = 1, \cdots, M$. The broadcast channel is degraded if user 1's scatterers are stronger than those of user 2's in the sense that $\tilde{H}_{\pi_{1k},Re}\tilde{H}_{\pi_{1j},Re} \geq_{st} \tilde{H}_{\pi_{2k},Re}\tilde{H}_{\pi_{2j},Re}$, and $\tilde{H}_{\pi_{1k},Im}\tilde{H}_{\pi_{1j},Im} \geq_{st} \tilde{H}_{\pi_{2k},Im}\tilde{H}_{\pi_{2j},Im}$, $\forall k, j \in \{1, \cdots, M\}$, for some permutation π_1 and π_2 of users 1's and 2's clusters.*

The condition of the number of clusters of channels 1 and 2 in Corollary 2 can be relaxed to two non-negative integer-valued random variables N and M, respectively, and the degradedness among the two channels is still valid, if $|\sum_{k=1}^{N} \tilde{H}_{1k}|^2 \geq_{st} |\sum_{k=1}^{M} \tilde{H}_{2k}|^2$ and if $N \geq_{st} M$, which can be proved with the aid of Theorem 1.A.4 in [15].

4.2.2. Example

Assume the magnitudes of a three-receiver GBC are independent Nakagami-m random variables with shape parameters m_1, m_2, and m_3, and spread parameters w_1, w_2, and w_3, respectively. From Corollary 1 we know that the broadcast channel is degraded if

$$\frac{\gamma\left(m_1, \frac{m_1 x}{w_1}\right)}{\Gamma(m_1)} \geq \frac{\gamma\left(m_2, \frac{m_2 x}{w_2}\right)}{\Gamma(m_2)} \geq \frac{\gamma\left(m_3, \frac{m_3 x}{w_3}\right)}{\Gamma(m_3)}, \forall x,$$

where $\gamma(s, x) = \int_0^x t^{s-1} e^{-t} dt$ is the incomplete gamma function and $\Gamma(s) = \int_0^\infty t^{s-1} e^{-t} dt$ is the ordinary gamma function [35]. An example satisfying the above inequality is $(m_1, w_1) = (1, 3)$, $(m_2, w_2) = (1, 2)$ and $(m_3, w_3) = (0.5, 1)$.

Remark 7. *The developed scheme for the MU channels can be easily extended to those with secrecy constraint from physical layer security. For example, wiretap channels (WTC) [36], broadcast channels with confidential messages (BCCM) [37], interference channels with secrecy constraint, etc. Compared to the discussion in Remark 1, the main difference to systems with secrecy constraint is that, here, we need to additionally check the validity of the same marginal property for related terms in the secrecy constraint, e.g., the strong or weak secrecy constraints, i.e., $H(W_i|Y_j)$ or $\frac{1}{n} H(W_i|Y_j)$, $i \neq j$. More specifically, by Definition of the conditional entropy, we can easily observe that these entropies only depend on p_{W_i, Y_j}, $i \neq j$, but not depend on p_{W_i, Y_i, Y_j}. Therefore, only marginal distributions affect the performances of channels with the additional secrecy constraint. So the developed scheme works for these cases.*

Remark 8. *Note that for either GBC, GWTC, or GBC-CM, we can also use the developed scheme to check the degradedness. However, the degradedness for characterizing the performances GWTC and GBC-CM are used in a converse way. For a degraded GWTC, we can identify that Gaussian input is optimal and also the non-zero secrecy capacity. In contrast, for a two-user GBC-CM if it is degraded, then one received signal is always weaker than the other one. This means that the secrecy capacity region will degenerate to the secrecy capacity, i.e., the secrecy capacity region does not exist. It is because, for GBC-CM, both receivers are legitimate receivers as well as the eavesdroppers, simultaneously. Therefore, the degraded cases of GBC-CM should be avoided.*

5. Extension to Multiple Antenna Cases

In this section we consider multiple-antenna at both channel input and output. We assume all nodes are equipped with the same number of antennas n_T. In the following we discuss cases with only one transmitter, e.g., GBC, GWTC, etc. The results can be easily extended to cases with multiple transmitters. Again the signals at receivers 1 and 2 can be respectively expressed as

$$Y_1 = H_1 X + Z_1, \tag{22}$$
$$Y_2 = H_2 X + Z_2, \tag{23}$$

where $Z_1 \sim \mathsf{CN}(0, I_{n_T})$ and $Z_2 \sim \mathsf{CN}(0, I_{n_T})$, H_1 and $H_2 \in \mathbb{C}^{n_T \times n_T}$ with entries varying over each code symbol. For the (fading) multiple-antenna cases, the description of the channels is by (random) matrices. How to order random matrices or which part of the matrices to be ordered is critical to be clarified. In the following we provide two methods to deal with this problem, including the aforementioned coupling scheme and the channel enhancement scheme.

5.1. Alignment by Usual Stochastic Order

From random matrix theory we know that the probability of a random matrix to be full rank approaches 1. In addition to the assumption of full channel state information at the receiver (CSIR),

we can construct an alternative channel which is full rank and does not change the capacity [1] by normalizing (22) and (23) equivalently as

$$Y_1' = X + Z_1',$$
(24)
$$Y_2' = X + Z_2',$$
(25)

where $Z_1' \sim \mathrm{CN}(0, A)$ and $Z_2' \sim \mathrm{CN}(0, B)$, $A \triangleq H_1^{-1}H_1^{-H}$, $B \triangleq H_2^{-1}H_2^{-H}$. For full CSIT and full CSIR cases, to make the Markov chain $X \to Y_1' \to Y_2'$ valid, i.e., to obtain a (stochastically) degraded wiretap channel, the constraint $B - A \succcurlyeq 0$ is sufficient (The reason that it is not necessary is, we may be able to use the channel enhancement scheme to obtain a degraded channel with $B \not\succcurlyeq A$). In the considered scenario we have full CSIR but only statistical CSIT. So we aim to construct an equivalent degraded channel by showing $P(B - A \succcurlyeq 0) = 1$ according to the coupling. In the following we will find the relation of the degradedness and the stochastic order among the eigenvalues of A and B. Note that in [15] the usual stochastic order in a vector (but not matrix) version is considered, where in the expression of $\mathrm{vec}(B) \leq_{st} \mathrm{vec}(A)$, the inequality is element-wise, i.e., $b_i' \leq a_i'$, $\forall i$, for $P(\mathrm{vec}(B') \leq \mathrm{vec}(A')) = 1$. However, we can not directly apply the multivariate usual stochastic order to our scenario. This is because that it will not guarantee the positive definiteness of $B - A$, which is required for the degradedness. Instead, it is sufficient to check the stochastic order of the eigenvalues of $A - B$, namely, $\Lambda_B \geq_{st} \Lambda_A$, to attain the existence of A' and B' in an equivalent channel, such that $P(B' - A' \succcurlyeq 0) = 1$ after using coupling.

We first transform $B' - A' \succcurlyeq 0$ into a form that we can simply connect it to the eigenvalues with the aid of the following lemmas.

Lemma 1 ([38]). *Let $Y \succ 0$ and Hermitian, and $X \succcurlyeq 0$ and Hermitian. $Y - X \succcurlyeq 0$ if and only if the eigenvalues of XY^{-1} all satisfy $\lambda_i \leq 1$.*

We then use the following Lemma to connect the eigenvalues of XY^{-1} to those of X and Y.

Lemma 2 ([39]). *If X and Y are $n \times n$ positive semidefinite Hermitian matrices, then*

$$\lambda_{max}(XY) \leq \lambda_{max}(X)\lambda_{max}(Y).$$
(26)

Then from Lemmas 1 and 2 we can derive the following theorem.

Theorem 6. *A sufficient condition to have a degraded multiple-antenna channel $X - Y_1' - Y_2'$ is*

$$\lambda_{min}(H_1 H_1^H) \geq_{st} \lambda_{max}(H_2 H_2^H).$$
(27)

Remark 9. *To have a degraded channel, (27) is a strict condition to satisfy. The reasons are (**1**) the condition $A \preccurlyeq B$ may not be necessary for the existence of a degraded channel. More specifically, for the full CSIT case, for arbitrary covariance matrices A and B, Ref. [1] proves that such channel can be transformed into a degraded one by the channel enhancement technique; (**2**) the usual stochastic ordering is sufficient but may not necessary, which can be seen from the SISOSE case [14].*

Remark 10. *Note that the fast fading channel with only statistical CSIT can be verified as a degraded one, if, there exists A' and B' such that $A' \succcurlyeq B'$ for each channel realization, where A' and B' are the covariance matrices of the equivalent noises at receivers 1 and 2. Then by Proposition 1 in [40] we know that solving the optimal input covariance matrix for a GWTC is a convex problem. For full CSIT cases we can use convex optimization tools to solve it numerically or some partial analytical results can be seen in [40,41], etc.*

In the following, we show a sufficient condition for channels with additional assumptions to have a degraded channel. Note that due to the additional assumptions, we can derive a less stringent sufficient condition than that in Theorem 6.

Theorem 7. *Let $U_1 D_1 V_1^H$ and $U_2 D_2 V_2^H$ be the singular value decompositions of H_1 and H_2, respectively. Assume that V_1 is independent to D_1 and U_1, and V_2 is independent to D_2 and U_2. Also assume that V_1^H and V_2^H have the same distribution. If $D_1 \geq_{st} D_2$, then there exists an equivalently degraded channel.*

Proof. To proceed, we form a new 1st user's channel as $H_1' = U_1 D_1 V_2^H$. We can check that the PDF of H_1' is the same as that of H_1 by the following

$$f_{H_1'} = f_{U_1 D_1 V_2^H} \overset{(a)}{=} f_{U_1 D_1 V_1^H} \overset{(b)}{=} f_{H_1}, \tag{28}$$

where (a) is due to the following. To calculate the distributions of $U_1 D_1 V_2^H$ and $U_1 D_1 V_1^H$, only the joint distributions f_{U_1, D_1, V_2^H} and f_{U_1, D_1, V_1^H} are needed. With the assumptions that V_1 and V_2 are independent to D_1 and U_1, we can further have $f_{U_1, D_1, V_1^H} = f_{U_1, D_1} f_{V_1^H}$ and $f_{U_1, D_1, V_2^H} = f_{U_1, D_1} f_{V_2^H}$. Since V_2^H and V_1^H have the same PDF, i.e., $f_{V_2^H} = f_{V_1^H}$, we know that $f_{U_1, D_1, V_2^H} = f_{U_1, D_1, V_1^H}$. (b) is by Definition of H_1. By (28) and the same marginal property, we know that this new channel with H_1' being the new 1st user's channel has the same ergodic secrecy capacity as that with H_1. Then following the same steps in Theorem 6, we can complete the proof. □

Remark 11. *Channel matrices with i.i.d. Gaussian entries are valid for the requirement in Theorem 7, i.e., V_1 is independent to D_1 and U_1. In particular, we can apply the LQ decomposition (LQD) on those channel matrices to get the right singular vector V_1 and V_2 which are the Q matrices of the LQD and are independent of the L matrix [42]. In addition, the random matrix Q follows the isotropic distribution (i.d.) with PDF [43]*

$$f(Q) = \frac{\Pi_{k=1}^{n_T} \Gamma(k)}{\pi^{\frac{n_T(n_T+1)}{2}}} \delta(Q^H Q - I_{n_T}), \tag{29}$$

where Γ is the gamma function and δ is the delta function.

In the following, we consider another condition on channel matrices with a special structure. We can prove that if the channels can be decomposed into i.d. unitary matrices, the channel is equivalent to a degraded one.

Theorem 8. *Let $H_1' = \Sigma_1^{1/2} H_1$, $H_2' = \Sigma_2^{1/2} H_2$. If H_1 and H_2 are i.d. and $\Sigma_1 \succeq \Sigma_2 \succ 0$, then (H_1', H_2') is equivalent to a degraded channel.*

Proof. We can form an equivalent channel as

$$Y_1' = H_1 X + Z_1',$$
$$Y_2' = H_2 X + Z_2',$$

where $Z_1' \sim \text{CN}(0, \Sigma_1^{-1})$ and $Z_2' \sim \text{CN}(0, \Sigma_2^{-1})$. After applying the eigenvalue decomposition, the covariance matrices of Z_1' and Z_2' can be expressed as $U_1 D_1 U_1^H$ and $U_2 D_2 U_2^H$, respectively, with $D_2 \succeq D_1$, by the monotonicity theorem (Theorem 8.4.9 in [44]). Now let $Z_2'' \sim \text{N}(0, U_2(D_2 - D_1)U_2^H)$ which is independent of Z_1' and Z_2'. We can form another equivalent channel at Eve as

$$\tilde{Y}_2 \triangleq U_2 U_1^H Y_1' + Z_2''$$

$$\overset{(a)}{=} U_2 U_1^H H_1 X + U_2 D_1^{1/2} W + Z_2''$$

$$\overset{(b)}{=} U_2 U_1^H H_1 X + Z_2''',$$

where in (a) $W \sim \mathsf{N}(\mathbf{0}, I_{n_T})$ and in (b) $Z_2''' \sim \mathsf{N}(\mathbf{0}, U_2 D_2 U_2^H)$, which has the same distribution as that of Z_2'. From the above it is clear that $\tilde{Y}_2$ is stochastically degraded of Y_1'. Since H_1 is i.d., we know $U_2 U_1^H H_1$ and H_1 have the same distribution. In addition, since H_1 and H_2 are i.d., we have $f_{\tilde{Y}_2|X} = f_{Y_2|X}$. By same marginal property, we know that the equivalent channel $\tilde{Y}_2$ has the same capacity as the original one. Thus, we conclude that the original channel is equivalent to a degraded one with the same secrecy capacity. □

Remark 12. *The constraint $\Sigma_1 \succeq \Sigma_2$ in Theorem 8 may be relaxed to $\Sigma_1 \not\succeq \Sigma_2$ by the deterministic channel enhancement, which will be shown in the next sub-section.*

5.2. Alignment by Channel Enhancement

In this section, we discuss how to apply the channel enhancement argument [30], which is originally designed for channels with full CSIT, to the considered model where there is only statistical CSIT and the channels are fast faded.

The channel enhancement technique, invented in [30], is a critical technique to prove that Gaussian input is optimal for MIMO Gaussian BC. Later [1] applies this technique to wiretap channels. However, there are major differences on the use of channel enhancement. In MIMO GBC, every sub-channel from the transmitter to each receiver is enhanced by reducing the corresponding noise covariance. In GWTC, however, only those Bob's sub-channels weaker than those of Eve are enhanced to be the same as Eve's. Therefore, an equivalently degraded WTC can be formed. For more detailed discussions please see [1]. Note that in both [1,30], perfect CSIT are required.

For fading channels which are not isotropically distributed, we use the following example to show that it is still possible to use channel enhancement to attain the capacity region/secrecy capacity.

Example 2: For the received signals

$$Y_1 = H\Sigma_1^{1/2} X + Z_1, \tag{30}$$

$$Y_2 = H\Sigma_2^{1/2} X + Z_2, \tag{31}$$

assume that the fading channel H has realizations $\{H_0, \{AH_0\} : A \in U(n^2)\}$, where $U(n)$ is the unitary group with degree n. Then it can be easily seen that we can apply the channel enhancement to the pair of channel realizations $(H_0 \Sigma_1^{1/2}, H_0 \Sigma_2^{1/2})$ to achieve the capacity region/secrecy capacity. A simple way to see it is that the receivers know A and multiplying Y_1 and Y_2 by unitary A does not change the capacity.

Following similar steps in Example 2, we can easily see that if the channels are respectively modeled as $\Sigma_1^{1/2} H$ and $\Sigma_2^{1/2} H$, channel enhancement can still work.

6. Other Stochastic Orders

In this section, we show an application of the Laplace transform order introduced in Definition 3 on solving the resource allocation problem for channels with multiple antennas to attain the capacity result. We consider the wiretap channel as an example. Note that the secrecy capacity of a multiple-antenna Gaussian WTC is a difference of two concave functions with respect to the channel input covariance matrix. Under perfect CSIT assumption, in [45] the authors proved that the maximum secrecy capacity coincides with the saddle point of a min-max problem, which is by considering a Sato-type outer bound setting, i.e., Bob additionally knows what Eve knows, in addition to an additional parameter,

i.e., a correlation matrix between the noises at Bob and Eve. Based on the min-max description, [40] then developed an algorithm to numerically solve this problem. An analytical solution is still unknown. In contrast, with statistical CSIT, optimal input distribution is open in general. In the following, we will review the result [13] that under i.i.d. Rayleigh fading, the Laplace transform order combined with completely monotone can help us to prove that uniform power allocation is optimal for multiple-input single-output singe-antenna at eavesdropper GWTC with statistical CSIT of both the legitimate and eavesdropper's channels.

First of all, we want to find the optimal input covariance matrix Σ_x by solving:

$$
\arg\max_{\Sigma_x} C_s = \arg\max_{\Sigma_x} \left(\mathbb{E}_g \left[\log \frac{\frac{\sigma_g^2}{\sigma_h^2} + g^H \Sigma_x g}{\frac{\sigma_g^2}{\sigma_h^2}} \right] - \mathbb{E}_g[\log(1 + g^H \Sigma_x g)] \right). \tag{32}
$$

After the derivation in [13] we can transform (32) into the following power allocation problem

$$
\max_D \mathbb{E}_g \left[\log \left(a + g^H D^* g \right) \right] - \mathbb{E}_g \left[\log \left(1 + g^H D^* g \right) \right], \tag{33}
$$

where we denote σ_g^2/σ_h^2 by a, which belongs to $[0,1)$ since we only need to consider the case where $\sigma_h > \sigma_g$. From Section V in [46], the optimal power allocation D satisfies $\text{Tr}(D) = P$. Then for any $D = [d_1, d_2, \cdots, d_{N_T}]$ where $\sum_{i=1}^{N_T} d_i = P$ and $d_i \geq 0, \forall i$. Here we introduce some results from the stochastic ordering theory [15] to proceed.

Definition 5 ([15]). *A function $\psi(x) : [0, \infty) \to \mathbb{R}$ is completely monotone if for all $x > 0$ and $n = 0, 1, 2, \cdots$, its derivative $\psi^{(n)}$ exists and $(-1)^n \psi^{(n)}(x) \geq 0$.*

Lemma 3 ([15]). *Let B_1 and B_2 be two nonnegative random variables. If $B_1 \leq_{LT} B_2$ then $\mathbb{E}[f(B_1)] \leq \mathbb{E}[f(B_2)]$, where the first derivative of a differentiable function f on $[0, \infty)$ is completely monotone, provided that the expectations exist.*

To solve (33), we let $B_1 = g^H D g$, $B_2 = g^H D^* g$, and $f(x) = \log(a + x) - \log(1 + x)$ to invoke Lemma 3. It can be easily verified that $\psi(x)$, defined as the first derivative of $f(x)$, is completely monotone by checking Definition 5. More specifically, the n-th derivative of ψ meets

$$
\psi^{(n)}(x) = \begin{cases} \frac{n!}{(a+x)^{n+1}} - \frac{n!}{(1+x)^{n+1}} > 0, & \text{if } n \text{ is even,} \\ \frac{-n!}{(a+x)^{n+1}} + \frac{n!}{(1+x)^{n+1}} < 0, & \text{if } n \text{ is odd,} \end{cases} \tag{34}
$$

when $x > 0$, since $a \in [0, 1)$. Now from Lemma 3 and Definition 3, we know that to prove (33) is equivalent to proving $\mathbb{E}[e^{-sB_1}] \geq \mathbb{E}[e^{-sB_2}]$ or

$$
\log \left(\frac{\mathbb{E}[e^{-sB_1}]}{\mathbb{E}[e^{-sB_2}]} \right) \geq 0, \forall s > 0.
$$

From [47], we know that

$$
\log \left(\frac{\mathbb{E}[e^{-sB_1}]}{\mathbb{E}[e^{-sB_2}]} \right) = \sum_{k=1}^{N_T} \log(1 + \sigma_g^2 d_k^* s) - \sum_{k=1}^{N_T} \log(1 + \sigma_g^2 d_k s). \tag{35}
$$

To show that the above is nonnegative, we resort to the majorization theory. Note that $\sum_{k=1}^{N_T} \log(1 + \sigma_{\check{g}}^2 \check{d}_k s)$ is a Schur-concave function [39] in $(\check{d}_1, \ldots, \check{d}_{N_T})$, $\forall s > 0$, and by Definition of majorization [39],

$$(d_1^*, \cdots, d_{N_T}^*) \triangleq (P/N_T, P/N_T, \cdots, P/N_T)$$
$$\prec (d_1, d_2, \cdots d_{N_T}), \tag{36}$$

where $\boldsymbol{b} \prec \boldsymbol{a}$ means that $\boldsymbol{b}$ is majorized by $\boldsymbol{a}$. Thus, from [39], we know that the RHS of (35) is nonnegative, $\forall s > 0$. Then (33) is valid, and $\boldsymbol{D}^*$ is optimal. Note that $\boldsymbol{D}^*$ is also the optimal input covariance matrix Σ_x since the optimal beamformer $\boldsymbol{U}$ can be selected as $\boldsymbol{I}$.

7. Conclusions

In this paper, we investigated the ergodic capacity results of the fast fading Gaussian memoryless multiuser channels when only the statistics of the channel state information are known at the transmitter. To achieve this goal, we resorted to classifying the random channels through their probability distributions, by which we are able to attain the capacity results. In particular, we derived sufficient conditions to attain some information theoretic channel orders such as degraded and very strong interference by combining the usual stochastic order and the same marginal property, such that the capacity regions can be simply characterized, which include Gaussian interference channels and Gaussian broadcast. An extension of the framework to channels with multiple-antenna was also considered. Practical examples illustrated the application of the derived results.

Author Contributions: Pin-Hsun Lin and Eduard A. Jorswieck conceived the project, formulate the problems, and contribute to different parts of the work; Pin-Hsun Lin performed the main mathematical analysis; Pin-Hsun Lin wrote the paper. Both authors have read and approved the final version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, T.; Shamai (Shitz), S. A note on the secrecy capacity of the multiple-antenna wiretap channel. *IEEE Trans. Inf. Theory* **2009**, *55*, 2547–2553.
2. Annapureddy, V.S.; Veeravalli, V.V. Gaussian interference networks: Sum capacity in the low-interference regime and new outer bounds on the capacity region. *IEEE Trans. Inf. Theory* **2009**, *55*, 3032–3050.
3. Sato, H. The capacity of the Gaussian interference channel under strong interference. *IEEE Trans. Inf. Theory* **1981**, *27*, 786–788.
4. Carleial, A.B. A case where interference does not reduce capacity. *IEEE Trans. Inf. Theory* **1975**, *21*, 569–570.
5. Han, T.S.; Kobayashi, K. A new achievable rate region for the interference channel. *IEEE Trans. Inf. Theory* **1981**, *27*, 49–60.
6. Liang, Y.; Poor, V.; Shamai (Shitz), S. Secure communication over fading channels. *IEEE Trans. Inf. Theory* **2008**, *54*, 2470–2492.
7. Sankar, L.; Shang, X.; Erkip, E.; Poor, H.V. Ergodic fading interference channels: Sum capacity and separability. *IEEE Trans. Inf. Theory* **2011**, *57*, 2605–2626.
8. Caire, G.; Shamai (Shitz), S. On the capacity of some channels with channel state information. *IEEE Trans. Inf. Theory* **1999**, *45*, 2007–2019.
9. Gamal, A.E.; Kim, Y.H. *Network Information Theory*; Cambridge University Press: Cambridge, UK, 2012.
10. Tse, D.N.C.; Yates, R.D. Fading broadcast channels with state information at the receivers. *IEEE Trans. Inf. Theory* **2012**, *58*, 3453–3471.
11. Vahid, A.; Maddah-Ali, M.A.; Avestimehr, A.S.; Zhu, Y. Binary Fading Interference Channel with No CSIT. *IEEE Trans. Inf. Theory* **2017**, *63*, 3565–3578.
12. Zhu, Y.; Guo, D. Ergodic fading Z-interference channels without state information at transmitters. *IEEE Trans. Inf. Theory* **2011**, *57*, 2627–2647.

13. Lin, S.C.; Lin, P.H. On ergodic secrecy capacity of multiple input wiretap channel with statistical CSIT. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 414–419.

14. Lin, P.H.; Jorswieck, E.A. On the fast fading Gaussian wiretap channel with statistical channel state information at the transmitter. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 46–58.

15. Shaked, M.; Shanthikumar, J.G. *Stochastic Orders*; Springer: Berlin, Germany, 2007.

16. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 1st ed.; Wiley: New York, NY, USA, 1991.

17. Rajan, A.; Tepedelenlioğlu, C. Stochastic ordering of fading channels through the shannon transform. *IEEE Trans. Inf. Theory* **2015**, *61*, 1619–1628.

18. Nguyen, V.M.; Kountouris, M. Performance Limits of Network Densification. Available online: https://arxiv.org/pdf/1611.07790.pdf (accessed on 21 September 2017).

19. Moser, S.M. Advanced Topics in Information Theory-Lecture Notes. Available online: http://moser-isi.ethz.ch/docs/atit_script_v210.pdf (accessed on 21 September 2017).

20. Körner, J.; Marton, K. Comparison of two noisy channels. In *Colloquia Mathematica Societatis, János Bolyai, 16, Topics in Information Theory*; János Bolyai Mathematical Society: Budapest, Romania, 1977; pp. 411–424.

21. Markur, A.; Polyanskiy, Y. Comparison of Channels: Criteria for Domination by a Symmetric Channel. Available online: https://arxiv.org/abs/1609.06877 (accessed on 21 September 2017).

22. Hajek, B. Notes for ECE 534: An Exploration of Random Processes For Engineers. Available online: http://www.ifp.illinois.edu/~hajek/Papers/randomprocJan14.pdf (accessed on 21 September 2017).

23. Ross, S.M.; Peköz, E.A. *A Second Course in Probability*; ProbabilityBookstore.com: Boston, MA, USA, 2007.

24. Thorisson, H. *Coupling, Stationarity, and Regeneration*; Springer: New York, NY, USA, 2000.

25. Nelson, R.B. *An Introduction to Copulas*, 2nd ed.; Springer: Berlin, Germany, 2006.

26. Lin, P.H.; Jorswieck, E.A.; Schaefer, R.F.; Mittelbach, M.; Janda, C.R. On stochastic orders and ergodic capacity results of fast fading multiuser channels with statistical channel state information at the transmitter. To be submitted.

27. Al-Naffouri, T.Y.; Hassibi, B. On the distribution of indefinite quadratic forms in Gaussian random variables. In Proceedings of the International Symposium on Information Theory, Seoul, Korea, 28 June–3 July 2009; pp. 1744–1748.

28. Marton, K. A coding theorem for the discrete memoryless broadcast channel. *IEEE Trans. Inf. Theory* **1979**, *25*, 306–311.

29. Nair, C.; Gamal, A.E. An outer bound to the capacity region of the broadcast channel. *IEEE Trans. Inf. Theory* **2007**, *53*, 350–355.

30. Weingarten, H.; Steinberg, Y.; Shamai (Shitz), S. The capacity region of the Gaussian multiple-input multiple-output broadcast channel. *IEEE Trans. Inf. Theory* **2006**, *52*, 3936–3964.

31. Dabbagh, A.D.; Love, D.J. Precoding for multiple antenna gaussian broadcast channels with successive zero-forcing. *IEEE Trans. Signal Process.* **2007**, *55*, 3837–3850.

32. Oechtering, T.; Jorswieck, E.; Wyrembelski, R.; Boche, H. On the optimal transmit strategy for the MIMO bidirectional broadcast channel. *IEEE Trans. Commun.* **2009**, *57*, 3817–3826.

33. Hellings, C.; Joham, M.; Utschick, W. Gradient-based power minimization in MIMO broadcast channels with linear precoding. *IEEE Trans. Signal Process.* **2012**, *60*, 877–890.

34. Abbe, E.; Zheng, L. A coordinate system for Gaussian networks. *IEEE Trans. Inf. Theory* **2012**, *58*, 721–733.

35. Abramowitz, M.M.; Stegun, I.A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*; Dover: New York, NY, USA, 1972.

36. Csiszár, I.; Korner, J. Broadcast channels with confidential messages. *IEEE Trans. Inf. Theory* **1978**, *24*, 339–348.

37. Liu, R.; Maric, I.; Spasojevic, P.; Yates, R.D. Discrete memoryless interference and broadcast channels with channels with confidential messages: Secrecy rate regions. *IEEE Trans. Inf. Theory* **2008**, *54*, 2493–2507.

38. Seber, G.A.F. *A Matrix Handbook for Statisticians*; Wiley: New York, NY, USA, 2008.

39. Marshall, A.W.; Olkin, I. *Inequalities: Theory of Majorization and Its Application*, 2nd ed.; Academic Press: New York, NY, USA, 1979.

40. Loyka, S.; Charalambous, C.D. An algorithm for global maximization of secrecy rates in Gaussian MIMO wiretap channels. *IEEE Trans. Commun.* **2015**, *63*, 2288–2299.

41. Fakoorian, S.A.A.; Swindlehurst, A.L. Full rank solutions for the MIMO Gaussian wiretap channel with an average power constraint. *IEEE Trans. Signal Process.* **2013**, *61*, 2620–2631.

42. Gupta, A.; Nagar, D. *Matrix Variate Distributions*; Chapman & Hall: London, UK, 2000.

43. Hassibi, B.; Marzetta, T.L. Multiple-antennas and isotropically random unitary inputs: The received signal density in closed form. *IEEE Trans. Inf. Theory* **2002**, *48*, 1473–1484.

44. Bernstein, D.S. *Matrix Mathematics*; Princeton University Press: Princeton, NJ, USA, 2009.

45. Khisti, A.; Wornell, G.W. Secure transmission with multiple antennas II: The MIMOME wiretap channel. *IEEE Trans. Inf. Theory* **2010**, *56*, 5515–5532.

46. Li, J.; Petropulu, A.P. On ergodic secrecy rate for Gaussian MISO wiretap channels. *IEEE Trans. Wirel. Commun.* **2011**, *10*, 1176–1187.

47. Mathai, A.M.; Provost, S.B. *Quadratic Forms in Random Variables*; Marcel Dekker: New York, NY, USA, 1992.