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
December 9, 2020

Contents

1 Contributions **2**

2 introduction **2**

3 Literature Review **3**

4 Wireless Channel Models **5**

4.1 Single user MIMO channel under independent fading 5

4.2 Multiuser MIMO channel 8

4.2.1 MIMO channel under correlated fading 9

4.2.2 Multi-user MIMO communication with correlated fading 11

4.2.3 Stieltjes transform 11

4.2.4 Shannon transform 12

4.2.5 Deterministic equivalent for the Stieltjes transform 12

4.2.6 Deterministic equivalent for the Shannon transform 18

5 Simulations and Results **21**

6 Conclusion and Remarks **24**

Appendices **24**

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

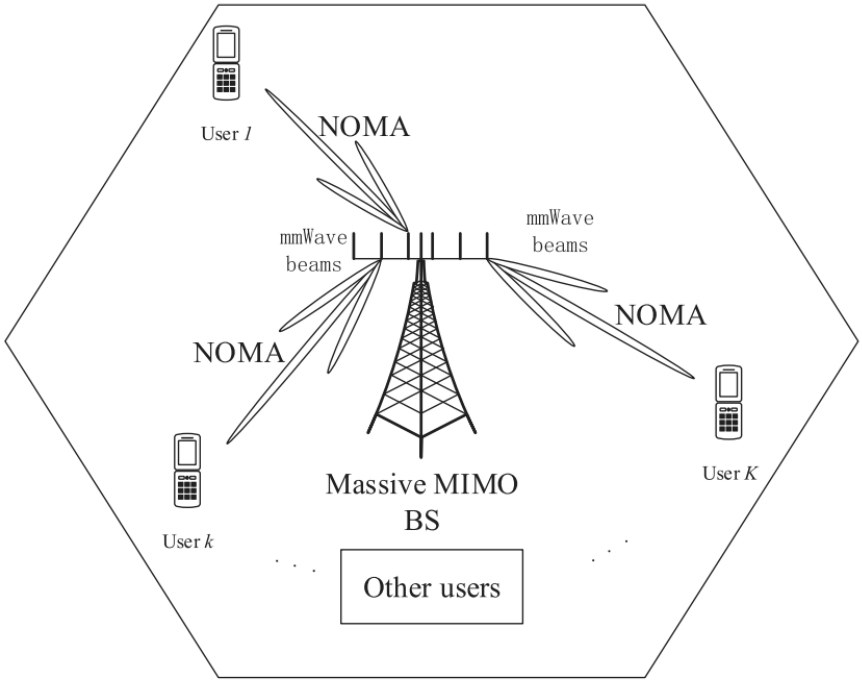

Figure 1: The proposed integrating model[14]

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

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

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 $\to inf$ 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:

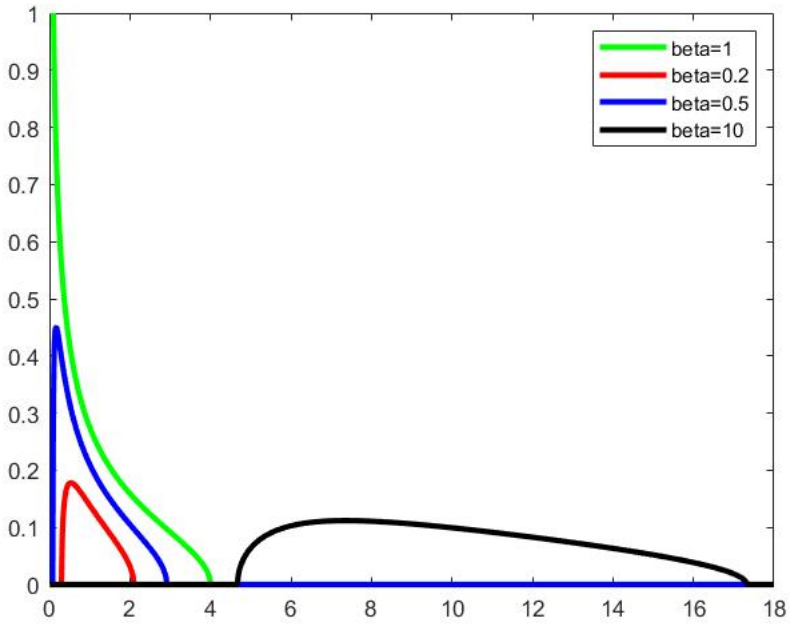

Figure 2: The limit distribution of $f_{\beta}(x)$

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.

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

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.

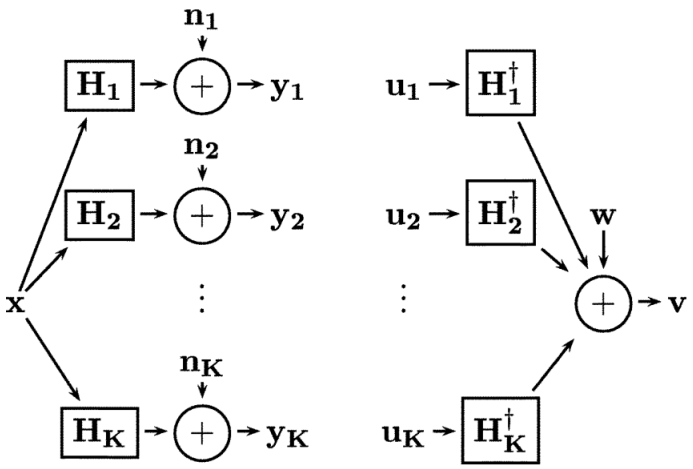

Figure 3: Models of MIMO BC (left) and MIMO MAC (right)

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_{ij}^T 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 $\frac{1}{M}tr\mathbf{R}_k \leq P_k$ 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

$$a < \min_k \liminf_N \quad c \leq \max_k \limsup_N \quad c < b \tag{19}$$

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 $\mathbf{X}'_\mathbf{k}\mathbf{U}$ 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 $\mathbf{X}_\mathbf{k}\mathbf{U}$ 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.

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

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

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

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 $\bar{\mathbf{X}} = \left(\frac{1}{\sqrt{M}}\bar{X}_{ij}\right)$.

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

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

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

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

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

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 $\left\|F^{\mathbf{S}+\mathbf{\Psi}^{\frac{R}{2}}\mathbf{X}\mathbf{\Psi}^{T}\mathbf{X}^{\dagger}\mathbf{\Psi}^{\frac{R}{2}}}-F^{\mathbf{S}+\mathbf{\Psi}^{\frac{R}{2}}\bar{\mathbf{X}}\mathbf{\Psi}^{T}\bar{\mathbf{X}}^{\dagger}\mathbf{\Psi}^{\frac{R}{2}}}\right\|\overset{a.s.}{\to}0$

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 $\underline{\mathbf{\Psi}^R}=\{\underline{\Psi_{ij}^R}\}$, then

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

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

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

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

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

It is also proven that $w_N^e \overset{a.s.}{\to} 0$ 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 $e_N(z)$,

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

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

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

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

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

and

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

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

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

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

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:

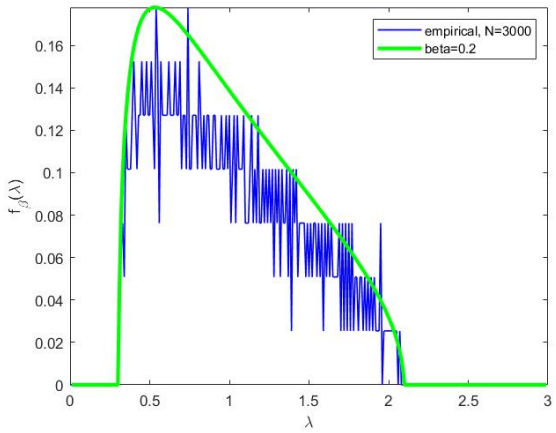

Figure 4: Empirical distribution fuction with $N = 3000$

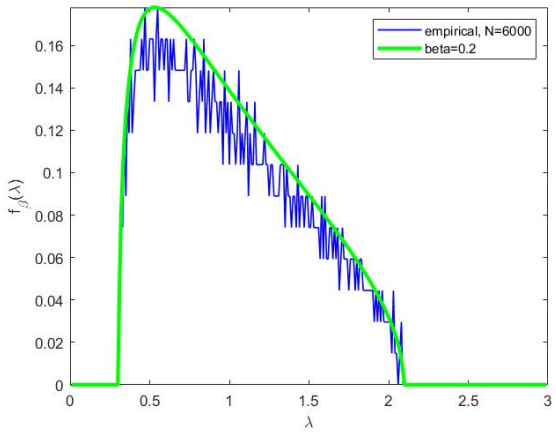

Figure 5: Empirical distribution fuction with $N = 6000$

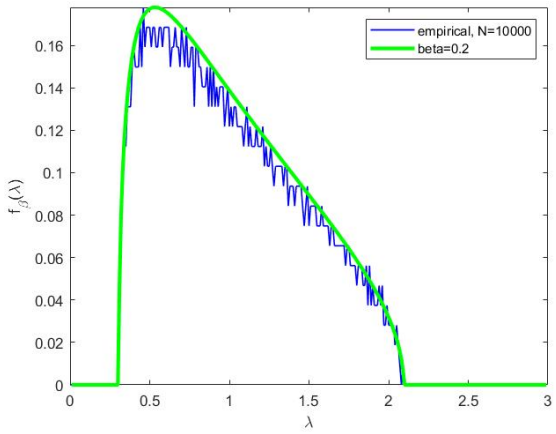

Figure 6: Empirical distribution fuction with $N = 10000$

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:

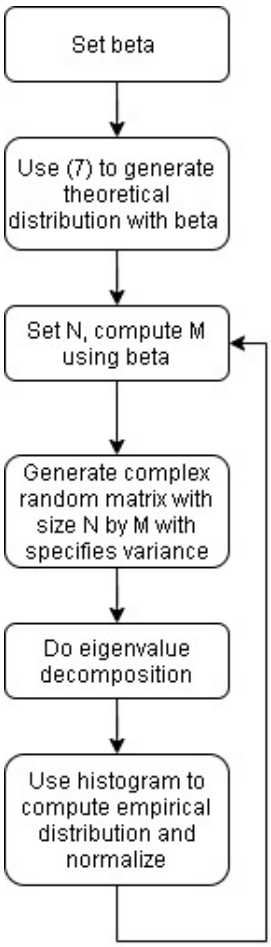

Figure 7: The flow chart of simulation of empirical distribution of eigenvalues

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

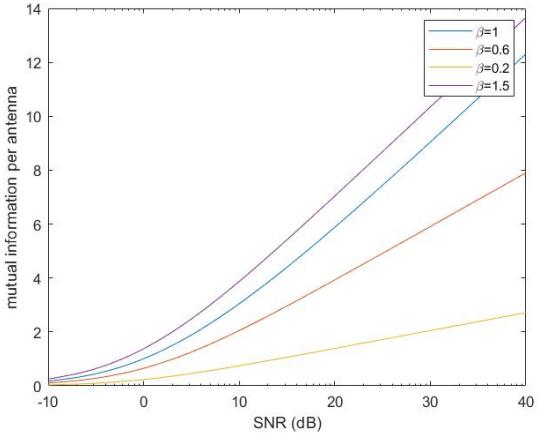

Figure 8: Mutual information per antenna of large matrix

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

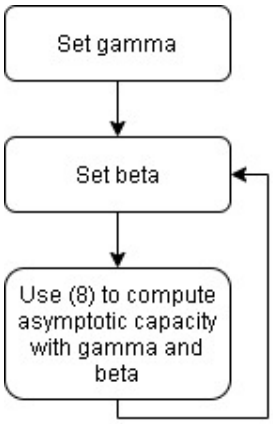

Figure 9: The flow chart of simulation of asymptotic capacity

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

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

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

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

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

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

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

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

Then

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

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

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

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: $\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\dagger$ According to the extension of Rayleigh's quotient, for any vector $\mathbf{x} \in \mathbb{C}^{N\times N}$we have

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

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 $\mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$. But $\mathbf{P}^{-1} \neq \mathbf{P}^{\dagger}$

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

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

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

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')
```

Appendix D

Capacity Scaling in MIMO Wireless Systems Under Correlated Fading

Chen-Nee Chuah, *Student Member, IEEE*, David N. C. Tse, *Member, IEEE*, Joseph M. Kahn, *Fellow, IEEE*, and Reinaldo A. Valenzuela, *Fellow, IEEE*

***Abstract*—Previous studies have shown that single-user systems employing n-element antenna arrays at both the transmitter and the receiver can achieve a capacity proportional to n, assuming *independent* Rayleigh fading between antenna pairs. In this paper, we explore the capacity of dual-antenna-array systems under *correlated fading* via theoretical analysis and ray-tracing simulations. We derive and compare expressions for the asymptotic growth rate of capacity with n antennas for both independent and correlated fading cases; the latter is derived under some assumptions about the scaling of the fading correlation structure. In both cases, the theoretic capacity growth is linear in n but the growth rate is 10–20% smaller in the presence of correlated fading. We analyze our assumption of separable transmit/receive correlations via simulations based on a ray-tracing propagation model. Results show that empirical capacities converge to the limit capacity predicted from our asymptotic theory even at moderate $n = 16$. We present results for both the cases when the transmitter does and does not know the channel realization.**

***Index Terms*—Asymptotic capacity growth, correlated fading, multiantenna arrays, multiple-input–multiple-output (MIMO) systems, ray tracing.**

I. Introduction

In response to the demand for higher bit rates in wireless local-area networks (LANs), researchers have explored the use of multiple-element arrays (MEAs) at both the transmitter and the receiver. Signals propagating through the wireless channel experience path loss and distortion due to multipath fading and additive noise. These impairments, along with the constraints of power and bandwidth, limit the system capacity. In the past, multiple antennas have been used at the receiver to combat multipath fading, e.g., using maximal-ratio combining [1], or to suppress interfering signals, e.g., using optimal combining [2]. Recent studies report that in single-user, point-to-point links, using MEAs at both transmitter and receiver increases the capacity significantly over single-antenna systems [3], [4].

Manuscript received June 13, 2000; revised June 3, 2001. The material in this paper was presented in part at IEEE GLOBECOM'98, Sydney, Australia, November 8–12, 1998.

C.-N. Chuah is with the Department of Electrical and Computer Engineering, University of California, Davis, Davis, CA 95616 USA (e-mail: chuah@ece.ucdavis.edu).

D. N. C. Tse and J. M. Kahn are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA 94720-1770 USA (e-mail: dtse@eecs.berkeley.edu; jmk@eecs.berkeley.edu).

R. A. Valenzuela is with Bell Laboratories, HOH R-255, Holmdel, NJ 07733 USA (e-mail: rav@dnrc.bell-labs.com).

Communicated by M. L. Honig, Associate Editor for Communications.

Publisher Item Identifier S 0018-9448(02)00635-1.

Foschini and Gans have analyzed the information-theoretic capacity of MEA systems in a narrow-band Rayleigh-fading environment [3]. They consider independent and identically distributed (i.i.d.) fading at different antenna elements, and assume that the transmitter does not know the channel. With n transmitting and n receiving antennas, the MEA mutual information with equal-power allocation I_n is reported to grow linearly with n for a given fixed average transmitter power. An MEA system achieves almost n more bits per hertz for every 3-dB increase in signal-to-noise ratio (SNR) at high SNR, compared to the single-antenna case, which only achieves one additional bit per hertz for every 3-dB increase in SNR.

In practice, correlation exists between the signals transmitted by or received at different antenna elements. Correlation can arise if the elements are not spaced sufficiently far apart. For example, Lee pointed out in [5] that in order to obtain a correlation coefficient at adjacent elements less than 0.7, the elements must be spaced by about 15–20 wavelengths in the broadside case and 70 wavelengths in the inline case. The presence of a dominant line-of-sight component can also affect the MEA capacities. It is important to understand the impact of these factors on MEA system capacity.

The goal of this paper is to explore the capacities of single-user MEA systems in a more realistic propagation environment, where the fading is correlated. We consider the performance in two scenarios: 1) the transmitter knows the channel, so that optimal transmit power allocation (also known as water filling) can be used; 2) the transmitter does not know the channel, so that equal power is allocated to each of the transmit antenna elements. In both cases, it is assumed that the receiver knows the channel perfectly. We study the behavior of MEA capacities through analysis and simulation.

The multiple-input–multiple-output (MIMO) fading channel is modeled as a random matrix H. The water-filling capacity C_n and the mutual information under equal power allocation I_n of a n by n system are random variables, being functions of the singular values of the random H. We construct a fading correlation structure assuming separable transmit/receive correlations, and derive the large system limiting distribution of the singular values of H in two cases: a) when the fades between different antenna pairs are independent and b) when these fades are correlated. Using these results, we show almost-sure convergence of the asymptotic growth rate C_n/n and I_n/n. In both the independent and correlated fading cases, the capacity and mutual information grow linearly with n but the growth rate is different when the fades are correlated. In particular, we show that under correlated fading, the growth rate of I_n is smaller at all SNRs

compared to the independent fading case, while the growth rate of C_n is smaller at high SNR but larger at low SNR.

Our hypothesized fading correlation structure is studied carefully via simulation based on a ray-tracing propagation model. We use the WiSE (Wireless System Engineering) [6] software tool to model explicitly the channel response between a transmitter and a receiver placed inside an office building. Comparing the empirical capacity distribution with the asymptotic theory, reasonable agreement is found even for moderate $n \leq 16$. Initial results can be found in [7] and [8]. We also quantify the capacity improvements achieved by water filling over the equal power strategy empirically at different SNR levels.

An alternative approach to ray-tracing simulations is to use scattering models [9], [12] to characterize the spatial fading correlations. In independent work, Shiu *et al*. quantify the effect of fading correlations on MEA capacity in [13] by employing such an abstract scattering model.

The remainder of this paper is organized as follows. In Section II, we model the channel as a MIMO system with flat frequency response. Using this mathematical model, we define information-theoretic capacity and mutual information of MEA systems in Section III, and analyze their asymptotic behavior as $n \to \infty$ in Section IV. In Section V, we present capacity estimates for the simulated channels and discuss the discrepancies between these results and the asymptotic capacities predicted by theory. We briefly describe how WiSE is used to model the indoor propagation environment that our numerical analysis is based on. We also include details about placements of transmitting and receiving MEAs, arrangement of antennas in an array, and basic assumptions about the antenna elements. Conclusions are presented in Section VI.

To simplify notations, we will focus exclusively on the case when the number of transmit antenna is equal to the number of receive antenna (n by n systems). The extension of the analytical results to the case with unequal number of transmit and receive antennas is straightforward.

II. Channel Model

The following notation is used throughout the paper: $'$ for vector transpose, $\dagger$ for transpose conjugate, $I_{n\times n}$ for the identity matrix, $E[\cdot]$ for expectation, and underline for vectors. All logarithms are with respect to base 2.

We consider a single-user,[1] point-to-point communication channel with n transmitting and n receiving antenna elements, denoted as an (n, n)-MEA system. We assume that the transmitted signal occupies a bandwidth W, over which the channel frequency response is essentially constant. For this assumption to be valid, W must be much smaller than the channel coherence bandwidth, which is approximately the reciprocal of the channel delay spread.[2] Since the maximum delay spread of our channels is about 25 ns, we require that W be much less than 40 MHz. Assuming zero excess bandwidth, this requires a symbol rate much less than 40 Mbaud.

[1] n transmitting antennas are colocated, and so are the receiving antennas.

[2] Here, delay spread refers to the difference in arrival times of the earliest and latest strong rays.

For the remaining analysis and discussions, we assume that the channel is linear and time-invariant and use the following discrete-time equivalent model:

$$\underline{Y} = H\underline{X} + \underline{Z}. \tag{1}$$

$\underline{X} = [x_1, x_2, \ldots, x_T]'$ is an $n \times 1$ vector whose jth component represents the signal transmitted by the jth antenna. Similarly, the received signal and received noise are represented by $n \times 1$ vectors, $\underline{Y}$ and $\underline{Z}$, respectively, where y_i and z_i represent the signal and noise received at the ith antenna. The complex path gain between transmitter j and receiver i is represented by $\{H_{ij}\colon i, j = 1, 2, \ldots, n\}$.

We further assume the following.

- The total average power (sum over all transmitting antennas) is P_{tot}, regardless of n.
- The noise vector $\underline{Z}$ is an additive white complex Gaussian random vector, whose entries $\{Z_i, i = 1, 2, \ldots, n\}$are i.i.d. circularly symmetric complex Gaussian random variables with variance

$$E[|Z|^2] = N_{\text{o}}W$$

  where W is the signal bandwidth.

We consider the following two cases.

1) H is known only to the receiver but not the transmitter. Power is distributed equally over all transmitting antennas in this case.
2) H is known at the transmitter and receiver, so that power allocation can be optimized to maximize the achievable rate over the channel.

In this work, we treat H as quasi-static. H is considered fixed for the whole duration of communication, so that the capacity is computed for each realization of H without time averaging. On the other hand, H changes if the receiver is moved from one place to the other, and we assume this will happen over a time scale much longer than the duration of communication. The associated capacity and mutual information C_n and I_n for each specific realization of H can be viewed as random variables. We are interested in studying the statistics of these random variables, in particular, the averages, $\overline{C}_n$ and $\overline{I}_n$ and the values at 5% channel outage, $C_n^{0.05}$ and $I_n^{0.05}$.

III. MEA Capacity and Mutual Information

Channel capacity is defined as the highest rate at which information can be sent with arbitrarily low probability of error. Since the channel H is considered quasi-static, it is reasonable to associate the capacity to a specific realization of H, given a fixed average total power P_{tot} and noise variance $N_{\text{o}}W$ (see Section II for channel model and assumptions). Throughout our analysis, we assume that $\{H_{ij}\colon i, j = 1, 2, \ldots, n\}$ are identically distributed with the variance normalized to be 1. Therefore, the average received SNR is defined as

$$\rho = \frac{P_{\text{tot}}}{N_{\text{o}}W}.$$

When n antennas are used, we denote the MEA capacity with water filling and mutual information with equal-power allocation as C_n and I_n, respectively. For the case with $n = 1$, the Shannon capacity is

$$C_1 = I_1 = \log\left(1 + \rho|H_{11}|^2\right) \text{ b/s/Hz.} \tag{2}$$

In the high-SNR regime, each 3-dB increase of ρ yields a capacity increase of 1 b/s/Hz.

A. Capacity With Water-Filling Power Allocation

In this subsection, we derive the MEA capacity C_n assuming the transmitter has perfect knowledge about the channel. With this knowledge of the channel, the total transmit power can be allocated in the most efficient way over the different transmitters to achieve the highest possible bit rate. Based on the model in Section II and definitions in [14], the MEA capacity with optimal power allocation is

$$C_n = \max_{Q} \log\det\left[I_{n\times n} + \rho HQH^{\dagger}\right] \text{ b/s/Hz} \tag{3}$$

where Q is the $n \times n$ covariance matrix of $\underline{X}$ and Q must satisfy the average power constraint

$$tr(Q) = \sum_{i=1}^{n} E\left[|x_i|^2\right] \le P_{\text{tot}}. \tag{4}$$

The optimal solution is

$$C_n = \sum_{i=1}^{n} \log(\lambda_i \mu)^{+} \tag{5}$$

where μ satisfies

$$\sum_{i} \left(\mu - \frac{1}{\lambda_i}\right)^{+} = \rho \tag{6}$$

and the λ_i are the eigenvalues of $HH^{\dagger}$.

The optimal solutions given in (5) and (6) are analogous to the optimal power allocation calculated through the water-filling algorithm for parallel Gaussian channels [14]. Intuitively, (5) and (6) suggest that the original MIMO channel can be decomposed into n parallel independent subchannels, and we allocate more power to the subchannels with higher SNR $\rho\lambda_i$. Here, μ is the "water level" that marks the height of the power that is poured into the "water vessel" formed by the function $\{1/\lambda_i, i = 1, 2, \ldots, n\}$. Each of these subchannels contributes to the total capacity through $\log 2(\lambda_i\mu)^{+}$. If $\lambda_i\mu \gg 1$, we say that this subchannel provides an effective mode of transmission and is called a *strong eigenmode*.

B. Mutual Information With Equal-Power Allocation

In this case, we assume that equal power is radiated from each transmitting antenna, which is a natural thing to do when the transmitter does not know the channel. The mutual information of (n, n)-MEAs with equal-power allocation is

$$I_n = \log\det\left[I_{n\times n} + \frac{\rho}{n} HH^{\dagger}\right] \text{ b/s/Hz.} \tag{7}$$

Applying singular-value decomposition to H, we can write (7) as

$$I_n = \sum_{i=1}^{n} \log\left(1 + \frac{\rho\lambda_i}{n}\right).$$

IV. Asymptotic Analysis

The capacity C_n and mutual information I_n depend on H, which is random in a fading environment. We analyze the asymptotic behavior of I_n and C_n as $n \to \infty$ for two cases: a) when the H_{ij} are independent, and b) when the H_{ij} are correlated. Our analysis is based on the channel model and properties described in Sections II and III. In all cases, we normalize $E[|H_{ij}|^2] = 1$ for all i, j. For clarity, let us use $I_n(H)$ and $C_n(H)$ to explicitly denote the dependency on H.

A. Independent Fading

We first assume that the path gains H_{ij} are i.i.d. for all i and j. We scale up the size of the MEA by letting n grow large. For each n, let F_n be the empirical distribution of the eigenvalues of $HH^{\dagger}$, i.e., for each λ

$$F_n(\lambda) := \frac{1}{n}\left|\left\{i\colon \lambda_t^{(i)} \le \lambda\right\}\right|$$

the fraction of squared singular values of H less than or equal to λ. Note that since the singular values are random, so is the empirical distribution. An important observation is that, from the expressions (5)–(7), both the capacity and the mutual information under equal-power allocation depends on H only through the empirical distribution of the eigenvalues. The asymptotic properties of the random variables $C_n(H)$ and $I_n(H)$ hinge on how the (random) empirical distribution of the singular values behaves as $n \to \infty$. We have the following theorem (see, e.g., [15]).

Theorem IV.1: Define $G_n(\lambda) := F_n(n\lambda)$. Then almost surely, G_n converges in distribution to a limit G^*, which has a density given by

$$g^*(\lambda) = \begin{cases} \frac{1}{\pi}\sqrt{\frac{1}{\lambda} - \frac{1}{4}}, & 0 \le \lambda \le 4 \\ 0, & \text{else.} \end{cases} \tag{8}$$

Moreover, if $\lambda_{\max}(HH^{\dagger})$ is the largest eigenvalue of $HH^{\dagger}$, then almost surely

$$\lim_{n\to\infty} \frac{\lambda_{\max}(HH^{\dagger})}{n} = 4.$$

This result says several interesting things. First, the scaling by n in the definition of G_n means that the eigenvalues are growing of the order of n. After rescaling, the random distribution converges to a *deterministic* limiting distribution, i.e., for large n, the empirical distribution of the eigenvalues looks similar for almost all realizations of H. Moreover, the limit does not depend on the distribution of the entries H_{ij}.

The asymptotic behavior of the mutual information $I_n(H)$ follows directly from this proposition

$$\frac{I_n(H)}{n} = \frac{1}{n}\sum_{i=1}^{n} \log\left(1+\frac{\rho}{n}\lambda_i\right) \to \int_0^4 \log(1+\rho\lambda) g^*(\lambda)\, d\lambda$$

where the convergence is almost surely. This observation was previously made by Foschini [3]. The integral can actually be computed in closed form, as was done in [16] in the context of a related capacity analysis problem for randomly spread code-division multiple-access (CDMA) systems.

$$\begin{aligned} I^*(\rho) &\equiv \int_0^4 \log(1+\rho\lambda) g^*(\lambda)\, d\lambda \\ &= 2\log\left(1+\sqrt{4\rho+1}\right) - \frac{\log e}{4\rho}\left(\sqrt{4\rho+1}-1\right)^2. \end{aligned} \tag{9}$$

We now turn to the water-filling capacity. By relabeling the parameter μ as μ_n/n we can rewrite (5) and (6) as

$$\frac{C_n(H)}{n} = \frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{\lambda_i}{n}\mu_n\right)^+$$

where μ_n satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mu_n - \frac{n}{\lambda_i}\right)^+ = \rho. \tag{10}$$

As $n \to \infty$, the empirical distribution of λ_i/n converges almost surely to a limit with density g^*. From (10), we see that μ_n converges to μ^* satisfying the equation

$$\int_0^4 \left(\mu^* - \frac{1}{\lambda}\right)^+ g^*(\lambda)\, d\lambda = \rho$$

and $C_n(H)/n$ converges almost surely to

$$C^*(\rho) \equiv \int_0^4 \log(\lambda\mu^*)^+ g^*(\lambda)\, d\lambda. \tag{11}$$

Thus, when both the transmitter and the receiver have perfect knowledge of the fading channel, the capacity scales like $nC^*(\rho)$, where $C^*(\rho)$ can be interpreted as the capacity of a fading channel with fading distribution g^* when water filling over the fading state is performed [17]. Similarly, when only the receiver has knowledge of the channel and the transmitter allocates an equal amount of power to each transmit antenna, the achievable mutual information scales like $nI^*(\rho)$, where $I^*(\rho)$ can be interpreted as the mutual information achieved by using constant transmit power in a fading channel with the gain distributed as g^*. We conclude that both $C_n(H)$ and $I_n(H)$ scale linearly with n but the rate of growth is larger for C_n than is for I_n. Moreover, if we let C_n^ϵ be the ϵ-outage capacity, i.e., such that

$$P(C_n(H) \le C_n^\epsilon) = \epsilon$$

then the above results implies that

$$\lim_{n\to\infty} \frac{C_n^\epsilon}{n} = C^*(\rho)$$

for all $\epsilon > 0$. This is because almost-sure convergence implies convergence in probability. This means that for large n, the capacity becomes insensitive to the realization of H. Similar comments apply to the scaling of $I_n(H)$.

We now compare $I^*(\rho)$ and $C^*(\rho)$ in both the low- and high-SNR regimes.

As a first-order approximation, at low SNR

$$\begin{aligned} I^*(\rho) &= \int_0^4 \log(1+\rho\lambda)) g^*(\lambda)\, d\lambda \\ &\approx \rho \int_0^4 \lambda g^*(\lambda)\, d\lambda \\ &= \rho. \end{aligned} \tag{12}$$

We observe that at low SNR, $I^*(\rho)$ depends only on the average SNR and not on the eigenvalue distribution g^*.

For the capacity, we calculate

$$\frac{dC^*(\rho)}{d\rho} = \frac{1}{\mu^*} \tag{13}$$

where μ^* is the water-filling level. As ρ approaches 0, μ^* approaches $\frac{1}{4}$. To first order, at low SNR

$$C^*(\rho) \approx 4\rho$$

and we conclude that

$$\lim_{\rho\to 0} \frac{C^*(\rho)}{I^*(\rho)} = 4.$$

Hence, the water-filling strategy affords a significant performance gain over the constant-power strategy at low SNR. The intuition is that when there is little transmit power, it is much more effective to expend it on the strongest eigenmode of the system (with gain 4) rather than spread the power evenly across all modes.

Next we consider the high-SNR regime. Using the explicit expression (9), we see that as $\rho \to \infty$

$$I^*(\rho) = \log(\rho/e) + o(1)$$

a result already noted in [3].

At high SNR, it is well known that the water-filling and the constant power strategies yield almost the same performance

$$\lim_{\rho\to\infty} [C^*(\rho) - I^*(\rho)] = 0$$

and hence $C^*(\rho)$ has the same high-SNR approximation of $\log(\rho/e)$.

Although water filling does not always give significant capacity improvements over the equal-power strategy, the perfect channel knowledge at the transmitter often leads to easier and more reliable implementations of the receiver, since the receiver can now be dealing with decoupled channels instead of having to perform cancellation and nulling.

B. Correlated Fading

1) Correlation Model: In the previous subsection, we assumed that the fades between different antenna pairs are independent of each other. We will now consider the situation when

the fading between antenna pairs is correlated. While the results we obtained for the independent fading case holds for *any distribution* of the individual H_{ij}, the results we present here for the correlated case are only for the case of a Rayleigh-fading model. Each of the H_{ij} are assumed to be complex, zero-mean, circular symmetric Gaussian random variables with variance $E[|H_{ij}|^2] = 1$. The H_{ij} are jointly Gaussian with the following covariance structure:

$$E[H_{pj}H_{qk}^*] = \Psi_{jk}^T \Psi_{pq}^R$$

where Ψ^T and Ψ^R are n by n covariance matrices. This fading model embodies three assumptions.

- The correlation between the fading from transmit antennas p and q to the same receive antenna is Ψ_{pq}^T and does not depend on the receive antenna; Ψ^T describes the *transmit correlation*.
- The correlation between the fading from a transmit antenna to receive antenna j and to receive antenna k is Ψ_{jk}^R and does not depend on the transmit antenna; Ψ^R describes the *receive correlation.*
- The correlation between the fading of two distinct antenna pairs is the product of the corresponding transmit correlation and receive correlation.

The first two assumptions are usually quite accurate when antenna elements are colocated in the same physical unit at the transmitter and also at the receiver. The product-form assumption is made for analytical tractability and can be thought of as a first-order approximation of the correlation structure when the fading from two transmit antennas to the same receive antenna and the fading from two receive antennas to the same transmit antenna is much more highly correlated than that between two distinct antenna pairs. This product form assumption is studied through simulations in Section V.

To consider the scaling of capacity and mutual information with the number of antennas, we need to make further assumptions on the covariance matrices Ψ^R and Ψ^T as the system scales. In particular, we assume that the empirical eigenvalue distributions of Ψ^R and Ψ^T converge in distribution to some limiting distributions F_R and F_T, respectively. This will be true if

1) the correlation between the fading at two antennas depends only on the relative and not absolute positions of the antennas;
2) the antennas are arranged in some regular arrays, such as square or linear grids, and as we scale up the number of antennas, the relative position of adjacent antennas are fixed; and
3) the correlation decays sufficiently fast over space.

For example, if the antennas are arranged in a linear array, Ψ^R and Ψ^T are Toeplitz. If the power spectral densities of the stationary processes

$$\{H_{j1}\colon j = 1, 2, \ldots\} \quad \text{and} \quad \{H_{1k}\colon k = 1, 2, \ldots\}$$

exist at all frequencies, then the limiting eigenvalue distributions F_T and F_R of Ψ^T and Ψ^R exist. For a given $x \geq 0$, $F_T(x)$ is the fraction of frequencies in the power spectral density of $\{H_{1k}\colon k = 1, 2, \ldots\}$ with power less than or equal to x. Motivated by this example, we will in general define

$$S_R(\omega) := F_R^{-1}(\omega), \quad S_T(\omega) := F_T^{-1}(\omega), \qquad \omega \in [0, 1].$$

S_R and S_T defined in this way is always nondecreasing from 0 to 1. One can think of S_R and S_T as power spectral densities except that the frequencies are reordered such that they are always nondecreasing functions of ω. In the results to be presented, the ordering is immaterial and only the distribution of powers is relevant. We also observe that

$$\int_0^1 S_R(\omega)\, d\omega = \int_0^1 S_R(\omega)\, d\omega = 1$$

because $E[|H_{ij}|^2] = 1$ for all i, j.

It should be noted that the power spectral densities of some fading correlation models may not exist at all frequencies. An example is Jakes' model [1], with the "U-shaped" power spectral density which is bounded over only a finite interval. The reason is that the autocorrelation function decays slowly as a function of distance r, like $\frac{1}{r}$. The results below do not apply to such models.

2) Analysis: The starting point of the analysis is that H can be factorized in the form $H = (\Psi^R)^{\frac{1}{2}} W (\Psi^T)^{\frac{1}{2}}$, where the entries of W are i.i.d. complex circular symmetric Gaussian with mean 0 and variance 1. Hence,

$$HH^\dagger = (\Psi^R)^{\frac{1}{2}} W \Psi^T W^\dagger (\Psi^R)^{\frac{1}{2}}.$$

For capacity analysis, we are interested in the eigenvalue distribution of $HH^\dagger$, or equivalently $\Psi^R W \Psi^T W^\dagger$. Now W is isotropic, i.e., UW and WU have the same distribution as W for any deterministic unitary matrix U. We can factorize $\Psi^T = U D_T U^\dagger$ and $\Psi^R = V D_R V^\dagger$, where U and V are unitary and D_T, D_R are diagonal. The fact that W is isotropic allows us to conclude that the matrix $\Psi^R W \Psi^T W^\dagger$ has the same eigenvalue distribution as $D_R W D_T W^\dagger$. It should be noted that as $n \to \infty$, the eigenvalue distributions of D_R and D_T converge to F_R and F_T, respectively.

Theorem IV.1 tells us that the distribution of the eigenvalues of $WW^\dagger$, scaled by $1/n$, converges for large n. It turns out that the eigenvalue distribution of $D_R W D_T W^\dagger$ converges as well under the same scaling. However, in this case, no explicit expression for the limiting distribution is available. Instead, it is given in terms of its *Steltjes' transform* [18]. The Steltjes' transform of a distribution G is defined by

$$m_G(z) := \int \frac{1}{\lambda - z}\, dG(\lambda)$$

for $z \in \mathcal{C}$ with $\Im(z) > 0$. It can be shown by an inversion theorem that the Steltjes' transform uniquely specifies a distribution. The following result yields a characterization of the limiting eigenvalue distribution of $D_T W D_R W^\dagger$.

Theorem IV.2: Let $\tilde{F}_n$ be the empirical eigenvalue distribution of $D_T W D_R W^\dagger$. Define $\tilde{G}_n(\lambda) := \tilde{F}_n(n\lambda)$. Then almost

surely, $\tilde{G}_n$ converges in distribution to a limit G^o, whose Steltjes' transform is given by

$$m_{G^o}(z) = \int_0^1 u(\theta, z)\, d\theta$$

where $u(\theta, z)$ is the unique solution to the functional fixed-point equation

$$u(\theta, z) = \frac{1}{-z + S_T(\theta)\int_0^1 \frac{S_R(\omega)}{1+S_R(\omega)\int_0^1 u(\phi, z)S_T(\phi)\, d\phi}\, d\omega}.$$

The proof of this result, which is based on random matrix results in Girko [18], is given in Appendix II.

Using this result, it can now be shown, exactly as in the independent fading case, that almost surely as $n \to \infty$

$$\frac{I_n(H)}{n} \to I^o(S_R, S_T, \rho)$$
$$\frac{C_n(H)}{n} \to C^o(S_R, S_T, \rho)$$

where

$$I^o(S_R, S_T, \rho) = \int_0^\infty \log(1+\rho\lambda)\, dG^o(\lambda) \qquad (14)$$

and

$$C^o(S_R, S_T, \rho) = \int_0^\infty \log(\lambda\mu^*)^+\, dG^o(\lambda) \qquad (15)$$

with μ^* satisfying

$$\int_0^\infty \left(\mu^* - \frac{1}{\lambda}\right)^+ dG^o(\lambda) = \rho.$$

The important conclusion is that even with correlation, the capacity and mutual information still scale linearly with n. However, the rate of growth is different from the independent fading case. It should be emphasized that this conclusion is valid only under the specific scaling assumptions we made.

The constants C^o and I^o depend on the limiting distribution G^o, which is only indirectly characterized via its Steltjes' transform in Theorem IV.2. The following result gives a more direct characterization of the constant I^o without involving the Steltjes' transform of G^o.

Theorem IV.3:

$$I^o(S_R, S_T, \rho) = \int_0^1 \log[1 + S_T(x)\beta(x)]\, d = x$$

and for each $x \in [0, 1]$, $\beta(x)$ is the unique solution β to the fixed-point equation

$$\beta = \int_0^1 \frac{S_R(\omega)}{\frac{1}{\rho} + S_R(\omega)\int_{1-x}^1 \frac{S_T(\phi)}{1+S_T(\phi)\beta}\, d\phi}\, d\omega \equiv f_c(\beta). \qquad (16)$$

For the special case of no correlation at the receiver (i.e., $S_R(\omega) = 1$ for all $\omega \in [0, 1]$), the fixed-point equation that β must satisfy is simplified to

$$\beta = \frac{1}{\frac{1}{\rho} + \int_{1-x}^1 \frac{S_T(\phi)}{1+S_T(\phi)\beta}\, d\phi} \equiv f_{\text{ind}}(\beta). \qquad (17)$$

The proof of this result exploits the fact that the mutual information I_n can be achieved by a combination of successive decoding and linear minimum mean-square error (MMSE) demodulation. A sketch of this proof can be found in Appendix III.

Does correlation always reduce capacity? Let us fix the transmit correlation S_T and compare the performance when there is correlation at the receiver and when there is none. Since the function $h(y) = y/(a + by)$ is concave for $a, b > 0$, it follows from Jensen's inequality that in Theorem IV.3

$$f_c(\beta) \le f_{\text{ind}}(\beta)$$

for all $\beta > 0$. Hence, for a given x, if $\beta_c = f_c(\beta_c)$ and $\beta_{\text{ind}} = f_{\text{ind}}(\beta_{\text{ind}})$, then $\beta_c \le f_{\text{ind}}(\beta_c)$. By monotonicity of the fixed-point equation $\beta = f_{\text{ind}}(\beta)$, this implies $\beta_c < \beta_{\text{ind}}$. Hence, correlation at the receiver always decreases I^o. By the reciprocity property (see Appendix I), it can be seen that correlation at the transmit antenna always reduces I^o for a fixed receive correlation S_R.

A more general statement can be made to compare the performance under two different power spectra. A nondecreasing spectrum S_1 is defined to be *more spread out* than a nondecreasing S_2 if

$$\int_0^1 S_1(\omega)\, d\omega = \int_0^1 S_2(\omega)\, d\omega = 1$$

and for every $\theta \in [0, 1]$

$$\int_\theta^1 S_1(\omega)\, d\omega \ge \int_\theta^1 S_2(\omega)\, d\omega.$$

Note that the flat spectrum corresponding to independent fading is the least spread out according to this definition: there is the same amount of power at all frequencies. This notion of "spreading out" (also called *majorization* [19]) can be taken as a measure of the strength of correlation: the more spread out the spectrum, the stronger the correlation.

In the theory of majorization, a real-valued function H is said to be *Schur-concave* (resp., Schur-convex) if S_1 is more (resp., less) spread out than S_2 implies that $H(S_1(\cdot)) < H(S_2(\cdot))$. A basic result says that the function:

$$H(S_1(\cdot)) = \int_0^1 h[S_1(\omega)]d\omega$$

is Schur-concave (resp., Schur-convex) if the function h is concave (resp., convex). Applying this result to our problem, it follows that the right-hand side of (16) is a Schur-concave function of S_T. It then follows that I^o is, in fact, a Schur-concave function of S_T, i.e., stronger correlation always decreases I^o.

The above discussion focuses on the effect of fading correlation on the mutual information I^o. But, in fact, something more basic is going on. It is shown in Appendix IV that the more spread out are the power spectra S_T and S_R, the more spread out is G^o, the limiting spectrum of $D_R W D_T W^\dagger$. The mutual information I^o (14), being a Schur-concave function of G^o, therefore, decreases with stronger correlation. However, the water-filling capacity is *not* a Schur-concave function of G^o, and hence similar conclusions cannot be drawn.

We now focus on the low-SNR and high-SNR regimes. At low SNR, it follows from (12) that I^o depends only on the average received SNR and does not depend on the eigenvalue distribution G^o. Hence, at low SNR, fading correlation has no effect on the mutual information achieved by the equal transmit power strategy. On the other hand, the water-filling capacity at low SNR is approximately the average received SNR amplified by the upper limit of the eigenvalue distribution G^o. Since the stronger the correlation, the more G^o is spread out, this upper limit increases and hence the water-filling capacity at low SNR actually increases as the correlation becomes stronger.

At high SNR, the difference $C^o - I^o$ approaches 0 and hence both are reduced by fading. To calculate I^o, let us make the substitution $\beta(x) = \eta_R(x)\rho + \epsilon(x, \rho)$. Taking the limit as $\rho \to \infty$ in (16), we see that $\epsilon(x, \rho) \to 0$ and $\eta_R(x)$ satisfies the fixed-point equation

$$1 = \int_0^1 \frac{S_R(\omega)}{\eta_R(x) + xS_R(\omega)}\, d\omega \tag{18}$$

for each $x \in [0, 1]$. The high-SNR approximation of I^o is therefore,

$$\begin{aligned} I^o(S_R, S_T, \rho) &= \int_0^1 \log[1 + \rho S_T(x)\eta(x)]\, dx + o(1) \\ &= \log\rho + \int_0^1 \log S_T(\omega)\, d\omega \\ &\quad + \int_0^1 \log \eta_R(x)\, dx + o(1). \end{aligned} \tag{19}$$

This can be simplified further. By the reciprocity property (see Appendix I), we know that $I_n(H) = I_n(H^\dagger)$. From this we can conclude that

$$I^o(S_T, S_R, \rho) = I^o(S_R, S_T, \rho).$$

Let us set $S_T(\omega) = 1$ for all $\omega \in [0, 1]$. From (19), we get

$$I(S_T, S_R, \rho) = \log\rho + \int_0^1 \log \eta_R(x)\, dx + o(1).$$

On the other hand,

$$\begin{aligned} I(S_R, S_T, \rho) &= \log\rho + \int_0^1 \log S_R(\omega)\, d\omega \\ &\quad + \int_0^1 \log(1 - x)\, dx + o(1). \end{aligned}$$

Equating these two expressions, we get

$$\int_0^1 \log \eta_R(x)\, dx = \int_0^1 \log S_R(\omega)\, d\omega + \int_0^1 \log(1 - x)\, dx.$$

Now

$$\int_0^1 \log(1 - x)\, dx = \log e.$$

Substituting these equations into (19), we get

$$\begin{aligned} I(S_T, S_R, \rho) &= \log(\rho/e) + \int_0^1 \log S_T(\omega)\, d\omega \\ &\quad + \int_0^1 \log S_R(\omega)\, d\omega + o(1). \end{aligned} \tag{20}$$

We observe that the first term is the high–SNR capacity for independent fading. Hence, the second and third terms represent the capacity penalty due to correlation at the transmit and receive sides, respectively. The fact that they are not positive follows directly from Jensen's inequality.

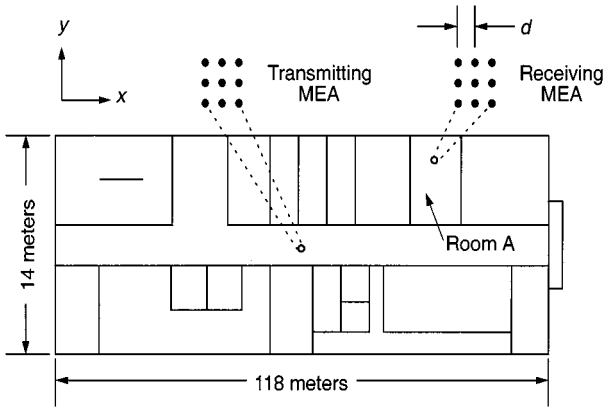

Fig. 1. Floor plan for the first floor of an office building at Crawford Hill, NJ. Receivers with antennas positioned in linear or square grids are placed randomly at 1000 locations in Room A. The transmitting MEA is placed with its adjacent sides parallel to x-axis and y-axis, respectively. The receiving MEA is placed in a random orientation at each of the sample locations.

V. Simulation Experiments

A. Methodology and Assumptions

We use the WiSE ray-tracing simulator [6] to construct random instances of channel matrix H for indoor wireless environment. WiSE allows us to specify the floor plan of a building (e.g., location of vertical walls, ceilings, corridors, etc.) and generate the corresponding propagation models inside the building. As described in [6], the reflection/refraction coefficients and scattering effect for different building materials are derived from a multilayer dielectric model. For our numerical study, we consider the indoor wireless environment of a two-floor office building at Crawford Hill, NJ (see Fig. 1). We place the transmitting MEA on the first floor ceiling near the middle of the office building throughout our study. Receiving MEAs are placed with random rotations at 1000 randomly chosen positions in Room A, which is at intermediate distance from the transmitter. We consider a carrier frequency of 5.2 GHz, i.e., wavelength $\lambda_o = 5.8$ cm. The MEAs consist of multiple omnidirectional antennas, arranged either in square grids or linear arrays within horizontal planes. The separation between antenna elements d is the same at both the transmitting and receiving MEAs. We consider $d = 0.5\lambda_o$ and $d = 5\lambda_o$, unless specified otherwise.

The power of the rays impinging on the receiving antennas is recorded when the carrier is launched from the transmitting MEA with power $10 \log P_{\max}$ dBm. The impulse response between a specific transmitting–receiving antenna pair is modeled as the vector sum of all the rays arriving at the receiving antenna as

$$g_{ij}(t) = \sum_{k=0}^{M} \sqrt{P_k} \cdot e^{i\theta_k} \cdot \delta_k(t - \tau_k) \tag{21}$$

where P_k, θ_k, and τ_k are the received power, phase angle, and time delay of the kth ray, respectively. M is the total number of rays and $\delta_k(t)$ is the delta impulse function. With narrow-band

assumption, we compute the frequency response at infinitesimally small bandwidth centered at the carrier frequency as

$$h_{ij} = \sum_{k=0}^{M} \sqrt{P_k} \cdot e^{i\theta_k} \cdot e^{i2\pi f_o \tau_k}. \tag{22}$$

H is computed using (22) and P_k, θ_k, and τ_k are obtained from the WiSE simulation. All the n^2 elements h_{ij} are complex numbers in this case.

Since H varies for different receiver locations, we estimate the channel variance σ^2 by averaging over 1000 realizations of H, and over all possible antenna pairs, j to i. We assume that the average received SNR ρ, as defined in Section III, should be high enough for low-error-rate communication. If SNR is too low, we need very complex codes to provide enough redundancy to combat the noise so that we can recover the desired signal with low error probability at the receiver. The practical constraints on analog-to-digital (A/D) converters that are available with current technology limit the maximum SNR that can be exploited effectively. Thus, we consider SNRs in the 18–22 dB range. For all our simulations, we assume W to be 10 MHz, and N_o to be -170 dBm/Hz, giving a total noise variance N_oW of -100.8 dBm. When we take expectation with respect to different realizations of H, we mean taking the ensemble average over the 1000 sample receiver locations. The capacities with and without water filling, C_n and I_n, are computed for different n. The results are presented in terms of complementary cumulative distribution functions (CCDFs), the averages $\overline{C}_n$ and $\overline{I}_n$, and capacities at 5% channel outage, $C_n^{0.05}$ and $I_n^{0.05}$.

B. Fading Correlation

As mentioned before, H_{ij} are correlated for finite separation between antenna elements. For an illustration, we consider the case of a two-antenna MEA system. 1000 realizations of the channel matrix H are generated using WiSE for different antenna spacing d. Using the notations in Section IV, Ψ_{12}^T and Ψ_{12}^R are determined. The magnitude of Ψ_{12}^T and Ψ_{12}^R resemble zero-ordered Bessel functions that decay very slowly, as shown in Fig. 2(a). At $d = 0.5\lambda_o$, a strong correlation of 0.65 exists between path gains originating from different transmitters. The correlation between path gains arriving at different receivers is 0.34. The asymmetry is due to the different local scattering environments around the transmitter and receiver.

In Section IV, we modeled the two-dimensional correlation function in product form. To verify the appropriateness of this approach, we plot the product of Ψ_{12}^T and Ψ_{12}^R in Fig. 2(b), together with the correlation $E[H_{11}H_{22}^*]$ inferred from WiSE simulation results. Close agreement is found consistently between these two curves over the range of d we consider. This implies that our assumption of separable transmit/receive correlations in Section IV-B1 is a reasonable first approximation.

C. Capacity of MEA Systems

In this subsection, we consider square MEAs, which are more compact than linear arrays for a given n. The receiver MEA is placed in Room A. Fig. 3 shows the CCDFs of C_n and I_n for $n = 1, 4, 9, 16, 25$, and 36, assuming $d = 0.5\lambda_o$ and $\rho = 18$ dB. Recall that C_n and I_n are defined as the capacity with optimal water-filling power allocation and with equal-power allocation, respectively. Examining Fig. 3, we see that as n increases, the CCDFs of both C_n and I_n shift to the right, indicating that MEAs yield a capacity gain that increases steadily with n. We see that as n increases, the horizontal gap between

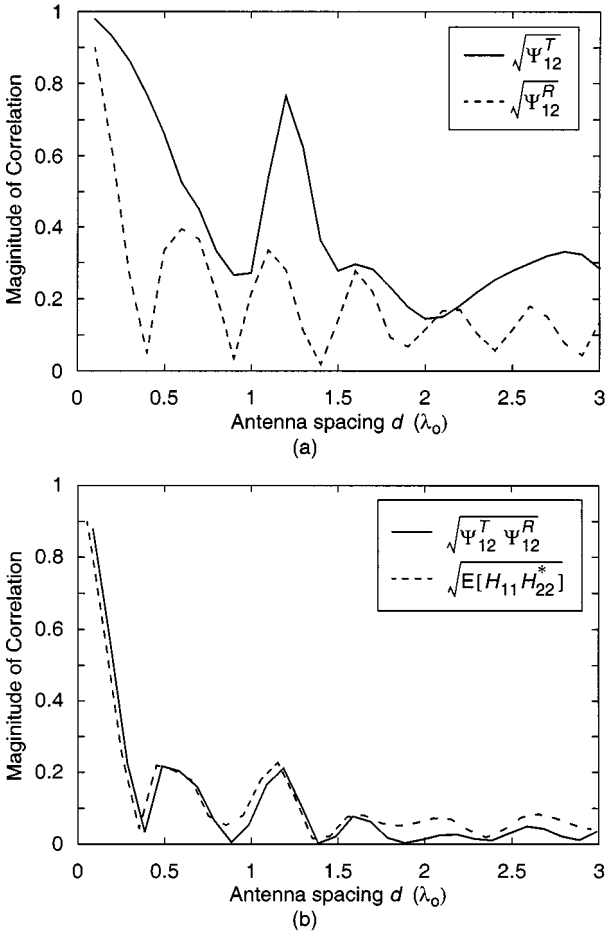

Fig. 2. (a) Magnitude of correlations Ψ_{12}^T and Ψ_{12}^R (as defined in Section IV) for antenna spacings ranging from 0 to $3\lambda_o$. (b) Magnitude of the normalized correlation $E[H_{11}H_{22}^*]$ compared to the magnitude of the product $\Psi_{12}^T\Psi_{12}^R$.

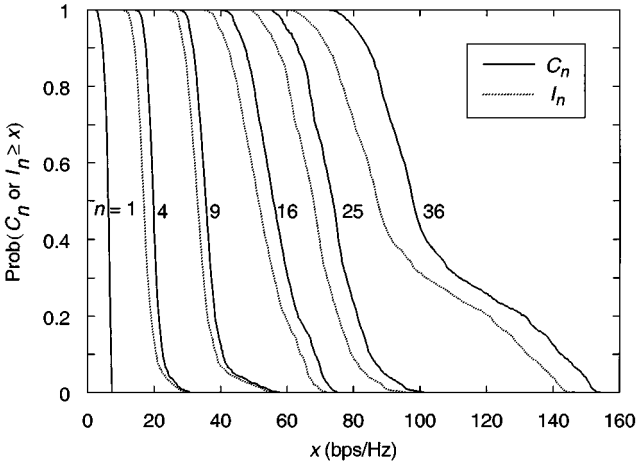

Fig. 3. The CCDFs of C_n (achieved via water filling) and I_n (with equal power allocation) for $n = 1, 4, 9, 16, 25$, and 36 at $\rho = 18$ dB. MEAs are arranged in square grids with $d = 0.5\lambda_o$.

TABLE I
THE PERCENTAGE DIFFERENCE: $(C_n^{0.05} - I_n^{0.05})/I_n^{0.05}$ FOR MEAS PLACED IN ROOM A. HERE $\rho = 18$ dB AND $d = 0.5\lambda_o$

Number of antennas, n	1	4	9	16	25	36
$C_n^{0.05}$ (bps/Hz)	5.9	20	36	57	75	106
$I_n^{0.05}$ (bps/Hz)	5.9	19	34	52	68	95
% difference	0	5.7	7.2	8.8	10	11

C_n and I_n increases, i.e., water filling yields a larger gain over equal-power allocation.

A reasonable performance indicator is the capacity that can be supported with 5% outage. Table I presents values of $C_n^{0.05}$ and $I_n^{0.05}$ extracted from the CCDFs shown in Fig. 3. When $n = 1$, $C_n^{0.05} = I_n^{0.05} = 5.9$ b/s/Hz. Increasing n can yield dramatic increases in $C_n^{0.05}$ and $I_n^{0.05}$. When $n = 4$, $C_n^{0.05} = 20$ b/s/Hz and $I_n^{0.05} = 19$ b/s/Hz, which are nearly three-and-a-half times higher than for $n = 1$. Increasing to $n = 36$, we obtain $C_n^{0.05} = 106$ b/s/Hz and $I_n^{0.05} = 95$ b/s/Hz, which are, respectively about 18 and 16 times higher than for $n = 1$.

Table I also presents $(C_n^{0.05} - I_n^{0.05})/I_n^{0.05}$, the fractional gain yielded by water filling over equal-power allocation. This fractional gain increases from 0 $(n = 1)$ to 11.3% $(n = 36)$.

The capacity improvement of water filling over equal-power allocation depends not only on n, but on the SNR ρ as well. Fig. 4(a) and (b) shows the ratios $\overline{C}_n/\overline{I}_n$ and $C_n^{0.05}/I_n^{0.05}$ versus ρ for $n = 4, 9, 16$. The figure assumes MEAs on square grids with $d = 0.5\lambda_o$. The receiving MEA is placed in Room A. The ratios $\overline{C}_n/\overline{I}_n$ and $C_n^{0.05}/I_n^{0.05}$ are substantial at low SNR ρ, and decrease asymptotically toward unity as ρ increases. When ρ is low, it is important to allocate the available power to the strongest subchannels, while as ρ increases, there is sufficient power to be distributed over all subchannels.

D. Asymptotic Behavior of MEA Capacities

In this section, we study the asymptotic behavior of the capacity as n grows large. We focus on the high-SNR regime, considering $\rho = 22$ dB. Since $C_n \approx I_n$ for high SNR, we consider only the water-filling capacity C_n here. We consider linear MEAs for two different values of antenna spacing: $d = 0.5\lambda_o$ and $d = 5\lambda_o$. The transmitting MEA is placed either parallel to the long dimension of the hallway (inline case) or perpendicular to it (broadside case). In all cases, the receiving MEA is placed in a random angular orientation in Room A. In this section, we consider the average capacities obtained on simulated channels $\overline{C}_n$ as opposed to the 5% outage capacity $C_n^{0.05}$ considered in the previous section.

In order to compute the asymptotic growth rate of capacity assuming independent fading, as derived in Section IV-A, we use simulated channel matrices H' whose entries H'_{ij} are generated by placing individual transmitter and receiver antenna elements at i.i.d. random locations in Room A, instead of placing them in a linear array separated by a fixed distance (e.g., $d = 0.5\lambda_o$) as in the regular case. With such an arrangement, the fades between antenna pairs are almost mutually independent. We use these matrices H' to estimate the variance σ^2 and the equivalent SNR ρ. We then compute the asymptotic growth rate C^*

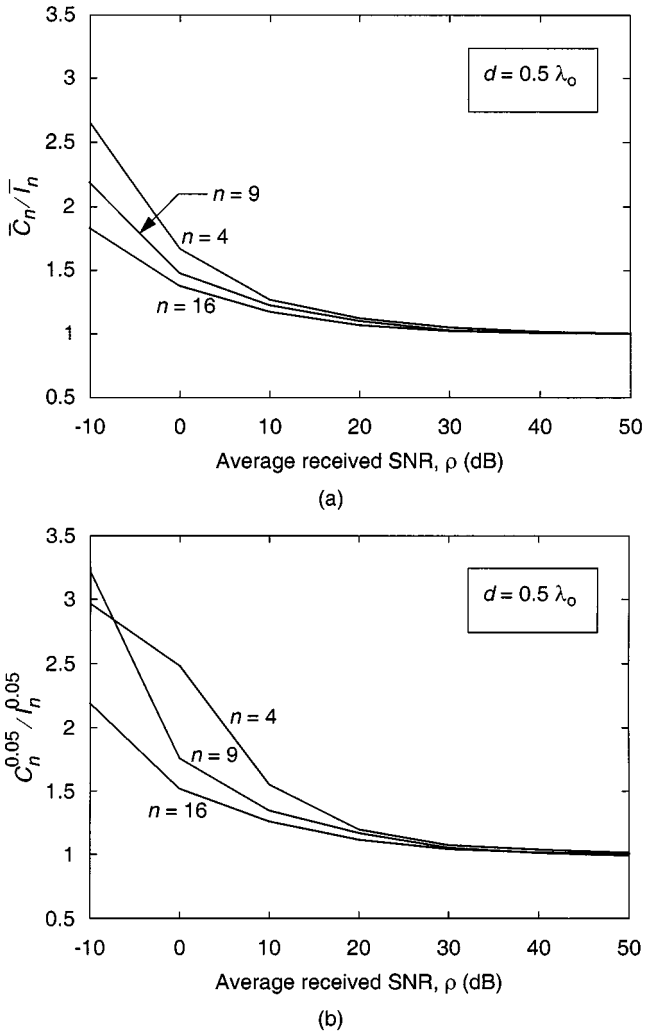

Fig. 4. (a) The ratio of average capacity with water filling to that with equal-power allocation $\overline{C}_n/\overline{I}_n$ at varying average received SNR ρ, for $n = 4, 9$, and 16. (b) Ratio of 5% outage capacity with water filling to that with equal-power allocation $C_n^{0.05}/I_n^{0.05}$ over different ρ, for $n = 4, 9$, and 16 . In both (a) and (b) MEAs at both the transmitter and the receiver are arranged in square grids.

using (11). To compute the growth rate of capacity including fading correlation, as derived in Section IV-B, we use simulated channel matrices H to estimate the variance of H_{ij} (i.e., σ^2), λ_i^R and λ_j^T. In this case, both the transmitting and the receiving MEAs are linear arrays, and individual elements are placed at a fixed spacing of either $0.5\lambda_o$ or $5\lambda_o$ apart. Recall that λ_i^R and λ_j^T are eigenvalues of Ψ^R and Ψ^T, respectively. For each case, we generate 1000 random channel matrices H and estimate the covariance matrices Ψ^R and Ψ^T as $E[H_{i\cdot}^{\dagger} H_{i\cdot}]$ and $E[H_{\cdot j} H_{\cdot j}^{\dagger}]$, respectively. Since the asymptotic growth rate C^o given by (15) is difficult to compute, we can approximate C^o by computing I^o (since $C^o \approx I^o$ at high SNR) as given by (20). We can approximate $S_T(w)$ and $S_R(w)$ in (20) with piecewise-linear curves, and replace the integrals with summations

$$C_n^o \approx I_n^o \approx \frac{1}{n}\left(\log(\rho) + \sum_{i=1}^{n} \log \lambda_i^R + \sum_{j=1}^{n} \log \lambda_j^T\right) \quad (23)$$

where λ_i^R's and λ_j^T's are eigenvalues of Ψ^R and Ψ^T, respectively.

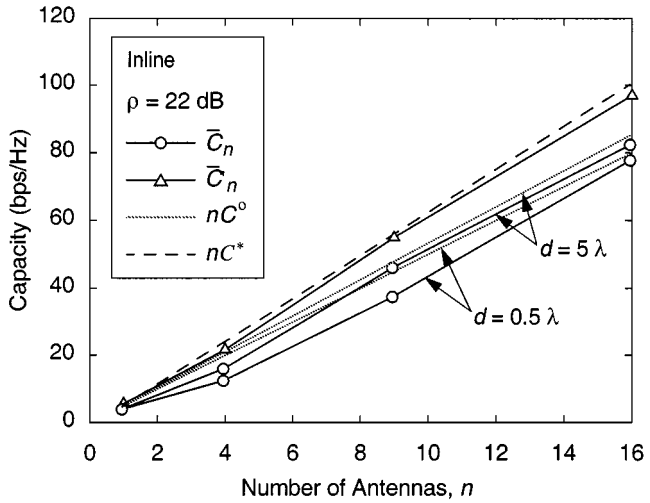

Fig. 5. Average capacity $\overline{C}_n$ versus n for inline case. We consider linear arrays with two antenna spacings: $d = 0.5\lambda_o$ and $d = 5\lambda_o$. The transmitting MEA is placed parallel to the x-axis. The receiving MEA is placed in a random orientation at 1000 random locations in Room A. $\overline{C}'_n$ is the average capacity obtained when the transmit and receive antenna elements lie at i.i.d. locations within their respective workspaces, i.e., they are not constrained to regular linear arrays. nC^* and nC^o are asymptotic results for correlated and independent H_{ij}, respectively.

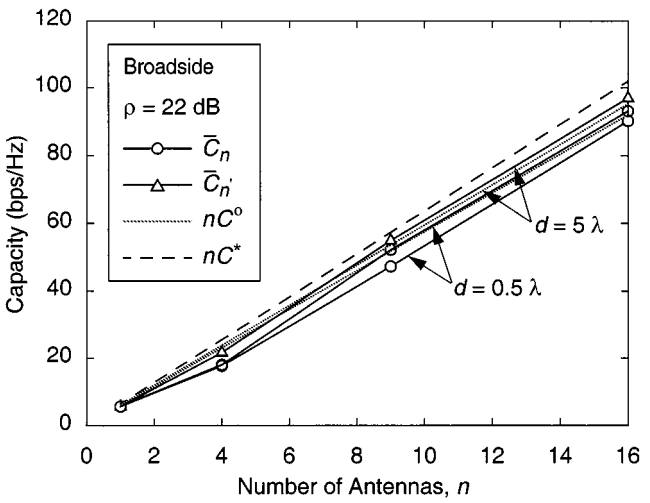

Fig. 6. Average capacity $\overline{C}_n$ versus n for the broadside case. We consider linear arrays with two antenna spacings: $d = 0.5\lambda_o$ and $d = 5\lambda_o$. The transmitting MEA is placed parallel to the y-axis. The receiving MEA is placed in a random orientation at 1000 random locations in Room A. $\overline{C}'_n$ is the average capacity obtained when the transmit and receive antenna elements lie at i.i.d. locations within their respective workspaces, i.e., they are not constrained to regular linear arrays. nC^* and nC^o are asymptotic results for correlated and independent H_{ij}, respectively.

We first consider the inline case. Fig. 5 shows the average capacity $\overline{C}_n$ versus n for $d = 0.5\lambda_o$ and $d = 5\lambda_o$. For each d, $\overline{C}_n$ grows roughly linearly with n, from about 5.57 b/s/Hz at $n = 1$ to about 79.3 b/s/Hz for $d = 0.5\lambda_o$ and 84.0 b/s/Hz for $d = 5\lambda_o$ at $n = 16$. At each n, $\overline{C}_n$ is larger for $d = 5\lambda_o$ than for $d = 0.5\lambda_o$, because the larger d reduces fading correlation, which was shown in Section IV-B to reduce capacity. In Fig. 5, we see that the asymptotic growth rate of capacity, assuming independent fading nC^*, significantly exceeds the observed average capacities $\overline{C}_n$ even for $d = 5\lambda_o$. The discrepancy between nC^* and $\overline{C}_n$ grows with increasing n. On the other hand, the asymptotic growth rates of capacity including correlation nC^o form better upper bounds for $\overline{C}_n$ than nC^* for both values of d. nC^o are in better agreement with $\overline{C}_n$ for $d = 5\lambda$ than $d = 0.5\lambda$, for all values of n.

To further explore the effects of fading correlation, we have computed channel matrices H'_{ij} in which the transmitter and receiver antenna elements lie in i.i.d. random locations within their respective workspaces, rather than lying in a regular linear or square array. Thus, the elements of H'_{ij} should be more nearly independent than those of H_{ij}. We have used these H'_{ij} to compute the average capacity $\overline{C}'_n$, which is also shown in Fig. 5. We observe that at each n, $\overline{C}'_n$ is larger than $\overline{C}_n$, and that $\overline{C}'_n$ is nearly as large as nC^*, the asymptotic growth rate assuming independent fading.

Similar results are obtained for the broadside case, as shown in Fig. 6. The average capacity $\overline{C}_n$ is generally higher than for the inline case for both $d = 0.5\lambda_o$ and $5\lambda_o$. Indeed, $\overline{C}_n$ in this case lies closer to the asymptotic value nC^* than $\overline{C}_n$ for the inline case. At $n = 16$ and $d = 5\lambda_o$, $\overline{C}_n/nC^* = 91.3\%$ for the broadside case, but only 82.4% for the inline case. This suggests that there is less correlation between path gains for the broadside configuration than for the inline configuration.

Our results indicate that fading correlation can significantly reduce MEA system capacity, even for antenna element spacing as large as $d = 5\lambda_o$. Moreover, our results show that the asymptotic growth rate nC^o, which considers correlation, provides a good estimate of the observed average capacity $\overline{C}_n$. This tends to validate the assumptions under which the formula for C^o was derived in Section IV-B, including the correlation model for the fades between different antenna pairs. If the assumptions in Section IV-B hold, C_n/n should converge almost surely to C^o (see (23)) in the limit of large n and high SNR. In Figs. 7 and 8, we illustrate this asymptotic behavior of C_n/n by plotting the empirical probability density functions (pdfs) of C_n/n for $n = 4$, 9, and 16, considering $\rho = 22$ dB. Fig. 7 considers $d = 0.5\lambda_o$ (where there is strong correlation between elements of H_{ij}), while Fig. 8 considers $d = 5\lambda_o$ (where there is less correlation between elements of H_{ij}). As n increases, the pdf becomes narrower and has a higher peak value, i.e., C_n/n becomes less random. In the limit of large n, we expect the pdf of C_n/n to converge to an impulse function centered at the value C^o. The narrowing pdfs in Figs. 7 and 8 illustrate the almost-sure convergence of C_n/n to C^o. Note that when $d = 5\lambda_o$, the pdfs are narrower and taller than when $d = 0.5\lambda_o$. This indicates that the rate of convergence is higher when d is larger, i.e., when the correlation between elements of H_{ij} is lower.

VI. Conclusion

MEA systems offers potentially large capacity gains over single-antenna systems. With perfect channel knowledge at the transmitter, power can be allocated optimally over different transmitting antennas (water filling) to achieve capacity C_n. The water-filling gain C_n/I_n is most significant when there are fewer strong eigenmodes, i.e., when the average received

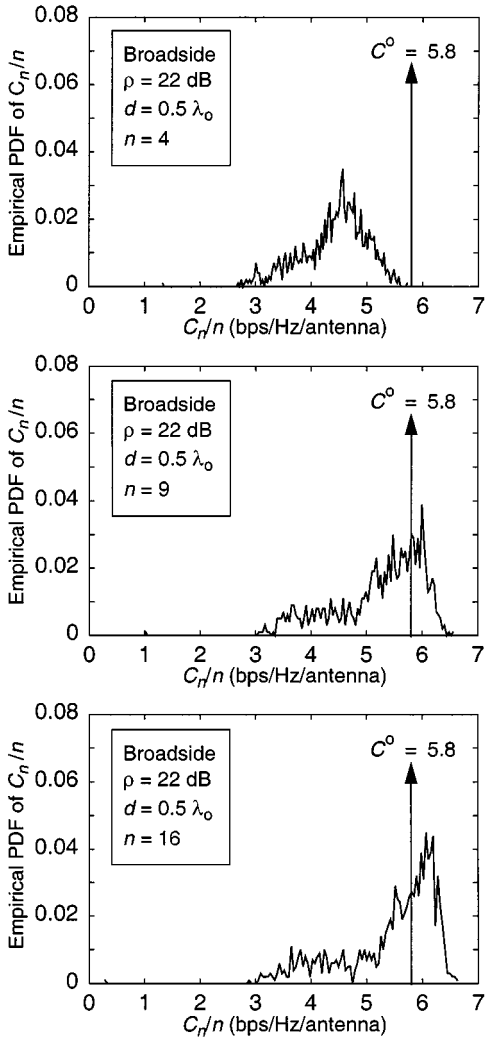

Fig. 7. Empirical pdf of the normalized capacity for $n = 4, 9,$ and 16. We consider linear arrays with antenna elements separated by $0.5\lambda_o$. The reference value is C^o as predicted by the asymptotic theory considering correlated H_{ij}.

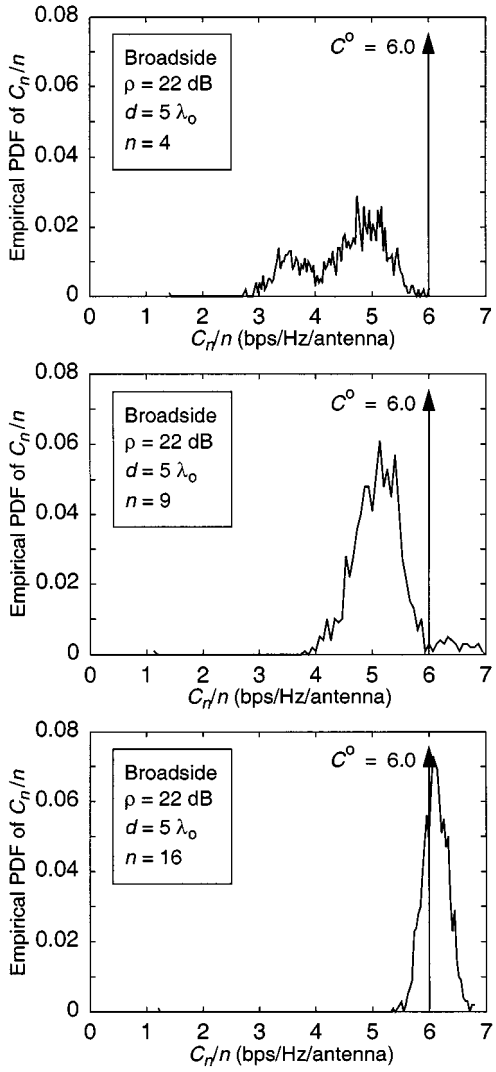

Fig. 8. Empirical pdf of the normalized capacity for $n = 4, 9,$ and 16. We consider linear arrays with antenna elements separated by $5\lambda_o$. The reference value is C^o as predicted by the asymptotic theory considering correlated H_{ij}.

SNR ρ is small. For example, $C_n^{0.05}/I_n^{0.05} = 3.5$ when $\rho = -10$ dB, but at $\rho = 50$ dB, water-filling gain is negligible, $C_n^{0.05}/I_n^{0.05} \approx 1$.

Assuming i.i.d. path gains between different antenna pairs, theoretical analysis shows that the capacity grows linearly with the number of antennas n in the limit of large n. In a more realistic propagation environment, correlation does exist between antenna pairs and affects the rate of growth of C_n and I_n, although it was shown that they still grow linearly with n. The rate of growth of I_n is reduced by correlation over the entire range of SNRs, while that for C_n is reduced by correlation at high SNR but is increased at low SNR. Our simulation results show that for $0.5\lambda_o$ antenna spacing, the simulated average capacity $\overline{C}_n$ is only 88.5% of the predicted value based on independent fading assumptions, nC^* for $n = 16$ in the case of broadside with $\rho =$ 22 dB. When the antenna spacing is increased to $d = 5\lambda_o$, the effect of correlations on total capacity is smaller: $\overline{C}_n/nC^* =$ 91.3%. The approximation based on our asymptotic results for correlated fading nC^o forms a close upper bound for the average capacity observed on simulated channels $\overline{C}_n$ at high SNR.

APPENDIX I
RECIPROCITY PROPERTY

We note here two reciprocity properties discussed in [20] which will be useful in some of the analysis in Section IV-B. These properties also make it very easy to extend the analytical results to the case with unequal number of transmit and receive antennas.

First, the (water-filling) capacity of the multiantenna channel with t transmit and r receive antennas and channel matrix H is the same as that of a system with r transmit and t receive antennas and channel matrix $H^\dagger$. This is because the nonzero singular values of $HH^\dagger$ and $H^\dagger H$ are identical. Second, the mutual information achieved with equal transmit power Q at each antenna in a system with r transmit and t receiver antennas and channel matrix H is the same as that achieved using transmit power Q in a system with r transmit and t receive antennas and channel matrix $H^\dagger$. This is because

$$I_n = \log\det(I_{r\times r} + QHH^\dagger) = \log\det(I_{t\times t} + QH^\dagger H).$$

It should be noted that for the water-filling capacity, the reciprocity is with respect to two systems with the same *total* transmit power, whereas for the equal-power mutual information, the reciprocity is with respect to two systems with the same power *per transmit antenna*.

APPENDIX II
PROOF OF THEOREM IV.2

The following theorem captures the essence of [18, Corollary 10.1.2], which is the key random matrix result we need.

Theorem AII.1: Let A_n be an $n \times n$ random matrix with independent entries which are zero-mean and satisfy the condition

$$n \operatorname{Var}(A_{ij}) < B$$

for some uniform bound $B < \infty$. Moreover, suppose we define for each n a function v_n: $[0, 1] \times [0, 1] \to \Re$ by

$$v_n(x, y) = n \operatorname{Var}(\mathbf{A}_{ij}),$$

i, j, satisfying

$$\frac{i}{n} \le x \le \frac{(i+1)}{n}, \qquad \frac{j}{n} \le y \le \frac{(j+1)}{n}$$

and that v_n converges uniformly to a limiting bounded function v. Then, the limiting eigenvalue distribution H^* of $A_n A_n^\dagger$ exists and its Steltjes' transform $m(z)$ is given by

$$m(z) = \int_0^1 u(x, z)\, dx \tag{24}$$

and $u(x, z)$ satisfies the equation,

$$u(x, z) = \frac{1}{-z + \int_0^1 \frac{v(x,y)}{1+\int_0^1 u(w,z)v(w,y)\,dw}\, dy}. \tag{25}$$

The solution of (25) exists and is unique in the class of functions $u(x, t) \ge 0$, analytical in z and continuous on $x \in [0, 1]$.

To prove Theorem IV.1, we can apply this random matrix result to

$$A_n := \frac{1}{\sqrt{n}} D_R^{1/2} W D_T^{1/2}.$$

The desired result follows by noting that, in this case, $v(x, y) = S_T(x) S_R(y)$.

APPENDIX III
PROOF OF THEOREM IV.3

Let us define

$$\tilde{W} := \frac{1}{\sqrt{n}} D_R^{1/2} W \tag{26}$$

and let $\tilde{w}_k$ be the kth column of $\tilde{W}$. The mutual information I_n has the same distribution as

$$\log\det\left(I + \rho \tilde{W} D_T \tilde{W}^\dagger\right).$$

This can be interpreted as the sum capacity of a vector multiple-access channel with n users and n degrees of freedom at the receiver, treating each of the n transmit antennas as a separate user. The transmit power of the kth user is $\rho(D_T)_{kk}$, and its vector of channel gain at the receiver is $\tilde{w}_k$.

The sum capacity can be achieved by a combination of successive cancellation and linear MMSE demodulation [21]: first one user is demodulated by a linear MMSE receiver and decoded, treating all the other users as interference; then the signal from that user is subtracted off and the process is repeated for the remaining users. Moreover, the sum capacity is achieved regardless of the decoding order among the users. Let us decode then in increasing order of the transmit powers of the users, and without loss of generality assume that the diagonal elements of D_T are in increasing order. Thus, we have

$$\log\det\left(I + \rho \tilde{W} D_T \tilde{W}^\dagger\right) = \sum_{k=1}^{n} \log(1 + \mathrm{SIR}_k), \tag{27}$$

where SIR_k is the signal-to-interference ratio (SIR) achieved when demodulating the kth user. From MMSE estimation theory it can be computed that

$$\mathrm{SIR}_k = \rho(D_T)_{kk} \tilde{w}^\dagger \left(\rho \tilde{W} D_T^{(k)} \tilde{W}^\dagger + I\right)^{-1} \tilde{w}$$

where

$$(D_T^{(k)})_{ii} = 0, \qquad \text{if } i \le k$$

and

$$(D_T^{(k)})_{ii} = (D_T)_{ii}, \qquad \text{if } i > k.$$

Substituting (26), we get

$$\mathrm{SIR}_k = \frac{(D_T)_{kk}}{n} w_k^\dagger \left(\frac{1}{n} W D_T^{(k)} W^\dagger + \frac{1}{\rho} D_R^{-1}\right)^{-1} w_k$$

where w_k is the kth column of W.

We now let $n \to \infty$ but keep $k = xn$ for x fixed. We wish to compute the asymptotic limit of SIR_k.

The entries of w_k are i.i.d. We can apply [22, Lemma 3.2] to conclude that

$$\mathrm{SIR}_k - (D_T)_{kk} \frac{1}{n} \operatorname{Tr}\left(W D_T^{(k)} W^\dagger + \frac{1}{\rho} D_R^{-1}\right)^{-1} \to 0 \tag{28}$$

in probability. The trace term is the Steltjes' transform of the empirical eigenvalue distribution of $W D_T^{(k)} W^\dagger + \frac{1}{\rho} D_R^{-1}$ evaluated at 0. The following theorem, due to Marcenko and Pastur [23] and refined by Silverstein and Bai [15], computes the limit.

Theorem AIII.1: Let A and B be two n by n diagonal matrices whose empirical distributions of the diagonal elements converge to F_A and F_B, respectively, as $n \to \infty$. Let m_A be the Steltjes' transform of F_A. Then, the empirical eigenvalue distribution of $\frac{1}{n} W B W^\dagger + A$ converges almost surely, and the Steltjes' transform $m(z)$ of the limiting distribution is the unique solution to the functional fixed-point equation

$$m(z) = m_A\left(z - \int \frac{\tau}{1 + \tau m(z)}\, dF_B(\tau)\right).$$

Applying this result to the above yields

$$\frac{1}{n} \operatorname{Tr}\left(W D_T^{(k)} W^\dagger + \frac{1}{\rho} D_R^{-1}\right)^{-1} \to m(0) \equiv \beta(x)$$

where $\beta(x)$ satisfies

$$\beta(x) = \int_0^1 \frac{S_R(\omega)}{\frac{1}{\rho} + S_R(\omega)\int_{1-x}^1 \frac{S_T(\phi)}{1+S_T(\phi)\beta(x)}\,d\phi}\,d\omega.$$

Hence, combining this with (27) and (28), we get

$$\frac{1}{n}\log\det\left(I+\rho\tilde{W}D_T\tilde{W}^\dagger\right) = \frac{1}{n}\sum_{k=1}^{n}\log(1+\mathrm{SIR}_k) \to \int_0^1 \log(1+\rho S_T(x)\beta(x))\,dx$$

thus establishing the desired result.

Appendix IV
A Majorization Result

In this appendix, we will explore the effect of correlation has on the limiting eigenvalue distribution of the key random matrix $D_R W D_T W^\dagger$. Just as for power spectrum, we can define an equivalent notion of a distribution being more spread out than another.

Theorem AIV.1: A distribution F is more spread out than G if they have the same expectation, and for every $\theta \in [0, 1]$

$$\int_\theta^1 F^{-1}(x)\,dx \geq \int_\theta^1 G^{-1}(x)\,dx.$$

The main result we want to prove is the following.

Theorem AIV.2: Let S_R, $\tilde{S}_R$ be two receiver correlation power spectra, and S_T, $\tilde{S}_T$ be two transmitter correlation power spectra. Let G and $\tilde{G}$ be the limiting eigenvalue distributions of $D_R W D_T W^\dagger$ and $\tilde{D}_R W \tilde{D}_T W^\dagger$, respectively, where the diagonal elements of D_R, $\tilde{D}_R$, D_T, $\tilde{D}_T$ approach the spectra S_R, $\tilde{S}_R$, S_T, $\tilde{S}_T$, respectively. If $\tilde{S}_R$ is more spread out than S_T and $\tilde{S}_T$ is more spread out than S_T, then $\tilde{G}$ is more spread out than G.

To prove the theorem, we need the following definition, which gives a slightly more general notion of Schur convexity than the one presented earlier in the main body of the paper.

Definition AIV.3: Consider a map $\mathcal{H}$: $\mathcal{L}^1[0, 1] \to \Re$, where $\mathcal{L}^1[0, 1]$ is the set of all integrable functions on $[0, 1]$. For a function $g \in \mathcal{L}^1[0, 1]$, let F_g be the empirical distribution of g, i.e., $F_g(x) = m\{u\colon g(u) \leq x\}$, where m is the Lebesque measure. The map $\mathcal{H}$ is said to be Schur-convex if for any two functions g_1 and g_2 such that F_{g_1} is more spread out than F_{g_2} implies that $\mathcal{H}(g_1) \geq \mathcal{H}(g_2)$.

The following is a key lemma in the proof of Theorem AIV.2.

Lemma AIV.4: Suppose the map $\mathcal{H}$: $\mathcal{L}^1[0, 1] \to \Re$ has the following properties.

1) For any g, $\mathcal{H}(g)$ depends on g only through the empirical distribution F_g.
2) $\mathcal{H}$ is convex, i.e.,

$$\mathcal{H}(\alpha g_1 + (1-\alpha)g_2) \leq \alpha\mathcal{H}(g_1) + (1-\alpha)\mathcal{H}(g_2)$$

or $\alpha \in [0, 1]$.

Then $\mathcal{H}$ is Schur-convex.

Proof: Fix n and consider the class of functions in $\mathcal{L}^1[0, 1]$ that are piecewise-constant functions on the intervals $[k/n, (k+1)/n)$, for $k = 0, \ldots, n-1$. The mapping $\mathcal{H}$ on this class can be viewed as a function of n variables, the n values that the functions in the class can take on. Condition 1) then implies that $\mathcal{H}$ is a symmetric function of these n variables. The Schur convexity of $\mathcal{H}$ on this class follows from [19, Proposition C.2, p. 67]. The Schur convexity of $\mathcal{H}$ on $\mathcal{L}^1[0, 1]$ follows from an approximating argument by taking n large. □

Proof of Theorem AIV.2: If A is an n by n Hermitian matrix and $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)$ are its ordered eigenvalues, then it is known that for all k

$$\sum_{i=1}^{k}\lambda_i(A) = \max_{UU^\dagger = I_{k\times k}} UAU^\dagger$$

where the maximization is over $k \times n$ complex matrices U. This extremal representation shows that $\sum_{i=1}^k \lambda_i(A)$ is a convex function of the entries of A. Applying this observation to the matrix $A = D_R H D_T H^\dagger$, it follows that $\sum_{i=1}^k \lambda_i(D_R H D_T H^\dagger)$ is a convex function of the entries of D_R, and also a convex function of the entries of D_T. From Theorem IV.2, almost surely the empirical eigenvalue distribution of $D_R H D_T H^\dagger$ converges to a limiting distribution G. This implies that for each fixed $\theta \in [0, 1]$

$$\sum_{i=1}^{\lfloor(1-\theta)n\rfloor}\lambda_i(D_R H D_T H^\dagger) \to \int_\theta^1 (G)^{-1}(x)\,dx := \mathcal{C}(g_R, g_T)$$

where g_R, $g_T \in \mathcal{L}^1[0, 1]$ are the limiting functions of the diagonal elements of D_R, D_T, respectively. The convexity of $\mathcal{C}$ both as a function of g_R and as a function of g_T follows from a limiting argument. Moreover, the dependence of $\mathcal{C}$ on g_T and g_R is only through their empirical distributions. Hence, from Lemma AIV.4, it follows that $\mathcal{C}$ is a Schur-convex function of g_R for a fixed g_T, and also a Schur-convex function of g_T for a fixed g_R. Hence, for a fixed S_R, if $\tilde{S}_T$ is more spread out than S_T, then

$$\int_\theta^1 \left(\tilde{G}\right)^{-1}(x)\,dx \geq \int_\theta^1 (G)^{-1}(x)\,dx.$$

This holds for all θ, and hence $\tilde{G}$ is more spread out than G; similarly, for a fixed S_T, if $\tilde{S}_R$ is more spread out than S_R, then $\tilde{G}$ is more spread out than G. This proves the theorem.

Acknowledgment

The authors wish to thank J. Ling and D. Chizhik for their assistance with the WiSE simulation tools. Discussions with G. J. Foschini, J. Salz, and D. Shiu have been enlightening and are greatly appreciated.

References

[1] W. Jakes, Jr., *Microwave Mobile Communications*. New York: Wiley, 1974.
[2] J. Winters, "Optimum combining for indoor radio systems with multiple users," *IEEE Trans. Commun.*, vol. 35, pp. 1222–1230, Nov. 1987.

[3] G. J. Foschini and M. J. Gans, "Capacity when using diversity at transmit and receive sites and the Rayleigh-faded matrix channel is unknown at the transmitter," in *Proc. WINLAB Workshop on Wireless Information Network*, Mar. 1996.
[4] ——, "On limits of wireless communication in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.
[5] W. C.-Y. Lee, "Effects on correlation between two mobile radio base-station antennas," *IEEE Trans. Commun.*, vol. COM-21, pp. 1214–1224, Nov. 1974.
[6] S. J. Fortune, D. H. Gay, B. W. Kernighan, O. Landron, R. A. Valenzuela, and M. H. Wright, "WiSE design of indoor wireless systems: Practical computation and optimization," *IEEE Comput. Sci. Eng.*, vol. 2, pp. 58–68, Mar. 1995.
[7] G. J. Foschini and R. A. Valenzuela, "Initial estimation of communication efficiency of indoor wireless channels," *Wireless Networks*, vol. 3, no. 2, pp. 141–154, 1997.
[8] C. N. Chuah, J. M. Kahn, and D. Tse, "Capacity of multi-antenna array systems in indoor wireless environment," in *Proc. IEEE GLOBECOM*, vol. 4, Nov. 1998, pp. 1894–1899.
[9] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Personal Commun.*, vol. 5, pp. 10–22, Feb. 1998.
[10] J. Salz and J. H. Winters, "Effect of fading correlation on adapting arrays in digital wireless communications," in *Proc. IEEE Vehicular Technology Conf.*, vol. 3, 1993, pp. 1758–1774.
[11] J. Fuhl, A. F. Molisch, and E. Bonek, "Unified channel model for mobile radio systems with smart antennas," in *Proc. Inst. Elec. Eng.–Radar, Sonar Navig.*, vol. 145, Feb. 1998, pp. 32–41.
[12] G. G. Raleigh and J. M. Cioffi, "Spatio–temporal coding for wireless communications," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.
[13] D. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502–513, Mar. 2000.
[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[15] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *J. Multivariate Anal.*, vol. 54, no. 2, pp. 175–192, 1995.
[16] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 622–640, Mar. 1999.
[17] A. Goldsmith and P. Varaiya, "Capacity of fading channel with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.
[18] V. L. Girko, *Theory of Random Determinants*. Norwell, MA: Kluwer, 1990.
[19] A. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.
[20] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecommun.*, vol. 10, no. 6, pp. 586–595, 1999.
[21] M. K. Varanasi and T. Guess, "Optimum decision feedback multiuser equalization and successive decoding achieves the total capacity of the Gaussian multiple-access channel," in *Proc. Asilomar Conf. Signals, Systems and Computers*, Nov. 1997, pp. 1405–1409.
[22] D. Tse and O. Zeitouni, "Performance of linear multiuser receivers in random environments," in *IEEE Communications Theory Mini-Conf.*, 1999, pp. 163–167.
[23] V. A. Marcenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math. USSR-Sbornik1*, pp. 457–483, 1967.

A Deterministic Equivalent for the Analysis of Correlated MIMO Multiple Access Channels

Romain Couillet, *Member, IEEE*, Mérouane Debbah, *Senior Member, IEEE*, and Jack W. Silverstein

***Abstract*—In this article, novel deterministic equivalents for the Stieltjes transform and the Shannon transform of a class of large dimensional random matrices are provided. These results are used to characterize the ergodic rate region of multiple antenna multiple access channels, when each point-to-point propagation channel is modelled according to the Kronecker model. Specifically, an approximation of all rates achieved within the ergodic rate region is derived and an approximation of the linear precoders that achieve the boundary of the rate region as well as an iterative water-filling algorithm to obtain these precoders are provided. An original feature of this work is that the proposed deterministic equivalents are proved valid even for strong correlation patterns at both communication sides. The above results are validated by Monte Carlo simulations.**

***Index Terms*—Deterministic equivalent, ergodic capacity, MAC, MIMO, optimal precoder, random matrix theory.**

I. Introduction

WHEN mobile networks were expected to run out of power and frequency resources while being simultaneously subject to a demand for higher transmission rates, Foschini [1] introduced the idea of multiple input multiple output (MIMO) communication schemes. Telatar [2] then predicted a growth of the channel capacity by a factor $\min(N, n)$ for an $N \times n$ MIMO system compared to the single-antenna case when the matrix-valued channel is modelled with independent and identically distributed (i.i.d.) standard Gaussian entries. In practical systems though, this linear gain can only be achieved for high SNR and for uncorrelated transmit and receive antenna arrays at both communication sides. Nevertheless, the current scarcity of available frequency resources has led to a widespread incentive for MIMO communications. Mobile terminal engineers now embed numerous antennas in small devices. Due to space limitations, this inevitably induces channel correlation and thus reduced transmission rates. An implication of the results introduced in this paper is the ability to study the performance of MIMO systems subject to strong correlation effects in multi-user and multi-cellular contexts, a question which is paramount to cellular service providers.

Although alternative communication models could be treated using similar mathematical expressions, such as cooperative and non-cooperative multi-cell communications with users equipped with multiple antennas, the present article investigates the MIMO multiple access channel (MIMO-MAC), where K multi-antenna mobile terminals transmit information to a single receiver. Under perfect channel state information at the transmitters (CSIT), the boundaries of the achievable rate region of the MIMO-MAC have been characterized by Yu *et al.* [3] who provide an iterative water-filling algorithm to obtain the sum rate maximizing precoders. However, to achieve perfect CSIT, the channel must be quasi-static during a sufficiently long period to allow feedback or pilot signalling from the receiver to the transmitters. For high mobility wireless services, this is often unacceptable. In this situation, the transmitters are often assumed to have statistical information about the random fast varying channels, such as first order moments of their distribution. The achievable rates are in this case the points lying in the *ergodic* rate region. It is however difficult to characterize the boundary of the ergodic rate region because it is difficult to compute the optimal precoders that reach the boundaries. In the single-user context, an algorithm was provided by Vu *et al.* in [4] to solve this problem. However, the technique of [4] is rather involved as it requires nested Monte Carlo simulations and does not provide any insight on the nature of the optimal precoders.

In the present article, we provide a parallel approach that consists in approximating the ergodic sum rate by *deterministic equivalents*. That is, for all finite system dimensions, we provide an approximation of the ergodic rates, which is accurate as the system dimensions grow asymptotically large. Furthermore, we provide an efficient way to derive an asymptotically accurate approximation of the optimal precoders. The mathematical field of large dimensional random matrices is particularly suited for this purpose, as it can provide deterministic equivalents of the achievable rates depending only on the relevant channel parameters, e.g., the long-term transmit and receive channel covariance matrices in the present situation, the deterministic line of sight components in Rician models as in [5] etc. The earliest notable result in line with the present study is due to Tulino *et al.* [6], who provide an expression of the asymptotic ergodic capacity of point-to-point MIMO communications when the random channel matrix is composed of i.i.d. Gaussian entries. In [7], Peacock *et al.* extend the asymptotic result of [6] in the context of multi-user communications by considering a K-user MAC with channels $\mathbf{H}_1, \ldots, \mathbf{H}_K$ modelled as Gaussian

Manuscript received July 15, 2009; revised August 21, 2010; accepted October 11, 2010. Date of current version May 25, 2011. The work of M. Debbah was supported by the European Commission, FP7 Network of Excellence in Wireless Communications NEWCOM++ and the French ANR Project SESAME. The work of J. W. Silverstein was supported by the U.S. Army Research Office under Grant W911NF-09-1-0266.

R. Couillet and M. Debbah are with the Alcatel-Lucent Chair on Flexible Radio, SUPÉLEC, 3 Rue Joliot-Curie, France (e-mail: romain.couillet@supelec.fr; romain.couillet@gmail.com; merouane.debbah@supelec.fr).

J. W. Silverstein is with the Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205 USA (e-mail: jack@math.ncsu.edu).

Communicated by A. Moustakas, Associate Editor for Communications.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2011.2133151

with a separable variance profile. This is, the entries of $\mathbf{H}_k$ are Gaussian independent with (i,j)-th entry of zero mean and variance $\sigma_{k,ij}^2$ that can be written as a product $\sigma_{k,ij}^2 = r_{k,i}t_{k,j}$ of a term depending on i and a term depending on j. The asymptotic eigenvalue distribution of this matrix model is derived, but neither any explicit expression of the sum rate is provided as in [6], nor is any ergodic capacity maximizing policy derived. In [8], Soysal *et al.* derive the sum rate maximizing precoder policy in the case of a MAC channel with K users whose channels $\mathbf{H}_1, \ldots, \mathbf{H}_k$ are one-side correlated zero mean Gaussian, in the sense that all rows of $\mathbf{H}_k$ have a common covariance matrix, different for each k.

In this article, we concentrate on the more general *Kronecker* channel model. This is, we assume a K-user MIMO-MAC, with channels $\mathbf{H}_1, \ldots, \mathbf{H}_K$, where each $\mathbf{H}_k$ can be written in the form of a product $\mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k^{\frac{1}{2}}$ where $\mathbf{X}_k$ has i.i.d. zero mean Gaussian entries and the left and right correlation matrices $\mathbf{R}_k$ and $\mathbf{T}_k$ are deterministic nonnegative definite Hermitian matrices. This model clearly covers the aforementioned channel models of [6], [7] and [8] as special cases. The Kronecker model is particularly suited to model communication channels that show transmit and receive correlations, different from one user to the next, in a rich scattering environment. Nonetheless, the Kronecker model is only valid in the absence of a line-of-sight component in the channel, when a sufficiently large number of scatterers is present in the communication medium to justify the i.i.d. aspect of the inner Gaussian matrix and when the channel is frequency flat over the transmission bandwidth. Using similar tools as those used in this article, many works have studied these channel models, mostly in a single-user context. We remind the main contributions, from which the present work borrows several ideas. In [5], [9], and [10], Hachem *et al.* study the point-to-point multi-antenna Rician channel model, i.e., non-central Gaussian matrices with a variance profile, for which they provide a deterministic equivalent of the ergodic capacity [5], the corresponding ergodic capacity-achieving input covariance matrix [9] and a central limit theorem for the ergodic capacity [10]. In [11], Moustakas *et al.* provide an expression of the mutual information in time varying frequency-selective Kronecker channels, using the replica method [12]. This result has been recently proved by Dupuy *et al.* in a yet unpublished work. Dupuy *et al.* then derived the expression of the capacity maximizing precoding matrix for the frequency-selective channel [13]. A more general frequency-selective channel model with non-separable variance profile is studied in [14] by Rashibi *et al.* using alternative tools from free probability theory. Of practical interest is also the theoretical work of Tse [15] on MIMO point-to-point capacity in both uncorrelated and correlated channels, which are validated by ray-tracing simulations.

The main contribution of this paper is summarized in two theorems contributing to the field of random matrix theory and enabling the evaluation of the ergodic rate region of the MIMO-MAC with Kronecker channels. We subsequently derive an iterative water-filling algorithm enabling the description of the boundaries of the rate region by providing an expression of the asymptotically optimal precoders. The remainder of this paper is structured as follows: in Section II, we provide a short summary of the main results and how they apply to multi-user wireless communications. In Section III, the two theorems are introduced, the complete proofs being left to the appendices. In Section IV, the ergodic rate region of the MIMO-MAC is studied. In this section, we introduce our third main result: an iterative water-filling algorithm to describe the boundary of the ergodic rate region of the MIMO-MAC. In Section V, we provide simulation results of the previously derived theoretical formulas. Finally, in Section VI, we give our conclusions.

A. Notation

In the following, boldface lower-case characters represent vectors, capital boldface characters denote matrices ($\mathbf{I}_N$ is the $N \times N$ identity matrix). X_{ij} denotes the (i,j) entry of $\mathbf{X}$. The Hermitian transpose is denoted $(\cdot)^{\mathsf{H}}$. The operators $\operatorname{tr}\mathbf{X}$, $|\mathbf{X}|$ and $\|\mathbf{X}\|$ represent the trace, determinant and spectral norm of matrix $\mathbf{X}$, respectively. The symbol $\mathrm{E}\cdot]$ denotes expectation. The notation $F^{\mathbf{Y}}$ stands for the (cumulative) distribution function of the eigenvalues of the Hermitian matrix $\mathbf{Y}$. The function $(x)^+$ equals $\max(x,0)$ for real x. For F, G two distribution functions, we denote $F \Rightarrow G$ the vague convergence of F to G. The notation $x_n \xrightarrow{\text{a.s.}} x$ denotes the almost sure convergence of the sequence x_n to x. The notation $\|F\|$ for the distribution function F is the supremum norm defined as $\|F\| = \sup_x F(x)$. The symbol $\mathbf{X} \geq 0$ for a square matrix $\mathbf{X}$ means that $\mathbf{X}$ is Hermitian nonnegative definite.

II. Scope and Summary of Main Results

In this section, we summarize the main results of this article and explain their impact on the study of the effects of channel correlation on the achievable communication rates in the present multi-user framework.

A. General Model

Consider a set of K wireless terminals, equipped with $n_1, \ldots, n_K$ antennas, respectively, which we refer to as the transmitters, and another device equipped with N antennas, which we call the receiver or the base station. We consider the uplink communication from the terminals to the base station. Denote $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$ the channel matrix between transmitter k and the receiver. Let $\mathbf{H}_k$ be defined as

$$\mathbf{H}_k = \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k^{\frac{1}{2}} \tag{1}$$

where $\mathbf{R}_k^{\frac{1}{2}} \in \mathbb{C}^{N \times N}$ and $\mathbf{T}_k^{\frac{1}{2}} \in \mathbb{C}^{n_k \times n_k}$ are nonnegative Hermitian matrices and $\mathbf{X}_k \in \mathbb{C}^{N \times n_k}$ is a realization of a random matrix with independent Gaussian entries of zero mean and variance $1/n_k$. In this scenario, the matrices $\mathbf{T}_k$ and $\mathbf{R}_k$ model the correlation present in the channel at transmitter k and at the receiver, respectively. This setup is depicted in Fig. 1.

It is important to underline that the correlation patterns emerge both from the inter-antenna spacings on the volume-limited transmit and receive radio devices and from the solid angles of transmitted and received signal energy. Even though the transmit antennas emit signals in an isotropic manner, only a limited solid angle of emission is effectively received, and the same holds for the receiver which captures signal energy in a non-isotropic manner. Given this propagation factor, it is clear that the transmit covariance matrices $\mathbf{R}_k$ matrices are not equal for all users. We nonetheless assume physically identical and

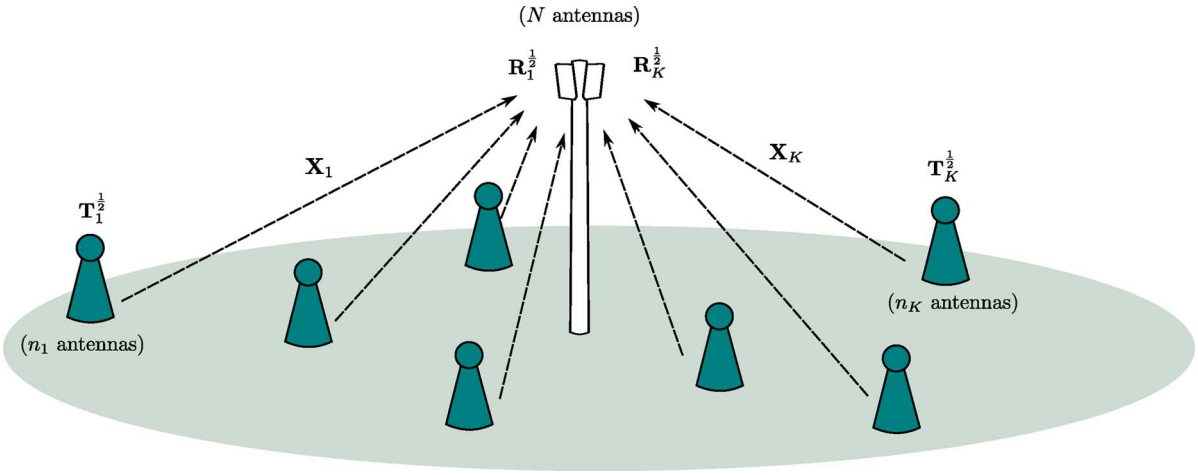

Fig. 1. Multi-antenna multiple access scenario with Kronecker channels.

interchangeable antennas on each device. We, therefore, claim that the diagonal entries of $\mathbf{R}_k$ and $\mathbf{T}_k$, i.e., the variance of the channel fading on every antenna, are identical and equal to one, which, along with the normalization of the Gaussian matrix $\mathbf{X}_k$, allows for a consistent definition of the SNR. As a consequence, $\operatorname{tr}\mathbf{R}_k = N$ and $\operatorname{tr}\mathbf{T}_k = n_k$. We will see that under these trace constraints the hypotheses made in Theorem 1 are always satisfied, therefore making Theorem 1 valid for all possible figures of correlation, including strongly correlated patterns. The hypotheses of Theorem 2, used to characterize the ergodic rate region of the MIMO-MAC, require additional mild assumptions, making Theorem 2 valid for most practical models of $\mathbf{R}_k$ and $\mathbf{T}_k$. These statements are of major importance and rather new since in other contributions, e.g., [5], [13], it is usually assumed that the correlation matrices have uniformly bounded spectral norms across N. Physically, this means that only low correlation patterns are allowed, excluding short distances between antennas and small solid angles of energy propagation. The counterpart of this interesting property is a theoretical reduction of the convergence rates of the derived deterministic equivalents, compared to those proposed in [5] and [13].

The rate performance of multi-cell or multi-user communication schemes is connected to the so-called Stieltjes transform and Shannon transform of matrices $\mathbf{B}_N$ of the type

$$\mathbf{B}_N = \sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k \mathbf{X}_k^{\mathsf{H}} \mathbf{R}_k^{\frac{1}{2}}. \tag{2}$$

We study these matrices using tools from the field of large dimensional random matrix theory [16]. Among these tools, we define the Stieltjes transform $m_N(z)$ of the Hermitian nonnegative definite matrix $\mathbf{B}_N \in \mathbb{C}^{N\times N}$, for $z \in \mathbb{C}\setminus\mathbb{R}^+$, as

$$m_N(z) = \int \frac{1}{\lambda - z} dF_N(\lambda) = \frac{1}{N}\operatorname{tr}\left(\mathbf{B}_N - z\mathbf{I}_N\right)^{-1}$$

where F_N denotes the (cumulative) distribution function of the eigenvalues of $\mathbf{B}_N$. The Stieltjes transform was originally used to characterize the asymptotic distribution of the eigenvalues of large dimensional random matrices [17]. From a wireless communications viewpoint, it can be used to characterize the signal-to-interference plus noise ratio of certain communication models, e.g., [18], [19]. A second interest of the Stieltjes transform in wireless communications is its link to the so-called Shannon transform $\mathcal{V}_N(x)$ of $\mathbf{B}_N$, that we define for $x > 0$ as

$$\begin{aligned}\mathcal{V}_N(x) &= \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\mathbf{B}_N\right) = \int_0^{+\infty} \log\left(1+\frac{\lambda}{x}\right) dF_N(\lambda) \\ &= \int_x^{+\infty} \left(\frac{1}{w} - m_N(-w)\right) dw.\end{aligned}$$

The Shannon transform is commonly used to provide approximations of capacity expressions in large dimensional systems, e.g., [6]. In the present work, the Shannon transform of $\mathbf{B}_N$ will be used to provide a deterministic approximation of the ergodic achievable rate of the MIMO-MAC.

Before introducing our main results, namely Theorem 1 and Theorem 2, which are rather technical and difficult to fathom without a preliminary explanation, we briefly describe these results in telecommunication terms and their consequences to the multi-user multi-cell communication models at hand.

B. Main Results

The main results of this work unfold as follows.

- Theorem 1 provides a deterministic equivalent $m_N^\circ(z)$ for the Stieltjes transform $m_N(z)$ of $\mathbf{B}_N$, under the assumption that N and n_k grow large with the same order of magnitude and the sequences of distribution functions $\{F^{\mathbf{T}_k}\}_{n_k}$ and $\{F^{\mathbf{R}_k}\}_N$ form *tight sequences* [20]. This is, we provide an approximation $m_N^\circ(z)$ of $m_N(z)$ which can be expressed without reference to the random $\mathbf{X}_k$ matrices and which is almost surely asymptotically exact when $N \to \infty$. The tightness hypothesis is the key assumption that allows degenerated $\mathbf{R}_k$ and $\mathbf{T}_k$ matrices to be valid in our framework, and that therefore allows us to study strongly correlated channel models.
- Theorem 2 provides a deterministic equivalent $\mathcal{V}_N^\circ(x)$ for the Shannon transform $\mathcal{V}_N(x)$ of $\mathbf{B}_N$. In this theorem, the assumptions on the $\mathbf{R}_k$ and $\mathbf{T}_k$ matrices are only slightly more constraining and of marginal importance for practical purposes. In particular, Theorem 2 theoretically allows the largest eigenvalues of $\mathbf{T}_k$ or $\mathbf{R}_k$ to grow linearly with N,

as the number of antennas increases, as long as the number of these large eigenvalues is of order $o(N)$.

- Based on Theorem 2, the precoders that maximize the deterministic equivalent of the ergodic sum rate of the MIMO-MAC are computed. Those precoders have the following properties:
  - — the eigenspace of the precoder for user k coincides with the eigenspace of the transmit channel correlation matrix at user k;
  - — the eigenvalues of the precoder for user k are the solution of a water-filling algorithm;
  - — as the system dimensions grow large, the mutual information achieved using these precoders becomes asymptotically close to the channel capacity.

The major practical interest of Theorems 1 and 2 lies in the possibility to analyze mutual information expressions for multidimensional channels, not as the averaged of stochastic variables depending on the random matrices $\mathbf{X}_k$ but as approximated deterministic expressions which do no longer feature the matrices $\mathbf{X}_k$. The study of those quantities is in general simpler than the study of the averaged stochastic expressions, which leads here to a simple derivation of the approximated rate optimal precoders.

In the next section, we introduce our theoretical results, whose proofs are left to the appendices.

III. Mathematical Preliminaries

In this section, we first introduce Theorem 1, which provides a deterministic equivalent for the Stieltjes transform of matrices $\mathbf{B}_N$ defined in (2). A deterministic equivalent for the Shannon transform of $\mathbf{B}_N$ is then provided in Theorem 2, before we discuss in detail how this last result can be used to characterize the performance of the MIMO-MAC with strong channel correlation patterns.

A. Main Results

Theorem 1: Let $K, N, n_1, \ldots, n_K$ be positive integers and let

$$\mathbf{B}_N = \sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k \mathbf{X}_k^{\mathsf{H}} \mathbf{R}_k^{\frac{1}{2}} + \mathbf{S} \tag{3}$$

be an $N \times N$ matrix with the following hypothesis for all $k \in \{1, \ldots, K\}$:

1) $\mathbf{X}_k \in \mathbb{C}^{N \times n_k}$ has i.i.d. entries $\frac{1}{\sqrt{n_k}} X_{ij}^k$, such that $\mathrm{E}[|X_{11}^k - \mathrm{E}X_{11}^k|^2] = 1$;
2) $\mathbf{R}_k^{\frac{1}{2}} \in \mathbb{C}^{N \times N}$ is a Hermitian nonnegative square root of the nonnegative definite Hermitian matrix $\mathbf{R}_k$;
3) $\mathbf{T}_k = \mathrm{diag}(\tau_1, \ldots, \tau_{n_k})$ with $\tau_i \geq 0$ for all i;
4) the sequences $\{F^{\mathbf{T}_k}\}_{n_k}$ and $\{F^{\mathbf{R}_k}\}_N$ are tight;[1]
5) $\mathbf{S} \in \mathbb{C}^{N \times N}$ is Hermitian nonnegative definite;
6) there exist $b > a > 0$ for which

$$a < \min_k \liminf_N c_k \leq \max_k \limsup_N c_k < b, \tag{4}$$

with $c_k = N/n_k$.

[1]this is, for all $\varepsilon > 0$, there exists $M > 0$ such that $F^{\mathbf{T}_k}(M) > 1 - \varepsilon$ and $F^{\mathbf{R}_k}(M) > 1 - \varepsilon$ for all n_k, N. See e.g., [20] for more details.

Also denote, for $z \in \mathbb{C} \setminus \mathbb{R}^+$, $m_N(z) = \int \frac{1}{\lambda - z} dF_N(\lambda)$, the Stieltjes transform of $\mathbf{B}_N$. Then, as all n_k and N grow large, with ratio c_k

$$m_N(z) - m_N^{\circ}(z) \xrightarrow{\text{a.s.}} 0 \tag{5}$$

where

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

and the functions $e_i(z)$, $i \in \{1, \ldots, K\}$, form the unique solution to the K equations

$$e_i(z) = \frac{1}{N} \mathrm{tr}\, \mathbf{R}_i \left(\mathbf{S} + \sum_{k=1}^{K} \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k \tau_k e_k(z)} \mathbf{R}_k - z \mathbf{I}_N \right)^{-1} \tag{6}$$

such that $\mathrm{sgn}(\Im[e_i(z)]) = \mathrm{sgn}(\Im[z])$ when $\Im[z] \neq 0$ and such that $e_i(z) > 0$ when z is real and negative.

Moreover, for any $\varepsilon > 0$, the convergence of (5) is uniform over any region of $\mathbb{C}$ bounded by a contour interior to

$$\mathbb{C} \setminus (\{z : |z| \leq \varepsilon\} \cup \{z = x + iv : x > 0, |v| \leq \varepsilon\}). \tag{7}$$

For all N, the function m_N° is the Stieltjes transform of a distribution function F_N°. Denoting F_N the empirical eigenvalue distribution function of $\mathbf{B}_N$, we finally have

$$F_N - F_N^{\circ} \Rightarrow 0 \tag{8}$$

weakly and almost surely as $N \to \infty$.

Proof: The proof of Theorem 1 is deferred to Appendix A. ∎

A few technical remarks are of order at his point.

Remark 1: In [21], Zhang derives an expression of the limiting eigenvalue distribution for the simpler case where $K = 1$ and $\mathbf{S} = 0$ but $\mathbf{T}_1$ is not constrained to be diagonal. Her work also uses a method based on the Stieltjes transform. Based on [21], it seems to the authors that Theorem 1 could well be extended to non-diagonal $\mathbf{T}_k$. However, proving so requires involved calculus, which we did not perform. Also, in [22], using the same techniques as in the proof provided in Appendix A, Silverstein *et al.* do not assume that the matrices $\mathbf{T}_k$ are nonnegative definite. Our result could be extended to this less stringent requirement on the central $\mathbf{T}_k$ matrices, although in this case Theorem 1 does not hold for z real negative. For application purposes though, it is fundamental that the Stieltjes transform of $\mathbf{B}_N$ exist for $z \in \mathbb{R}^-$.

We now claim that, under proper initialization, for $z \in \mathbb{C} \setminus \mathbb{R}^+$, a classical fixed-point algorithm converges surely to the solution of (6).

Proposition 1: For $z \in \mathbb{C} \setminus \mathbb{R}^+$, the output $\{e_1^n, \ldots, e_K^n\}$ of fixed-point algorithm described in Table I converges surely to the unique solution $\{e_1(z), \ldots, e_K(z)\}$ of (6), such that $\mathrm{sgn}(\Im[e_i(z)]) = \mathrm{sgn}(\Im[z])$ when $\Im[z] \neq 0$ and such that $e_i(z) > 0$ when $z < 0$, for all i.

Proof: The proof of Proposition 1, inspired by the work of Dupuy *et al.* [13] in the context of frequency-selective channel models, is provided in Appendix E. ∎

TABLE I
FIXED-POINT ALGORITHM CONVERGING TO THE SOLUTION OF (6)

Define $\varepsilon > 0$, the convergence threshold and $n \geq 0$, the iteration step. For all $k \in \{1, \ldots, K\}$, set $e_j^0 = -1/z$ and $e_j^{-1} = \infty$.
while $\max_j \{|e_j^n - e_j^{n-1}|\} > \varepsilon$ **do**
  for $k \in \{1, \ldots, K\}$ **do**
    Compute

$$e_k^{n+1} = \frac{1}{N} \operatorname{tr} \mathbf{R}_k \left(\mathbf{S} + \sum_{j=1}^{K} \int \frac{\tau_j dF^{\mathbf{T}_j}(\tau_j)}{1 + c_j \tau_j e_j^n} \mathbf{R}_j - z\mathbf{I}_N \right)^{-1} \tag{9}$$

  end for
  Assign $n \leftarrow n + 1$
end while

Different hypotheses will be used in the applications of Theorem 1 provided in Section IV. For practical reasons, we will in particular need the entries of $\mathbf{X}_k$ will be Gaussian, the matrices $\mathbf{T}_k$ to be non-diagonal and $\mathbf{S} = 0$. This entails the following corollary:

Corollary 1: Let $K, N, n_1, \ldots, n_K$ be positive integers and let

$$\mathbf{B}_N = \sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k \mathbf{X}_k^{\mathsf{H}} \mathbf{R}_k^{\frac{1}{2}} \tag{10}$$

be an $N \times N$ matrix with the following hypothesis for all $k \in \{1, \ldots, K\}$:

1) $\mathbf{X}_k \in \mathbb{C}^{N \times n_k}$ has i.i.d. Gaussian entries $\frac{1}{\sqrt{n_k}} X_{ij}^k$, with $\mathrm{E}[X_{11}^k] = 0$ and $\mathrm{E}[|X_{11}^k|^2] = 1$;
2) $\mathbf{R}_k^{\frac{1}{2}} \in \mathbb{C}^{N \times N}$ is a Hermitian nonnegative square root of the nonnegative definite Hermitian matrix $\mathbf{R}_k$;
3) $\mathbf{T}_k \in \mathbb{C}^{n_k \times n_k}$ is a nonnegative definite Hermitian matrix;
4) $\{F^{\mathbf{T}_k}\}_{n_k}$ and $\{F^{\mathbf{R}_k}\}_N$ form tight sequences;
5) there exist $b > a > 0$ for which

$$a < \min_k \liminf_N c_k \leq \max_k \limsup_N c_k < b \tag{11}$$

with $c_k = N/n_k$.

Also denote, for $x > 0$, $m_N(-x) = \frac{1}{N} \operatorname{tr}(\mathbf{B}_N + x\mathbf{I}_N)^{-1}$. Then, as all N and n_k grow large (while K is fixed)

$$m_N(-x) - m_N^{\circ}(-x) \xrightarrow{\text{a.s.}} 0$$

where

$$m_N^{\circ}(-x) = \frac{1}{N} \operatorname{tr} \left(x \left[\mathbf{I}_N + \sum_{k=1}^{K} \delta_k(-x) \mathbf{R}_k \right] \right)^{-1}$$

and the set of functions $\{e_i(-x), \delta_i(-x)\}$, $i \in \{1, \ldots, K\}$, form the unique solution to the equations

$$e_i(-x) = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(x \left[\mathbf{I}_N + \sum_{k=1}^{K} \delta_k(-x) \mathbf{R}_k \right] \right)^{-1}$$

$$\delta_i(-x) = \frac{1}{n_i} \operatorname{tr} \mathbf{T}_i \left(x \left[\mathbf{I}_{n_i} + c_i e_i(-x) \mathbf{T}_i \right] \right)^{-1}$$

such that $e_i(-x) > 0$ for all i.

Proof: Since the $\mathbf{X}_k$ are Gaussian, the joint distribution of the entries of $\mathbf{X}_k \mathbf{U}$ coincides with that of $\mathbf{X}_k$, for $\mathbf{U}$ any $n_k \times n_k$ unitary matrix. Therefore, $\mathbf{X}_k \mathbf{T}_k \mathbf{X}_k^{\mathsf{H}}$ in Theorem 1 can be substituted by $\mathbf{X}_k (\mathbf{U}\mathbf{T}_k\mathbf{U}^{\mathsf{H}}) \mathbf{X}_k^{\mathsf{H}}$ without compromising the final result. As a consequence, the $\mathbf{T}_k$ can be taken nonnegative definite Hermitian and the result of Theorem 1 holds. It then suffices to replace $\delta_i(-x)$ in the expression of $e_i(-x)$ to fall back on the result of Theorem 1. ∎

The deterministic equivalent of the Stieltjes transform m_N of $\mathbf{B}_N$ is then extended to a deterministic equivalent of the Shannon transform of $\mathbf{B}_N$ in the following result.

Theorem 2: Let $x > 0$ and $\mathbf{B}_N$ be a random Hermitian matrix as defined in Corollary 1 with the following additional assumptions:

1) there exists $\alpha > 0$ and a sequence $r_1, r_2, \ldots$, such that, for all N,

$$\max_{1 \leq k \leq K} \max(\lambda_{r_N+1}^{\mathbf{T}_k}, \lambda_{r_N+1}^{\mathbf{R}_k}) \leq \alpha,$$

where $\lambda_1^{\mathbf{Z}} \geq \cdots \geq \lambda_N^{\mathbf{Z}}$ denote the ordered eigenvalues of the $N \times N$ matrix $\mathbf{Z}$.
2) denoting b_N an upper-bound on the spectral norm of the $\mathbf{T}_k$ and $\mathbf{R}_k$, $k \in \{1, \ldots, K\}$, and $\beta > 0$ a constant such that $\beta > \frac{Kb}{a}(1 + \sqrt{a})^2$, $a_N = b_N^2 \beta$ satisfies

$$r_N \log(1 + a_N/x) = o(N).$$

Then, for large N, n_k, the Shannon transform $\mathcal{V}_N(x) = \int \log(1 + \frac{1}{x}\lambda) dF_N(\lambda)$ of $\mathbf{B}_N$, satisfies

$$\mathcal{V}_N(x) - \mathcal{V}_N^{\circ}(x) \xrightarrow{\text{a.s.}} 0$$

where

$$\mathcal{V}_N^{\circ}(x) = \sum_{k=1}^{K} \frac{1}{N} \log\det\left(\mathbf{I}_{n_k} + c_k e_k(-x) \mathbf{T}_k\right) + \frac{1}{N} \log\det\left(\mathbf{I}_N + \sum_{k=1}^{K} \delta_k(-x) \mathbf{R}_k\right) - x \sum_{k=1}^{K} \delta_k(-x) e_k(-x). \tag{12}$$

Proof: The proof of Theorem 2 is provided in Appendix B. ∎

Note that this last result is consistent both with [6] when the transmission channels are i.i.d. Gaussian and with [23] when $K = 2$. This result is also similar in nature to the expressions

obtained in [5] for the multi-antenna Rician channel model and with [11] in the case of frequency-selective channels. We point out that the expressions obtained in [11], [13] and [9], when the entries of the $\mathbf{X}_k$ matrices are Gaussian distributed, suggest a faster convergence rate of the deterministic equivalent of the Stieltjes and Shannon transforms than the one obtained here. Indeed, while we show here a convergence of order $o(1)$ (which is in fact refined to $o(1/\log^k N)$ for any k in Appendix A), in those works the convergence is proved to be of order $O(1/N^2)$.

However, contrary to the above contributions, we allow the $\mathbf{R}_k$ and $\mathbf{T}_k$ matrices to be more general than uniformly bounded in spectral norm. This is thoroughly discussed in the section below.

B. Kronecker Channel With Strong Correlation Patterns

Theorem 1 and Corollary 1 require $\{F^{\mathbf{R}_k}\}_N$ and $\{F^{\mathbf{T}_k}\}_{n_k}$ to form tight sequences. Remark that, because of the trace constraint $\frac{1}{N}\operatorname{tr}\mathbf{R}_k = 1$, all sequences $\{F^{\mathbf{R}_k}\}_N$ are necessarily tight (the same reasoning naturally holds for $\mathbf{T}_k$). Indeed, given $\varepsilon > 0$, take $M = 2/\varepsilon$; $N[1 - F^{\mathbf{R}_k}(M)]$ is the number of eigenvalues in $\mathbf{R}_k$ larger than $2/\varepsilon$, which is necessarily less than or equal to $N\varepsilon/2$ from the trace constraint, leading to $1 - F^{\mathbf{R}_k}(M) \leq \varepsilon/2$ and then $F^{\mathbf{R}_k}(M) \geq 1 - \varepsilon/2 > 1 - \varepsilon$. The same naturally holds for the $\mathbf{T}_k$ matrices. Observe now that Condition 2) in Theorem 2 requires a stronger assumption on the correlation matrices. Under the trace constraint, a sufficient assumption for Condition 2) is that there exists $\alpha > 0$, such that the number of eigenvalues in $\mathbf{R}_k$ greater than α is of order $o(N/\log N)$. This is a mild assumption, which may not be verified for some very specific choices of $\{F^{\mathbf{R}_k}\}_N$.[2] Nonetheless, most conventional models for the $\mathbf{R}_k$ and $\mathbf{T}_k$, even when showing strong correlation properties, satisfy the assumptions of Theorem 2. We mention in particular the following examples:

- if all $\mathbf{R}_k$ and $\mathbf{T}_k$ have uniformly bounded spectral norm, then there exists $\alpha > 0$ such that all eigenvalues of $\mathbf{R}_k$ and $\mathbf{T}_k$ are less then α for all N. This implies $r_N = 0$ for all N and therefore the condition is trivially satisfied. Our model is therefore compatible with loosely correlated antenna structures;
- in contrast, when antennas are densely packed on a volume-limited device, the correlation matrices $\mathbf{R}_k$ and $\mathbf{T}_k$ tend to be asymptotically of finite rank, see e.g., [24] in the case of a dense circular array. That is, for any given $\alpha > 0$, for all large N, the number r_N of eigenvalues greater than α is finite, while a_N defined in Theorem 2 is of order N^2. This implies $r_N \log(1 + a_N/x) = O(\log N) = o(N)$ and therefore volume-limited devices with densely packed antennas are consistent with our framework;
- for 1-D, 2-D, and 3-D antenna arrays with neighbors separated by half the wavelength as discussed by Moustakas *et al.* in [25], the correlation figures have a peculiar behaviour. In a linear array of antenna, $O(N)$ eigenvalues are of order of magnitude $O(1)$, the remaining eigenvalues being small. In a 2-D grid of antennas, $O(\sqrt{N})$ eigenvalues are of order $O(\sqrt{N})$, the remaining eigenvalues being close to zero. Finally, in a 3-D parallelepiped of antennas, $O(N^{\frac{2}{3}})$ eigenvalues are of order $O(N^{\frac{1}{3}})$, the remaining eigenvalues being close to 0 also. As such, in the p-dimensional scenario, we can approximate r_N by $N^{\frac{p-1}{p}}$, a_N by $N^{\frac{2}{p}}$ and we have

$$r_N \log(1 + a_N/x) \sim N^{\frac{p-1}{p}} \log N = o(N)$$

  so that the multidimensional antenna arrays with close antennas separated by half the wavelength also satisfy the hypotheses of Theorem 2.

As a consequence, a wide scope of antenna correlation models enter our deterministic equivalent framework, which comes again at the price of a slower theoretical convergence of the difference $\mathcal{V}_N - \mathcal{V}_N^\circ$.

We now move to practical applications of the above results and more specifically to the determination of the ergodic rate region of the MIMO-MAC.

IV. Rate Region of the MIMO-MAC

In this section, we successively apply Theorem 2 to approximate the ergodic mutual information for all deterministic precoders, and then we determine the precoders that maximize this approximated mutual information. This gives an approximation of all points on the boundary of the MIMO-MAC rate region. We also introduce an iterative power allocation algorithm to obtain explicitly the optimal precoders.

A. Deterministic Equivalent of the Mutual Information

Consider the wireless multiple access channel as described in Section II and depicted in Fig. 1. We denote $c_k = N/n_k$ the ratio between the number of antennas at the receive base station and the number of transmit antennas of user k. Denote $\mathbf{s}_k \in \mathbb{C}^{n_k}$ the Gaussian signal transmitted by user k, such that $\mathrm{E}[\mathbf{s}_k] = 0$ and $\mathrm{E}[\mathbf{s}_k\mathbf{s}_k^{\mathsf{H}}] = \mathbf{P}_k$, with $\frac{1}{n_k}\operatorname{tr}\mathbf{P}_k \leq P_k$ where P_k is the total power of transmitter k, $\mathbf{y} \in \mathbb{C}^N$ the signal received at the base station and $\mathbf{n}$ the additive white Gaussian noise of variance $\mathrm{E}[\mathbf{n}\mathbf{n}^{\mathsf{H}}] = \sigma^2\mathbf{I}_N$. We recall that the Kronecker channel between user k and the base station is denoted $\mathbf{H}_k = \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k^{\frac{1}{2}}$, with the entries of $\mathbf{X}_k \in \mathbb{C}^{N\times n_k}$ Gaussian independent of zero mean and variance $1/n_k$ and $\mathbf{R}_k$, $\mathbf{T}_k$ deterministic. The received vector $\mathbf{y}$ is therefore given by

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

Suppose that the $\mathbf{H}_k$ channels are varying fast and that the transmitters in the MAC only have statistical channel state information about the $\mathbf{H}_k$ in the sense that user k only knows the long term statistics $\mathbf{R}_1, \ldots, \mathbf{R}_K$ and $\mathbf{T}_1, \ldots, \mathbf{T}_K$. In this

[2] As a counter-example, take $N = 2^p + 1$ and the eigenvalues of $\mathbf{R}_k$ to be

$$2^{p-1}, \underbrace{p, \ldots, p}_{\frac{2^p-1}{p}}, \underbrace{0, \ldots, 0}_{2^p - \frac{2^p-1}{p}}.$$

The largest eigenvalue is of order N so that a_N is of order N^2, and the number r_N of eigenvalues larger than any $\alpha > 0$ for N large is of order $\frac{2^{p-1}}{p} \sim \frac{N}{\log(N)}$. Therefore $r_N \log(1 + a_N/x) = O(N)$ here.

case, for a noise variance equal to σ^2, the per-antenna ergodic MIMO-MAC rate region $\mathcal{C}_{\rm MAC}$ is given by [26]

$$\mathcal{C}_{\rm MAC} = \bigcup_{\substack{\frac{1}{n_i}\operatorname{tr}(\mathbf{P}_i)\leq P_i \\ \mathbf{P}_i\geq 0 \\ i=1,\ldots,K}} \left\{ \{R_i, 1\leq i\leq K\} : \sum_{i\in\mathcal{S}} R_i \leq \mathrm{E}\mathcal{V}_N(\mathbf{P}_{i_1},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}};\sigma^2), \forall \mathcal{S}\subset\{1,\ldots,K\} \right\}$$

with the expectation taken over the joint random variable $(\mathbf{X}_1,\ldots,\mathbf{X}_K)$, $\mathcal{S}=\{i_1,\ldots,i_{|\mathcal{S}|}\}$, and where we introduced the notation

$$\mathcal{V}_N(\mathbf{P}_{i_1},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}};\sigma^2) \triangleq \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\sum_{i\in\mathcal{S}}\mathbf{H}_i\mathbf{P}_i\mathbf{H}_i^{\mathsf{H}}\right).$$

Now, assuming the $\mathbf{T}_k$, $\mathbf{P}_k$ and $\mathbf{R}_k$ satisfy the hypotheses of Theorem 2, we have

$$\mathcal{V}_N(\mathbf{P}_{i_1},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}};x) - \mathcal{V}_N^{\circ}(\mathbf{P}_{i_1},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}};x)\to 0$$

as $N, n_{i_1},\ldots,n_{i_{|\mathcal{S}|}}$ grow large for some sequence $\{\mathcal{V}_N^{\circ}(\mathbf{P}_{i_1},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}};x)\}_N$, on a subset of measure 1 of the probability space Ω that engenders $(\mathbf{X}_{i_1},\ldots,\mathbf{X}_{i_{|\mathcal{S}|}})$. Integrating this expression over Ω therefore leads to

$$\mathrm{E}\mathcal{V}_N^{\circ}(\mathbf{P}_{i_1},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}};x) - \mathcal{V}_N^{\circ}(\mathbf{P}_{i_1},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}};x)\to 0.$$

We can therefore apply Theorem 2 to determine the ergodic rate region $\mathcal{C}_{\rm MAC}$ of the MIMO-MAC. We specifically have

$$\begin{aligned}\mathrm{E}\mathcal{V}_N(\mathbf{P}_{i_1},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}};\sigma^2) - \Bigg[&\sum_{k\in\mathcal{S}}\frac{1}{N}\log\det\left(\mathbf{I}_{n_k}+c_k e_k(-\sigma^2)\mathbf{T}_k\mathbf{P}_k\right) \\ &+\frac{1}{N}\log\det\left(\mathbf{I}_N+\sum_{k\in\mathcal{S}}\delta_k(-\sigma^2)\mathbf{R}_k\right) \\ &-\sigma^2\sum_{k\in\mathcal{S}}\delta_k(-\sigma^2)e_k(-\sigma^2)\Bigg]\to 0 \end{aligned} \tag{13}$$

with $e_i(-\sigma^2)$ and $\delta_i(-\sigma^2)$ the unique positive solutions of

$$\begin{aligned} e_i(-\sigma^2) &= \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\sigma^2\left[\mathbf{I}_N+\sum_{k\in\mathcal{S}}\delta_k(-\sigma^2)\mathbf{R}_k\right]\right)^{-1} \\ \delta_i(-\sigma^2) &= \frac{1}{n_i}\operatorname{tr}\mathbf{T}_i^{\frac{1}{2}}\mathbf{P}_i\mathbf{T}_i^{\frac{1}{2}}\left(\sigma^2\left[\mathbf{I}_{n_i}+c_i e_i(-\sigma^2)\mathbf{T}_i^{\frac{1}{2}}\mathbf{P}_i\mathbf{T}_i^{\frac{1}{2}}\right]\right)^{-1}. \end{aligned} \tag{14}$$

This provides a deterministic equivalent for all points in the MIMO-MAC rate region, i.e., for all $\mathbf{P}_1,\ldots,\mathbf{P}_K$ precoders.

B. Rate Maximization

Now we wish to determine for which precoders the boundary of the MIMO-MAC rate region is reached. This requires here to determine the rate optimal precoding matrices $\mathbf{P}_{i_1},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}$, for all $\mathcal{S}\subset\{1,\ldots,K\}$. To this end, we first need the following result.

Proposition 2: If at least one of the correlation matrices $\mathbf{T}_k$, $k\in\mathcal{S}$, is invertible, then $\mathcal{V}_N^{\circ}$ is a strictly concave function in $\mathbf{P}_{i_1},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}$.

Proof: The proof of Proposition 2 is provided in Appendix C. ∎

Without loss of generality, for any k, since the $\mathbf{X}_k$ matrices are standard Gaussian, and therefore of unitarily invariant joint distribution, $\mathbf{T}_k$ can be assumed diagonal. If $\mathbf{T}_k$ is not of full rank then it can be reduced into a matrix of smaller size, such that the resulting matrix is invertible, without changing the problem at hand. We therefore assume all $\mathbf{T}_k$ matrices to be of full rank from now on. From Proposition 2, we then immediately prove that the $|\mathcal{S}|$-ary set of matrices $\{\mathbf{P}_{i_1}^{\circ},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}^{\circ}\}$ which maximizes the deterministic equivalent of the ergodic sum rate over the set $\mathcal{S}$ is unique. In a very similar way as in [9], we then show that the matrices $\mathbf{P}_k^{\circ}$, $k\in\mathcal{S}$, have the following properties.

Proposition 3: For every $k\in\mathcal{S}$, denote $\mathbf{T}_k=\mathbf{U}_k\bar{\mathbf{T}}_k\mathbf{U}_k^{\mathsf{H}}$ the spectral decomposition of $\mathbf{T}_k$ with $\mathbf{U}_k$ unitary and $\bar{\mathbf{T}}_k=\operatorname{diag}(t_{k,1},\ldots,t_{k,n_k})$. Then the precoders $\mathbf{P}_{i_1}^{\circ},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}^{\circ}$ which maximize the right-hand side of (13) satisfy:

1) $\mathbf{P}_k^{\circ}=\mathbf{U}_k\bar{\mathbf{P}}_k^{\circ}\mathbf{U}_k^{\mathsf{H}}$, with $\bar{\mathbf{P}}_k^{\circ}$ diagonal, i.e., the eigenspace of $\mathbf{P}_k^{\circ}$ is the same as the eigenspace of $\mathbf{T}_k$;
2) denoting, for all k, $e_k^{\circ}=e_k(-\sigma^2)$ as in (14) for $\mathbf{P}_k=\mathbf{P}_k^{\circ}$, the i^{th} diagonal entry $p_{k,i}^{\circ}$ of $\bar{\mathbf{P}}_k^{\circ}$ satisfies

$$p_{k,i}^{\circ}=\left(\mu_k-\frac{1}{c_k e_k^{\circ} t_{ki}}\right)^+ \tag{15}$$

where the μ_k are evaluated such that $\frac{1}{n_k}\operatorname{tr}\mathbf{P}_k=P_k$.

In Table II, we provide an iterative water-filling algorithm to obtain the p_{ki}°.

Proof: The proof of Proposition 3 is provided in Appendix D. ∎

Remark 2: In [9], it is proved that the convergence of this algorithm implies its convergence towards the correct limit. The line of reasoning in [9] can be directly adapted to the current situation so that, if the iterative water-filling algorithm of Table II converges, then $\mathbf{P}_{i_1},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}$ converge to the matrices $\mathbf{P}_{i_1}^{\circ},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}^{\circ}$. However, similar to [9], it is difficult to prove the sure convergence of the water-filling algorithm. Nonetheless, extensive simulations suggest that convergence is always attained.

For the set $\mathcal{S}$ under consideration, denote now $\mathbf{P}_{i_1}^{\star},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}^{\star}$ the true sum rate maximizing precoders. Then, if $\mathbf{P}_{i_1}^{\circ},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}^{\circ}$ and $\mathbf{P}_{i_1}^{\star},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}^{\star}$ are such that Condition 1) of Theorem 2 is satisfied with the sets $\{\mathbf{T}_1,\ldots,\mathbf{T}_K\}$ and $\{\mathbf{R}_1,\ldots,\mathbf{R}_K\}$ replaced by $\{\mathbf{T}_{i_1}\mathbf{P}_{i_1}^{\circ},\ldots,\mathbf{T}_{i_{|\mathcal{S}|}}\mathbf{P}_{i_{|\mathcal{S}|}}^{\circ}\}$ (or $\{\mathbf{T}_{i_1}\mathbf{P}_{i_1}^{\star},\ldots,\mathbf{T}_{i_{|\mathcal{S}|}}\mathbf{P}_{i_{|\mathcal{S}|}}^{\star}\}$) and $\{\mathbf{R}_{i_1},\ldots,\mathbf{R}_{i_{|\mathcal{S}|}}\}$, respectively, we have from Theorem 2

$$\begin{aligned}&\mathcal{V}_N(\mathbf{P}_{i_1}^{\star},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}^{\star};\sigma^2)-\mathcal{V}_N(\mathbf{P}_{i_1}^{\circ},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}^{\circ};\sigma^2) \\ &=\left(\mathcal{V}_N(\mathbf{P}_{i_1}^{\star},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}^{\star};\sigma^2)-\mathcal{V}_N^{\circ}(\mathbf{P}_{i_1}^{\star},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}^{\star};\sigma^2)\right) \\ &\quad+\left(\mathcal{V}_N^{\circ}(\mathbf{P}_{i_1}^{\star},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}^{\star};\sigma^2)-\mathcal{V}_N^{\circ}(\mathbf{P}_{i_1}^{\circ},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}^{\circ};\sigma^2)\right) \\ &\quad+\left(\mathcal{V}_N^{\circ}(\mathbf{P}_{i_1}^{\circ},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}^{\circ};\sigma^2)-\mathcal{V}_N(\mathbf{P}_{i_1}^{\circ},\ldots,\mathbf{P}_{i_{|\mathcal{S}|}}^{\circ};\sigma^2)\right)\end{aligned}$$

TABLE II
ITERATIVE WATER-FILLING ALGORITHM

Define $\eta > 0$ the convergence threshold and $l \geq 0$ the iteration step. At step $l = 0$, for all $k \in \mathcal{S}$, $i \in \{1, \ldots, n_k\}$, set $p_{k,i}^0 = P_k$.
while $\max_{k,i}\{|p_{k,i}^l - p_{k,i}^{l-1}|\} > \eta$ **do**
  For $k \in \mathcal{S}$, define e_k^{l+1} as the solution of (6) for $z = -\sigma^2$ and $\mathbf{P}_k$ with eigenvalues $p_{k,1}^l, \ldots, p_{k,n_k}^l$, obtained from the fixed-point algorithm of Table I.
  for $k \in \mathcal{S}$ **do**
    for $i = 1 \ldots, n_k$ **do**
      Set $p_{k,i}^{l+1} = \left(\mu_k - \frac{1}{c_k e_k^{l+1} t_{ki}}\right)^+$, with μ_k such that $\frac{1}{n_k} \operatorname{tr} \mathbf{P}_k = P_k$.
    end for
  end for
  Assign $l \leftarrow l + 1$
end while

where both right-hand side differences of the type $\mathcal{V}_N - \mathcal{V}_N^\circ$ tend to zero, while the left-hand side term is positive by definition of $\mathbf{P}_k^\star$ and the remaining right-hand side term is negative by definition of the $\mathbf{P}_k^\circ$. This finally ensures that

$$\mathcal{V}_N(\mathbf{P}_{i_1}^\star, \ldots, \mathbf{P}_{i_{|\mathcal{S}|}}^\star; \sigma^2) - \mathcal{V}_N(\mathbf{P}_{i_1}^\circ, \ldots, \mathbf{P}_{i_{|\mathcal{S}|}}^\circ; \sigma^2) \to 0$$

as $N, n_{i_1}, \ldots, n_{i_{|\mathcal{S}|}}$ grow large with uniformly bounded ratios. Therefore, the mutual information obtained based on the precoders $\mathbf{P}_{i_1}^\circ, \ldots, \mathbf{P}_{i_{|\mathcal{S}|}}^\circ$ is asymptotically close to the capacity achieved with the ideal precoders $\mathbf{P}_{i_1}^\star, \ldots, \mathbf{P}_{i_{|\mathcal{S}|}}^\star$. Finally, if, for all sets $\mathcal{S} \subset \{1, \ldots, K\}$, the matrices $\mathbf{T}_k$, $\mathbf{R}_k$ and the resulting $\mathbf{P}_k^\star, \mathbf{P}_k^\circ$ satisfy the mild conditions of Theorem 2, then all points of the boundary of the MIMO-MAC rate region can be given a deterministic equivalent.

This concludes this application section. In the following section, we provide simulation results that confirm the accuracy of the deterministic equivalents as well as the validity of the hypotheses made on the $\mathbf{T}_k$ and $\mathbf{R}_k$ matrices.

V. SIMULATIONS AND RESULTS

In the following, we apply the results obtained in Section IV to provide comparative simulation results between ergodic rate regions, sum rates and their respective deterministic equivalents, for non negligible channel correlations on both communication sides. We provide simulation results in the context of a two-user MIMO-MAC, with N antennas at the base station and $n_1 = n_2$ antennas at the user terminals. The antennas are placed on a possibly multidimensional array, antenna i being located at $\mathbf{x}_i \in \mathbb{R}^3$. We further assume that both terminals are physically identical. To model the transmit and receive correlation matrices, we consider both the effect of the distances between adjacent antennas at the user terminals and at the base station, and the effect of the solid angles of effective energy transmission and reception. We assume a channel model where signals are transmitted and received isotropically in the vertical direction, but transmitted and received under an angle π in the horizontal direction. We then model the entries of the correlation matrices from a natural extension of Jakes model [27] with privileged direction of signal departure and arrival. Denoting λ the transmit signal wavelength, $T_{1_{ab}}$, the entry (a, b) of the matrix $\mathbf{T}_1$, is

$$T_{1_{ab}} = \int_{\theta_{\min}^{(\mathbf{T}_1)}}^{\theta_{\max,}^{(\mathbf{T}_1)}} \exp\left(\frac{2\pi i}{\lambda} ||\mathbf{x}_a - \mathbf{x}_b|| \cos(\theta)\right) d\theta$$

with $[\theta_{\min}^{(\mathbf{T}_1)}, \theta_{\max}^{(\mathbf{T}_1)}]$ the effective horizontal directions of signal propagation. With similar notations for the other correlation matrices, we choose $\theta_{\min}^{(\mathbf{T}_1)} = 0$, $\theta_{\max}^{(\mathbf{T}_1)} = \pi$, $\theta_{\min}^{(\mathbf{T}_2)} = \pi/3$, $\theta_{\max}^{(\mathbf{T}_2)} = 4\pi/3$, $\theta_{\min}^{(\mathbf{R}_1)} = 2\pi/3$, $\theta_{\max}^{(\mathbf{R}_1)} = -2\pi/3$, $\theta_{\min}^{(\mathbf{R}_2)} = \pi$ and $\theta_{\max}^{(\mathbf{R}_2)} = 0$.

We start by simulating the MIMO-MAC rate region obtained when $n_1 = n_2 = N$, either for $N = 2$, $N = 4$ or $N = 8$, for a linear antenna array with distance $\lambda/10$ between close antennas, under identity or optimal precoding policies and with signal to noise ratio of 20 dB or -10 dB. Simulation results, averaged over 10,000 channel realizations are compared with the deterministic equivalents. This is depicted in Fig. 2. The deterministic equivalents of the rate regions appear to approximate the true rate regions extremely well, even for very small system dimensions. The case $N = 8$ shows in particular a perfect match. We note that increasing the number of antennas on both communication side provides a greater gain when using the optimal precoding policy. We also observe that, while increasing the number of antennas tends to reduce the individual per-antenna rate under uniform power allocation, using an optimal precoding policy significantly increases the per-antenna rate. This phenomenon is particularly accentuated in the low SNR regime. This confirms the observations made by Vishwanath *et al.* in [28], according to which the efficiency of individual antennas can grow as the correlation grows at low SNR. This is, for a given number of antennas, by increasing correlation and systematically applying optimal precoding, strong eigenmodes emerge over which data can be directed. This leads to higher rates than for an uncorrelated antenna array for which there are no such strong eigenmodes.

In order to test the robustness of the proposed deterministic equivalents to strongly correlated channel conditions, we then compare in Fig. 3 the MIMO-MAC ergodic sum rate to the associated deterministic equivalents when precoder optimization is performed or not and when the antenna grids are either 1-D regular arrays or 3-D regular cubes. In all situations, the distance between neighboring antennas is half the wavelength and the signal to noise ratio is taken to be 20 dB. We take the number of antennas on both sides to be successively 1, 8, 27, 64, and 125. We first observe that the deterministic equivalents are extremely accurate in this scenario. We confirm also the behaviour of the antenna efficiency, which saturates for the 1-D array and

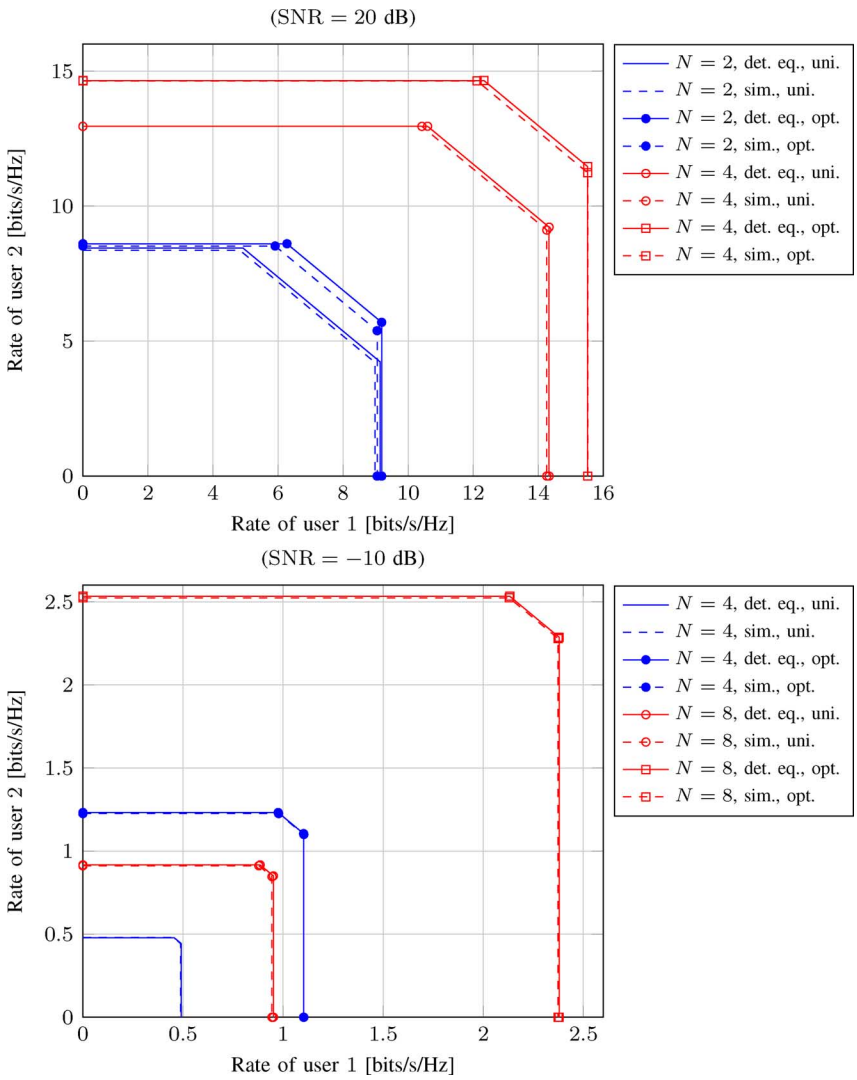

Fig. 2. Rate region for a two-user MIMO-MAC when the antennas are placed on a linear array with distance between close antennas is $\lambda/10$. The number of antennas $N = n_1 = n_2$ is taken equal to 2, 4, or 8. Simulations (sim.) are compared against deterministic equivalents (det. eq.). Uniform power allocation across transmit antennas (uni.) as well as optimal sum rate maximizing precoding (opt.) are considered. The SNR is 20 dB in the top figure and -10 dB in the bottom figure.

decreases fast for the 3-D antenna array, similar to what was observed in [25] for the single-user case. In terms of performance, a large improvement of the achievable sum rate is observed, especially in the 3-D case, when the optimal precoding policy is applied. Using optimal precoding policies can therefore significantly reduce the negative impact of antenna correlation even in this high SNR regime.

VI. Conclusion

In this article, we analyzed the performance of multi-antenna multi-user wireless communications and more particularly the multi-antenna multiple access channel, while taking into account the correlation effects due to close antennas and reduced solid angles of energy transmission and reception. The analytic approach is based on novel results, based on recent tools from the field of large dimensional random matrix theory. From these tools, we provide on the one hand a deterministic equivalent of the per-antenna mutual information of the MIMO-MAC for arbitrary precoders under possibly strongly correlated channel conditions and on the other hand an approximation of the rate maximizing precoders along with an iterative water-filling algorithm to compute these precoders. In particular, while theoretical results prove the asymptotic accuracy of our model in the case of dense antenna packing or multidimensional antenna array on either communication side, simulations concur and suggest that the deterministic equivalents are moreover extremely accurate for very small system dimensions. These results can be used both from a practical side to easily derive optimal precoders and from a theoretical side to quantify the gains achieved by optimal power allocation policies in strongly correlated MIMO channels.

Appendix A
Proof of Theorem 1

For ease of read, the proof will be divided into several sections.

We first consider the case $K = 1$, whose generalization to $K \geq 1$ is given in Appendix A-E. Therefore, in the coming sections, we drop the useless indexes.

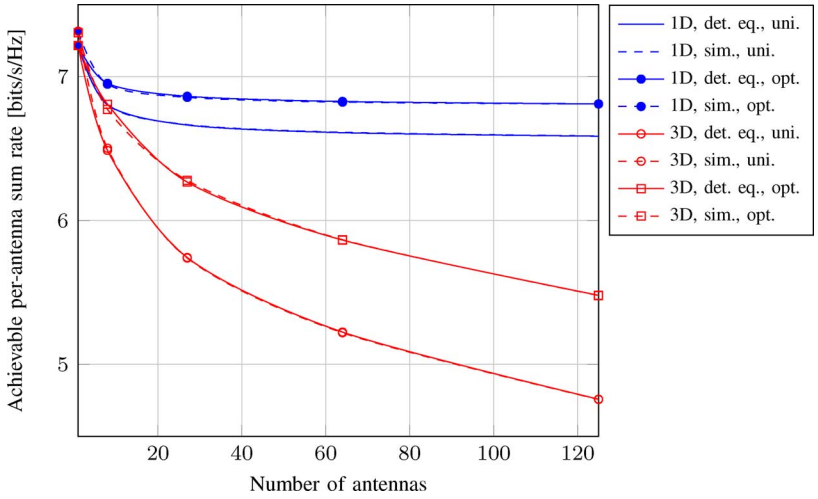

Fig. 3. Per-antenna sum rate for a two-user MIMO-MAC when the antennas are placed on a line (1-D) or a cubic array (3-D). The number of antennas satisfy $N = n_1 = n_2$ and range from 1 to 125. Simulations (sim.) are compared against deterministic equivalents (det. eq.). The distance between close antennas is $\lambda/2$. The SNR is 20 dB. Uniform power allocation (uni.) as well as optimal sum rate maximizing precoding (opt.) are considered.

A) Truncation and Centralization: We begin with the truncation and centralization steps which will replace $\mathbf{X}$, $\mathbf{R}$ and $\mathbf{T}$ by matrices with bounded entries, more suitable for analysis; the difference of the Stieltjes transforms of the original and new $\mathbf{B}_N$ converging to zero. Since vague convergence of distribution functions is equivalent to the convergence of their Stieltjes transforms, it is sufficient to show the original and new empirical distribution functions of the eigenvalues approach each other almost surely in the space of subprobability measures on $\mathbb{R}$ with respect to the topology which yields vague convergence.

Let $\widetilde{X}_{ij} = X_{ij}\mathbf{1}_{\{|X_{ij}|<\sqrt{N}\}} - \mathrm{E}(X_{ij}\mathbf{1}_{\{|X_{ij}|<\sqrt{N}\}})$ and $\widetilde{\mathbf{X}} = \left(\frac{1}{\sqrt{n}}\widetilde{X}_{ij}\right)$. Then, from c), Lemma 1 and a), Lemma 3, it follows exactly as in the initial truncation and centralization steps in [22] and [29] (which provide more details in their appendices), that

$$\|F_N - F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}}\widetilde{\mathbf{X}}\mathbf{T}\widetilde{\mathbf{X}}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}}\| \xrightarrow{\text{a.s.}} 0$$

as $N \to \infty$.

Let now $\overline{X}_{ij} = \widetilde{X}_{ij} \cdot \mathbf{1}_{\{|X_{ij}|<\ln N\}} - \mathrm{E}(\widetilde{X}_{ij}\mathbf{1}_{I\{|X_{ij}|<\ln N\}})$ and $\overline{\mathbf{X}} = \left(\frac{1}{\sqrt{n}}\overline{X}_{ij}\right)$. This is the final truncation and centralization step, which will be practically handled the same way as in [22], which some minor modifications, given presently.

For any Hermitian non-negative definite $r \times r$ matrix $\mathbf{A}$, let $\lambda_i^{\mathbf{A}}$ denote its i-th smallest eigenvalue of $\mathbf{A}$. With $\mathbf{A} = \mathbf{U}\,\mathrm{diag}(\lambda_1^{\mathbf{A}}, \ldots, \lambda_r^{\mathbf{A}})\mathbf{U}^{\mathsf{H}}$ its spectral decomposition, let for any $\alpha > 0$

$$\mathbf{A}^{\alpha} = \mathbf{U}\,\mathrm{diag}(\lambda_1^{\mathbf{A}}\mathbf{1}_{\{\lambda_r^{\mathbf{A}}\le\alpha\}}, \ldots, \lambda_r^{\mathbf{A}}\mathbf{1}_{\{\lambda_1\le\alpha\}})\mathbf{U}^{\mathsf{H}}.$$

Then for any $N \times N$ matrix $\mathbf{Q}$, we get from 1) and 2), Lemma 3

$$\begin{aligned}
&\|F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}}\mathbf{Q}\mathbf{T}\mathbf{Q}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}} - F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}\alpha}Q\mathbf{T}^{\alpha}\mathbf{Q}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}\alpha}}\| \\
&\le \frac{2}{N}\mathrm{rank}(\mathbf{R}^{\frac{1}{2}} - \mathbf{R}^{\frac{1}{2}\alpha}) + \frac{1}{N}\mathrm{rank}(\mathbf{T} - \mathbf{T}^{\alpha}) \\
&= \frac{2}{N}\sum_{i=1}^{N}\mathbf{1}_{\{\lambda_i^{\mathbf{R}}>\alpha\}} + \frac{1}{N}\sum_{i=1}^{n}\mathbf{1}_{\{\lambda_i^{\mathbf{T}}>\alpha)\}} \\
&= 2F^{\mathbf{R}}((\alpha,\infty)) + \frac{1}{c_N}F^{\mathbf{T}}((\alpha,\infty)).
\end{aligned}$$

Therefore, from the assumptions 4) and 6) in Theorem 1, we have for any sequence $\{\alpha_N\}$ with $\alpha_N \to \infty$

$$\|F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}}\mathbf{Q}\mathbf{T}\mathbf{Q}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}} - F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}\alpha_N}\mathbf{Q}\mathbf{T}^{\alpha_N}\mathbf{Q}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}\alpha_N}}\| \to 0 \quad (16)$$

as $N \to \infty$.

A metric D on probability measures defined on $\mathbb{R}$, which induces the topology of vague convergence, is introduced in [22] to handle the last truncation step. The matrices studied in [22] are essentially $\mathbf{B}_N$ with $\mathbf{R} = \mathbf{I}_N$. Following the steps beginning at [22, (3.4)], we see in our case that when α_N is chosen so that as $N \to \infty$, $\alpha_N \uparrow \infty$

$$\begin{gathered}
\alpha_N^8(\mathrm{E}|X_{11}^2\mathbf{1}_{\{X_{11}|\ge\ln N\}} + N^{-1}) \to 0 \\
\text{and} \\
\sum_{N=1}^{\infty}\frac{\alpha_N^{16}}{N^2}\left(\mathrm{E}|X_{11}|^4\mathbf{1}_{\{|X_{11}|<\sqrt{N})\}} + 1\right) < \infty.
\end{gathered}$$

We will get

$$D(F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}\alpha_N}\widetilde{\mathbf{X}}\mathbf{T}^{\alpha_N}\widetilde{\mathbf{X}}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}\alpha_N}}, F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}\alpha_N}\overline{\mathbf{X}}\mathbf{T}^{\alpha_N}\overline{\mathbf{X}}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}\alpha_N}}) \xrightarrow{\text{a.s.}} 0 \quad (17)$$

as $N \to \infty$.

Since $\mathrm{E}|\overline{X}_{11}|^2 \to 1$ as $N \to \infty$ we can rescale and replace $\overline{\mathbf{X}}$ with $\overline{\mathbf{X}}/\sqrt{\mathrm{E}|\overline{X}_{11}|^2}$, whose components are bounded by $k\ln N$ for some $k > 2$. Let $\log N$ denote logarithm of N with base $e^{1/k}$ (so that $k\ln N = \log N$). Therefore, from (16) and (17) we can assume that for each N the X_{ij} are i.i.d., $\mathrm{E}X_{11} = 0$, $\mathrm{E}|X_{11}|^2 = 1$, and $|X_{ij}| \le \log N$.

Later on the proof will require a restricted growth rate on both $\|\mathbf{R}\|$ and $\|\mathbf{T}\|$. We see from (16) that we can also assume

$$\max(\|\mathbf{R}\|, \|\mathbf{T}\|) \le \log N. \quad (18)$$

B) Deterministic Approximation of $m_N(z)$: Write $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$, $\mathbf{x}_i \in \mathbb{C}^N$ and let $\mathbf{y}_j = (1/\sqrt{n})\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j$. Then we can write

$$\mathbf{B}_N = \mathbf{S} + \sum_{j=1}^{n} \tau_j \mathbf{y}_j \mathbf{y}_j^{\mathsf{H}}.$$

We assume $z \in \mathbb{C}^+$ and let $v = \Im[z]$. Define

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

and

$$p_N = -\frac{1}{nz}\sum_{j=1}^{n}\frac{\tau_j}{1+c_N\tau_j e_N} = \int \frac{-\tau}{z(1+c_N\tau e_N)} dF^{\mathbf{T}}(\tau).$$

Write $\mathbf{B}_N = \mathbf{O}\Lambda\mathbf{O}^{\mathsf{H}}$, $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_N)$, its spectral decomposition. Let $\underline{\mathbf{R}} = \{\underline{R}_{ij}\} = \mathbf{O}^{\mathsf{H}}\mathbf{R}\mathbf{O}$. Then

$$e_N = (1/N)\operatorname{tr}\underline{\mathbf{R}}(\Lambda - z\mathbf{I}_N)^{-1} = (1/N)\sum_{i=1}^{N}\frac{\underline{R}_{ii}}{\lambda_i - z}.$$

We therefore see that e_N is the Stieltjes transform of a measure on the nonnegative reals with total mass $(1/N)\operatorname{tr}\mathbf{R}$. It follows that both $e_N(z)$ and $ze_N(z)$ map $\mathbb{C}^+$ into $\mathbb{C}^+$. This implies that $p_N(z)$ and $zp_N(z)$ map $\mathbb{C}^+$ into $\mathbb{C}^+$ and, as $z \to \infty$, $zp_N(z) \to -(1/n)\operatorname{tr}\mathbf{T}$. Therefore, from Lemma 6, we also have p_N the Stieltjes transform of a measure on the nonnegative reals with total mass $(1/n)\operatorname{tr}\mathbf{T}$. From (18), it follows that

$$|e_N| \leq v^{-1}\log N \tag{19}$$

and

$$\left|\int \frac{\tau}{(1+c_N\tau e_N)} dF^{\mathbf{T}}(\tau)\right| = |zp_N(z)| \leq |z|v^{-1}\log N. \tag{20}$$

More generally, from Lemma 6, any function of the form

$$\frac{-\tau}{z(1+m(z))}$$

where $\tau \geq 0$ and $m(z)$ is the Stieltjes transform of a finite measure on $\mathbb{R}^+$, is the Stieltjes transform of a measure on the nonnegative reals with total mass τ. It follows that

$$\left|\frac{\tau}{1+m(z)}\right| \leq \tau|z|v^{-1}. \tag{21}$$

Fix now $z \in \mathbb{C}^+$. Let $\mathbf{B}_{(j)} = \mathbf{B}_N - \tau_j\mathbf{y}_j\mathbf{y}_j^{\mathsf{H}}$. Define $\mathbf{D} = -z\mathbf{I}_N + \mathbf{S} - zp_N(z)\mathbf{R}$. We write

$$\mathbf{B}_N - z\mathbf{I}_N - \mathbf{D} = \sum_{j=1}^{n}\tau_j\mathbf{y}_j\mathbf{y}_j^{\mathsf{H}} + zp_N\mathbf{R}.$$

Taking inverses and using Lemma 4 we have

$$\begin{aligned}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \mathbf{D}^{-1} &= \sum_{j=1}^{n}\tau_j\mathbf{D}^{-1}\mathbf{y}_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \\ &\quad + zp_N\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \\ &= \sum_{j=1}^{n}\tau_j\frac{\mathbf{D}^{-1}\mathbf{y}_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} \\ &\quad + zp_N\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}.\end{aligned}$$

Taking traces and dividing by N, we have

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

where

$$\begin{aligned}d_j = &\frac{(1/N)\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} \\ &- \frac{(1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}}{1+c_N\tau_j e_N}.\end{aligned}$$

Multiplying both sides of the above matrix identity by $\mathbf{R}$, and then taking traces and dividing by N, we find

$$\frac{1}{N}\operatorname{tr}\mathbf{D}^{-1}\mathbf{R} - e_N(z) = \frac{1}{n}\sum_{j=1}^{n}\tau_j d_j^e \equiv w_N^e$$

where

$$\begin{aligned}d_j^e = &\frac{(1/N)\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} \\ &- \frac{(1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}^{-1}}{1+c_N\tau_j e_N}.\end{aligned}$$

We then show that, for any $k > 0$, almost surely

$$\lim_{N\to\infty}(\log^k N)w_N^m = 0 \tag{22}$$

and

$$\lim_{n\to\infty}(\log^k N)\, w_N^e = 0. \tag{23}$$

Notice that for each j, $\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j$ can be viewed as the Stieltjes transform of a measure on $\mathbb{R}^+$. Therefore, from (21) we have

$$\left|\frac{1}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j}\right| \leq \frac{|z|}{v}. \tag{24}$$

For each j, let $e_{(j)} = e_{(j)}(z) = (1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}$, and

$$p_{(j)} = p_{(j)}(z) = \int \frac{-\tau}{z(1+c_N\tau e_{(j)})} dF^{\mathbf{T}}(\tau)$$

both being Stieltjes transforms of measures on $\mathbb{R}^+$, along with the integrand for each τ.

Using Lemma 4, Equations (18) and (21), we have

$$\begin{aligned}&|zp_N - zp_{(j)}| \\ &\quad = |e_N - e_{(j)}|c_N\left|\int \frac{\tau^2}{(1+c_N\tau e_N)(1+c_N\tau e_{(j)})} dF^{\mathbf{T}}(\tau)\right| \\ &\quad \leq \frac{c_N|z|^2\log^3 N}{Nv^3}.\end{aligned} \tag{25}$$

Let $\mathbf{D}_{(j)} = -z\mathbf{I}_N + \mathbf{S} - zp_{(j)}(z)\mathbf{R}$. Notice that $(\mathbf{B}_N - z\mathbf{I}_N)^{-1}$ and $(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}$ are bounded in spectral norm by v^{-1} and, from Lemma 8, the same holds true for $\mathbf{D}^{-1}$ and $\mathbf{D}_{(j)}^{-1}$.

In order to handle both w_N^m, d_j and w_N^e, d_j^e at the same time, we shall denote by $\mathbf{E}$ either $\mathbf{T}$ or $\mathbf{I}_N$, and w_N, d_j for now will denote either the original w_N^m, d_j or w_N^e, d_j^e. Write $d_j = d_j^1 + d_j^2 + d_j^3 + d_j^4$, where

$$\begin{aligned}
d_j^1 &= \frac{(1/N)\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} \\
&\quad - \frac{(1/N)\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} \\
d_j^2 &= \frac{(1/N)\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} \\
&\quad - \frac{(1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} \\
d_j^3 &= \frac{(1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} \\
&\quad - \frac{(1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}^{-1}}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} \\
d_j^4 &= \frac{(1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}^{-1}}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} \\
&\quad - \frac{(1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}^{-1}}{1+c_N\tau_j e_N}.
\end{aligned}$$

From Lemma 4, Equations (18), (24), and (25), we have the first equation shown at the bottom of the page. From Lemma 7, there exists $\bar{K} > 0$ such that we get the second equation shown at the bottom of the page. All three moments when multiplied by n times any power of $\log N$, are summable. Applying standard arguments using the Borel-Cantelli lemma and Boole's inequality (on $4n$ events), we conclude that, for any $k > 0$ $\log^k N \max_{j\le n}\tau_j d_j \xrightarrow{\text{a.s.}} 0$ as $N \to \infty$. Hence (22) and (23).

C) Existence and Uniqueness of $m_N^\circ(z)$: We show now that for any N, n, $\mathbf{S}$, $\mathbf{R}$, $N\times N$ nonnegative definite and $\mathbf{T} = \operatorname{diag}(\tau_1,\ldots,\tau_N)$, $\tau_k \ge 0$ for all $1 \le k \le N$, there exists a unique e with positive imaginary part for which

$$e = \frac{1}{N}\operatorname{tr}\left(\mathbf{S} + \left[\int \frac{\tau}{1+c_N\tau e}dF^{\mathbf{T}}(\tau)\right]\mathbf{R} - z\mathbf{I}_N\right)^{-1}\mathbf{R}. \tag{26}$$

For existence we consider the subsequences $\{N_j\}$, $\{n_j\}$ with $N_j = jN$, $n_j = jn$, so that c_{N_j} remains c_N, form the block diagonal matrices

$$\mathbf{R}_{N_j} = \operatorname{diag}(\mathbf{R},\mathbf{R},\ldots,\mathbf{R}), \quad \mathbf{S}_{N_j} = \operatorname{diag}(\mathbf{S},\mathbf{S},\ldots,\mathbf{S}) \tag{27}$$

both $jN\times jN$ and

$$\mathbf{T}_{N_j} = \operatorname{diag}(\mathbf{T},\mathbf{T},\ldots,\mathbf{T}) \tag{28}$$

of size $jn \times jn$. We see that $F^{\mathbf{T}_{N_j}} = F^{\mathbf{T}}$ and the right side of (26) remains unchanged for all N_j. Consider a realization where $w_{N_j}^e \to 0$ as $j\to\infty$. We have $|e_{N_j}(z)| = |(jN)^{-1}\operatorname{tr}\mathbf{R}(\mathbf{B}_{jN} - z\mathbf{I}_N)^{-1}| \le v^{-1}\log N$, remaining bounded as $j\to\infty$. Consider then a subsequence for which e_{N_j} converges to, say, e. From (21), we see that

$$\left|\frac{\tau}{1+c_N\tau e_{N_j}}\right| \le \tau|z|v^{-1}$$

so that from the dominated convergence theorem we have

$$\int \frac{\tau}{1+c_N\tau e_{N_j}(z)}dF^{\mathbf{T}}(\tau) \to \int \frac{\tau}{1+c_N\tau e}dF^{\mathbf{T}}(\tau)$$

along this subsequence. Therefore e solves (26).

$$\begin{aligned}
\tau_j|d_j^1| &\le \frac{1}{N}\|\mathbf{x}_j\|^2\frac{c_N\log^7 N|z|^3}{Nv^7} \\
\tau_j|d_j^2| &\le |z|v^{-1}\frac{\log N}{N}\left|\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j - \operatorname{tr}\mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}\right| \\
\tau_j|d_j^3| &\le \frac{|z|\log^3 N}{vN}\left(\frac{1}{v^2} + \frac{c_N|z|^2\log^3 N}{v^6}\right) \to 0, \text{ as } n\to\infty \\
\tau_j|d_j^4| &\le \frac{|z|c_N\log^4 N}{Nv^3}\left(|\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j - \operatorname{tr}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}| + \frac{\log N}{v}\right)
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}|\frac{1}{N}\|\mathbf{x}_j\|^2 - 1|^6 &\le KN^{-3}\log^{12} N \\
\mathrm{E}\frac{1}{N^6}|\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j - \operatorname{tr}\mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}|^6 &\le KN^{-3}v^{-12}\log^{24} N \\
\mathrm{E}\frac{1}{N^6}|\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j - \operatorname{tr}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}|^6 &\le KN^{-3}v^{-6}\log^{18} N
\end{aligned}$$

We now show uniqueness. Let e be a solution to (26) and let $e_2 = \Im[e]$. Recalling the definition of $\mathbf{D}$ we write

$$e = \frac{1}{N}\operatorname{tr}\left(\mathbf{D}^{-1}\mathbf{R}\mathbf{D}^{-\mathsf{H}}\left(\mathbf{S}+\left[\int \frac{\tau}{1+c_N\tau e^*}dF^{\mathbf{T}}(\tau)\right]\mathbf{R}-z^*\mathbf{I}\right)\right).$$

We see that since both $\mathbf{R}$ and $\mathbf{S}$ are Hermitian nonnegative definite, $\operatorname{tr}\left(\mathbf{D}^{-1}\mathbf{R}\mathbf{D}^{-\mathsf{H}}\mathbf{S}\right)$ is real and nonnegative. Therefore we can write

$$\begin{aligned} e_2 &= \frac{1}{N}\operatorname{tr}\left(\mathbf{D}^{-1}\mathbf{R}(\mathbf{D}^{\mathsf{H}})^{-1}\left(\left[\int \frac{c_N\tau^2 e_2}{|1+c_N\tau e|^2}dF^{\mathbf{T}}(\tau)\right]\mathbf{R}+v\mathbf{I}_N\right)\right) \\ &= e_2\alpha + v\beta \end{aligned} \tag{29}$$

where we denoted

$$\alpha = \frac{1}{N}\operatorname{tr}\left(\mathbf{D}^{-1}\mathbf{R}(\mathbf{D}^{\mathsf{H}})^{-1}\left[\int \frac{c_N\tau^2}{|1+c_N\tau e|^2}dF^{\mathbf{T}}(\tau)\right]\mathbf{R}\right)$$
$$\beta = \frac{1}{N}\operatorname{tr}\left(\mathbf{D}^{-1}\mathbf{R}(\mathbf{D}^{\mathsf{H}})^{-1}\right).$$

Let $\underline{e}$ be another solution to (26), with $\underline{e}_2 = \Im[\underline{e}]$, and analogously we can write $\underline{e}_2 = \underline{e}_2\underline{\alpha} + v\underline{\beta}$. Let $\underline{\mathbf{D}}$ denote $\mathbf{D}$ with e replaced by $\underline{e}$. Then we have $e - \underline{e} = \gamma(e - \underline{e})$, where

$$\gamma = \int \frac{c_N\tau^2}{(1+c_N\tau e)(1+c_N\tau\underline{e})}dF^{\mathbf{T}}(\tau)\frac{\operatorname{tr}\mathbf{D}^{-1}\mathbf{R}\underline{\mathbf{D}}^{-1}\mathbf{R}}{N}.$$

If $\mathbf{R}$ is the zero matrix, then $\gamma = 0$, and $e = \underline{e}$ would follow. For $\mathbf{R} \neq 0$ we use Cauchy-Schwarz to find

$$\begin{aligned} |\gamma| &\leq \left(\int \frac{c_N\tau^2}{|1+c_N\tau e|^2}dF^{\mathbf{T}}(\tau)\frac{\operatorname{tr}\mathbf{D}^{-1}\mathbf{R}(\mathbf{D}^{\mathsf{H}})^{-1}\mathbf{R}}{N}\right)^{\frac{1}{2}} \\ &\quad\times\left(\int \frac{c_N\tau^2}{|1+c_N\tau\underline{e}|^2}dF^{\mathbf{T}}(\tau)\frac{\operatorname{tr}\underline{\mathbf{D}}^{-1}\mathbf{R}(\underline{\mathbf{D}}^{\mathsf{H}})^{-1}\mathbf{R}}{N}\right)^{\frac{1}{2}} \\ &= \alpha^{\frac{1}{2}}\underline{\alpha}^{\frac{1}{2}} \\ &= \left(\frac{e_2\alpha}{e_2\alpha+v\beta}\right)^{\frac{1}{2}}\left(\frac{\underline{e}_2\underline{\alpha}}{\underline{e}_2\underline{\alpha}+v\underline{\beta}}\right)^{\frac{1}{2}}. \end{aligned}$$

Necessarily, β and $\underline{\beta}$ are positive since $\mathbf{R} \neq 0$. Therefore, $|\gamma| < 1$ so we must have $e = \underline{e}$. For $z < 0$ and $e > 0$, the same calculus can be performed, with γ remaining the same. The step (29) is changed by evaluating e, instead of e_2, using the same technique. We obtain the same α while β is replaced by another positive scalar. We therefore still have that $\gamma < 1$.

D) Termination of the Proof: Let e_N° denote the solution to (26). We show now for any $\ell > 0$, almost surely

$$\lim_{N\to\infty} \log^\ell N(e_N - e_N^\circ) = 0. \tag{30}$$

Let $e_2^\circ = \Im[e_N^\circ]$, and $\alpha^\circ = \alpha_N^\circ, \beta^\circ = \beta_N^\circ$ be the values as above for which $e_2^\circ = e_2^\circ\alpha^\circ + v\beta^\circ$. We have, using (18) and (21)

$$\begin{aligned} e_2^\circ\alpha_N^\circ/\beta_N^\circ &\leq e_2^\circ c_N \log N\int \frac{\tau^2}{|1+c_N\tau e_N^\circ|^2}dF^{\mathbf{T}}(\tau) \\ &= -\log N\Im\left[\int \frac{\tau}{1+c_N\tau e_N^\circ}dF^{\mathbf{T}}(\tau)\right] \\ &\leq \log^2 N|z|v^{-1}. \end{aligned}$$

Therefore

$$\begin{aligned} \alpha^\circ &= \left(\frac{e_2^\circ\alpha^\circ}{e_2^\circ\alpha^\circ+v\beta^\circ}\right) \\ &= \left(\frac{e_2^\circ\alpha^\circ/\beta^\circ}{v+e_2^\circ\alpha^\circ/\beta^\circ}\right) \\ &\leq \left(\frac{\log^2 N|z|}{v^2+\log^2 N|z|}\right). \end{aligned} \tag{31}$$

Let $\mathbf{D}^\circ$, $\mathbf{D}$ denote $\mathbf{D}$ as above with e replaced by, respectively e_N° and e_N. We have

$$e_N = \frac{1}{N}\operatorname{tr}\mathbf{D}^{-1}\mathbf{R} - w_N^e.$$

With $e_2 = \Im[e_N]$, we write as above

$$\begin{aligned} e_2 &= \frac{1}{N}\operatorname{tr}\left(\mathbf{D}^{-1}\mathbf{R}\mathbf{D}^{-\mathsf{H}}\left(\left[\int \frac{c_N\tau^2 e_2}{|1+c_N\tau e_N|^2}dF^{\mathbf{T}}(\tau)\right]\mathbf{R}+v\mathbf{I}_N\right)\right) \\ &\quad - \Im[w_N^e] \\ &= e_2\alpha + v\beta - \Im w_N^e. \end{aligned}$$

We have as above $e_N - e_N^\circ = \gamma(e_N - e_N^\circ) + w_N^e$ where now

$$|\gamma| \leq \alpha^{\circ\frac{1}{2}}\alpha^{\frac{1}{2}}.$$

Fix an $\ell > 0$ and consider a realization for which $\log^{\ell'} N\, w_N^e \to 0$, where $\ell' = \max(\ell+1, 4)$ and n large enough so that

$$|w_N^e| \leq \frac{v^3}{4c_N|z|^2\log^3 N}. \tag{32}$$

Suppose $\beta \leq \frac{v^2}{4c_N|z|^2\log^3 N}$. Then by (18) and (21), we get

$$\alpha \leq c_N v^{-2}|z|^2\log^3 N\beta \leq 1/4,$$

which implies $|\gamma| \leq 1/2$. Otherwise, we get from (31) and (32)

$$\begin{aligned} |\gamma| &\leq \alpha^{\circ\frac{1}{2}}\left(\frac{e_2\alpha}{e_2\alpha+v\beta-\Im[w_N^e]}\right)^{\frac{1}{2}} \\ &\leq \left(\frac{\log N|z|}{v^2+\log N|z|}\right)^{\frac{1}{2}}. \end{aligned}$$

Therefore for all N large

$$\begin{aligned} \log^\ell N|e_N - e_N^\circ| &\leq \frac{(\log^\ell N)w_N^e}{1-\left(\frac{\log^2 N|z|}{v^2+\log^2 N|z|}\right)^{\frac{1}{2}}} \\ &\leq 2v^{-2}(v^2+\log^2 N|z|)(\log^\ell N)w_N^e \\ &\to 0 \end{aligned}$$

as $n \to \infty$. Therefore (30) follows.

Let $m_N^\circ = N^{-1}\operatorname{tr}\mathbf{D}^\circ$. We finally show

$$m_N - m_N^\circ \xrightarrow{\text{a.s.}} 0 \tag{33}$$

as $n \to \infty$. Since $m_N = N^{-1} \operatorname{tr} \mathbf{D}^{-1} - w_N^m$, we have

$$m_N - m_N^\circ = \gamma(e_N - e_N^\circ) - w_N^m$$

where now

$$\gamma = \int \frac{c_N \tau^2}{(1 + c_N \tau e_N)(1 + c_N \tau e_N^\circ)} dF^{\mathbf{T}}(\tau) \frac{\operatorname{tr} \mathbf{D}^{-1} \mathbf{R} \mathbf{D}^{\circ -1}}{N}.$$

From (18) and (21), we get $|\gamma| \leq c_N |z|^2 v^{-4} \log^3 N$. Therefore, from (22) and (30), we get (33).

Returning to the original assumptions on X_{11}, $\mathbf{T}$, and $\mathbf{R}$, for each of a countably infinite collection of z with positive imaginary part, possessing a cluster point with positive imaginary part, we have (33). Therefore, by Vitali's convergence theorem, [30, page 168], for any $\varepsilon > 0$ we have with probability one $m_N(z) - m_N^\circ(z) \to 0$ uniformly in any region of $\mathbb{C}$ bounded by a contour interior to

$$\mathbb{C} \setminus (\{z : |z| \leq \epsilon\} \cup \{z = x + iv : x > 0, |v| \leq \varepsilon\}).$$

If $\mathbf{S} = f(\mathbf{R})$, meaning the eigenvalues of $\mathbf{R}$ are changed via f in the spectral decomposition of $\mathbf{R}$, then we have

$$m_N^\circ(z) = \int \frac{1}{f(r) + r \int \frac{\tau}{1 + c_N \tau e_N^\circ(z)} dF^{\mathbf{T}}(\tau) - z} dF^{\mathbf{R}}(r)$$
$$e_N^\circ(z) = \int \frac{r}{f(r) + r \int \frac{\tau}{1 + c_N \tau e_N^\circ(z)} dF^{\mathbf{T}}(\tau) - z} dF^{\mathbf{R}}(r).$$

E) Extension to $K \geq 1$: Suppose now

$$\mathbf{B}_N = \mathbf{S} + \sum_{k=1}^K \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k \mathbf{X}_k^{\mathsf{H}} \mathbf{R}_k^{\frac{1}{2}}$$

where K remains fixed, $\mathbf{X}_k$ is $N \times n_k$ satisfying 1, the $\mathbf{X}_k$'s are independent, $\mathbf{R}_k$ satisfies 2) and 4), $\mathbf{T}_k$ is $n_k \times n_k$ satisfying 3) and 4), $c_k = N/n_k$ satisfies 6), and $\mathbf{S}$ satisfies 5). After truncation and centralization we may assume the same condition on the entries of the $\mathbf{X}_k$'s, and the spectral norms of the $\mathbf{R}_k$'s and the $\mathbf{T}_k$'s. Write $\mathbf{y}_{k,j} = (1/\sqrt{n_k}) \mathbf{R}_k^{\frac{1}{2}} \mathbf{x}_{k,j}$, with $\mathbf{x}_{k,j}$ denoting the j-th column of $\mathbf{X}_k$, and let $\tau_{k,j}$ denote the j-th diagonal element of $\mathbf{T}_k$. Then we can write

$$\mathbf{B}_N = \mathbf{S} + \sum_{k=1}^K \sum_{j=1}^{n_k} \tau_{k,j} \mathbf{y}_{k,j} \mathbf{y}_{k,j}^{\mathsf{H}}.$$

Define

$$e_{N,k} = e_{N,k}(z) = (1/N) \operatorname{tr} \mathbf{R}_k (\mathbf{B}_N - z \mathbf{I}_N)^{-1}$$
and
$$p_k = -\frac{1}{n_k z} \sum_{j=1}^{n_k} \frac{\tau_{k,j}}{1 + c_k \tau_{k,j} e_{N,k}} = \int \frac{-\tau_k}{1 + c_k \tau_k e_{N,k}} dF^{\mathbf{T}_k}(\tau_k).$$

We see $e_{N,k}$ and p_k have the same properties as e_N and p_N. Let $\mathbf{B}_{k,(j)} = \mathbf{B}_N - \tau_{k,j} \mathbf{y}_{k,j} \mathbf{y}_{k,j}^{\mathsf{H}}$. Define $\mathbf{D} = -z \mathbf{I}_N + \mathbf{S} - \sum_{k=1}^K z p_k(z) \mathbf{R}_k$. We write

$$\mathbf{B}_N - z \mathbf{I}_N - \mathbf{D} = \sum_{k=1}^K \left(\sum_{j=1}^{n_k} \tau_{k,j} \mathbf{y}_{k,j} \mathbf{y}_{k,j}^{\mathsf{H}} + z p_k(z) \mathbf{R}_k \right).$$

Taking inverses and using Lemma 4, we have

$$\begin{aligned} &\mathbf{D}^{-1} - (\mathbf{B}_N - z \mathbf{I}_N)^{-1} \\ &\quad = \sum_{k=1}^K \left(\sum_{j=1}^{n_k} \tau_{k,j} \mathbf{D}^{-1} \mathbf{y}_{k,j} \mathbf{y}_{k,j}^{\mathsf{H}} (\mathbf{B}_N - z \mathbf{I}_N)^{-1} \right. \\ &\qquad \left. + z p_k \mathbf{D}^{-1} \mathbf{R}_k (\mathbf{B}_N - z \mathbf{I}_N)^{-1} \right) \\ &\quad = \sum_{k=1}^K \left(\sum_{j=1}^{n_k} \tau_{k,j} \frac{\mathbf{D}^{-1} \mathbf{y}_{k,j} \mathbf{y}_{k,j}^{\mathsf{H}} (\mathbf{B}_{k,(j)} - z \mathbf{I}_N)^{-1}}{1 + \tau_{k,j} \mathbf{y}_{k,j}^{\mathsf{H}} (\mathbf{B}_{k,(j)} - z \mathbf{I}_N)^{-1} \mathbf{y}_{k,j}} \right. \\ &\qquad \left. + z p_k \mathbf{D}^{-1} \mathbf{R}_k (\mathbf{B}_N - z \mathbf{I}_N)^{-1} \right). \end{aligned}$$

Taking traces and dividing by N, we have

$$(1/N) \operatorname{tr} \mathbf{D}^{-1} - m_N(z) = \sum_{k=1}^K \frac{1}{n_k} \sum_{j=1}^{n_k} \tau_{k,j} d_{k,j} \equiv w_N^m,$$

where

$$d_{k,j} = \frac{(1/N) \mathbf{x}_{k,j}^{\mathsf{H}} \mathbf{R}_k^{\frac{1}{2}} (\mathbf{B}_{k,(j)} - z \mathbf{I}_N)^{-1} \mathbf{D}^{-1} \mathbf{R}_k^{\frac{1}{2}} \mathbf{x}_{k,j}}{1 + \tau_{k,j} \mathbf{y}_{k,j}^{\mathsf{H}} (\mathbf{B}_{k,(j)} - z \mathbf{I}_N)^{-1} \mathbf{y}_{k,j}} - \frac{(1/N) \operatorname{tr} \mathbf{R}_k (\mathbf{B}_N - z \mathbf{I}_N)^{-1} \mathbf{D}^{-1}}{1 + c_k \tau_{k,j} e_{N,k}}.$$

For a fixed $\underline{k} \in \{1, \ldots, K\}$, we multiply the above matrix identity by $\mathbf{R}_{\underline{k}}$, take traces and divide by N. Thus we get

$$(1/N) \operatorname{tr} \mathbf{D}^{-1} \mathbf{R}_{\underline{k}} - e_{\underline{k}}(z) = \sum_{k=1}^K \frac{1}{n_k} \sum_{j=1}^{n_k} \tau_{k,j} d_{\underline{k}kj}^e \equiv w_{\underline{k}}^e$$

where

$$d_{\underline{k}kj}^e = \frac{(1/N) \mathbf{x}_{k,j}^{\mathsf{H}} \mathbf{R}_k^{\frac{1}{2}} (\mathbf{B}_{k,(j)} - z \mathbf{I}_N)^{-1} \mathbf{R}_{\underline{k}} \mathbf{D}^{-1} \mathbf{R}_k^{\frac{1}{2}} \mathbf{x}_{k,j}}{1 + \tau_{k,j} \mathbf{y}_{k,j}^{\mathsf{H}} (\mathbf{B}_{k,(j)} - z \mathbf{I}_N)^{-1} \mathbf{y}_{k,j}} - \frac{(1/N) \operatorname{tr} \mathbf{R}_k (\mathbf{B}_N - z \mathbf{I}_N)^{-1} \mathbf{R}_{\underline{k}} \mathbf{D}^{-1}}{1 + c_k \tau_{k,j} e_{N,k}}.$$

In exactly the same way as in the case with $K = 1$ we find that for any nonnegative ℓ, $\log^\ell N w_N^m$ and the $\log^\ell w_{\underline{i}}^e$'s converge almost surely to zero. By considering block diagonal matrices as before with N, n_i's, $\mathbf{S}$, $\mathbf{R}_i$'s and $\mathbf{T}_i$'s all fixed we find that

there exist $e_1^\circ, \ldots, e_K^\circ$ with positive imaginary parts for which for each i

$$e_i^\circ = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\mathbf{S} + \sum_{k=1}^{K} \left[\int \frac{\tau}{1 + c_k \tau e_k^\circ} dF^{\mathbf{T}_k}(\tau) \right] \mathbf{R}_k - z \mathbf{I}_N \right)^{-1}. \tag{34}$$

Let us verify uniqueness. Let $\mathbf{e}^\circ = (e_1^\circ, \ldots, e_K^\circ)^\mathsf{T}$, and let $\mathbf{D}^\circ$ denote the matrix in (34) whose inverse is taken (essentially $\mathbf{D}$ after the $e_{N,i}$'s are replaced by the e_i°'s). Let for each j, $e_{j,2}^\circ = \Im e_j^\circ$, and $\mathbf{e}_2^\circ = (e_{1,2}^\circ, \ldots, e_{K,2}^\circ)^\mathsf{T}$. Then, noticing that for each i, $\operatorname{tr} \mathbf{S} \mathbf{D}^{\circ -\mathsf{H}} \mathbf{R}_i \mathbf{D}^{\circ -1}$ is real and nonnegative (positive whenever $\mathbf{S} \neq 0$) and $\operatorname{tr} \mathbf{D}^{\circ -\mathsf{H}} \mathbf{R}_i \mathbf{D}^{\circ -1}$ and $\operatorname{tr} \mathbf{R}_j \mathbf{D}^{\circ -\mathsf{H}} \mathbf{R}_i \mathbf{D}^{\circ -1}$ are real and positive for all i, j, we have (35), as shown at the bottom of the page.

Let $\mathbf{C}^\circ = (c_{ij}^\circ)$, $\mathbf{b}^\circ = (b_1^\circ, \ldots, b_N^\circ)^\mathsf{T}$, where

$$c_{ij}^\circ = \frac{1}{N} \operatorname{tr} \mathbf{R}_j \mathbf{D}^{\circ -\mathsf{H}} \mathbf{R}_i \mathbf{D}^{\circ -1} c_j \int \frac{\tau^2}{|1 + c_j \tau e_j^\circ|^2} dF^{\mathbf{T}_j}(\tau)$$

and

$$b_i^\circ = \frac{1}{N} \operatorname{tr} \mathbf{D}^{\circ -\mathsf{H}} \mathbf{R}_i \mathbf{D}^{\circ -1}.$$

Therefore, we have that $\mathbf{e}_2^\circ$ satisfies

$$\mathbf{e}_2^\circ = \mathbf{C}^\circ \mathbf{e}_2^\circ + v \mathbf{b}^\circ. \tag{36}$$

We see that each $e_{j,2}^\circ$, c_{ij}°, and b_j° are positive. Therefore, from Lemma 9 we have $\rho(\mathbf{C}^\circ) < 1$.

Let $\underline{\mathbf{e}}^\circ = (\underline{e}_1^\circ, \ldots, \underline{e}_K^\circ)^\mathsf{T}$ be another solution to (34), with $\underline{\mathbf{e}}_2^\circ$, $\underline{\mathbf{D}}^\circ$, $\underline{\mathbf{C}}^\circ = (\underline{c}_{ij}^\circ)$, $\underline{\mathbf{b}}^\circ$ defined analogously, so that (36) holds and $\rho(\underline{\mathbf{C}}^\circ) < 1$. We have for each i

$$e_i^\circ - \underline{e}_i^\circ = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \mathbf{D}^{\circ -1} \times \sum_{j=1}^{K} (e_j^\circ - \underline{e}_j^\circ) c_j \int \frac{\tau^2}{(1 + c_j \tau e_j^\circ)(1 + c_j \tau \underline{e}_j^\circ)} dF^{\mathbf{T}_j}(\tau) \mathbf{R}_j \underline{\mathbf{D}}^{\circ -1}.$$

Thus, with $\mathbf{A} = (a_{ij})$ where

$$a_{ij} = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \mathbf{D}^{\circ -1} \mathbf{R}_j \underline{\mathbf{D}}^{\circ -1} c_j \int \frac{\tau^2}{(1 + c_j \tau e_j^\circ)(1 + c_j \tau \underline{e}_j^\circ)} dF^{\mathbf{T}_j}(\tau) \tag{37}$$

we have

$$\mathbf{e}^\circ - \underline{\mathbf{e}}^\circ = \mathbf{A} (\mathbf{e}^\circ - \underline{\mathbf{e}}^\circ) \tag{38}$$

which means, if $\mathbf{e}^\circ \neq \underline{\mathbf{e}}^\circ$, then $\mathbf{A}$ has an eigenvalue equal to 1.

Applying Cauchy-Schwarz, we have

$$\begin{aligned} |a_{ij}| &\leq \left(\frac{1}{N} \mathbf{R}_i \mathbf{D}^{\circ -1} \mathbf{R}_j \mathbf{D}^{\circ -\mathsf{H}} \int \frac{\tau^2}{|1 + c_j \tau e_j^\circ|^2} dF^{\mathbf{T}_j}(\tau) \right)^{\frac{1}{2}} \\ &\quad \times \left(\frac{1}{N} \mathbf{R}_i \underline{\mathbf{D}}^{\circ -1} \mathbf{R}_j \underline{\mathbf{D}}^{\circ -\mathsf{H}} \int \frac{\tau^2}{|1 + c_j \tau \underline{e}_j^\circ|^2} dF^{\mathbf{T}_j}(\tau) \right)^{\frac{1}{2}} \\ &= {c_{ij}^\circ}^{1/2} {\underline{c}_{ij}^\circ}^{1/2}. \end{aligned}$$

Therefore, from Lemmas 10 and 11 we get

$$\rho(\mathbf{A}) \leq \rho(c_{ij}^{\circ \frac{1}{2}} \underline{c}_{ij}^{\circ \frac{1}{2}}) \leq \rho(\mathbf{C}^\circ)^{\frac{1}{2}} \rho(\underline{\mathbf{C}}^\circ)^{\frac{1}{2}} < 1.$$

A contradiction to the statement $\mathbf{A}$ has an eigenvalue equal to 1. Consequently we have $\mathbf{e} = \underline{\mathbf{e}}$.

The same reasoning can be applied to $z < 0$, with $e_i^\circ > 0$. In this case matrix $\mathbf{A}$ remains the same. The step (35) is now replaced by taking e_i°, instead of its imaginary part, using the same line of reasoning. This leads to the same matrix $\mathbf{C}^\circ$ with (36) remaining true with $\mathbf{b}^\circ$ replaced by another positive vector. The conclusion $\rho(\mathbf{A}) < 1$ therefore remains.

Let $\mathbf{e}_N = (e_{N,1}, \ldots, e_{N,K})^\mathsf{T}$ and $\mathbf{e}_N^\circ = (e_{N,1}^\circ, \ldots, e_{N,K}^\circ)$ denote the vector solution to (34) for each N. We will show for any $\ell > 0$, almost surely

$$\lim_{N \to \infty} \log^\ell N (\mathbf{e}_N - \mathbf{e}_N^\circ) \to \mathbf{0}. \tag{39}$$

We have

$$\mathbf{e}_N^\circ = \left(\frac{1}{N} \operatorname{tr} \mathbf{R}_1 \mathbf{D}^{\circ -1}, \ldots, \frac{1}{N} \operatorname{tr} \mathbf{R}_K \mathbf{D}^{\circ -1} \right)^\mathsf{T}.$$

Let $\mathbf{w}^e = \mathbf{w}_N^e = -(w_1^e, \ldots, w_K^e)^\mathsf{T}$. Then we can write

$$\mathbf{e}_N = \left(\frac{1}{N} \operatorname{tr} \mathbf{R}_1 \mathbf{D}^{-1}, \ldots, \frac{1}{N} \operatorname{tr} \mathbf{R}_K \mathbf{D}^{-1} \right)^\mathsf{T} + \mathbf{w}^e.$$

Therefore

$$\mathbf{e}_N - \mathbf{e}_N^\circ = \mathbf{A}(N) (\mathbf{e}_N - \mathbf{e}_N^\circ) + \mathbf{w}^e$$

$$\begin{aligned} e_{i,2}^\circ &= \Im \left[\frac{1}{N} \operatorname{tr} \left(\mathbf{S} + \sum_{j=1}^{K} \left[\int \frac{\tau}{1 + c_j \tau \overline{e}_j^\circ} dF^{\mathbf{T}_j}(\tau) \right] \mathbf{R}_j - z^* \mathbf{I} \right) \mathbf{D}^{\circ -\mathsf{H}} \mathbf{R}_i \mathbf{D}^{\circ -1} \right] \\ &= \sum_{j=1}^{K} e_{j,2}^\circ \frac{1}{N} \operatorname{tr} \mathbf{R}_j \mathbf{D}^{\circ -\mathsf{H}} \mathbf{R}_i \mathbf{D}^{\circ -1} c_j \int \frac{\tau^2}{|1 + c_j \tau e_j^\circ|^2} dF^{\mathbf{T}_j}(\tau) + \frac{v}{N} \operatorname{tr} \mathbf{D}^{\circ -\mathsf{H}} \mathbf{R}_i \mathbf{D}^{\circ -1} \end{aligned} \tag{35}$$

where $\mathbf{A}(N) = (a_{ij}(N))$ with

$$a_{ij}(N) = \frac{1}{N}\operatorname{tr}\mathbf{R}_i\mathbf{D}^{-1}\mathbf{R}_j\mathbf{D}^{\circ -1}c_j \times \int \frac{\tau^2}{(1+c_j\tau e_{N,j})(1+c_j\tau e^{\circ}_{N,j})}dF^{\mathbf{T}_j}(\tau).$$

We let $\mathbf{e}^{\circ}_{N,2}$, $b^{\circ}_{ij}(N)$, $\mathbf{C}^{\circ}(N)$, $b^{\circ}_{N,i}$, and $\mathbf{b}^{\circ}_N$, denote the quantities from above, reflecting now their dependence on N. Let $\mathbf{C}(N) = (c_{ij}(N))$ be $K \times K$ with

$$c_{ij}(N) = \frac{1}{N}\operatorname{tr}\mathbf{R}_j\mathbf{D}^{-\mathsf{H}}\mathbf{R}_i\mathbf{D}^{-1}c_j \int \frac{\tau^2}{|1+c_j\tau e_{N,j}|^2}dF^{\mathbf{T}_j}(\tau).$$

Let $\mathbf{e}_{N,2} = \Im[\mathbf{e}_N]$ and $\mathbf{w}^e_2 = \Im[\mathbf{w}^e]$. Define $\mathbf{b}_N = (b_{N,1},\ldots,b_{N,K})^{\mathsf{T}}$ with

$$b_{N,i} = \frac{1}{N}\operatorname{tr}\mathbf{D}^{-\mathsf{H}}\mathbf{R}_i\mathbf{D}^{-1}.$$

Then, as above we find that

$$\mathbf{e}_{N,2} = \mathbf{C}(N)\mathbf{e}_{N,2} + v\mathbf{b}_N + \mathbf{w}^e_2. \tag{40}$$

Using (18) and (21), we see there exists a constant $K_1 > 0$ for which

$$\begin{aligned} c^{\circ}_{ij}(N) &\le K_1\log^3 N b^{\circ}_{N,i} \\ &\text{and} \\ c_{ij}(N) &\le K_1\log^3 N b_{N,i} \\ c_{ij}(N) &\le K_1\log^4 N \end{aligned}$$

for each i,j. Therefore, from (36), we see there exists $\hat{K} > 0$ for which

$$e^{\circ}_{N,i} \le \hat{K}\log^4 N v b^{\circ}_{N,i}. \tag{41}$$

Let $\mathbf{x}$ be such that $\mathbf{x}^{\mathsf{T}}$ is a left eigenvector of $\mathbf{C}^{\circ}(N)$ corresponding to eigenvalue $\rho(\mathbf{C}^{\circ}(N))$, guaranteed by Lemma 12. Then from(40), we have

$$\mathbf{x}^{\mathsf{T}}\mathbf{e}^{\circ}_{N,2} = \rho(\mathbf{C}^{\circ}(N))\mathbf{x}^{\mathsf{T}}\mathbf{e}^{\circ}_{N,2} + v\mathbf{x}^{\mathsf{T}}\mathbf{b}^{\circ}_N. \tag{42}$$

Using (42), we have

$$1-\rho(\mathbf{C}^{\circ}(N)) = \frac{v\mathbf{x}^{\mathsf{T}}\mathbf{b}^{\circ}_N}{\mathbf{x}^{\mathsf{T}}\mathbf{e}^{\circ}_{N,2}} \ge (\hat{K}\log^4 N)^{-1}. \tag{43}$$

Fix an $\ell > 0$ and consider a realization for which $\log^{\ell+3+p} N\mathbf{w}^e_N \to 0$, as $N \to \infty$, where $p \ge 12K-7$. We will show for all N large

$$\rho(\mathbf{C}(N)) \le 1 + (\hat{K}\log^4 N)^{-1}. \tag{44}$$

For each N, we rearrange the entries of $\mathbf{e}_{N,2}$, $v\mathbf{b}_m + \mathbf{w}^e_2$, and $\mathbf{C}(n)$ depending on whether the i^{th} entry of $v\mathbf{b}_m + \mathbf{w}^e_2$ is greater than, or less than or equal to zero. We can therefore assume

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11}(N) & \mathbf{C}_{12}(N) \\ \mathbf{C}_{21}(N) & \mathbf{C}_{22}(N) \end{pmatrix}$$

where $\mathbf{C}_{11}(N)$ is $k_1 \times k_1$, $\mathbf{C}_{22}(N)$ is $k_2 \times k_2$, $\mathbf{C}_{12}(N)$ is $k_1 \times k_2$, and $\mathbf{C}_{21}(N)$ is $k_2 \times k_1$. From Lemma 9, we have $\rho(\mathbf{C}_{11}(N)) < 1$. If $vb_{N,i} + \mathbf{w}^e_{2,i} \le 0$, then necessarily $vb_{N,i} \le |\mathbf{w}^e_N| \le K_1(\log n)^{-(3+p)}$, and so from (41), we have the entries of $\mathbf{C}_{21}(N)$ and $\mathbf{C}_{22}(N)$ bounded by $K_1(\log N)^{-p}$. We may assume for all N large $0 < k_1 < K$, since otherwise we would have $\rho(\mathbf{C}(N)) < 1$.

We seek an expression for $\det(\mathbf{C}(N) - \lambda\mathbf{I}_N)$ in which Lemma 14 can be used. We consider N large enough so that, for $|\lambda| \ge 1/2$, we have $(\mathbf{C}_{22}(N) - \lambda\mathbf{I}_N)^{-1}$ existing with entries uniformly bounded, and have the equation shown at the bottom of the page.

We see then that for $\lambda = \rho(\mathbf{C}(N))$ real and greater than 1

$$\det(\mathbf{C}_{11}(N) - \lambda\mathbf{I} - \mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda\mathbf{I})^{-1}\mathbf{C}_{21}(N)) \tag{45}$$

must be zero.

Notice that from (41), the entries of $\mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda\mathbf{I})^{-1}\mathbf{C}_{21}(N)$ can be made smaller than any negative power of $\log N$ for p sufficiently large. Notice also that the diagonal elements of $\mathbf{C}_{11}(N)$ are all less than 1. From this, Lemma 13 and (41), we see that $\rho(\mathbf{C}(N)) \le K_1\log^4 N$. The determinant in (45) can be written as

$$\det(\mathbf{C}_{11}(N) - \lambda\mathbf{I}) + g(\lambda)$$

where $g(\lambda)$ is a sum of products, each containing at least one entry from $\mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda\mathbf{I})^{-1}\mathbf{C}_{21}(N)$. Again, from (41), we see that for all $|\lambda| \ge 1/2$, $g(\lambda)$ can be made smaller than any negative power of $\log N$ by making p sufficiently large. Choose p so that $|g(\lambda)| < (\hat{K}\log N)^{-4k_1}$ for these λ. It is clear that any $p > 8k_1 + 4$ will suffice. Let $\lambda_1,\ldots,\lambda_{k_1}$

$$\begin{aligned} \det(\mathbf{C}(N) - \lambda\mathbf{I}) &= \det\left[\begin{pmatrix} \mathbf{I} & -\mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda\mathbf{I})^{-1} \\ 0 & \mathbf{I} \end{pmatrix}\begin{pmatrix} \mathbf{C}_{11}(N) - \lambda\mathbf{I} & \mathbf{C}_{12}(N) \\ \mathbf{C}_{21}(N) & \mathbf{C}_{22}(N) - \lambda\mathbf{I} \end{pmatrix}\right] \\ &= \det\begin{pmatrix} \mathbf{C}_{11}(N) - \lambda\mathbf{I} - \mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda\mathbf{I})^{-1}\mathbf{C}_{21}(N) & 0 \\ \mathbf{C}_{21}(N) & \mathbf{C}_{22}(N) - \lambda\mathbf{I} \end{pmatrix} \\ &= \det(\mathbf{C}_{11}(N) - \lambda\mathbf{I} - \mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda\mathbf{I})^{-1}\mathbf{C}_{21}(N))\det(\mathbf{C}_{22}(N) - \lambda\mathbf{I}) \end{aligned}$$

denote the eigenvalues of $\mathbf{C}_{11}$. Since $\rho(\mathbf{C}_{11}) < 1$, we see that for $|\lambda| \geq (\widehat{K}\log N)^{-4}$, we have

$$|\det(\mathbf{C}_{11}(N) - \lambda\mathbf{I})| = |\prod_{i=1}^{k_1}(\lambda_i - \lambda)| > (\widehat{K}\log N)^{-4k_1}.$$

Thus, with $f(\lambda) = \det(\mathbf{C}_{11}(N) - \lambda\mathbf{I})$, a polynomial, and $g(\lambda)$ being a rational function, we have the conditions of Lemma 14 being met on any rectangle C, with vertical lines going through $((\widehat{K}\log N)^{-4}, 0)$ and $(K_1(\log N)^4, 0)$. Therefore, since $f(\lambda)$ has no zeros inside C, neither does $\det(\mathbf{C}(N) - \lambda\mathbf{I})$. Thus we get (44). As before, we see that

$$|a_{ij}(N)| \leq c_{ij}^{1/2}(N) c_{ij}^{\circ\,1/2}(N).$$

Therefore, from (43), (44), and Lemmas 10 and 11, we have for all N large

$$\rho(\mathbf{A}(N)) \leq \left(\frac{\widehat{K}^2\log^8 N - 1}{\widehat{K}^2\log^8 N}\right)^{\frac{1}{2}}. \tag{46}$$

For these N, we have then $\mathbf{I} - \mathbf{A}(N)$ invertible, and so

$$\mathbf{e}_N - \mathbf{e}_N^0 = (\mathbf{I} - \mathbf{A}(N))^{-1}\mathbf{w}^e.$$

By (18) and (21), we have the entries of $\mathbf{A}(N)$ bounded by $K_1\log^4 N$. Notice also, from (46)

$$\begin{aligned}|\det(\mathbf{I}-\mathbf{A}(N))| &\geq (1 - \rho(\mathbf{A}(N))^K \\ &\geq \left(\widehat{K}^2\log^8 N\left(1+\frac{\widehat{K}^2\log^8 N - 1}{\widehat{K}^2\log^8 N}\right)^{\frac{1}{2}}\right)^{-K} \\ &\geq (2\widehat{K}^2\log^8 N)^{-K}.\end{aligned}$$

When considering the inverse of a square matrix in terms of its adjoint divided by its determinant, we see that the entries of $(\mathbf{I} - \mathbf{A}(N))^{-1}$ are bounded by

$$\frac{(K-1)!K_1(\log N)^{4(K-1)}}{|\det(\mathbf{I} - \mathbf{A}(N))|} \leq K_3(\log N)^{12K-4}.$$

Therefore, since $p \geq 12K - 7$ ($> 8k_1 + 4$), (39) follows on this realization, an event which occurs with probability one.

Letting $m_N^\circ = \frac{1}{N}\operatorname{tr}\mathbf{D}^{\circ-1}$, we have

$$m_N - m_N^\circ = \vec{\gamma}^{\mathsf{T}}(\mathbf{e}_N - \mathbf{e}_N^\circ)$$

where $\vec{\gamma} = (\gamma_1, \ldots, \gamma_K)^{\mathsf{T}}$ with

$$\gamma_j = \int \frac{c_N\tau^2}{(1+c_N\tau e_{N,j})(1+c_n\tau e_{N,j}^\circ)} dF^{\mathbf{T}_N}(\tau)\frac{\operatorname{tr}\mathbf{D}^{-1}\mathbf{R}_j\mathbf{D}^{\circ-1}}{N}.$$

From (18) and (21), we get each $|\gamma_j| \leq c_N|z|^2 v^{-4}\log^3 N$. Therefore from (39) and the fact that $w_N^m \to 0$, we have

$$m_N - m_N^\circ \to 0,$$

almost surely, as $N \to \infty$.

This completes the proof.

APPENDIX B
PROOF OF THEOREM 2

We first prove that $\mathcal{V}_N^\circ(x)$ as defined in (12) verifies

$$\mathcal{V}_N^\circ(x) = \int_x^\infty \left(\frac{1}{w} - m_N^\circ(-w)\right) dw \tag{47}$$

and then we prove that, under the conditions of Theorem 2, $\mathcal{V}^\circ(x)$ defined as such verifies

$$\mathcal{V}_N^\circ(x) - \mathcal{V}_N(x) \xrightarrow{\text{a.s.}} 0. \tag{48}$$

A) Proof of (47): First, write $e_i(z)$ under the symmetric form

$$\begin{aligned}e_i(z) &= \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(-z\left[\mathbf{I}_N + \sum_{k=1}^K \delta_k\mathbf{R}_k\right]\right)^{-1} \\ \delta_i(z) &= \frac{1}{n_i}\operatorname{tr}\mathbf{T}_i\left(-z\left[\mathbf{I}_{n_i} + c_i e_i(z)\mathbf{T}_i\right]\right)^{-1}\end{aligned}$$

and then for $m_N^\circ(z)$,

$$m_N^\circ(z) = \frac{1}{N}\operatorname{tr}\left(-z\left[\mathbf{I}_N + \sum_{k=1}^K \delta_k\mathbf{R}_k\right]\right)^{-1}.$$

Now, notice that

$$\begin{aligned}\frac{1}{z} - m_N^\circ(-z) &= \frac{1}{N}\left((z\mathbf{I})^{-1} - \left(z\left[\mathbf{I}_N + \sum_{k=1}^K \delta_k\mathbf{R}_k\right]\right)^{-1}\right) \\ &= \sum_{k=1}^K \delta_k(-z)\cdot e_k(-z).\end{aligned}$$

Since the Shannon transform $\mathcal{V}(x)$ satisfies $\mathcal{V}(x) = \int_x^{+\infty}[w^{-1} - m_N(-w)]dw$, we need to find an integral form for $\sum_{k=1}^K \delta_k(-z)\cdot e_k(-z)$. Notice now that

$$\begin{aligned}&\frac{d}{dz}\frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k=1}^K \delta_k(-z)\mathbf{R}_k\right) \\ &= -z\sum_{k=1}^K e_k(-z)\cdot\delta_k'(-z) \\ &\frac{d}{dz}\frac{1}{N}\log\det\left(\mathbf{I}_{n_k} + c_k e_k(-z)\mathbf{T}_k\right) \\ &= -z\cdot e_k'(-z)\cdot\delta_k(-z) \\ &\frac{d}{dz}\left(z\sum_{k=1}^K \delta_k(-z)e_k(-z)\right) \\ &= \sum_{k=1}^K \delta_k(-z)e_k(-z) - z\sum_{k=1}^K \delta_k'(-z)\cdot e_k(-z) + \delta_k(-z)\cdot e_k'(-z).\end{aligned}$$

Combining the last three lines, we have

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

$$-\sum_{k=1}^{K}\frac{1}{N}\log\det\left(\mathbf{I}_{n_k}+c_k e_k(-z)\mathbf{T}_k\right)+z\sum_{k=1}^{K}\delta_k(-z)e_k(-z)\Bigg]$$

which after integration leads to

$$\int_z^{+\infty}\left(\frac{1}{w}-m_N^{\circ}(-w)\right)dw=\frac{1}{N}\log\det\left(\mathbf{I}_N+\sum_{k=1}^{K}\delta_k(-z)\mathbf{R}_k\right)+\sum_{k=1}^{K}\frac{1}{N}\log\det\left(\mathbf{I}_{n_k}+c_k e_k(-z)\mathbf{T}_k\right)-z\sum_{k=1}^{K}\delta_k(-z)e_k(-z) \tag{49}$$

which is exactly the right-hand side of (12).

B) Proof of (48): Consider now the existence of a non-random α and for each N a non-negative integer r_N for which

$$\max_{i\le K}\max(\lambda_{r_N+1}^{\mathbf{T}_i},\lambda_{r_N+1}^{\mathbf{R}_i})\le\alpha$$

(eigenvalues also arranged in non-increasing order). Then for each i

$$\lambda_{2r_N+1}^{\mathbf{R}_i^{\frac{1}{2}}\mathbf{X}_i\mathbf{T}_i\mathbf{X}_i^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}}=(s_{2r_N+1}^{\mathbf{R}_i^{\frac{1}{2}}\mathbf{X}_i\mathbf{T}_i^{\frac{1}{2}}})^2\le\alpha^2\|\mathbf{X}_i\mathbf{X}_i^{\mathsf{H}}\|$$

and then we have, from Lemma 15

$$\lambda_{2Kr_N+1}^{\mathbf{B}_N}\le\alpha^2(\|\mathbf{X}_1\mathbf{X}_1^{\mathsf{H}}\|+\cdots+\|\mathbf{X}_K\mathbf{X}_K^{\mathsf{H}}\|).$$

We can in fact consider that the spectral norms of the $\mathbf{X}_i$ are bounded in the limit. Either Gaussian assumptions on the components, or finite fourth moment, but all coming from doubly infinite arrays (remember though that we need the right-unitary invariance structure of $\mathbf{X}_i$). Because of assumption 5 in Corollary 1, we can, by enlarging the sample space, assume each $\mathbf{X}_i$ is embedded in an $N\times n_i'$ matrix $\mathbf{X}_i'$, where $N/n_i'\to a$ as $N\to\infty$. Then, with probability one (see, e.g., [16])

$$\limsup_N\lambda_{2Kr_N+1}^{\mathbf{B}_N}\le\limsup_N\alpha^2(\|\mathbf{X}_1'\mathbf{X}_1'^{\mathsf{H}}\|+\cdots+\|\mathbf{X}_K'\mathbf{X}_K'^{\mathsf{H}}\|)\le\alpha^2\frac{Kb}{a}(1+\sqrt{a})^2. \tag{50}$$

Let a° be any real greater than $\alpha^2K\frac{b}{a}(1+\sqrt{a})^2$.

Since $\mathbf{S}=0$ here, it follows as in [22] that $\{F^{\mathbf{B}_n}\}$ is almost surely tight. Let F_N° denote the distribution function having Stieltjes transform m_N°, and let f on $[0,\infty)$ be a continuous function. Then the function

$$f_{a^\circ}(x)=\begin{cases}f(x), & x\le a^\circ\\ f(a^\circ), & x>a^\circ\end{cases}$$

is bounded and continuous. Therefore, with probability 1

$$\int f_{a^\circ}(x)dF_N(x)-\int f_{a^\circ}(x)dF_N^\circ(x)\to 0$$

as $N\to\infty$.

Suppose now $r_N=o(N)$. Then, since almost surely there are at most $2Kr_N$ eigenvalues greater than a° for all N large, any converging subsequence of $\{F_N^\circ\}$ must have some mass lying on $[0,a^\circ]$. This implies, with probability 1

$$\frac{1}{N}\sum_{\lambda_i\le a^\circ}f(\lambda_i)-\int_{[0,a^\circ]}f(x)dF_N^\circ(x)\to 0$$

as $N\to\infty$.

Let b_N be a bound on the spectral norms of the $\mathbf{T}_i$ and $\mathbf{R}_i$. Then

$$\|\mathbf{B}_n\|\le b_N^2(\|\mathbf{X}_1'\mathbf{X}_1'^{\mathsf{H}}\|+\cdots+\|\mathbf{X}_K'\mathbf{X}_K'^{\mathsf{H}}\|). \tag{51}$$

Fix a number $\beta>\frac{Kb}{a}(1+\sqrt{a})^2$, and let $a_N=b_N^2\beta$. Suppose also that f is increasing and that $f(a_N)r_N=o(N)$. Then

$$\int f(x)dF^{\mathbf{B}_n}(x)-\frac{1}{N}\sum_{\lambda_i\le a^\circ}f(\lambda_i)\to 0$$

almost surely, as $N\to\infty$. Therefore, with probability 1

$$\int f(x)dF_N(x)-\int_{[0,a^\circ]}f(x)dF_N^\circ(x)\to 0,$$

as $N\to\infty$.

For any N we consider, for $j=1,2,\ldots$, the $jN\times jN$ matrix $\mathbf{B}_{N,j}$ formed, as before, from block diagonal matrices and $jN\times jn_i$ matrices of i.i.d. variables. Then with probability 1, $F^{\mathbf{B}_{N,j}}$ converges weakly to F_N° as $j\to\infty$. Properties on the eigenvalues of $\mathbf{B}_{N,j}$ will thus yield properties of F_N°.

By considering the bound on $\|\mathbf{B}_{n,j}\|$ analogous to (51), we must have $F_N^\circ(a_N)=1$ for all N large.

Similar to (50) we see that, with probability 1

$$\limsup_j\lambda_{2Kjr_N+1}^{\mathbf{B}_{N,j}}\le a^2((1+\sqrt{c_1})^2+\cdots+(1+\sqrt{c_K})^2)$$

this latter number being less than a° for all N large.

At this point, we will use the fact that for probability measures P_N, P on $\mathbb{R}$ with P_N converging weakly to P, we have (see, e.g., [31])

$$\liminf_N P_N(G)\ge P(G)$$

for any open set G. Thus, with $G=(a^\circ,\infty)$, we see that, with probability 1, for all N large

$$F_N^\circ((a^\circ,\infty))=1-F_N^\circ(a^\circ)\le\liminf_j F^{\mathbf{B}_{N,j}}((a^\circ,\infty))\le 2Kr_N/N.$$

Therefore, for all N large

$$\int_{(a^\circ,\infty)}f(x)dF_N^\circ(x)\le f(a_N)2Kr_N/N\to 0$$

as $N\to\infty$.

Therefore, we conclude that, $\int f(x)dF_N^\circ(x)$ is bounded, and with probability 1

$$\int f(x)dF_N(x) - \int f(x)dF_N^\circ(x) \to 0,$$

as $N \to \infty$. This concludes the proof.

Appendix C
Proof of Proposition 2

The proof stems from the following result,

Proposition 4: $f(\mathbf{P}_1, \ldots, \mathbf{P}_K)$ is a strictly concave matrix in the Hermitian nonnegative definite matrices $\mathbf{P}_1, \ldots, \mathbf{P}_K$, if and only if, for any couples $(\mathbf{P}_{1_a}, \mathbf{P}_{1_b}), \ldots, (\mathbf{P}_{K_a}, \mathbf{P}_{K_b})$ of Hermitian nonnegative definite matrices, the function

$$\phi(\lambda) = f\left(\lambda\mathbf{P}_{1_a} + (1-\lambda)\mathbf{P}_{1_b}, \ldots, \lambda\mathbf{P}_{K_a} + (1-\lambda)\mathbf{P}_{K_b}\right)$$

is strictly concave.

Denote

$$\bar{\mathcal{V}}_N^\circ(\lambda) = \mathcal{V}_N^\circ(\lambda\mathbf{P}_{1_a} + (1-\lambda)\mathbf{P}_{1_b}, \ldots, \lambda\mathbf{P}_{|\mathcal{S}|_a} + (1-\lambda)\mathbf{P}_{|\mathcal{S}|_b})$$

and consider a set $(\delta_k, e_k, \mathbf{P}_{i_1}, \ldots, \mathbf{P}_{i_{|\mathcal{S}|}})$ which satisfies the system of (53)–(55). Then, from remark (56) and (57)

$$\begin{aligned}\frac{d\bar{\mathcal{V}}_N^\circ}{d\lambda} &= \sum_{k\in\mathcal{S}} \frac{\partial\bar{V}}{\partial\delta_k}\frac{\partial\delta_k}{\partial\lambda} + \frac{\partial\bar{V}}{\partial e_k}\frac{\partial e_k}{\partial\lambda} + \frac{\partial\bar{V}}{\partial\lambda} \\ &= \frac{\partial\bar{V}}{\partial\lambda}\end{aligned}$$

where

$$\bar{V} : (\delta_1, \ldots, \delta_{|\mathcal{S}|}, e_1, \ldots, e_{|\mathcal{S}|}, \lambda) \mapsto \bar{\mathcal{V}}_N^\circ(\lambda). \tag{52}$$

Mere derivations of $\bar{V}$ lead then to

$$\frac{\partial^2\bar{V}}{\partial\lambda^2} = -\sum_{i\in\mathcal{S}} (c_i^2 e_i^2)\frac{1}{N}\operatorname{tr}\left(\mathbf{I} + c_i e_i \mathbf{R}_i\mathbf{P}_i\right)^{-2}\left(\mathbf{R}_i(\mathbf{P}_{i_a} - \mathbf{P}_{i_b})\right)^2.$$

Since $e_i > 0$ on the strictly negative real axis, if any of the $\mathbf{R}_i$'s is positive definite, then, for all nonnegative definite couples $(\mathbf{P}_{i_a}, \mathbf{P}_{i_b})$, such that $\mathbf{P}_{i_a} \neq \mathbf{P}_{i_b}$, $\bar{\mathcal{V}}_N'' < 0$. Then, from Proposition 4, the deterministic approximate $\mathcal{V}_N^\circ$ is strictly concave in $\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}$ if any of the $\mathbf{R}_i$ matrices is invertible.

Appendix D
Proof of Proposition 3

The proof of Proposition 3 recalls the proof from [9], Proposition 5. Let us define the functions

$$\begin{aligned}\mathcal{V}_N^\circ(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}) &= \sum_{k\in\mathcal{S}} \frac{1}{N}\log\det\left(\mathbf{I}_{n_k} + c_k e_k \mathbf{R}_k\mathbf{P}_k\right) \\ &\quad + \frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k\in\mathcal{S}}\delta_k\mathbf{T}_k\right) \\ &\quad - \sigma^2\sum_{k=1}^{K}\delta_k(-\sigma^2)e_k(-\sigma^2)\end{aligned} \tag{53}$$

where

$$e_i = e_i(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}) = \frac{1}{N}\operatorname{tr}\mathbf{T}_i\left(\sigma^2\left[\mathbf{I}_N + \sum_{k\in\mathcal{S}}\delta_k\mathbf{T}_k\right]\right)^{-1} \tag{54}$$

$$\delta_i = \delta_i(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}) = \frac{1}{n_i}\operatorname{tr}\mathbf{R}_i\mathbf{P}_i\left(\sigma^2\left[\mathbf{I}_{n_i} + c_i e_i(z)\mathbf{R}_i\mathbf{P}_i\right]\right)^{-1} \tag{55}$$

and

$$V : (\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}, \delta_1, \ldots, \delta_{|\mathcal{S}|}, e_1, \ldots, e_{|\mathcal{S}|}) \mapsto \mathcal{V}_N^\circ(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}).$$

Then we need only prove that, for all $k \in \mathcal{S}$

$$\begin{aligned}\frac{\partial V}{\partial\delta_k}(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}, \delta_1^\circ, \ldots, \delta_{|\mathcal{S}|}^\circ, e_1^\circ, \ldots, e_{|\mathcal{S}|}^\circ) &= 0 \\ \frac{\partial V}{\partial e_k}(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}, \delta_1^\circ, \ldots, \delta_{|\mathcal{S}|}^\circ, e_1^\circ, \ldots, e_{|\mathcal{S}|}^\circ) &= 0.\end{aligned}$$

Remark then that

$$\begin{aligned}&\frac{\partial V}{\partial\delta_k}(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}, \delta_1, \ldots, \delta_{|\mathcal{S}|}, e_1, \ldots, e_{|\mathcal{S}|}) \\ &\quad = \frac{1}{N}\operatorname{tr}\left[\left(\mathbf{I} + \sum_{i\in\mathcal{S}}\delta_i\mathbf{T}_i\right)^{-1}\mathbf{T}_k\right] - \sigma^2 e_k\end{aligned} \tag{56}$$

$$\begin{aligned}&\frac{\partial V}{\partial e_k}(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}, \delta_1, \ldots, \delta_{|\mathcal{S}|}, e_1, \ldots, e_{|\mathcal{S}|}) \\ &\quad = c_k\frac{1}{N}\operatorname{tr}\left[\left(\mathbf{I} + c_k e_k \mathbf{R}_i\mathbf{P}_i\right)^{-1}\mathbf{R}_k\mathbf{P}_k\right] - \sigma^2\delta_k\end{aligned} \tag{57}$$

both being null whenever, for all k, $e_k = e_k(-\sigma^2, \mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|})$ and $\delta_k = \delta_k(-\sigma^2, \mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|})$, which is true in particular for the unique power optimal solution $\mathbf{P}_1^\circ, \ldots, \mathbf{P}_{|\mathcal{S}|}^\circ$ whenever $e_k = e_k^\circ$ and $\delta_k = \delta_k^\circ$.

When, for all k, $e_k = e_k^\circ$, $\delta_k = \delta_k^\circ$, the maximum of V over the $\mathbf{P}_k$ is then obtained by maximizing the expressions $\log\det(\mathbf{I}_{n_k} + c_k e_k^\circ \mathbf{R}_k\mathbf{P}_k)$ over $\mathbf{P}_k$. From the inequality (see, e.g., [2])

$$\det(\mathbf{I}_{n_k} + c_k e_k^\circ \mathbf{R}_k\mathbf{P}_k) \leq \prod_{i=1}^{n_k}(\mathbf{I}_{n_k} + c_k e_k^\circ \mathbf{R}_k\mathbf{P}_k)_{ii}$$

where, only here, we denote $(\mathbf{X})_{ii}$ the entry (i,i) of matrix $\mathbf{X}$. The equality is obtained if and only if $\mathbf{I}_{n_k} + c_k e_k^\circ \mathbf{R}_k\mathbf{P}_k$ is diagonal. The equality case arises for $\mathbf{P}_k$ and $\mathbf{R}_k = \mathbf{U}_k\mathbf{D}_k\mathbf{U}_k^{\mathsf{H}}$ co-diagonalizable. In this case, denoting $\mathbf{P}_k = \mathbf{U}_k\mathbf{Q}_k\mathbf{U}_k^{\mathsf{H}}$, the entries of $\mathbf{Q}_k$, constrained by $\frac{1}{n_k}\operatorname{tr}(\mathbf{Q}_k) = P_k$ are solutions of the classical optimization problem under constraint

$$\sup_{\substack{\mathbf{Q}_k \\ \frac{1}{n_k}\operatorname{tr}(\mathbf{Q}_k)\leq P_k}} \log\det\left(\mathbf{I}_{n_k} + c_k e_k^\circ \mathbf{Q}_k\mathbf{D}_k\right)$$

whose solution is given by the classical water-filling algorithm. Hence, (15).

APPENDIX E
PROOF OF PROPOSITION 1

The convergence of the fixed-point algorithm follows the same line of proof as the uniqueness in Section A-E. We prove the convergence for $z \in \mathbb{C}^+$, although this can be easily generalized. If one considers the difference $\mathbf{e}^{n+1} - \mathbf{e}^n$, where $\mathbf{e}^n = (e_1^n, \ldots, e_K^n)$, instead of $\mathbf{e}^\circ - \underline{\mathbf{e}}^\circ$, the same development as in Section A-E leads to

$$\mathbf{e}^{n+1} - \mathbf{e}^n = \mathbf{A}_n(\mathbf{e}^n - \mathbf{e}^{n-1})$$

for $n \geq 1$, where $\mathbf{A}_n$ is defined, similarly as in (37), as $\mathbf{A}_n = (a_{ij}^n)$, with a_{ij}^n defined by

$$a_{ij}^n = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \mathbf{D}_{n-1}{}^{-1} \mathbf{R}_j \mathbf{D}_n{}^{-1} c_j \times \int \frac{\tau^2}{(1 + c_j \tau e_j^{n-1})(1 + c_j \tau e_j^n)} dF^{\mathbf{T}_j}(\tau),$$

where $\mathbf{D}_n$ is $\mathbf{D}$ for $e_j(z)$ replaced by $e_j^n(z)$.

From Cauchy-Schwarz inequality, and the different bounds on the $\mathbf{D}_n$, $\mathbf{R}_k$ and $\mathbf{T}_k$ matrices used so far, we have

$$a_{ij}^n \leq \frac{|z|^2 c_j}{v^4} \frac{\log N^4}{N}$$

with $v = \Im[z]$. Denoting $c_0 = \max(c_j)$, we then have that

$$\max_j \left(e_j^{n+1} - e_j^n \right) < K \frac{|z|^2 c_0}{v^4} \leq \frac{\log N^4}{N} \max_j \left(e_j^n - e_j^{n-1} \right).$$

Let $0 < \varepsilon < 1$, and take now a countable set $\{z_1, z_2, \ldots\}$, $v_k = \Im[z_k]$, such that $K \frac{|z_k|^2 c_0}{v_k^4} \frac{\log N^4}{N} < 1 - \varepsilon$ for all z_k (this is possible by letting $v_k > 0$ be large enough). On this countable set, the sequences $\{\mathbf{e}^n\}$ are therefore Cauchy sequences on $\mathbb{C}^K$: they all converge. Since the e_j^n are holomorphic and bounded on every compact set included in $\mathbb{C} \setminus \mathbb{R}^+$, from Vitali's convergence theorem [30], the function $e_j^n(z)$ converges on such compact sets. Now, from the fact that we forced the initialization step to be $e_j^0 = -1/z$, e_j^0 is the Stieltjes transform of a distribution function at point z. It now suffices to verify that, if e_j^n is the Stieltjes transform of a distribution function at point z, then so is e_j^{n+1}. This requires to verify that $z \in \mathbb{C}^+$, $e_j^n \in \mathbb{C}^+$ implies $e_j^{n+1} \in \mathbb{C}^+$, $z \in \mathbb{C}^+$, $z e_j^n \in \mathbb{C}^+$ implies $z e_j^{n+1} \in \mathbb{C}^+$, and $\lim_{y \to \infty} -y e_j^n(iy) < \infty$ implies that $\lim_{y \to \infty} -y e_j^n(iy) < \infty$. This follows directly from the definition of e_j^n. From the dominated convergence theorem, we then also have that the limit of e_j^n is a Stieltjes transform that is solution to (6). From the uniqueness of the Stieltjes transform, solution to (6) (this follows from the pointwise uniqueness on $\mathbb{C}^+$ and the fact that the Stieltjes transform is holomorphic on all compact sets of $\mathbb{C} \setminus \mathbb{R}^+$), we then have that e_j^n converges for all j and $z \in \mathbb{C} \setminus \mathbb{R}^+$, if e_j^0 is initialized at a Stieltjes transform.

APPENDIX F
USEFUL LEMMAS

In this section, we gather most of the known or new lemmas which are needed in various places in the proof of Appendices A–E.

The statements in the following Lemma are well-known

Lemma 1:

1) For rectangular matrices $\mathbf{A}$, $\mathbf{B}$ of the same size
$$\operatorname{rank}(\mathbf{A} + \mathbf{B}) \leq \operatorname{rank}(\mathbf{A}) + \operatorname{rank}(\mathbf{B}).$$
2) For rectangular matrices $\mathbf{A}$, $\mathbf{B}$ for which $\mathbf{AB}$ is defined,
$$\operatorname{rank}(\mathbf{AB}) \leq \min(\operatorname{rank}(\mathbf{A}), \operatorname{rank}(\mathbf{B})),$$
3) For rectangular $\mathbf{A}$, $\operatorname{rank}(\mathbf{A})$ is less than the number of non-zero entries of $\mathbf{A}$.

Lemma 2: (Lemma 2.4 of [22]) For $N \times N$ Hermitian matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\|F^{\mathbf{A}} - F^{\mathbf{B}}\| \leq \frac{1}{N} \operatorname{rank}(\mathbf{A} - \mathbf{B}).$$

From these two lemmas we get the following.

Lemma 3: Let $\mathbf{S}$, $\mathbf{A}$, $\overline{\mathbf{A}}$, be Hermitian $N \times N$, $\mathbf{Q}$, $\overline{\mathbf{Q}}$ both $N \times n$, and $\mathbf{B}$, $\overline{\mathbf{B}}$ both Hermitian $n \times n$. Then:

1)
$$\|F^{\mathbf{S}+\mathbf{AQBQ}^{\mathsf{H}}\mathbf{A}} - F^{\mathbf{S}+\mathbf{A}\overline{\mathbf{Q}}\mathbf{B}\overline{\mathbf{Q}}^{\mathsf{H}}\mathbf{A}}\| \leq \frac{2}{N} \operatorname{rank}(\mathbf{Q} - \overline{\mathbf{Q}});$$

2)
$$\|F^{\mathbf{S}+\mathbf{AQBQ}^{\mathsf{H}}\mathbf{A}} - F^{\mathbf{S}+\overline{\mathbf{A}}\mathbf{QBQ}^{\mathsf{H}}\overline{\mathbf{A}}}\| \leq \frac{2}{N} \operatorname{rank}(\mathbf{A} - \overline{\mathbf{A}});$$

3)
$$\|F^{\mathbf{S}+\mathbf{AQBQ}^{\mathsf{H}}\mathbf{A}} - F^{\mathbf{S}+\mathbf{AQ}\overline{\mathbf{B}}\mathbf{Q}^{\mathsf{H}}\mathbf{A}}\| \leq \frac{1}{N} \operatorname{rank}(\mathbf{B} - \overline{\mathbf{B}}).$$

Lemma 4: For $N \times N$ $\mathbf{A}$, $\tau \in \mathbb{C}$ and $\mathbf{r} \in \mathbb{C}^N$ for which $\mathbf{A}$ and $\mathbf{A} + \tau \mathbf{r}\mathbf{r}^{\mathsf{H}}$ are invertible,

$$\mathbf{r}^{\mathsf{H}}(\mathbf{A} + \tau \mathbf{r}\mathbf{r}^{\mathsf{H}})^{-1} = \frac{1}{1 + \tau \mathbf{r}^{\mathsf{H}} \mathbf{A}^{-1} \mathbf{r}} \mathbf{r}^{\mathsf{H}} \mathbf{A}^{-1}.$$

This result follows from $\mathbf{r}^{\mathsf{H}} \mathbf{A}^{-1}(\mathbf{A} + \tau \mathbf{r}\mathbf{r}^{\mathsf{H}}) = (1 + \tau \mathbf{r}^{\mathsf{H}} \mathbf{A}^{-1} \mathbf{r}) \mathbf{r}^{\mathsf{H}}$.

Moreover, we recall [22. Lemma 2.6].

Lemma 5: Let $z \in \mathbb{C}^+$ with $v = \Im[z]$, $\mathbf{A}$ and $\mathbf{B}$ $N \times N$ with $\mathbf{B}$ Hermitian, and $\mathbf{r} \in \mathbb{C}^N$. Then

$$\left| \operatorname{tr} \left((\mathbf{B} - z\mathbf{I}_N)^{-1} - (\mathbf{B} + \mathbf{r}\mathbf{r}^{\mathsf{H}} - z\mathbf{I}_N)^{-1} \right) \mathbf{A} \right| = \left| \frac{\mathbf{r}^{\mathsf{H}} (\mathbf{B} - z\mathbf{I}_N)^{-1} \mathbf{A} (\mathbf{B} - z\mathbf{I}_N)^{-1} \mathbf{r}}{1 + \mathbf{r}^{\mathsf{H}} (\mathbf{B} - z\mathbf{I}_N)^{-1} \mathbf{r}} \right| \leq \frac{\|\mathbf{A}\|}{v}.$$

If $z < 0$, we also have

$$\left| \operatorname{tr} \left((\mathbf{B} - z\mathbf{I}_N)^{-1} - (\mathbf{B} + \mathbf{r}\mathbf{r}^{\mathsf{H}} - z\mathbf{I}_N)^{-1} \right) \mathbf{A} \right| \leq \frac{\|\mathbf{A}\|}{|z|}.$$

From [32, Lemma 2.2] and [33, Theorems A.2, A.4, A.5], we have the following.

Lemma 6: If f is analytic on $\mathbb{C}^+$, both $f(z)$ and $zf(z)$ map $\mathbb{C}^+$ into $\mathbb{C}^+$, and there exists a $\theta \in (0, \pi/2)$ for which $zf(z) \to c$, finite, as $z \to \infty$ restricted to $\{w \in \mathbb{C} : \theta < \arg w < \pi - \theta\}$, then $c < 0$ and f is the Stieltjes transform of a measure on the nonnegative reals with total mass $-c$.

Also, from [22], we need

Lemma 7: Let $\mathbf{y} = (y_1, \ldots, y_N)^{\mathsf{T}}$ with the y_i's i.i.d. such that $\mathrm{E}y_1 = 0$, $\mathrm{E}|y_1|^2 = 1$ and $y_1 \leq \log N$, and $\mathbf{A}$ an $N \times N$ matrix independent of $\mathbf{y}$, then

$$\mathrm{E}|\mathbf{y}^{\mathsf{H}}\mathbf{A}\mathbf{y} - \operatorname{tr}\mathbf{A}|^6 \leq K\|\mathbf{A}\|^6 N^3 \log^{12} N$$

where K does not depend on N, $\mathbf{A}$, nor on the distribution of y_1.

Additionally, we need

Lemma 8: Let $\mathbf{D} = \mathbf{A} + i\mathbf{B} + iv\mathbf{I}$, where $\mathbf{A}, \mathbf{B}$ are $N \times N$ Hermitian, B is also positive semi-definite, and $v > 0$. Then $\|\mathbf{D}^{-1}\| \leq v^{-1}$.

Proof: We have $\mathbf{D}\mathbf{D}^{\mathsf{H}} = (\mathbf{A} + i\mathbf{B})(\mathbf{A} - i\mathbf{B}) + v^2\mathbf{I} + 2v\mathbf{B}$. Therefore the eigenvalues of $\mathbf{D}\mathbf{D}^{\mathsf{H}}$ are greater or equal to v^2, which implies the singular values of $\mathbf{D}$ are greater or equal to v, so that the singular values of $\mathbf{D}^{-1}$ are less or equal to v^{-1}. We therefore get our result. ∎

From Theorem 2.1 of [34],

Lemma 9: Let $\rho(\mathbf{C})$ denote the spectral radius of the $N \times N$ matrix $\mathbf{C}$ (the largest of the absolute values of the eigenvalues of $\mathcal{C}$). If $\mathbf{x}, \mathbf{b} \in \mathbb{R}^N$ with the components of $\mathbf{C}$, $\mathbf{x}$, and $\mathbf{b}$ all positive, then the equation $\mathbf{x} = \mathbf{C}\mathbf{x} + \mathbf{b}$ implies $\rho(\mathbf{C}) < 1$.

From [35, Theorem 8.1.18],

Lemma 10: Suppose $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ are $N \times N$ with b_{ij} nonnegative and $|a_{ij}| \leq b_{ij}$. Then

$$\rho(\mathbf{A}) \leq \rho((|a_{ij}|)) \leq \rho(\mathbf{B}).$$

Also, from [36, Lemma 5.7.9],

Lemma 11: Let $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ be $N \times N$ with a_{ij}, b_{ij} nonnegative. Then

$$\rho((a_{ij}^{\frac{1}{2}} b_{ij}^{\frac{1}{2}})) \leq (\rho(\mathbf{A}))^{\frac{1}{2}} (\rho(\mathbf{B}))^{\frac{1}{2}}.$$

And [35, Theorems 8.2.2 and 8.3.1].

Lemma 12: If $\mathbf{C}$ is a square matrix with nonnegative entries, then $\rho(\mathbf{C})$ is an eigenvalue of $\mathbf{C}$ having an eigenvector $\mathbf{x}$ with nonnegative entries. Moreover, if the entries of $\mathbf{C}$ are all positive, then $\rho(\mathbf{C}) > 0$ and the entries of $\mathbf{x}$ are all positive.

From [36], we also need Theorem 6.1.1,

Lemma 13: Gersgorin's Theorem All the eigenvalues of an $N \times N$ matrix $\mathbf{A} = (a_{ij})$ lie in the union of the N disks in the complex plane, the i^{th} disk having center a_{ii} and radius $\sum_{j \neq i} |a_{ij}|$.

Theorem 3.42 of [30],

Lemma 14: Rouche's Theorem If $f(z)$ and $g(z)$ are analytic inside and on a closed contour C of the complex plane, and $|g(z)| < |f(z)|$ on C, then $f(z)$ and $f(z) + g(z)$ have the same number of zeros inside C.

In order to prove Theorem 2, we also need, from [37]

Lemma 15: Consider a rectangular matrix $\mathbf{A}$ and let $s_i^{\mathbf{A}}$ denote the i^{th} largest singular value of $\mathbf{A}$, with $s_i^{\mathbf{A}} = 0$ whenever $i > \operatorname{rank}(A)$. Let m, n be arbitrary non-negative integers. Then for $\mathbf{A}$, $\mathbf{B}$ rectangular of the same size

$$s_{m+n+1}^{\mathbf{A}+\mathbf{B}} \leq s_{m+1}^{\mathbf{A}} + s_{n+1}^{\mathbf{B}},$$

and for $\mathbf{A}$, $\mathbf{B}$ rectangular for which $\mathbf{AB}$ is defined

$$s_{m+n+1}^{AB} \leq s_{m+1}^{A} s_{n+1}^{B}.$$

As a corollary, for any integer $r \geq 0$ and rectangular matrices $\mathbf{A}_1, \ldots, \mathbf{A}_K$, all of the same size

$$s_{Kr+1}^{\mathbf{A}_1 + \cdots + \mathbf{A}_K} \leq s_{r+1}^{\mathbf{A}_1} + \cdots + s_{r+1}^{\mathbf{A}_K}.$$

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Feb. 1999.

[3] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[4] M. Vu and A. Paulraj, "Capacity optimization for Rician correlated MIMO wireless channels," in *Proc. IEEE Conf. Rec. Asilomar Conf. Signals, Systems, and Computers (Asilomar'05)*, Pacific Grove, CA, 2005, pp. 133–138.

[5] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," *Ann. Appl. Probabil.*, vol. 17, no. 3, pp. 875–930, 2007.

[6] A. M. Tulino and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2491–2509, Jul. 2005.

[7] M. J. M. Peacock, I. B. Collings, and M. L. Honig, "Eigenvalue distributions of sums and products of large random matrices via incremental matrix expansions," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2123–2138, May 2008.

[8] A. Soysal and S. Ulukus, "Optimality of beamforming in fading MIMO multiple access channels," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 1171–1183, Apr. 2009.

[9] J. Dumont, W. Hachem, P. Loubaton, and J. Najim, "On the capacity achieving covariance matrix for Rician MIMO channels: An asymptotic approach," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1048–1069, Mar. 2010.

[10] W. Hachem, P. Loubaton, and J. Najim, "A CLT for information theoretic statistics of Gram random matrices with a given variance profile," *Ann. Probabil.*, vol. 18, no. 6, pp. 2071–2130, Dec. 2008.

[11] A. L. Moustakas and S. H. Simon, "On the outage capacity of correlated multiple-path MIMO channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 3887–3903, Nov. 2007.

[12] M. Mézard, G. Parisi, and M. Virasoro, "Spin glass theory and beyond," *Phys. Today*, vol. 41, no. 12, 1988.

[13] F. Dupuy and P. Loubaton, "On the capacity achieving covariance matrix for frequency selective MIMO channels using the asymptotic approach," *IEEE Trans. Inf. Theory*, submitted for publication.

[14] R. R. Far, T. Oraby, W. Bryc, and R. Speicher, "On slow-fading MIMO systems with nonseparable correlation," *IEEE Trans. Inf. Theory*, vol. 54, no. 2, pp. 544–553, Feb. 2008.

[15] C. N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[16] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, ser. Statistics, 2nd ed. New York: Springer, 2009.

[17] V. A. Marcenko and L. A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math USSR-Sbornik*, vol. 1, no. 4, pp. 457–483, Apr. 1967.

[18] D. N. C. Tse and S. V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 641–657, Feb. 1999.

[19] M. Debbah, W. Hachem, P. Loubaton, and M. de Courville, "MMSE analysis of certain large isometric random precoded systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1293–1311, May 2003.

[20] P. Billingsley, *Probability and Measure*, 3rd ed. Hoboken, NJ: Wiley, 1995.
[21] L. Zhang, "Spectral Analysis of Large Dimensional Random Matrices," Ph.D. dissertation, National University of Singapore, Singapore, 2006.
[22] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *J. Multivariate Anal.*, vol. 54, no. 2, pp. 175–192, 1995.
[23] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large N analysis," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.
[24] T. Pollock, T. Abhayapala, and R. Kennedy, "Antenna saturation effects on dense array MIMO capacity," in *Proc. IEEE Int. Conf. Communications (ICC'03)*, Anchorage, AK, 2003, pp. 2301–2305.
[25] A. L. Moustakas, H. U. Baranger, L. Balents, A. M. Sengupta, and S. Simon, "Communication through a diffusive medium: Coherence and capacity," *Science*, vol. 287, pp. 287–290, 2000.
[26] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 684–702, May 2003.
[27] R. Clarke, "A statistical theory of mobile-radio reception," *Bell Syst. Tech. J.*, vol. 47, no. 6, pp. 957–1000, 1968.
[28] S. Viswanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcsat channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.
[29] R. B. Dozier and J. W. Silverstein, "On the empirical distribution of eigenvalues of large dimensional information plus noise-type matrices," *J. Multivariate Anal.*, vol. 98, no. 4, pp. 678–694, 2007.
[30] E. C. Titchmarsh, *The Theory of Functions*. New York: Oxford University Press, 1939.
[31] P. Billingsley, *Convergence of Probability Measures*. Hoboken, NJ: Wiley, 1968.
[32] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*. Providence, RI: American Mathematical Society, 1970.
[33] M. K. Krein and A. A. Nudelman, *The Markov Moment Problem and Extremal Problems*. Providence, RI: American Mathematical Society, 1997.
[34] E. Seneta, *Non-negative Matrices and Markov Chains*, Second ed. New York: Springer-Verlag, 1981.
[35] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York: Cambridge University Press, 1985.
[36] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. New York: Cambridge University Press, 1991.
[37] K. Fan, "Maximum properties and inequalities for the eigenvalues of completely continuous operators," in *Proc. Nat. Academy of Sciences of the United States of America*, 1951, vol. 37, pp. 760–766.

Romain Couillet (M'11) was born in Abbeville, France. He received the M.Sc. degree in mobile communications from the Eurecom Institute, Sophia Antipolis, France, in 2007, and the M.Sc. degree in communication systems from Telecom ParisTech, Paris, France in 2007, and the Ph.D. degree from Supélec, Gif-sur-Yvette, France, in November 2010.

From 2007 to 2010, he was an Algorithm Development Engineer with ST-Ericsson, working on the Long Term Evolution Advanced (LTE-A) project. He is currently an Assistant Professor in the EDF Chair ''Systems Science and the Energetic Challenge'' with Supelec and Centrale Paris, France. His research topics are in information theory, signal processing, complex systems, and random matrix theory.

Dr. Couillet is the recipient of the Valuetools 2008 Best Student Paper award and of the 2011 EEA/GdR ISIS/GRETSI Best Ph.D. Thesis award.

Mérouane Debbah (SM'07) was born in Madrid, Spain. He received the M.Sc and Ph.D. degrees from the Ecole Normale Supérieure de Cachan, France, in 1999 and 2002, respectively.

From 1999 to 2002, he was with Motorola Labs, working on wireless local area networks and prospective fourth-generation systems. From 2002 to 2003, he was appointed Senior Researcher at the Vienna Research Center for Telecommunications (FTW), Vienna, Austria. From 2003 to 2007, he was an Assistant Professor with the Mobile Communications Department, Institut Eurecom, Sophia Antipolis, France. He is presently a Professor at Supelec, Gif-sur-Yvette, France, and holder of the Alcatel-Lucent Chair on Flexible Radio. His research interests are in information theory, signal processing, and wireless communications.

Dr. Debbah was the recipient of the "Mario Boella" prize award in 2005, the 2007 General Symposium IEEE GLOBECOM best paper award, the Wi-Opt 2009 best paper award, the 2010 Newcom++ best paper award, as well as the Valuetools 2007, Valuetools 2008, and CrownCom2009 best student paper awards. He is a WWRF Fellow.

Jack W. Silverstein received the B.A. degree in mathematics from Hofstra University, Hempstead, NY, in 1971 and the M.S. and Ph.D. degrees in applied mathematics from Brown University, Providence, RI, in 1973 and 1975, respectively.

After postdoctoral work and teaching at Brown University, in 1978 he began a tenure track position in the Department of Mathematics, North Carolina State University, Raleigh, where he has been a Professor since 1994. His research interests are in probability theory, with emphasis on the spectral behavior of large-dimensional random matrices.

Dr. Silverstein was elected Fellow of the Institute of Mathematical Statistics in 2007.

Capacity Limits of MIMO Channels

Andrea Goldsmith, *Senior Member, IEEE*, Syed Ali Jafar, *Student Member, IEEE*, Nihar Jindal, *Student Member, IEEE*, and Sriram Vishwanath, *Student Member, IEEE*

Invited Paper

***Abstract*—We provide an overview of the extensive recent results on the Shannon capacity of single-user and multiuser multiple-input multiple-output (MIMO) channels. Although enormous capacity gains have been predicted for such channels, these predictions are based on somewhat unrealistic assumptions about the underlying time-varying channel model and how well it can be tracked at the receiver, as well as at the transmitter. More realistic assumptions can dramatically impact the potential capacity gains of MIMO techniques. For time-varying MIMO channels there are multiple Shannon theoretic capacity definitions and, for each definition, different correlation models and channel information assumptions that we consider. We first provide a comprehensive summary of ergodic and capacity versus outage results for single-user MIMO channels. These results indicate that the capacity gain obtained from multiple antennas heavily depends on the available channel information at either the receiver or transmitter, the channel signal-to-noise ratio, and the correlation between the channel gains on each antenna element. We then focus attention on the capacity region of the multiple-access channels (MACs) and the largest known achievable rate region for the broadcast channel. In contrast to single-user MIMO channels, capacity results for these multiuser MIMO channels are quite difficult to obtain, even for constant channels. We summarize results for the MIMO broadcast and MAC for channels that are either constant or fading with perfect instantaneous knowledge of the antenna gains at both transmitter(s) and receiver(s). We show that the capacity region of the MIMO multiple access and the largest known achievable rate region (called the dirty-paper region) for the MIMO broadcast channel are intimately related via a duality transformation. This transformation facilitates finding the transmission strategies that achieve a point on the boundary of the MIMO MAC capacity region in terms of the transmission strategies of the MIMO broadcast dirty-paper region and *vice-versa*. Finally, we discuss capacity results for multicell MIMO channels with base station cooperation. The base stations then act as a spatially diverse antenna array and transmission strategies that exploit this structure exhibit significant capacity gains. This section also provides a brief discussion of system level issues associated with MIMO cellular. Open problems in this field abound and are discussed throughout the paper.**

***Index Terms*—Antenna correlation, beamforming, broadcast channels (BCs), channel distribution information (CDI), channel state information (CSI), multicell systems, multiple-access channels (MACs), multiple-input multiple-output (MIMO) channels, multiuser systems, Shannon capacity.**

Manuscript received November 8, 2002; revised January 31, 2003. This work was supported in part by the Office of Naval Research (ONR) under Grants N00014-99-1-0578 and N00014-02-1-0003. The work of S. Vishwanath was supported by a Stanford Graduate Fellowship.

The authors are with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: andrea@wsl.stanford.edu; syed@wsl.stanford.edu; njindal@wsl.stanford.edu; sriram@wsl.stanford.edu).

Digital Object Identifier 10.1109/JSAC.2003.810294

I. Introduction

WIRELESS systems continue to strive for ever higher data rates. This goal is particularly challenging for systems that are power, bandwidth, and complexity limited. However, another domain can be exploited to significantly increase channel capacity: the use of multiple transmit and receive antennas. Pioneering work by Winters [81], Foschini [20], and Telatar [69] ignited much interest in this area by predicting remarkable spectral efficiencies for wireless systems with multiple antennas when the channel exhibits rich scattering and its variations can be accurately tracked. This initial promise of exceptional spectral efficiency almost "for free" resulted in an explosion of research activity to characterize the theoretical and practical issues associated with multiple-input multiple-output (MIMO) wireless channels and to extend these concepts to multiuser systems. This tutorial summarizes the segment of this recent work focused on the capacity of MIMO systems for both single-users and multiple users under different assumptions about spatial correlation and channel information available at the transmitter and receiver.

The large spectral efficiencies associated with MIMO channels are based on the premise that a rich scattering environment provides independent transmission paths from each transmit antenna to each receive antenna. Therefore, for single-user systems, a transmission and reception strategy that exploits this structure achieves capacity on approximately $\min(M, N)$ separate channels, where M is the number of transmit antennas and N is the number of receive antennas. Thus, capacity scales linearly with $\min(M, N)$ relative to a system with just one transmit and one receive antenna. This capacity increase requires a scattering environment such that the matrix of channel gains between transmit and receive antenna pairs has full rank and independent entries and that perfect estimates of these gains are available at the receiver. Perfect estimates of these gains at both the transmitter and receiver provides an increase in the constant multiplier associated with the linear scaling. Much subsequent work has been aimed at characterizing MIMO channel capacity under more realistic assumptions about the underlying channel model and the channel estimates available at the transmitter and receiver. The main question from both a theoretical and practical standpoint is whether the enormous capacity gains initially predicted by Winters, Foschini, and Telatar can be obtained in more realistic operating scenarios and what specific gains result from adding more antennas and/or a feedback link to feed receiver channel information back to the transmitter.

MIMO channel capacity depends heavily on the statistical properties and antenna element correlations of the channel. Recent work has developed both analytical and measurement-based MIMO channel models along with the corresponding capacity calculations for typical indoor and outdoor environments [26]. Antenna correlation varies drastically as a function of the scattering environment, the distance between transmitter and receiver, the antenna configurations, and the Doppler spread [1], [65]. As we shall see, the effect of channel correlation on capacity depends on what is known about the channel at the transmitter and receiver: correlation sometimes increases capacity and sometimes reduces it [16]. Moreover, channels with very low correlation between antennas can still exhibit a "keyhole" effect where the rank of the channel gain matrix is very small, leading to limited capacity gains [12]. Fortunately, this effect is not prevalent in most environments. The impact of channel statistics in the low-power (wideband) regime has interesting properties as well: recent results in this area can be found in [71].

We focus on MIMO channel capacity in the Shannon theoretic sense. The Shannon capacity of a single-user time-invariant channel is defined as the maximum mutual information between the channel input and output. This maximum mutual information is shown by Shannon's capacity theorem to be the maximum data rate that can be transmitted over the channel with arbitrarily small error probability. When the channel is time-varying channel capacity has multiple definitions, depending on what is known about the channel state or its distribution at the transmitter and/or receiver and whether capacity is measured based on averaging the rate over all channel states/distributions or maintaining a constant fixed or minimum rate. Specifically, when the instantaneous channel gains, called the channel state information (CSI), are known perfectly at both transmitter and receiver, the transmitter can adapt its transmission strategy relative to the instantaneous channel state. In this case, the Shannon (ergodic) capacity is the maximum mutual information averaged over all channel states. This ergodic capacity is typically achieved using an adaptive transmission policy where the power and data rate vary relative to the channel state variations. Other capacity definitions for time-varying channels with perfect transmitter and receiver CSI include outage capacity and minimum-rate capacity. These capacities require a fixed or minimum data rate in all nonoutage channel states, which is needed for applications with delay-constrained data where the data rate cannot depend on channel variations (except in outage states, where no data is transmitted). The average rate associated with outage or minimum rate capacity is typically smaller than ergodic capacity due to the additional constraints associated with these definitions. This tutorial will focus on ergodic capacity in the case of perfect transmitter and receiver CSI.

When only the channel distribution is known at the transmitter (receiver) the transmission (reception) strategy is based on the channel distribution instead of the instantaneous channel state. The channel coefficients are typically assumed to be jointly Gaussian, so the channel distribution is specified by the channel mean and covariance matrices. We will refer to knowledge of the channel distribution as channel distribution information (CDI). We assume throughout the paper that CDI is always perfect, so there is no mismatch between the CDI at the transmitter or receiver and the true channel distribution. When only the receiver has perfect CSI the transmitter must maintain a fixed-rate transmission strategy optimized with respect to its CDI. In this case, ergodic capacity defines the rate that can be achieved based on averaging over all channel states [69]. Alternatively, the transmitter can send at a rate that cannot be supported by all channel states: in these poor channel states the receiver declares an outage and the transmitted data is lost. In this scenario, each transmission rate has an outage probability associated with it and capacity is measured relative to outage probability[1] (capacity CDF) [20]. An excellent tutorial on fading channel capacity for single antenna channels can be found in [4]. For single-user MIMO channels with perfect transmitter and receiver CSI the ergodic and outage capacity calculations are straightforward since the capacity is known for every channel state. Thus, for single-user MIMO systems the tutorial will focus on capacity results assuming perfect CDI at the transmitter and perfect CSI or CDI at the receiver. Although there has been much recent progress in this area, many open problems remain.

In multiuser channels, capacity becomes a K-dimensional region defining the set of all rate vectors $(R_1, \ldots, R_K)$ simultaneously achievable by all K users. The multiple capacity definitions for time-varying channels under different transmitter and receiver CSI and CDI assumptions extend to the capacity region of the multiple-access channel (MAC) and broadcast channel (BC) in the obvious way [28], [48], [49], [70]. However, these MIMO multiuser capacity regions, even for time-invariant channels, are difficult to find. Few capacity results exist for time-varying multiuser MIMO channels, especially under the realistic assumption that the transmitter(s) and/or receiver(s) have CDI only. Therefore, the tutorial focus for MIMO multiuser systems will be on ergodic capacity under perfect CSI at the transmitter and receiver, with a brief discussion of the known results and open problems for other capacity definitions and CSI/CDI assumptions.

Note that the MIMO techniques described herein are applicable to any channel described by a matrix. Matrix channels describe not only multiantenna systems but also channels with crosstalk [85] and wideband channels [72]. While the focus of this tutorial is on memoryless channels (flat-fading), the results can also be extended to channels with memory (ISI) using well-known methods for incorporating the channel delay spread into the channel matrix [59], as will be discussed in the next section.

Many practical MIMO techniques have been developed to capitalize on the theoretical capacity gains predicted by Shannon theory. A major focus of such work is space-time coding: recent work in this area is summarized in [21]. Other techniques for MIMO systems include space–time modulation [30], [33], adaptive modulation and coding [10], space–time

[1]Note that an outage under perfect CSI at the receiver only is different than an outage when both transmitter and receiver have perfect CSI. Under receiver CSI only an outage occurs when the transmitted data cannot be reliably decoded at the receiver, so that data is lost. When both the transmitter and receiver have perfect CSI the channel is not used during outage (no service), so no data is lost.

TABLE I
TABLE OF ABBREVIATIONS

CSI	Channel State Information
CDI	Channel Distribution Information
CSIT	Transmitter Channel State Information
CSIR	Receiver Channel State Information
CDIT	Transmitter Channel Distribution Information
CDIR	Receiver Channel Distribution Information
ZMSW	Zero Mean Spatially White
CMI	Channel Mean Information
CCI	Channel Covariance Information
DPC	Dirty Paper Coding
MAC	Multiple-Access Channel
BC	Broadcast Channel

equalization [2], [51], space–time signal processing [3], space–time CDMA [14], [34], and space–time OFDM [50], [52], [82]. An overview of the recent advances in these areas and other practical techniques along with their performance can be found in [25].

The remainder of this paper is organized as follows. In Section II, we discuss the capacity of single-user MIMO systems under different assumptions about channel state and distribution information at the transmitter and receiver. This section also describes the optimality of beamforming and training issues. Section III describes the capacity region of the MIMO MAC and the "dirty-paper" achievable region of the MIMO BC, along with a duality connection between these regions. The capacity of multicell systems under dirty paper coding (DPC) and opportunistic beamforming is discussed in Section IV, as well as tradeoffs between capacity, diversity, and sectorization. Section V summarizes these capacity results and describes some remaining open problems and design questions associated with MIMO systems.

A note on notation: We use boldface to denote matrices and vectors and $\mathbb{E}[\cdot]$ for expectation. $|\mathbf{S}|$ denotes the determinant and $\mathbf{S}^{-1}$ the inverse of a square matrix $\mathbf{S}$. For any general matrix $\mathbf{M}$, $\mathbf{M}^\dagger$ denotes the conjugate transpose and $\mathrm{Tr}(\mathbf{M})$ denotes the trace. $\mathbf{I}$ denotes the identity matrix and $\mathrm{diag}(\lambda_i)$ denotes a diagonal matrix with the (i,i) entry equal to λ_i. For symmetric matrices the notation $\mathbf{Q} \geq 0$ implies that $\mathbf{Q}$ is positive semidefinite.

A table of abbreviations used throughout the paper is given in Table I.

II. SINGLE-USER MIMO

In this section, we focus on the capacity of single-user MIMO channels. While most wireless systems today support multiple users, single-user results are still of much interest for the insight they provide and their application to channelized systems, where users are allocated orthogonal resources (time, frequency bands, etc.). MIMO channel capacity is also much easier to derive for single users than for multiple users. Indeed, single-user MIMO capacity results are known for many cases, where the corresponding multiuser problems remain unsolved. In particular, very little is known about multiuser capacity without the assumption of perfect channel state information at the transmitter (CSIT) and at the receiver (CSIR). While there remain many open problems in obtaining the single-user capacity under general assumptions of CSI and CDI, for several interesting cases the solution is known. This section will give an overview of known results for single-user MIMO channels with particular focus on special cases of CDI at the transmitter, as well as the receiver. We begin with a description of the channel model and the different CSI and CDI models we consider, along with their motivation.

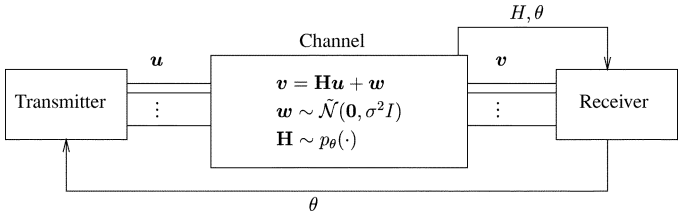

Fig. 1. MIMO channel with perfect CSIR and distribution feedback.

A. Channel Model

Consider a transmitter with M transmit antennas and a receiver with N receive antennas. The channel can be represented by the $N \times M$ matrix $\mathbf{H}$. The $N \times 1$ received signal $\mathbf{v}$ is equal to

$$\mathbf{v} = \mathbf{H}\mathbf{u} + \mathbf{w} \tag{1}$$

where $\mathbf{u}$ is the $M \times 1$ transmitted vector and $\mathbf{w}$ is the $N \times 1$ additive white circularly symmetric complex Gaussian noise vector, normalized so that its covariance matrix is the identity matrix. The normalization of any nonsingular noise covariance matrix $\mathbf{K_w}$ to fit the above model is as straightforward as multiplying the received vector $\mathbf{u}$ with $\mathbf{K_w}^{-1/2}$ to yield the effective channel $\mathbf{K_w}^{-1/2}\mathbf{H}$ and a white noise vector.

The CSI is the channel matrix $\mathbf{H}$. Thus, with perfect CSIT or CSIR, the channel matrix $\mathbf{H}$ is assumed to be known perfectly and instantaneously at the transmitter or receiver, respectively. When the transmitter or receiver knows the channel state perfectly, we also assume that it knows the distribution of this state perfectly, since the distribution can be obtained from the state observations.

1) Perfect CSIR and CDIT: The perfect CSIR and CDIT model is motivated by the scenario where the channel state can be accurately tracked at the receiver and the statistical channel model at the transmitter is based on CDI fed back from the receiver. This distribution model is typically based on receiver estimates of the channel state and the uncertainty associated with these estimates. Fig. 1 illustrates the underlying communication model in this scenario, where $\tilde{\mathcal{N}}$ denotes the complex Gaussian distribution.

The salient features of the model are as follows.

- Conditioned on the parameter θ that defines the channel distribution, the channel realizations H at different time instants are independent identically distributed (i.i.d.).

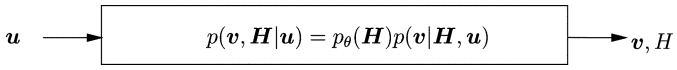

Fig. 2. MIMO channel with perfect CSIR and CDIT (θ fixed).

- In a wireless system the channel statistics change over time due to mobility of the transmitter, receiver, and the scattering environment. Thus, θ is time-varying.
- The statistical model depends on the time scale of interest. For example, in the short term, the channel coefficients may have a nonzero mean and one set of correlations reflecting the geometry of the particular propagation environment. However, over a long term the channel coefficients may be described as zero-mean and uncorrelated due to the averaging over several propagation environments. For this reason, uncorrelated, zero-mean channel coefficients is a common assumption for the channel distribution in the absence of distribution feedback or when it is not possible to adapt to the short-term channel statistics. However, if the transmitter receives frequent updates of θ and it can adapt to these time-varying short-term channel statistics then capacity is increased relative to the transmission strategy associated with just the long-term channel statistics. In other words, adapting the transmission strategy to the short-term channel statistics increases capacity. In the literature adaptation to the short-term channel statistics (the feedback model of Fig. 1) is referred to by many names including mean and covariance feedback, imperfect feedback and partial CSI [38], [40], [42], [45], [46], [56], [66], [76].
- The feedback channel is assumed to be free from noise. This makes the CDIT a deterministic function of the CDIR and allows optimal codes to be constructed directly over the input alphabet [8].
- For each realization of θ the conditional average transmit power is constrained as $\mathbb{E}[||\boldsymbol{u}||^2|\Theta = \theta] \leq P$.
- The ergodic capacity C of the system in Fig. 1 is the capacity $C(\theta)$ averaged over the different θ realizations

$$C = \mathbb{E}_\theta[C(\theta)]$$

where $C(\theta)$ is the ergodic capacity of the channel shown in Fig. 2. This figure represents a MIMO channel with perfect CSI at the receiver and only CDI about the *constant* distribution θ at the transmitter. Channel capacity calculations generally implicitly assume CDI at both the transmitter and receiver except for special channel classes, such as the compound channel or arbitrarily varying channel. This implicit knowledge of θ is justified by the fact that the channel coefficients are typically modeled based on their long-term average distribution. Alternatively, θ can be obtained by the feedback model of Fig. 1. Thus, motivated by the distribution feedback model of Fig. 1, we will provide capacity results for the system model of Fig. 2 under different distribution (θ) models. For clarity, we explicitly state when CDI is available at either the transmitter or receiver, to contrast with the case where CSI is also available.

Computation of $C(\theta)$ for general $p_\theta(\cdot)$ is a hard problem. Almost all research in this area has focused on three special cases for this distribution: zero-mean spatially white channels, spatially white channels with nonzero mean, and zero-mean channels with nonwhite channel covariance. In all three cases, the channel coefficients are modeled as complex jointly Gaussian random variables. Under the zero-mean spatially white (ZMSW) model, the channel mean is zero and the channel covariance is modeled as white, i.e., the channel elements are assumed to be i.i.d. random variables. This model typically captures the long-term average distribution of the channel coefficients averaged over multiple propagation environments. Under the channel mean information (CMI) model, the mean of the channel distribution is nonzero while the covariance is modeled as white with a constant scale factor. This model is motivated by a system where the channel state is measured imperfectly at the transmitter, so the CMI reflects this measurement and the constant factor reflects the estimation error. Under the channel covariance information (CCI) model, the channel is assumed to be varying too rapidly to track its mean, so the mean is set to zero and the information regarding the relative geometry of the propagation paths is captured by a nonwhite covariance matrix. Based on the underlying system model shown in Fig. 1, in the literature the CMI model is also called mean feedback and the CCI model is also called covariance feedback. Mathematically, the three distribution models for $\mathbf{H}$ can be described as follows:

Zero-Mean Spatially White (ZMSW): $\mathbb{E}[\mathbf{H}] = \mathbf{0}$, $\mathbf{H} = \mathbf{H}^{\mathbf{w}}$;
Channel Mean Information (CMI): $\mathbb{E}[\mathbf{H}] = \overline{\mathbf{H}}$, $\mathbf{H} = \overline{\mathbf{H}} + \sqrt{\alpha}\mathbf{H}^{\mathbf{w}}$;
Channel Covariance Information (CCI): $\mathbb{E}[\mathbf{H}] = \overline{\mathbf{0}}$, $\mathbf{H} = (\mathbf{R}^r)^{1/2}\mathbf{H}^{\mathbf{w}}(\mathbf{R}^t)^{1/2}$.

Here, $\mathbf{H}^{\mathbf{w}}$ is an $N \times M$ matrix of i.i.d. zero mean, unit variance complex circularly symmetric Gaussian random variables. The channel mean $\overline{\mathbf{H}}$ and α are constants that may be interpreted as the channel estimate based on the feedback and the variance of the estimation error, respectively. $\mathbf{R}^r$ and $\mathbf{R}^t$ are called the receive and transmit fade covariance matrices. Although not completely general, this simple correlation model has been validated through recent field measurements as a sufficiently accurate representation of the fade correlations seen in actual cellular systems [13]. Under CMI the channel mean $\overline{\mathbf{H}}$ and the variance of the estimation error α are assumed known and under CCI the transmit and receive covariance matrices $\mathbf{R}^r$ and $\mathbf{R}^t$ are assumed known.

2) CDIT and CDIR: In highly mobile channels, the assumption of perfect CSI at the receiver can be unrealistic. Thus, we now consider a model where both transmitter and receiver only have information about the channel distribution. Even for a rapidly fluctuating channel where reliable channel estimation is not possible, it might be possible for the receiver to track the short-term distribution of the channel fades, as the channel distribution changes much more slowly than the channel itself. The estimated distribution can be made available to the transmitter through a feedback channel. Fig. 3 illustrates the underlying communication model.

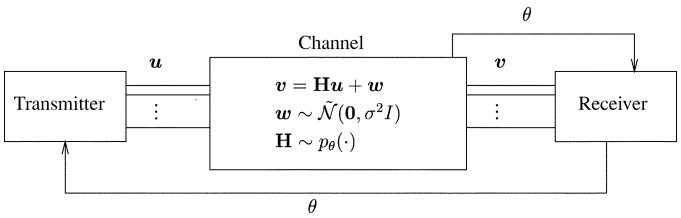

Fig. 3. MIMO channel with CDIR and distribution feedback.

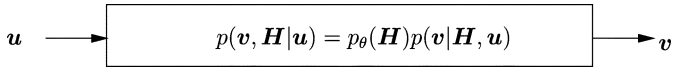

Fig. 4. MIMO channel with CDIT and CDIR (θ fixed).

Note that the estimation of the channel statistics at the receiver is captured in the model as a genie that provides the receiver with the correct channel distribution. The feedback channel represents the same information being made available to the transmitter simultaneously. This model is slightly optimistic because in practice the receiver estimates θ only from the received signal $\boldsymbol{v}$ and therefore will not have a perfect estimate.

As in the previous section, the ergodic capacity turns out to be the expected value (expectation over θ) of the ergodic capacity $C(\theta)$, where $C(\theta)$ is the ergodic capacity of the channel in Fig. 4. In this figure, θ is constant and known at both the transmitter and receiver (CDIT and CDIR). As in the previous section, the computation of $C(\theta)$ is difficult for general θ, so we restrict ourselves to the same three channel distribution models described in the previous subsection: the ZMSW, CMI, and CCI models.

Next, we summarize the single-user MIMO capacity results under various assumptions on CSI and CDI.

B. Constant MIMO Channel Capacity

When the channel is constant and known perfectly at the transmitter and the receiver, the capacity is

$$C = \max_{\mathbf{Q}:\mathrm{Tr}(\mathbf{Q})=P} \log\left|\mathbf{I}_N + \mathbf{HQH}^\dagger\right| \tag{2}$$

where $\mathbf{Q}$ is the input covariance matrix. Telatar [69] showed that the MIMO channel can be converted to parallel, noninterfering single-input single-output (SISO) channels through a singular value decomposition (SVD) of the channel matrix. The SVD yields $\min(M,N)$ parallel channels with gains corresponding to the singular values σ_i^2 of $\mathbf{H}$. Waterfilling the transmit power over these parallel channels leads to the power allocation

$$P_i = \left(\mu - \frac{1}{\sigma_i^2}\right)^+,\ 1 \le i \le \min(M,N) \tag{3}$$

where μ is the waterfill level, P_i is the power in the ith eigenmode of the channel and x^+ is defined as $\max(x,0)$. The channel capacity is shown to be

$$C = \sum_{i}^{\min(M,N)} (\log(\mu\sigma_i^2))^+. \tag{4}$$

Although the constant channel model is relatively easy to analyze, wireless channels in practice are not fixed or constant. Instead, due to the changing propagation environment wireless channels vary over time, assuming values over a continuum. The capacity of fading channels is investigated next.

C. Fading MIMO Channel Capacity

With slow fading, the channel may remain approximately constant long enough to allow reliable estimation of the channel state at the receiver (perfect CSIR) and timely feedback of this state information to the transmitter (perfect CSIT). However, in systems with moderate to high user mobility, the system designer is inevitably faced with channels that change rapidly. Fading models where only the channel distribution is available to the receiver (CDIR) and/or transmitter (CDIT) are more applicable to such systems. Capacity results under various assumptions regarding CSI and CDI are summarized in this section.

1) Capacity With Perfect CSIT and Perfect CSIR: Perfect CSIT and perfect CSIR model a fading channel that changes slow enough to be reliably measured by the receiver and fed back to the transmitter without significant delay. The ergodic capacity of a flat-fading channel with perfect CSIT and CSIR is simply the average of the capacities achieved with each channel realization. The capacity for each channel realization is given by the constant channel capacity expression in the previous section. Thus, the fading MIMO channel capacity assuming perfect channel knowledge at both transmitter and receiver is

$$C = \mathbb{E}_{\mathbf{H}}\left[\max_{\mathbf{Q}:\mathrm{Tr}(\mathbf{Q})=P} \log\left|\mathbf{I}_N + \mathbf{HQH}^\dagger\right|\right]. \tag{5}$$

2) Capacity With Perfect CSIR and CDIT: ZMSW Model: Seminal work by Foschini and Gans [22] and Telatar [69] addressed the case of perfect CSIR and a ZMSW channel distribution at the transmitter. Recall that in this case, the channel matrix $\mathbf{H}$ is assumed to have i.i.d. complex Gaussian entries (i.e., $\mathbf{H} \sim \mathbf{H^w}$). As described in the introduction, the two relevant capacity definitions in this case are capacity versus outage (capacity CDF) and ergodic capacity. For any given input covariance matrix the input distribution that achieves the ergodic capacity is shown in [22] and [69] to be complex vector Gaussian, mainly because the vector Gaussian distribution maximizes the entropy for a given covariance matrix. This leads to the transmitter optimization problem—i.e., finding the optimum input covariance matrix to maximize ergodic capacity subject to a transmit power (trace of the input covariance matrix) constraint. Mathematically, the problem is to characterize the optimum $\mathbf{Q}$ to maximize

$$C = \max_{\mathbf{Q}:\mathrm{Tr}(\mathbf{Q})=P} C(\mathbf{Q}) \tag{6}$$

where

$$C(\mathbf{Q}) \triangleq \mathbb{E}_{\mathbf{H}}\left[\log\left|\mathbf{I}_N + \mathbf{HQH}^\dagger\right|\right] \tag{7}$$

is the mutual information with the input covariance matrix $\mathbb{E}[\mathbf{uu}^\dagger] = \mathbf{Q}$ and the expectation is with respect to the channel matrix $\mathbf{H}$. The mutual information $C(\mathbf{Q})$ is achieved by transmitting independent complex circular Gaussian symbols along the eigenvectors of $\mathbf{Q}$. The powers allocated to each eigenvector are given by the eigenvalues of $\mathbf{Q}$.

It is shown in [22] and [69] that the optimum input covariance matrix that maximizes ergodic capacity is the scaled identity matrix, i.e., the transmit power is divided equally among all the transmit antennas. Thus, the ergodic capacity is given by

$$C = \mathbb{E}_{\mathbf{H}} \left[\log \left| \mathbf{I}_N + \frac{P}{M} \mathbf{H}\mathbf{H}^{\dagger} \right| \right]. \tag{8}$$

An integral form of this expectation involving Laguerre polynomials is derived in [69]. If M and N simultaneously become large, capacity is seen to grow *linearly* with $\min(M, N)$. Expressions for the growth rate constant can be found in [32] and [69].

Telatar [69] conjectures that the optimal input covariance matrix that maximizes capacity versus outage is a diagonal matrix with the power equally distributed among a *subset* of the transmit antennas. The principal observation is that as the capacity CDF becomes steeper, capacity versus outage increases for low outage probabilities and decreases for high outage probabilities. This is reflected in the fact that the higher the outage probability, the smaller the number of transmit antennas that should be used. As the transmit power is shared equally between more antennas the expectation of C increases (so the ergodic capacity increases) but the tails of its distribution decay faster. While this improves capacity versus outage for low outage probabilities, the capacity versus outage for high outages is decreased. Usually, we are interested in low outage probabilities[2] and, therefore, the usual intuition for outage capacity is that it increases as the diversity order of the channel increases, i.e., as the capacity CDF becomes steeper. Foschini and Gans [22] also propose a layered architecture to achieve these capacities with scalar codes. This architecture, called Bell Labs Layered Space–Time (BLAST), shows enormous capacity gains over single antenna systems. For example, at 1% outage, 12 dB signal-to-noise ratio (SNR) and with 12 antennas, the spectral efficiency is shown to be 32 b/s/Hz as opposed to the spectral efficiencies of around 1 b/s/Hz achieved in present day single antenna systems. While the channel models in [22] and [69] assume uncorrelated and frequency flat fading, practical channels exhibit both correlated fading, as well as frequency selectivity. The need to estimate the capacity gains of BLAST for practical systems in the presence of channel fade correlations and frequency selective fading sparked off measurement campaigns reported in [24] and [55]. The measured capacities are found to be about 30% smaller than would be anticipated from an idealized model. However, the capacity gains over single antenna systems are still overwhelming.

3) Capacity With Perfect CSIR and CDIT: CMI and CCI Models: Recent results indicate that for MIMO channels the capacity improvement resulting from some knowledge of the short-term channel statistics at the transmitter can be substantial. These results have ignited much interest in the capacity of MIMO channels with perfect CSIR and CDIT under general distribution models. In this section, we focus on the cases of CMI and CCI channel distributions, corresponding to distribution feedback of the channel mean or covariance matrix. Key results on the capacity of such channels have been recently obtained by several authors including Madhow and Visotsky [76], Trott and Narula [58], [57], Jafar and Goldsmith [42], [40], [38], Jorsweick and Boche [45], [46], and Simon and Moustakas [56], [66].

Mathematically the problem is defined by (6) and (7), with the distribution on $\mathbf{H}$ determined by the CMI or CCI. The optimum input covariance matrix in general can be a full rank matrix which implies either vector coding across the antenna array or transmission of several scalar codes in parallel with successive interference cancellation at the receiver. Limiting the rank of the input covariance matrix to unity, called *beamforming*, essentially leads to a scalar coded system which has a significantly lower complexity for typical array sizes.

The complexity versus capacity tradeoff is an interesting aspect of capacity results under CDIT. The ability to use scalar codes to achieve capacity under CDIT for different channel distribution models, also called optimality of beamforming, captures this tradeoff and has been the topic of much research in itself. Note that vector coding refers to fully unconstrained signaling schemes for the memoryless MIMO Gaussian channel. Every symbol period, a channel use corresponds to the transmission of a vector symbol comprised of the inputs to each transmit antenna. Ideally, while decoding vector codewords the receiver needs to take into account the dependencies in both space and time dimensions and therefore the complexity of vector decoding grows exponentially in the number of transmit antennas. A lower complexity implementation of the vector coding strategy is also possible in the form of several scalar codewords being transmitted in parallel. It is shown in [38] that without loss of capacity, any input covariance matrix, regardless of its rank, can be treated as several scalar codewords encoded independently at the transmitter and decoded successively at the receiver by subtracting out the contribution from previously decoded codewords at each stage. However, well-known problems associated with successive decoding and interference subtraction, e.g., error propagation, render this approach unsuitable for use in practical systems. It is in this context that the question of optimality of beamforming becomes important. Beamforming transforms the MIMO channel into a single-input single-output (SISO) channel. Thus, well established scalar codec technology can be used to approach capacity and since there is only one beam, interference cancellation is not needed. In the summary given below, we include the results on both the transmitter optimization problem, as well as the optimality of beamforming.

Multiple-Input Single-Output (MISO) Channels: We first consider systems that use a single receive antenna and multiple transmit antennas. The channel matrix is rank one. With perfect CSIT and CSIR, for every channel matrix realization it is possible to identify the only nonzero eigenmode of the channel accurately and beamform along that mode. On the other hand, with perfect CSIR and CDIT under the ZMSW model, it was shown by Foschini and Gans [22] and Telatar [69] that the optimal input covariance matrix is a multiple of the identity matrix. Thus, the inability of the transmitter to identify the nonzero channel eigenmode forces a strategy, where the power is equally distributed in all directions.

[2]The capacity for high outage probabilities becomes relevant for schemes that transmit only to the best user. For such schemes, it is shown in [6] that increasing the number of transmit antennas reduces the average sum capacity.

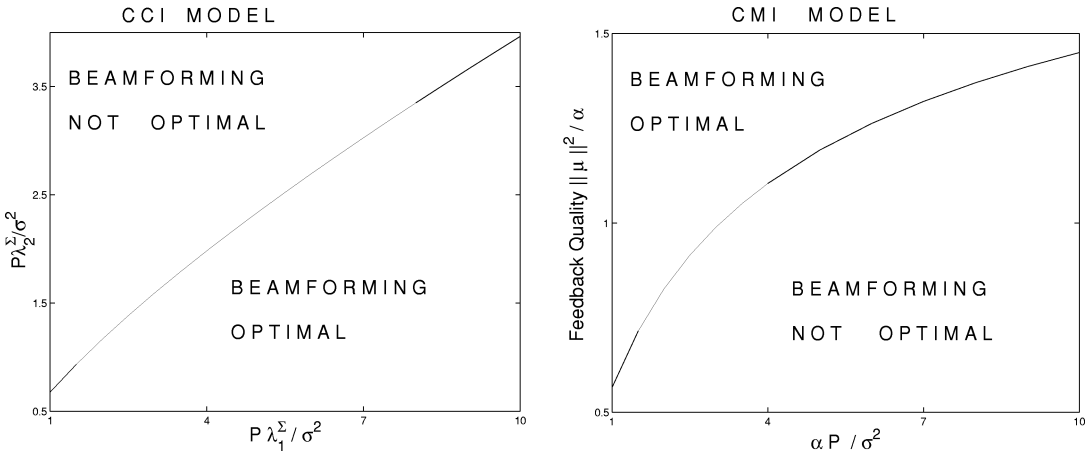

Fig. 5. Plot of necessary and sufficient conditions (9). **<Author: Fig. 5 not cited in text>**

For a system using a single receive antenna and multiple transmit antennas, the transmitter optimization problem under CSIR and CDIT is solved by Visotsky and Madhow in [76] for the distribution models of CMI and CCI. For the CMI model ($\mathbf{H} \sim \tilde{\mathcal{N}}(\overline{\mathbf{H}}, \alpha\mathbf{I})$) the principal eigenvector of the optimal input covariance matrix $\mathbf{Q^o}$ is found to be along the channel mean vector and the eigenvalues corresponding to the remaining eigenvectors are shown to be equal. When beamforming is optimal, all power is allocated to the principal eigenvector. For the CCI model ($\mathbf{H} \sim \tilde{\mathcal{N}}(\mathbf{0}, \mathbf{R^t})$) the eigenvectors of the optimal input covariance matrix $\mathbf{Q^o}$ are shown to be along the eigenvectors of the transmit fade covariance matrix and the eigenvalues are in the same order as the corresponding eigenvalues of the transmit fade covariance matrix. Moreover, Visotsky and Madhow's numerical results indicate that beamforming is close to the optimal strategy when the quality of feedback improves, i.e., when the channel uncertainty α decreases under CMI or when a stronger channel mode can be identified under CCI. We will discuss quality of feedback in more detail below. Under CMI, Narula and Trott [58] point out that there are cases where the capacity is actually achieved via beamforming. While they do not obtain fully general necessary and sufficient conditions for when beamforming is a capacity achieving strategy, they develop partial answers to the problem for two transmit antennas.

A general condition that is both necessary and sufficient for optimality of beamforming is obtained by Jafar and Goldsmith in [40] for both the CMI and CCI models. The result can be stated as follows.

The ergodic capacity can be achieved with a unit rank matrix if and only if the following condition is true:

$$\mathbb{E}_{w_1}\left[\frac{1}{1+P\lambda_1 w_1}\right] \leq \frac{1}{1+P\lambda_2} \tag{9}$$

where for the CCI model

1) $\lambda_1 > \lambda_2$ are the two largest eigenvalues of the channel fade covariance matrix $\mathbf{R^t}$;
2) w_1 is exponential distributed with unit mean, i.e., $w_1 \sim e^{-w1}$;

and for the CMI model

1) $\lambda_1 = \lambda_2 = \alpha$;
2) w_1 has a noncentral chi-squared distribution. More precisely, $w_1 \sim e^{-||\mu||^2/\alpha - w_1} I_0\left(2||\mu||\sqrt{w_1/\alpha}\right)$ where $I_0(\cdot)$ is the zeroth-order modified Bessel function of the first kind.

Further, for the CCI model the expectation can be evaluated to express (9) explicitly in closed form as

$$\frac{1}{P\lambda_1} e^{1/P\lambda_1} \Gamma(0, \frac{1}{P\lambda_1}) \leq \frac{1}{1+P\lambda_2}. \tag{10}$$

The optimality conditions are plotted in Fig. 5. For the CCI model the optimality of beamforming depends on the two largest eigenvalues λ_1, λ_2 of the transmit fade covariance matrix and the transmit power P. Beamforming is found to be optimal when the two largest eigenvalues of the transmit covariance matrix are sufficiently disparate or the transmit power P is sufficiently low. Since beamforming corresponds to using the principal eigenmode alone, this is reminiscent of waterpouring solutions where only the deepest level gets all the water when it is sufficiently deeper than the next deepest level and when the quantity of water is small enough. For the CMI model the optimality of beamforming is found to depend on transmit power P and the *quality of feedback* associated with the mean information, which is defined mathematically as the ratio $||\overline{\mathbf{H}}||^2/\alpha$ of the norm squared of the channel mean vector $\overline{\mathbf{H}}$ and the channel uncertainty α. As the transmit power P is decreased or the quality of feedback improves beamforming becomes optimal. As mentioned earlier, for perfect CSIT (uncertainty $\alpha \to 0$ so quality of feedback $\to \infty$) the optimal input strategy is beamforming, while in the absence of mean feedback (quality of feedback $\to 0$ so the CMI model becomes the ZMSW model), as shown by Telatar [69], the optimal input covariance has full rank, i.e., beamforming is necessarily suboptimal. Note that [40], [57], [58], and [76] assume a single receive antenna. Next, we summarize the analogous capacity results for MIMO channels.

MIMO Channels: With multiple transmit and receive antennas, capacity with CSIR and CDIT under the CCI model with spatially white fading at the receiver ($\mathbf{R^r} = I$) is obtained by

Jafar and Goldsmith in [42]. Like the single receive antenna case the capacity achieving input covariance matrix is found to have the eigenvectors of the transmit fade covariance matrix and the eigenvalues are in the same order as the corresponding eigenvalues of the transmit fade covariance matrix. Jafar and Goldsmith also presented in closed form a mathematical condition that is both necessary and sufficient for optimality of beamforming in this case. The same necessary and sufficient condition is also derived independently by Jorsweick and Boche in [45] and Simon and Moustakas in [66]. In [46], Jorsweick and Boche extend these results to incorporate fade correlations at the receiver as well. Their results show that while the receive fade correlation matrix does not affect the eigenvectors of the optimal input covariance matrix, it does affect the eigenvalues. The general condition for optimality of beamforming found by Jorsweick and Boche depends upon the two largest eigenvalues of the transmit covariance matrix and all the eigenvalues of the receive covariance matrix.

Capacity under the CMI model with multiple transmit and receive antennas is solved by Jafar and Goldsmith in [38] when the channel mean has rank one and is extended to general channel means by Moustakas and Simon in [67]. Similar to the MISO case, the principal eigenvector of the optimal input covariance matrix and of the channel mean are the same and the eigenvalues of the remaining eigenvectors are equal. For the case where the channel mean has unit rank, a necessary and sufficient condition for optimality of beamforming is also determined in [38].

These results summarize our discussion of channel capacity with CDIT and perfect CSIR under different channel distribution models. From these results we notice that the benefits of adapting to distribution information regarding CMI or CCI fed back from the receiver to the transmitter are twofold. Not only does the capacity increase with more information about the channel distribution, but this feedback also allows the transmitter to identify the stronger channel modes and achieve this higher capacity with simple scalar codewords.

We conclude this section with a discussion on the growth of capacity with number of antennas. With perfect CSIR and CDIT under the ZMSW channel distribution, it was shown by Foschini and Gans [22] and by Telatar [69] that the channel capacity grows linearly with $\min(M, N)$. This linear increase occurs whether the transmitter knows the channel perfectly (perfect CSIT) or only knows its distribution (CDIT). The proportionality constant of this linear increase, called the rate of growth, has also been characterized in [15], [31], [68], [69]. Chuah *et al.* [15] show that with perfect CSIR and CSIT, the rate of growth of capacity with $\min(M, N)$ is reduced by channel fading correlations at high SNR but is increased at low SNR. They also show that the mutual information under CSIR increases linearly with $\min(M, N)$ even when a spatially white transmission strategy is used on a correlated fading channel, although the slope is reduced relative to the uncorrelated fading channel. As we will see in the next section, the assumption of perfect CSIR is crucial for the linear growth behavior of capacity with the number of antennas.

In the next section, we explore the capacity when only CDI is available at the transmitter and the receiver.

4) Capacity With CDIT and CDIR: ZMSW Model: We saw in the last section that with perfect CSIR, channel capacity grows linearly with the minimum of the number of transmit and receive antennas. However, reliable channel estimation may not be possible for a mobile receiver that experiences rapid fluctuations of the channel coefficients. Since user mobility is the principal driving force for wireless communication systems, the capacity behavior with CDIT and CDIR under the ZMSW distribution model (i.e., $\mathbf{H}$ is distributed as $\mathbf{H}^w$ with no knowledge of $\mathbf{H}$ at either the receiver or transmitter) is of particular interest. In this section, we summarize some MIMO capacity results in this area.

One of the first papers to address the MIMO capacity with CDIR and CDIT under the ZMSW model is [53] by Marzetta and Hochwald. They model the channel matrix components as i.i.d. complex Gaussian random variables that remain constant for a coherence interval of T symbol periods after which they change to another independent realization. Capacity is achieved when the $T \times M$ transmitted signal matrix is equal to the product of two statistically independent matrices: a $T \times T$ isotropically distributed unitary matrix times a certain $T \times M$ random matrix that is diagonal, real, and nonnegative. This result enables them to determine capacity for many interesting cases. Marzetta and Hochwald show that, for a fixed number of antennas, as the length of the coherence interval increases, the capacity approaches the capacity obtained as if the receiver knew the propagation coefficients. However, perhaps the most surprising result in [53] is the following: In contrast to the linear growth of capacity with $\min(M, N)$ under the perfect CSIR assumption, [53] showed that in the absence of CSIT and CSIR, capacity does not increase at all as the number of transmit antennas is increased beyond the length of the coherence interval T. The MIMO capacity for this model was further explored by Zheng and Tse in [89]. They show that at high SNRs capacity is achieved using no more than $M^\star = \min(M, N, \lfloor T/2 \rfloor)$ transmit antennas. In particular, having more transmit antennas than receive antennas does not provide any capacity increase at high SNR. Zheng and Tse also find that for each 3-dB SNR increase, the capacity gain is $M^\star(1 - M^\star/T)$.

Notice that [53], [89] assume block fading models, i.e., the channel fade coefficients are assumed to be constant for a block of T symbol durations. Hochwald and Marzetta extend their results to continuous fading in [54] where, within each independent T-symbol block, the fading coefficients have an arbitrary time correlation. If the correlation vanishes beyond some lag τ, called the *correlation time* of the fading, then it is shown in [54] that increasing the number of transmit antennas beyond $\min(\tau, T)$ antennas does not increase capacity. Lapidoth and Moser [47] explored the channel capacity of this CDIT/CDIR model for the ZMSW distribution at high SNR without the block fading assumption. In contrast to the results of Zheng and Tse for block fading, Lapidoth and Moser show that without the block fading assumption, the channel capacity grows only double logarithmically in SNR. This result is shown to hold under very general conditions, even allowing for memory and partial receiver side information.

5) Capacity With CDIR and CDIR: CCI Model: The results in [53] and [89] seem to leave little hope of achieving the high

capacity gains predicted for MIMO systems when the channel cannot be accurately estimated at the receiver and the channel distribution follows the ZMSW model. However, before resigning ourselves to these less-than-optimistic results we note that these results assume a somewhat pessimistic model for the channel distribution. That is because most channels when averaged over a relatively small area have either a nonzero mean or a nonwhite covariance. Thus, if these distribution parameters can be tracked, the channel distribution corresponds to either the CMI or CCI model.

Recent work by Jafar and Goldsmith [37] addresses the MIMO channel capacity with CDIT and CDIR under the CCI distribution model. The channel matrix components are modeled as spatially correlated complex Gaussian random variables that remain constant for a coherence interval of T symbol periods after which they change to another independent realization based on the spatial correlation model. The channel correlations are assumed to be known at the transmitter and receiver. As in the case of spatially white fading (ZMSW model), Jafar and Goldsmith show that with the CCI model the capacity is achieved when the $T \times M$ transmitted signal matrix is equal to the product of a $T \times T$ isotropically distributed unitary matrix, a statistically independent $T \times M$ random matrix that is diagonal, real and nonnegative and the matrix of the eigenvectors of the transmit fade covariance matrix $\mathbf{R}^t$. It is shown in [37] that the channel capacity is independent of the smallest $(M - T)^+$ eigenvalues of the transmit fade covariance matrix, as well as the eigenvectors of the transmit and receive fade covariance matrices $\mathbf{R}^t$ and $\mathbf{R}^r$. Also, in contrast to the results for the spatially white fading model where adding more transmit antennas beyond the coherence interval length ($M > T$) does not increase capacity, [37] shows that additional transmit antennas always increase capacity as long as their channel fading coefficients are spatially correlated. Thus, in contrast to the results in favor of independent fades with perfect CSIR, these results indicate that with CCI at the transmitter and the receiver, transmit fade correlations can be beneficial, making the case for minimizing the spacing between transmit antennas when dealing with highly mobile, fast fading channels that cannot be accurately measured. Mathematically, [37] proves that for fast fading channels ($T = 1$), capacity is a Schur-concave function of the vector of eigenvalues of the transmit fade correlation matrix. The maximum possible capacity gain due to transmitter fade correlations is shown to be $10 \log_{10} M$ db.

6) Frequency Selective Fading Channels: While flat fading is a realistic assumption for narrowband systems where the signal bandwidth is smaller than the channel coherence bandwidth, broadband communications involve channels that experience frequency selective fading. Research on the capacity of MIMO systems with frequency selective fading typically takes the approach of dividing the channel bandwidth into parallel flat fading channels and constructing an overall block diagonal channel matrix with the diagonal blocks given by the channel matrices corresponding to each of these subchannels. Under perfect CSIR and CSIT, the total power constraint then leads to the usual closed-form waterfilling solution. Note that the waterfill is done simultaneously over both space and frequency. Even SISO frequency selective fading channels can be represented by the MIMO system model (1) in this manner [59]. For MIMO systems, the matrix channel model is derived by Bolcskei, Gesbert and Paulraj in [5] based on an analysis of the capacity behavior of OFDM-based MIMO channels in broadband fading environments. Under the assumption of perfect CSIR and CDIT for the ZMSW model, their results show that in the MIMO case, unlike the SISO case, frequency selective fading channels may provide advantages over flat fading channels not only in terms of ergodic capacity but also in terms of capacity versus outage. In other words, MIMO frequency selective fading channels are shown to provide both higher diversity gain and higher multiplexing gain than MIMO flat-fading channels. The measurements in [55] show that frequency selectivity makes the CDF of the capacity steeper and, thus, increases the capacity for a given outage as compared with the flat-frequency case, but the influence on the ergodic capacity is small.

7) Training for Multiple-Antenna Systems: The results summarized in the previous sections indicate that CSI plays a crucial role in the capacity of MIMO systems. In particular, the capacity results in the absence of CSIR are strikingly different and often quite pessimistic compared with those that assume perfect CSIR. To recapitulate, with perfect CSIR and CDIT MIMO channel capacity is known to increase linearly with $\min(M, N)$ when the CDIT assumes the ZMSW or CCI distribution models. However, in fast fading when the channel changes so rapidly that it cannot be estimated reliably at the receiver (CDIR only) the capacity does not increase with the number of transmit antennas at all for $M > T$ where T is the channel decorrelation time. Also at high SNR under the ZMSW distribution model, capacity with perfect CSIR and CDIT increases logarithmically with SNR, while the capacity with CDIR and CDIT increases only double logarithmically with SNR. Thus, CSIR is critical for obtaining the high capacity benefits of multiple-antenna wireless links. CSIR is often obtained by sending known training symbols to the receiver. However, with too little training the channel estimates are poor, whereas with too much training there is no time for data transmission before the channel changes. So the key question to ask is how much training is needed in multiple-antenna wireless links. This question itself is the title of the paper [29] by Hassibi and Hochwald where they compute a lower bound on the capacity of a channel that is learned by training and maximize the bound as a function of the receive SNR, fading coherence time, and number of transmitter antennas. When the training and data powers are allowed to vary, the optimal number of training symbols is shown to be equal to the number of transmit antennas—which is also the smallest training interval length that guarantees meaningful estimates of the channel matrix. When the training and data powers are instead required to be equal, the optimal training duration may be longer than the number of antennas. Hassibi and Hochwald also show that training-based schemes can be optimal at high SNR, but are suboptimal at low SNR.

D. Open Problems in Single-User MIMO

The results summarized in this section form the basis of our understanding of channel capacity under different CSI and CDI

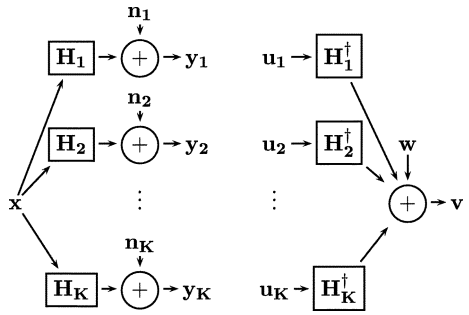

Fig. 6. System models of the (left) MIMO BC and the (right) MIMO MAC channels.

assumptions. These results serve as useful indicators for the benefits of incorporating training and feedback schemes in a MIMO wireless link to obtain CSIR/CDIT and CSIT/CDIT, respectively. However, our knowledge of MIMO capacity with CDI only is still far from complete, even for single-user systems. We conclude this section by pointing out some of the many open problems.

1) Combined CCI and CMI: Capacity under CDIT and perfect CSIR is unsolved under a combined CCI and CMI distribution model even with a single receive antenna.
2) CCI: With perfect CSIR and CDIT capacity is not known under the CCI model for completely general correlations.
3) CDIR: Almost all cases with only CDIR are open problems.
4) Outage capacity: Most results for CDI only at either the transmitter or receiver are for ergodic capacity. Capacity versus outage has proven to be less analytically tractable than ergodic capacity and contains an abundance of open problems.

III. Multiuser MIMO

In this section, we consider the two basic multiuser MIMO channel models: the MIMO MAC and the MIMO BC. Since the capacity region of a general MAC has been known for quite a while, there are many results on the MIMO MAC for both constant channels and fading channels with different CSI and CDI assumptions at the transmitters and receivers. The MIMO BC, however, is a relatively new problem for which capacity results have only recently been found. As a result, the field is much less developed, but we summarize the recent results in the area. Interestingly, the MIMO MAC and MIMO BC have been shown to be duals, as we will discuss in Section III-C2.

A. System Model

To describe the MAC and BC models, we consider a cellular-type system in which the base station has M antennas and each of the K mobiles has N antennas. The downlink of this system is a MIMO BC and the uplink is a MIMO MAC. We will use $\mathbf{H}_i$ to denote the *downlink* channel matrix from the base station to user i. Assuming that the same channel is used on the uplink and downlink, the *uplink* matrix of user i is $\mathbf{H}_i^\dagger$. A picture of the system model is shown in Fig. 6.

In the MAC, let $\mathbf{u_k} \in \mathbb{C}^{N\times 1}$ be the transmitted signal of user (i.e., mobile) k. Let $\mathbf{v} \in \mathbb{C}^{M\times 1}$ denote the received signal and $\mathbf{w} \in \mathbb{C}^{M\times 1}$ the noise vector where $\mathbf{w} \sim \tilde{\mathcal{N}}(0, \mathbf{I})$ is circularly symmetric complex Gaussian with identity covariance. The received signal at the base station is then equal to

$$\begin{aligned}\mathbf{v} =& \mathbf{H_1}^\dagger \mathbf{u_1} + \cdots + \mathbf{H}_K^\dagger \mathbf{u}_K + \mathbf{w} \\ =& \mathbf{H}^\dagger \begin{bmatrix} \mathbf{u_1} \\ \vdots \\ \mathbf{u}_K \end{bmatrix} + \mathbf{w} \quad \text{where} \quad \mathbf{H}^\dagger = \begin{bmatrix} \mathbf{H_1}^\dagger & \cdots & \mathbf{H}_K^\dagger \end{bmatrix}.\end{aligned}$$

In the MAC, each user (i.e., mobile) is subject to an individual power constraint of P_k. The transmit covariance matrix of user k is defined to be $\mathbf{Q}_k \triangleq \mathbb{E}[\mathbf{u_k}\mathbf{u_k}^\dagger]$. The power constraint implies $\mathrm{Tr}(\mathbf{Q}_k) \le P_k$ for $k = 1, \ldots, K$.

In the BC, let $\mathbf{x} \in \mathbb{C}^{M\times 1}$ denote the transmitted vector signal (from the base station) and let $\mathbf{y_k} \in \mathbb{C}^{N\times 1}$ be the received signal at receiver (i.e., mobile) k. The noise at receiver k is represented by $\mathbf{n_k} \in \mathbb{C}^{N\times 1}$ and is assumed to be circularly symmetric complex Gaussian noise ($\mathbf{n_k} \sim N(0, \mathbf{I})$). The received signal of User k is equal to

$$\mathbf{y_k} = \mathbf{H}_k \mathbf{x} + \mathbf{n}_k. \tag{11}$$

The transmit covariance matrix of the input signal is $\boldsymbol{\Sigma}_x \triangleq \mathbb{E}[\mathbf{x}\mathbf{x}^\dagger]$. The base station is subject to an average power constraint P, which implies $\mathrm{Tr}(\boldsymbol{\Sigma}_x) \le P$.

B. MIMO Multiple-Access Channel

In this section, we summarize capacity results on the multiple-antenna MAC. We first analyze the constant channel scenario and then consider the fading channel. Since the capacity region of a general MAC is known, the expressions for the capacity of a constant MAC are quite straightforward. For the fading case, one must consider different assumptions about the CSI and CDI available at the transmitter and receiver. We consider three cases: perfect CSIR and CSIT, perfect CSIR and CDIT, and CDIT and CDIR. As above, under CDI, we consider three different distribution models: the ZMSW, CMI, and CCI models.

1) Constant Channel: The capacity of any MAC can be written as the convex closure of the union of rate regions corresponding to every product input distribution $p(u_1)\cdots p(u_K)$ satisfying the user-by-user power constraints [18]. For the Gaussian MIMO MAC, however, it has been shown that it is sufficient to consider only Gaussian inputs and that the convex hull operation is not needed [11], [86]. For any set of powers $\mathbf{P} = (P_1, \ldots, P_K)$, the capacity of the MIMO MAC is shown in (12), at the bottom of the next page. The ith user transmits a zero-mean Gaussian with spatial covariance matrix $\mathbf{Q}_i$. Each set of covariance matrices $(\mathbf{Q}_1, \ldots, \mathbf{Q}_K)$ corresponds to a K-dimensional polyhedron (i.e.)

$$\left\{(R_1, \ldots, R_K) : \sum_{i\in S} R_i \le \frac{1}{2}\log\left|\mathbf{I} + \sum_{i\in S}\mathbf{H}_i^\dagger \mathbf{Q}_i \mathbf{H}_i\right| \quad \forall S \subseteq \{1, \ldots, K\}\right\}$$

and the capacity region is equal to the union (over all covariance matrices satisfying the trace constraints) of

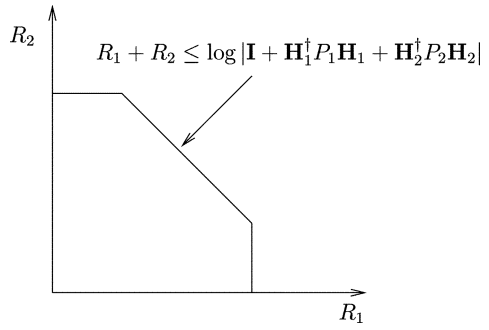

Fig. 7. Capacity region of MIMO MAC for $N = 1$.

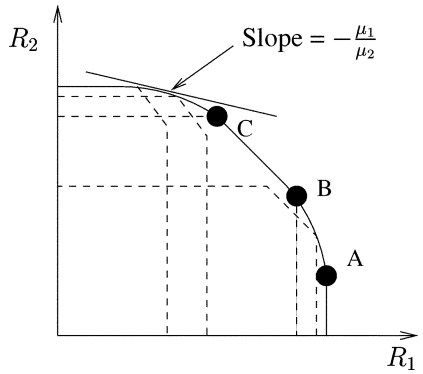

Fig. 8. Capacity region of MIMO MAC for $N > 1$.

all such polyhedrons. The corner points of each polyhedron can be achieved by *successive decoding*, in which users' signals are successively decoded and subtracted out of the received signal. For the two-user case, each set of covariance matrices corresponds to a pentagon, similar in form to the capacity region of the scalar Gaussian MAC. The corner point where $R_1 = \log|\mathbf{I} + \mathbf{H}_1^\dagger \mathbf{Q}_1 \mathbf{H}_1|$ and $R_2 = \log|\mathbf{I} + \mathbf{H}_1^\dagger \mathbf{Q}_1 \mathbf{H}_1 + \mathbf{H}_2^\dagger \mathbf{Q}_2 \mathbf{H}_2| - R_1 = \log|\mathbf{I} + (\mathbf{I} + \mathbf{H}_1^\dagger \mathbf{Q}_1 \mathbf{H}_1)^{-1} \mathbf{H}_2^\dagger \mathbf{Q}_2 \mathbf{H}_2|$ corresponds to decoding user 2 first (i.e., in the presence of interference from user 1) and decoding user 1 last (without interference from user 2). Successive decoding can reduce a complex multiuser detection problem into a series of single-user detection steps [27].

The capacity region of a MIMO MAC for the single transmit antenna case ($N = 1$) is shown in Fig. 7. When $N = 1$, the covariance matrix of each transmitter is a scalar equal to the transmitted power. Clearly, each user should transmit at full power. Thus, the capacity region for a K-user MAC for $N = 1$ is the set of all rate vectors $(R_1, \ldots, R_K)$ satisfying

$$\sum_{i \in S} R_i \le \frac{1}{2} \log \left| \mathbf{I} + \sum_{i \in S} \mathbf{H}_i^\dagger P_i \mathbf{H}_i \right| \quad \forall S \subseteq \{1, \ldots, K\}. \tag{13}$$

For the two-user case, this reduces to the simple pentagon seen in Fig. 7.

When $N > 1$, however, a union must be taken over all covariance matrices. Intuitively, the set of covariance matrices that maximize R_1 are different from the set of covariance matrices that maximize the sum rate. In Fig. 8, a MAC capacity region for $N > 1$ is shown. Notice that the region is equal to the union of pentagons (each pentagon corresponding to a different set of transmit covariance matrices), a few of which are shown with dashed lines in the figure. The boundary of the capacity region is in general curved, except at the sum rate point, where the boundary is a straight line [86]. Each point on the curved portion of the boundary is achieved by a *different* set of covariance matrices. At point A, user 1 is decoded last and achieves his single-user capacity by choosing $\mathbf{Q}_1$ as a water-fill of the channel $\mathbf{H}_1$ (independent of $\mathbf{H}_2$ or $\mathbf{Q}_2$). User 2 is decoded first, in the presence of interference from user 1, so $\mathbf{Q}_2$ is chosen as a waterfill of the channel $\mathbf{H}_2$ and the interference from user 1. The sum-rate corner points B and C are the two corner points of the pentagon corresponding to the sum-rate optimal covariance matrices $\mathbf{Q}_1^{\text{sum}}$ and $\mathbf{Q}_2^{\text{sum}}$. At point B user 1 is decoded last, whereas at point C user 2 is decoded last. Thus, points B and C are achieved using the same covariance matrices but different decoding orders.

Next, we focus on characterizing the optimal covariance matrices $(\mathbf{Q}_1, \ldots, \mathbf{Q}_K)$ that achieve different points on the boundary of the MIMO MAC capacity region. Since the MAC capacity region is convex, it is well known from convex theory that the boundary of the capacity region can be fully characterized by maximizing the function $\mu_1 R_1 + \cdots + \mu_K R_K$ over all rate vectors in the capacity region and for all nonnegative priorities $(\mu_1, \ldots, \mu_K)$ such that $\sum_{i=1}^K \mu_i = 1$. For a fixed set of priorities $(\mu_1, \ldots, \mu_K)$, this is equivalent to finding the point on the capacity region boundary that is tangent to a line whose slope is defined by the priorities. See the tangent line in Fig. 8 for an example. The structure of the MAC capacity region implies that all boundary points of the capacity region are corner points of polyhedrons corresponding to different sets of covariance matrices. Furthermore, the corner point should correspond to successive decoding in order of *increasing* priority, i.e., the user with the highest priority should be decoded last and, therefore, sees no interference [70], [73]. Thus, the problem of finding the boundary point on the capacity region associated with priorities $\mu_1, \ldots, \mu_K$ assumed to be in descending order (users can be arbitrarily re-numbered to satisfy this condition) can be written as

$$\max_{\mathbf{Q}_1, \ldots, \mathbf{Q}_K} \mu_K \log \left| \mathbf{I} + \sum_{l=1}^{K} \mathbf{H}_l^\dagger \mathbf{Q}_l \mathbf{H}_l \right| + \sum_{i=1}^{K-1} (\mu_i - \mu_{i+1}) \log \left| \mathbf{I} + \sum_{l=1}^{i} \mathbf{H}_l^\dagger \mathbf{Q}_l \mathbf{H}_l \right|$$

subject to power constraints on the trace of each of the covariance matrices. Note that the covariances that maximize the func-

$$\mathcal{C}_{MAC}\left(\mathbf{P}; \mathbf{H}^\dagger\right) \triangleq \bigcup_{\{\mathbf{Q}_i \ge 0,\ \mathrm{Tr}(\mathbf{Q}_i) \le P_i \ \forall i\}} \left\{ \begin{array}{l} (R_1, \ldots, R_K): \\ \sum_{i \in S} R_i \le \frac{1}{2} \log \left| \mathbf{I} + \sum_{i \in S} \mathbf{H}_i^\dagger \mathbf{Q}_i \mathbf{H}_i \right| \ \ \forall S \subseteq \{1, \ldots, K\} \end{array} \right\} \tag{12}$$

tion above are the *optimal* covariances. The most interesting and useful feature of the optimization problem above is that the objective function is concave in the covariance matrices. Thus, efficient convex optimization tools exist that solve this problem numerically [7]. A more efficient numerical technique to find the sum-rate maximizing (i.e., $\mu_1 = \cdots = \mu_K$) covariance matrices, called iterative waterfilling, was developed by Yu *et al.* [86]. This technique is based on the Karush Kuhn Tucker (KKT) optimality conditions for the sum-rate maximizing covariance matrices. These conditions indicate that the sum-rate maximizing covariance matrix of any user in the system should be the single-user water-filling covariance matrix of its own channel with noise equal to the actual noise plus the interference from the other $K-1$ transmitters.

2) Fading Channels: As in the single-user case, the capacity of the MIMO MAC where the channel is time-varying depends on the definition of capacity and the availability of CSI and CDI at the transmitters and the receiver. The capacity with perfect CSIT and CSIR is very well studied, as is the capacity with perfect CSIR and CDIT under the ZMSW distribution model. However, little is known about the capacity of the MIMO MAC with CDIT at either the transmitter or receiver under the CMI or CCI distribution models. Some results on the optimum distribution for the single antenna case with CDIT and CDIR under the ZMSW distribution can be found in [62].

With perfect CSIR and CSIT the system can be viewed as a set of *parallel* non interfering MIMO MACs (one for each fading state) sharing a common power constraint. Thus, the ergodic capacity region can be obtained as an average of these parallel MIMO MAC capacity regions [87], where the averaging is done with respect to the channel statistics. The iterative waterfilling algorithm of [86] easily extends to this case, with joint space and time waterfilling.

The capacity region of a MAC with perfect CSIR and CDIT under the ZMSW distribution model was found in [23] and [63]. In this case, Gaussian inputs are optimal and the ergodic capacity region is equal to the time average of the capacity obtained at each fading instant with a constant transmit policy (i.e., a constant covariance matrix for each user). Thus, the ergodic capacity region is given by

$$\bigcup_{\{\mathbf{Q}_i \geq 0,\ \mathrm{Tr}(\mathbf{Q}_i) \leq P_i \ \ \forall i\}} \left\{ (R_1, \ldots, R_K) : \sum_{i \in S} R_i \leq \frac{1}{2} \mathbb{E}_{\mathbf{H}} \left[\log \left| \mathbf{I} + \sum_{i \in S} \mathbf{H}_i^{\dagger} \mathbf{Q}_i \mathbf{H}_i \right| \right] \quad \forall S \subseteq \{1, \ldots, K\} \right\}.$$

If the channel matrices $\mathbf{H}_i$ have i.i.d. complex Gaussian entries and each user has the same power constraint, then the optimal covariances are scaled versions of the identity matrix [69].

There has also been some work on capacity with perfect CSIR and CDIT under the CCI distribution model [41]. In this paper, Jafar and Goldsmith determine the optimal transmit covariance matrices when there is transmit antenna correlation that is known at the transmitters. This topic has yet to be fully investigated.

Asymptotic results on the sum capacity of MIMO MAC channels with the number of receive antennas and the number of transmitters increasing to infinity were obtained by Telatar [69] and by Viswanath *et al.* [80]. MIMO MAC sum capacity with perfect CSIR and CDIT under the ZMSW distribution model (i.e., each transmitter's channel is distributed as $\mathbf{H}^{\mathbf{w}}$) is found to grow *linearly* with $\min(M, NK)$ [69]. Thus, for systems with large numbers of users, increasing the number of receive antennas at the base station (M) while keeping the number of mobile antennas (N) constant can lead to linear growth. Sum capacity with perfect CSIR and CSIT also scales linearly with $\min(M, NK)$, but perfect CSIT is of decreasing value as the number of receive antennas increases [32], [80]. Furthermore, the limiting distribution of the sum capacity with perfect CSIR and CSIT was found to be Gaussian by Hochwald and Vishwanath [32].

C. MIMO Broadcast Channel

In this section, we summarize capacity results on the multiple-antenna BC. When the transmitter has only one antenna, the Gaussian broadcast channel is a degraded broadcast channel (i.e., the users can be absolutely ranked by their channel strength), for which the capacity region is known [18]. However, when the transmitter has more than one antenna, the Gaussian broadcast channel is generally nondegraded.[3] The capacity region of general nondegraded broadcast channels is unknown, but the seminal work of Caire and Shamai [9] and subsequent research on this problem have shed a great deal of light on this channel and the sum capacity of the MIMO BC has been found. In subsequent sections, we focus mainly on the constant channel, but we do briefly discuss the fading channel as well which is still an open problem. Note that the BC transmitter (i.e., the base station) has M antennas and each receiver has N antennas, as described in Section III-A.

1) Dirty Paper Coding (DPC) Achievable Rate Region: An achievable region for the MIMO BC was first obtained for the $N = 1$ case by Caire and Shamai [9] and later extended to the multiple-receive antenna case by Yu and Cioffi [83] using the idea of DPC [17]. The basic premise of DPC is as follows. If the transmitter (but not the receiver) has perfect, noncausal knowledge of additive Gaussian interference in the channel, then the capacity of the channel is the same as if there was no additive interference, or equivalently as if the receiver also had knowledge of the interference. DPC is a technique that allows noncausally known interference to be "presubtracted" at the transmitter, but in such a way that the transmit power is not increased. A more practical (and more general) technique to perform this presubtraction is the cancelling for known interference technique found by Erez *et al.* in [19].

In the MIMO BC, DPC can be applied at the transmitter when choosing codewords for different receivers. The transmitter first picks a codeword (i.e., $\mathbf{x_1}$) for receiver 1. The transmitter then chooses a codeword for receiver 2 (i.e., $\mathbf{x_2}$) with full (noncausal) knowledge of the codeword intended for receiver 1. Therefore,

[3]The multiple-antenna broadcast channel is nondegraded because users receive different strength signals from different transmit antennas. See [18] for a precise definition of degradedness.

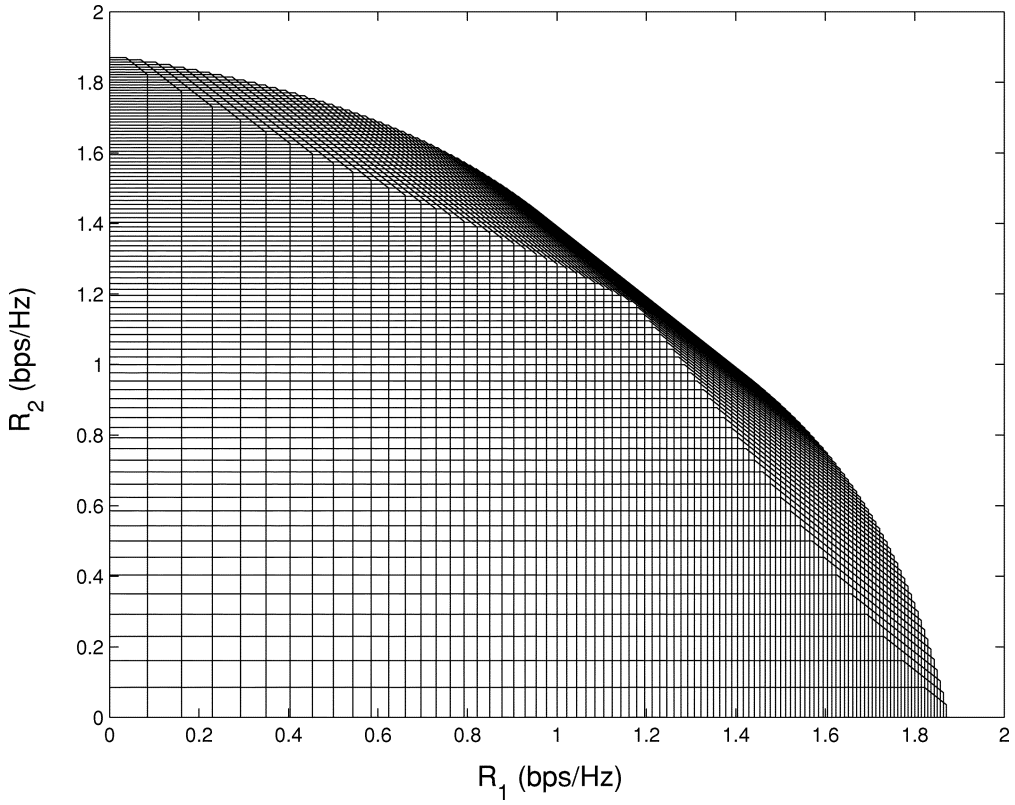

Fig. 9. Dirty paper rate region, $\mathbf{H}_1 = [1\ 0.5]$, $\mathbf{H}_2 = [0.5\ 1]$, $P = 10$.

the codeword of user 1 can be presubtracted such that receiver 2 does not see the codeword intended for receiver 1 as interference. Similarly, the codeword for receiver 3 is chosen such that receiver 3 does not see the signals intended for receivers 1 and 2 (i.e $\mathbf{x_1} + \mathbf{x_2}$) as interference. This process continues for all K receivers. If user $\pi(1)$ is encoded first, followed by user $\pi(2)$, etc., the following is an achievable rate vector:

$$R_{\pi(i)} = \frac{1}{2}\log\frac{\left|\mathbf{I} + \mathbf{H}_{\pi(i)}\left(\sum_{j\geq i}\boldsymbol{\Sigma}_{\pi(j)}\right)\mathbf{H}_{\pi(i)}^{\dagger}\right|}{\left|\mathbf{I} + \mathbf{H}_{\pi(i)}\left(\sum_{j>i}\boldsymbol{\Sigma}_{\pi(j)}\right)\mathbf{H}_{\pi(i)}^{\dagger}\right|} \quad i=1,\ldots,K. \tag{14}$$

The dirty paper region $\mathcal{C}_{\mathrm{DPC}}(P,\mathbf{H})$ is defined as the convex hull of the union of all such rates vectors over all positive semi-definite covariance matrices $\boldsymbol{\Sigma}_1,\ldots,\boldsymbol{\Sigma}_K$ such that $\mathrm{Tr}(\boldsymbol{\Sigma}_1 + \cdots \boldsymbol{\Sigma}_K) = \mathrm{Tr}(\boldsymbol{\Sigma}_x) \leq P$ and over all permutations $(\pi(1),\ldots,\pi(K))$:

$$\mathcal{C}_{\mathrm{DPC}}(P,\mathbf{H}) \triangleq Co\left(\bigcup_{\pi,\boldsymbol{\Sigma}_i}\mathbf{R}(\pi,\boldsymbol{\Sigma}_i)\right) \tag{15}$$

where $\mathbf{R}(\pi,\Sigma_i)$ is given by (14). The transmitted signal is $\mathbf{x} = \mathbf{x_1} + \cdots + \mathbf{x_K}$ and the input covariance matrices are of the form $\boldsymbol{\Sigma}_i = \mathbb{E}[\mathbf{x_i}\mathbf{x_i}^{\dagger}]$. From the dirty paper result we find that $\mathbf{x}_1,\ldots,\mathbf{x}_K$ are uncorrelated, which implies $\boldsymbol{\Sigma}_x = \boldsymbol{\Sigma}_1 + \cdots + \boldsymbol{\Sigma}_K$.

One important feature to notice about the dirty paper rate equations in (14) is that the rate equations are neither a concave nor convex function of the covariance matrices. This makes numerically finding the dirty paper region very difficult, because generally a brute force search over the entire space of covariance matrices that meet the power constraint must be conducted. The dirty paper rate region for a two-user channel with $M = 2$ and $N = 1$ is shown in Fig. 9.

Note that DPC and successive decoding (i.e., interference cancellation by the receiver instead of the transmitter) are completely equivalent capacity-wise for scalar channels, but this equivalence does not hold for MIMO channels. It has been shown [36] that the achievable region with successive decoding is contained within the DPC region.

2) MAC-BC Duality: In [74], Vishwanath, Jindal, and Goldsmith showed that the dirty paper rate region of the multiantenna BC with power constraint P is equal to the union of capacity regions of the dual MAC, where the union is taken over all individual power constraints that sum to P

$$\mathcal{C}_{\mathrm{DPC}}(P,\mathbf{H}) = \bigcup_{\mathbf{P}:\sum_{i=1}^{K}P_i=P}\mathcal{C}_{\mathrm{MAC}}\left(P_1,\ldots,P_K,\mathbf{H}^{\dagger}\right). \tag{16}$$

This is the multiple-antenna extension of the previously established duality between the scalar Gaussian broadcast and multiple-access channels [44]. In addition to the relationship between the two rate regions, for any set of covariance matrices in the MAC/BC (and the corresponding rate vector), [74] provides an explicit set of transformations to find covariance matrices in

the BC/MAC that achieve the same rates. The union of MAC capacity regions in (16) is easily seen to be the same expression as in (12) but with the constraint $\sum_{i=1}^{K} \mathrm{Tr}(\mathbf{Q}_i) \leq P$ instead of $\mathrm{Tr}(\mathbf{Q}_i) \leq P_i \ \ \forall i$ (i.e., a sum constraint instead of individual constraints).

The MAC-BC duality is very useful from a numerical standpoint because the dirty paper region leads to nonconcave rate functions of the covariances, whereas the rates in the dual MAC are *concave* functions of the covariance matrices. Thus, the optimal MAC covariances can be found using standard convex optimization techniques and then transformed to the corresponding optimal BC covariances using the MAC-BC transformations given in [74]. A specialized algorithm to find the optimal MAC covariances can be found in [35]. An algorithm based on the iterative waterfilling algorithm [86] that finds the sum rate optimal covariances is given in [43].

The dirty paper rate region is shown in Fig. 9 for a channel with two-users, $M = 2$ and $N = 1$. Notice that the dirty paper rate region shown in Fig. 9 is actually a union of MAC regions, where each MAC region corresponds to a different set of individual power constraints. Since $N = 1$, each of the MAC regions is a pentagon, as discussed in Section III-B1. Similar to the MAC capacity region, the boundary of the DPC region is curved, except at the sum-rate maximizing portion of the boundary. For the $N = 1$ case, duality also indicates that rank-one covariance matrices (i.e., beamforming) are optimal for DPC. This fact is not obvious from the dirty paper rate equations, but follows from the transformations of [74] which find BC covariances that achieve the same rates as a set of MAC covariance matrices (which are scalars in the $N = 1$ case).

Duality also allows the MIMO MAC capacity region to be expressed as an intersection of the dual dirty paper BC rate regions [74, Corollary 1]

$$\mathcal{C}_{\mathrm{MAC}}\left(P_1, \ldots, P_K, \mathbf{H}^{\dagger}\right) = \bigcap_{\boldsymbol{\alpha}>0} \mathcal{C}_{\mathrm{DPC}}\left(\sum_{i=1}^{K} \frac{P_i}{\alpha_i}; \left[\sqrt{\alpha_1}\mathbf{H}_1^T \cdots \sqrt{\alpha_K}\mathbf{H}_K^T\right]^T\right). \tag{17}$$

3) Optimality of DPC: DPC was first shown to achieve the *sum-rate* capacity of the MIMO BC for the two-user, $M = 2$, $N = 1$ channel by Caire and Shamai [9]. This was shown by proving that the Sato upper bound [61] on the broadcast channel sum-rate capacity is achievable using DPC. The sum-rate optimality of DPC was extended to the multiuser channel with $N = 1$ by Viswanath and Tse [79] and to the more general $N > 1$ case by Vishwanath *et al.* [74] and Yu and Cioffi [84].

It has also recently been conjectured that the DPC rate region is the actual capacity region of the multiple-antenna broadcast channel. Significant progress toward proving this conjecture is made in [75] and [77].

4) Fading Channels: Most of the capacity problems for fading MIMO BCs are still open, with the exception of sum-rate capacity with perfect CSIR and CSIT. In this case, as for the MIMO MAC, the MIMO BC can be split into parallel channels with an overall power constraint (see Li and Goldsmith [48] for a treatment of the scalar case).

Asymptotic results for the sum-rate capacity of the MIMO BC for $N = 1$ under the ZSMW model can be obtained by combining the asymptotic results for the sum-rate capacity of the MIMO MAC with duality [32]. Thus, the role of transmitter side information reduces with the growth in the number of transmit antennas and, hence, the sum capacity of the MIMO BC with K users and M transmit antennas tends to the sum capacity of a single-user system with only receiver CSI and M receive antennas and K transmit antennas, which is given by $\mathbb{E}_{\mathbf{H}} \log|\mathbf{I} + P/K\mathbf{H}\mathbf{H}^{\dagger}|$. Thus, the asymptotic growth under CSIR and CSIT or CDIT under the ZMSW model is linear as $\mathfrak{C} \min(M, K)$ and the growth rate constant $\mathfrak{C}$ can be found in [32]. As seen for the MIMO MAC, for systems with large numbers of users, increasing the number of transmit antennas at the base station (M) while keeping the number of mobile antennas (N) constant can lead to linear growth.

D. Open Problems in Multiuser MIMO

Multiuser MIMO has been the primary focus of research in recent years, mainly due to the large number of open problems in this area. Some of these are as follows.

1) BC with perfect CSIR and CDIT: The broadcast channel capacity is only known when both the transmitter and the receivers have perfect knowledge of the channel.
2) CDIT and CDIR: Since perfect CSI is rarely possible, a study of capacity with CDI at both the transmitter(s) and receiver(s) for both MAC and BC is of great practical relevance.
3) Non-DPC techniques for BC: DPC is a very powerful capacity-achieving scheme, but it appears quite difficult to implement in practice. Thus, non-DPC multiuser transmissions schemes for the downlink (such as downlink beamforming [60]) are also of practical relevance.

IV. Multicell MIMO

The MAC and the BC are information theoretic abstractions of the uplink and the downlink of a single cell in a cellular system. However, a cellular system, by definition, consists of many cells. Due to the fundamental nature of wireless propagation, transmissions in a cell are not limited to within that cell. Users and base stations in adjacent cells experience interference from each other. Also, since the base stations are typically not mobile themselves there is the possibility for the base stations to communicate through a high-speed reliable connection, possibly consisting of optical fiber links capable of very high data rates. This opens up the opportunity for base stations to cooperate in the way they process different users' signals. Analysis of the capacity of the cellular network, explicitly taking into account the presence of multiple cells, multiple users and multiple antennas, and the possibilities of cooperation between base stations is inevitably a hard problem and runs into several long-standing unsolved problems in network information theory. However, such an analysis is also of utmost importance because it defines a common benchmark that can be used to gauge the efficiency of any practical scheme, in the same way that the capacity of a single-user link serves as a measure of the performance of practical schemes. There has been some recent

research in this area that extends the single-cell MAC and BC results to multiple cells. In this section, we summarize some of these results.

The key to the extension of single-cell results to multiple-cell systems is the assumption of perfect cooperation between base stations. Conceptually, this allows the multiple base stations to be treated as physically distributed antennas of one composite base station. Specifically, consider a group of B coordinated cells, each with M antennas and K mobiles, each with N antennas. If we define $\mathbf{H}_{i,j} \in \mathcal{C}^{N\times M}$ to be the *downlink* channel of user i from base station j, then the *composite* downlink channel of user i is $\mathbf{F}_i = [\mathbf{H}_{i,1} \cdots \mathbf{H}_{i,B}]$ and the composite uplink channel is $\mathbf{F}_i^\dagger$. The received signal of user i can then be written as $\mathbf{Y}_i = \mathbf{F}_i\mathbf{W} + \mathbf{n}_i$, where $\mathbf{W}$ is the composite transmitted signal defined as $\mathbf{W}^T = [\mathbf{W}_1^T \cdots \mathbf{W}_B^T]$. Here, we let $\mathbf{W}_j$ represent the transmit signal from base j.

First, let us consider the uplink. As pointed out by Jafar *et al.* [36] the single-cell MIMO MAC capacity region results apply to this system in a straightforward way. Thus, by assuming perfect data cooperation between the base stations, the multiple-cell uplink is easily seen to be equal to the MAC capacity region of the composite channel, defined as $\mathcal{C}_{\mathrm{MAC}}(\mathbf{F}_1^\dagger, \ldots, \mathbf{F}_K^\dagger; P_1, \ldots, P_K)$ in (12), where the power constraints of the ith mobile is P_i.

On the downlink, since the base stations can cooperate perfectly, DPC can be used over the entire transmitted signal (i.e., across base stations) in a straightforward manner. The application of DPC to a multiple-cell environment with cooperation between base stations is pioneered in recent work by Shamai and Zaidel [64]. For one antenna at each user and each base station, they show that a relatively simple application of DPC can enhance the capacity of the cellular downlink. While capacity computations are not the focus of [64], they do show that their scheme is asymptotically optimal at high SNRs.

The MIMO downlink capacity is explored by Jafar and Goldsmith in [39]. Note that the multicell downlink can be solved in a similar way as the uplink. But this requires perfect data and *power cooperation* between the base stations. If we let $\mathbf{X}_{i,j}$ represent the transmit vector for User i from base station j, the composite transmit vector intended for User i is $\mathbf{X}_i^T = [\mathbf{X}_{i,1}^T \cdots \mathbf{X}_{i,B}^T]$. Thus, the composite covariance of user i is defined as $\Sigma_i = \mathbb{E}[\mathbf{X}_i\mathbf{X}_i^T]$. The covariance matrix of the entire transmitted signal is $\Sigma_x = \sum_{i=1}^K \Sigma_i$. Assuming perfect data cooperation between the base stations, DPC can be applied to the composite vectors intended for different users. Thus, the dirty paper region described in Section III-C1, (15), can be achieved in the multicell downlink.

While data cooperation is a justifiable assumption for capacity computations in the sense that it captures the possibility of base stations cooperating among themselves as described earlier in this section, in practice each base station has its own power constraint. The per-base power constraint can be expressed as $\mathbb{E}[\mathbf{W}_j^\dagger\mathbf{X}_j] \le P_j$, where P_j is the power constraint at base j. Thus power cooperation, or pooling the transmit power for all the base stations to have one overall transmit power constraint, is not realistic. Note that on the uplink the base stations are only receiving signals and, therefore, no power cooperation is required. The per-base power constraints restrict consideration to covariance matrices such that $\sum_{i=1}^K \mathrm{Tr}(\mathbb{E}[\mathbf{X}_{i,j}\mathbf{X}_{i,j}^\dagger]) \le P_j j = 1, \ldots, B$. This is equivalent to a constraint of P_1 on the sum of the first M diagonal entries of Σ_x, a constraint of P_2 on the sum on the next M diagonal entries of Σ_x, etc. These constraints are considerably stricter than a constraint on the trace of Σ_x as in the single-cell case.

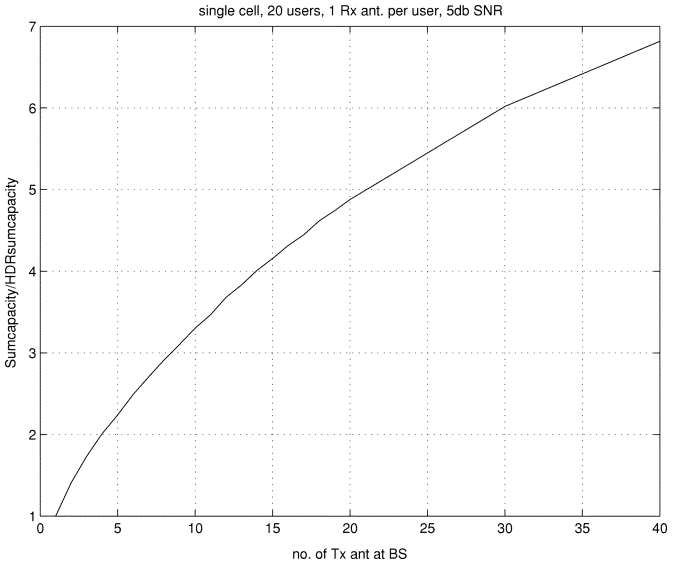

Fig. 10. Optimal sum rate relative to HDR.

Though DPC yields an achievable region, it has not been shown to achieve the capacity region or even the sum-rate capacity with per-base power constraints. Additionally, the MAC-BC duality (Section III-C2) which greatly simplified calculation of the dirty paper region does not apply under per-base power constraints. Thus, even generating numerical results for the multicell downlink is quite challenging.

However, data and power cooperation does give a simple upper bound on the capacity of the network. Based on numerical comparisons between this upper bound and a lower bound on capacity derived in [39], Jafar and Goldsmith find that the simple upper bound with power and data cooperation is also a good measure of the capacity with data cooperation alone.

Note that current wireless systems use the high data rate (HDR) protocol and transmit to only one user at a time on the downlink, where this best user is chosen to maximize the average system data rate. In contrast, DPC allows the base station to transmit to many users simultaneously. This is particularly advantageous when the number of transmit antennas at the base station is much larger than the number of receive antennas at each user—a common scenario in current cellular systems. To illustrate the advantages of DPC over HDR, even for a single cell, the relative gains of optimal DPC over a strategy that serves only the best user at any time are shown in Fig. 10. Note that this single-cell model is equivalent to the multicell system with no cooperation between base stations so that the interference from other cells is treated as noise. With cooperation between base stations the gains are expected to be even more significant as DPC reduces the overall interference by making some users invisible to others.

The capacity results described in this section address just a few out of many interesting questions in the design of a cellular system with multiple antennas. Multiple antennas can be used not only to enhance the capacity of the system but also to drive down the probability of error through diversity combining. Recent work by Zheng and Tse [88] unravels a fundamental diversity versus multiplexing tradeoff in MIMO systems. Also, instead of using isotropic transmit antennas on the downlink and transmitting to many users, it may be simpler to use directional antennas to divide the cell into sectors and transmit to one user within each sector. The relative impact of CDIT and/or CDIR on each of these schemes is not fully understood. Although in this paper we focus on the physical layer, smart schemes to handle CDIT can also be found at higher layers. An interesting example is the idea of opportunistic beamforming [78]. In the absence of CSIT, the transmitter randomly chooses the beamforming weights. With enough users in the system, it becomes very likely that these weights will be nearly optimal for one of the users. In other words, a random beam selected by the transmitter is very likely to be pointed toward a user if there are enough users in the system. Instead of feeding back the channel coefficients to the transmitter the users simply feed back the SNRs they see with the current choice of beamforming weights. This significantly reduces the amount of feedback required. By randomly changing the weights frequently, the scheme also treats all users fairly.

V. Conclusion

We have summarized recent results on the capacity of MIMO channels for both single-user and multiuser systems. The great capacity gains predicted for such systems can be realized in some cases, but realistic assumptions about channel knowledge and the underlying channel model can significantly mitigate these gains. For single-user systems the capacity under perfect CSI at the transmitter and receiver is relatively straightforward and predicts that capacity grows linearly with the number of antennas. Backing off from the perfect CSI assumption makes the capacity calculation much more difficult and the capacity gains are highly dependent on the nature of the CSI/CDI, the channel SNR, and the antenna element correlations. Specifically, assuming perfect CSIR, CSIT provides significant capacity gain at low SNRs but not much at high SNRs. The insight here is that at low SNRs it is important to put power into the appropriate eigenmodes of the system. Interestingly, with perfect CSIR and CSIT, antenna correlations are found to increase capacity at low SNRs and decrease capacity at high SNRs. Finally, under CDIT and CDIR for a zero-mean spatially white channel, at high SNRs capacity grows relative to only the double log of the SNR with the number of antennas as a constant additive term. This rather poor capacity gain would not typically justify adding more antennas. However, at moderate SNRs the growth relative to the number of antennas is less pessimistic.

We also examined the capacity of MIMO broadcast and multiple-access channels. The capacity region of the MIMO MAC is well-known and can be characterized as a convex optimization problem. Duality allows the DPC achievable region for the MIMO BC, a nonconvex region, to be computed from the MIMO MAC capacity region. These capacity and achievable regions are only known for ergodic capacity under perfect CSIT and CSIR. Relatively little is known about the MIMO MAC and BC regions under more realistic CSI assumptions. A multicell system with base station cooperation can be modeled as a MIMO BC (downlink) or MIMO MAC (uplink), where the antennas associated with each base station are pooled by the system. Exploiting this antenna structure leads to significant capacity gains over HDR transmission strategies.

There are many open problems in this area. For single-user systems the problems are mainly associated with CDI only at either the transmitter or receiver. Most capacity regions associated with multiuser MIMO channels remain unsolved, especially ergodic capacity and capacity versus outage for the MIMO BC under perfect receiver CSI only. There are very few existing results for CDI at either the transmitter or receiver for any multiuser MIMO channel. Finally, the capacity of cellular systems with multiple antennas remains a relatively open area, in part because the single-cell problem is mostly unsolved and in part because the Shannon capacity of a cellular system is not well-defined and depends heavily on frequency assumptions and propagation models. Other fundamental tradeoffs in MIMO cellular designs such as whether antennas should be used for sectorization, capacity gain, or diversity are not well understood. In short, we have only scratched the surface in understanding the fundamental capacity limits of systems with multiple transmitter and receiver antennas, as well as the implications of these limits for practical system designs. This area of research is likely to remain timely, important, and fruitful for many years to come.

References

[1] A. Abdi and M. Kaveh, "A space-time correlation model for multielement antenna systems in mobile fading channels," *IEEE J. Select. Areas Commun.*, vol. 20, pp. 550–561, Apr. 2002.

[2] N. Al-Dhahir, "Overview and comparison of equalization schemes for space-time-coded signals with application to EDGE," *IEEE Trans. Signal Processing*, vol. 50, pp. 2477–2488, Oct. 2002.

[3] N. Al-Dhahir, C. Fragouli, A. Stamoulis, W. Younis, and R. Calderbank, "Space-time processing for broadband wireless access," *IEEE Commun. Mag.*, vol. 40, pp. 136–142, Sept. 2002.

[4] E. Biglieri, J. Proakis, and S. S. Shitz, "Fading channels: Information theoretic and communication aspects," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.

[5] H. Bolcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, pp. 225–234, Feb. 2002.

[6] S. Borst and P. Whiting, "The use of diversity antennas in high-speed wireless systems: Capacity gains, fairness issues, multi-user scheduling," *Bell Labs Tech. Mem.*, 2001.

[7] S. Boyd and L. Vandenberghe. (2001) Introduction to Convex Optimization With Engineering Applications. [Online]. Available: www.stanford.edu/~boyd/cvxbook.html

[8] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2007–2019, Sept. 1999.

[9] ——, "On achievable rates in a multi-antenna broadcast downlink," in *Proc. 38th Annual Allerton Conf. Communications, Control, Computing*, Oct. 2000, pp. 1188–1193.

[10] S. Catreux, V. Erceg, D. Gesbert, and R. W. Heath, "Adaptive modulation and MIMO coding for broadband wireless data networks," *IEEE Commun. Mag.*, vol. 40, pp. 108–115, June 2002.

[11] R. Cheng and S. Verdu, "Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling," *IEEE Trans. Inform. Theory*, vol. 39, pp. 773–785, May 1993.

[12] D. Chizhik, G. Foschini, M. Gans, and R. Valenzuela, "Keyholes, correlations and capacities of multielement transmit and receive antennas," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 361–368, Apr. 2002.

[13] D. Chizhik, J. Ling, P. Wolniansky, R. Valenzuela, N. Costa, and K. Huber, "Multiple input multiple output measurements and modeling in Manhattan," in *Proc. IEEE Vehicular Technology Conf.*, 2002, pp. 107–110.

[14] Chong and L. Milstein, "The performance of a space-time spreading CDMA system with channel estimation errors," in *Proc. Int. Communications Conf.*, Apr. 2002, pp. 1793–1797.

[15] C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inform. Theory*, vol. 48, pp. 637–650, Mar. 2002.

[16] C.-N. Chuah, D. N. Tse, J. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inform. Theory*, vol. 48, pp. 637–650, Mar. 2002.

[17] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, pp. 439–441, May 1983.

[18] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[19] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for cancelling known interference," in *Proc. Int. Symp. Information Theory Applications*, Nov. 2000, pp. 681–684.

[20] G. J. Foschini, "Layered space-time architecture for wireless communication in fading environments when using multi-element antennas," *Bell Labs Tech. J.*, pp. 41–59, 1996.

[21] G. J. Foschini, D. Chizhik, M. Gans, C. Papadias, and R. A. Valenzuela, "Analysis and performance of some basic spacetime architectures," *IEEE J. Select. Areas Commun., Special Issue on MIMO Systems*, pt. I, vol. 21, pp. 303–320, Apr. 2003.

[22] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.: Kluwer Academic Press*, no. 6, pp. 311–335, 1998.

[23] R. G. Gallager, "An inequality on the capacity region of multiaccess fading channels," in *Communication and Cryptography—Two Sides of One Tapestry*. Boston, MA: Kluwer, 1994, pp. 129–139.

[24] M. J. Gans, N. Amitay, Y. S. Yeh, H. Xu, T. Damen, R. A. Valenzuela, T. Sizer, R. Storz, D. Taylor, W. M. MacDonald, C. Tran, and A. Adamiecki, "Outdoor BLAST measurement system at 2.44 GHz: Calibration and initial results," *IEEE J. Select. Areas Commun.*, vol. 20, pp. 570–581, Apr. 2002.

[25] D. Gesbert, M. Shafi, D. S. Shiu, P. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Select. Areas Commun. Special Issue on MIMO Systems*, pt. I, vol. 21, pp. 281–302, Apr. 2003.

[26] L. Greenstein, J. Andersen, H. Bertoni, S. Kozono, D. Michelson, and W. Tranter, "Channel and propagation models for wireless system design I and II," *IEEE J. Select. Areas Commun.*, vol. 20, Apr./Aug. 2002.

[27] T. Guess and M. K. Varanasi, "Multiuser decision-feedback receivers for the general Gaussian multiple-access channel," in *Proc. Allerton Conf. Communications, Control, Computing*, Monticello, IL, Oct. 1996, pp. 190–199.

[28] S. Hanly and D. Tse, "Multiaccess fading channels-Part II: Delay-limited capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2816–2831, Nov. 1998.

[29] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inform. Theory*, vol. 49, pp. 951–963, Apr. 2003.

[30] ——, "Cayley differential unitary space-time codes," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1485–1503, June 2002.

[31] B. Hochwald, T. L. Marzetta, and V. Tarokh, "Multi-antenna channel-hardening and its implications for rate feedback and scheduling," IEEE Trans. Inform. Theory, 2002, submitted for publication.

[32] B. Hochwald and S. Vishwanath, "Space-time multiple access: Linear growth in sum rate," in *Proc. 40th Allerton Conf. Communications, Control, Computing*, Monticello, IL, Oct. 2002.

[33] M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1962–1973, Sept. 2000.

[34] H. Huang, H. Viswanathan, and G. J. Foschini, "Multiple antennas in cellular CDMA systems: Transmission, detection and spectral efficiency," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 383–392, July 2002.

[35] H. C. Huang, S. Venkatesan, and H. Viswanathan, "Downlink capacity evaluation of cellular networks with known interference cancellation," in *Proc. DIMACS Workshop on Signal Processing Wireless Communications*, DIMACS Center, Rutgers Univ., Oct. 7–9, 2002.

[36] S. Jafar, G. Foschini, and A. Goldsmith, "Phantomnet: Exploring optimal multicellular multiple antenna systems," in *Proc. Vehicular Technology Conf.*, 2002, pp. 261–265.

[37] S. Jafar and A. Goldsmith. Multiple-Antenna Capacity in Correlated Rayleigh Fading With no Side Information. [Online]. Available: http://wsl.stanford.edu/publications.html.

[38] ——, "Transmitter optimization and optimality of beamforming for multiple antenna systems with imperfect feedback," *IEEE Trans. Wireless Commun.*, submitted for publication.

[39] S. A. Jafar and A. Goldsmith, "Transmitter optimization for multiple antenna cellular systems," in *Proc. Int. Symp. Information Theory*, June 2002, p. 50.

[40] S. A. Jafar and A. J. Goldsmith, "On optimality of beamforming for multiple antenna systems with imperfect feedback," in *Proc. Int. Symp. Information Theory*, June 2001, p. 321.

[41] ——, "Vector mac capacity region with covariance feedback," in *Proc. Int. Symp. Information Theory*, June 2001, p. 321.

[42] S. A. Jafar, S. Vishwanath, and A. J. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *Proc. Int. Conf. Communications*, vol. 7, 2001, pp. 2266–2270.

[43] N. Jindal, S. Jafar, S. Vishwanath, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," in *Proc. Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, Nov. 3–6, 2002.

[44] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," in *Proc. Int. Symp. Inform. Theory*, June 2002, p. 500.

[45] E. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE J. Select. Areas Commun.*, submitted for publication.

[46] ——, "Optimal transmission with imperfect channel state information at the transmit antenna array," *Wireless Personal Commun.*, submitted for publication.

[47] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multi-antenna systems on flat fading channels," *IEEE Trans. Inform. Theory*, submitted for publication.

[48] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels-Part I: Ergodic capacity," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1083–1102, Mar. 2001.

[49] ——, "Capacity and optimal resource allocation for fading broadcast channels-Part II: Outage capacity," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1103–1127, Mar. 2001.

[50] Y. Li, J. Winters, and N. Sollenberger, "MIMO-OFDM for wireless communication: Signal detection with enhanced channel estimation," *IEEE Trans. Commun.*, pp. 1471–1477, Sept. 2002.

[51] A. Lozano and C. Papadias, "Layered space-time receivers for frequency-selective wireless channels," *IEEE Trans. Commun.*, vol. 50, pp. 65–73, Jan. 2002.

[52] B. Lu, X. Wang, and Y. Li, "Iterative receivers for space-time block-coded OFDM systems in dispersive fading channels," *IEEE Trans. Wireless Commun.*, vol. 1, pp. 213–225, Apr. 2002.

[53] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.

[54] ——, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–564, Mar. 2000.

[55] A. F. Molisch, M. Stienbauer, M. Toeltsch, E. Bonek, and R. S. Thoma, "Capacity of MIMO systems based on measured wireless channels," *IEEE J. Select. Areas Commun.*, vol. 20, pp. 561–569, Apr. 2002.

[56] A. Moustakas and S. Simon. Optimizing Multi-Transmitter Single-Receiver (MISO) Antenna Systems With Partial Channel Knowledge. [Online]. Available: http://mars.bell-labs.com.

[57] A. Narula, M. Trott, and G. Wornel, "Performance limits of coded diversity methods for transmitter antenna arrays," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2418–2433, Nov. 1999.

[58] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple antenna data transmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1423–1436, Oct. 1998.

[59] G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.

[60] F. Rashid-Farrokhi, K. R. Liu, and L. Tassiulas, "Transit beamforming and power control for cellular wireless systems," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1437–1450, Oct. 1998.

[61] H. Sato, "An outer bound on the capacity region of the broadcast channel," *IEEE Trans. Inform. Theory*, vol. 24, pp. 374–377, May 1978.
[62] S. Shamai and T. L. Marzetta, "Multiuser capacity in block fading with no channel state information," *IEEE Trans. Inform. Theory*, vol. 48, pp. 938–942, Apr. 2002.
[63] S. Shamai and A. D. Wyner, "Information-theoretic considerations for symmetric, cellular, multiple-access fading channels," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1877–1991, Nov. 1997.
[64] S. Shamai and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. IEEE Vehicular Technology Conf.*, May 2001, pp. 1745–1749.
[65] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multi-element antenna systems," *IEEE Trans. Commun.*, vol. 48, pp. 502–513, Mar. 2000.
[66] S. Simon and A. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. Select. Areas Commun.*, vol. 21, pp. 406–417, Apr. 2003.
[67] ——, "Optimality of beamforming in multiple transmitter multiple receiver communication systems with partial channel knowledge," in *Proc. DIMACS Workshop Signal Proessing Wireless Communications*, DIMACS Center, Rutgers Univ., Oct. 7–9, 2002.
[68] P. J. Smith and M. Shafi, "On a Gaussian approximation to the capacity of wireless MIMO systems," in *Proc. Int. Conf. Communications*, Apr. 2002, pp. 406–410.
[69] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm. ETT*, vol. 10, no. 6, pp. 585–596, Nov. 1999.
[70] D. Tse and S. Hanly, "Multiaccess fading channels-Part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2796–2815, Nov. 1998.
[71] A. Tulino, A. Lozano, and S. Verdu, "Capacity of multi-antenna channels in the low power regime," in *Proc. IEEE Information Theory Workshop*, Oct. 2002, pp. 192–195.
[72] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1319–1343, June 2002.
[73] S. Vishwanath, S. Jafar, and A. Goldsmith, "Optimum power and rate allocation strategies for multiple access fading channels," in *Proc. Vehicular Technology Conf.*, May 2000, pp. 2888–2892.
[74] S. Vishwanath, N. Jindal, and A. Goldsmith, "On the capacity of multiple input multiple output broadcast channels," in *Proc. Int. Conf. Communications*, Apr. 2002, pp. 1444–1450.
[75] S. Vishwanath, G. Kramer, S. Shamai(Shitz), S. A. Jafar, and A. Goldsmith, "Outer bounds for multi-antenna broadcast channels," in *Proc. DIMACS Workshop on Signal Processing Wireless Communications*, DIMACS Center, Rutgers Univ., Oct. 7–9, 2002.
[76] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2632–2639, Sept. 2001.
[77] P. Viswanath and D. Tse, "On the capacity of the multi-antenna broadcast channel," in *Proc. DIMACS workshop on Signal Processing Wireless Communications*, DIMACS Center, Rutgers Univ., Oct. 7–9, 2002.
[78] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1277–1294, June 2002.
[79] P. Viswanath and D. N. Tse, "Sum capacity of the multiple antenna Gaussian broadcast channel," in *Proc. Int. Symp. Information Theory*, June 2002, p. 497.
[80] P. Viswanath, D. N. Tse, and V. Anantharam, "Asymptotically optimal water-filling in vector multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 47, pp. 241–267, Jan. 2001.
[81] J. Winters, "On the capacity of radio communication systems with diversity in a Rayleigh fading environment," *IEEE J. Select. Areas Commun.*, vol. 5, pp. 871–878, June 1987.
[82] Y. Xin and G. Giannakis, "High-rate space-time layered OFDM," *IEEE Commun. Lett.*, pp. 187–189, May 2002.
[83] W. Yu and J. Cioffi, "Trellis precoding for the broadcast channel," in *Proc.Global Communications Conf.*, Oct. 2001, pp. 1344–1348.
[84] W. Yu and J. M. Cioffi, "Sum capacity of a Gaussian vector broadcast channel," in *Proc. Int. Symp. Information Theory*, June 2002, p. 498.
[85] W. Yu, G. Ginis, and J. Cioffi, "An adaptive multiuser power control algorithm for VDSL," in *Proc. Global Communications Conf.*, Oct. 2001, pp. 394–398.
[86] W. Yu, W. Rhee, S. Boyd, and J. Cioffi, "Iterative water-filling for vector multiple access channels," in *Proc. IEEE Int. Symp. Information Theory*, 2001, p. 322.
[87] W. Yu, W. Rhee, and J. Cioffi, "Optimal power control in multiple access fading channels with multiple antennas," in *Proc. Int. Conf. Communications*, 2001, pp. 575–579.
[88] L. Zheng and D. Tse, "Optimal diversity-multiplexing tradeoff in multiple antenna channels," in *Proc. Allerton Conf. Communications, Control, Computing*, Monticello, IL, Oct. 2001, pp. 835–844.
[89] L. Zheng and D. N. Tse, "Packing spheres in the Grassmann manifold: A geometric approach to the noncoherent multi-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359–383, Feb. 2002.

Andrea Goldsmith (S'90–M'93–SM'99) received the B.S., M.S., and Ph.D. degrees in electrical engineering from University of California, Berkeley, in 1986, 1991, and 1994, respectively.

She was an Assistant Professor in the Department of Electrical Engineering, California Institute of Technology (Caltech), Pasadena, from 1994 to 1999. In 1999, she joined the Electrical Engineering Department, Stanford University, Stanford, CA, where she is currently an Associate Professor. Her industry experience includes affiliation with Maxim Technologies, Santa Clara, CA, from 1986 to 1990, where she worked on packet radio and satellite communication systems and with AT&T Bell Laboratories, Holmdel, NJ, from 1991 to 1992, where she worked on microcell modeling and channel estimation. Her research includes work in capacity of wireless channels and networks, wireless information and communication theory, multiantenna systems, joint source and channel coding, cross-layer wireless network design, communications for distributed control and adaptive resource allocation for cellular systems and *ad-hoc* wireless networks.

Dr. Goldsmith is a Terman Faculty Fellow at Stanford University and a recipient of the Alfred P. Sloan Fellowship, the National Academy of Engineering Gilbreth Lectureship, a National Science Foundation CAREER Development Award, the Office of Naval Research Young Investigator Award, a National Semiconductor Faculty Development Award, an Okawa Foundation Award, and the David Griep Memorial Prize from University of California, Berkeley. She was an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS from 1995 to 2002 and has been an Editor for the IEEE WIRELESS COMMUNICATIONS MAGAZINE since 1995. She is also an Elected Member of Stanford's Faculty Senate and the Board of Governors for the IEEE Information Theory Society.

Syed Ali Jafar (S'99) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology (IIT), Delhi, in 1997 and the M.S. degree in electrical engineering from California Institute of Technology (Caltech), Pasadena, in 1999. He is a Graduate Research Assistant in the Wireless Systems Lab, Stanford University, Stanford, CA, and is currently working toward the Ph.D. degree in electrical engineering.

He was a Summer Intern in the Wireless Communications Group of Lucent Bell Laboratories, Holmdel, NJ, in 2001 and has two pending patents resulting from that work. He was also an Engineer in the satellite networks division of Hughes Software Systems, India, from 1997 to 1998. His research interests include spread-spectrum systems, multiple antenna systems, and multiuser information theory.

Nihar Jindal (S'99) received the B.S. degree in electrical engineering and computer science from University of California, Berkeley, in 1999 and the M.S. degree in electrical engineering from Stanford University, Stanford, CA, in 2001, and is currently working toward the Ph.D. degree at the same university.

His industry experience includes summer internships at Intel Corporation, Santa Clara, CA, in 2000 and at Lucent Bell Labs, Holmdel, NJ, in 2002. His research interests include multiple-antenna channels and multiuser information theory and their applications to wireless communication.

Sriram Vishwanath (S'99) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology (IIT), Madras, in 1998 and the M.S. degree in electrical engineering from the California Institute of Technology (Caltech), Pasadena, in 1999. He is a graduate fellow currently working toward the Ph.D. degree in electrical engineering at Stanford University, Stanford, CA.

His research interests include information and coding theory, with a focus on multiple antenna systems. His industry experience includes work at National Semiconductor Corporation, Santa Clara, CA, in the Summer of 2000 and at the Lucent Bell Labs, Murray Hill, NJ, during the Summer of 2002.

Information Capacity of a Random Signature Multiple-Input Multiple-Output Channel

Predrag B. Rapajic and Dan Popescu

***Abstract*—A closed-form expression for information capacity of the random signature multiple-input multiple-output channel is given. A direct calculation of the capacity is provided additional to the proof involving the capacity of minimum mean-square-error systems provided in [8].**

***Index Terms*—Code-division multiple access, communication systems, interchannel interference, MIMO systems, multibeam antennas, multiuser channels, space-division multiplexing.**

I. Introduction

A MULTIPLE-INPUT multiple-output (MIMO) linear channel model is common for many multiple-access communications scenarios, such as 1) direct-sequence code-division multiple-access system (CDMA) [7]; 2) cellular mobile system with multisensor (antenna array) reception in base stations [5]; and 3) multicellular system with joint-multiuser detection (in contrast to cell-by-cell separate detection) [1].

The first reliable estimate of the random signature (RS) CDMA channel capacity was given in [4] in the form of upper and lower bound. The explicit formula for RS-CDMA channel capacity is given in [8]. The explicit formula for the capacity is found as integral/sum of the capacities in a system where ideal cancelers with minimum mean-square-error (MMSE) prefiltering are used. This solution is mathematically short and elegant. However, historically first development of the explicit capacity formula was given as a tedious direct solution without involving the argument of MMSE cancellation, [6]. A slight improvement over the original derivation of the direct solution for the capacity formula is given in this paper.

II. System Definition

A linear MIMO communication channel is represented by the following equation:

$$\boldsymbol{y} = \boldsymbol{S}\boldsymbol{x} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{y}$ is the N-dimensional vector of signal outputs. The symbols sent by K users are represented as K-dimensional vector $\boldsymbol{x}$ and $\boldsymbol{n}$ is a N-dimensional vector representing zero mean additive white Gaussian noise (AWGN) with the covariance matrix $E\{\boldsymbol{n}\boldsymbol{n}^H\} = \boldsymbol{I}\sigma^2$. The matrix $\boldsymbol{S}$ is the channel signature matrix, where its components $s_{k,n}$, $k = 1, \cdots, K$, $n = 1, \cdots, N$, are complex numbers. All signals are treated as complex numbers. Referring to the system dimensionality as N, means that N complex dimensions are considered which is in fact $2N$ real dimensions. We consider the *proper complex random processes* [2] where the real and the imaginary part of the system could be treated as two mutually orthogonal systems with the same dimensionality. Factor $1/2$ is removed from the capacity formula and this convention is kept throughout the paper. The capacity of a MIMO channel depicted by (1) is

$$C = \log\left|\boldsymbol{I} + \frac{\boldsymbol{S}\boldsymbol{V}_{\boldsymbol{x}}\boldsymbol{S}^H}{\sigma^2}\right| \tag{2}$$

where $\boldsymbol{V}_{\boldsymbol{x}} = E\{\boldsymbol{x}\boldsymbol{x}^H\}$ is the symbol vector covariance matrix, $|\cdot|$ denotes matrix determinant, and $(\cdot)^H$ denotes conjugate transpose of a matrix [9].

Our objective is to find explicit formula for the MIMO system information capacity where $\boldsymbol{S}$ is the random matrix called RS-MIMO system.

III. Expression for RS-MIMO Channel Capacity

In deriving expression for RS-MIMO channel capacity, we start from (2). If the symbol powers are normalized to 1, i.e., $\boldsymbol{V}_{\boldsymbol{x}} = \boldsymbol{I}$, then (2) takes the form

$$C = \log\left|\boldsymbol{I} + \frac{\boldsymbol{S}\boldsymbol{S}^H}{\sigma^2}\right| = \sum_{n=1}^{N} \log\left(1 + \frac{\lambda_n}{\sigma^2}\right) \tag{3}$$

where λ_n, $n = 1, \cdots, N$, are eigenvalues of the matrix $\boldsymbol{Z} = \boldsymbol{S}\boldsymbol{S}^H$.

If the normalized number of users is defined as $\rho = K/N$, depending on how the number of users K measures with respect to the system dimensionality N, the following cases are of interest. The **unsaturated system,** where $K < N$, $(0 < \rho < 1)$, and the **oversaturated system,** where $K > N$, $(1 < \rho < \infty)$. For notational reasons during cumbersome derivations, sometimes it is more convenient to use $y = 1/\rho$ as a reciprocal value of ρ, rather than ρ itself.

It is possible to find the probability density function $f(x)$ of eigenvalues λ_n when N and K are large [3]. For $0 < y \leq 1$, this is a continuous density function

$$\begin{aligned} f_y(x) &= \frac{\sqrt{(x - a(y))(b(y) - x)}}{2\pi y x}, \qquad \text{for } a(y) < x < b(y) \\ &= 0, \qquad \text{otherwise} \end{aligned} \tag{4}$$

where $a(y) = 1 + y - 2\sqrt{y}$ and $b(y) = 1 + y + 2\sqrt{y}$. Then, with the probability equal to 1, summation in (3) could be replaced by Lebesgue integral over probability measure given by [4]

$$E\{C\} = N \int_0^{\infty} \log\left(1 + \frac{x}{\sigma^2}\right) dF_y(x). \tag{5}$$

Paper approved by G. Caire, the Editor for Multiuser Detection and CDMA of the IEEE Communications Society. Manuscript received June 16, 1999; revised December 17, 1999 and October 15, 1999.

The authors are with the School of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney, NSW 2052, Australia.

Publisher Item Identifier S 0090-6778(00)07100-2.

A. Capacity of the Oversaturated System

For the oversaturated system, it makes practical sense to limit the total signal power of the system and make it independent of the number of users. The power limiting condition is imposed on matrix $\boldsymbol{Z}$, ($\boldsymbol{Z} = (1/K)\boldsymbol{S}\boldsymbol{S}^H$). The total average power of such a system is N and the total number of dimensions of this system is N regardless of the number of users K. For the oversaturated system, where $\rho > 1$, i.e., $0 < y < 1$, the expression for eigenvalue distribution (4) applies directly. Equation (5) becomes

$$E\{C\} = C_O = NC_o$$

where C_O is the oversaturated RS-MIMO system capacity. The oversaturated RS-MIMO capacity per a system dimension C_o is given in a closed form by

$$C_o = \log \frac{w_y}{\sigma^2} + \frac{1-y}{y} \log\left(\frac{1}{1-v_y}\right) - \frac{v_y}{y} \tag{6}$$

with

$$\begin{aligned} w_y &= w(y, \sigma^2) \\ &= \frac{1}{2}\left(1+y+\sigma^2 + \sqrt{(1+y+\sigma^2)^2 - 4y}\right) \qquad (7) \\ v_y &= v(y, \sigma^2) \\ &= \frac{1}{2}\left(1+y+\sigma^2 - \sqrt{(1+y+\sigma^2)^2 - 4y}\right) \qquad (8) \end{aligned}$$

where $y = N/K = 1/\rho$ is the normalized number of system dimensions per user (inverse of the normalized number of users per system dimension) and σ^2 is the AWGN variance. The proof is given in the Appendix.

B. Capacity of the Unsaturated System

For the unsaturated system, the number of dimensions is larger than the number of users. From a practical engineering point of view, it is useful to compare unsaturated RS-MIMO system with an orthogonal system with the same number of users and the same power limitation per user. Now, we limit the signal power per **user,** i.e., impose following condition on matrix $\boldsymbol{Z}$, ($\boldsymbol{Z} = (1/N)\boldsymbol{S}\boldsymbol{S}^H$). That means that the power per user rather than the power per system dimension is limited to 1. Please note that the difference for unsaturated system power limitation is per user, where for oversaturated system, power limitation is per system dimension (or which is the same, the total system power independent of the number of users). Matrix $\boldsymbol{Z}$ has dimensionality $N \times N$, but only K of its eigenvalues are different from 0. In fact, those K nonzero, eigenvalues are exactly equal to the eigenvalues of matrix $(1/N)\boldsymbol{S}^H\boldsymbol{S}$ which is $K \times K$ dimensional matrix. Then, with λ being a stochastic root of $(1/N)\boldsymbol{S}^H\boldsymbol{S}$ and $\rho = K/N$, For $0 < \rho \leq 1$, probability density function of λ is a continuous distribution with density

$$\begin{aligned} f_\rho(x) &= \frac{\sqrt{(x-a(\rho))(b(\rho)-x)}}{2\pi\rho x}, \qquad \text{for } a(\rho) < x < b(\rho) \\ &= 0, \qquad \text{otherwise} \end{aligned}$$

where $a(\rho) = 1 + \rho - 2\sqrt{\rho}$ and $b(\rho) = 1 + \rho + 2\sqrt{\rho}$. This statement follows verbatim if K is replaced by N, N is replaced by K, and y is replaced by ρ. The clear benefit is that the capacity per **dimension** formula (6) of the oversaturated RS-MIMO could be used verbatim as the capacity per **user** of the RS-MIMO unsaturated system without lengthy derivations.

For the unsaturated system, $\rho < 1$, the average capacity formula per user is

$$C_u = \log \frac{w_\rho}{\sigma^2} + \frac{1-\rho}{\rho} \log\left(\frac{1}{1-v_\rho}\right) - \frac{v_\rho}{\rho} \tag{9}$$

with

$$\begin{aligned} w_\rho &= w(\rho, \sigma^2) \\ &= \frac{1}{2}\left(1+\rho+\sigma^2 + \sqrt{(1+\rho+\sigma^2)^2 - 4\rho}\right) \\ v_\rho &= v(\rho, \sigma^2) \\ &= \frac{1}{2}\left(1+\rho+\sigma^2 - \sqrt{(1+\rho+\sigma^2)^2 - 4\rho}\right) \end{aligned}$$

where $\rho = K/N$ is the normalized number of users and σ^2 is the AWGN variance.

APPENDIX

By starting with (5)

$$\begin{aligned} C_o &= \int_0^\infty \log\left(1 + \frac{x}{\sigma^2}\right) dF_y(x) \\ &= \int_0^\infty \log(\sigma^2 + x)\, dF_y(x) + \log\frac{1}{\sigma^2} \int_0^\infty dF_y(x) \\ &= C_{\sigma^2} - \log(\sigma^2). \end{aligned}$$

With substitution $\alpha = \sigma^2$ (in order to simplify notation) and for $\alpha > 0$ and $0 < y < 1$ (oversaturated system) integral

$$C_\alpha = \int_0^\infty \log(\alpha + x)\, dF_y(x)$$

could be written as

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

where $a(y) = 1 + y - 2\sqrt{y}$ and $b(y) = 1 + y + 2\sqrt{y}$. With substitution, $x = 1 + y - 2\sqrt{y}\cos t$ becomes

$$C_\alpha = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\log(1+y+\alpha-2\sqrt{y}\cos t)}{1+y-2\sqrt{y}\cos t} \cdot \left(1 - \frac{e^{2it}+e^{-2it}}{2}\right) dt. \tag{10}$$

It is possible to find constants w, u, $|u| < 1$, such that

$$1 + y + \alpha - 2\sqrt{y}\cos t = w(1 + u^2 - 2u\cos t). \tag{11}$$

Since (11) has to be satisfied for every t, constants w and u have to satisfy

$$\begin{aligned} \alpha+1+y &= w(1+u^2) \\ \sqrt{y} &= wu. \end{aligned} \tag{12}$$

By eliminating w from previous set of equations, a quadratic equation with respect to u is obtained as

$$u^2 - \frac{1+y+\alpha}{\sqrt{y}}\,u + 1 = 0 \tag{13}$$

with the following roots:

$$u_{1,2} = \frac{1}{2\sqrt{y}}\left(1+y+\alpha \pm \sqrt{(1+y+\alpha)^2-4y}\right).$$

Since $u<1$, then the only valid solution is

$$u = \frac{1}{2\sqrt{y}}\left(1+y+\alpha - \sqrt{(1+y+\alpha)^2-4y}\right). \tag{14}$$

Observation: Indeed, $u<1$. Since $(1+y+\alpha)-4y \geq 0$ for $\alpha,\, y>0$ the roots $u_1,\, u_2$ of (13) are real, then with $u_1u_2=1$ must be $u<1$.

Substituting u in (12), we get

$$w = \frac{1}{2}\left(1+y+\alpha+\sqrt{(1+y+\alpha)^2-4y}\right). \tag{15}$$

Starting with (10), we may write

$$C_\alpha = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\log w + \log(1+u^2-2u\cos t)}{1+y-2\sqrt{y}\cos t} \cdot\left(1-\frac{e^{2it}+e^{-2it}}{2}\right) dt.$$

Then

$$C_\alpha = I_1 + I_2$$

where

$$\begin{aligned} I_1 &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\log w}{1+y-2\sqrt{y}\cos t}\left(1-\frac{e^{2it}+e^{-2it}}{2}\right) dt \\ &= \log w \end{aligned}$$

and

$$I_2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\log(1+u^2-2u\cos t)}{1+y-2\sqrt{y}\cos t}\left(1-\frac{e^{2it}+e^{-2it}}{2}\right) dt.$$

The following identity is used in further derivations:

$$\begin{aligned} \log(1+u^2-2u\cos t) &= \log(1-ue^{it})(1-ue^{-it}) \\ &= -\sum_{m=1}^{\infty}\frac{u^m}{m}e^{imt} - \sum_{m=1}^{\infty}\frac{u^m}{m}e^{-imt}. \end{aligned} \tag{16}$$

The function

$$p_r(t) = \frac{1}{2\pi}\cdot\frac{1-r^2}{1+r^2-2r\cos t}, \qquad 0<r<1$$

has the representation

$$\frac{1}{2\pi}\sum_{s=-\infty}^{\infty} r^{|s|}e^{ist}. \tag{17}$$

Using (16) and $p_r(t)/1-r^2$, with $r=\sqrt{y}$ from (17), we get

$$\begin{aligned} I_2 &= \frac{1}{2\pi(1-y)}\int_{-\pi}^{\pi} -\sum_{s=-\infty}^{\infty}\sqrt{y}^{|s|}e^{ist} - \sum_{m=1}^{\infty}\frac{u^m}{m} \\ &\quad\cdot(e^{-imt}+e^{-imt})\left(1-\frac{e^{2it}+e^{-2it}}{2}\right) dt \\ &= \frac{1}{2\pi(1-y)}\int_{-\pi}^{\pi}\sum_{k=-\infty}^{\infty} g_k(u,\,y,\,t)e^{ikt} \\ &\quad\cdot\left(1-\frac{e^{2it}+e^{-2it}}{2}\right) dt. \end{aligned}$$

We note that $\int_{-\pi}^{\pi} e^{-imt}\,dt = 0$ for $m\neq 0$, i.e., we need only consider the constant terms under the integral sign. This means that the only relevant terms of $g_k(u,\,y,\,t)$, $k=-\infty\cdots,\, -2,\, 1,\, 0,\, 1,\, 2,\cdots,\,\infty$, are $g_2(u,\,y,\,t)e^{-2it}$, $g_0(u,\,y,\,t)$, and $g_{-2}(u,\,y,\,t)e^{2it}$. All other terms are negligible.

Using the $(m,\,s)$-terms $(1,\,1)$, $(2,\,2)$, $(3,\,3),\cdots$, and $(1,\,-1)$, $(2,\,-2)$, $(3,\,-3),\cdots$, respectively, we get

$$\begin{aligned} &g_0(u,\,y,\,t) \\ &= 2\left(-u\sqrt{y} - \frac{\left(u\sqrt{y}\right)^2}{2} - \frac{\left(u\sqrt{y}^3\right)}{3} - \frac{\left(u\sqrt{y}\right)^4}{4} - \cdots\right) \\ &= 2\log\left(1-u\sqrt{y}\right). \end{aligned}$$

Using the $(m,\,s)$-terms $(1,\,-3)$, $(2,\,-4)$, $(3,\,-5),\cdots$, and $(1,\,-1)$, $(2,\,0)$, $(3,\,1),\cdots$, we get

$$\begin{aligned} g_{-2}(y,\,t) &= -y\left(\sqrt{y}u + \frac{\left(u\sqrt{y}\right)^2}{2} + \frac{\left(u\sqrt{y}\right)^3}{3} + \cdots\right) \\ &\quad -\left(u\sqrt{y} + \frac{u^2}{2} + \frac{u^3\sqrt{y}}{3} + \frac{u^4\sqrt{y}^2}{4} + \cdots\right) \\ &= y\log\left(1-u\sqrt{y}\right) - u\sqrt{y} + \frac{u}{\sqrt{y}} + \frac{1}{y} \\ &\quad\cdot\log\left(1-u\sqrt{y}\right) \\ &= \left(y+\frac{1}{y}\right)\log\left(1-u\sqrt{y}\right) + \frac{u}{\sqrt{y}} - u\sqrt{y}. \end{aligned}$$

Finally, using the $(m,\,s)$-terms $(1,\,3)$, $(2,\,4)$, $(3,\,5),\cdots$, and $(1,\,1)$, $(2,\,0)$, $(3,\,-1),\cdots$, we get

$$\begin{aligned} g_2(y,\,t) &= -y\left(\sqrt{y}u + \frac{\left(u\sqrt{y}\right)^2}{2} + \frac{\left(u\sqrt{y}\right)^3}{3} + \cdots\right) \\ &\quad -\left(u\sqrt{y} + \frac{u^2}{2} + \frac{u^3\sqrt{y}}{3} + \frac{u^4\sqrt{y}^2}{4} + \cdots\right) \\ &= y\log\left(1-u\sqrt{y}\right) - u\sqrt{y} + \frac{u}{\sqrt{y}} + \frac{1}{y} \\ &\quad\cdot\log\left(1-u\sqrt{y}\right) \\ &= \left(y+\frac{1}{y}\right)\log\left(1-u\sqrt{y}\right) + \frac{u}{\sqrt{y}} - u\sqrt{y}. \end{aligned}$$

By using the remaining terms containing e^{imt} for $m = -2$, $m = 0$, $m = 2$, we may write the integral as

$$I_2 = \frac{1}{2\pi(1-y)} \int_{-\pi}^{\pi} 2\log(1 - u\sqrt{y})\, dt + -\frac{1}{2\pi(1-y)} \int_{-\pi}^{\pi} \left(\left(y + \frac{1}{y} \right) \log(1 - u\sqrt{y}) + \left(\frac{u}{\sqrt{y}} - u\sqrt{y} \right) \right) \frac{(e^{2it} + e^{-2it})^2}{2}\, dt$$

$$I_2 = \frac{2\log(1-u\sqrt{y})}{(1-y)} - \frac{1}{2\pi(1-y)} \int_{-\pi}^{\pi} \left(\left(y + \frac{1}{y} \right) \cdot \log(1 - u\sqrt{y}) + \left(\frac{u}{\sqrt{y}} - u\sqrt{y} \right) \right) dt$$

$$I_2 = \frac{1}{(1-y)} \left(\left(2 - y - \frac{1}{y} \right) \log(1 - u\sqrt{y}) - (1-y) \frac{u}{\sqrt{y}} \right)$$

$$I_2 = \frac{1}{(1-y)} \left(\left(\frac{(1-y)^2}{y} \right) \log \frac{1}{1-u\sqrt{y}} - (1-y) \frac{u}{\sqrt{y}} \right).$$

With improved notation, $v = u\sqrt{y}$ from (14) follows:

$$I_2 = \frac{1-y}{y} \log\left(\frac{1}{1-v} \right) - \frac{v}{y}.$$

Then

$$C_\alpha = I_1 + I_2 = \log w + \frac{1-y}{y} \log\left(\frac{1}{1-v} \right) - \frac{v}{y}.$$

which is desired result.

With the return to the original substitution, $\alpha = \sigma^2$, the capacity formula in terms of the signal-to-noise ratio is

$$C_o = \int_0^{\infty} \log\left(1 + \frac{x}{\sigma^2} \right) dF_y(x) = C_\alpha - \log(\sigma^2).$$

Then, the final result takes the form of (6)–(8). **Q.E.D.**

ACKNOWLEDGMENT

The authors wish to thank Prof. S. Verdu for his valuable comments and support during preparation of this paper.

REFERENCES

[1] A. D. Wyner, "Shannon-theoretic approach to a Gaussian cellular multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1713–1727, Nov. 1994.

[2] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1293–1302, July 1993.

[3] D. Jonsson, "Some limit theorems for the eigenvalues of a sample covariance matrix," *J. Multivariate Anal.*, vol. 12, pp. 1–38, 1982.

[4] A. J. Grant and P. D. Alexander, "Random sequence multisets for synchronous for code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2832–2836, Nov. 1998.

[5] P. B. Rapajic, "Information capacity of the space division multiple access mobile communication system," *Wireless Pers. Commun.*, vol. 11, no. 1, pp. 131–159, Oct. 1999.

[6] P. B. Rapajic and D. Popescu, "Derivation of the closed form information capacity equation of the random signature multiple-input multiple-output Gaussian channel," in *Proc. 1999 IEEE Information Theory Workshop*, Kruger National Park, South Africa, June 20–25, 1999, p. 96.

[7] S. Verdu, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[8] S. Verdu and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 622–640, Mar. 1999.

[9] S. Verdu, "Capacity region of Gaussian CDMA channels: The symbol-synchronous case," in *Proc. 24th Allerton Conf.*, Oct. 1986, pp. 1025–1034.

Capacity Analysis of NOMA With mmWave Massive MIMO Systems

Di Zhang, *Student Member, IEEE*, Zhenyu Zhou, *Member, IEEE*, Chen Xu, *Member, IEEE*, Yan Zhang, *Senior Member, IEEE*, Jonathan Rodriguez, *Senior Member, IEEE*, and Takuro Sato, *Fellow, IEEE*

***Abstract*— Non-orthogonal multiple access (NOMA), millimeter wave (mmWave), and massive multiple-input-multiple-output (MIMO) have been emerging as key technologies for fifth generation mobile communications. However, less studies have been done on combining the three technologies into the converged systems. In addition, how many capacity improvements can be achieved via this combination remains unclear. In this paper, we provide an in-depth capacity analysis for the integrated NOMA-mmWave-massive-MIMO systems. First, a simplified mmWave channel model is introduced by extending the uniform random single-path model with angle of arrival. Afterward, we divide the capacity analysis into the low signal to noise ratio (SNR) and high-SNR regimes based on the dominant factors of signal to interference plus noise ratio. In the noise-dominated low-SNR regime, the capacity analysis is derived by the deterministic equivalent method with the Stieltjes–Shannon transform. In contrast, the statistic and eigenvalue distribution tools are invoked for the capacity analysis in the interference-dominated high-SNR regime. The exact capacity expression and the low-complexity asymptotic capacity expression are derived based on the probability distribution function of the channel eigenvalue. Finally, simulation results validate the theoretical analysis and demonstrate that significant capacity improvements can be achieved by the integrated NOMA-mmWave-massive-MIMO systems.**

***Index Terms*— mmWave, NOMA, massive MIMO, capacity analysis, Stieltjes and Shannon transform, statistics and probability analysis.**

Manuscript received October 15, 2016; revised February 21, 2017; accepted March 12, 2017. Date of publication April 27, 2017; date of current version June 19, 2017. This work was supported in part by the National Science Foundation of China under Grant 61601180 and Grant 61601181, in part by the Fundamental Research Funds for the Central Universities under Grant 2016MS17 and Grant 2017MS13, in part by the Beijing Natural Science Foundation under Grant 4174104, in part by the Beijing Outstanding Young Talent under Grant 2016000020124G081, and in part by the projects 240079/F20 funded by the Research Council of Norway. *(Corresponding author: Zhenyu Zhou.)*

D. Zhang is with the State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China, and also with GITS/GITI, Waseda University, Tokyo 169-0072, Japan (e-mail: di_zhang@fuji.waseda.jp).

Z. Zhou and C. Xu are with the State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China (e-mail: zhenyu_zhou@ncepu.edu.cn; chen.xu@ncepu.edu.cn).

Y. Zhang is with the Department of Informatics, University of Oslo, N-0373 Oslo, Norway, and also with the Simula Research Laboratory, Fornebu, Norway (e-mail: yanzhang@ieee.org).

J. Rodriguez is with the Institutõ de Telecomunicações, 3810-193 Aveiro, Portugal, and also with the University of South Wales, Pontypridd CF37 1DL, U.K. (e-mail: jonathan@av.it.pt).

T. Sato is with GITS/GITI, Waseda University, Tokyo 169-0072, Japan (email: t-sato@waseda.jp).

Digital Object Identifier 10.1109/JSAC.2017.2699059

I. Introduction

WITH even higher transmission rate claimed by the fifth generation (5G) wireless communications, spectrum efficiency (SE) [1]–[4] and energy efficiency (EE) [5], [6] are categorized as two main topics of the study. In which, millimeter wave (mmWave) [1], non-orthogonal multiple access (NOMA) [3], [4], and massive multi-input-multi-output (massive MIMO, also known as large-scale antenna system [7], large-scale MIMO [8]–[10]) [2] are noticed a lot both in academia and industry. MmWave refers to the frequency with $30 \sim 300$ GHz [11]. With shorter propagation distance and even higher frequency, the propagation characteristics are normally different from existing macro waves in use. In this regard, the propagation characteristics and channel model were intensively investigated at the incipient stage of mmWave studies [1], [12].

Other than the mmWave, NOMA was proposed to alleviate the spectrum bottleneck by invoking the superposition coding of multiple users in the same frequency, thereby, enhancing the systems' SE performance [13]. In NOMA studies, transmit power values amongst users are exploited to separate signals belonging to different users. Research topics such as beamforming design [14], user pairing [15], and power allocation [16], etc., were intensively investigated.

On the other hand, massive MIMO was introduced as well to tumble down the 5G's SE and EE requirement toughies [17]. The benefit of massive MIMO is that, with even larger antenna number, thermal noise and fast fading effects can be averaged out [2]. Besides the studies on SE issue, prior studies on massive MIMO's EE issue mostly focused on the effective engaged component selection method design, energy harvesting and content sharing technologies and their optimization methods, such as the work in [18]–[21].

A. Related Work

The related antecedent work is summarized as follows. For mmWave communications, the channel models were characterized and analyzed in [1], [12], and [22]. In [1], the angle of departure (AoD), angel of arrive (AoA), and channel gain characteristics of mmWave channels were estimated for both light of sight (LOS) and non light of sight (NLOS) paths to obtain a general channel model. In [12], a structured compressive sensing (SCS)-based channel estimation scheme was proposed. In which, the angular sparsity was employed

as well to reduce the required pilot overhead. In [22], the mmWave channel was estimated with mmWave band (from 28 to 73 GHz) propagation characteristics. Besides, the cellular capacity was evaluated based on the experiment data collected from New York City. The simplified uniform random single-path (UR-SP) model was adopted for optimal beamforming design in mmWave communications [23]. Alkhateeb *et al.* [24], investigated the Kronecker channel model and proposed a hybrid pre-coding method for mmWave communications.

For the aspect of integrating NOMA with massive MIMO, a joint antenna selection and user scheduling algorithm was proposed in [25]. Numerical results showed that the proposed algorithm achieved better search efficiency in single-band two-user scenario. The simplified and limited feedback scenario for NOMA-MIMO systems was studied in [26]; where the NOMA-MIMO channel was decomposed into multiple NOMA-SISO channels. In the study of [27], the outage probability was investigated for NOMA-massive-MIMO systems. To integrate the massive MIMO with mmWave, the majority prior work focused on the beamforming design. For example, an interference-aware (IA) beam selection scheme was proposed by [28]; it can achieve near-optimal sum rate with better EE performance compared with conventional schemes. However, fewer studies have been done on NOMA-mmWave or on the integration of mmWave, NOMA and massive MIMO. From an intuitionistic perspective, by integrating these three, better SE and sum rate performances can be obtained. However, by what scale this increment can be is still ambiguous.

For the capacity analysis, various studies have been done before. For instance, the reservation-based random access wireless network capacity was investigated in [29] with addressed upper and lower bound expressions. Recently, the random matrix theory (RMT) tools are intensively noticed and have been vastly used to tame the performance of massive MIMO systems. The majority work of this focused on the closed-form expressions for critical parameter analysis, such as the ergodic capacity, higher-order capacity moments, and outage probability [27], [30]–[36]. Among the various mathematical tools provided by RMT, the deterministic equivalent method was introduced by [30], [31], [37], and [38]. In which, the Stieltjes-Shannon transform method [30]–[32], Gaussian method [34], and free probability theory [33] play critical roles. On the other hand, the statistics and probability analysis method was widely applied for capacity [35] and outage probability analysis [27] in massive MIMO systems. But still, for the more complicated NOMA-mmWave-massive-MIMO systems, these methods cannot be adopted directly, especially with the NOMA decoding scheme. Low-complexity asymptotic method is of great importance in this case.

B. Contributions

The aforementioned work plays vital role and lies solid foundation for the study of mmWave, NOMA and massive MIMO. In this paper, we step further to study the integrated NOMA-mmWave-massive-MIMO systems and provide a theoretical analysis on the achievable capacity. The main contributions of this paper are summarized as follows:

- The model of the integrated NOMA-mmWave-massive-MIMO systems is systematically introduced. To settle down the intractable characteristics with mmWave channel, a simplified mmWave channel model is introduced by extending the UR-SP model with AoA.
- For the capacity analysis, it is divided into the low signal to noise ratio (SNR) and high-SNR regimes to simplify the analysis. In the noise-dominated low-SNR regime, the capacity analysis is derived by the deterministic equivalent method with the Stieltjes-Shannon transform. During the analytical process, we provide mathematical proofs for the relationship between the Stieltjes transform and the Shannon transform. In the interference-dominated high-SNR regime, the deterministic equivalent method is no longer valid. In this regard, the exact capacity expression as well as a low-complexity special case (the numbers of paths, antennas, and user terminals are equal) expression are derived with the statistic and eigenvalue distribution tools.
- We evaluate the derived capacity expressions under both low-SNR and high-SNR regimes, and investigate the impacts of the numbers of LOS paths, antennas, and user terminals on the system performance. In the low-SNR regime, it is found that SNR and user number have positive correlations with the systems' capacity performance. This significantly outperforms the existing long term evolution (LTE) systems especially under the cell-edge scenario. In the high-SNR regime, numerical results manifest the matching relationship between the asymptotic PDF expression and the exact PDF expression. In addition, we find that the number of LOS paths has positive but ignorable effect to the capacity increment.

C. Organizations

The rest of this paper is organized as follows. The systems' model as well as the channel model are introduced by section II. The capacity analysis in the low-SNR regime is investigated by section III. Afterwards, section IV provides the capacity analysis for the high-SNR regime. The numerical results are given in section V. The main results and discussions are provided by section VI. All of the mathematical proofs are given by the Appendices.

D. Notations

Throughout the paper, the uppercase boldface letters, lowercase boldface letters, and normal letters are used to represent the matrix, vector, and scalar quantity, respectively. Furthermore, $\mathbb{C}$ and $\mathbb{R}$ denote the sets of complex and real numbers, respectively. $\mathbf{A}^H$ denotes the Hermitian transposition of a matrix $\mathbf{A}$. $\mathbf{A}_{i,j}$ is the (i, j)-th entry of a matrix $\mathbf{A}$ with the i-th row and j-th column. Additionally, $tr(\mathbf{A})$, $\det(\mathbf{A})$, and $\mathbb{E}(\mathbf{A})$ denote the trace, determinant, and expectation of the matrix $\mathbf{A}$, respectively. Moreover, $\mathbf{A}^{-1}$ is the inverse transpose of matrix $\mathbf{A}$. Finally, inf and sup are used to denote the infimum and supremum.

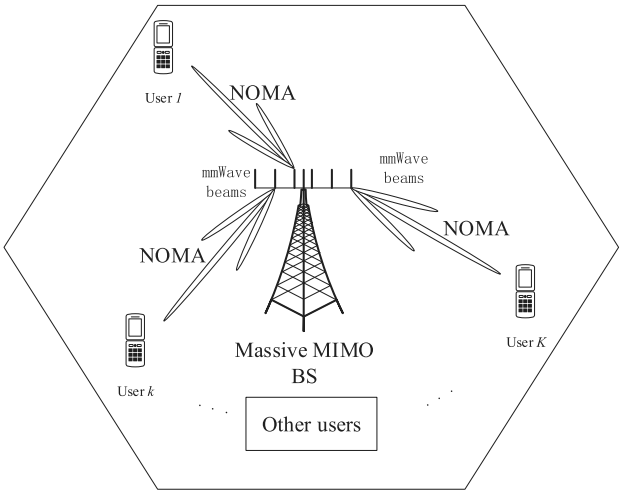

Fig. 1. The conceptual model of the NOMA-mmWave-massive-MIMO systems.

II. System Model

In this section, firstly, the NOMA-mmWave-massive-MIMO systems model in the downlink is elaborated. We further propose a simplified mmWave channel model by extending the UR-SP model with AoA and express it in matrix form.

A. The NOMA-mmWave-Massive-MIMO Systems' Model

The NOMA-mmWave-massive-MIMO systems' model in the downlink that focused by this study is shown in Fig. 1. It consists of a massive MIMO BS that serving less but multiple user terminals (with number K). As shown here, the transmission is carried out in the mmWave frequency with mmWave beams. In addition, within each beam, the NOMA encoding scheme is utilized to encode the transmit signal.

With NOMA encoding scheme, the same spectrum resource block [13] is shared by multiple users within the same user group (it is assumed that the NOMA users within each frequency resource block is one user group). Among different user groups, orthogonal frequency correlations are assumed to isolate the inter-channel interference. The optimal power allocation problem of NOMA-MIMO has been investigated by prior studies [15], [39]. In this paper, we focus on the capacity analysis of the proposed NOMA-mmWave-massive-MIMO systems by assuming that power value is different amongst different users with regards to the NOMA concept [27]. This is due to the fact that the optimal power allocation study based on the scenario that the transmission rate requirement of each user is given beforehand [15], [39]. In addition, this capacity expression can also be applied for the optimal power allocation scenario while giving the transmission rate requirement of each user and the component carrier (CC) bandwidth. The optimal power allocation study can be done in future based on the NOMA-mmWave-massive-MIMO system s' model. At the receiver side, user can make use of SIC [27], [40] to remove the interferences from other users with higher orders. The remaining information from low order users is treated as interference.[1] With perfect orthogonal characteristics among channels of different user groups, the inter-channel interference caused by users in different groups can be ignored. Thus co-channel interference is mainly from users in the same group with a lower order.

With mmWave frequency in hand, much wider bandwidth can be allocated compared with macro wave frequency used by LTE and prior generations. It was estimated that the CC bandwidth can be up to 1 GHz or even more with mmWave [43]. In line with Shannon theory for achievable transmission rate, with better channel condition, wider CC bandwidth will yield faster rate. The 5G's claiming rate can be easily met with mmWave in this regard, albeit the specific frequency allocation and usage method of mmWave in 5G is still on discussion with international telecommunications union-radio communication (ITU-R).

Assuming the NOMA power allocation for each user as $P_i, i \in [1, K]$, the received signal is given by

$$\mathbf{y} = \mathbf{H}^H \mathbf{x} + \mathbf{n} = \sum_{i=1}^{K} P_i \mathbf{H}^H \mathbf{s} + \mathbf{n} \tag{1}$$

where $\mathbf{H} \in \mathbb{C}^{N \times K}$ is the channel model from N transmit antennas to K user terminals, $\mathbf{x} \in \mathbb{C}^{N \times 1}$ is the transmit information at the transmit side, which consists of the transmit signal $\mathbf{s}_i$ as well as the transmit power P_i. In addition, $\mathbf{n} \in \mathbb{C}^{K \times 1}$ yields the additive white Gaussian noise (AWGN). Moreover, without loss of generality, it is assumed that $N \geq K$. This is due to the fact that the transmit antenna number is usually larger than the receive antenna number in massive MIMO systems. It is also assumed that the transmit signal $\mathbf{s}$ is normalized, which means that each column of $\mathbf{s}$ obeys $\mathbb{E}[\mathbf{s}_i] = 0$, and $\mathbb{E}[\mathbf{s}_i \mathbf{s}_i^H] = 1$.

With this model in hand, to determinate the capacity performance of NOMA-mmWave-massive-MIMO systems, the channel model should be set forth. Given a constant normalized noise value within the channel assumption, capacity performance is largely determined by the allocated power to each user and the channel model [2].

B. The Proposed MmWave Channel Model

In line with prior studies from [12], [44], the mmWave channel model with a three dimensional (3D) transmission background has to take into consideration the channel gain, the AoD at the transmitter, and the AoA at the receiver. Taking an example, the mmWave channel response for the k-th user can be given as

$$\mathbf{h}_k = \sqrt{N} \left\{ \frac{\beta_k^0 \mathbf{d}(\theta_k^0) \mathbf{a}(\phi_k^0)}{\sqrt{1 + d_k^{\beta_k^0}}} + \sum_{i=1}^{M} \frac{\beta_k^i \mathbf{d}(\theta_k^i) \mathbf{a}(\phi_k^i)}{\sqrt{1 + d_k^{\beta_k^i}}} \right\}, \tag{2}$$

where $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K]$. Besides, integer $k \in [1, K]$ is the user index, and integer $i \in [1, M]$, the NLOS path index. β_k^i denotes the channel gain of user k for the i-th NLOS

[1] Note that NOMA decoding order can either with regard to the user orders [41], or with a reversed order with regard to the SNR [42]. Here in this study, we focus on the first scheme.

path, which can be assumed to obey a complex Gaussian distribution. Similarly, β_k^0 is the channel gain of the LOS path. M is the total number of NLOS paths. d_k denotes the distance between the BS and the k-th user. θ represents the normalized AoD of each path (by LOS or NLOS), which follows

$$\mathbf{d}(\theta) = \frac{1}{\sqrt{N}}[1, e^{-j\pi\theta}, \cdots, e^{-j\pi(N-1)\theta}]. \tag{3}$$

Similarly, the normalized AoA of each path follows

$$\mathbf{a}(\phi) = \frac{1}{\sqrt{N}}[1, e^{-j\pi\phi}, \cdots, e^{-j\pi(K-1)\phi}]^T. \tag{4}$$

However, according to prior findings in [46] and [47], mmWave transmission is highly susceptible to obstructions due to its vulnerable characteristic to diffraction and path loss. Thus mmWave beam highly relays on the LOS paths while applying in the wireless communications. With those antecedent studies in hand, by ignoring the NLOS components in (2) and assuming L LOS paths, a simplified model according to the UR-SP channel model [23], mmWave channel response for k-th user can be re-elaborated by

$$\mathbf{h}_k = \sqrt{N}\sum_{i=1}^{L}\frac{\beta_k^i \mathbf{d}(\theta_k^i)}{\sqrt{1+d_k^{\beta_k^i}}}. \tag{5}$$

Although the array steering vector (AoD vector) is included in the UR-SP channel model, the AoA factor is neglected in this model. Thus, in this study, we further extend the UR-SP model by taking the AoA vector into consideration. This results the channel response for the k-th user as

$$\mathbf{h}_k = \sqrt{N}\sum_{i=1}^{L}\frac{\mathbf{d}(\theta_k^i)\beta_k^i \mathbf{a}(\phi_k^i)}{\sqrt{1+d_k^{\beta_k^i}}}. \tag{6}$$

Additionally, by ignoring the difference bringing in by the shape of transmitter and receiver with an correlation free case both at transmit and receiver sides, the channel model $\mathbf{H}$ in high-dimensional matrix form can be given as

$$\mathbf{H} = \mathbf{DBA}, \tag{7}$$

where $\mathbf{D} \in \mathbb{C}^{N\times L}, \mathbf{B} \in \mathbb{C}^{L\times L}, \mathbf{A} \in \mathbb{C}^{L\times K}$. $\mathbf{B} = \eta\boldsymbol{\beta}$ with $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_L]$, and $\eta = [\eta_1, \ldots, \eta_k, \ldots, \eta_K]$. $\mathbf{D} = [\mathbf{d}(\theta_0), \ldots, \mathbf{d}(\theta_{L-1})]$, $\mathbf{A} = [\mathbf{a}(\phi_0), \ldots, \mathbf{a}(\phi_{L-1})]^T$. Here η_k is the coefficient given as

$$\eta_k = \frac{\sqrt{N}}{\sqrt{1+d_k^{\beta_k^i}}}. \tag{8}$$

Here it is further assumed that the distance differences amongst different users can be absorbed into the β_k^i effects. This is because that with randomly generated $d_k^{\beta_k^i}$, η_k as an coefficient can be denoted with the randomly generated β_k^i in the analysis.

III. Capacity Analysis in the Noise-Dominated Low-SNR Regime

It is still intractable to directly analyze the capacity of the integrated NOMA-mmWave-massive-MIMO systems even with the simplified channel model due to the prohibited analysis complexity. That is, on the one hand, SIC is employed in NOMA to remove the interference and detect the desired signal. On the other hand, the mmWave channel model with multiple paths makes the analysis even tough. Thus, we divide the capacity analysis into low-SNR and high-SNR regimes, which can be adapted for various application scenarios such as cell edge, cell center, and etc. In addition, by summing up the two regimes, the majority conditions of cellular area can be covered.

In the low-SNR regime, the impact of the co-channel interference is trivial, and the dominant factor to each user's SINR value is the noise. In comparison, the dominant factor of SINR in the high-SNR regime is the co-channel interferences from other users. In this section, the capacity in the low-SNR regime is analyzed first by the deterministic equivalent method with the Stieltjes-Shannon transform [31], [47].

In the noise-dominated low-SNR regime, it is assumed that $\mathbf{D}$, $\mathbf{A}$ are the diagonal matrices with optimal beamforming of the NOMA-mmWave-massive-MIMO systems. This is a reasonable assumption due to the fact that only the direct beam targeting on the desired user can be the effective beam for transmission. All of the other AoD beams have no actual contribution for user k's transmission. In the low-SNR regime, the interference can be neglected compared to the noise power. The Stieltjes transform $S_{\mathbf{B}_N}(z)$ of matrix $\mathbf{B}_N$ can be employed for the capacity analysis in low-SNR regime,[2] which is defined as

$$\begin{aligned} S_{\mathbf{B}_N}(z) &= \frac{1}{N}tr(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \\ &= \int \frac{1}{\lambda - z} dF_{\mathbf{B}_N}(\lambda) \\ &\overset{a.s.}{\to} \int \frac{1}{\lambda - z} dF_N(\lambda). \end{aligned} \tag{9}$$

In which the Hermitian non-negative definite matrix $\mathbf{B}_N$ is defined as

$$\mathbf{B}_N = \mathbf{HH}^H = \mathbf{DBA}^2\mathbf{B}^H\mathbf{D}^H. \tag{10}$$

$z \in \mathbb{C} - \mathbb{R}^+ \equiv \{z \in \mathbb{C}, \Im(z) > 0\}$, and $\mathbf{I}_N$ is an identity matrix. In addition, $F_{\mathbf{B}_N}(\lambda)$ is the eigenvalue empirical distribution function (EDF) of $\mathbf{B}_N$. With N and K growing large, $F_{\mathbf{B}_N}(\lambda)$ converges to $F_N(\lambda)$ (the determinant eigenvalue CDF of $\mathbf{B}_N$) with probability 1 according to the Glivenko-Cantelli theorem [48].

The importance of the Stieltjes transform lies in its link to the Shannon transform $\mathcal{V}_{\mathbf{B}_N}(z)$ of $\mathbf{B}_N$, where the Shannon transform is directly linked with the capacity expression, which can be derived from the mutual information analysis in MIMO systems [37], [49].

[2]Note that here in this study, it is assumed that each user equipped with one receive antenna.

Theorem 1: The relationship between Stieltjes transform and Shannon transform of $\mathbf{B}_N$ can be given as

$$\begin{aligned}\mathcal{V}_{\mathbf{B}_N}(z) &= \int_0^{+\infty} \log\left(1+\frac{\lambda}{z}\right) dF_N(\lambda) \\ &= \int_z^{+\infty} \left(\frac{1}{w} - S_{\mathbf{B}_N}(-w)\right) dw. \end{aligned} \tag{11}$$

Proof: Please see Appendix A. ∎

By assuming perfect channel state information at the receiver side (CSIR), the mutual information can be described by

$$I_{\mathbf{G}_N}(\sigma^2) = \mathbb{E}\left\{\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{B}_N\right)\right\}. \tag{12}$$

In addition, the relationship between the Shannon transform and the ergodic mutual information is

$$I_{\mathbf{G}_N}(\sigma^2) = N\mathcal{V}_{\mathbf{B}_N}(\sigma^2). \tag{13}$$

Based on Theorem 1, the ergodic mutual information can be further determined by the Stieltjes transform on condition that $F_{\mathbf{B}}(\lambda) \to F_N(\lambda)$, which will be discussed in the following analysis. Before delving into detail analysis, a lemma and a hypothesis are given to clarify the constraints of this approach that used in this study.

Lemma 1: The sequences of $F_{\mathbf{D}}$, $F_{\mathbf{B}}$ and $F_{\mathbf{A}}$ (EDF of matrix $\mathbf{D}, \mathbf{B}, \mathbf{A}$) are tight, where $\mathbf{D}$ and $\mathbf{A}$ are the diagonal matrices as claimed before. Additionally, $\mathbf{B}$ is the random matrix with i.i.d. Gaussian entries of zero mean and covariance $\frac{1}{L}$.

Proof: Please see Appendix B. ∎

The following hypothesis holds as well:

1) By defining $c = \frac{N}{K}$ and assuming $0 < a < b < \infty$, we have the following inequalities

$$a < \min_K \liminf_N c < \max_K \limsup_N c < b. \tag{14}$$

This hypothesis is to claim that the value of c has infimum and supremum with regards to K, N. This is reasonable by assuming that c is constant within the region $[0, +\infty]$. On this point, interested readers can refer to [31] and [32] and the references therein.

With all of those in hand, the mutual information can be straightforwardly obtained based on the Stieltjes transform of the channel matrix $\mathbf{B}_N$, which can be solved by using the link between $F_{\mathbf{B}_N}(\lambda)$ and $F_N(\lambda)$ of $\mathbf{B}_N$. Thus, the main problem is to find such a $F_N(\lambda)$. It is proved by prior studies [31], [47] that the difference between $F_{\mathbf{B}_N}(\lambda)$ and $F_N(\lambda)$ converges vaguely to zero:

$$S_{\mathbf{B}_N}(z) - S_N(z) \overset{a.s.}{\to} 0, \quad \text{for } z \in \mathbb{C} - \mathbb{R}^+, \tag{15}$$

where

$$S_N(z) \equiv \int_{\mathbb{R}^+} \frac{1}{\lambda - z} dF_N(\lambda) \tag{16}$$

By following the study in [50], given the noise variance σ^2 and the power matrix of each user $\mathbf{P}_k$, the deterministic equivalent of mutual information can be derived by using Lemma 1 and the hypothesis defined in (14) as

$$I(\sigma^2) = \bigcup_{\substack{\frac{1}{N}tr\mathbf{P}_k \le P_k, \\ \mathbf{P}_k \ge 0, \\ k\in\mathcal{S}}} \left\{\sum_{k\in\mathcal{S}} C_k \le \mathbb{E}\left\{\mathcal{V}_N(\mathbf{P}_k;\sigma^2)\right\}\right\}, \tag{17}$$

here $\mathcal{S} = \{1, \ldots, K\}$, C_k is the capacity of the k-th user. The Shannon transform is given by

$$\mathcal{V}_N(\mathbf{P}_k;\sigma^2) \overset{\text{def.}}{=} \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\sum_{k\in\mathcal{S}}\mathbf{H}_k\mathbf{P}_k\mathbf{H}_k^H\right). \tag{18}$$

In this case, by assuming α a constant value, i.e., $0 < \alpha < \infty$, the spectral norm satisfies

$$\max\{\|\mathbf{D}\|, \|\mathbf{A}\|, \|\mathbf{H}\mathbf{H}^H\|\} \le \alpha. \tag{19}$$

By following the prior studies in [31], [34], the deterministic expression of the ergodic capacity can be given as

$$\begin{aligned} C \;&\le\; \sum_{k=1}^{K}\mathbb{E}\left\{\mathcal{V}_N(\mathbf{P}_k;\sigma^2)\right\} \\ &\overset{\text{a. s.}}{\to} \frac{1}{N}\sum_{k=1}^{K}\log\det(\mathbf{I} + ce_k(-\sigma^2)\mathbf{A}_k^2\mathbf{P}_k) \\ &\quad + \frac{1}{N}\log\det(\mathbf{I} + \sum_{k=1}^{K} f_k(-\sigma^2)\mathbf{D}_k^2) \\ &\quad - \sigma^2\sum_{k=1}^{K} f_k(-\sigma^2)e_k(-\sigma^2). \end{aligned} \tag{20}$$

where $e_k(-\sigma^2)$ and $f_k(-\sigma^2)$ are the unique positive solutions of the following symmetric equalities

$$e_k(-\sigma^2) = \frac{1}{N}tr\mathbf{D}_k^2(\sigma^2[\mathbf{I} + \sum_{k=1}^{K} f_k(-\sigma^2)\mathbf{D}_k^2])^{-1}, \tag{21}$$

$$f_k(-\sigma^2) = tr\mathbf{A}_k\mathbf{P}_k\mathbf{A}_k(\sigma^2[\mathbf{I} + ce_k(-\sigma^2)\mathbf{A}_k\mathbf{P}_k\mathbf{A}_k])^{-1}. \tag{22}$$

The sum rate supremum of the NOMA-mmWave-massive-MIMO systems can be addressed by (20) with $e_k(-\sigma^2)$ and $f_k(-\sigma^2)$ the unique solutions of the equalities given in (21) and (22), where the iterative algorithm to obtain these solutions can be found by [32], [34], and [52].

As discussed before, this deterministic equivalent method is only valid in the low-SNR scenario. In the following analysis, we further focus the study in high SNR regime. In that case, the interferences mostly come from neighboring users within the user group. Then after SIC, prior deterministic equivalent method with Shannon-Stieltjes transform is not valid for analysis. This is because that with Shannon-Stieltjes transform, it is assumed that for each user, only the channel noise is existing. Thus given the transmission and channel noise power values, only the channel matrix is the determinant variable for achievable capacity. The high-SNR capacity analysis is addressed in the following section. Finally by summarizing these two regimes, we can approach the majority scenarios in integrating NOMA-mmWave-massive-MIMO systems. The

comprehensive closed-form capacity expression in the existence of co-channel interference and channel noise is way complex, which is left for further study.

IV. Capacity Analysis in the Interference-Dominated High-SNR Regime

The high-SNR regime is investigated in this section. To surround the systems' capacity in high-SNR regime, an alternative method with statistics and probability analysis is adopted based on the channel distribution analysis. Firstly, by employing the SIC [41] to perfectly cancel the co-channel interferences with higher orders, the SINR of user k can be given as

$$SINR_k \Leftrightarrow \frac{P_k \mathbf{H}\mathbf{H}^H}{\sum_{k'=1,k'\neq k}^{K} P_{k'}\mathbf{H}\mathbf{H}^H + \sigma^2} \overset{SIC}{\Leftrightarrow} \frac{P_k \mathbf{H}\mathbf{H}^H}{\sum_{k'=1}^{k-1} P_{k'}\mathbf{H}\mathbf{H}^H + \sigma^2}. \tag{23}$$

Here the first expression is logically defined as the received SINR for each user k without the SIC. After the SIC, the second equality can be straightforwardly arrived. The capacity of the NOMA-mmWave-massive-MIMO systems within high-SNR regime then, can be approximated as

$$C = \sum_{k=1}^{K} \mathbb{E}\left\{\log\det\left(\mathbf{I}_N + \frac{P_k \mathbf{H}\mathbf{H}^H}{\sum_{k'=1}^{k-1} P_{k'}\mathbf{H}\mathbf{H}^H + \sigma^2}\right)\right\} \overset{high\ SNR}{\approx} \sum_{k=1}^{K} \mathbb{E}\left\{\log\det\left(\mathbf{I}_N + \frac{P_k \mathbf{H}\mathbf{H}^H}{\sum_{k'=1}^{k-1} P_{k'}\mathbf{H}\mathbf{H}^H}\right)\right\}. \tag{24}$$

In the following, a tractable capacity expression is derived by employing the tools of statistics and probability analysis method [35]. First of all, the capacity expression can be divided into the power allocation part and channel characteristic part by Theorem 4.1.

Theorem 2: The ergodic capacity of NOMA-mmWave-massive-MIMO systems in high-SNR regime is

$$C = \frac{1}{\ln 2}\sum_{k=1}^{K}\left\{\ln\left(\frac{\sum_{k'=1}^{k} P_{k'}}{\sum_{k'=1}^{k} P_{k'} - P_k}\right) + \ln\int_0^{+\infty} \lambda f(\lambda) d\lambda\right\}, \tag{25}$$

where λ is the eigenvalue of $\mathbf{H}\mathbf{H}^H$, and $f(\lambda)$ is the PDF of λ.

Proof: Please see Appendix C. ∎

As shown, the first part of Theorem 2 is about the power ratio with NOMA scheme, where the second part is about the eigenvalue and its PDF of $\mathbf{H}\mathbf{H}^H$. Once the NOMA power allocation is given, the first part will be determined, and the capacity mainly depends on the second part with $f(\lambda)$. The exact expression of $f(\lambda)$ is pursued and given by the following lemma 2.

Lemma 2: By exploring the knowledge of probability analysis, the unconditional PDF of $\mathbf{H}\mathbf{H}^H$ can be calculated as

$$f(\lambda) = \frac{1}{\prod_{i=1}^{L}\Gamma^2(L-i+1)\prod_{i=1}^{L}\Gamma(N-i+1)L} \times \sum_{j=L-K+1}^{L}\sum_{i=1}^{L}(-1)^{i+j}\frac{\lambda^{K-L+j-1}}{\Gamma(K-L+j)}\det(\mathbf{M})N_\lambda(i). \tag{26}$$

where $\mathbf{M}_{i,j}$ is the (i, j)-th minor of matrix $\mathbf{M} \in \mathbb{C}^{L\times L}$, whose entry is given as

$$\mathbf{M}_{i,j} = \Gamma(i+j-1)\Gamma(N-L+j). \tag{27}$$

Additionally, the expression of $N_\lambda(j)$ is

$$N_\lambda(i) = \int_0^{+\infty} 4x^{N+L-2K+i-2}e^{-\frac{\lambda}{x^2}}K_{L-N+i-1}(2x)dx, \tag{28}$$

with $K_m(n)$ is the modified Bessel function of its second kind.

Proof: Please see Appendix D. ∎

The exact capacity expression is acquired by substituting lemma 2 into (25), and $\int_0^{+\infty}\lambda f(\lambda)d\lambda$ is given as (29), shown at the bottom of the next page.

As shown in (25) and (29), although the exact expression of the capacity is derived, but the expression is complex as a non closed-form expression with integral. Fortunately, to obtain the closed-form expression of the eigenvalues' PDF of $\mathbf{H}\mathbf{H}^H$, the study from [52] gives an asymptotic expression under a similar condition. By following the deduction procedure, although it is still difficult to obtain the closed-form expression under condition $N \neq K \neq L$, but when $N = L = K$, the expression is reduced to [35]

$$f(\lambda) = \frac{1}{\pi}\sqrt{g^2(\lambda) + \frac{1}{4\lambda^2 g(\lambda)}}, \tag{30}$$

where $g^2(\lambda)$ is given as [35]

$$g^2(\lambda) = \frac{\sqrt[3]{64^2\lambda^8}(1-i'\sqrt{3})}{384\lambda^4}\sqrt[3]{\frac{-27+\sqrt{27^2-27\frac{16^2}{\lambda}}}{2}} + \frac{\sqrt[3]{64^2\lambda^8}(1+i'\sqrt{3})}{384\lambda^4}\sqrt[3]{\frac{-27-\sqrt{27^2-27\frac{16^2}{\lambda}}}{2}}. \tag{31}$$

Here i' yields the unit imaginary number. On condition that $f(\lambda) \geq 0$, by combining the (30) and (31), we have $\lambda_{max} = \frac{16^2}{27}$. This gives the expression of $f(\lambda)$ as

$$f(\lambda) = \frac{\left(1+i'\sqrt{3}\right)\sqrt[3]{\lambda^8}\sqrt[3]{-\sqrt{729-\frac{6912}{\lambda}}-27}}{24\sqrt[3]{2}\lambda^4} + \frac{\left(1-i'\sqrt{3}\right)\sqrt[3]{\lambda^8}\sqrt[3]{\sqrt{729-\frac{6912}{\lambda}}-27}}{24\sqrt[3]{2}\lambda^4}. \tag{32}$$

Thus by substituting (32) into (25), the ergodic capacity can be rewritten as

$$C = \frac{1}{\ln 2}\sum_{k=1}^{K}\left\{\ln\left(\frac{\sum_{k'=1}^{k} P_{k'}}{\sum_{k'=1}^{k} P_{k'} - P_k}\right) + \ln\int_0^{\frac{16^2}{27}} \lambda f(\lambda) d\lambda\right\}, \quad (33)$$

where $f(\lambda)$ is given by (32).

In summary, the exact capacity expression defined in (25) is well suited for general cases in NOMA-mmWave-massive-MIMO systems despite its high computation complexity. For the special case that $N = L = K$, the asymptotic capacity expression with (32) can be employed, which will be verified through numerical results.

V. Numerical Results

The capacity performance of NOMA-mmWave-massive-MIMO systems is evaluated in this section. The low-SNR regime analysis is evaluated firstly. The matrix $\mathbf{B}$ is randomly generated with zero mean and covariance $\frac{1}{L}$, in line with the hypothesis and assumptions of the prior analysis. With SNR value given by 0 dB and -10 dB (the SNR value is given by averaged over all users in each simulation), the capacity performance is given by Fig. 2. As shown here, with SNR increasing, the achievable capacity is also increased. In addition, with NOMA user number increasing, the capacity difference between 0 dB and -10 dB becomes even greater. This is due to the fact that, in this simulation, it is assumed that the user allocated power value increases with user index increasing. Thus more user yields greater averaged power value (total allocated power divided by engaged user number), which in turns, better capacity performance. It is worth noting that the capacity performance of the NOMA-mmWave-massive-MIMO systems outperforms the existing LTE systems ($0.07 \sim 0.12$ bits/s/Hz of the cell-edge, which yields the low-SNR regime) [53]. For instance, by 10 users and -10 dB, the achievable capacity value is almost 10 times compared with prior LTE systems in low-SNR regime. This is mainly due to the NOMA encoding scheme with multiple users of each frequency resource block, and the power allocation method of this simulation.

To verify the correctness of the PDF deductions in this study, the exact eigenvalue PDF expression in (26) is compared with the asymptotic PDF expression in (32). The results are shown in Fig. 3 and Fig. 4 for $N = L = K = 2$ and $N = L = K = 6$, respectively. By observing Fig. 3 and Fig. 4, it is clear that the asymptotic PDF expression is in good

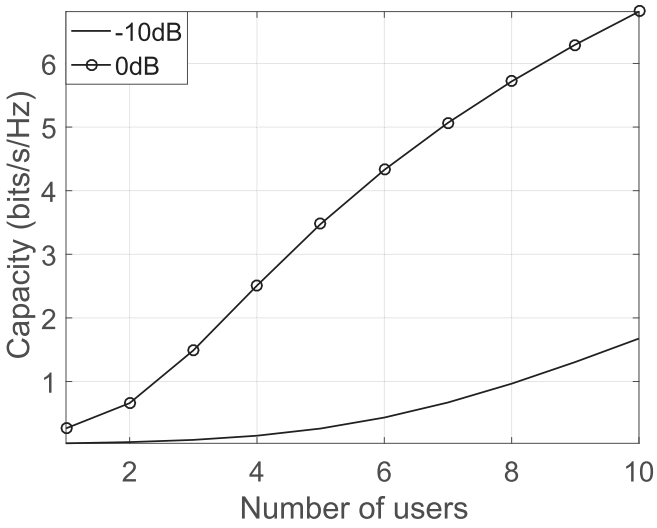

Fig. 2. Capacity performance of the low-SNR scenario. The calculation is based on (20), (21) and (22).

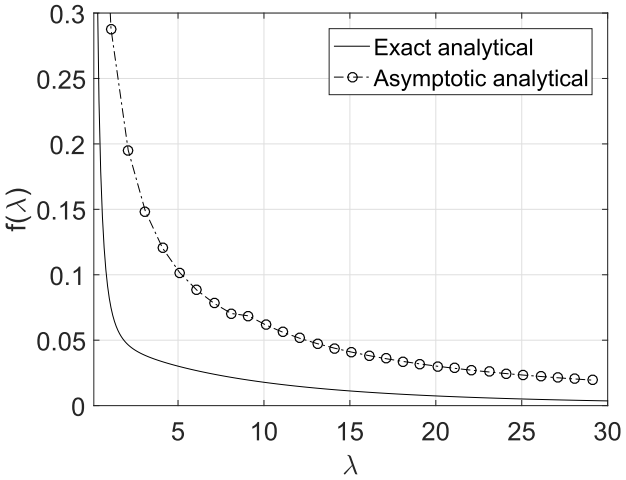

Fig. 3. Analytical comparison of the exact and asymptotic eigenvalue PDFs ($N = L = K = 2$). The exact analytical curve is based on (26), where asymptotic analytical curve is based on (32).

$$\begin{aligned}\int_0^{+\infty} \lambda f(\lambda) d\lambda &= \int_0^{+\infty} \frac{\lambda}{\prod_{i=1}^{L}\Gamma^2(L-i+1)\prod_{i=1}^{L}\Gamma(N-i+1)L} \\ &\quad \times \sum_{j=L-K+1}^{L}\sum_{i=1}^{L}(-1)^{i+j}\frac{\lambda^{K-L+j-1}}{\Gamma(K-L+j)}\det(\mathbf{M})N_\lambda(i)d\lambda \\ &= \frac{1}{\prod_{i=1}^{L}\Gamma^2(L-i+1)\prod_{i=1}^{L}\Gamma(N-i+1)L}\sum_{j=L-K+1}^{L}\sum_{i=1}^{L}(-1)^{i+j} \\ &\quad \times \int_0^{+\infty}\frac{\det(\mathbf{M})\lambda^{K-L+j}}{\Gamma(K-L+j)}N_\lambda(i)d\lambda. \end{aligned} \quad (29)$$

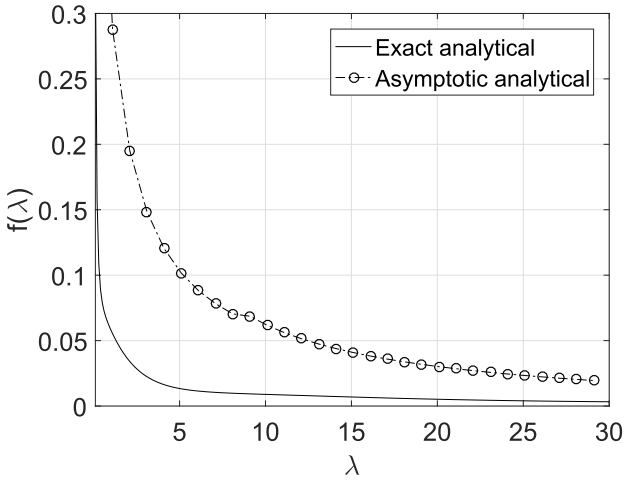

Fig. 4. Comparison of the exact and asymptotic eigenvalue PDFs ($N = L = K = 6$). The exact analytical curve is based on (26), where asymptotic analytical curve is based on (32).

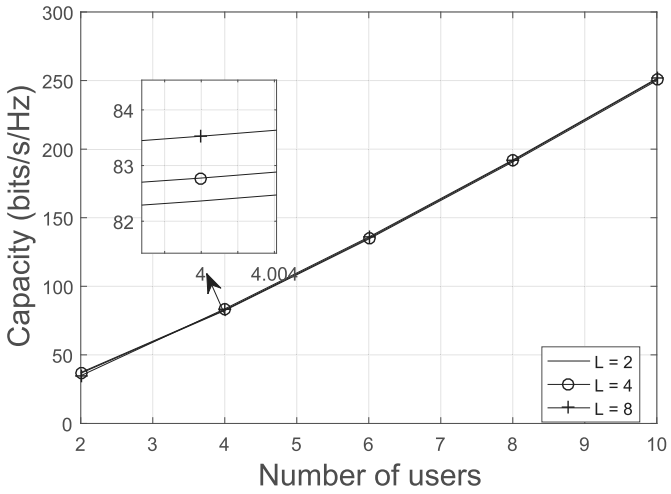

Fig. 6. Capacity performance of high-SNR scenario with the exact eigenvalue PDF ($N = 20$). The calculation is based on (25) and (29).

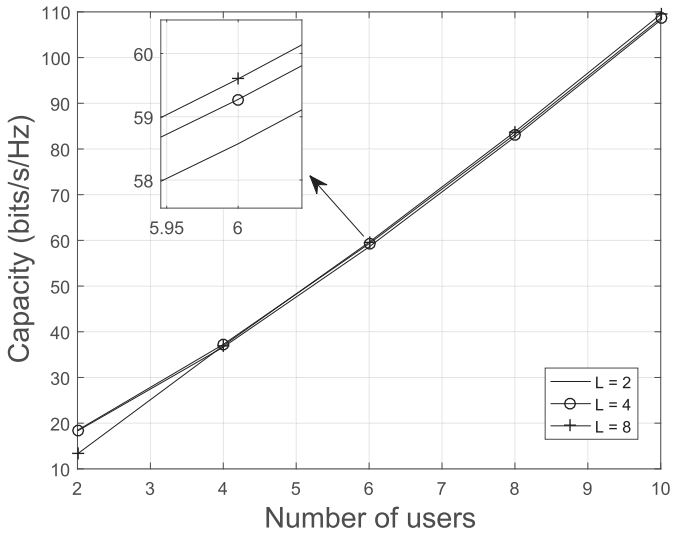

Fig. 5. Capacity performance of high-SNR scenario with the exact eigenvalue PDF ($N = 10$). The calculation is based on (25) and (32).

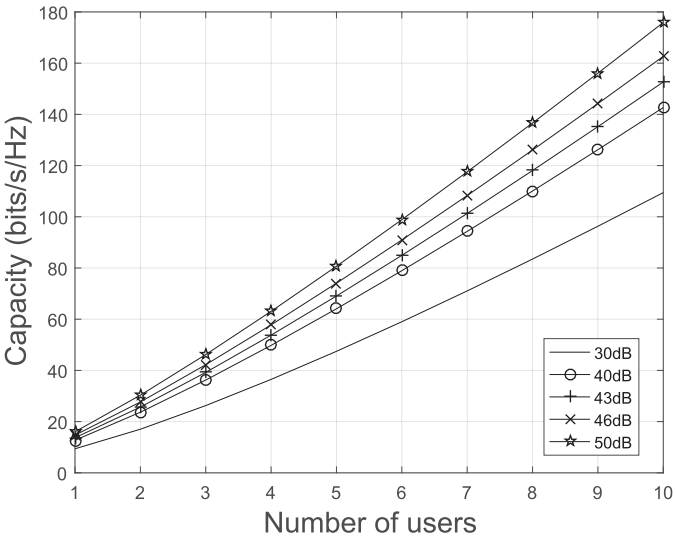

Fig. 7. Capacity performance of high-SNR scenario with the asymptotic eigenvalue PDF. The calculation is based on (33) and (29).

agreement with the exact eigenvalue PDF expression in low and high eigenvalue regions. However, there is a disagreement within other region. In addition, the disagreement grows large with the numbers (N, L, K) growing. Thus the asymptotic PDF is more feasible to adopt with small N, L, K values. The asymptotic expression, albeit results in larger difference with numbers (N, L, K) growing large, but consumes much lesser time while adopting. For instance, to plot the Fig. 4 with $N = L = K = 6$ (Intel Xeon Processor E3-1241 v3, 8 M Cache, 3.50 GHz, 16 G RAM), the consuming time is 4.913563 s with exact PDF expression. In contrast, the asymptotic PDF expression consumes 0.000655 s.

The capacity performances of NOMA-mmWave-massive-MIMO systems with the exact PDF expression are shown in Fig. 5 and Fig. 6 for $N = 10$ and $N = 20$, respectively. The SNR is set to be 30 dB for both simulations. By comparing these two figures, it is clear that with antenna number growing, the capacity performance is enhanced. The reason behind is that more transmit antennas bring in more degree of freedom [4], which is in tune with previous studies [2], [32]. In addition, it is observed that little capacity improvement can be achieved by increasing the number of LOS paths. This is mainly because of the channel hardening effects [54]. Besides, the co-channel interferences from neighboring users and the correlative effects at transmit and receiver sides are enhanced with LOS path number increasing.

Fig. 7 further verifies the relationship between the high-SNR capacity and SNR according to the asymptotic expression. Simulation results show that the high-SNR capacity of the NOMA-mmWave-massive-MIMO systems significantly outperforms the existing LTE systems (with a capacity of 30 bit/s/Hz) [53]. For example, a 240% capacity improvement when $K = 10$, SNR $= 50$ dB. Thus, it is clear that the integration of the three key technologies demonstrates dramatic potential to meet the SE requirement of 5G. On the

other hand, it is shown that although the capacity increases monotonically as the transmit power increasing; the performance improvement becomes saturated when the SNR is sufficiently high. The reason is that co-channel interferences from neighboring users are also increased as the transmit power increasing. Although the capacity improvement can be obtained by just increasing the power, simulation results demonstrate that there is a tradeoff between energy consumption and capacity improvement.

VI. Discussion and Conclusion

The capacity performance of NOMA-mmWave-massive-MIMO systems was investigated in this study. We divided the capacity analysis into the noise-dominated low-SNR regime and the interference-dominated high-SNR regime. The deterministic expressions were given by the analysis for both low-SNR and high-SNR regimes. Additionally, a low-complexity asymptotic capacity expression was given based on the asymptotic PDF of channel eigenvalues. Simulation results indicated that enormous capacity improvement can be achieved compared to the existing LTE systems. We also found that the user number has a strong positive impact on the capacity improvement. This is due to the non-orthogonal user multiplexing in the same frequency resource block enabled by NOMA. In comparison, little capacity improvement can be achieved by increasing the number of LOS paths. Numerical results also revealed that there exists a tradeoff between energy consumption and capacity improvement.

In addition, with much wider bandwidth that provided by mmWave, even higher systems' sum rate increment can be obtained, which yields an attractive perspective for 5G. For instance, as the possible 1 GHz CC bandwidth, under ideal condition, the achievable throughput will be $100 \sim 200$ Gbit/s with the NOMA-mmWave-massive MIMO systems. This integration results in negative effect to the systems' deployment by its denser BS deployment with massive MIMO and vulnerable transmission beams with mmWave. Albeit the deployment of small cell BS with massive MIMO is on the test, but the various obstructions will bring in great challenges for integrating those small cells with mmWave. The SIC encoding equipment of NOMA is another challenge for application at the receiver side. Other than the theoretical analysis in this paper, the optimal power allocation scheme of NOMA-mmWave-massive-MIMO systems can be another interesting topic, which is left for future study.

Appendix A

Proof: To prove the relationship between Shannon transform and Stieltjes transform, some notes should be stated beforehand.

Note 1: use ln for the log base e. For $b > 0$, we have

$$\ln(1+b) = \int_0^1 \frac{b}{1+bt} dt. \tag{34}$$

Note 2: Instead of the distribution function $dF(x)$, for convenience, we use $\rho(x)dx$ as the density. In this case, for $z \to \infty$, say in the upper-half plane, we have

$$\int_0^{+\infty} \rho(\lambda) d\lambda = 1. \tag{35}$$

Firstly, the Shannon transform is defined as

$$\mathcal{V}(x) = \int_0^{+\infty} \log(1+\frac{\lambda}{x}) \rho(\lambda) d\lambda. \tag{36}$$

In this case, the differential result of this equality can be given as

$$\frac{d\mathcal{V}(x)}{dx} = -\frac{1}{\log e} \int_0^{+\infty} \frac{\frac{\lambda}{x^2}\rho(\lambda)}{1+\frac{\lambda}{x}} d\lambda. \tag{37}$$

Furthermore, by multiplying x on both sides, we have

$$\begin{aligned} x\frac{d\mathcal{V}(x)}{dx} &= -\frac{1}{\log e} \int_0^{+\infty} \frac{\lambda\rho(\lambda)}{x+\lambda} d\lambda \\ &= -\frac{1}{\log e} \int_0^{+\infty} \frac{(\lambda+x-x)\rho(\lambda)}{x+\lambda} d\lambda \\ &= -\frac{1}{\log e} \left(1 - x\int_0^{+\infty} \frac{\rho(\lambda)}{x+\lambda} d\lambda\right). \end{aligned} \tag{38}$$

It is noticed that a Stieltjes transform appeared by the last part of the equality's right side. This result gives

$$x\frac{d\mathcal{V}(x)}{dx} = -\frac{1}{\log e}\left(1 - xS(-x)\right). \tag{39}$$

Which is the link between Shannon transforma and Stieltjes transform given by [55] and similar literatures. In addition, it is noticed that in alternative literature (for instance, [31], [47]), the $\log e$ factor is omitted to arrive the equivalence given by Theorem 3.1. Thus by omitting the factor and unpacking the log according to (34), we have,

$$\begin{aligned} \mathcal{V}(x) &\approx \int_0^{+\infty} \rho(\lambda) \int_0^1 \left(\frac{\frac{\lambda}{x}}{1+\frac{\lambda}{x}t} dt\right) d\lambda \\ &= \int_0^{+\infty} \rho(\lambda) \left(\int_0^1 \frac{\lambda}{x+\lambda t} dt\right) d\lambda. \end{aligned} \tag{40}$$

Let $t = \frac{1}{\omega}$, since $t \in [0,1], \omega \in [0,\infty)$, we see that

$$\begin{aligned} \mathcal{V}(x) &= \int_0^{+\infty} \rho(\lambda) \left(\int_0^1 \frac{\lambda}{x+\lambda\frac{1}{\omega}} d\frac{1}{\omega}\right) d\lambda \\ &= \int_0^{+\infty} \rho(\lambda) \left(\int_1^{\infty} \left(\frac{\lambda}{\omega x+\lambda}\right) \frac{d\omega}{\omega}\right) d\lambda. \end{aligned} \tag{41}$$

By changing the variable with $\Omega = \omega x$, whereas $\omega = \frac{\Omega}{x}, d\omega = \frac{d\Omega}{x}$, and exchanging the λ and ω integration, we have

$$\begin{aligned} \mathcal{V}(x) &= \int_0^{+\infty} \rho(\lambda) \left(\int_1^{\infty} \left(\frac{\lambda}{\omega x+\lambda}\right) \frac{d\omega}{\omega}\right) d\lambda \\ &\overset{a}{=} \int_0^{+\infty} \rho(\lambda) \left(\int_x^{\infty} \left(\frac{\lambda}{\Omega+\lambda}\right) \frac{1}{\Omega} d\Omega\right) d\lambda \\ &= \int_0^{+\infty} \frac{1}{\Omega}\rho(\lambda) \left(\int_x^{\infty} \left(\frac{\lambda+\Omega-\Omega}{\Omega+\lambda}\right) d\Omega\right) d\lambda \\ &= \int_x^{+\infty} \left(\frac{1}{\Omega}\rho(\lambda) d\lambda - \int_o^{\infty} \left(\frac{\rho(\lambda)}{\Omega+\lambda}\right) d\lambda\right) d\Omega, \end{aligned} \tag{42}$$

where a denotes the exchange of Ω with ωx. Thus consequently we have

$$\mathcal{V}(x) = \int_x^{+\infty} (\frac{1}{\omega} - S(-\omega))d\omega. \tag{43}$$

This completes the proof. ∎

APPENDIX B

Proof: As the proof procedures are similar for $F_{\mathbf{D}}$, $F_{\mathbf{B}}$, and $F_{\mathbf{A}}$, here we only give the proof of the tightness of $F_{\mathbf{A}}$. Without loss of generality, assuming $\mathbf{A} \sim \mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, its PDF can be given as

$$p(x; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^2|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right). \tag{44}$$

The CDF can be given while doing the integral to $\mathbf{x}$, which is

$$F(x; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^2|\boldsymbol{\Sigma}|}} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_k} \times \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) dx_1 \ldots dx_k. \tag{45}$$

Hence by following the proof procedure of sequence tightness, we have proof that for $\epsilon > 0$ and $N_\epsilon > 0$, for all k, the following inequality holds

$$\begin{aligned} &F_k(N_\eta; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= \frac{1}{\sqrt{(2\pi)^2|\boldsymbol{\Sigma}|}} \int_{-\infty}^{N_{\epsilon 1}} \cdots \int_{-\infty}^{N_{\epsilon k}} \\ &\quad \times \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) dN_{\epsilon 1} \ldots dN_{\epsilon k} \\ &> 1 - \eta. \end{aligned} \tag{46}$$

That is

$$\begin{aligned} \epsilon &> 1 - F_k(N_\epsilon; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= 1 - \frac{1}{\sqrt{(2\pi)^2|\boldsymbol{\Sigma}|}} \int_{-\infty}^{N_{\epsilon 1}} \cdots \int_{-\infty}^{N_{\epsilon k}} \\ &\quad \times \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) dN_{\epsilon 1} \ldots dN_{\epsilon k}. \end{aligned} \tag{47}$$

As known, $-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$ is a quadratic form of $\mathbf{x}$, and $\boldsymbol{\Sigma}$ is positive definite. Thus for any $x \neq \mu$, we have

$$-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) < 0. \tag{48}$$

This implies that both $p(x; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $F(x; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ are monotone decreasing functions of x. Thus by assuming a vector $\mathbf{b}$ with $0 < b \leq N$, when $b \to 0$ we have

$$\epsilon > 1 - F_k(b; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \to \epsilon > 1 - F_k(N_\epsilon; \boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{49}$$

By following the Glivenko-Cantelli theorem, we can arrive at the tightness conclusion with EDF $F_{\mathbf{A}}$. The proof of the tightness characteristics of sequences $F_{\mathbf{D}}$ and $F_{\mathbf{B}}$ is similar, which is omitted here.

This completes the proof. ∎

APPENDIX C

Proof: As stated by (24), the achievable transmission rate of each user k can be given as

$$C_k = \mathbb{E}\left\{\log\det\left(\mathbf{I}_N + \frac{P_k \mathbf{H}\mathbf{H}^H}{\sum_{k'=1}^{k-1} P_{k'} \mathbf{H}\mathbf{H}^H}\right)\right\}. \tag{50}$$

While reducing of fractions to a common denominator, we have

$$C_k = \mathbb{E}\left\{\log\det\left(\frac{\sum_{k'=1}^{k-1} P_{k'} \mathbf{H}\mathbf{H}^H + P_k \mathbf{H}\mathbf{H}^H}{\sum_{k'=1}^{k-1} P_{k'} \mathbf{H}\mathbf{H}^H}\right)\right\}. \tag{51}$$

Additionally, this can be further written as

$$\begin{aligned} C_k &= \mathbb{E}\left\{\log\det\left(\frac{\sum_{k'=1}^{k} P_{k'} \mathbf{H}\mathbf{H}^H}{\sum_{k'=1}^{k-1} P_{k'} \mathbf{H}\mathbf{H}^H}\right)\right\} \\ &= \mathbb{E}\left\{\log\det\left(\frac{\sum_{k'=1}^{k} P_{k'} \mathbf{H}\mathbf{H}^H}{\sum_{k'=1}^{k} P_{k'} \mathbf{H}\mathbf{H}^H - P_k \mathbf{H}\mathbf{H}^H}\right)\right\} \\ &= \log\left\{\left(\frac{\sum_{k'=1}^{k} P_{k'}}{\sum_{k'=1}^{k} P_{k'} - P_k}\right)\mathbb{E}\left(\det \mathbf{H}\mathbf{H}^H\right)\right\}. \end{aligned} \tag{52}$$

Exchanging the base of logarithm to the last equality, and following the prior studies by [27], [42], and [57] will yield the following equality

$$\begin{aligned} C_k &= \log\left\{\left(\frac{\sum_{k'=1}^{k} P_{k'}}{\sum_{k'=1}^{k} P_{k'} - P_k}\right)\mathbb{E}\left(\det \mathbf{H}\mathbf{H}^H\right)\right\} \\ &= \frac{1}{\ln 2}\left\{\ln\left(\frac{\sum_{k'=1}^{k} P_{k'}}{\sum_{k'=1}^{k} P_{k'} - P_k}\right) \right. \\ &\quad \left. + \ln\int_0^{+\infty} \lambda f(\lambda)\right\}. \end{aligned} \tag{53}$$

While summarizing the achievable rate to all K users, it will be the result in *Theorem 4.1*.

This completes the proof. ∎

APPENDIX D

Proof: Here the eigenvalue decomposition (ED) method is utilized. The difference between the singular eigenvalue decomposition (SVD) and ED methods is that, SVD yields the rotation transform while ED is not. Since $\mathbf{B}$ is a normal square random matrix, ED will yield two unitary matrices plus a diagonal matrix. The analysis is simplified via this method. Inspired by the prior studies in [27], [57], and [58], the eigenvalue decomposition of $\mathbf{B}$ can be given as

$$\mathbf{B} = \mathbf{Q}\mathbf{D}_1\mathbf{Q}^H, \tag{54}$$

whereas $\mathbf{Q}$ is the unitary matrix, and $\mathbf{D}_1$ is the diagonal matrix. With this in hand, for $\mathbf{BA}$, the following equality holds

$$\begin{aligned} (\mathbf{BA})(\mathbf{BA})^H &= \mathbf{Q}\mathbf{D}_1\mathbf{Q}^H\mathbf{A}\mathbf{A}^H\mathbf{Q}\mathbf{D}_1^H\mathbf{Q}^H \\ &= \mathbf{Q}\mathbf{D}_1\tilde{\mathbf{A}}\tilde{\mathbf{A}}^H\mathbf{D}_1^H\mathbf{Q}^H \triangleq \mathbf{Q}\mathbf{W}_0\mathbf{Q}^H. \end{aligned} \tag{55}$$

This gives the matrix $\mathbf{W}_0$ a central Wishart matrix with K non-zero eigenvalues defined as $0 < \chi_1 < \ldots < \chi_K < \infty$. By denoting the eigenvalues of $\mathbf{B}\mathbf{B}^H$ as $0 < \upsilon_1 < \ldots < \upsilon_L < \infty$, in line with prior study [56], the CDF of the largest

eigenvalue of $(\mathbf{BA})(\mathbf{BA})^H$ conditioned on $\mathbf{B}$ can be given as [56], [58]

$$F_{\chi_{max}}(x|\mathbf{B}) = \frac{(-1)^{K(L-K)}\det(\Delta(x))}{\det(\mathbf{V})\prod_{i=1}^{K}\Gamma(K-i+1)}. \tag{56}$$

where $\Delta(x)$ is an $L \times L$ matrix with entries

$$\Delta(x)_{i,j} = \begin{cases} (-\frac{1}{v_j})^{L-K-i}, & \text{for } i \le L-K, \\ v_j^{L-i+1}\gamma\,(L-i+1, \frac{xL}{v_j}), & \text{for } i > L-K. \end{cases} \tag{57}$$

Additionally, $\mathbf{V}$ is a $L \times L$ matrix defined as [27]

$$\det(\mathbf{V}) = \left(\prod_{i=1}^{L} v_i^K\right)\prod_{1\le l\le k\le L}\left(\frac{1}{v_k}-\frac{1}{v_l}\right). \tag{58}$$

By some manipulations with regard to the Vandermonde determinant identity, it can be further written as

$$\begin{aligned}\det(\mathbf{V}) &= \left(\prod_{i=1}^{L} v_i^K\right)(-1)^{\frac{L(L-1)}{2}}\frac{\prod_{1\le i\le j\le L}(v_j-v_i)}{\prod_{i=1}^{L} v_i^{L-1}} \\ &= \left(\prod_{i=1}^{L} v_i^{K-L+1}\right)\prod_{1\le i\le j\le L}(v_i-v_j)\end{aligned} \tag{59}$$

On the other hand, by following the prior studies in [57] and [59], for the square matrix $\mathbf{B} \in \mathbb{C}^{L\times L}$ here in this paper, the joint PDF of the eigenvalue $0 < v_1 < \ldots < v_L < \infty$ of the matrix constituted by $\mathbf{BB}^H$ is given by

$$f_v(\mathbf{D}_1) = \frac{e^{-\sum_{i=1}^{L} v_i}\prod_{i<j}^{L}(v_j-v_i)^2}{\prod_{i=1}^{L}\Gamma(L-i+1)^2}. \tag{60}$$

To this end, the unconditional CDF of $0 < \chi_1 < \ldots < \chi_K < \infty$ will be

$$F_{\chi_{max}}(x) = \int_{\mathcal{U}} F_{\chi_{max}}(x|\mathbf{B}) f_v(\mathbf{D}_1) d v_1, \ldots d v_L, \tag{61}$$

where $\mathcal{U} \triangleq \{0 < \chi_1 < \ldots < \chi_K < \infty\}$, this gives

$$F_{\chi_{max}}(x) = \frac{(-1)^{K(L-K)}\det(\mathbf{D}(x))}{\prod_{k=1}^{K}\Gamma(K-k+1)\prod_{i=1}^{L}\Gamma(L-i+1)^2}, \tag{62}$$

where $\mathbf{D}(x)$ is given as

$$\begin{aligned}\mathbf{D}(x) = \int_{\mathcal{U}} &\det(\Delta(x))e^{-\sum_{i=1}^{L} v_i} \\ &\times\prod_{i=1}^{L} v_i^{L-K-1}\prod_{i<j}^{L}(v_i-v_j)d v_1, \ldots d v_L.\end{aligned} \tag{63}$$

Observation has that

$$\prod_{i<j}^{L}(v_i-v_j) = \det(v_j^{i-1}). \tag{64}$$

By following the analysis in [56], $\mathbf{D}(x)$ is finally given as

$$\mathbf{D}(x)_{i,j} = \begin{cases} (-1)^{L-K-i}\Gamma(i+j-1), & \text{for } i \le L-K, \\ \int_0^{+\infty} e^t t^{2L-K-i+j-1}\gamma\,(L-i+1, \frac{xL}{t})dt, & \\ \text{for } i > L-K. & \end{cases} \tag{65}$$

To determine the second expression of (65), it is noticed that $\gamma\,(\cdot,\cdot)$ is defined as [27], [56]

$$\gamma\,(a,x) = \int_0^x t^{a-1}e^{-t}dt = (a-1)!\left(1-e^{-x}\sum_{i=0}^{a-1}\frac{x^i}{i!}\right). \tag{66}$$

Furthermore, observation from [59] has that

$$\int_0^{+\infty} x^{\alpha-1}e^{-\beta x-\frac{\gamma}{x}}dx = 2(\frac{\gamma}{\beta})^{\frac{\alpha}{2}}K_\alpha(2\sqrt{\beta\gamma}), \tag{67}$$

where $K_a(b)$ as the modified Bessel function of the first kind. By substituting this (67) and (66) into (65), with tremendous calculation, the determinant expression of its second part can be finally obtained as

$$\begin{aligned}&\int_0^{+\infty} e^t t^{2L-K-i+j-1}\gamma\,(L-i+1, \frac{xL}{t})dt \\ &= (L-i)!\Bigg[\Gamma(2L-K-i+j) - \sum_{i}^{L-i}\frac{(xL)^i}{i!} \\ &\quad\times 2(xL)^{\frac{2L-K-2i+j}{2}}K_{2L-K-2i+j}(2\sqrt{xL})\Bigg], \text{ for } i > L-K.\end{aligned} \tag{68}$$

Thus the PDF of $0 < \chi_1 < \ldots < \chi_K < \infty$ can be obtained as

$$f_\chi(x) = \frac{(-1)^{K(L-K)}\frac{d}{dx}[\det(\mathbf{D}(x))]}{\prod_{k=1}^{K}\Gamma(K-k+1)\prod_{i=1}^{L}\Gamma(L-i+1)^2}. \tag{69}$$

In line with [60], the unordered PDF of eigenvalues $\lambda_1, \ldots, \lambda_K$ of $(\mathbf{DBA})(\mathbf{DBA})^H$ conditioned on $\mathbf{BA}$ is

$$\begin{aligned}f_\lambda(\lambda|\mathbf{BA}) &= \frac{1}{L\prod_{i<j}^{K}(\chi_j-\chi_i)} \\ &\quad\times\sum_{m=L-K+1}^{L}\frac{\lambda^{K-L+m-1}}{\Gamma(K-L+m-1)}\det(\mathbf{G})\end{aligned} \tag{70}$$

whereas $\mathbf{G}$ is a $L \times L$ matrix with entries

$$\mathbf{G}_{i,j} = \begin{cases} \chi_j^{i-1}, & for\ i \ne j \\ \chi_j^{L-K-1}e^{-\frac{\lambda}{\chi_j}}. & for\ i = j. \end{cases} \tag{71}$$

Thus by using $f(\lambda) = f_\lambda(\lambda|\mathbf{BA})f_\chi(x)$ and integrating it to all χ, we can finally obtain the result.

This completes the proof. ∎

ACKNOWLEDGMENT

The author would like to thank Prof. Yang Chen with the University of Macau for the valuable discussion and help on the deduction of the relationship between Stieltjes and Shannon transform.

REFERENCES

[1] T. S. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.
[2] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
[3] R. G. Gallager, *An Inequality Capacity Region Multiaccess Multipath Channels*. New York, NY, USA: Springer, 1994, pp. 129–139.

[4] D. Tse and P. Viswanath, *Fundamentals Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[5] Z. Chang *et al.*, "Energy efficient resource allocation for wireless power transfer enabled collaborative mobile clouds," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3438–3450, Dec. 2016.

[6] Z. Chang, T. Ristaniemi, and Z. Niu, "Radio resource allocation for collaborative OFDMA relay networks with imperfect channel state information," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2824–2835, May 2014.

[7] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

[8] J. Zhang, L. Dai, X. Zhang, E. Björnson, and Z. Wang, "Achievable rate of rician large-scale MIMO channels with transceiver hardware impairments," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8800–8806, Oct. 2016.

[9] L. Dai, X. Gao, X. Su, S. Han, C. L. I, and Z. Wang, "Low-complexity soft-output signal detection based on gauss–seidel method for uplink multiuser large-scale MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4839–4845, Oct. 2015.

[10] L. Dai, Z. Wang, and Z. Yang, "Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 251–263, Feb. 2013.

[11] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, "Modeling and analyzing millimeter wave cellular systems," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 403–430, Jan. 2017.

[12] Z. Gao, L. Dai, and Z. Wang, "Channel estimation for MmWave massive MIMO based access and backhaul in ultra-dense network," in *Proc. IEEE ICC*, May 2016, pp. 1–6.

[13] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE VTC*, Jun. 2013, pp. 1–5.

[14] J. Choi, "Minimum power multicast beamforming with superposition coding for multiresolution broadcast and application to NOMA systems," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 791–800, Mar. 2015.

[15] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[16] Z. Yang, Z. Ding, P. Fan, and Z. Ma, "Outage performance for dynamic power allocation in hybrid non-orthogonal multiple access systems," *IEEE Commun. Lett.*, vol. 20, no. 8, pp. 1695–1698, Aug. 2016.

[17] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[18] Z. Zhou, M. Dong, K. Ota, R. Shi, Z. Liu, and T. Sato, "Game-theoretic approach to energy-efficient resource allocation in device-to-device underlay communications," *IET Commun.*, vol. 9, no. 3, pp. 375–385, Feb. 2015.

[19] C. Xu *et al.*, "Efficiency resource allocation for device-to-device underlay communication systems: A reverse iterative combinatorial auction based approach," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 348–358, Sep. 2013.

[20] D. Zhang, Z. Zhou, S. Mumtaz, J. Rodriguez, and T. Sato, "One integrated energy efficiency proposal for 5G IoT communications," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1346–1354, Dec. 2016.

[21] Z. Chang, S. Zhou, T. Ristaniemi, and Z. Niu, "Collaborative mobile clouds: An energy efficient paradigm for content sharing," *IEEE Wireless Commun.*, to be published.

[22] M. R. Akdeniz *et al.*, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[23] G. Lee, Y. Sung, and J. Seo, "Randomly-directional beamforming in millimeter-wave multiuser MISO downlink," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1086–1100, Feb. 2016.

[24] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[25] X. Liu and X. Wang, "Efficient antenna selection and user scheduling in 5G massive MIMO-NOMA system," in *Proc. IEEE VTC*, May 2016, pp. 1–5.

[26] Z. Ding and H. V. Poor, "Design of massive-MIMO-NOMA with limited feedback," *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 629–633, May 2016.

[27] D. Zhang, K. Yu, Z. Wen, and T. Sato, "Outage probability analysis of NOMA within massive MIMO systems," in *Proc. IEEE VTC*, May 2016, pp. 1–5.

[28] X. Gao, L. Dai, Z. Chen, Z. Wang, and Z. Zhang, "Near-optimal beam selection for beamspace MmWave massive MIMO systems," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 1054–1057, May 2016.

[29] A. Vinel, Q. Ni, D. Staehle, and A. Turlikov, "Capacity analysis of reservation-based random access for broadband wireless access networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 2, pp. 172–181, Feb. 2009.

[30] M. Debbah and R. R. Müller, "MIMO channel modeling and the principle of maximum entropy," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1667–1690, May 2005.

[31] R. Couillet, M. Debbah, and J. W. Silverstein, "A deterministic equivalent for the analysis of correlated MIMO multiple access channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3493–3514, Jun. 2011.

[32] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[33] A. Lu, X. Gao, and C. Xiao, "A free deterministic equivalent for the capacity of MIMO MAC with distributed antenna sets," in *Proc. IEEE ICC*, Jun. 2015, pp. 1751–1756.

[34] J. Zhang, C.-K. Wen, S. Jin, X. Gao, and K.-K. Wong, "On capacity of large-scale MIMO multiple access channels with distributed sets of correlated antennas," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 133–148, Feb. 2013.

[35] A. Firag, P. J. Smith, and M. R. McKay, "Capacity analysis of MIMO three product channels," in *Proc. AusCTW*, Feb. 2010, pp. 13–18.

[36] Z. Pan, J. Lei, Y. Zhang, X. Sun, and S. Kwong, "Fast motion estimation based on content property for low-complexity H.265/HEVC encoder," *IEEE Trans. Broadcast.*, vol. 62, no. 3, pp. 675–684, Sep. 2016.

[37] J. W. Silverstein and A. M. Tulino, "Theory of large dimensional random matrices for engineers," in *Proc. IEEE 19th Int. Symp. Spread Spectr. Techn. Appl.*, Aug. 2006, pp. 458–464.

[38] Z. Zhou, Y. Wang, Q. M. J. Wu, C. N. Yang, and X. Sun, "Effective and efficient global context verification for image copy detection," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 1, pp. 48–63, Jan. 2017.

[39] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, "A general power allocation scheme to guarantee quality of service in downlink and uplink NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7244–7257, Nov. 2016.

[40] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[41] J. Choi, "H-ARQ based non-orthogonal multiple access with successive interference cancellation," in *Proc. IEEE GLOBECOM*, Nov. 2008, pp. 1–5.

[42] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[43] T. S. Rappaport, R. W. Heath, Jr., R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*. Westford, MA, USA: Prentice-Hall, 2014.

[44] M. N. Kulkarni, A. Ghosh, and J. G. Andrews, "A comparison of MIMO techniques in downlink millimeter wave cellular networks with hybrid beamforming," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1952–1967, May 2016.

[45] H. Zhao *et al.*, "28 GHz millimeter wave cellular communication measurements for reflection and penetration loss in and around buildings in New York city," in *Proc. IEEE ICC*, Jun. 2013, pp. 5163–5167.

[46] S. Collonge, G. Zaharia, and G. E. Zein, "Influence of the human activity on wide-band characteristics of the 60 GHz indoor radio channel," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 2396–2406, Nov. 2004.

[47] C.-K. Wen, G. Pan, K.-K. Wong, M. Guo, and J.-C. Chen, "A deterministic equivalent for the analysis of non-Gaussian correlated MIMO multiple access channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 329–352, Jan. 2013.

[48] J. Dehardt, "Generalizations of the Glivenko-Cantelli theorem," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 2050–2055, Dec. 1971.

[49] N. Letzepis and A. Grant, "Shannon transform of certain matrix products," in *Proc. IEEE ISIT*, Jun. 2007, pp. 1646–1650.

[50] J. Gong, S. Zhou, Z. Zhou, and Z. Niu, "Joint optimization of content caching and push in renewable energy powered small cells," in *Proc. IEEE ICC*, May 2016, pp. 1–6.
[51] C.-K. Wen, S. Jin, and K.-K. Wong, "On the sum-rate of multiuser MIMO uplink channels with jointly-correlated Rician fading," *IEEE Trans. Commun.*, vol. 59, no. 10, pp. 2883–2895, Oct. 2011.
[52] R. Müller, "On the asymptotic eigenvalue distribution of concatenated vector-valued fading channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 2086–2091, Jul. 2002.
[53] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*, 2nd ed. Hoboken, NJ, USA: Wiley, 2011.
[54] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1893–1909, Sep. 2004.
[55] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.
[56] S. Jin, M. R. McKay, K. K. Wong, and X. Gao, "Transmit beamforming in Rayleigh product MIMO channels: Capacity and performance analysis," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5204–5221, Oct. 2008.
[57] H. Zhang, S. Jin, M. R. McKay, X. Zhang, and D. Yang, "High-SNR performance of MIMO multi-channel beamforming in double-scattering channels," *IEEE Trans. Commun.*, vol. 59, no. 6, pp. 1621–1631, Jun. 2011.
[58] M. Kang and M. S. Alouini, "Impact of correlation on the capacity of MIMO channels," in *Proc. IEEE ICC*, vol. 4. May 2003, pp. 2623–2627.
[59] D. Zwillinger and V. Moll, *Table of Integrals, Series, and Products*, 8th ed. San Diego, CA, USA: Academic, 2015.
[60] S. Jin, R. McKay, C. Zhong, and K.-K. Wong, "Ergodic capacity analysis of amplify-and-forward MIMO dual-hop systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2204–2224, May 2010.

Di Zhang (S'13) received the degree (Hons.) in 2013. He is currently pursuing the Ph.D. degree with Waseda University, Tokyo, Japan. His research interests include massive MIMO, green communications, information theory, and signal processing. He is a Student Member of the IEICE. He served as a TPC member of the IEEE VTC and WCNC workshops in 2016.

Zhenyu Zhou (S'06–M'11) received the M.E. and Ph.D. degrees from Waseda University, Tokyo, Japan, in 2008 and 2011, respectively. From 2012 to 2013, he was the Chief Researcher with the Department of Technology, KDDI, Tokyo. Since 2013, he has been an Associate Professor with the School of Electrical and Electronic Engineering, North China Electric Power University, China. He has also been a Visiting Scholar with the Tsinghua–Hitachi Joint Laboratory on Environment-Harmonious ICT, University of Tsinghua, Beijing, since 2014. His research interests include green communications and smart grid. He is a member of the IEICE and the CSEE. He received the Young Researcher Encouragement Award from the IEEE Vehicular Technology Society in 2009. He served as the Workshop Co-Chair of the IEEE ISADS 2015 and a TPC member of the IEEE VTC 2017, the IEEE VTC 2016, the IEEE ICC 2016, the IEEE ICC 2015, the Globecom 2015, the ACM Mobimedia 2015, and the IEEE Africon 2015.

Chen Xu (S'12–M'15) received the B.S. degree from the Beijing University of Posts and Telecommunications in 2010 and the Ph.D. degree from Peking University in 2015. She is currently a Lecturer with the School of Electrical and Electronic Engineering, North China Electric Power University, China. Her research interests mainly include wireless resource allocation and management, game theory, optimization theory, heterogeneous networks, and smart grid communication. She received the Best Paper Award at the International Conference on Wireless Communications and Signal Processing in 2012 and the IEEE Leonard G. Abraham Prize 2016.

Yan Zhang (M'05–SM'10) received the Ph.D. degree from the School of Electrical and Electronics Engineering, Nanyang Technological University, Singapore. He is currently a Full Professor with the Department of Informatics, University of Oslo, Oslo, Norway. His current research interests include next-generation wireless networks leading to 5G, green and secure cyber-physical systems, such as smart grid, healthcare, and transport. He is an IEEE Vehicular Technology Society (VTS) Distinguished Lecturer. He is also a Senior Member of the IEEE ComSoc, the IEEE CS, the IEEE PES, and the IEEE VTS. He is a fellow of the IET. He is an Associate Technical Editor of the *IEEE Communications Magazine*, an Editor of the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, an Editor of the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, and an Associate Editor of the IEEE ACCESS. He serves as the Chair in a number of conferences, including the IEEE GLOBECOM 2017, the IEEE PIMRC 2016, the IEEE CloudCom 2016, the IEEE ICCC 2016, the IEEE CCNC 2016, the WCSP 2016, the IEEE SmartGridComm 2015, and the IEEE CloudCom 2015. He serves as a TPC member for numerous international conference, including the IEEE INFOCOM, the IEEE ICC, the IEEE GLOBECOM, and the IEEE WCNC.

Jonathan Rodriguez (M'04–SM'13) received the master's and Ph.D. degrees in electronic and electrical engineering from the University of Surrey, U.K., in 1998 and 2004, respectively. In 2005, he was a Researcher with the Institutõ de Telecomunicações. He was a Senior Researcher with the Institutõ de Telecomunicações in 2008, where he established the 4TELL Research Group targeting the next generation mobile networks with key interests on energy efficient design, cooperative strategies, security, and electronic circuit design. He has served as a Project Coordinator for major international research projects, such as the Eureka LOOP and the FP7 C2POWER, while acting as the Technical Manager of the FP7 COGEU and the FP7 SALUS. He is currently leading the H2020-ETN SECRET Project, a European Training Network on 5G communications. He joined the University of South Wales, U.K., in 2017, where he was appointed Professor of Mobile Communications. He has authored over 350 scientific works, including nine book editorials. He is a Chartered Engineer (CEng) and Fellow of the IET (FIET) since 2015.

Takuro Sato (F'13) received the Ph.D. degree in electronics engineering from Niigata University, Niigata, Japan, in 1993. Since 1995, he has been a Professor with the Niigata Institute of Technology, where his research focused on CDMA, OFDM, personal communication systems, and other related areas. In 2004, he joined GITS/GITI, Waseda University, Tokyo, as a Professor, where he is currently serving as the Dean. He has been involved in the research on PCM transmission equipment development, mobile communications, data transmission, and digital signal processing. He has developed the wideband CDMA system for personal communication system and joined the PCS Standardization Committee in the USA and Japan. His contributions are mainly in high speed cellular modem standardization for ITU, 2.4GHz PCS for ITA and wireless LAN of the IEEE 802.11. He has authored 11 books and over 200 papers. His current research interests include next generation mobile communications, wireless communications, ICN/CCN technology, ICT in smart grid, and their global standardizations. He is a fellow of the IEICE. He served as the Chair of the IEEE ISADS 2015, the IEEE Globecom 2015, and various other conferences and journal editors.

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.