Capacity of MIMO Channels

Under Independent and Correlated Fading

Zichao Zhang

Project report of ELG 7177
MIMO Communications

uOttawa

Department of Electrical Engineering and Computer Science
University of Ottawa

March 14, 2022

Contents

Abstract

Extevsive studies have shown that using multiple element array (MEA) at both transmitter and receiver in wireless communication can bring significant gain to theoretical channel capacity. Such multiple-input-multiple-output (MIMO) technology has been used for a long time and massively applied in modern wireless mobile communication systems. In this report we repeat the study in several papers to show models of MIMO communication under fading channel, correlated channel and multi-user channel. Then we analyse the capacity of them.

1 Contributions

I completed the missing steps of proofs of theorems, and completed middle steps in derivation. I did simulation about empirical eigenvalue distribution and asymptotic capacity under independent fading. I marked all the contributions with red rectangles.

2 introduction

Wireless communication systems have changed greatly over generations, with data transmission rate evolving from kbps scale to Gbps scale, researchers still seek for even higher data rate. In Shannon's channel coding theory, maximum possible transmission rate for a certain channel is bounded, or the spectral efficiency is bounded. The limit is determined by signal to noise ratio (SNR) according to Shannon's theory. In order to increase channel capacity, early researchers like Foschini [3] have predicted optimistic performance for MIMO systems, which uncovered great potential for this technique and stimulated research interest. With proper design of the MEA system, we can achieve spectral efficiency almost linear to the number of antennas. However, wireless channel is more complicated than ordinary wired channel. According to the properties of radio propagation, signal waves reach the receiver from various directions, causing interference to each other, causing the amplitude of receiving signal varies with time, which is called the fading phenomenon [13]. Real life scenarios such as reflection from buildings, diffraction from objects and scattering caused by rough surfaces can cause pathloss. Empirical models like the log-distance path loss model can be used to approximate the real channel in analysis. Furthermore, the mobility of receiver causes multipath propagation bringing about rapid change in received signal power. In spite of all these impairments, the capacity of these channels can still be solved by similiar method used in capacity analysis of toy MIMO channel model.

In this report, we first show the channel model of Rayleigh fading, which is independent fading channel, then we analyse the asymptotic capacity of it in 4.1. After that, we go to multiuser MIMO channel, introduce MIMO multiple access channel (MAC) model and MIMO broadcast channel (BC). When dealing with correlated channel, the Kronecker model is used to deal with correlation. A deterministic method is used in 4.2.2 with Shannon transform and Stieltjes transform. Simulation of empirical distribution of eigenvalues of Hermitian matrix is conducted in 5.

3 Literature Review

In this section, we make a comparation between four papers, they are the papers I refered to at the beginning. However, most of my work is based on [1], [9], [6], [2] and [4]. Paper [1] talked about single user MIMO communication system under independent fading and correlated fading, their channel model is N by N channel. They analyzed performance of system under fading channel with optimal power allocation (water filling) and equal power allocation (isotropic signaling). Capacity is achieved with water-filling, mutual information is evaluated with isotropic signaling. Different from [2], they first analysed capacity of single user MIMO system without fading, and compare the waterfilling gain in low SNR and high SNR case. In the paper, they found the largest eigenvalue of matrix $\mathbf{HH}^\dagger$ divided by N approaches 4 when N grows to infinity, it is actually the theorem in 7, we can easily compute the largest eigenvalue will be 4 with $b = (1 + \sqrt{\beta})^2$ with $\beta = 1$. They analyzed performance using scaled capacity and scaled mutual information, they are nothing new but capacity per receiving antenna and mutual information per receiving antenna. The scaled mutual information in low SNR is obtained by using equation in the form of 6, then apply first order approximation. Then scaled capacity in low SNR regime is obtained with asymptotic analysis. Gain of transmitting strategy is about 4 at low SNR. This corresponds to what we talked in lecture, water-filling is prefered in low SNR regime. But their analysis for high-SNR regime simply used the well known fact that water-filling has almost same performance with isotropic signaling, I believe it's better to give some explaination even if it's fact. Notice that the asymptotic analysis is done in very large size matrix (almost infinity), then the eigenvalue distribution of channel matrix converges to a fixed distribution, in this case whether the channel matrix is random or not does not matter because for large size, capacity becomes insensitive to the realization of channel matrix. For correlated fading, the Kronecker model is applied to help analysis, that model is explained later in this report. Because the asymptotic distribution of eigenvalues of the product of several matrices, the Stieltjes transform is used because by an inversion theorem, the Stieltjes transform can specify the distribution. The simulation they performed is in indoor environment and the ray-tracing software is only available for Bell lab, that makes implement simulation harder. Besides, multiuser MIMO systems are not discussed. They discovered that compared to independent fading, correlation decreases the per-antenna mutual information and capacity in high-SNR, but correlation will increase capacity in low-SNR.

In paper [4], they made a review about research results of MIMO channel capacity, it focuses on ergodic capacity of single user MIMO systems with perfect CSI both at transmitter and receiver, it briefly discussed the case that only CDI is available. Then they introduced the results of research on capacity region of MIMO MAC and MIMO BC. It doesn't have technical details nor their own contribution to the field, but it is a very comprehensive review of the industry. Furthermore, it lists open problems to be solved and acts as indicator for future research.

The third paper here [12] analyzed achievable capacity for mmWave QSM system using a 3-D statistical channel model for outdoor mmWave communications, they used QSM modulator, Doubledirectional model as 3D mmWave channel model and MLOptimum detector. Monte Carlo simulations are conducted to study the capacity performance, they used lognormal fitting for the 3D mmWave empirical channel model. They found that Presence of light

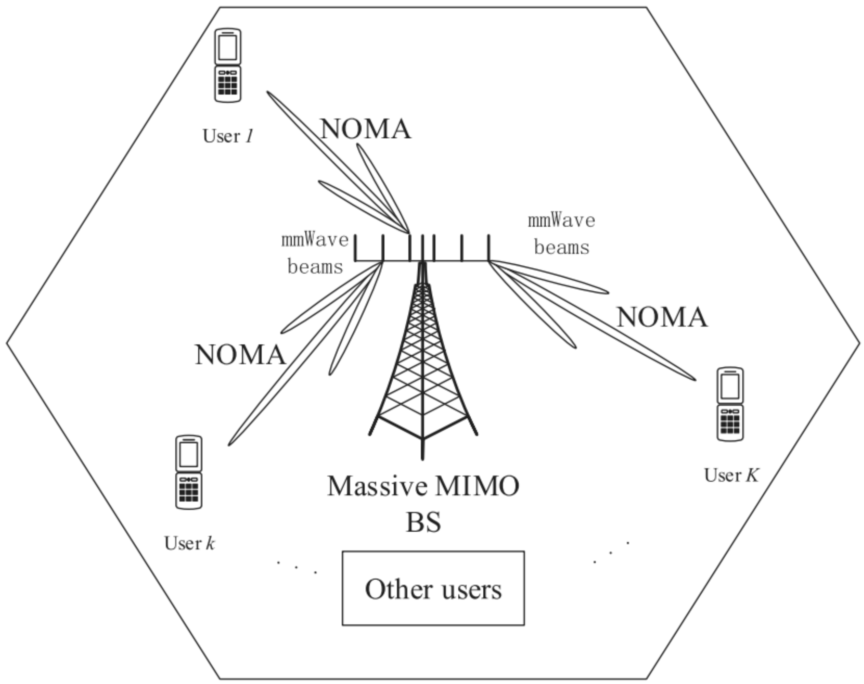

Figure 1: The proposed integrating model[14]

of sight (LOS) path increases correlation, mmWave signal is received from different subpaths, it decreases correlation. However, many theorems proposed by them do not have closed-form solution, and the conditions of achieving the capacity proposed by them is unclear.

The last paper here is [14], they integrated mmWave, NOMA and massive MIMO together to compose a model then analyze the capacity, an extended channel model for mmWave is proposed. However, their notation is chaos, nothing makes any sense and definition of symbols is not clear, proofs are not understandable. However, the model proposed is valuable, the model integrates three technologies. The base station uses massive MIMO beams to do transmission and receiving, transmission use mmWave frequency and each beam encodes with NOMA encoding scheme. The model is shown in Fig. 1. They also extended the UR-SP mmWave channel model by AoA, it is simple extension but the notation makes readers wonder how they did the extension. After they extended the channel model, they used Kronecker model to decompose the channel model into the form of the product of three matrices. However, despite the product form is correct, if this model is used in the analysis of correlated fading, strong conditions are required. Like in [2], they also used Kronecker model, but the matrices on the both sides are all covariance matrices, which means they are Hermitian matrices and Hermition propery is also used in the derivation of solutions. So I doubt the decomposition used in [14] is feasible or not. In addition, they directly applied the result obtained in [6], namely, equation 33, by simply replacing covariance matrices with decomposed matrices of channel model. This replacment is not explained and no justification is given, so I can not confirm whether it's reasonable or not.

The paper [6] is about closed-form solution of capacity under independent fading, it is fully described in the appendix A and the appendix includes almost everything in the paper and some comments and explaination of proof is provided. Besides, missing steps are completed by me and some typos are also corrected.

4 Wireless Channel Models

The significant gain in channel capacity of MIMO channel brought by adding more transmitting and receiving antennas is under the condition that transmission is diversed enough that allows the channel to be transformed into multiple independent channels [4]. In other words, the rank of channel is full rank for the best, the capacity will be propotional to the number of antennas, otherwise the capacity will be similiar to SIMO or MISO when MIMO channel is only rank one.

The capacity of time invariant channel derived by Shannon's theorem is the maximum mutual information that can be transmitted over the channel, but for time-varing channel, multiple definitions exist in the literature. According to [4], when the channel state information (CSI) is known perfectly at both receiver and transmitter, we can use adaptive methods to adjust transmission strategy to fit the changing channel state, e. g., water filling algorithm is a proper strategy. In this case, the ergodic channel capacity is defined as the expectation of capacity over all channel realizations. We can also define capacity as outage capacity and minimum-rate capacity. When the channel information isn't known at the transmitter or receiver, we can always assume the channel distribution to be Gaussian, then the channel matrix can be easily specified by its mean and variance.

4.1 Single user MIMO channel under independent fading

We can assume there are many independent signal paths at the receiver resulted from scattering or reflection, and wih random amplitude. According to central limit theorem, given that the channel gain between any two transmitting receiving antenna pairs are independent and the distribution of this channel gain is unknown, we can model every entry of channel matrix as Gaussian random variable [8]. In this scenario, we can let $\mathbf{H}$ to be our channel matrix with size N×M and assume that every entry of $\mathbf{H}$, H_{ij}, is a Gaussian variable with mean 0 and variance 1.

The channel model is defined as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{\xi} \tag{1}$$

$\mathbf{y} = [y_1, y_2, \cdots, y_N]^T$ and $\mathbf{x} = [x_1, x_2, \cdots, x_M]^T$ are received signal and transmitted signal respectively, $\boldsymbol{\xi}$ is the received noise, with zero mean and $\mathbb{E}[\boldsymbol{\xi}\boldsymbol{\xi}^T] = \sigma_0^2\mathbf{I}_n$. We define mutual information in as $\mathbf{I}(\mathbf{H})$, the $\mathbf{H}$ is given to show the dependence. The Channel capacity given $\mathbf{H}$ is

$$\begin{aligned}\mathbf{C}(\mathbf{H}) &= \underset{p(\mathbf{x})}{max}\mathbf{I}(\mathbf{H}) \\ &= \underset{Q}{max} \log |\mathbf{I}_n + \mathbf{H}\mathbf{Q}\mathbf{H}^\dagger|\end{aligned} \tag{2}$$

The matrix $\mathbf{Q}$ is the covariance matrix of $\mathbf{x}$. The optimal solution is given by

$$C^*(\mathbf{H}) = \sum_{i=1}^{N} \log(1 + \lambda_i d_i)_+ \tag{3}$$

λ_i are eigenvalues of $\mathbf{HH}^\dagger$ and d_i are power allocated to the ith eigenmode. $(x)_+ = max(x, 0)$ is the positive part. It is under the power constraint

$$tr(\mathbf{Q}) \leq P$$

It is easily shown that both the capacity and mutual information depend on the empirical distribution of eigenvalues of $\mathbf{HH}^\dagger$. If we study the case that transmitter performing equal-power allocation, and normalize the mutual information, it will be simplified to the following form .

$$\begin{aligned}\mathbf{I}(\mathbf{H}) &= \frac{1}{N} \log |\mathbf{I}_n + \gamma \mathbf{HH}^\dagger| \\ &= \frac{1}{N} \sum_{i=1}^{N} \log(1 + \gamma \lambda_i(\mathbf{HH}^\dagger))\end{aligned} \tag{4}$$

$\lambda_i(\mathbf{HH}^\dagger)$ are the eigenvalues of $\lambda_i(\mathbf{HH}^\dagger)$ and γ is defined as

$$\gamma = \frac{N\mathbb{E}\|\mathbf{x}\|^2}{M\mathbb{E}\|\boldsymbol{\xi}\|^2}$$

According to [9], define the empirical cumulative distribution function of eigenvalues of $\mathbf{HH}^\dagger$ to be

$$F^N_{\mathbf{HH}^\dagger}(x) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\{\lambda_i(\mathbf{HH}^\dagger) < x\} \tag{5}$$

$\mathbb{1}$ is the indicator function. And we notice that (4) can be written as

$$\mathbf{I}(\mathbf{H}) = \int_0^\infty \log(1 + \gamma x) dF^N_{\mathbf{HH}^\dagger}(x) \tag{6}$$

The intuition behind it is

$$\begin{aligned}\mathbf{I}(\mathbf{H}) &= \frac{1}{N} \sum_{i=1}^{N} \log(1 + \gamma \lambda_i(\mathbf{HH}^\dagger)) \\ &= \frac{1}{N} \sum_{i=1}^{N} \log(1 + \gamma \lambda) \delta(\lambda - \lambda_i(\mathbf{HH}^\dagger)) \\ &= \int_0^\infty \log(1 + \gamma x) f^N_{\mathbf{HH}^\dagger}(x) dx \\ &= \int_0^\infty \log(1 + \gamma x) dF^N_{\mathbf{HH}^\dagger}(x)\end{aligned}$$

The above explaination is not justified with serious mathematical theorem, it can act as a way of understanding. (6) can be pushed forward by applying theorems in random matrix theory. According to [9], when the entries of channel matrix $\mathbf{H}$ are i.i.d. and have zero mean, variance $\frac{1}{N}$, when M, N $\rightarrow inf$ with fixed retio $\frac{M}{N} = \beta$, the empirical distribution of eigenvalues of $\mathbf{H}\mathbf{H}^{\dagger}$ converges almost surely in distribution to a non-random limit. And the density function of it is

$$f_{\beta}(x) = (1-\beta)_{+}\delta(x) + \frac{\sqrt{(x-a)_{+}(b-x)_{+}}}{2\pi x} \tag{7}$$

where $a = (1-\sqrt{\beta})^2$ and $b = (1+\sqrt{\beta})^2$. The distribution of this limit can be shown in the following figure:

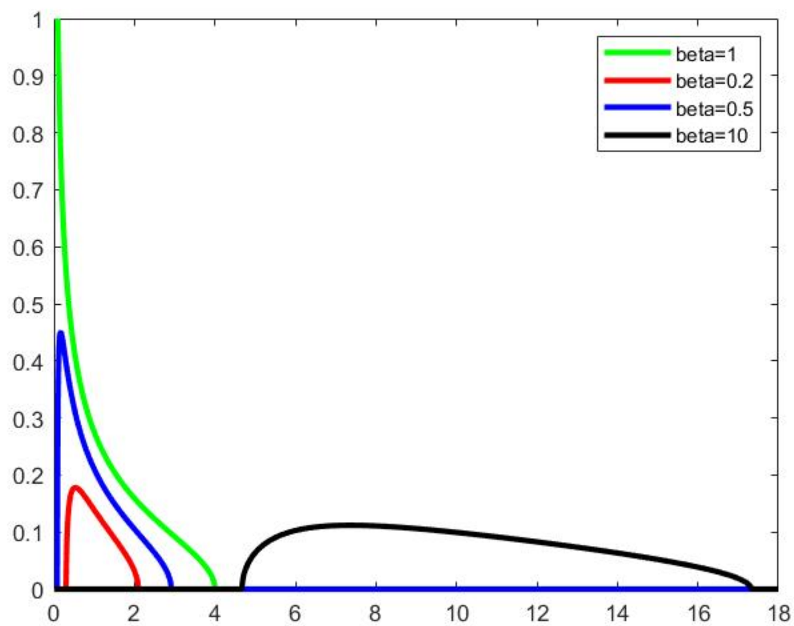

Figure 2: The limit distribution of $f_{\beta}(x)$

The asymptotic analysis used in [1] and [14] used this limit distribution that M and N approaches infinity to show the capacity result, the expression of asymptotic capacity is given below, it actually has closed form solution, and the proof of it is in [6], but we also give it in the Appendix A.

$$\begin{aligned}\frac{1}{N}\log|\mathbf{I}_n+\gamma\mathbf{H}\mathbf{H}^\dagger| \rightarrow &\int_a^b \log(1+\gamma x)f_\beta(x)dx \\ &= \beta\log\left(1+\gamma-\frac{1}{4}\mathcal{F}(\gamma,\beta)\right) \\ &+\log\left(1+\gamma\beta-\frac{1}{4}\mathcal{F}(\gamma,\beta)\right) \\ &-\frac{\mathcal{F}(\gamma,\beta)}{4\gamma}\end{aligned} \tag{8}$$

with

$$\mathcal{F}(x,z)=\left(\sqrt{x(1+\sqrt{z})^2+1}-\sqrt{x(1-\sqrt{z})^2+1}\right)^2 \tag{9}$$

Such kind of matrix has a cheering feature: it shows ergodicy. It means when the size of the matrix is large enough, we just need one realization of channel matrix to get the limit distribution, namely the empirical distribution function of any realization converges to the same asymptotic distribution. This is meaningful because it tells us as the channel matrix grows, the realization of it becomes trivial. In this case, capacity and mutual information are only relevant to the distribution of eigenvalues of $\mathbf{H}\mathbf{H}^\dagger$, which is determinant. Then we have $C(\mathbf{H})=C$ and $I(\mathbf{H})=I$.

4.2 Multiuser MIMO channel

In this section, we talk about the two models of multi-user MIMO, the MIMO broadcast channels (BC) and the MIMO multiple access channels (MAC). As indicated by the name, the BC channel models downlink transmission and MAC models uplink transmission. Imagine a base station (BS) with N antennas, there are K users in the cell, each equipped with M antennas. The users are sparse in space, so we don't consider inter-user correlation. The channel distribution for each user is different considering distance with the BS, azimuth angle, scattering environment and so on, the channel matrix for every user is different. So we denote channel matrix of each user k by $\mathbf{H}_k$ and we are using the same channel matrix in downlink and uplink for simplification. The figure below from [4] shows graphical representation of the system model.

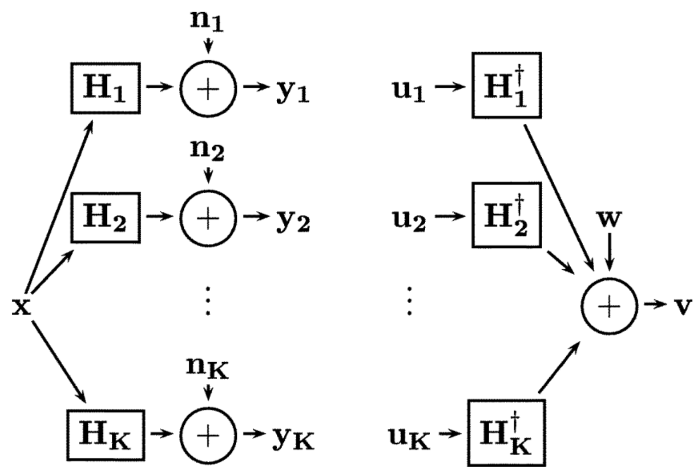

Figure 3: Models of MIMO BC (left) and MIMO MAC (right)

In the BC model, the BS transmits signal $\mathbf{x}$ to users, $\mathbf{x} \in \mathbb{C}^{N\times 1}$, received signal of each user $\mathbf{y}_k \in \mathbb{C}^{M\times 1}$, the noise for each user $\mathbf{n}_k \in \mathbb{C}^{M\times 1}$ and channel matrix $\mathbf{H}_k \in \mathbb{C}^{M\times N}$. This is also the downlink transmission. We have the relationship

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x} + \mathbf{n}_k \tag{10}$$

For the MAC model, users transmit signal to the BS, assume signal transmitted by kth user is $\mathbf{u}_k \in \mathbb{C}^{M\times 1}$, the noise in the receiverside $\mathbf{w} \in \mathbb{C}^{N\times 1}$, and received signal $\mathbf{v} \in \mathbb{C}^{N\times 1}$, we have the following relationship

$$\mathbf{v} = \sum_{k=1}^{K} \mathbf{H}_k^\dagger \mathbf{u}_k + \mathbf{w} \tag{11}$$

Note that the size of channel matrix here is slightly different with previously defined, we will stick to $\mathbf{H} \in \mathbb{C}^{N\times M}$. Next we first a model used when dealing with MIMO channel with correlated fading, it is called the Kronecker model, then we apply this method to multiuser MIMO communication with correlated fading, analysing its capacity, and mainly base on the study of [2].

4.2.1 MIMO channel under correlated fading

This part of discussion is mainly from [1], here we only talk about correlated Rayleigh fading. Correlated fading is the correlation between two transmitting receiving antenna pairs not zero, in other words, the correlation of two different elements of channel matrix is not zero, and it's easy to infer that elements of channel matrix are not independent to each other any more. But we still assume the elements of $\mathbf{H}$ are complex Gaussian with zero mean and $\mathbb{E}\left[|H_{ij}|^2\right] = 1$. It is discovered that when measuring the correlation between two paths of transmitting and receiving antennas, the correlation between paths that from two transmitting antenna to the same receiving antenna or the other way around is much stronger than that of two distinct paths. So we can use correlation between transmitting or receiving antennas to rerpresent the correlation. We denote the covariance matrix of transmitter by

$\mathbf{\Psi}^T$, and the covariance matrix of receiver by $\mathbf{\Psi}^R$. Furthermore, $\mathbf{\Psi}^T$ is of size $M \times M$ and $\mathbf{\Psi}^R$ is of size $N \times N$. Then we can denote the correlation between two transmission paths by the correlation inside trnsmitter and receiver.

$$\mathbb{E}\big[H_{ip}H_{jq}^*\big] = \Psi_{ij}^R\Psi_{pq}^T \tag{12}$$

The correlation between two transmission paths, namely two channel matrix elements, is represented by trnasmitter correlation and receiver correlation. Then we say channel matrix can be factorized to Kronecker model

$$\mathbf{H} = (\mathbf{\Psi}^R)^{\frac{1}{2}}\mathbf{X}(\mathbf{\Psi}^T)^{\frac{1}{2}} \tag{13}$$

Where the elements of $\mathbf{X}$ are i.i.d. complex Gaussian random variables and with mean 0 and variance 1. This is true and we justify it here, but first, simplify the notation $(\mathbf{\Psi}^T)^{\frac{1}{2}}$ to $\mathbf{\Psi}^{\frac{T}{2}}$ and do the same to $(\mathbf{\Psi}^R)^{\frac{1}{2}}$.

$$\begin{aligned}
\mathbb{E}\big[H_{ip}H_{jq}^*\big] &= \mathbb{E}\Bigg[\Bigg(\sum_{m=1}^{M}\sum_{l=1}^{N}\Psi_{il}^{\frac{R}{2}}W_{lm}\Psi_{mp}^{\frac{T}{2}}\Bigg)\Bigg(\sum_{b=1}^{M}\sum_{a=1}^{N}\Psi_{ja}^{\frac{R}{2}}W_{ab}\Psi_{bq}^{\frac{T}{2}}\Bigg)\Bigg] \\
&= \mathbb{E}\Bigg[\Bigg(\sum_{m=1}^{M}\sum_{l=1}^{N}\Psi_{il}^{\frac{R}{2}}W_{lm}\Psi_{mp}^{\frac{T}{2}}\Psi_{jl}^{\frac{R}{2}}W_{lm}\Psi_{mq}^{\frac{T}{2}}\Bigg)\Bigg] \\
&= \sum_{m=1}^{M}\sum_{l=1}^{N}\Psi_{il}^{\frac{R}{2}}\Psi_{mp}^{\frac{T}{2}}\Psi_{jl}^{\frac{R}{2}}\Psi_{mq}^{\frac{T}{2}} \\
&= \Bigg(\sum_{m=1}^{M}\Psi_{il}^{\frac{R}{2}}\Psi_{jl}^{\frac{R}{2}}\Bigg)\Bigg(\sum_{l=1}^{N}\Psi_{mp}^{\frac{T}{2}}\Psi_{mq}^{\frac{T}{2}}\Bigg) \\
&= \Bigg(\sum_{m=1}^{M}\Psi_{il}^{\frac{R}{2}}\Psi_{lj}^{\frac{R}{2}}\Bigg)\Bigg(\sum_{l=1}^{N}\Psi_{pm}^{\frac{T}{2}}\Psi_{mq}^{\frac{T}{2}}\Bigg) \\
&= \Psi_{ij}^R\Psi_{pq}^T
\end{aligned}$$

Note that if $m \neq b$, $l \neq a$, $\mathbb{E}\big[X_{lm}X_{ab}\big] = 0$. Thus, we proved that the form (11) is feasible.

This definition can be written in the form of Kronecker product, namely, $\mathbf{H} \sim \mathcal{CN}(0, \mathbf{\Psi}^T \otimes \mathbf{\Psi}^R)$, where $\otimes$ denotes the Kronecker product.

*proof:*Let's first flatten the matrix $\mathbf{H}$ to get a column vector $\mathbf{h}$ and then we get the covariance matrix of $\mathbf{H}$ by $\mathbb{E}[\mathbf{h}\mathbf{h}^\dagger]$, we call it $\mathbf{R}$, this notation only valid in this proof. Then we use $\mathbf{h}_i$ to denote the ith column of $\mathbf{H}$, then divide $\mathbf{R}$, get a $N \times N$ matrix with element being $M \times M$ submatrix. Then we denote the ijth submatrix of $\mathbf{R}$ by $\mathbf{R}_{ij}$, we have the following

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \cdots & \mathbf{R}_{1N} \\ \vdots & \ddots & \vdots \\ \mathbf{R}_{N1} & \cdots & \mathbf{R}_{NN} \end{bmatrix}$$

Now let's see the klth element of matrix $\mathbf{R}_{ij}$, we denote it as $(R_{ij})_{kl}$, notice that $(R_{ij})_{kl} = \mathbb{E}[H_{ki}H_{lj}] = \Psi_{kl}^R\Psi_{ij}^T$, if we change the index kl inside matrix $(R_{ij})_{kl}$, we can find that the factor Ψ^T_{ij} doesn't change, and we know $(R_{ij})_{kl}$ and $\mathbf{\Psi}^R$ have the same size, then we can write $\mathbf{R}$ as the Kronecker product of $\mathbf{\Psi}^T$ and $\mathbf{\Psi}^R$.

$$\mathbf{R} = \mathbf{\Psi}^T \otimes \mathbf{\Psi}^R$$

4.2.2 Multi-user MIMO communication with correlated fading

In this part we mainly discuss MIMO MAC model. When we have perfect CSI at transmitter (CSIT), we can use iterative water-filling algorithm [10] to find capacity region. However, to obtain CSI we need feedback link to the transmitter and it has to be fast enough, namely, the CSI feed back has to be completed before the channel changes and this usually requires channel to be quasi-static for a relatively long time, which is not realistic in high mobility user situation. In this case, it is more reasonable to have channel distribution information at the transmitter (CDIT). In [2], they proposed an deterministic equivalent method to approximate the ergodic capacity, using asymptotic method, which is accurate when the system dimension is very large.

Using the model in 11, but make modifications on notation for consistancy with [2].

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{s}_k + \mathbf{n} \tag{14}$$

We consider a BS with K users, each user is equipped with M antennas, $\beta = \frac{N}{M}$ is the ratio between number of antennas at the receiver (BS) and that at the transmitter (user equipment). The signal transmittered by user k is denoted by $\mathbf{s}_k \in \mathbb{C}^{M\times 1}$, we can assume signal is Gaussian and $\mathbb{E}[\mathbf{s_k}] = \mathbf{0}$ and $\mathbb{E}[\mathbf{s}_k\mathbf{s}_k^\dagger] = \mathbf{R}_k$. $\mathbf{y} \in \mathbb{C}^{N\times 1}$. $\mathbf{n}$ is the AWGN with zero mean and variance $\mathbb{E}[\mathbf{n}\mathbf{n}^\dagger] = \sigma^2\mathbf{I}$. Recall that the channel can be interpreted as the form of (13), we use $\mathbf{\Psi}^{\frac{R}{2}}\mathbf{X}\mathbf{\Psi}^{\frac{T}{2}}$ to represent channel matrix, and entries of $\mathbf{X}$ are i.i.d. Gaussian distribution with zero mean and variance $\frac{1}{M}$, the $\mathbf{\Psi}^{\frac{R}{2}}$ and $\mathbf{\Psi}^{\frac{T}{2}}$ are $N\times N$ and $M\times M$ deterministic Hermitian matrices respectively. Notice that here we pre-scale the matrix $\mathbf{X}$ by $\frac{1}{\sqrt{M}}$, before that, variance of entries of $\mathbf{X}$ is 1. Because if we investigate the distributions of eigenvalues of $\mathbf{X}\mathbf{X}^\dagger$, scaled by $\frac{1}{M}$, converges to deterministic distribution for large matrix size. We don't want to change anything in the definition 14, so we put the scaling factor M in $\mathbf{s}$, and we have power constraint $\frac{1}{M}tr\mathbf{R}_k \le P_k$ where P_k is the power constraint of kth user equipment (UE).

Assume the channels are varying fast, and the transmitters only know long-term distribution information about the channel, which is, the $\mathbf{\Psi}_k^R$'s and $\mathbf{\Psi}_k^T$'s. Before we study the capacity of this channel, we first formally introduce two transforms

4.2.3 Stieltjes transform

The Stieltjes transform $S_{\mathbf{HH}^\dagger}(z)$ of a hermitian matrix $\mathbf{HH}^\dagger$ is defined as

$$\begin{aligned} S_{\mathbf{HH}^\dagger}(z) &= \frac{1}{N}tr(\mathbf{HH}^\dagger - z\mathbf{I}_N)^{-1} \\ &= \int \frac{1}{\lambda - z} dF_{\mathbf{HH}^\dagger}(\lambda) \end{aligned} \tag{15}$$

where $F_{\mathbf{HH}^\dagger}(\lambda)$ is the empirical cumulative distribution function of eigenvalues of $\mathbf{HH}^\dagger$. The Stieltjes transform is used widely in wireless communication studies, it was firstly used to characterise the asymptotic distribution of eigenvalues, it is also closely connected to the Shannon transform, which is of critical importance in finding the channel capacity.

4.2.4 Shannon transform

As was shown in (6), the Shannon transform is defined in a similiar form:

$$\begin{aligned}\mathcal{V}_{\mathbf{HH}^\dagger}(z) &= \frac{1}{N}\log\left|\mathbf{I}_N + \frac{\mathbf{HH}^\dagger}{z}\right| \\ &= \int_0^{+\infty}\log\left(1+\frac{\lambda}{z}\right)dF_{\mathbf{HH}^\dagger}(\lambda) \\ &= \int_z^{+\infty}\left(\frac{1}{w} - S_{\mathbf{HH}^\dagger}(-w)\right)dw \end{aligned} \tag{16}$$

With Shannon transform, according to [2], the ergodic rate region for MIMO-MAC is given by

$$\mathcal{C}_{MAC} = \bigcup_{\substack{\frac{1}{M}tr\mathbf{R}_k \le P_k \\ \mathbf{R}_k \ge 0 \\ i=1,\cdots,K}} \left\{\{R_k, 1\le k\le K\} : \sum_{k\in\mathcal{S}} R_i \le \mathbb{E}\mathcal{V}(\mathbf{R}_{k_1},\cdots,\mathbf{R}_{k_{|\mathcal{S}|}};\sigma^2), \forall\mathcal{S}\subset\{1,\cdots,K\}\right\} \tag{17}$$

Where $\mathcal{S} = \{k_1,\cdots,k_{|\mathcal{S}|}\}$ and as the notation defined,

$$\mathcal{V}(\mathbf{R}_{k_1},\cdots,\mathbf{R}_{k_{|\mathcal{S}|}};\sigma^2) \triangleq \frac{1}{N}\log\left(\mathbf{I}_N + \frac{1}{\sigma^2}\sum_{k\in\mathcal{S}}\mathbf{H}_k\mathbf{R}_k\mathbf{H}_k^\dagger\right)$$

As we can see, Shannon transform is the key factor in this equivalence. Later we will use these two transforms to find the asymptotic capacity of correlated channel.

4.2.5 Deterministic equivalent for the Stieltjes transform

According to the study of channel capacity, the matrix $\mathbb{E}[\mathbf{yy}^\dagger]$ is of great importance, so we use (14) to expand $\mathbf{yy}^\dagger$, notice that $\mathbf{s}'_k s$ and $\mathbf{n}$ are mutually independent, so if we take expectation, many factors will become 0, we only consider those non-zero factors in the expectation. The matrix of interest is

$$\mathbf{B}_N = \sum_{k=1}^{K}\mathbf{\Psi}_k^{\frac{R}{2}}\mathbf{X}_k\mathbf{\Psi}_k^T\mathbf{X}_k^\dagger\mathbf{\Psi}_k^{\frac{R}{2}} + \mathbf{S} \tag{18}$$

Before we begin to find the closed form solution, we need to make several assumptions, according to [2],

1) $\mathbf{X}_k \in \mathbb{C}^{N\times M}$ has i.i.d. entries $\frac{1}{\sqrt{M_k}}X'_{k,ij}$, and $Var(X'_{k,ij}) = 1$

2) $\mathbf{\Psi}_k^{\frac{R}{2}} \in \mathbb{C}^{N\times N}$ is the square root of the positive semi-definite Hermitian matrix $\mathbf{\Psi}_k^R$

3) $\mathbf{\Psi}_k^T = diag(\tau_1, \cdots, \tau_M)$ and $\tau_i > 0, \forall i$
4) The sequences $\{F^{\mathbf{\Psi}_k^T}\}_M$ and $\{F^{\mathbf{\Psi}_k^R}\}_N$ are tight
5) $\mathbf{S} \in \mathbb{C}^{N \times N}$ is positive semi-definite Hermitian matrix
6) $c = \frac{N}{M}$, then there exist $b > a > 0$ that

$$a < \underset{k}{min} \lim \underset{N}{inf} \quad c \leq \underset{k}{max} \lim \underset{N}{sup} \quad c < b \tag{19}$$

Remark: The matrix matrix before normalization is $\mathbf{X}_k'$, it has entries following standard Gaussian distribution. If we have a unitary matrix $\mathbf{U} \in \mathbb{C}^{M \times M}$, the entries in matrix $\mathbf{X_k'U}$ are still standard Gaussian [8]. Besides, we can easily know $\mathbf{X}_k'$ and $\mathbf{X}_k$ has the same distribution. It means we can replace $\mathbf{X}_k$ in (18) by $\mathbf{X_kU}$ and the original matrix becomes

$$\mathbf{B}_N = \sum_{k=1}^{K} \mathbf{\Psi}_k^{\frac{R}{2}} \mathbf{X}_k (\mathbf{U}\mathbf{\Psi}_k^T \mathbf{U}^\dagger) \mathbf{X}_k^\dagger \mathbf{\Psi}_k^{\frac{R}{2}} + \mathbf{S}$$

Thus, even if in real case, the transmitter correlation matrix $\mathbf{\Psi}_k^T$ isn't diagonal, we can still use eigenvalue decomposition (EVD) to diagonalize it and only the diagonal matrix is of interest.

We provide definition of tight sequences in assumption 4):

A sequence of probability measures $\mathbb{P}_n$ on metric space $(\mathcal{S}, d)$ is called tight if $\forall \epsilon > 0$, there exists N and a compact set $\mathcal{K} \subset \mathcal{S}$, such that $\mathbb{P}_n(\mathcal{K}) > 1 - \epsilon, \forall n > N$.

This basically means that as M and N grows large, the sequence of empirical distribution $\{F^{\mathbf{\Psi}_k^T}\}_M$ and $\{F^{\mathbf{\Psi}_k^R}\}_N$ is restricted in a certain range, not spreading out over the whole $\mathbb{R}^+$. For example, the size ratio β for $\mathbf{\Psi}_k^R$ and $\mathbf{\Psi}_k^T$ is 1, we can find the corresponding asymptotic distribution in Fig. 2, it is constrained in $(0, 1]$, this means a compact set $\mathcal{K} = [0, 10^{10}]$ (10^{10} is an large enough number chosen arbitrarily) is enough to contain all possible distribution range of the sequence of empirical distributions $\{F^{\mathbf{\Psi}_k^T}\}_M$ and $\{F^{\mathbf{\Psi}_k^R}\}_N$.

In assumption 6), the concept lim sup (limit superior) is defined as follows:

Let $\{a_n\}$ be a sequence bounded from below, for arbitrary N, the subsequence

$$a_N = \{a_n : n \geq N\}$$

Then we take superior on every subsequence, and the superior of subsequence k is denoted by β_k, at last we take the limit of this superior as N goes to infinity, then we get the lim sup.

$$\lim sup\{a_n\} = \lim_{k \to \infty} \beta_k \tag{20}$$

Intuitively, we can come up with a non-convergent sequence, e.g., $a_n = \pm 1$, it does not converge to any number, but in this case the sequence composed of superior of subsequences converges. We can think of the $\lim\sup$ as the sup in the long run. The $\lim\inf$ is defined in the same way but it's the opposite. In 6), we assume the $\lim\sup$ and $\lim\inf$ of c are bounded, it means when N and M grow large, the ratio between then is bounded by its $\lim\sup$ and $\lim\inf$, furthermore bounded by a and b. It is important assumption, it assures that we are not dealing with extreme cases like infinity.

We first consider the case only one user, the matrices $\mathbf{X}$, $\mathbf{\Psi}^R$ and $\mathbf{\Psi}^T$ need to be truncated and centralized so that they have bounded elements. We show the truncated distribution function converges to that of the original matrix. For $N \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, we have

$$\left\|F^{\mathbf{A}\mathbf{A}^\dagger} - F^{\mathbf{B}\mathbf{B}^\dagger}\right\| \leq \frac{1}{N} rank(\mathbf{A} - \mathbf{B}) \tag{21}$$

Now we extend this, let c be any real number, both $\mathbf{A} + c\mathbf{I}$ and $\mathbf{B} + c\mathbf{I}$ are non-egative definite, for any $x \in \mathbb{R}$,

$$F^{\mathbf{A}}(x) - F^{\mathbf{B}}(x) = F^{(\mathbf{A}+c\mathbf{I})^2}((x+c)^2) - F^{(\mathbf{B}+c\mathbf{I})^2}((x+c)^2)$$

This is because if we diagonalize $\mathbf{A}$ and $\mathbf{B}$, we can know the entries of diagonal matrix of $\mathbf{A} + c\mathbf{I}$ and $\mathbf{B} + c\mathbf{I}$ are $\lambda_i^{\mathbf{A}} + c$ and $\lambda_i^{\mathbf{B}} + c$ where $\lambda_i^{\mathbf{A}}$ and $\lambda_i^{\mathbf{B}}$ are eigenvalues of $\mathbf{A}$ and $\mathbf{B}$. So $\|F^{\mathbf{A}} - F^{\mathbf{B}}\| = \|F^{(\mathbf{A}+c\mathbf{I})^2} - F^{(\mathbf{B}+c\mathbf{I})^2}\|$. According to (24),

$$\left\|F^{\mathbf{A}} - F^{\mathbf{B}}\right\| \leq \frac{1}{N} rank(\mathbf{A} + c\mathbf{I} - \mathbf{B} - c\mathbf{I}) = \frac{1}{N} rank(\mathbf{A} - \mathbf{B}) \tag{22}$$

Notice that $\mathbf{\Psi}^R$ and $\mathbf{\Psi}^T$ are full rank matrices, because they are covariance matrices, this is used in the following demonstration.

First let $\tilde{X}_{ij} = X_{ij}\mathbb{1}_{\{|X_{ij}|<\sqrt{N}\}} - \mathbb{E}(X_{ij}\mathbb{1}_{\{|X_{ij}|<\sqrt{N}\}})$ and $\tilde{\mathbf{X}} = \left(\frac{1}{\sqrt{M}}\tilde{X}_{ij}\right)$ and $\bar{X}_{ij} = \tilde{X}_{ij}\mathbb{1}_{\{|X_{ij}|<\ln N\}} - \mathbb{E}(\tilde{X}_{ij}\mathbb{1}_{\{|X_{ij}|<\ln N\}})$, then $\bar{\mathbf{X}} = \left(\frac{1}{\sqrt{M}}\bar{X}_{ij}\right)$.

$$\begin{aligned}
\left\|F^{\mathbf{S}+\mathbf{\Psi}^{\frac{R}{2}}\mathbf{X}\mathbf{\Psi}^T\mathbf{X}^\dagger\mathbf{\Psi}^{\frac{R}{2}}} - F^{\mathbf{S}+\mathbf{\Psi}^{\frac{R}{2}}\bar{\mathbf{X}}\mathbf{\Psi}^T\bar{\mathbf{X}}^\dagger\mathbf{\Psi}^{\frac{R}{2}}}\right\| &\leq \frac{1}{N} rank\big(\mathbf{S} + \mathbf{\Psi}^{\frac{R}{2}}\mathbf{X}\mathbf{\Psi}^T\mathbf{X}^\dagger\mathbf{\Psi}^{\frac{R}{2}} - \mathbf{S} - \mathbf{\Psi}^{\frac{R}{2}}\bar{\mathbf{X}}\mathbf{\Psi}^T\bar{\mathbf{X}}^\dagger\mathbf{\Psi}^{\frac{R}{2}}\big) \\
&= \frac{1}{N} rank\big(\mathbf{\Psi}^{\frac{R}{2}}(\mathbf{X}\mathbf{\Psi}^T\mathbf{X}^\dagger - \bar{\mathbf{X}}\mathbf{\Psi}^T\bar{\mathbf{X}}^\dagger)\mathbf{\Psi}^{\frac{R}{2}}\big) \\
&= \frac{1}{N} rank\big(\mathbf{X}\mathbf{\Psi}^T\mathbf{X}^\dagger - \bar{\mathbf{X}}\mathbf{\Psi}^T\bar{\mathbf{X}}^\dagger\big) \\
&= \frac{1}{N} rank\big((\mathbf{X}\mathbf{\Psi}^{\frac{T}{2}})(\mathbf{X}\mathbf{\Psi}^{\frac{T}{2}})^\dagger - (\bar{\mathbf{X}}\mathbf{\Psi}^{\frac{T}{2}})(\bar{\mathbf{X}}\mathbf{\Psi}^{\frac{T}{2}})^\dagger\big) \\
&= \frac{1}{N} rank\big(\mathbf{X}\mathbf{\Psi}^{\frac{T}{2}} - \bar{\mathbf{X}}\mathbf{\Psi}^{\frac{T}{2}}\big)\big(\mathbf{X}\mathbf{\Psi}^{\frac{T}{2}} - \bar{\mathbf{X}}\mathbf{\Psi}^{\frac{T}{2}}\big)^\dagger
\end{aligned}$$

For any matrix, $rank(\mathbf{A}) = rank(\mathbf{A^T})$, so $\bar{\mathbf{X}}\mathbf{\Psi}^T\mathbf{X}^\dagger - \mathbf{X}\mathbf{\Psi}^T\bar{\mathbf{X}}^\dagger$ doesn't contribute to the rank, then

$$\begin{aligned}
&\frac{1}{N} rank\big(\mathbf{X}\mathbf{\Psi}^{\frac{T}{2}} - \bar{\mathbf{X}}\mathbf{\Psi}^{\frac{T}{2}}\big)\big(\mathbf{X}\mathbf{\Psi}^{\frac{T}{2}} - \bar{\mathbf{X}}\mathbf{\Psi}^{\frac{T}{2}}\big)^\dagger \\
&\leq \frac{1}{N} rank\big(\mathbf{X}\mathbf{\Psi}^{\frac{T}{2}}\big)\big(\mathbf{X}\mathbf{\Psi}^{\frac{T}{2}} - \bar{\mathbf{X}}\mathbf{\Psi}^{\frac{T}{2}}\big)^\dagger + \frac{1}{N}\big(\bar{\mathbf{X}}\mathbf{\Psi}^{\frac{T}{2}}\big)\big(\mathbf{X}\mathbf{\Psi}^{\frac{T}{2}} - \bar{\mathbf{X}}\mathbf{\Psi}^{\frac{T}{2}}\big)^\dagger
\end{aligned}$$

It's easy to observe that $rank(\mathbf{X}) \geq rank(\mathbf{X} - \bar{\mathbf{X}})$ and so does $\bar{\mathbf{X}}$. So

$$\begin{aligned}
&\leq \frac{1}{N} rank\big(\mathbf{X}\mathbf{\Psi}^{\frac{T}{2}}\big)\big(\mathbf{X}\mathbf{\Psi}^{\frac{T}{2}} - \bar{\mathbf{X}}\mathbf{\Psi}^{\frac{T}{2}}\big)^\dagger + \frac{1}{N}\big(\bar{\mathbf{X}}\mathbf{\Psi}^{\frac{T}{2}}\big)\big(\mathbf{X}\mathbf{\Psi}^{\frac{T}{2}} - \bar{\mathbf{X}}\mathbf{\Psi}^{\frac{T}{2}}\big)^\dagger \\
&= \frac{2}{N} rank\big(\mathbf{X} - \bar{\mathbf{X}}\big)(\mathbf{\Psi}^{\frac{T}{2}}) \\
&= \frac{2}{N} rank\big(\mathbf{X} - \bar{\mathbf{X}}\big)
\end{aligned}$$

Now we consider when N goes to infinity.

$$\begin{aligned}\left\|F^{\mathbf{S}+\mathbf{\Psi}^{\frac{R}{2}}\mathbf{X}\mathbf{\Psi}^{T}\mathbf{X}^{\dagger}\mathbf{\Psi}^{\frac{R}{2}}} - F^{\mathbf{S}+\mathbf{\Psi}^{\frac{R}{2}}\bar{\mathbf{X}}\mathbf{\Psi}^{T}\bar{\mathbf{X}}^{\dagger}\mathbf{\Psi}^{\frac{R}{2}}}\right\| &\leq \frac{2}{N}rank\big(\mathbf{X}-\bar{\mathbf{X}}\big)\\ &\leq \frac{2}{N}\left|\left\{(i,j) : |X_{i,j} \geq \frac{\sqrt{n}}{2}; i\leq N, j\leq n\right\}\right|\\ &\triangleq \frac{2}{N}\xi\end{aligned}$$

We write the following according to [11]

$$\eta = P\bigg(|X_{ii}| \geq \ln N\bigg) = o\bigg(\frac{1}{(2\ln^2 N)^2}\bigg)$$

According to Hoeffding's inequality, for any $\epsilon > 0$, when N is large enough,

$$P(\left\|F^{\mathbf{S}+\mathbf{\Psi}^{\frac{R}{2}}\mathbf{X}\mathbf{\Psi}^{T}\mathbf{X}^{\dagger}\mathbf{\Psi}^{\frac{R}{2}}} - F^{\mathbf{S}+\mathbf{\Psi}^{\frac{R}{2}}\bar{\mathbf{X}}\mathbf{\Psi}^{T}\bar{\mathbf{X}}^{\dagger}\mathbf{\Psi}^{\frac{R}{2}}}\right\| \geq \epsilon) \leq P(\frac{2}{N}\xi \geq \epsilon)$$

where Hoeffding inequality,

$$\begin{aligned}P(\xi \geq (\eta+\epsilon)N) &= \exp(-2\epsilon^2 N)\\ &= P(\frac{2}{N(2\ln N)^2}\xi \geq \frac{2\eta+2\epsilon}{(2\ln N)^2})\\ &\approx P(\frac{2}{N(2\ln N)^2}\xi \geq \frac{2\epsilon}{(2\ln N)^2})\\ &= P(\frac{2}{N}\xi \geq 2\epsilon)\\ &\approx P(\frac{2}{N}\xi \geq \epsilon)\end{aligned}$$

So when N is large enough, we have $\left\|F^{\mathbf{S}+\mathbf{\Psi}^{\frac{R}{2}}\mathbf{X}\mathbf{\Psi}^{T}\mathbf{X}^{\dagger}\mathbf{\Psi}^{\frac{R}{2}}} - F^{\mathbf{S}+\mathbf{\Psi}^{\frac{R}{2}}\bar{\mathbf{X}}\mathbf{\Psi}^{T}\bar{\mathbf{X}}^{\dagger}\mathbf{\Psi}^{\frac{R}{2}}}\right\| \xrightarrow{a.s.} 0$

We first find a deterministic approximation of the Stieltjes transform of $\mathbf{B}_N$,we call it $m_N(z)$. Write $\mathbf{y}_j = (\frac{1}{\sqrt{M}})\mathbf{\Psi}^{\frac{R}{2}}\mathbf{x}_j$, where $\mathbf{x}_j$ denotes the jth column of $\mathbf{X}$, then let τ_j denote the jth diagonal element of matrix $\mathbf{\Psi}^T$. Write matrix $\mathbf{B}_N$ in the following form

$$\mathbf{B}_N = \mathbf{S} + \sum_{j=1}^{M}\tau_j\mathbf{y}_j\mathbf{y}_j^{\dagger} \tag{23}$$

Define

$$e_N = e_N(z) = \frac{1}{N}tr\mathbf{\Psi}^R(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$$

We know that $\mathbf{B}_N$ is diagonalizable, so we assume $\mathbf{B}_N = \mathbf{O}\mathbf{\Lambda}\mathbf{O}^{\dagger}$, $\mathbf{\Lambda} = diag(\lambda_1, \cdots, \lambda_N)$ is its eigenvalues. $\frac{1}{N}tr\mathbf{\Psi}^R(\mathbf{B}_N - z\mathbf{I}_N)^{-1} = \frac{1}{N}\mathbf{O}^{\dagger}\mathbf{\Psi}^R\mathbf{O}^{\dagger}(\mathbf{\Lambda} - z\mathbf{I}_N)^{-1}$. Denote $\mathbf{O}^{\dagger}\mathbf{\Psi}^R\mathbf{O}^{\dagger}$ by $\underline{\mathbf{\Psi}^R} = \{\underline{\Psi^R_{ij}}\}$, then

$$e_N(z) = \frac{1}{N}\sum_{i=1}^{N}\frac{\underline{\Psi^R_{ii}}}{\lambda_i - z}$$

We can see $e_N(z)$ is the Stieltjes transform of a measure on $\mathbb{R}^+$ and total mass is $\frac{1}{N}tr\underline{\mathbf{\Psi}^R} = \frac{1}{N}tr\mathbf{\Psi}^R$. Furthermore, define

$$\begin{aligned} p_N &= -\frac{1}{Mz}\sum_{j=1}^{M}\frac{\tau_j}{1+\beta\tau_j e_N} \\ &= \int \frac{-\tau}{z(1+\beta\tau e_N)}dF^{\mathbf{\Psi}^T}(\tau) \end{aligned}$$

Denote $\mathbf{B}_{(j)} = \mathbf{B}_N - \tau_j \mathbf{y}_j \mathbf{y}_j^\dagger$, and define $\mathbf{D} = -z\mathbf{I}_N + \mathbf{S} - zp(z)\mathbf{\Psi}^R$. Recall that $m_N(z) = \frac{1}{N}tr(\mathbf{B}_N - z\mathbf{I})$, we will use this representation in the following. Write

$$\begin{aligned} \mathbf{D}^{-1} - (\mathbf{B}_N - z\mathbf{I})^{-1} &= \mathbf{D}^{-1}\Big((\mathbf{B}_N - z\mathbf{I}) - \mathbf{D}\Big)(\mathbf{B}_N - z\mathbf{I})^{-1} \\ &= \mathbf{D}^{-1}\big(\sum_{j=1}^{M}\tau_j\mathbf{y}_j\mathbf{y}_j^\dagger + zp(z)\mathbf{\Psi}^R\big)(\mathbf{B}_N - z\mathbf{I})^{-1} \\ &= \sum_{j=1}^{M}\tau_j\mathbf{D}^{-1}\mathbf{y}_j\mathbf{y}_j^\dagger(\mathbf{B}_N - z\mathbf{I})^{-1} + zp(z)\mathbf{D}^{-1}\mathbf{\Psi}^R(\mathbf{B}_N - z\mathbf{I})^{-1} \end{aligned}$$

Using *Lemma 2*, we have

$$= \sum_{j=1}^{M}\tau_j\frac{\mathbf{D}^{-1}\mathbf{y}_j\mathbf{y}_j^\dagger(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}}{1+\tau_j\mathbf{y}_j^\dagger(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{y}_j} + zp(z)\mathbf{D}^{-1}\mathbf{\Psi}^R(\mathbf{B}_N - z\mathbf{I})^{-1}$$

Taking traces and devide by N, we know that

$$tr(\mathbf{y}\mathbf{y}^\dagger(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}) = \mathbf{y}^\dagger(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{y}$$

Then

$$\frac{1}{N}tr\left(\sum_{j=1}^{M}\tau_j\frac{\mathbf{D}^{-1}\mathbf{y}_j\mathbf{y}_j^\dagger(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}}{1+\tau_j\mathbf{y}_j^\dagger(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{y}_j}\right) = \frac{1}{NM}\sum_{j=1}^{M}\tau_j\frac{\mathbf{x}_j^\dagger\mathbf{\Psi}^{\frac{R}{2}}(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{D}^{-1}\mathbf{\Psi}^{\frac{R}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{y}_j^\dagger(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{y}_j}$$

and

$$\frac{1}{N}tr\Big(zp_N(z)\mathbf{D}^{-1}\mathbf{\Psi}^R(\mathbf{B}_N - z\mathbf{I})^{-1}\Big) = -\frac{1}{NM}\sum_{j=1}^{M}\tau_j\frac{tr\mathbf{\Psi}^R(\mathbf{B}_N - z\mathbf{I})^{-1}\mathbf{D}^{-1}}{1+\beta\tau_j e_N}$$

So we can write

$$\frac{1}{N}tr\mathbf{D}^{-1} - m_N(z) = \frac{1}{M}\sum_{j=1}^{M}\tau_j d_j = w_N^m$$

where

$$d_j = \frac{(1/N)\mathbf{x}_j^\dagger\mathbf{\Psi}^{\frac{R}{2}}(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{D}^{-1}\mathbf{\Psi}^{\frac{R}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{y}_j^\dagger(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{y}_j} - \frac{(1/N)tr\mathbf{\Psi}^R(\mathbf{B}_N - z\mathbf{I})^{-1}\mathbf{D}^{-1}}{1+\beta\tau_j e_N}$$

it was proven that as N goes to infinity, w_N^m goes to 0 almost surely, it means we can use $\frac{1}{N}tr\mathbf{D}^{-1}$ to approximate the empirical distribution of eigenvalues of $\mathbf{B}_N$. Now we need to solve e_N

$$\begin{aligned}
&\frac{1}{N}\bigg(tr\mathbf{D}^{-1}\mathbf{\Psi}^R - tr(\mathbf{B}_N - z\mathbf{I})^{-1}\mathbf{\Psi}^R\bigg) = \frac{1}{N}tr\mathbf{D}^{-1}\mathbf{\Psi}^R - e_N(z)\\
&= \frac{1}{M}\sum_{j=1}^{M}\tau_j\bigg(\frac{(1/N)\mathbf{x}_j^\dagger\mathbf{\Psi}^{\frac{R}{2}}(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{\Psi}^R\mathbf{D}^{-1}\mathbf{\Psi}^{\frac{R}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{y}_j^\dagger(\mathbf{B}_{(j)} - z\mathbf{I})^{-1}\mathbf{y}_j} - \frac{(1/N)tr\mathbf{\Psi}^R(\mathbf{B}_N - z\mathbf{I})^{-1}\mathbf{\Psi}^R\mathbf{D}^{-1}}{1+\beta\tau_j e_N}\bigg)\\
&= \frac{1}{M}\sum_{j=1}^{M}\tau_j d_j^e\\
&= w_N^e
\end{aligned}$$

It is also proven that $w_N^e \xrightarrow{a.s.} 0$ as N grows to infinity. So we can approximate $e_N(z)$ with $\frac{1}{N}tr\mathbf{D}^{-1}\mathbf{\Psi}^R$. However, $\frac{1}{N}tr\mathbf{D}^{-1}\mathbf{\Psi}^R$ still contains $e_N(z)$,

$$e_N = \frac{1}{N}tr\bigg(\mathbf{S} + \bigg[\int\frac{\tau}{1+\beta\tau e_N}dF^{\mathbf{\Psi}^T}(\tau)\bigg]\mathbf{\Psi}^R - z\mathbf{I}_N\bigg)^{-1}\mathbf{\Psi}^R$$

There exists solution to this function and the solution is unique.

We extend the form to multiple user, say, K users. Suppose number of users $K \geq 1$, we can assume here inter user interference does not exist and signals transmitted from different users are independent. Write $\mathbf{y}_{k,j} = (\frac{1}{\sqrt{M}})\mathbf{\Psi}_k^{\frac{R}{2}}\mathbf{x}_{k,j}$, with $\mathbf{x}_{k,j}$ denotes the jth column of $\mathbf{X}_k$, then let $\tau_{k,j}$ denote the jth diagonal element of matrix $\mathbf{\Psi}_k^T$. Then matrix $\mathbf{B}_N$ for multiuser can be written in the following form

$$\mathbf{B}_N = \mathbf{S} + \sum_{k=1}^{K}\sum_{j=1}^{M}\tau_{k,j}\mathbf{y}_{k,j}\mathbf{y}_{k,j}^\dagger \tag{24}$$

The definition

$$e_{N,k} = e_{N,k}(z) = \frac{1}{N}tr\mathbf{\Psi}_k^R(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$$

We still have $\frac{1}{N}tr\mathbf{\Psi}_k^R(\mathbf{B}_N - z\mathbf{I}_N)^{-1} = \frac{1}{N}\mathbf{O}^\dagger\mathbf{\Psi}_k^R\mathbf{O}^\dagger(\mathbf{\Lambda} - z\mathbf{I}_N)^{-1}$. Denote $\mathbf{O}^\dagger\mathbf{\Psi}_k^R\mathbf{O}^\dagger$ by $\underline{\mathbf{\Psi}_k^R} = \{\underline{\Psi_{k,ij}^R}\}$, then

$$e_{N,k}(z) = \frac{1}{N}\sum_{i=1}^{N}\frac{\underline{\Psi_{k,ii}^R}}{\lambda_i - z}$$

Furthermore, define

$$\begin{aligned}
p_k &= -\frac{1}{Mz}\sum_{j=1}^{M}\frac{\tau_{k,j}}{1+\beta\tau_{k,j}e_{N,k}}\\
&= \int\frac{-\tau_k}{z(1+\beta\tau_k e_{N,k})}dF^{\mathbf{\Psi}_k^T}(\tau_k)
\end{aligned}$$

Denote $\mathbf{B}_{k,(j)} = \mathbf{B}_N - \tau_{k,j}\mathbf{y}_{k,j}\mathbf{y}_{k,j}^{\dagger}$, and define $\mathbf{D} = -z\mathbf{I}_N + \mathbf{S} - \sum_{k=1}^{K} zp_k(z)\mathbf{\Psi}_k^R$. With similiar procedure, we still approximate $m_N(z)$ with $\frac{1}{N}tr\mathbf{D}^{-1}$, and approximate $e_N(z)$ with $\frac{1}{N}tr\mathbf{D}^{-1}\mathbf{\Psi}^R$, but

$$m_N(z) = \frac{1}{N}tr\left(\mathbf{S} + \sum_{k=1}^{K}\int \frac{\tau_k dF^{\mathbf{\Psi}_k^T}(\tau_k)}{1+\beta\tau_k e_k(z)}\mathbf{\Psi}_k^R - z\mathbf{I}_N\right)^{-1}$$

We have K functions $e_i(z), i \in \{1, \cdots, K\}$ form unique solutions to the K equations

$$e_i(z) = \frac{1}{N}tr\mathbf{\Psi}_i^R\left(\mathbf{S} + \sum_{k=1}^{K}\int \frac{\tau_k dF^{\mathbf{\Psi}_k^T}(\tau_k)}{1+\beta\tau_k e_k(z)}\mathbf{\Psi}_k^R - z\mathbf{I}_N\right)^{-1}$$

4.2.6 Deterministic equivalent for the Shannon transform

First we prove the interconnection between Stieltjes tansform and Shannon transform, it is proven in [14]. We use ln for log base e, and for $b > 0$, we have

$$\ln(1+b) = \int_0^1 \frac{b}{1+bt}dt \tag{25}$$

We denote the empirical distribution function of a matrix by $F(\lambda)$, then use $f(\lambda)$ as the derivative of it (we assume it is differentiable), and $\int_0^{+\infty} f(\lambda)d\lambda = 1$. The Shannon transform is defined as

$$\mathcal{V}(z) = \int_0^{+\infty} \log(1+\frac{\lambda}{z})f(\lambda)d\lambda \tag{26}$$

Take derivative on both sides,

$$\begin{aligned}
\frac{d\mathcal{V}}{dz} &= -\frac{1}{\log e}\int_0^{+\infty} \frac{\frac{\lambda}{z^2}f(\lambda)}{1+\frac{\lambda}{z}}d\lambda \\
z\frac{d\mathcal{V}}{dz} &= -\frac{1}{\log e}\int_0^{+\infty} \frac{(\lambda+z-z)f(\lambda)}{z+\lambda}d\lambda \\
&= -\frac{1}{\log e}\left(1 - z\int_0^{+\infty}\frac{f(\lambda)}{z+\lambda}d\lambda\right) \\
&= -\frac{1}{\log e}(1 - z\mathcal{S}(-z))
\end{aligned} \tag{27}$$

That is the connection between the two transforms, sometimes we omit the factor $\log e$, using (25) and (26),

$$\begin{aligned}
\mathcal{V}(z) &\approxeq \int_0^{+\infty} f(\lambda)\int_0^1 \left(\frac{\frac{\lambda}{z}}{1+\frac{\lambda}{z}}dt\right)d\lambda \\
&= \int_0^{+\infty} f(\lambda)\left(\int_0^1 \frac{\lambda}{z+\lambda t}dt\right)d\lambda
\end{aligned} \tag{28}$$

Let $t = \frac{1}{\omega}, \omega \in [0, \infty)$,

$$
\begin{aligned}
\mathcal{V}(z) &= \int_0^{+\infty} f(\lambda)\bigg(\int_0^1 \frac{\lambda}{z + \lambda\frac{1}{\omega}} d\frac{1}{\omega}\bigg) d\lambda \\
&= \int_0^{+\infty} f(\lambda)\bigg(\int_1^{\infty} \frac{\lambda}{\omega z + \lambda} \frac{d\omega}{\omega}\bigg) d\lambda \\
&\overset{\Omega=\omega z}{=} \int_0^{+\infty} f(\lambda)\bigg(\int_z^{\infty} \frac{\lambda}{\Omega + \lambda} \frac{d\Omega}{\Omega}\bigg) d\lambda \\
&= \int_0^{+\infty} f(\lambda)\bigg(\int_z^{\infty} \frac{\lambda + \Omega - \Omega}{\Omega + \lambda} \frac{d\Omega}{\Omega}\bigg) d\lambda \\
&= \int_z^{\infty} \bigg(\int_0^{+\infty} \frac{f(\lambda)}{\Omega} d\lambda - \int_0^{\infty} \frac{f(\lambda)}{\Omega + \lambda} d\lambda\bigg) d\Omega \\
&= \int_z^{\infty} (\frac{1}{\Omega} - \mathcal{S}(-\Omega)) d\Omega
\end{aligned} \tag{29}
$$

So we proved the relationship between them. With this relationship, we can find a closed form solution of channel capacity. Notice that in (16), we need the Stieltjes trnasform of $\sum_{k\in\mathcal{S}} \mathbf{H}_k \mathbf{R}_k \mathbf{H}_k^\dagger$, so here we actually need to analyse the matrix

$$
\mathbf{B}_N = \sum_{k=1}^{K} \mathbf{\Psi}_k^{\frac{R}{2}} \mathbf{X}_k (\mathbf{U}\mathbf{\Psi}_k^T \mathbf{U}^\dagger) \mathbf{X}_k^\dagger \mathbf{\Psi}_k^{\frac{R}{2}}
$$

According to [2], the Stieltjes transform of it becomes

$$
m_N(-z) = \frac{1}{N} tr\bigg(z\bigg[\mathbf{I}_N + \sum_{k=1}^{K} \delta_k(-z)\mathbf{\Psi}_k^R\bigg]\bigg)^{-1} \tag{30}
$$

where

$$
\delta_i(-z) = \frac{1}{M} tr \mathbf{\Psi}_i^T (z[\mathbf{I}_M + \beta e_i(-z)\mathbf{\Psi}_i^T])^{-1} \tag{31}
$$

$$
e_i(-z) = \frac{1}{N} tr \mathbf{\Psi}_i^R \bigg(z\bigg[\mathbf{I}_N + \sum_{k=1}^{K} \delta_k(-z)\mathbf{\Psi}_k^R\bigg]\bigg)^{-1} \tag{32}
$$

Now we just need to use the relationship between Shannon transform and Stieltjes transform to compute the capacity. Notice that

$$\begin{aligned}
\frac{1}{z} - m_N(-z) &= \frac{1}{N}tr\Bigg((z\mathbf{I})^{-1} - \bigg(z\Big[\mathbf{I}_N + \sum_{k=1}^{K}\delta_k(-z)\mathbf{\Psi}_k^R\Big]\bigg)^{-1}\Bigg) \\
&= \frac{1}{N}tr\Bigg((z\mathbf{I})^{-1}\bigg(\Big(z\Big[\mathbf{I}_N + \sum_{k=1}^{K}\delta_k(-z)\mathbf{\Psi}_k^R\Big]\Big) - (z\mathbf{I})\bigg)\bigg(z\Big[\mathbf{I}_N + \sum_{k=1}^{K}\delta_k(-z)\mathbf{\Psi}_k^R\Big]\bigg)^{-1}\Bigg) \\
&= \frac{1}{N}tr\Bigg(\frac{1}{z}\mathbf{I}^{-1}\bigg(z\sum_{k=1}^{K}\delta_k(-z)\mathbf{\Psi}_k^R\bigg)\bigg(z\Big[\mathbf{I}_N + \sum_{k=1}^{K}\delta_k(-z)\mathbf{\Psi}_k^R\Big]\bigg)^{-1}\Bigg) \\
&= \frac{1}{N}tr\bigg(\sum_{k=1}^{K}\delta_k(-z)\mathbf{\Psi}_k^R\bigg)\bigg(z\Big[\mathbf{I}_N + \sum_{k=1}^{K}\delta_k(-z)\mathbf{\Psi}_k^R\Big]\bigg)^{-1} \\
&= \sum_{k=1}^{K}\delta_k(-z)\frac{1}{N}tr\mathbf{\Psi}_k^R\bigg(z\Big[\mathbf{I}_N + \sum_{k=1}^{K}\delta_k(-z)\mathbf{\Psi}_k^R\Big]\bigg)^{-1} \\
&= \sum_{k=1}^{K}\delta_k(-z)e_k(-z)
\end{aligned}$$

And we have invertible square differentiable matrix Σ, $(\log\det(\Sigma))' = tr(\Sigma^{-1}\Sigma')$, so

$$\begin{aligned}
\frac{d}{dz}\frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k=1}^{K}\delta_k(-z)\mathbf{\Psi}_k^R\right) &= \frac{1}{N}tr\bigg(\sum_{k=1}^{K}\delta_k'(-z)\mathbf{\Psi}_k^R\bigg)\bigg(\mathbf{I}_N + \sum_{k=1}^{K}\delta_k(-z)\mathbf{\Psi}_k^R\bigg)^{-1} \\
&= \sum_{k=1}^{K}\delta_k'(-z)\frac{1}{N}tr\mathbf{\Psi}_k^R\bigg(z\Big[\mathbf{I}_N + \sum_{k=1}^{K}\delta_k(-z)\mathbf{\Psi}_k^R\Big]\bigg)^{-1} \\
&= -z\sum_{k=1}^{K}e_k(-z)\delta_k'(-z)
\end{aligned}$$

and

$$\begin{aligned}
\frac{d}{dz}\frac{1}{N}\log\det(\mathbf{I}_M + \beta e_k(-z)\mathbf{\Psi}_k^T) &= \frac{1}{N}tr(\mathbf{I}_M + \beta e_k(-z)\mathbf{\Psi}_k^T)^{-1}(\beta e_k'(-z)\mathbf{\Psi}_k^T) \\
&= \beta e_k'(-z)\frac{1}{N}tr\mathbf{\Psi}_k^T(\mathbf{I}_M + \beta e_k(-z)\mathbf{\Psi}_k^T)^{-1} \\
&= \frac{N}{M}e_k'(-z)\frac{1}{N}tr\mathbf{\Psi}_k^T(\mathbf{I}_M + \beta e_k(-z)\mathbf{\Psi}_k^T)^{-1} \\
&= -ze_k'(-z)\delta_k(-z)
\end{aligned}$$

Now we need to find the equivalence of $\sum_{k=1}^{K}\delta_k(-z)e_k(-z)$, we have

$$\frac{d}{dz}\bigg(z\sum_{k=1}^{K}\delta_k(-z)e_k(-z)\bigg) = \sum_{k=1}^{K}\delta_k(-z)e_k(-z) - z\sum_{k=1}^{K}\Big[\delta_k'(-z)e_k(-z) + \delta_k(-z)e_k'(-z)\Big]$$

$$\sum_{k=1}^{K}\delta_k(-z)e_k(-z) = z\sum_{k=1}^{K}\Big[\delta_k'(-z)e_k(-z) + \delta_k(-z)e_k'(-z)\Big] + \frac{d}{dz}\bigg(z\sum_{k=1}^{K}\delta_k(-z)e_k(-z)\bigg)$$

According to the derivation before, we can solve $\sum_{k=1}^{K} \delta_k(-z)e_k(-z)$:

$$\sum_{k=1}^{K} \delta_k(-z)e_k(-z) =$$

$$\frac{d}{dz}\left[\frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k=1}^{K}\delta_k(-z)\mathbf{\Psi}_k^R\right) + \sum_{k=1}^{K}\frac{1}{N}\log\det(\mathbf{I}_M + \beta e_k(-z)\mathbf{\Psi}_k^T) - z\sum_{k=1}^{K}\delta_k(-z)e_k(-z)\right]$$

$$= \frac{1}{z} - m_N(-z)$$

Take integral,

$$\mathcal{V}(z) = \int_z^{\infty}\left(\frac{1}{\omega} - m_N(-\omega)\right)d\omega =$$

$$\frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k=1}^{K}\delta_k(-z)\mathbf{\Psi}_k^R\right) + \sum_{k=1}^{K}\frac{1}{N}\log\det(\mathbf{I}_M + \beta e_k(-z)\mathbf{\Psi}_k^T) - z\sum_{k=1}^{K}\delta_k(-z)e_k(-z) \tag{33}$$

5 Simulations and Results

The simulation of multi-user MIMO channel capacity isn't conducted because of the difficulty of solving the function (32), here we only do simulation of empirical distribution of eigenvalues (e. d. f.) for large matrices. In order to see if the e.d.f. approximates the asymptotic distribution, we set the size of random channel matrices to be $N \times M$, and $M = 0.2 * N$, which means $\beta = 0.2$. Besides, let $N = 3000, 6000, 10000$ respectively to see the asymptotic performance of the distribution. The entries of the random matrix has i.i.d. complex Gaussian distribution with zero mean and variance $\frac{1}{N}$, but notice that when generating random numbers, we need to generate real part and complex part separately and each of them follows Gaussian distribution with zero mean and variance $\frac{1}{2N}$. At the same time, we put the asymptotic distribution (7) as a reference. Here are results of simulation:

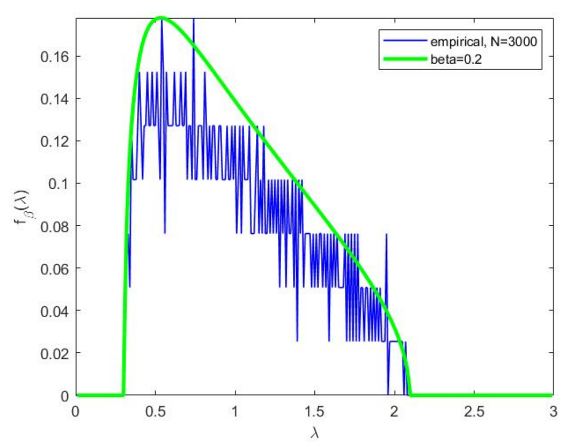

Figure 4: Empirical distribution fuction with $N = 3000$

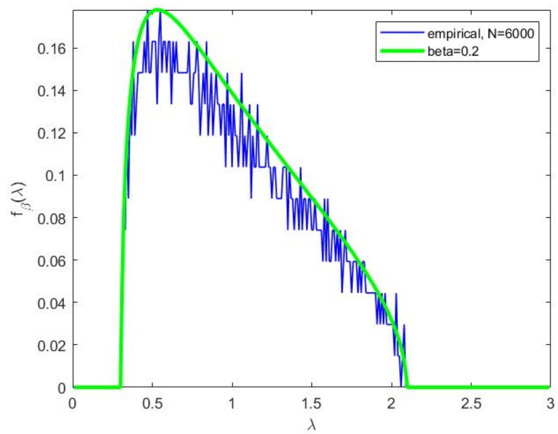

Figure 5: Empirical distribution fuction with $N = 6000$

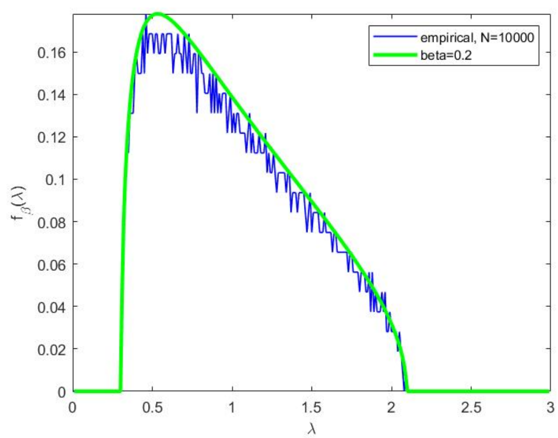

Figure 6: Empirical distribution fuction with $N = 10000$

We can see that the e.d.f. is approaching the asymptotic distribution as N grows larger, so we can approximate the real distribution using the asymptotic one when N is large, to compute the channel capacity.

The flow chart of simulation is as follows:

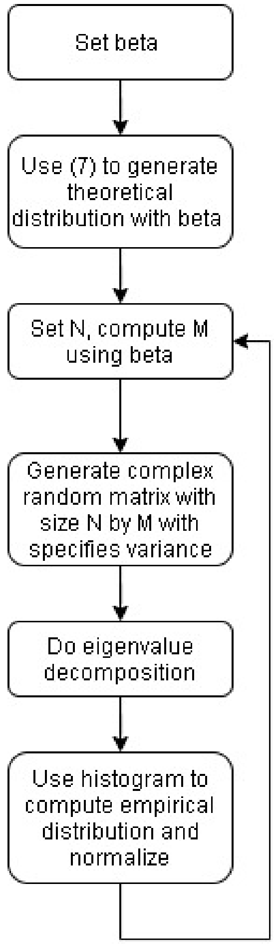

Figure 7: The flow chart of simulation of empirical distribution of eigenvalues

In order to see the performance of asymptotic capacity under different channel matrix shape, I set β to be 1.5, 1, 0.6, 0.2 and see the result

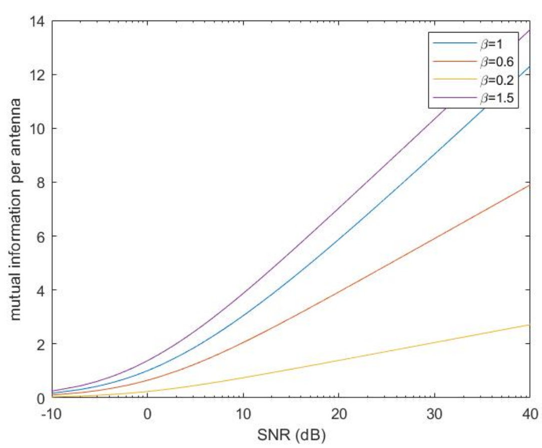

Figure 8: Mutual information per antenna of large matrix

The result is obtained using asymptotic distribution, by doing so we assume that size of channel matrix is large enough, but according to [9], with size 50×50, we can get results very close to asymptotic capacity. We fix N, and increasing β means increasing the number of antennas. It is clear that fix number of antenna, the capacity will increase if we increase antenna, and if we fix transmitting power, increase the number of antennas will also increase capacity because of antenna gain.

The flow chart is

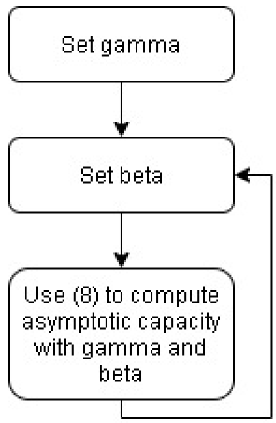

Figure 9: The flow chart of simulation of asymptotic capacity

6 Conclusion and Remarks

In this report, we discussed several channel models in MIMO communication and their analyzing method, the optimal precoder of achieving these capacity is not mentioned, more details can be found in [2] and [9]. Under independent fading, namely, Rayleigh fading, we use asymptotic method to analyse the capacity, when there is correlation between distinct communication links, we treat correlation inside transmitting antennas and receiving antennas as the major correlation, and extend the results to multiple user scenario. According to [1], under independent fading, when we use water-filling or isotropic signaling, the mutual information (if we use water-filling at low SNR, it is capacity) both grows linearly with the number of antennas, no matter SNR is low or high, but if there's correlation, the growing speed of mutual information using isotropic signaling is reduced, and growing speed of capacity when using water-filling is increased under low-SNR and decreased under high-SNR.

Appendices

Appendix A

Here we give the formal proof for the closed form solution in (8). In [6], the proof discussed situation $\beta > 1$ and $\beta < 1$, but the result of them are the same, so here we make no distinction. The proof shown below is different fron the reference, some modifications were made to make the proof more reasonable and more straight forward. Besides, for simplification, we assume high SNR scheme so the mutual information of isotropic signaling is actually channel capacity and power of every transmitting antenna is . First we denote capacity per receiving antenna by $C_0 = \frac{C}{N}$, then

$$\begin{aligned} C_0 &= \int_0^\infty \log(1+\gamma x) dF^M_{\mathbf{HH}^\dagger} \\ &= \int_{a(\beta)}^{b(\beta)} \log(1+\gamma x) f_\beta(x) dx \\ &= \int_{a(\beta)}^{b(\beta)} \log(\frac{1}{\gamma}+x) f_\beta(x) dx + \beta \log \gamma \end{aligned} \tag{34}$$

Notice that

$$\int_{a(\beta)}^{b(\beta)} \frac{\sqrt{(x-a(\beta))(b(\beta)-x)}}{2\pi x} dx = \beta$$

In order to simplify notation, we can replace $\frac{1}{\gamma}$ with α, then we got

$$C_\alpha = \int_{a(\beta)}^{b(\beta)} \log(\alpha+x) \frac{\sqrt{(x-a(\beta))(b(\beta)-x)}}{2\pi x} dx$$

We have $a(\beta) = 1+\beta-2\sqrt{\beta}$ and $a(\beta) = 1+\beta-2\sqrt{\beta}$ $b(\beta) = 1+\beta+2\sqrt{\beta}$. Then substitute x with $1+\beta-2\sqrt{\beta}\cos t$, we have $t \in [0,\pi]$. Then

$$C_\alpha = \frac{\beta}{\pi} \int_0^\pi \frac{\log(1+\beta+\alpha-2\sqrt{\beta}\cos t)}{1+\beta-2\sqrt{\beta}\cos t} \times \left(1 - \frac{e^{2jt}+e^{-2jt}}{2}\right) dt \tag{35}$$

Find w, u, $|u|<1$ that

$$1+\beta+\alpha-2\sqrt{\beta}\cos t = w(1+u^2-2u\cos t) \tag{36}$$

Then

$$\begin{aligned} \alpha+1+\beta &= w(1+u^2) \\ \sqrt{\beta} &= wu \end{aligned}$$

We have

$$u^2 - \frac{1+\beta+\alpha}{\sqrt{\beta}} u + 1 = 0 \tag{37}$$

Because $u<1$, we only have one valid solution

$$u = \frac{1}{2\sqrt{\beta}} \left(1+\beta+\alpha-\sqrt{(1+\beta+\alpha)^2-4\beta}\right) \tag{38}$$

And use this to solve w, we have

$$w = \frac{1}{2}\left(1+\beta+\alpha+\sqrt{(1+\beta+\alpha)^2}-4\beta\right)$$

Then

$$C_\alpha = \frac{\beta}{\pi}\int_0^\pi \frac{\log w + \log(1+u^2-2u\cos t)}{1+\beta-2\sqrt{\beta}\cos t} \times \left(1-\frac{e^{2jt}+e^{-2jt}}{2}\right)dt$$

Let

$$C_\alpha = I_1 + I_2$$

and

$$\begin{aligned}
I_1 &= \frac{\beta}{\pi}\int_0^\pi \frac{\log w}{1+\beta-2\sqrt{\beta}\cos t} \times \left(1-\frac{e^{2jt}+e^{-2jt}}{2}\right)dt \\
I_2 &= \frac{\beta}{\pi}\int_0^\pi \frac{\log(1+u^2-2u\cos t)}{1+\beta-2\sqrt{\beta}\cos t} \times \left(1-\frac{e^{2jt}+e^{-2jt}}{2}\right)dt
\end{aligned}$$

The first integral I_1 can be solved as follows

$$\begin{aligned}
I_1 =& \frac{\beta\log w}{\pi}\int_0^\pi \frac{1-\cos 2t}{1+\beta-2\sqrt{\beta}\cos t}dt \\
=& \frac{2\beta\log w}{2\sqrt{\beta}\pi}\int_0^\pi \frac{1-\cos^2 t}{\frac{1+\beta}{2\sqrt{\beta}}-\cos t}dt \\
\overset{a=\frac{1+\beta}{\sqrt{\beta}}}{=}& \frac{\sqrt{\beta}\log w}{\pi}\int_0^\pi \frac{1-\cos^2 t}{a-\cos t}dt \\
\because& (a-\cos t)(\frac{1}{a}+\cos t) = 1-\cos^2 t + (a-\frac{1}{a})\cos t \\
\therefore =& \frac{\sqrt{\beta}\log w}{\pi}\int_0^\pi \left[(\cos t+\frac{1}{a}) - (a-\frac{1}{a})\frac{\cos t}{a-\cos t}\right]dt \\
=& \frac{\sqrt{\beta}\log w}{\pi}\left[\frac{\pi}{a} - (a-\frac{1}{a})\int_0^\pi \frac{a}{a-\cos t}dt\right] \\
=& \frac{\sqrt{\beta}\log w}{\pi}\left[\frac{\pi}{a} - (a-\frac{1}{a})(\frac{a\pi}{\sqrt{a^2-1}}-\pi)\right] \\
=& \sqrt{\beta}\log w[a-\sqrt{a^2-1}] \\
=& \sqrt{\beta}\log w\left(\frac{1+\beta}{2\sqrt{\beta}} - \frac{1-\beta}{2\sqrt{\beta}}\right) \\
=& \beta\log w
\end{aligned} \tag{39}$$

Now we proceed to I_2, before we start, we need to approximate $\log(1+u^2-2u\cos t)$ with Taylor series:

$$\begin{aligned}
\log(1+u^2-2u\cos t) =& \log(1-ue^{jt})(1-ue^{jt}) \\
=& -\sum_{m=1}^{\infty}\frac{u^m}{m}e^{jmt} - \sum_{m=1}^{\infty}\frac{u^m}{m}e^{-jmt}
\end{aligned}$$

and

$$
\begin{aligned}
p_r(t) =& \frac{1-r^2}{1+r^2-2r\cos t}, \quad 0<r<1 \\
=& \sum_{s=-\infty}^{\infty} r^{|s|}e^{ist}
\end{aligned}
$$

Because

$$
\begin{aligned}
\sum_{s=-\infty}^{\infty} r^{|s|}e^{ist} &= \sum_{s=-\infty}^{-1} r^{-s}e^{jst} + \sum_{s=1}^{\infty} r^{s}e^{jst} + 1 \\
&= \frac{re^{-jt}}{1-re^{-jt}} + \frac{re^{jt}}{1-re^{jt}} + 1 \\
&= \frac{re^{-jt}+re^{jt}-r+1-2r\cos t+r^2}{1-2r\cos t+r^2} \\
&= \frac{1-r^2}{1+r^2-2r\cos t}
\end{aligned}
$$

simply let $\sqrt{\beta}=r$, then

$$
\begin{aligned}
I_2 =& \frac{\beta}{\pi}\int_0^{\pi} \frac{\log(1+u^2-2u\cos t)}{1+\beta-2\sqrt{\beta}\cos t} \times \left(1-\frac{e^{2jt}+e^{-2jt}}{2}\right)dt \\
=& \frac{\beta}{\pi(1-\beta)}\int_0^{\pi}\left(-\sum_{m=1}^{\infty}\frac{u^m}{m}e^{jmt}-\sum_{m=1}^{\infty}\frac{u^m}{m}e^{-jmt}\right)\left(\sum_{s=-\infty}^{\infty}\sqrt{\beta}^{|s|}e^{ist}\right)\left(1-\frac{e^{2jt}+e^{-2jt}}{2}\right)dt
\end{aligned}
\tag{40}
$$

Note that $\int_{-\pi}^{\pi} e^{-jmt}dt=0$ for $m\neq 0$, and the function inside integral is symmetric about t, in order to have the form of integral over a period, we have

$$
I_2 = \frac{\beta}{2\pi(1-\beta)}\int_{-\pi}^{\pi}\left(-\sum_{m=1}^{\infty}\frac{u^m}{m}e^{jmt}-\sum_{m=1}^{\infty}\frac{u^m}{m}e^{-jmt}\right)\left(\sum_{s=-\infty}^{\infty}\sqrt{\beta}^{|s|}e^{ist}\right)\left(1-\frac{e^{2jt}+e^{-2jt}}{2}\right)dt
$$

Then only some factors are non-zero, we take (m,s) pairs $(1,1)$, $(2,2)$, $(3,3)$, $\cdots$, $(1,-1)$, $(2,-2)$, $(3,-3)$, $\cdots$ and we get

$$
\frac{\beta}{\pi(1-\beta)}\int_{-\pi}^{\pi}\left(u\sqrt{\beta}+\frac{(u\sqrt{\beta})^2}{2}+\frac{(u\sqrt{\beta})^3}{3}+\cdots\right) = \frac{\beta}{\pi(1-\beta)}\int_{-\pi}^{\pi}\log(1-u\sqrt{\beta})dt
$$

And then choose (m,s) pairs $(1,-3)$, $(2,-4)$, $(3,-5)$, $\cdots$, $(1,-1)$, $(2,0)$, $(3,1)$, $\cdots$ and we get

$$
\begin{aligned}
&\frac{\beta}{4\pi(1-\beta)}\int_{-\pi}^{\pi}\left(-\beta(u\sqrt{\beta}+\frac{(u\sqrt{\beta})^2}{2}+\frac{(u\sqrt{\beta})^3}{3}+\cdots)-(u\sqrt{\beta}+\frac{u^2}{2}+\frac{u^3\sqrt{\beta}}{3}+\frac{u^4\sqrt{\beta}^2}{4}+\cdots)\right)dt \\
&\qquad = \frac{\beta}{4\pi(1-\beta)}\int_{-\pi}^{\pi}\left((\beta+\frac{1}{\beta})\log(1-u\sqrt{\beta})+\frac{u}{\sqrt{\beta}}-u\sqrt{\beta}\right)dt
\end{aligned}
$$

Finally, choose (m,s) pairs $(1,3)$, $(2,4)$, $(3,5)$, $\cdots$, $(1,1)$, $(2,0)$, $(3,-1)$, $\cdots$ and we get

$$\begin{aligned}\frac{\beta}{4\pi(1-\beta)}\int_{-\pi}^{\pi}\Big(-\beta\big(u\sqrt{\beta}+\frac{(u\sqrt{\beta})^2}{2}+\frac{(u\sqrt{\beta})^3}{3}+\cdots\big)-\big(u\sqrt{\beta}+\frac{u^2}{2}+\frac{u^3\sqrt{\beta}}{3}+\frac{u^4\sqrt{\beta}^2}{4}+\cdots\big)\Big)dt\\ =\frac{\beta}{4\pi(1-\beta)}\int_{-\pi}^{\pi}\Big((\beta+\frac{1}{\beta})\log(1-u\sqrt{\beta})+\frac{u}{\sqrt{\beta}}-u\sqrt{\beta}\Big)dt\end{aligned}$$

Then

$$\begin{aligned}I_2&=\frac{\beta}{2\pi(1-\beta)}\int_{-\pi}^{\pi}2\log(1-u\sqrt{\beta})-(\beta+\frac{1}{\beta})\log(1-u\sqrt{\beta})+u\sqrt{\beta}-u\frac{1}{\sqrt{\beta}}dt\\&=\frac{\beta}{1-\beta}\Big[(2-\beta-\frac{1}{\beta})(1-u\sqrt{\beta})+u\frac{\beta-1}{\sqrt{\beta}}\Big]\\&=\frac{\beta}{1-\beta}\Big[\frac{(1-\beta)^2}{\beta}\log\frac{1}{1-u\sqrt{\beta}}-u\frac{1-\beta}{\sqrt{\beta}}\Big]\\&\overset{u\sqrt{\beta}=v}{=}(1-\beta)\log\frac{1}{1-v}-v\end{aligned}\tag{41}$$

Then

$$\begin{aligned}C&=C_\alpha+\beta\log\gamma\\&=I_1+I_2+\beta\log\gamma\\&=\beta\log w+(1-\beta)\log\frac{1}{1-v}-v+\beta\log\gamma\end{aligned}\tag{42}$$

By defining

$$\begin{aligned}\mathcal{F}(\gamma,\beta)&=\left(\sqrt{\gamma(1+\sqrt{\beta})^2+1}-\sqrt{\gamma(1-\sqrt{\beta})^2+1}\right)^2\\&=2+2\gamma(1+\beta)-2\sqrt{(1+\gamma(1+\beta))^2-4\gamma^2\beta}\end{aligned}\tag{43}$$

and remember

$$w=\frac{1+\alpha+\beta+\sqrt{(1+\alpha+\beta)^2-4\beta}}{2}\tag{44}$$

$$v=\frac{1+\alpha+\beta-\sqrt{(1+\alpha+\beta)^2-4\beta}}{2}\tag{45}$$

plug (44) and (45) into (42), we have

$$
\begin{aligned}
&\beta\log\gamma w(1-v)+\log\frac{1}{1-v}-v\\
&=\beta\log\gamma\Bigg[\frac{\big(1+\alpha+\beta+\sqrt{(1+\alpha+\beta)^2-4\beta}\big)\big(1-\alpha-\beta+\sqrt{(1+\alpha+\beta)^2-4\beta}\big)}{4}\Bigg]\\
&+\log\frac{2}{1-\alpha-\beta+\sqrt{(1+\alpha+\beta)^2-4\beta}}-\frac{1+\alpha+\beta-\sqrt{(1+\alpha+\beta)^2-4\beta}}{2}\\
&=\beta\log\frac{1}{4}\Bigg[2\sqrt{(1+\gamma+\gamma\beta)^2-4\gamma^2\beta}+2(1+\gamma+\gamma\beta)-4\gamma\beta\Bigg]\\
&+\log 2\gamma\frac{\gamma-1-\gamma\beta-\sqrt{(1+\gamma+\gamma\beta)^2-4\gamma^2\beta}}{(1-\gamma+\gamma\beta)^2-(\gamma+1+\gamma\beta)^2+4\gamma^2\beta}\\
&+\frac{1+\gamma+\gamma\beta-\sqrt{(1+\gamma+\gamma\beta)^2-4\gamma^2\beta}}{2\gamma}\\
&=\beta\log\frac{1}{4}\Bigg(2+2\gamma(1+\beta)-\mathcal{F}(\gamma,\beta)+2(\gamma+1+\gamma\beta)-4\gamma\beta\Bigg)\\
&+\log\Bigg(-\frac{1}{4}\big[2\gamma-2-2\gamma\beta+\mathcal{F}(\gamma,\beta)-2-2\gamma-2\gamma\beta\big]\Bigg)-\frac{\mathcal{F}(\gamma,\beta)}{4\gamma}\\
&=\beta\log\big(1+\gamma-\frac{1}{4}\mathcal{F}(\gamma,\beta)\big)+\log\big(1+\gamma\beta-\frac{1}{4}\mathcal{F}(\gamma,\beta)\big)-\frac{\mathcal{F}(\gamma,\beta)}{4\gamma} \qquad (46)
\end{aligned}
$$

The final result is in the form of (8), but now let's consider the base of the logarism. Earlier we used Taylor series to expand I_2, we should transform the base 2 log to natural log, so instead of the equation (40), we should have it multiplied by $\log_2 e$ so that the expansion is precise. Furthermore, in (41) we should have $(1-\beta)\log\frac{1}{1-v}-\log_2 ev$ and at last, in (46) we should have the last factor $-\frac{\log_2 e\mathcal{F}(\gamma,\beta)}{4\gamma}$.

Appendix B

According to [7], we have the following useful lemmas

Lemma 1:

1) Matrices $\mathbf{A}$ and $\mathbf{B}$ in the same size

$$rank(\mathbf{A}+\mathbf{B})\leq rank(\mathbf{A})+rank(\mathbf{B})$$

2) Matrices A:$M\times N$ and B:$N\times K$

$$rank(\mathbf{AB})\leq min(rank(\mathbf{A}),rank(\mathbf{B}))$$

Lemma 2:

For $N\times N$ matrix $\mathbf{B}$, $\tau\in\mathbb{C}$, $\mathbf{r}\in\mathbb{C}^{N\times 1}$, and assume $\mathbf{B}$ and $\mathbf{B}+\tau\mathbf{r}\mathbf{r}^\dagger$ are invertible,

$$
\begin{aligned}
\mathbf{r}^\dagger\mathbf{B}^{-1}(\mathbf{B}+\tau\mathbf{r}\mathbf{r}^\dagger)&=\mathbf{r}^\dagger+\tau\mathbf{r}^\dagger\mathbf{B}^{-1}\mathbf{r}\mathbf{r}^\dagger\\
&=(1+\tau\mathbf{r}^\dagger\mathbf{B}^{-1}\mathbf{r})\mathbf{r}^\dagger\\
\frac{1}{(1+\tau\mathbf{r}^\dagger\mathbf{B}^{-1}\mathbf{r})}\mathbf{r}^\dagger\mathbf{B}^{-1}&=\mathbf{r}^\dagger(\mathbf{B}+\tau\mathbf{r}\mathbf{r}^\dagger)^{-1}
\end{aligned}
$$

Lemma 3:

From Lemma 2.6 in [7], assume $z \in \mathbb{C}^+$, with $v = Imz$, $\mathbf{A}$ and $\mathbf{B}$ $N \times N$ matrices, $\mathbf{B}$ Hermitian, $\tau \in \mathbb{R}$ and $\mathbf{r} \in \mathbb{C}^{N\times 1}$,

$$\left|tr\big((\mathbf{B}-z\mathbf{I})^{-1}-(\mathbf{B}+\tau\mathbf{r}\mathbf{r}^{\dagger}-z\mathbf{I})^{-1}\big)\mathbf{A}\right| \leq \|\mathbf{A}\|/v$$

proof.

$$\begin{aligned}
&\left|tr\big((\mathbf{B}-z\mathbf{I})^{-1}-(\mathbf{B}+\tau\mathbf{r}\mathbf{r}^{\dagger}-z\mathbf{I})^{-1}\big)\mathbf{A}\right| \\
&=\left|tr(\mathbf{B}-z\mathbf{I})^{-1}\big(\mathbf{I}-(\mathbf{B}-z\mathbf{I})(\mathbf{B}+\tau\mathbf{r}\mathbf{r}^{\dagger}-z\mathbf{I})^{-1}\big)\mathbf{A}\right| \\
&=\left|tr(\mathbf{B}-z\mathbf{I})^{-1}\big((\mathbf{B}+\tau\mathbf{r}\mathbf{r}^{\dagger}-z\mathbf{I})-(\mathbf{B}-z\mathbf{I})\big)(\mathbf{B}+\tau\mathbf{r}\mathbf{r}^{\dagger}-z\mathbf{I})^{-1}\mathbf{A}\right| \\
&=\left|tr(\mathbf{B}-z\mathbf{I})^{-1}\tau\mathbf{r}\mathbf{r}^{\dagger}(\mathbf{B}+\tau\mathbf{r}\mathbf{r}^{\dagger}-z\mathbf{I})^{-1}\mathbf{A}\right|
\end{aligned}$$

From *Lemma 2* we know that

$$\begin{aligned}
&\left|tr(\mathbf{B}-z\mathbf{I})^{-1}\tau\mathbf{r}\mathbf{r}^{\dagger}((\mathbf{B}-z\mathbf{I})+\tau\mathbf{r}\mathbf{r}^{\dagger})^{-1}\mathbf{A}\right| \\
&=\left|\frac{\tau tr(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}\mathbf{r}^{\dagger}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{A}}{1+\tau\mathbf{r}^{\dagger}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}}\right|
\end{aligned}$$

Notice that

$$tr\mathbf{r}\mathbf{r}^{\dagger}\mathbf{A} = \mathbf{r}^{\dagger}\mathbf{A}\mathbf{r}$$

Then we have

$$\left|\frac{\tau tr(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}\mathbf{r}^{\dagger}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{A}}{1+\tau\mathbf{r}^{\dagger}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}}\right| = \left|\tau\frac{\mathbf{r}^{\dagger}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{A}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}}{1+\tau\mathbf{r}^{\dagger}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}}\right|$$

If $\mathbf{A}$ is Hermitian, we can diagonalize it into unitary matrices and a diagonal matrix: $\mathbf{U\Lambda U}^{\dagger}$ According to the extension of Rayleigh's quotient, for any vector $\mathbf{x} \in \mathbb{C}^{N\times N}$we have

$$\mathbf{x}^{\dagger}\mathbf{A}\mathbf{x} = \mathbf{x}^{\dagger}\mathbf{U\Lambda U}^{\dagger}\mathbf{x} = (\mathbf{U}^{\dagger}\mathbf{x})^{\dagger}\mathbf{\Lambda}(\mathbf{U}^{\dagger}\mathbf{x}) = \mathbf{y}^{\dagger}\mathbf{\Lambda}\mathbf{y}$$

Where $\mathbf{y} = \mathbf{U}^{\dagger}\mathbf{x}$ and $\mathbf{\Lambda} = diag\{\lambda_1, \cdots, \lambda_N\}$. The eigenvalues are ranked descendent order, which means λ_1 is the biggest eigenvalue of $\mathbb{A}$. And note that in this case $\lambda_1 = \|\mathbf{A}\|$

$$\mathbf{y}^{\dagger}\mathbf{\Lambda}\mathbf{y} = \sum_{j=1}^{N}\lambda_j|y_j|^2 \leq \lambda_1\sum_{j=1}^{N}|y_j|^2 = \lambda_1\|\mathbf{y}\|^2$$

We have

$$\|\mathbf{y}\|^2 = \mathbf{y}^{\dagger}\mathbf{y} = (\mathbf{U}^{\dagger}\mathbf{x})^{\dagger}(\mathbf{U}^{\dagger}\mathbf{x}) = \mathbf{x}^{\dagger}\mathbf{U}\mathbf{U}^{\dagger}\mathbf{x} = \mathbf{x}^{\dagger}\mathbf{x} = \|\mathbf{x}\|^2$$

So

$$\mathbf{x}^{\dagger}\mathbf{A}\mathbf{x} \leq \lambda_1\|\mathbf{x}\|^2$$

In the case that $\mathbf{A}$ is not Hermitian, we can still diagonalize it with $\mathbf{P\Lambda P}^{-1}$. But $\mathbf{P}^{-1} \neq \mathbf{P}^{\dagger}$

$$\mathbf{x}^{\dagger}\mathbf{A}\mathbf{x} = \mathbf{x}^{\dagger}\mathbf{P\Lambda P}^{-1}\mathbf{x} = \mathbf{x}^{\dagger}\big(\sum_{j=1}^{N}\lambda_j \mathbf{p}_j\mathbf{q}_j\big)\mathbf{x}$$

where $\mathbf{p}_j$ is the jth column vector of $\mathbf{P}$ and $\mathbf{q}_j$ is the jth **row** vector of $\mathbf{P}^{-1}$. $\mathbf{\Lambda}$ is a diagonal matrix with entries the eigenvalues of $\mathbf{A}$. Then

$$\mathbf{x}^{\dagger}\big(\sum_{j=1}^{N}\lambda_j\mathbf{p}_j\mathbf{q}_j\big)\mathbf{x} = \sum_{j=1}^{N}\lambda_j\big(\mathbf{x}^{\dagger}\mathbf{p}_j\big)\big(\mathbf{q}_j\mathbf{x}\big) \leq \lambda_1\sum_{j=1}^{N}\big(\mathbf{x}^{\dagger}\mathbf{p}_j\big)\big(\mathbf{q}_j\mathbf{x}\big) = \lambda_1\mathbf{x}^{\dagger}\mathbf{PIP}^{-1}\mathbf{x} = \lambda_1\|\mathbf{x}\|^2$$

So

$$\left|\tau\frac{\mathbf{r}^{\dagger}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{A}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}}{1+\tau\mathbf{r}^{\dagger}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}}\right| \leq \|\mathbf{A}\||\tau|\frac{\|(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}\|^2}{|1+\tau\mathbf{r}^{\dagger}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}|}$$

write $\mathbf{B} = \sum\lambda_i^{\mathbf{B}}\mathbf{e}_i\mathbf{e}_i^{\dagger}$, $\mathbf{e}_i$'s are orthonormal eigenvectors of $\mathbf{B}$.

$$\begin{aligned}
\|(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}\|^2 &= \|\sum\frac{\mathbf{e}_i\mathbf{e}_i^{\dagger}}{\lambda_i^{\mathbf{B}}-z}\mathbf{r}\|^2 \\
&= \left(\sum\frac{\mathbf{e_i}\mathbf{e_i}^{\dagger}}{\lambda_i^{\mathbf{B}}-z}\mathbf{r}\right)^{\dagger}\left(\sum\frac{\mathbf{e_i}\mathbf{e_i}^{\dagger}}{\lambda_i^{\mathbf{B}}-z}\mathbf{r}\right) \\
&= \sum\frac{(\mathbf{e}_i^{\dagger}\mathbf{r})^*(\mathbf{e}_i^{\dagger}\mathbf{r})\mathbf{e}_i^{\dagger}\mathbf{e}_i}{(\lambda_i^{\mathbf{B}}-z)^*(\lambda_i^{\mathbf{B}}-z)} \\
&= \sum\frac{|\mathbf{e}_i^{\dagger}\mathbf{r}|^2}{|\lambda_i^{\mathbf{B}}-z|^2}
\end{aligned}$$

Assume $z = r + jv$ and we we the inequality $|z| \geq Im(z)$

$$\begin{aligned}
|1+\tau\mathbf{r}^{\dagger}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}| &= |\tau||\frac{1}{\tau}+\mathbf{r}^{\dagger}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}| \\
&\geq |\tau|Im\left[\sum\frac{|\mathbf{e}_i^{\dagger}\mathbf{r}|^2}{(\lambda_i^{\mathbf{B}}-r)-jv}\right] \\
&= |\tau|v\sum\frac{|\mathbf{e}_i^{\dagger}\mathbf{r}|^2}{|\lambda_i^{\mathbf{B}}-z|^2}
\end{aligned}$$

Therefore, we have

$$\|\mathbf{A}\||\tau|\frac{\|(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}\|^2}{|1+\tau\mathbf{r}^{\dagger}(\mathbf{B}-z\mathbf{I})^{-1}\mathbf{r}|} \leq \|\mathbf{A}\|\frac{|\tau|\sum\frac{|\mathbf{e}_i^{\dagger}\mathbf{r}|^2}{|\lambda_i^{\mathbf{B}}-z|^2}}{|\tau|v\sum\frac{|\mathbf{e}_i^{\dagger}\mathbf{r}|^2}{|\lambda_i^{\mathbf{B}}-z|^2}} = \frac{\|\mathbf{A}\|}{v}$$

Appendix C

Here is the code for empirical distribution of eigenvalues of random matrix

```
N=10000;
M=0.2*N;
beta=M/N;
x = 0:0.01:3-0.01;
a = (1 - sqrt(beta))^2;
b = (1 + sqrt(beta))^2;
xa = max(0, x - a);
xb = max(0, b - x);
fbeta = sqrt(xa .* xb) ./ (2 * pi * x);

A=normrnd(0,1/sqrt(2*N),N,M)+1i*normrnd(0,1/sqrt(2*N),N,M);
BN=A*A';
eigenv=eig(BN);
%eignon0=eigenv(eigenv>=0.000001);
%plot(x,eigenv)
edges = linspace(0, 3, 300);
histeig=hist(eigenv,x);
histeig(1)=0;
ratio=max(histeig)/max(fbeta);
histeig=histeig/ratio;

plot(x, histeig,'b', 'LineWidth', 1, 'displayname', 'empirical, N=100
ylim([0,max(histeig)])
hold on

plot(x, fbeta, 'g', 'LineWidth', 2.5, 'displayname', 'beta=0.2')
xlabel('\lambda')
ylabel('f_\beta(\lambda)')
legend
```

Here is the code for asymptotic capacity simulation

```
beta=1;
gamma=0.1:0.1:10000;
I=zeros(1,length(gamma));
for i=1:length(gamma)
I(i)=beta*log2(1+gamma(i)-(1/4)* ...
(sqrt(gamma(i)*(1+sqrt(beta))^2+1)-
sqrt(gamma(i)*(1-sqrt(beta))^2+1))^2)...
+log2(1+gamma(i)*beta-(1/4)* ...
(sqrt(gamma(i)*(1+sqrt(beta))^2+1)-
sqrt(gamma(i)*(1-sqrt(beta))^2+1))^2)...
-(1/4)* ...
(sqrt(gamma(i)*(1+sqrt(beta))^2+1)-
sqrt(gamma(i)*(1-sqrt(beta))^2+1))^2/gamma(i);
end
```

```
beta2=0.6;
I2=zeros(1,length(gamma));
for i=1:length(gamma)
I2(i)=beta2*log2(1+gamma(i)-(1/4)* ...
(sqrt(gamma(i)*(1+sqrt(beta2))^2+1)-
sqrt(gamma(i)*(1-sqrt(beta2))^2+1))^2)...
+log2(1+gamma(i)*beta2-(1/4)* ...
(sqrt(gamma(i)*(1+sqrt(beta2))^2+1)-
sqrt(gamma(i)*(1-sqrt(beta2))^2+1))^2)...
-(1/4)* ...
(sqrt(gamma(i)*(1+sqrt(beta2))^2+1)-
sqrt(gamma(i)*(1-sqrt(beta2))^2+1))^2/gamma(i);
end

beta3=0.2;
I3=zeros(1,length(gamma));
for i=1:length(gamma)
I3(i)=beta3*log2(1+gamma(i)-(1/4)* ...
(sqrt(gamma(i)*(1+sqrt(beta3))^2+1) -
sqrt(gamma(i)*(1-sqrt(beta3))^2+1))^2)...
+log2(1+gamma(i)*beta3-(1/4)* ...
(sqrt(gamma(i)*(1+sqrt(beta3))^2+1) -
sqrt(gamma(i)*(1-sqrt(beta3))^2+1))^2)...
-(1/4)* ...
(sqrt(gamma(i)*(1+sqrt(beta3))^2+1) -
sqrt(gamma(i)*(1-sqrt(beta3))^2+1))^2/gamma(i);
end

beta4=1.5;
I4=zeros(1,length(gamma));
for i=1:length(gamma)
I4(i)=beta4*log2(1+gamma(i)-(1/4)* ...
(sqrt(gamma(i)*(1+sqrt(beta4))^2+1)-
sqrt(gamma(i)*(1-sqrt(beta4))^2+1))^2)...
+log2(1+gamma(i)*beta4-(1/4)* ...
(sqrt(gamma(i)*(1+sqrt(beta4))^2+1)-
sqrt(gamma(i)*(1-sqrt(beta4))^2+1))^2)...
-(1/4)* ...
(sqrt(gamma(i)*(1+sqrt(beta4))^2+1)-
sqrt(gamma(i)*(1-sqrt(beta4))^2+1))^2/gamma(i);
end

%plot(gamma, I)
semilogx(gamma, I,'displayname','\beta=1')
```

```
hold on
semilogx(gamma, I2, 'displayname', '\beta=0.6')
hold on
semilogx(gamma, I3, 'displayname', '\beta=0.2')
hold on
semilogx(gamma, I4, 'displayname', '\beta=1.5')
legend
xticks([0.1 1 10 100 1000 10000])
xticklabels({'-10','0','10','20','30','40'})
xlabel('SNR (dB)')
ylabel('mutual information per antenna')
```

References

[1] Chen-Nee Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela. Capacity scaling in mimo wireless systems under correlated fading. *IEEE Transactions on Information Theory*, 48(3):637–650, 2002.

[2] R. Couillet, M. Debbah, and J. W. Silverstein. A deterministic equivalent for the analysis of correlated mimo multiple access channels. *IEEE Transactions on Information Theory*, 57(6):3493–3514, 2011.

[3] G. Foschini. Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. *Bell Labs Technical Journal*, 1:41–59, 1996.

[4] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of mimo channels. *IEEE Journal on Selected Areas in Communications*, 21(5):684–702, 2003.

[5] G. Lee, Y. Sung, and J. Seo. Randomly-directional beamforming in millimeter-wave multiuser miso downlink. *IEEE Transactions on Wireless Communications*, 15(2):1086–1100, 2016.

[6] P. B. Rapajic and D. Popescu. Information capacity of a random signature multiple-input multiple-output channel. *IEEE Transactions on Communications*, 48(8):1245–1248, 2000.

[7] J.W. Silverstein and Z.D. Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *Journal of Multivariate Analysis*, 54(2):175 – 192, 1995.

[8] David Tse and Pramod Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[9] Antonia M. Tulino and Sergio Verd. Random matrix theory and wireless communications. *Foundations and Trends in Communications and Information Theory*, 1(1):9–10, 2004.

[10] Wei Yu, Wonjong Rhee, S. Boyd, and J. M. Cioffi. Iterative water-filling for gaussian vector multiple-access channels. *IEEE Transactions on Information Theory*, 50(1):145–152, 2004.

[11] Y.Q. Yin. Limiting spectral distribution for a class of random matrices. *Journal of Multivariate Analysis*, 20(1):50 – 68, 1986.

[12] A. Younis, N. Abuzgaia, R. Mesleh, and H. Haas. Quadrature spatial modulation for 5g outdoor millimeterwave communications: Capacity analysis. *IEEE Transactions on Wireless Communications*, 16(5):2882–2890, 2017.

[13] D. Yuan. "wireless communication channels part 1—definition and concept". class lecture. school of information science and engineering. shandong university. qingdao. apr., 12, 2019.

[14] D. Zhang, Z. Zhou, C. Xu, Y. Zhang, J. Rodriguez, and T. Sato. Capacity analysis of noma with mmwave massive mimo systems. *IEEE Journal on Selected Areas in Communications*, 35(7):1606–1618, 2017.