A deterministic equivalent for the capacity analysis of correlated multi-user MIMO channels

Romain Couillet[†,‡], Mérouane Debbah[‡], Jack Silverstein[⊥]

Abstract

This paper provides the analysis of capacity expressions for multi-user and multi-cell wireless communication schemes when the transmitters and the receivers have a large number of correlated antennas. Our main contribution mathematically translates into a deterministic equivalent of the Shannon transform of a class of large dimensional random matrices; the latter are sums of Gram matrices with separable variance profiles, the covariance matrices of which possibly have asymptotically large eigenvalues. On the applicative side, this class of large matrices is used in this contribution to model (i) multi-antenna multiple access (MAC) and broadcast channels (BC) with transmit and receive channel correlation, (ii) multiple-input multiple-output (MIMO) communications with inter-cell interference and channel correlation both at the base stations and at the receivers. The theoretical capacity formulas obtained for these models extend the classical results on multi-user MIMO capacities in independent and identically distributed (i.i.d.) Gaussian channels to the more realistic Gaussian channels with separable variance profile. On an information theoretical viewpoint, this article provides: in scenario (i), an asymptotic description of the MAC and BC rate regions as a function only of the transmit and receive correlation matrices (thus independently of the random channel realizations); an asymptotic expression of the capacity-maximizing covariance matrices in the uplink MAC and downlink BC; a water-filling algorithm which, upon convergence, is proved to converge to the capacity-achieving power allocation at the transmitters both in MAC and BC. In scenario (ii), the article provides: an expression of the single-user decoding capacity and minimum mean square error (MMSE) decoding capacity when interference is treated as Gaussian noise with a known variance profile; for single-user decoding, the capacity-achieving antenna power allocation policy at the transmitter.

I. Introduction

When mobile networks were some time ago expected to run out of power and frequency resources while being simultaneously subject to an increasing demand of achievable data rates, Foschini [6] and Telatar [7] introduced the notion of multiple input multiple output (MIMO) systems and predicted a growth of capacity performance of $\min(N, n)$ for a communication between an n-antenna transmitter and an N-antenna receiver when the propagation

[†]ST-Ericsson, 505 Route des Lucioles, 06560 Sophia Antipolis

[‡]Alcatel-Lucent Chair on Flexible Radio, Supélec, 3 rue Joliot-Curie, 91192 Gif sur Yvette

[⊥]Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27695-8205

channel matrix model is formed of independent and identically (i.i.d.) Gaussian entries. However, in practical systems, this tremendous multiplexing gain can only be provided for large signal-to-interference plus noise ratios (SINR) and for uncorrelated transmit and receive antenna arrays at both communication sides. In present wireless mobile networks, the scarcity of available frequency resources has led to a widespread incentive for MIMO communications. Due to space limitations, mobile designers now embed more and more antennas in small devices, which inevitably spawns non-negligible correlation patterns at the antenna arrays and thus non-negligible effects on the achievable transmission rates. Since MIMO systems come along with a tremendous increase in signal processing requirements (which naturally imply a large non-linear increase in power consumption), both base station and mobile manufacturers need to accurately assess the exact cost of bit rate increases on finite size devices, prone to experience strong transmit and receive correlation. Our intention is to evaluate the antenna efficiency, i.e. the mean per-antenna achievable rate, for different communication models detailed in the following.

Multi-cell and multi-user systems are among the scenarios of main interest to cellular service providers. The scope of the present study (which, we will see, can be extended to a larger set of mobile communications scenarios) lies in the following two wireless communication systems, which the authors consider of most valuable interest,

1) the multiple access channels (MAC) in which K transmitters, hereafter assimilated to mobile terminal users, transmit information to a unique receiver, hereafter referred to as the base station; and the dual broadcast channels (BC) in which the base station multi-casts information to the K users. While the major scientific breakthroughs in multi-antenna broadcast channels are quite recent, see e.g. [20], the practical applications are foreseen to arise in a near future with the so-called multi-user MIMO techniques to be used in future long term evolution standards, e.g. [1].
2) the single-user decoding and minimum mean square error (MMSE) decoding [8] in multi-cell scenarios. In most current mobile communication systems, the wireless networks are composed of multiple overlapping cells, controlled by non-cooperating base stations. In these conditions, the achievable rates for every user in a cell, assuming no intra-cell interference, corresponds to the capacity of the single-user decoding scheme in which interfering signals are treated as Gaussian noise with a known correlation pattern. However, single-user decoders are not linear decoders and are often replaced in practical applications by the simpler linear MMSE decoders; these decoders maximize the signal-to-interference plus noise ratio (SINR) at the receiver.

The achievable rate region of MIMO MAC and BC channels for generic channels have been known since the successive contributions [20]-[21], who established an important duality link between MAC rate regions and BC rate regions in vector channels, both in single antenna and MIMO channels. These important contributions provide the general descriptions of the rate regions under fixed wireless channels, without any underlying transmission model. As a consequence, the exact expressions obtained are often intractable and generally do not provide any insight of the various channel parameters involved on the resulting achievable data rate. The mathematical field of large random matrices allows one to circumvent this issue by providing closed-form approximations of achievable rates as a function of the most relevant channel parameters only[1]. The most notable result in the line of the present study for random channels is due to Tulino [2], which provides an asymptotic capacity expression of point-to-point MIMO systems when the channel random matrix is composed of i.i.d. Gaussian entries. This result naturally extends to multiple users by dividing the MIMO channel matrix into K sub-matrices; this therefore allows one to obtain a description of the MAC and BC rate regions for the uncorrelated Gaussian channels, which in shown in [2] to depend only (i) on the signal to noise ratio (SNR) at the receivers and (ii) on the ratio $c = N/n$ between the number N of transmit antennas at the base station and the number n of receive antennas at all receivers. Tulino also provides an expression of the capacity-achieving power allocation policy at the base station. In [24], Ulukus derives the capacity-achieving power allocation policy for a finite number of antennas at all transmit/receive devices in the case of K users whose channels $\mathbf{H}_k$, $1 \leq k \leq K$, are modelled as Kronecker channels, i.e. Gaussian with separable variance profile, $\mathbf{H}_k = \mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k^{\frac{1}{2}}$, where $\mathbf{R}_k$, $\mathbf{T}_k$ are Hermitian nonnegative definite[2] and $\mathbf{X}_k$ is random with i.i.d. Gaussian entries; however, Ulukus (i) does not provide a theoretical large dimensional expression of the resulting capacity, (ii) makes the strong assumption that all $\mathbf{R}_k$ matrices are equal. It is in fact rather straightforward to observe that, under the assumption that all all $\mathbf{R}_k$ share the same eigenspace, the result from [2] is extensible to the Kronecker channel model, see e.g. [3]. When the $\mathbf{R}_k$ matrices have no trivial relationship, the problem is more complex and requires different mathematical tools. Those tools allow us in the present work to obtain a deterministic equivalent of the per-antenna rate regions associated to the channels $\mathbf{H}_k$: this is our main result, which we provide in Theorem 2. In fact, it is important to note that the final formula of Theorem 2 is already found and used by Chen, Equation (32) in [11]; however, the latter is provided without any proof, nor any hypotheses on the considered matrices, and merely stems from the previous Equation (6) which is in valid when all $\mathbf{R}_k$ matrices have the same eigenspace, but which is incorrect in the general case, as will become clear in the following. Chen also provides the iterative water-filling algorithm which we will also obtain in the course of this paper (see Table I); however, the convergence of this algorithm to the correct capacity, which we will prove, is not provided in [11]. Regarding multi-cell networks, to the authors' knowledge, few contributions treat simultaneously the problem of multi-cell interference in more structured channel models than i.i.d. Gaussian channels. In [13], the authors carry out the performance analysis of TDMA-based networks with inter-cell interference. In [14], a random matrix approach is used to study large CDMA-based networks with inter-cell interference. In our particular MIMO context, it is important to mention the work of Moustakas [10] who conjectures an analytic solution to the single-user decoding problem with channel correlation and interference with known variance profile, using the replica method [12].[3]

Practical channel models, as recalled earlier, prove more challenging than mere i.i.d. Gaussian channels. Small distances between antennas embedded in the wireless devices as well as solid angles of transmission and reception of signal energy tend to correlate (potentially strongly) the emitted and received signals. A largely spread channel

[1]such as the transmit/receive covariance matrices, the deterministic line of sight components etc.

[2]$\mathbf{R}_k^{\frac{1}{2}}$ and $\mathbf{T}_k^{\frac{1}{2}}$ are then defined as their unique nonnegative definite square root.

[3]we use here the term 'conjecture' instead of 'demonstrate' since the exactness of the replica method has not yet been mathematically proven.

model, which naturally unfolds from the knowledge of a correlation pattern at the transmitter and at the receiver [35] is the so-called Kronecker model, which mathematically translates into random matrices with i.i.d. Gaussian entries with separable correlation profile, as previously recalled. This model is of particular interest when no line of sight component is present in the channel and when a sufficiently large number of scatterers is found in the communication medium. This is the model we will consider in the following study. However note that, in their substantial contributions [26]-[27], Loubaton et al. provide a deterministic equivalent of capacity of a point-to-point MIMO channel [26] and its corresponding capacity-achieving input covariance matrix [27] when the channel matrix $\mathbf{H}$ is modelled as Ricean, i.e. $\mathbf{H} = \mathbf{A} + \mathbf{X}$ where $\mathbf{A}$ is some deterministic matrix, standing for the line-of-sight component, and $\mathbf{X}$ is Gaussian with general variance profile ($\mathrm{E}[|X_{ij}|^2] = \sigma_{ij}$). Of particular practical interest is also the theoretical work of Tse [9] 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 summarizes into two major mathematical theorems, contributing to the field of random matrix theory, and an important water-filling algorithm enabling to describe the boundaries of the MAC and BC rate regions. On a purely mathematical viewpoint, we provide a deterministic equivalent of the Stieltjes transform of the sum $\mathbf{B}_N$ of K Gaussian matrices with separable variance profile $\mathbf{R}_k^{\frac{1}{2}}\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k^{\mathsf{H}}\mathbf{R}_k^{\frac{1}{2}}$, $\mathbf{R}_k$, $\mathbf{T}_k$ nonnegative Hermitian, $\mathbf{X}_k$ i.i.d. and unitary invariant, plus a deterministic nonnegative Hermitian matrix $\mathbf{S}$,

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

This result extends the recent results from Taricco [5] and Honig [3] in which, respectively, all $\mathbf{R}_k$ matrices are equal or share the same eigenspace. The main consequence of letting the $\mathbf{R}_k$ matrices unconstrained is that no limit (for large N) eigenvalue distribution of $\mathbf{B}_N$ is available, even when the $\mathbf{R}_k$ and $\mathbf{T}_k$ matrices have a limit eigenvalue distribution. Therefore, classical results from random matrix theory are not usable here. The second major result lies in a deterministic equivalent of the Shannon transform of $\mathbf{B}_N$, obtained by integration of the deterministic equivalent of the Stieltjes transform. This result extends [2] to multiple channels with separable variance profile. For both deterministic equivalents of the Stieltjes and Shannon transforms, and contrary to most previous contributions in similar articles, we do not restrict here the eigenvalues of the $\mathbf{R}_k$ and $\mathbf{T}_k$ matrices to be uniformly (over the matrix sizes) bounded. In the present contribution, those eigenvalues can theoretically grow at a restricted rate, to be defined later. In practice, this allows to study a wider scope of channel models than in the previously mentioned studies; in particular, this study covers the usual case when the $\mathbf{R}_k$ and $\mathbf{T}_k$ matrices are constrained by $\mathrm{tr}(\mathbf{R}_k) = N$, $\mathrm{tr}(\mathbf{T}_k) = n_k$ (N and n_k the respective sizes of $\mathbf{R}_k$ and $\mathbf{T}_k$).

The remainder of this paper is structured as follows: in Section II, we provide a quick summary of our results and discuss the extent of their applicability. In Section III, we provide the two main mathematical theorems needed for practical applications in the subsequent sections. The complete proofs of both theorems are provided in the appendix. In Section IV, the rate region of MAC and BC channels and the capacity of single-user decoding and MMSE decoding with inter-cell interference are studied. In this section, we will introduce our third main result: an iterative water-filling algorithm to describe the boundary of the MAC and BC rate regions. In Section V, we provide simulation results of the previously derived theoretical formulas and discuss the advantages and limitations of the deterministic equivalents. Finally, in Section VI, we give our conclusions.

Notation: In the following, boldface lower-case symbols 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 $\mathbf{E}[\cdot]$ denotes expectation. The notation $F^{\mathbf{Y}}$ stands for the empirical distribution 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.

II. Scope and Summary of Main Results

In this section, we summarize the main results of this paper and explain how they naturally help to study, in the present multi-cell multi-user framework, the effects of channel correlation on the antenna *rate efficiency*, which we define as the mean achievable rate provided by each transmit/receive antenna.

A. General Model

Consider a set of K wireless entities which purpose is to communicate with another wireless device, either in downlink or in uplink, at the 'best' possible rate performance[4]. For clarity and without generality restriction, consider here the uplink scenario. Denote $\mathbf{H}_k$ the channel matrix model between the k^{th} transmitter k and the receiver. Then, assuming a large number of scatterers in the medium between both entities and no line-of-sight vision, it is often accurate to consider that $\mathbf{H}_k$ is modelled as Kronecker,

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

where, as already mentioned, the $N \times N$ matrix $\mathbf{R}_k^{\frac{1}{2}}$ and the $n_k \times n_k$ matrix $\mathbf{T}_k^{\frac{1}{2}}$ are respectively the only nonnegative square roots of the Hermitian nonnegative matrices $\mathbf{R}_k$ and $\mathbf{T}_k$, and $\mathbf{X}_k$ is a realization of a random Gaussian matrix[5]. The matrices $\mathbf{T}_k$ and $\mathbf{R}_k$ in this scenario model the correlation present in signals transmitted by user k and signals received at the receiver. It is important to stress out that those correlation patterns emerge both from the one-to-one antenna spacings on the volume limited devices and from the solid angles of useful transmitted and received energy. Without this second factor, it would make sense that all $\mathbf{R}_k$ are equals, which was claimed for instance in [24]. However, this would mean that signals are received isotropically at the receiver, which turns out to be often too strong an assumption to characterize practical communication channels. This being said, the only sensible restriction that one can make on the matrices $\mathbf{R}_k$ and $\mathbf{T}_k$ is for their diagonal entries to be less or equal to one, and then for their spectral norms to be less than N and n_k respectively. We will see that our results require different hypothesis, which nonetheless allow to treat some channel models with strong correlations. The

[4]we purposely are cryptic on the term 'best' for one might consider very different performance criteria, such as achievable sum-rate, max-min rate etc.

[5]Gaussian matrices often refer to random matrices with i.i.d. Gaussian entries.

assumption often considered, e.g. in [26]-[27], is that the spectral norms of $\mathbf{R}_k$ and $\mathbf{T}_k$, for all k, are uniformly bounded (over the matrix sizes), which means that only low correlation patterns can be studied; in the present work, we will allow a possibly large portion of the eigenvalues of $\mathbf{R}_k$, $\mathbf{T}_k$ to grow at a potentially high rate, along with growing N, n_k. More precisely, we provide a sufficient (but not necessary) condition on the number and growing rate of the eigenvalues of $\mathbf{R}_k$ and $\mathbf{T}_k$ which ensures the asymptotic accuracy of the deterministic equivalents under study. Under this condition, a wider scope of channel correlation models can be studied than in most previous contributions, including the classical case when $\operatorname{tr}(\mathbf{R}_k) = N$, $\operatorname{tr}(\mathbf{T}_k) = n_k$.

As will be evidenced by the information theoretic applications of Sections IV-A and IV-B, most multi-cell or multi-user capacity performance rely more or less directly on the so-called Stieltjes transform $m_N(z)$ of matrices $\mathbf{B}_N^{\mathcal{S}}$ of the type

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

for some given subset $\mathcal{S}$ of $\{1,\dots,K\}$.

The Stieltjes transform m_N of the $N \times N$ Hermitian matrix $\mathbf{B}_N^{\mathcal{S}}$ is defined over $\mathbb{C} \setminus \mathbb{R}^+$ as

$$m_N(z) = \int \frac{1}{\lambda - z} dF^{\mathbf{B}_N^{\mathcal{S}}}(\lambda) = \frac{1}{N} \operatorname{tr}\left(\mathbf{B}_N^{\mathcal{S}} - z\mathbf{I}_N\right)^{-1} \tag{4}$$

Of importance is the Shannon transform, denoted $\mathcal{V}(x)$, of $\mathbf{B}_N^{\mathcal{S}}$, which we define, for $x > 0$, as

$$\begin{aligned}\mathcal{V}(x) &= \frac{1}{N} \log\det\left(\mathbf{I}_N + \frac{1}{x}\mathbf{B}_N^{\mathcal{S}}\right) &(5)\\ &= \int_0^{+\infty} \log(1+\lambda x^{-1}) dF^{\mathbf{B}_N^{\mathcal{S}}}(\lambda) &(6)\\ &= \int_x^{+\infty} \left(\frac{1}{w} - m_N(-w)\right) dw &(7)\end{aligned}$$

B. Main results

The main results of this work related to random matrix theory come as follows,

- we first present a theorem, namely Theorem 1, which provides a deterministic equivalent of the Stieltjes transform of generalized $\mathbf{B}_N^{\mathcal{S}}$-like matrices, under the assumption that, with growing N and n_k (with non trivial ratio $0 < c = N/n \ll N$) the sequences $\{F^{\mathbf{T}_k}\}_{n_k}$ and $\{F^{\mathbf{R}_k}\}_N$ are tight. This is, we provide an approximation of $m_N(z)$ which does not depend on the realization of the $\mathbf{X}_k$ matrices. The tight sequence assumption allows some degenerated cases of very strong correlation in $\mathbf{R}_k$ and $\mathbf{T}_k$, which exhibit an eigenvalue of order of magnitude N.[6]
- we then provide in Theorem 2 a deterministic equivalent $\mathcal{V}^{(0)}(x)$ of the Shannon transform $\mathcal{V}(x)$ of $\mathbf{B}_N^{\mathcal{S}}$. For this theorem, the assumptions on the $\mathbf{R}_k$ and $\mathbf{T}_k$ matrices are more constraining, since we need to assume here that the maximum number r_N of the eigenvalues of $\mathbf{T}_k$, $\mathbf{R}_k$ which grow unbounded and their maximum

[6]remark, for instance, that the scenario in which the device volume is fixed while the number of antennas N grows inside this volume, is prone to exhibit such degenerated behaviour. This case can therefore be treated here.

value b_N satisfy $\log(1+b_N^2\beta/x)r_N = o(N)$ for some constant β to be defined in Theorem 2. However, this assumption is largely less restrictive than the hypothesis of uniformly (with respect to N) bounded spectral norms of $\mathbf{T}_k$ and $\mathbf{R}_k$. For instance, we can theoretically allow

– the matrices $\mathbf{T}_k$ and $\mathbf{R}_k$ to verify $\operatorname{tr}(\mathbf{R}_k) = O(N)$, $\operatorname{tr}(\mathbf{T}_k) = O(N)$, which is of practical interest here.
– all but a finite number of the eigenvalues of $\mathbf{T}_k$, $\mathbf{R}_k$ to be uniformly bounded, the largest of them growing like e^N. For our specific study, this case is less relevant though.

The major practical interest of Theorems 1 and 2 lies in the possibility to analyze capacity expressions, no longer as stochastic variables depending on the matrices $\mathbf{X}_k$ but as approximations of deterministic quantities. The study of those quantities are in general simpler than the study of the stochastic expressions, even if the deterministic results are actually solutions of involved implicit equations (see Section III). In particular, remember that our problematic introduced in Section I is to study the trade-off 'capacity gain' versus 'cost' of additional transmit/receive antennas. For this reason, the typical figures of performance sought for are the per-(transmit or receive) antenna normalized capacity, sum rate or rate region, i.e. the mean efficiency of an antenna of the transmit/receive array. Those are again related to the Stieltjes and Shannon transform of $\mathbf{B}_N^{\mathcal{S}}$-like matrices. However, we will not provide in this study asymptotic sum-capacity expressions, i.e. N times the Shannon transform, for which asymptotic accuracy of the deterministic equivalents cannot be verified; on this topic, the reader is strongly advised to refer to [27].

In such practical applications as the rate region of BC, the Shannon transform expressions (5) are needed to calculate the corner points in the rate region; because of MAC-BC duality, to determine the BC rate region, one needs to treat the dual MAC uplink problem, see Section IV-A; in order to decouple the influence of antenna correlation and transmit covariance matrices at the transmitters, the correlation matrix $\mathbf{R}_k$ at the k^{th} BC receiver (and then the k^{th} MAC transmitter) is replaced by the product $\mathbf{R}_k^{\frac{1}{2}}\mathbf{P}_k\mathbf{R}_k^{\frac{1}{2}}$, with $\mathbf{R}_k$ translating the channel correlation at the transmitter and $\mathbf{P}_k$ the transmit signal covariance matrix. Finding out the matrices $\mathbf{P}_k$ which maximize (5) is a very difficult problem for finite N, which, up to the authors' knowledge, has not been solved, unless all the $\mathbf{T}_k$ matrices are equal [24]. In Section IV-A, we determine the matrices $\mathbf{P}_k^\star$ which maximize the deterministic equivalent $\mathcal{V}^{(0)}$ of $\mathcal{V}$ in (5). The following points are of noticeable importance,

- the eigenspaces of the capacity-maximizing $\mathbf{P}_k^\star$ matrices (in the deterministic equivalent) coincide respectively with the eigenspaces of the correlation matrices $\mathbf{R}_k$ at the receivers.
- the eigenvalues of the $\mathbf{P}_k^\star$ matrices are solution of a classical optimization problem. They are given by a water-filling process.
- we provide an iterative water-filling algorithm for the eigenvalues of $\mathbf{P}_k$ which, upon convergence, are proved to converge to the optimal eigenvalues of $\mathbf{P}_k^\star$, for all k. This is to say, we do not prove that the algorithm converges (though after intensive simulations, it always converged) but we state and can prove that, if it does, then it converges towards the sought solution.

It is important to understand here that, contrary to [24], the above result does not imply that the capacity-maximizing power allocation in the finite N regime consists in aligning the eigenvectors of the $\mathbf{P}_k$ matrices to those of the transmit correlation matrices; we show that this strategy does optimize the deterministic equivalent though.

III. Mathematical Preliminaries

In this section, we first introduce Theorem 1, which provides a deterministic equivalent for the Stieltjes transform of matrices $\mathbf{B}_N$ of the type (1). The underlying assumptions of this theorem are made as large as possible for mathematical completeness. The Shannon transform of $\mathbf{B}_N$ is then provided in Theorem 2, under tighter assumptions on the matrices $\mathbf{X}_k$, $\mathbf{R}_k$ and $\mathbf{T}_k$.

Theorem 1: Let $K \in \mathbb{N}^*$ be some fixed positive integer. For some $N \in \mathbb{N}^*$, let

$$\mathbf{B}_N = \sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k \mathbf{X}_k^{\mathsf{H}} \mathbf{R}_k^{\frac{1}{2}} + \mathbf{S} \tag{8}$$

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

1) $\mathbf{X}_k = \left(\frac{1}{\sqrt{n_k}} X_{ij}^k\right)$ is $N \times n_k$ with the X_{ij}^k identically distributed complex for all N, i, j, independent for each fixed N, and $\mathrm{E}|X_{11}^k - \mathrm{E}X_{11}^k|^2 = 1$,
2) $\mathbf{R}_k^{\frac{1}{2}}$ is the $N \times N$ Hermitian nonnegative definite square root of the nonnegative definite Hermitian matrix $\mathbf{R}_k$,
3) $\mathbf{T}_k = \mathrm{diag}(\tau_1, \ldots, \tau_{n_k})$ is $n_k \times n_k$, $n_k \in \mathbb{N}^*$, diagonal with $\tau_i \geq 0$,
4) The sequences $\{F^{\mathbf{T}_k}\}_{n_k \geq 1}$ and $\{F^{\mathbf{R}_k}\}_{N \geq 1}$ are tight, i.e. for all $\varepsilon > 0$, there exists $M_0 > 0$ such that $M > M_0$ implies $F^{\mathbf{T}_k}([M, \infty)) < \varepsilon$ and $F^{\mathbf{R}_k}([M, \infty)) < \varepsilon$ for all n_k, N,
5) $\mathbf{S}$ is $N \times N$ Hermitian positive definite,
6) There exists $b > a > 0$ for which

$$a \leq \liminf_N c_k \leq \limsup_N c_N \leq b \tag{9}$$

with $c_k = N/n_k$.

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

$$m_N(z) - m_N^{(0)}(z) \xrightarrow{\text{a.s.}} 0 \tag{10}$$

where

$$m_N^{(0)}(z) = \frac{1}{N} \operatorname{tr} \left(\mathbf{S} + \sum_{k=1}^{K} \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k \tau_k e_k(z)} \mathbf{R}_k - z\mathbf{I}_N \right)^{-1} \tag{11}$$

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

$$e_i(z) = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\mathbf{S} + \sum_{k=1}^{K} \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k \tau_k e_k(z)} \mathbf{R}_k - z\mathbf{I}_N \right)^{-1} \tag{12}$$

such that $\mathrm{sgn}(\Im[e_i(z)]) = \mathrm{sgn}(\Im[z])$.

Moreover, for any $\varepsilon > 0$, the convergence of Equation (10) is uniform over any region of $\mathbb{C}$ bounded by a contour interior to

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

For all N, the function $m_N^{(0)}$ is the Stieltjes transform of a distribution F_N^0. Denoting $F^{\mathbf{B}_N}$ the empirical eigenvalue distribution of $\mathbf{B}_N$, we finally have

$$F^{\mathbf{B}_N} - F_N^0 \Rightarrow 0 \tag{13}$$

weakly as $N \to \infty$.

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

Looser hypothesis will be used in the applications of Theorem 1 provided in Section IV. We will specifically need the corollary hereafter,

Corollary 1: Let $K \in \mathbb{N}$ be some positive integer. For some $N \in \mathbb{N}^*$, let

$$\mathbf{B}_N = \sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k \mathbf{X}_k^{\mathsf{H}} \mathbf{R}_k^{\frac{1}{2}} \tag{14}$$

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

1) $\mathbf{X}_k = \left(\frac{1}{\sqrt{n_k}} X_{ij}^k \right)$ is $N \times n_k$ with joint distribution invariant to right unitary product, where the X_{ij}^k are independent and identically distributed for each i, j, N, with finite fourth order moment.
2) $\mathbf{R}_k^{\frac{1}{2}}$ is the $N \times N$ Hermitian nonnegative definite square root of the nonnegative definite Hermitian matrix $\mathbf{R}_k$,
3) $\mathbf{T}_k$ is an $n_k \times n_k$ nonnegative definite Hermitian matrix,
4) The sequences $\{F^{\mathbf{T}_k}\}_{n_k \geq 1}$ and $\{F^{\mathbf{R}_k}\}_{N \geq 1}$ are tight.
5) There exists $b > a > 0$ for which

$$a \leq \liminf_N c_k \leq \limsup_N c_N \leq b \tag{15}$$

with $c_k = N/n_k$.

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

$$m_N(x) - m_N^{(0)}(x) \xrightarrow{\text{a.s.}} 0 \tag{16}$$

where

$$m_N^{(0)}(x) = \frac{1}{N} \operatorname{tr} \left(\sum_{k=1}^{K} \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k \tau_k e_k(z)} \mathbf{R}_k - x\mathbf{I}_N \right)^{-1} \tag{17}$$

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

$$e_i(z) = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{k=1}^{K} \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k \tau_k e_k(z)} \mathbf{R}_k - z\mathbf{I}_N \right)^{-1} \tag{18}$$

such that $\operatorname{sgn}(\Im[e_i(z)]) = \operatorname{sgn}(\Im[z])$.

Proof: Since the $\mathbf{X}_k$'s are invariant to right-unitary product, the joint distribution of $\mathbf{X}_k\mathbf{U}$ coincides 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 $\mathbf{T}_k$'s can be taken non diagonal nonnegative definite Hermitian and the result of Theorem 1 still holds true. ■

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 be some strictly positive real number. Let $\mathbf{B}_N$ be a random Hermitian matrix as defined in Corollary 1 with the following additional assumptions

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

$$\max_{1\leq k\leq K} \max(\lambda_{r_N+1}^{\mathbf{T}_k}, \lambda_{r_N+1}^{\mathbf{R}_k}) \leq \alpha \tag{19}$$

where $\lambda_1^{\mathbf{X}} \geq \ldots \geq \lambda_N^{\mathbf{X}}$ denote the ordered eigenvalues of the $N \times N$ matrix $\mathbf{X}$.

2) denoting b_N an upper-bound on the spectral norm of the $\mathbf{T}_k$'s and $\mathbf{R}_k$'s, $k \in \{1, \ldots, K\}$, and β some real such that $\beta > K(b/a)(1+\sqrt{a})^2$, then $a_N = b_N^2\beta$ satisfies

$$r_N \log(1 + a_N/x) = o(N) \tag{20}$$

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

$$\mathcal{V}(x) - \mathcal{V}^{(0)}(x) \xrightarrow{\text{a.s.}} 0 \tag{21}$$

where

$$\begin{aligned}\mathcal{V}^{(0)}(x) =& \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\sum_{k=1}^{K}\mathbf{R}_k\int\frac{\tau_k}{1+c_ke_k(-x)\tau_k}dF^{\mathbf{T}_k}(\tau_k)\right)\\ &+\sum_{k=1}^{K}\frac{1}{c_k}\int\log\left(1+c_ke_k(-x)\tau_k\right)dF^{\mathbf{T}_k}(\tau_k)\\ &+x\cdot m_N^{(0)}(-x)-1\end{aligned} \tag{22}$$

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

Remark 1: Remind that, since Gaussian matrices are right-unitary invariant with entries of finite fourth moment, the special case of $\mathbf{X}_k$ matrices with Gaussian i.i.d. entries fits the requirements of Corollary 1 and Theorem 2. In the practical applications of the following section, we shall always consider this hypothesis.

Remark 2: Note that this last result is consistent with the work of Telatar [7] and Tulino [2] when $K=1$ and the transmission channels are respectively Gaussian with no correlation or with separable variance profile. The latter is easy to observe from the form of Equation (206) which, up to a Stieltjes-η variable change, is exactly the expression in [2] when $K=1$ with separable variance profile.

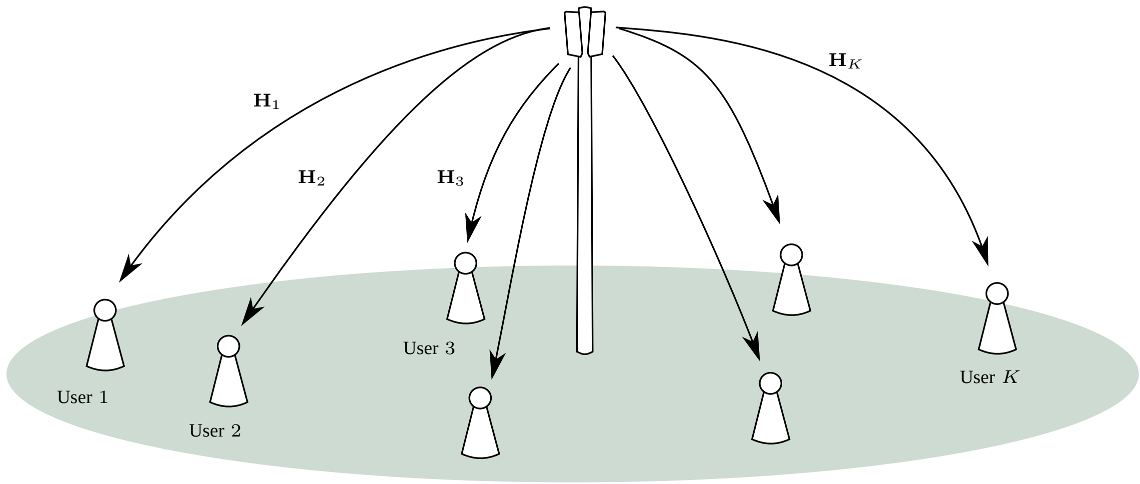

Fig. 1. Downlink scenario in multi-user broadcast channel

IV. APPLICATIONS

In this section, we provide applications of Theorems 1 and 2 to two studies in the field of wireless communications. First, in Section IV-A, we derive an expression of the rate region of multi-antenna multiple access and broadcast channels for given correlation matrices at the transmit and receive devices. An iterative power allocation algorithm is then introduced which is proved, upon convergence, to maximize the deterministic equivalent of the rate region corner points. Then, in Section IV-B, we provide an analytical expression of the capacity of the single user and MMSE decoders in wireless MIMO networks with inter-cell interference. For the single user decoder, a derivation of the optimal power allocation for the sum rate maximization is also provided.

A. Rate Region of Broadcast Channels

1) System Model: Consider a wireless multi-user channel with $K \geq 1$ users indexed from 1 to K, controlled by a single base station. User k is equipped with n_k antennas while the base station is equipped with N antennas. We additionally denote $c_k = N/n_k$. This situation is depicted in Figure 1.

Even if we will longly discuss the MAC channel between the K users and the base station, our prior objective is to characterize the BC channel between the base station and the users. For this, we denote $\mathbf{s} \in \mathbb{C}^N$, $\mathrm{E}[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{P}$, the signal transmitted by the base station, with power constraint $\mathrm{tr}(\mathbf{P}) \leq P$, $P > 0$; $\mathbf{y}_k \in \mathbb{C}^{n_k}$ the signal received by user k and $\mathbf{n}_k \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_{n_k})$ the noise vector received by user k.[7] The fading MIMO channel between the base station and user k is denoted $\mathbf{H}_k \in \mathbb{C}^{N \times n_k}$. Moreover we assume that $\mathbf{H}_k$ has a separable variance profile,

[7]up to a scaling of the power constraints of the individual users, setting the same noise variance σ^2 on each receive antenna for every user does not restrict the generality and simplifies the theoretical expressions.

i.e. can be decomposed as

$$\mathbf{H}_k = \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k^{\frac{1}{2}} \tag{23}$$

with $\mathbf{R}_k \in \mathbb{C}^{n_k \times n_k}$ the (Hermitian) correlation matrix at receiver k with respect to the channel $\mathbf{H}_k$, $\mathbf{T}_k \in \mathbb{C}^{N \times N}$ the correlation matrix at the base station for link $\mathbf{H}_k$ and $\mathbf{X}_k \in \mathbb{C}^{n_k \times N}$ a random matrix with Gaussian independent entries of variance $1/n_k$. In the Kronecker model, $\mathbf{T}_k$ and $\mathbf{R}_k$ satisfy $\mathrm{tr}(\mathbf{T}_k) = N$ and $\mathrm{tr}(\mathbf{R}_k) = n_k$.

With the assumptions above, the downlink communication model unfolds

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{s} + \mathbf{n} \tag{24}$$

Denoting equivalently $\mathbf{s}_k$ the signal transmitted in the dual uplink by user k, such that $\mathrm{E}[\mathbf{s}_k \mathbf{s}_k^{\mathsf{H}}] = \mathbf{P}_k$, $\mathrm{tr}(\mathbf{P}_k) \leq P_k$, $\mathbf{y}$ and $\mathbf{n}$ the signal and the noise received by the base station, we have the converse uplink model

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{H}_k^{\mathsf{H}} \mathbf{s}_k + \mathbf{n}_k \tag{25}$$

In the following, we will derive the BC rate region by means of the MAC-BC duality [20]. We then consider first the achievable MAC rate region.

2) MAC Rate Region: The (per-receive antenna normalized) rate region $\mathcal{C}_{\mathrm{MAC}}(P_1, \ldots, P_K; \mathbf{H}^{\mathsf{H}})$ of the MAC channel $\mathbf{H}^{\mathsf{H}}$ under respective transmit power constraints $P_1, \ldots, P_K$ for users 1 to K respectively and compound channel $\mathbf{H}^{\mathsf{H}} = [\mathbf{H}_1^{\mathsf{H}} \ldots \mathbf{H}_K^{\mathsf{H}}]$, is given in [22], and reads

$$\mathcal{C}_{\mathrm{MAC}}(P_1, \ldots, P_K; \mathbf{H}^{\mathsf{H}}) = \bigcup_{\substack{\mathrm{tr}(\mathbf{P}_i) \leq P_i \\ \mathbf{P}_i \geq 0 \\ i=1,\ldots,K}} \left\{ \{R_i, 1 \leq i \leq K\} : \sum_{i \in \mathcal{S}} R_i \leq \frac{1}{N} \log \left| \mathbf{I} + \frac{1}{\sigma^2} \sum_{i \in \mathcal{S}} \mathbf{H}_i^{\mathsf{H}} \mathbf{P}_i \mathbf{H}_i \right|, \forall \mathcal{S} \subset \{1, \ldots, K\} \right\} \tag{26}$$

For any set $\mathcal{S} \subset \{1, \ldots, K\}$, thanks to Theorem 2, we have approximately, for N, n_k large,

$$\begin{aligned} \frac{1}{N} \log \left| \mathbf{I}_N + \frac{1}{\sigma^2} \sum_{i \in \mathcal{S}} \mathbf{H}_i^{\mathsf{H}} \mathbf{P}_i \mathbf{H}_i \right| &= \frac{1}{N} \log\det \left(\mathbf{I}_N + \frac{1}{\sigma^2} \sum_{k \in \mathcal{S}} \mathbf{T}_k \int \frac{r_k}{1 + c_k e_k(-\sigma^2) r_k} dF^{\mathbf{R}_k \mathbf{P}_k}(r_k) \right) \\ &\quad + \sum_{k \in \mathcal{S}} \frac{1}{c_k} \int \log\left(1 + c_k e_k(-\sigma^2) r_k\right) dF^{\mathbf{R}_k \mathbf{P}_k}(r_k) \\ &\quad + \sigma^2 \cdot m_{\mathcal{S}}^{(0)}(-\sigma^2) - 1 \end{aligned} \tag{27}$$

where

$$m_{\mathcal{S}}(-\sigma^2) = \frac{1}{N} \mathrm{tr} \left(\sum_{k \in \mathcal{S}} \int \frac{r_k dF^{\mathbf{R}_k \mathbf{P}_k}(r_k)}{1 + c_k r_k e_k(-\sigma^2)} \mathbf{T}_k + \sigma^2 \mathbf{I}_N \right)^{-1} \tag{28}$$

and the e_i's satisfy

$$e_i(-\sigma^2) = \frac{1}{N} \mathrm{tr}\, \mathbf{T}_i \left(\sum_{k \in \mathcal{S}} \int \frac{r_k dF^{\mathbf{R}_k \mathbf{P}_k}(r_k)}{1 + c_k r_k e_k(-\sigma^2)} \mathbf{T}_k + \sigma^2 \mathbf{I}_N \right)^{-1} \tag{29}$$

From these equations, the complete MAC capacity region can be described. To go further, we first need the following result, *Proposition 1:* If any of the correlation matrices $\mathbf{R}_k$, $k \in \mathcal{S}$ is invertible, then the right-hand side of the capacity approximate (27) is a strictly concave function of $\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}$.

Proof: The proof of Proposition 1 is provided in Appendix C. ■

From Proposition 1, we immediately prove that the $|\mathcal{S}|$-ary set of capacity-achieving matrices $(\mathbf{P}_1^\star, \ldots, \mathbf{P}_{|\mathcal{S}|}^\star)$ is unique, provided that any of the $\mathbf{R}_k$'s is invertible. In a very similar way as in [27], we will show that the matrices $\mathbf{P}_k^\star$, $k \in \{1, \ldots, |\mathcal{S}|\}$, have the following properties: (i) their eigenspaces are the same as those of their corresponding $\mathbf{R}_k$'s, (ii) their eigenvalues are the solutions of a classical water-filling problem.

Proposition 2: For every $k \in \mathcal{S}$, denote $\mathbf{R}_k = \mathbf{U}_k \mathbf{D}_k \mathbf{U}_k^{\mathsf{H}}$ the spectral decomposition of $\mathbf{R}_k$ with $\mathbf{U}_k$ unitary and $\mathbf{D}_k = \operatorname{diag}(r_{k1}, \ldots, r_{kn_k})$ diagonal. Then the capacity-achieving signal covariance matrices $\mathbf{P}_1^\star, \ldots, \mathbf{P}_{|\mathcal{S}|}^\star$ satisfy

1) $\mathbf{P}_k^\star = \mathbf{U}_k \mathbf{Q}_k^\star \mathbf{U}_k^{\mathsf{H}}$, with $\mathbf{Q}_k^\star$ diagonal; i.e. the eigenspace of $\mathbf{P}_k^\star$ is the same as the eigenspace of $\mathbf{R}_k$.
2) denoting $\delta_k^\star = \delta_k(-\sigma^2, \mathbf{P}_k^\star)$ and $e_k^\star = e_k(-\sigma^2, \mathbf{P}_k^\star)$, the i^{th} diagonal entry $q_{ki}^\star$ of $\mathbf{Q}_k^\star$ satisfies
$$q_{ki}^\star = \left(\mu_k - \frac{1}{c_k e_k^\star r_{ki}}\right)^+ \tag{30}$$
where the μ_k's are evaluated such that $\operatorname{tr}(\mathbf{Q}_k) = P_k$.

Proof: The proof of Proposition 2 recalls the proof from Najim, Prop. 4 in [27]. We essentially need to show that, at point $(\delta_1^\star, \ldots, \delta_{|\mathcal{S}|}^\star, e_1^\star, \ldots, e_{|\mathcal{S}|}^\star)$, the derivative of (27) along any $\mathbf{Q}_k$ is the same whether the $\delta_k^\star$'s and the $e_k^\star$'s are fixed or vary with $\mathbf{Q}_k$. In other words, using the form (206) for the capacity, let us define the functions
$$\begin{aligned}\mathcal{V}^{(0)}(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}) = &\sum_{k\in\mathcal{S}} \frac{1}{N} \log\det\left(\mathbf{I}_{n_k} + c_k e_k \mathbf{R}_k \mathbf{P}_k\right) \\ &+ \frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k\in\mathcal{S}} \delta_k \mathbf{T}_k\right) \\ &- \sigma^2 \sum_{k=1}^{K} \delta_k(-\sigma^2) e_k(-\sigma^2)\end{aligned} \tag{31}$$
where
$$e_i = e_i(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}) = \frac{1}{N} \operatorname{tr} \mathbf{T}_i \left(\sigma^2 \left[\mathbf{I}_N + \sum_{k\in\mathcal{S}} \delta_k \mathbf{T}_k\right]\right)^{-1} \tag{32}$$
$$\delta_i = \delta_i(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}) = \frac{1}{n_i} \operatorname{tr} \mathbf{R}_i \mathbf{P}_i \left(\sigma^2 \left[\mathbf{I}_{n_i} + c_i e_i(z) \mathbf{R}_i \mathbf{P}_i\right]\right)^{-1} \tag{33}$$
and $V : (\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}, \delta_1, \ldots, \delta_{|\mathcal{S}|}, e_1, \ldots, e_{|\mathcal{S}|}) \mapsto \mathcal{V}^{(0)}(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|})$. Then we need only prove that, for all $k \in \mathcal{S}$,
$$\frac{\partial V}{\partial \delta_k}(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}, \delta_1^\star, \ldots, \delta_{|\mathcal{S}|}^\star, e_1^\star, \ldots, e_{|\mathcal{S}|}^\star) = 0 \tag{34}$$
$$\frac{\partial V}{\partial e_k}(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}, \delta_1^\star, \ldots, \delta_{|\mathcal{S}|}^\star, e_1^\star, \ldots, e_{|\mathcal{S}|}^\star) = 0 \tag{35}$$
Remark then that
$$\frac{\partial V}{\partial \delta_k}(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}, \delta_1, \ldots, \delta_{|\mathcal{S}|}, e_1, \ldots, e_{|\mathcal{S}|}) = \frac{1}{N} \operatorname{tr}\left[\left(\mathbf{I} + \sum_{i\in\mathcal{S}} \delta_i \mathbf{T}_i\right)^{-1} \mathbf{T}_k\right] - \sigma^2 e_k \tag{36}$$
$$\frac{\partial V}{\partial e_k}(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}, \delta_1, \ldots, \delta_{|\mathcal{S}|}, e_1, \ldots, e_{|\mathcal{S}|}) = c_k \frac{1}{N} \operatorname{tr}\left[\left(\mathbf{I} + c_k e_k \mathbf{R}_i \mathbf{P}_i\right)^{-1} \mathbf{R}_k \mathbf{P}_k\right] - \sigma^2 \delta_k \tag{37}$$

At initialization, for all $k \in \mathcal{S}$, $\mathbf{Q}_k = \frac{Q_k}{n_k}\mathbf{I}_{n_k}$, $\delta_k = 1$, $e_k = 1$.
while the $\mathbf{P}_k$'s have not converged **do**
  for $k \in \mathcal{S}$ **do**
    Set (δ_k, e_k) as solution of (32), (33)
    for $i = 1\ldots, n_k$ **do**
      Set $q_{k,i} = \left(\mu_k - \frac{1}{c_k e_k r_{ki}}\right)^+$, with μ_k such that $\operatorname{tr}\mathbf{Q}_k = P_k$.
    end for
  end for
end while

TABLE I

ITERATIVE WATER-FILLING ALGORITHM FOR POWER OPTIMIZATION

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 $\mathbf{P}_1^\star, \ldots, \mathbf{P}_{|\mathcal{S}|}^\star$ whenever $e_k = e_k^\star$ and $\delta_k = \delta_k^\star$.

When, for all k, $e_k = e_k^\star$, $\delta_k = \delta_k^\star$, the maximum of V over the $\mathbf{P}_k$'s is then obtained by maximizing the expressions $\log\det(\mathbf{I}_{n_k} + c_k e_k^\star \mathbf{R}_k \mathbf{P}_k)$ over $\mathbf{P}_k$. By Hadamard's inequality,

$$\det(\mathbf{I}_{n_k} + c_k e_k^\star \mathbf{R}_k \mathbf{P}_k) \leq \prod_{i=1}^{n_k} \|(\mathbf{I}_{n_k} + c_k e_k^\star \mathbf{R}_k \mathbf{P}_k)_i\|_2 \tag{38}$$

where, only here, we denote $(\mathbf{X})_i$ the i^{th} column of matrix $\mathbf{X}$. The equality is obtained if and only if $\mathbf{I}_{n_k} + c_k e_k^\star \mathbf{R}_k \mathbf{P}_k$ 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 $\operatorname{tr}(\mathbf{Q}_k) = P_k$ are solutions of the classical optimization problem under constraint,

$$\sup_{\substack{\mathbf{Q}_k \\ \operatorname{tr}(\mathbf{Q}_k) \leq P_k}} \log\det\left(\mathbf{I}_{n_k} + c_k e_k^\star \mathbf{Q}_k \mathbf{D}_k\right) \tag{39}$$

whose solution is given by the classical water-filling algorithm. Hence (30). ■

We then propose an iterative water-filling algorithm to obtain the power allocation policy which maximizes the right-hand side of (27). This is provided in Table I.

In [27], it is proven that the convergence of this algorithm ensures its convergence towards the only capacity-achieving power allocation policy. However, as recalled in [27], it is difficult to prove the convergence of the algorithm in Table I in general. Nonetheless, extensive simulations initialized with $\delta_k = 1$, $e_k = 1$, for all k, showed the algorithm in Table I always converged.

Remark 3: It is important to stress out the fact that we did not prove that the classical water-filling solution maximizes the true capacity for N finite. We merely showed that, for large N, the true capacity expression on the left-hand side of (27) is closely approximated by an expression (the right-hand side of (27)) whose maximum is achieved by the iterative water-filling algorithm of Table I. The optimality of the water-filling solution for finite N is proven, e.g. in [24], for the special case when all $\mathbf{R}_k$ are equal.

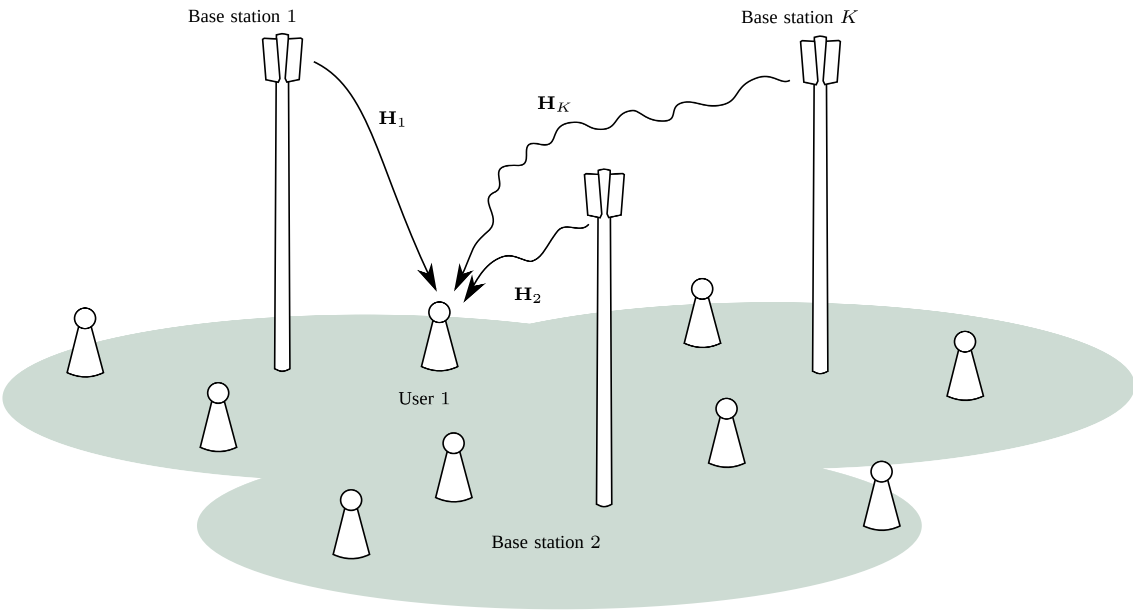

Fig. 2. Downlink multi-cell scenario

3) BC Rate Region: The capacity region of the broadcast multi-antenna channel has been recently shown [23] to be achieved by the dirty paper coding (DPC) algorithm. This region $\mathcal{C}_{\mathrm{BC}}(P;\mathbf{H})$, for a transmit power constraint P over the compound channel $\mathbf{H}$, is shown by duality arguments to be the set [20]

$$\mathcal{C}_{\mathrm{BC}}(P;\mathbf{H}) = \bigcup_{\sum_{k=1}^{K} P_k \leq P} \mathcal{C}_{\mathrm{MAC}}(P_1,\ldots,P_K;\mathbf{H}^{\mathsf{H}}) \tag{40}$$

which is easily obtained from Equation (26).

B. Multi-User MIMO

1) Signal Model: In this section we study the per-antenna rate performance of wireless networks including a multi-antenna transmitter and a multi-antenna receiver, the latter of which is interfered by several multi-antenna transmitters. This scheme is well-suited to multi-cell wireless networks with orthogonal intra-cell and interfering inter-cell transmissions, both in downlink and in uplink. In particular, this encompasses

- multi-cell uplink: consider a K-cell network; the base station of a cell indexed by $i \in \{1,\ldots,K\}$ receives data from a terminal user in this cell[8] and is interfered by $K-1$ users transmitting on the same physical resource from remote cells indexed by $j \in \{1,\ldots,K\}$, $j \neq i$.
- multi-cell downlink: the user being allocated a given time/frequency resource in a cell indexed by $i \in \{1,\ldots,K\}$ receives data from its dedicated base-station and is interfered by $K-1$ base stations in neighboring cells indexed by $j \in \{1,\ldots,K\}$, $j \neq i$. This situation is depicted in Figure 2.

[8]this user is allocated a given time/frequency resource, which is orthogonal to time/frequency resources of the other users in the cell; e.g. the multi-access protocol is OFDMA.

In the following, in order not to confuse both scenarios, only the downlink scheme is considered. However, one must keep in mind that the provided results can easily be adapted to the uplink case.

Consider a wireless mobile network with $K \geq 1$ cells indexed from 1 to K, controlled by *non-physically connected* base stations. We assume that, on a particular time or frequency resource, each base station serves only one user; therefore the base station and the user of cell j will also be indexed by j. Without loss of generality, we focus our attention on user 1, equipped with $N \gg K$ antennas and hereafter referred to as *the user* or *the receiver*. Every base station $j \in \{1, \ldots, K\}$ is equipped with $n_j \gg K$ antennas. Similarly to previous sections, we denote $c_j = N/n_j$.

Denote $\mathbf{s}_j \in \mathbb{C}^{n_j}$, $\mathbf{E}[\mathbf{s}_j \mathbf{s}_j^{\mathsf{H}}] = \mathbf{I}_{n_j}$, the signal transmitted by base station j, $\mathbf{y} \in \mathbb{C}^N$ and $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I}_N)$ the signal and noise vectors received by the user. The fading MIMO channel between base station j and the user is denoted $\mathbf{H}_j \in \mathbb{C}^{N \times n_j}$. Moreover, we assume that $\mathbf{H}_j$ is Gaussian with a separable variance profile, given by Equation (23).

With the assumptions above, the communication model unfolds

$$\mathbf{y} = \mathbf{H}_1 \mathbf{s}_1 + \sum_{j=2}^{K} \mathbf{H}_j \mathbf{s}_j + \mathbf{n} \tag{41}$$

where $\mathbf{s}_1$ is the useful signal (from base station 1) and $\mathbf{s}_j$, $j \geq 2$, constitute interfering signals.

2) Single User Decoding:

a) Uniform Power Allocation: If the receiving user considers the signals from the $K-1$ interfering transmitters as Gaussian noise with a known variance pattern[9], then base station 1 can transmit with arbitrarily low decoding error at a per-receive antenna rate $\mathcal{C}_{\mathrm{SU}}(\sigma^2)$ given by

$$\mathcal{C}_{\mathrm{SU}}(\sigma^2) = \frac{1}{N} \log |\mathbf{I}_N + \frac{1}{\sigma^2} \sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}}| - \frac{1}{N} \log |\mathbf{I}_N + \frac{1}{\sigma^2} \sum_{j=2}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}}| \tag{42}$$

Assume that N and the n_i, $i \in \{1, \ldots, K\}$, are large compared to K and such that the maximum eigenvalues of $\mathbf{R}_i$ or $\mathbf{T}_i$ are bounded away from N. From Corollary 1, we define the functions $m^{i,(0)}$ as the approximated Stieltjes transforms of $\sum_{j=i}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}}$, $i \in \{1, 2\}$,

$$m^{i,(0)}(z) = \frac{1}{N} \operatorname{tr} \left(\sum_{k=1}^{K} \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k \tau_k e_k^i(z)} \mathbf{R}_k - z \mathbf{I}_N \right)^{-1} \tag{43}$$

where, for all $j \in \{1, \ldots, K\}$, $e_j^i(z)$ is solution of the fixed-point equation

$$e_j^i(z) = \frac{1}{N} \operatorname{tr} \mathbf{R}_j \left(\sum_{k=i}^{K} \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k \tau_k e_k(z)} \mathbf{R}_k - z \mathbf{I}_N \right)^{-1} \tag{44}$$

[9]in practice, the noise variance does not require the knowledge of the other base station to user channels, as it can be inferred from data sensing in idle mode.

From Theorem 2, we then have approximately

$$\begin{aligned}\mathcal{C}_{\rm SU}(\sigma^2) =& \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\sum_{k=1}^{K}\mathbf{R}_k\int\frac{\tau_k}{1+c_k e_k^1(-\sigma^2)\tau_k}dF^{\mathbf{T}_k}(\tau_k)\right)\\ &-\frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{\sigma^2}\sum_{k=2}^{K}\mathbf{R}_k\int\frac{\tau_k}{1+c_k e_k^2(-\sigma^2)\tau_k}dF^{\mathbf{T}_k}(\tau_k)\right)\\ &+\sum_{k=1}^{K}\frac{1}{c_k}\int\log\left(1+c_k e_k^1(-\sigma^2)\tau_k\right)dF^{\mathbf{T}_k}(\tau_k)\\ &-\sum_{k=2}^{K}\frac{1}{c_k}\int\log\left(1+c_k e_k^2(-\sigma^2)\tau_k\right)dF^{\mathbf{T}_k}(\tau_k)\\ &+\sigma^2\cdot\left[m^{1,(0)}(-\sigma^2)-m^{2,(0)}(-\sigma^2)\right]\end{aligned}\tag{45}$$

b) Power Optimization for Single-User Decoding: In this section we wish to perform power allocation so to maximize the single-user decoding capacity $\mathcal{C}_{\rm SU}(\sigma^2)$ along $\mathbf{P}_1$, the signal variance at the main transmitter. We therefore replace the matrices $\mathbf{T}_j$ by $\mathbf{T}_j^{\frac{1}{2}}\mathbf{P}_j\mathbf{T}_j^{\frac{1}{2}}$ in Equation (45).

Let us rewrite $\mathcal{C}_{\rm SU}(\sigma^2)$ like

$$\mathcal{C}_{\rm SU}(\sigma^2) = \frac{1}{N}\log\left|\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{A}^{-\frac{1}{2}}\mathbf{R}_1^{\frac{1}{2}}\mathbf{X}_1\mathbf{T}_1^{\frac{1}{2}}\mathbf{P}_1\mathbf{T}_1^{\frac{1}{2}}\mathbf{X}_1^{\sf H}\mathbf{R}_1^{\frac{1}{2}}\mathbf{A}^{-\frac{1}{2}}\right|\tag{46}$$

with $\mathbf{A} = \mathbf{I}_N + \frac{1}{\sigma^2}\sum_{j>1}\mathbf{H}_j\mathbf{P}_j\mathbf{H}_j^{\sf H}$.

According to [25], the optimal power allocation strategy in such a model is $\mathbf{P}_1 = \sum_{i=1}^{n_1}p_i\mathbf{v}_i^{\sf H}\mathbf{v}_i$ with $\mathbf{V} = [\mathbf{v}_1,\ldots,\mathbf{v}_n]$ the eigenvector matrix in the spectral decomposition of $\mathbf{T}_1 = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{\sf H}$, $\boldsymbol{\Lambda}$ diagonal, where the p_i's satisfy

$$\begin{cases} p_i = 0 & (1/\alpha_i)-1 \leq \frac{1}{n_1}\sum_{l=1}^{n_1}(1-\alpha_l)\\ p_i = \frac{1-\alpha_i}{\frac{1}{n_1}\sum_{l=1}^{n_1}(1-\alpha_l)} & \text{otherwise}\end{cases}\tag{47}$$

for α_i defined as

$$\alpha_i = \left(1+\frac{1}{\sigma^2}\mathbf{h}_i^{\sf H}\mathbf{A}^{-\frac{1}{2}}\left[\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{A}^{-\frac{1}{2}}\mathbf{H}_{-i}\mathbf{P}_1\mathbf{H}_{-i}^{\sf H}\mathbf{A}^{-\frac{1}{2}}\right]^{-1}\mathbf{A}^{-\frac{1}{2}}\mathbf{h}_i\right)^{-1}\tag{48}$$

$\mathbf{h}_i$ is the i^{th} column of $\mathbf{H}_1$, and $\mathbf{H}_{-i}$ is $\mathbf{H}_1$ with column i removed.

Denoting $\mathbf{h}_i = \mathbf{R}_1^{\frac{1}{2}}\mathbf{x}_i$, we have $\mathbf{x}_i$ centered Gaussian with covariance $T_{1_{ii}}/n_1\mathbf{I}_N$ and independent of $\mathbf{R}_1^{\frac{1}{2}}\mathbf{A}^{-\frac{1}{2}}\left[\mathbf{I}_N + \frac{1}{\sigma^2}\mathbf{A}^{-\frac{1}{2}}\mathbf{H}_{-i}\mathbf{P}_1\mathbf{H}_{-i}^{\sf H}\mathbf{A}^{-\frac{1}{2}}\right]^{-1}\mathbf{A}^{-\frac{1}{2}}\mathbf{R}_1^{\frac{1}{2}}$. Therefore, asymptotically on N, from Lemma 7,

$$\alpha_i = \left(1+\frac{T_{1_{ii}}}{\sigma^2 N}\operatorname{tr}\mathbf{R}_1\left[\mathbf{A}+\frac{1}{\sigma^2}\mathbf{H}_{-i}\mathbf{P}_1\mathbf{H}_{-i}^{\sf H}\right]^{-1}\right)^{-1}\tag{49}$$

$$= \left(1+\frac{T_{1_{ii}}}{\sigma^2 N}\operatorname{tr}\mathbf{R}_1\left[\mathbf{A}+\frac{1}{\sigma^2}\mathbf{H}_1\mathbf{P}_1\mathbf{H}_1^{\sf H}\right]^{-1}\right)^{-1}\tag{50}$$

$$= \left(1+\frac{T_{1_{ii}}}{\sigma^2 N}\operatorname{tr}\mathbf{R}_1\left[\mathbf{I}_N+\frac{1}{\sigma^2}\sum_{j=1}^{K}\mathbf{H}_j\mathbf{P}_j\mathbf{H}_j^{\sf H}\right]^{-1}\right)^{-1}\tag{51}$$

$$= \left(1+T_{1_{ii}}e_1^1(-\sigma^2)\right)^{-1}\tag{52}$$

This leads to the power allocation

$$\begin{cases} p_i = 0 & T_{1_{ii}} e_1^1(-\sigma^2) \leq \frac{1}{n_1}\sum_{l=1}^{n_1} 1 - \left(1 + T_{1_{ll}} e_1^1(-\sigma^2)\right)^{-1} \\ p_i = \frac{1-\left(1+T_{1_{ii}} e_1^1(-\sigma^2)\right)^{-1}}{\frac{1}{n_1}\sum_{l=1}^{n_1} 1-\left(1+T_{1_{ll}} e_1^1(-\sigma^2)\right)^{-1}} & \text{otherwise} \end{cases} \tag{53}$$

In particular, if the diagonal entries of $\mathbf{T}_1$ are all equal (as is often the case in practice), then the optimal power allocation policy is the trivial equal power allocation.

3) MMSE Decoding: Achieving $\mathcal{C}_{\mathrm{SU}}$ requires non-linear processing at the receiver, such as successive MMSE interference cancellation techniques. A suboptimal linear technique, the MMSE decoder, is often used instead.

The communication model in this case reads

$$\mathbf{y} = \left(\sum_{j=1}^{k} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} + \sigma^2 \mathbf{I}_N\right)^{-1} \mathbf{H}_1^{\mathsf{H}} \left(\sum_{j=1}^{k} \mathbf{H}_j \mathbf{s}_j + \mathbf{n}\right) \tag{54}$$

where $\left(\sum_{j=1}^{k} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} + \sigma^2 \mathbf{I}_N\right)^{-1} \mathbf{H}_1^{\mathsf{H}}$ is the MMSE linear filter at the receiver. Each entry of $\mathbf{y}$ will be processed individually.

This technique makes it possible to transmit data reliably at any rate inferior to the per-antenna MMSE capacity $\mathcal{C}_{\mathrm{MMSE}}$,

$$\mathcal{C}_{\mathrm{MMSE}}(\sigma^2) = \frac{1}{N}\sum_{i=1}^{n_1} \log(1+\gamma_i) \tag{55}$$

where, denoting $\mathbf{h}_j \in \mathbb{C}^{n_j}$ the j^{th} column of $\mathbf{H}_1$ and $\mathbf{R}_1^{\frac{1}{2}} \mathbf{x}_j = \mathbf{h}_j$, the SINR γ_i expresses as

$$\gamma_i = \frac{\mathbf{h}_i^{\mathsf{H}} \left(\sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} + \sigma^2 \mathbf{I}_N\right)^{-1} \mathbf{h}_i}{1 - \mathbf{h}_i^{\mathsf{H}} \left(\sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} + \sigma^2 \mathbf{I}_N\right)^{-1} \mathbf{h}_i} \tag{56}$$

$$= \mathbf{h}_i^{\mathsf{H}} \left(\sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} - \mathbf{h}_i \mathbf{h}_i^{\mathsf{H}} + \sigma^2 \mathbf{I}_N\right)^{-1} \mathbf{h}_i \tag{57}$$

$$= \mathbf{x}_i^{\mathsf{H}} \mathbf{R}_i^{\frac{1}{2}} \left(\sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} - \mathbf{h}_i \mathbf{h}_i^{\mathsf{H}} + \sigma^2 \mathbf{I}_N\right)^{-1} \mathbf{R}_i^{\frac{1}{2}} \mathbf{x}_i \tag{58}$$

where Equation (57) comes from a direct application of the matrix inversion lemma. With these notations, $\mathbf{x}_i$ has i.i.d. complex Gaussian entries with variance $T_{1_{ii}}/n_i$ and the inner matrix of the right-hand side of (58) is independent of $\mathbf{x}_i$ (since the entries of $\mathbf{H}_1 \mathbf{H}_1^{\mathsf{H}} - \mathbf{h}_i \mathbf{h}_i^{\mathsf{H}}$ are independent of the entries $\mathbf{h}_i$). Applying Lemma 7, for N large,

$$\gamma_i = \frac{T_{1_{ii}}}{N} \operatorname{tr} \mathbf{R}_1 \left(\sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} - \mathbf{h}_i \mathbf{h}_i^{\mathsf{H}} + \sigma^2 \mathbf{I}_N\right)^{-1} \tag{59}$$

From Lemma 5, the rank 1 perturbation $(-\mathbf{h}_i \mathbf{h}_i^{\mathsf{H}})$ does not affect asymptotically the trace in (59). Therefore, approximately,

$$\gamma_i = \frac{T_{1_{ii}}}{N} \operatorname{tr} \mathbf{R}_1 \left(\sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} + \sigma^2 \mathbf{I}_N\right)^{-1} \tag{60}$$

Noting that $e_1^1(z)$ in Equation (44) corresponds to the normalized trace in Equation (60), we finally have the compact expression for $\mathcal{C}_{\mathrm{MMSE}}$,

$$\mathcal{C}_{\mathrm{MMSE}}(\sigma^2) = \frac{1}{N}\sum_{i=1}^{n_1} \log\left(1 + T_{1_{ii}} e_1^1(-\sigma^2)\right) \tag{61}$$

In practice, when no specific power allocation strategy is applied and when no exotic transmit correlation is present, $T_{1_{ii}} = P/n_k$ the average power per transmit symbol, and the capacity becomes $\mathcal{C}_{\mathrm{MMSE}} = \frac{1}{c_1} \cdot \log(1 + P/n_k \cdot e_1(-\sigma^2))$.

V. Simulations and Results

In the following, we apply the results obtained in Sections IV-A and IV-B for the broadcast channel rate region and the multi-user MIMO cases respectively.

A. BC Rate Region

First, we provide simulation results in the context of a two-user broadcast channel, with $N = 8$ transmit antennas and $n_1 = n_2 = 4$ receive antennas, all placed in linear arrays. The distance d^{T} between subsequent transmit antennas is such that $d^{\mathrm{T}}/\lambda = 10$, λ the signal median wavelength; between adjacent receive antennas, the distance d^{R} is the same for users 1 and 2 and such that $d^{\mathrm{R}}/\lambda = 1/4$. The variance profile is generated thanks to Jakes' model, with privileged directions of signal departure and arrival. For instance, the entry (a, b) of matrix $\mathbf{T}_1$ is

$$T_{1_{ab}} = \int_{\theta_{\mathrm{min}}^{(\mathbf{T}_1)}}^{\theta_{\mathrm{max},}^{(\mathbf{T}_1)}} \exp\left(2\pi i |a-b| \frac{d^{\mathrm{T}}}{\lambda}\cos(\theta)\right) d\theta \tag{62}$$

where, in our case, with obvious notations, $\theta_{\mathrm{min}}^{(\mathbf{R}_1)} = 0$, $\theta_{\mathrm{max}}^{(\mathbf{R}_1)} = \pi/2$, $\theta_{\mathrm{min}}^{(\mathbf{R}_2)} = \pi/3$, $\theta_{\mathrm{max}}^{(\mathbf{R}_2)} = 5\pi/6$, $\theta_{\mathrm{min}}^{(\mathbf{T}_1)} = 2\pi/3$, $\theta_{\mathrm{max},}^{(\mathbf{T}_1)} = -5\pi/6$, $\theta_{\mathrm{min}}^{(\mathbf{T}_2)} = \pi$, $\theta_{\mathrm{max}}^{(\mathbf{T}_2)} = -\pi/2$.

In Figure 3, we take a 20 dB value for the SNR and compare the rate region obtained from the convex optimization scheme of Section IV to the equal power allocation strategy. Notice that, in spite of the strong correlation exhibiting very low eigenvalues in $\mathbf{R}_1$ and $\mathbf{R}_2$, very little is gained by our power allocation scheme compared to uniform power allocation. Quite to the contrary, in Figure 4, we consider a -5 dB SNR, and observe a substantial gain in capacity from the optimal iterative water-filling algorithm of Table I compared to equal power allocation on the transmit antenna array.

In Figure 5, a comparison is made between the theoretical and simulated rate regions for a 20 dB SNR. The latter is obtained from $1,000$ averaged Monte Carlo simulations for every transmit power pair (P_1, P_2). We observe an almost perfect fit, even for these low $N = 8$, $n_1 = n_2 = 4$ numbers of transmit and receive antennas.

B. Multi-User MIMO

We now apply Equations (45) and (61) to the downlink of a two-cell network. The capacity analyzed here is the per-antenna achievable rate on the link between base station 1 and the user, the latter of which is interfered by base station 2. The relative power of the interfering signal from base station 2 is on average Γ times that of user

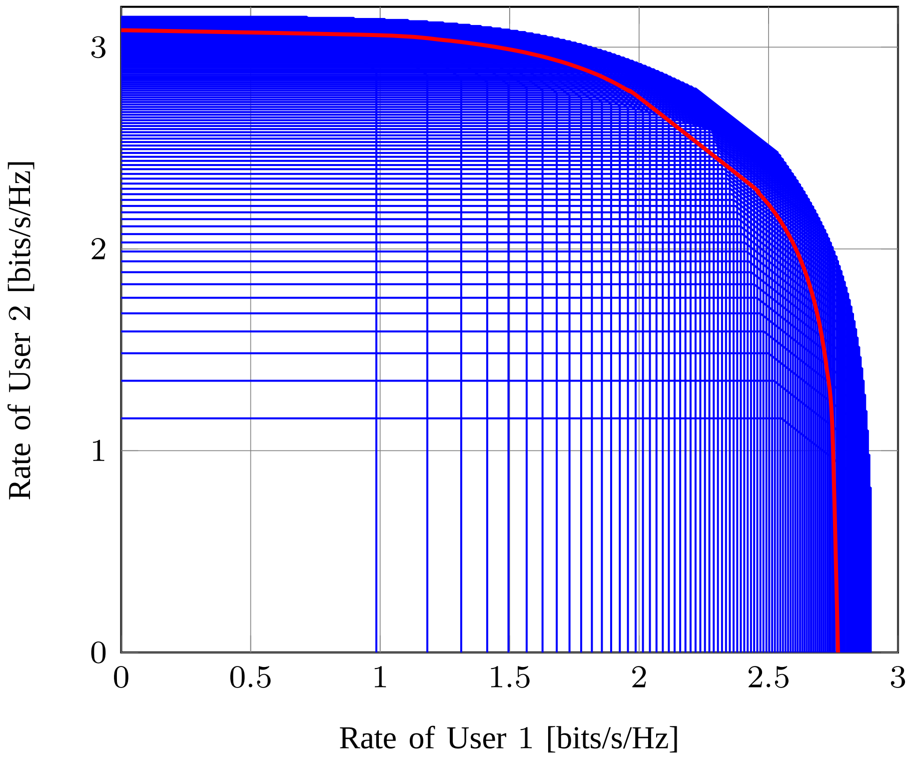

Fig. 3. (Per-antenna) rate region $\mathcal{C}_{\mathrm{BC}}$ for $K = 2$ users, $N = 8$, $n_1 = n_2 = 4$, $\mathrm{SNR} = 20$ dB, random transmit-receive solid angle of aperture $\pi/2$, $d_{\mathrm{T}}/\lambda = 10$, $d_{\mathrm{R}}/\lambda = 1/4$. In thick line, capacity limit when $\mathbf{E}[\mathbf{s}\mathbf{s}^{\mathsf{H}}] = \mathbf{I}_N$.

1. Both base stations 1 and 2 are equipped with linear arrays of n antennas and the user with a linear array of N antennas. The correlation matrices $\mathbf{T}_i$ at the transmission and $\mathbf{R}_i$ at the reception, $i \in \{1, 2\}$, are also modeled thanks to the generalized Jake's model.

In Figure 6, we took $N = 16$, $\Gamma = 0.25$ and we consider single-user decoding at the receiver. For every realization of $\mathbf{T}_i$, $\mathbf{R}_i$, 1000 channel realizations are processed to produce the simulated ergodic capacity and compared to the theoretical capacity (61). Those capacities are then averaged over 100 realizations of $\mathbf{T}_i$, $\mathbf{R}_i$, varying in the random choice of $\theta_{\min}^{(\mathbf{R}_i)}$, $\theta_{\min}^{(\mathbf{T}_i)}$, $\theta_{\max}^{(\mathbf{R}_i)}$ and $\theta_{\max}^{(\mathbf{T}_i)}$ with constraint $\theta_{\max}^{(\mathbf{R}_i)} - \theta_{\min}^{(\mathbf{R}_i)} = \theta_{\max}^{(\mathbf{T}_i)} - \theta_{\min}^{(\mathbf{T}_i)} = \pi/2$, while the distance between antennas indexed by a and b are $d_{ab}^{\mathrm{T}_i} = 10\lambda|a-b|$ at the transmitters and $d_{ab}^{\mathrm{R}} = 2\lambda|a-b|$ at the receiver. The SNR ranges from -5 dB to 30 dB, and $n \in \{8, 16\}$. We observe here that Monte-Carlo simulations perfectly match the capacity obtained from Equation (45).

In Figure 7, with the same assumptions as previously, we apply MMSE decoding at the base station. Here, a slight difference is observed in the high SNR regime between theory and practice. This was somehow expected, since the large N approximations in Lemmas 7 and 5 especially are very loose for σ^2 close to $\mathbb{R}^-$. To cope with this gap, many more antennas must be used. We also observe a significant difference in performance between the optimal single-user and the suboptimal linear MMSE decoders, especially in the high SNR region. Therefore, in wireless networks, when interfering cells are treated as Gaussian correlated noise at the cell-edge, i.e. where the

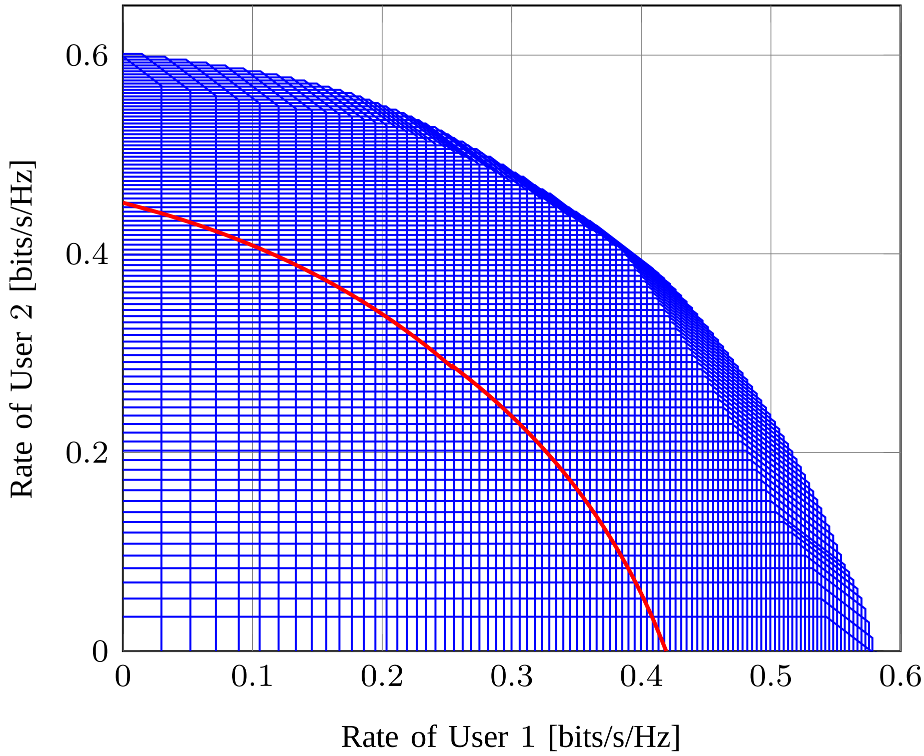

Fig. 4. (Per-antenna) rate region $\mathcal{C}_{\rm BC}$ for $K = 2$ users, $N = 8$, $n_1 = n_2 = 4$, $\rm SNR = -5$ dB, random transmit-receive solid angle of aperture $\pi/2$, $d_{\rm T}/\lambda = 10$, $d_{\rm R}/\lambda = 1/4$. In thick line, capacity limit when $\mathbf{E}[\mathbf{s}\mathbf{s}^{\sf H}] = \mathbf{I}_N$.

interference is maximum, the MMSE decoder provides tremendous performance loss.

Finally, in Figure 8, we study the effect of the angle spread of energy transmission and reception. As previously recalled, in most contributions treating antenna correlation, only the distances between the antennas is considered as relevant to the signal correlation. We provide hereafter a classical situation for which angles of departure and arrival are of critical importance to the system rate performance. The situation is similar to that of Figure 6 with $n = 8$ ($N = 16$), $\rm SNR = 20$ dB; the median directional of arrival and departure (for all devices) are taken from $\pi/32$ (grazing angle) to $\pi/2$ (signal transmitted/received in front), while the angle spread, on the x-axis, varies from $\pi/32$ to $\pi/2$. First, note importantly that the directions of departure or arrival are significant to the system performance: for instance, with a quite realistic angle spread of $\pi/4$, the single-user decoding capacity shows a three-fold increase from grazing to orthogonal angles.[10] We observe that, in all cases, the per-antenna capacity grows with the angle spread, which therefore offers some sort of diversity gain. Also, grazing angles provide significantly less diversity gain than square angles of departure or arrival.

[10]remember that we assume isotropic energy emission/reception in the top-bottom direction and we consider here a restrictive angle of captured energy in the horizontal plane so that, in reality, one might expect even more staggering results.

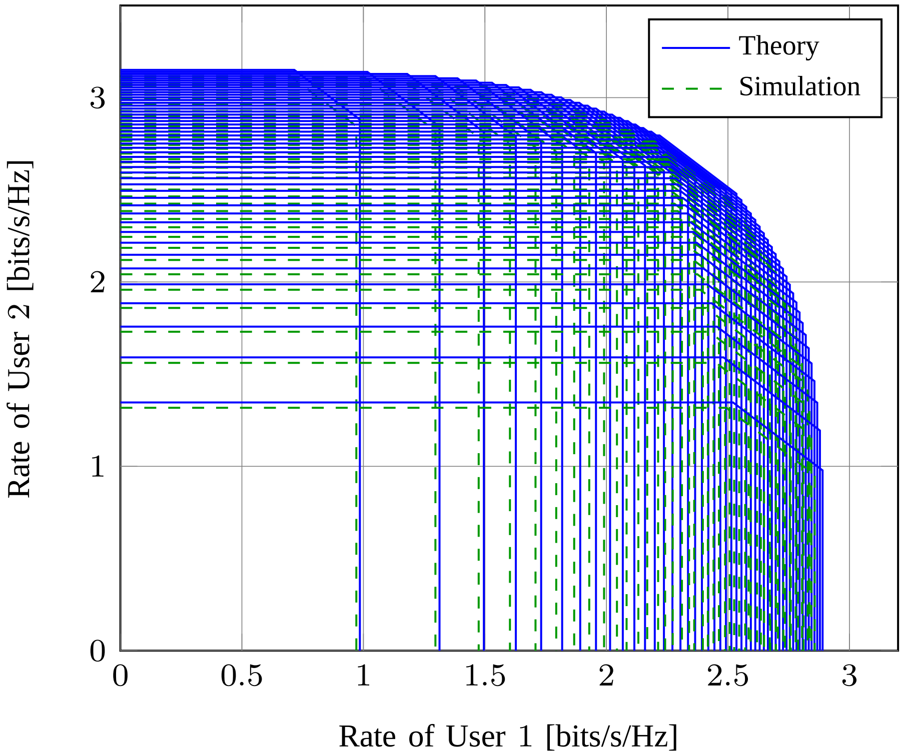

Fig. 5. (Per-antenna) rate region $\mathcal{C}_{\rm BC}$ for $K = 2$ users, theory against simulation, $N = 8$, $n_1 = n_2 = 4$, $\rm SNR = 20$ dB, random transmit-receive solid angle of aperture $\pi/2$, $d_{\rm T}/\lambda = 10$, $d_{\rm R}/\lambda = 1/4$.

VI. Conclusion

In this contribution, we proposed to analyze the per-antenna rate performance of a wide family of multi-antenna communication schemes including multiple cells and/or multiple users, and taking into account the correlation effects due to antenna closeness and directional energy transmission/reception. As an introductory example, we studied the rate region of MAC and BC channels, as well as the uplink and downlink capacity of multi-cell networks with interference. Our main results stem from a novel mathematical deterministic equivalent of the Stieltjes transform and the Shannon transform of a certain type of large random matrices. Based on these new tools, an accurate analysis of the effects of correlation can be directly translated into the antenna efficiency of multi-user multi-cell systems. We also provided an iterative water-filling algorithm to achieve the capacity boundary of the MAC and BC rate regions, which proved in simulation to give tremendous capacity gains in the low SNR regime.

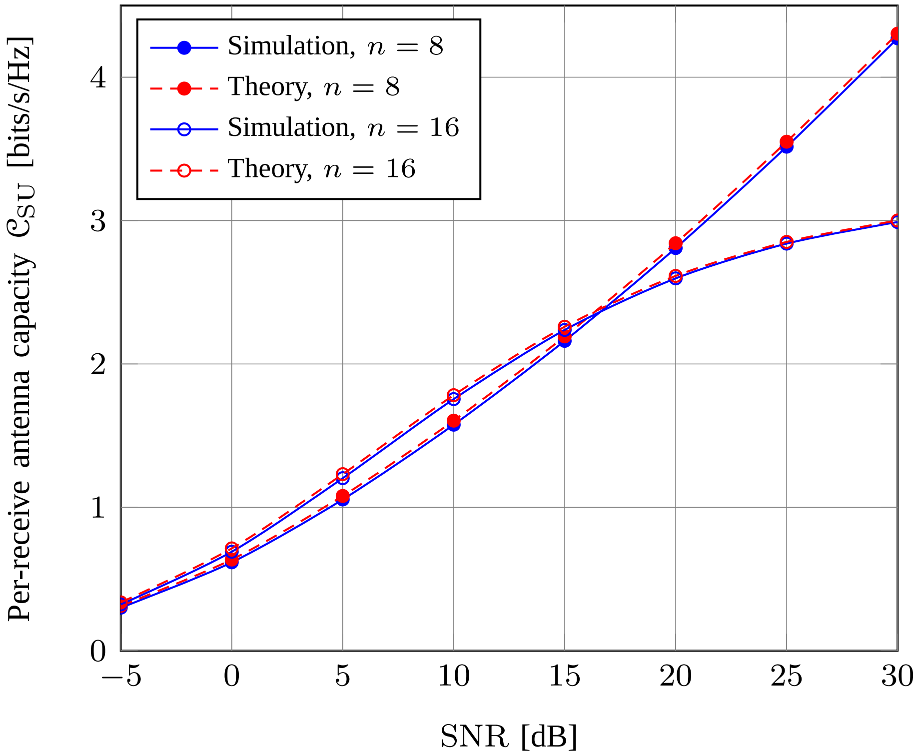

Fig. 6. Capacity of point-to-point MIMO in two-cell uplink, single-user decoding, $N = 16$, $n \in \{8, 16\}$, $\Gamma = 25\%$.

Appendix A
Proof of Theorem 1

Proof: 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.

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 [4] and [19] (which provide more details in their appendices), that

$$|F^{\mathbf{B}_N} - F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}}\widetilde{\mathbf{X}}\mathbf{T}\widetilde{\mathbf{X}}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}}| \xrightarrow{\text{a.s.}} 0 \tag{63}$$

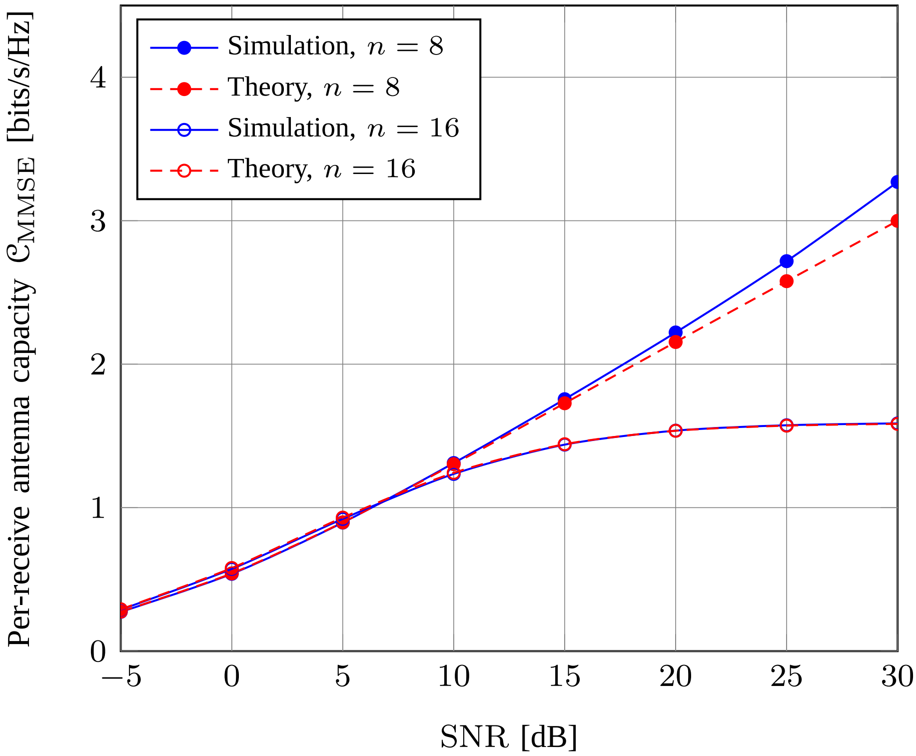

Fig. 7. Capacity of point-to-point MIMO in two-cell uplink, with MMSE decoder, $N = 16$, $n \in \{8, 16\}$, $\Gamma = 25\%$.

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 [4], 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 $\alpha > 0$

$$\mathbf{A}^{\alpha} = \mathbf{U}\,\mathrm{diag}(\lambda_1^{\mathbf{A}}\mathbf{1}_{\{\lambda_r^{\mathbf{A}}\leq\alpha\}}, \ldots, \lambda_r^{\mathbf{A}}\mathbf{1}_{\{\lambda_1\leq\alpha\}})\mathbf{U}^{\mathsf{H}} \tag{64}$$

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

$$\|F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}}\mathbf{Q}\mathbf{T}\mathbf{Q}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}} - F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}\alpha}Q\mathbf{T}^{\alpha}\mathbf{Q}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}\alpha}}\| \leq \frac{2}{N}\mathrm{rank}(\mathbf{R}^{\frac{1}{2}} - \mathbf{R}^{\frac{1}{2}\alpha}) + \frac{1}{N}\mathrm{rank}(\mathbf{T} - \mathbf{T}^{\alpha}) \tag{65}$$

$$= \frac{2}{N}\sum_{i=1}^{N}\mathbf{1}_{\{\lambda_i^{\mathbf{R}}>\alpha\}} + \frac{1}{N}\sum_{i=1}^{n}\mathbf{1}_{\{\lambda_i^{\mathbf{T}}>\alpha)\}} \tag{66}$$

$$= 2F^{\mathbf{R}}((\alpha,\infty)) + \frac{1}{c_N}F^{\mathbf{T}}((\alpha,\infty)) \tag{67}$$

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

$$\|F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}}\mathbf{Q}\mathbf{T}\mathbf{Q}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}} - F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}\alpha_N}\mathbf{Q}\mathbf{T}^{\alpha_N}\mathbf{Q}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}\alpha_N}}\| \to 0 \tag{68}$$

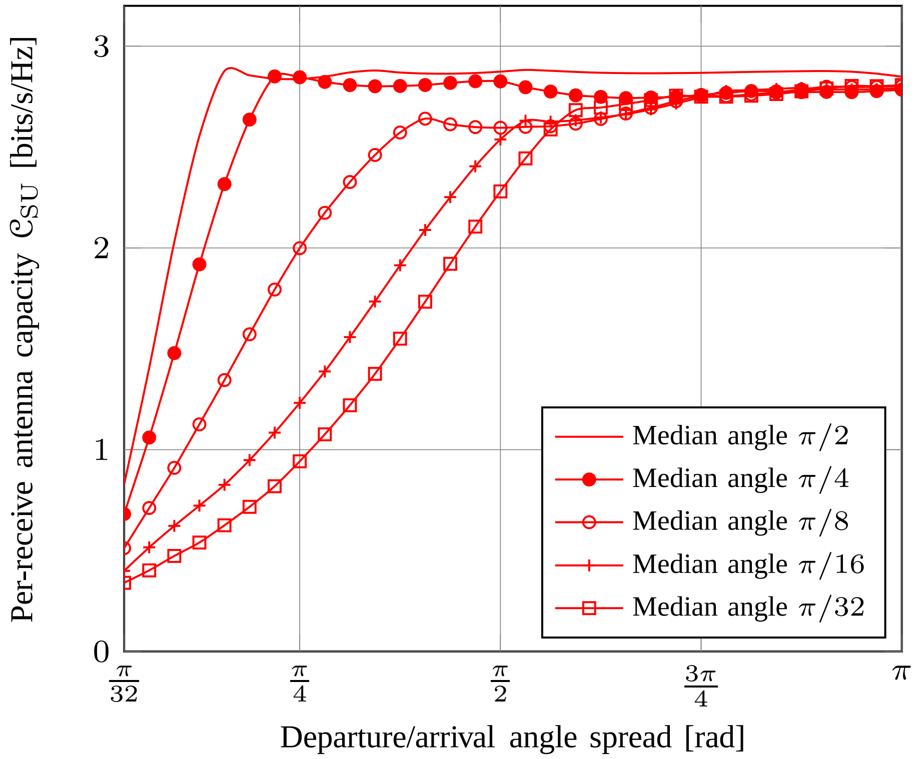

Fig. 8. Capacity of point-to-point MIMO in two-cell uplink, single-user decoding, $N = 16$, $n = 8$, $\Gamma = 25\%$, $\mathrm{SNR} = 20$ dB, for different median angles of departure/arrival.

as $N \to \infty$.

A metric D on probability measures defined on $\mathbb{R}$, which induces the topology of vague convergence, is introduced in [4] to handle the last truncation step. The matrices studied in [4] are essentially $\mathbf{B}_N$ with $\mathbf{R} = \mathbf{I}_N$. Following the steps beginning at (3.4) in [4], we see in our case that when α_N is chosen so that as $N \to \infty$, $\alpha_N \uparrow \infty$,

$$\alpha_N^8(\mathrm{E}|X_{11}^2 \mathbf{1}_{\{X_{11}|\geq \ln N\}} + N^{-1}) \to 0 \tag{69}$$

and

$$\sum_{N=1}^{\infty} \frac{\alpha_N^{16}}{N^2} \left(\mathrm{E}|X_{11}|^4 \mathbf{1}_{\{|X_{11}|<\sqrt{N})\}} + 1\right) < \infty \tag{70}$$

We will get

$$D(F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}\alpha_N}\widetilde{\mathbf{X}}\mathbf{T}^{\alpha_N}\widetilde{\mathbf{X}}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}\alpha_N}}, F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}\alpha_N}\overline{\mathbf{X}}\mathbf{T}^{\alpha_N}\overline{\mathbf{X}}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}\alpha_N}}) \xrightarrow{\text{a.s.}} 0 \tag{71}$$

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 (68) and (71) 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 $|X_{ij}| \leq \log N$.

Later on the proof will require a restricted growth rate on both $\|\mathbf{R}\|$ and $\|\mathbf{T}\|$. We see from (68) that we can also assume

$$\max(\|\mathbf{R}\|, \|\mathbf{T}\|) \leq \log N \tag{72}$$

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

$$\mathbf{B}_N = \mathbf{S} + \sum_{j=1}^{n} \tau_j \mathbf{y}_j \mathbf{y}_j^{\mathsf{H}} \tag{73}$$

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

$$e_N = e_N(z) = (1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \tag{74}$$

and

$$p_N = -\frac{1}{nz}\sum_{j=1}^{n} \frac{\tau_j}{1 + c_N \tau_j e_N} = \int \frac{-\tau}{z(1 + c_N \tau e_N)} dF^{\mathbf{T}}(\tau) \tag{75}$$

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

$$e_N = (1/N)\operatorname{tr}\underline{\mathbf{R}}(\Lambda - z\mathbf{I}_N)^{-1} = (1/N)\sum_{i=1}^{N} \frac{\underline{R}_{ii}}{\lambda_i - z} \tag{76}$$

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 (72), it follows that

$$|e_N| \leq v^{-1}\log N \tag{77}$$

and

$$\left|\int \frac{\tau}{(1 + c_N \tau e_N)} dF^{\mathbf{T}}(\tau)\right| = |zp_N(z)| \leq |z|v^{-1}\log N \tag{78}$$

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

$$\frac{-\tau}{z(1 + m(z))} \tag{79}$$

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

$$\left|\frac{\tau}{1 + m(z)}\right| \leq \tau|z|v^{-1} \tag{80}$$

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

$$\mathbf{B}_N - z\mathbf{I}_N - \mathbf{D} = \sum_{j=1}^{n} \tau_j \mathbf{y}_j \mathbf{y}_j^{\mathsf{H}} + zp_N\mathbf{R} \tag{81}$$

Taking inverses and using Lemma 4 we have

$$(\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \mathbf{D}^{-1} = \sum_{j=1}^{n} \tau_j \mathbf{D}^{-1}\mathbf{y}_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} + zp_N\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \tag{82}$$

$$= \sum_{j=1}^{n} \tau_j \frac{\mathbf{D}^{-1}\mathbf{y}_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}}{1 + \tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} + zp_N\mathbf{D}^{-1}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \tag{83}$$

Taking traces and dividing by N, we have

$$\frac{1}{N}\operatorname{tr}\mathbf{D}^{-1} - m_N(z) = \frac{1}{n}\sum_{j=1}^{n}\tau_j d_j \equiv w_N^m \tag{84}$$

where

$$d_j = \frac{(1/N)\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1 + \tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} - \frac{(1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}}{1 + c_N\tau_j e_N} \tag{85}$$

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

$$\frac{1}{N}\operatorname{tr}\mathbf{D}^{-1}\mathbf{R} - e_N(z) = \frac{1}{n}\sum_{j=1}^{n}\tau_j d_j^e \equiv w_N^e \tag{86}$$

where

$$d_j^e = \frac{(1/N)\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1 + \tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} - \frac{(1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}\mathbf{D}^{-1}}{1 + c_N\tau_j e_N} \tag{87}$$

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

$$\lim_{N\to\infty}(\log^k N)w_N^m = 0 \tag{88}$$

and

$$\lim_{n\to\infty}(\log^k N)\, w_N^e = 0 \tag{89}$$

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 (80) we have

$$\left|\frac{1}{1 + \tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j}\right| \le \frac{|z|}{v}. \tag{90}$$

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

$$p_{(j)} = p_{(j)}(z) = \int \frac{-\tau}{z(1 + c_N\tau e_{(j)})} dF^{\mathbf{T}}(\tau) \tag{91}$$

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

Using Lemma 4, Equations (72) and (80), we have

$$|zp_N - zp_{(j)}| = |e_N - e_{(j)}|c_N \left|\int \frac{\tau^2}{(1 + c_N\tau e_N)(1 + c_N\tau e_{(j)})} dF^{\mathbf{T}}(\tau)\right| \le \frac{c_N|z|^2\log^3 N}{Nv^3}. \tag{92}$$

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

$$d_j^1 = \frac{(1/N)\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} - \frac{(1/N)\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} \tag{93}$$

$$d_j^2 = \frac{(1/N)\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} - \frac{(1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} \tag{94}$$

$$d_j^3 = \frac{(1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} - \frac{(1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}^{-1}}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} \tag{95}$$

$$d_j^4 = \frac{(1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}^{-1}}{1+\tau_j\mathbf{y}_j^{\mathsf{H}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_j} - \frac{(1/N)\operatorname{tr}\mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}^{-1}}{1+c_N\tau_j e_N} \tag{96}$$

From Lemma 4, Equations (72), (90) and (92), we have

$$\tau_j|d_j^1| \le \frac{1}{N}\|\mathbf{x}_j\|^2\frac{c_N\log^7 N|z|^3}{Nv^7} \tag{97}$$

$$\tau_j|d_j^2| \le |z|v^{-1}\frac{\log N}{N}\left|\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j - \operatorname{tr}\mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}\right| \tag{98}$$

$$\tau_j|d_j^3| \le \frac{|z|\log^3 N}{vN}\left(\frac{1}{v^2} + \frac{c_N|z|^2\log^3 N}{v^6}\right) \to 0, \text{ as } n\to\infty \tag{99}$$

$$\tau_j|d_j^4| \le \frac{|z|c_N\log^4 N}{Nv^3}\left(|\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j - \operatorname{tr}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}| + \frac{\log N}{v}\right) \tag{100}$$

From Lemma 7, there exists $\bar{K} > 0$ such that,

$$\mathrm{E}|\frac{1}{N}\|\mathbf{x}_j\|^2 - 1|^6 \le KN^{-3}\log^{12} N \tag{101}$$

$$\mathrm{E}\frac{1}{N^6}|\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j - \operatorname{tr}\mathbf{R}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{E}\mathbf{D}_{(j)}^{-1}|^6 \le KN^{-3}v^{-12}\log^{24} N \tag{102}$$

$$\mathrm{E}\frac{1}{N^6}|\mathbf{x}_j^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_j - \operatorname{tr}\mathbf{R}^{\frac{1}{2}}(\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}^{\frac{1}{2}}|^6 \le KN^{-3}v^{-6}\log^{18} N \tag{103}$$

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 Equations (88) and (89).

C. Existence and uniqueness of $m_N^{(0)}(z)$

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

$$e = \frac{1}{N}\operatorname{tr}\left(\mathbf{S} + \left[\int\frac{\tau}{1+c_N\tau e}dF^{\mathbf{T}}(\tau)\right]\mathbf{R} - z\mathbf{I}_N\right)^{-1}\mathbf{R} \tag{104}$$

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

$$\mathbf{R}_{N_j} = \operatorname{diag}(\mathbf{R},\mathbf{R},\ldots,\mathbf{R}),\quad \mathbf{S}_{N_j} = \operatorname{diag}(\mathbf{S},\mathbf{S},\ldots,\mathbf{S}) \tag{105}$$

both $jN\times jN$ and

$$\mathbf{T}_{N_j} = \operatorname{diag}(\mathbf{T},\mathbf{T},\ldots,\mathbf{T}) \tag{106}$$

of size $jn \times jn$.

We see that $F^{\mathbf{T}_{N_j}} = F^{\mathbf{T}}$ and the right side of (104) remains unchanged for all N_j. Consider a realization where $w^e_{N_j} \to 0$ as $j \to \infty$. We have $|e_{N_j}(z)| = |(jN)^{-1} \operatorname{tr} \mathbf{R}(\mathbf{B}_{jN} - z\mathbf{I}_N)^{-1}| \leq v^{-1} \log N$, remaining bounded as $j \to \infty$. Consider then a subsequence for which e_{N_j} converges to, say, e. From (80), we see that

$$\left| \frac{\tau}{1 + c_N \tau e_{N_j}} \right| \leq \tau |z| v^{-1} \tag{107}$$

so that from the dominated convergence theorem we have

$$\int \frac{\tau}{1 + c_N \tau e_{N_j}(z)} dF^{\mathbf{T}}(\tau) \to \int \frac{\tau}{1 + c_N \tau e} dF^{\mathbf{T}}(\tau) \tag{108}$$

along this subsequence. Therefore e solves (104).

We now show uniqueness. Let e be a solution to (104) and let $e_2 = \Im[e]$. Recalling the definition of $\mathbf{D}$ we write

$$e = \frac{1}{N} \operatorname{tr} \left(\mathbf{D}^{-1} \mathbf{R} \mathbf{D}^{-\mathsf{H}} \left(\mathbf{S} + \left[\int \frac{\tau}{1 + c_N \tau e^*} dF^{\mathbf{T}}(\tau) \right] \mathbf{R} - z^* \mathbf{I} \right) \right) \tag{109}$$

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

$$e_2 = \frac{1}{N} \operatorname{tr} \left(\mathbf{D}^{-1} \mathbf{R} (\mathbf{D}^{\mathsf{H}})^{-1} \left(\left[\int \frac{c_N \tau^2 e_2}{|1 + c_N \tau e|^2} dF^{\mathbf{T}}(\tau) \right] \mathbf{R} + v \mathbf{I}_N \right) \right) = e_2 \alpha + v \beta \tag{110}$$

where we denoted

$$\alpha = \frac{1}{N} \operatorname{tr} \left(\mathbf{D}^{-1} \mathbf{R} (\mathbf{D}^{\mathsf{H}})^{-1} \left[\int \frac{c_N \tau^2}{|1 + c_N \tau e|^2} dF^{\mathbf{T}}(\tau) \right] \mathbf{R} \right) \tag{111}$$

$$\beta = \frac{1}{N} \operatorname{tr} \left(\mathbf{D}^{-1} \mathbf{R} (\mathbf{D}^{\mathsf{H}})^{-1} \right) \tag{112}$$

Let $\underline{e}$ be another solution to (104), 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

$$\gamma = \int \frac{c_N \tau^2}{(1 + c_N \tau e)(1 + c_N \tau \underline{e})} dF^{\mathbf{T}}(\tau) \frac{\operatorname{tr} \mathbf{D}^{-1} \mathbf{R} \underline{\mathbf{D}}^{-1} \mathbf{R}}{N} \tag{113}$$

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

$$|\gamma| \leq \left(\int \frac{c_N \tau^2}{|1 + c_N \tau e|^2} dF^{\mathbf{T}}(\tau) \frac{\operatorname{tr} \mathbf{D}^{-1} \mathbf{R} (\mathbf{D}^{\mathsf{H}})^{-1} \mathbf{R}}{N} \right)^{\frac{1}{2}} \left(\int \frac{c_N \tau^2}{|1 + c_N \tau \underline{e}|^2} dF^{\mathbf{T}}(\tau) \frac{\operatorname{tr} \underline{\mathbf{D}}^{-1} \mathbf{R} (\underline{\mathbf{D}}^{\mathsf{H}})^{-1} \mathbf{R}}{N} \right)^{\frac{1}{2}} \tag{114}$$

$$= \alpha^{\frac{1}{2}} \underline{\alpha}^{\frac{1}{2}} \tag{115}$$

$$= \left(\frac{e_2 \alpha}{e_2 \alpha + v \beta} \right)^{\frac{1}{2}} \left(\frac{\underline{e}_2 \underline{\alpha}}{\underline{e}_2 \underline{\alpha} + v \underline{\beta}} \right)^{\frac{1}{2}} \tag{116}$$

Necessarily β and $\underline{\beta}$ are positive since $\mathbf{R} \neq 0$. Therefore $|\gamma| < 1$ so we must have $e = \underline{e}$.

D. Termination of the proof

Let e_N^0 denote the solution to (104). We show now for any $\ell > 0$, almost surely

$$\lim_{N \to \infty} \log^{\ell} N (e_N - e_N^0) = 0 \tag{117}$$

Let $e_2^0 = \Im[e_N^0]$, and $\alpha^0 = \alpha_N^0$, $\beta^0 = \beta_N^0$ be the values as above for which $e_2^0 = e_2^0\alpha^0 + v\beta^0$. We have, using (72) and (80),

$$e_2^0\alpha_N^0/\beta_N^0 \leq e_2^0 c_N \log N \int \frac{\tau^2}{|1+c_N\tau e_N^0|^2} dF^{\mathbf{T}}(\tau) \tag{118}$$

$$= -\log N \Im\left[\int \frac{\tau}{1+c_N\tau e_N^0} dF^{\mathbf{T}}(\tau)\right] \tag{119}$$

$$\leq \log^2 N|z|v^{-1} \tag{120}$$

Therefore

$$\alpha^0 = \left(\frac{e_2^0\alpha^0}{e_2^0\alpha^0 + v\beta^0}\right) \tag{121}$$

$$= \left(\frac{e_2^0\alpha^0/\beta^0}{v + e_2^0\alpha^0/\beta^0}\right) \tag{122}$$

$$\leq \left(\frac{\log^2 N|z|}{v^2 + \log^2 N|z|}\right) \tag{123}$$

Let $\mathbf{D}^0$, $\mathbf{D}$ denote $\mathbf{D}$ as above with e replaced by, respectively e_N^0 and e_N. We have

$$e_N = \frac{1}{N}\operatorname{tr}\mathbf{D}^{-1}\mathbf{R} - w_N^e \tag{124}$$

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

$$e_2 = \frac{1}{N}\operatorname{tr}\left(\mathbf{D}^{-1}\mathbf{R}\mathbf{D}^{-\mathsf{H}}\left(\left[\int \frac{c_N\tau^2 e_2}{|1+c_N\tau e_N|^2} dF^{\mathbf{T}}(\tau)\right]\mathbf{R} + v\mathbf{I}_N\right)\right) - \Im[w_N^e] \tag{125}$$

$$= e_2\alpha + v\beta - \Im w_N^e \tag{126}$$

We have as above $e_N - e_N^0 = \gamma(e_N - e_N^0) + w_N^e$ where now

$$|\gamma| \leq {\alpha^0}^{\frac{1}{2}}\alpha^{\frac{1}{2}} \tag{127}$$

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

$$|w_N^e| \leq \frac{v^3}{4c_N|z|^2\log^3 N} \tag{128}$$

Suppose $\beta \leq \frac{v^2}{4c_N|z|^2\log^3 N}$. Then by Equations (72) and (80) we get

$$\alpha \leq c_N v^{-2}|z|^2\log^3 N\beta \leq 1/4 \tag{129}$$

which implies $|\gamma| \leq 1/2$. Otherwise we get from (121) and (128)

$$|\gamma| \leq {\alpha^0}^{\frac{1}{2}}\left(\frac{e_2\alpha}{e_2\alpha + v\beta - \Im[w_N^e]}\right)^{\frac{1}{2}} \tag{130}$$

$$\leq \left(\frac{\log N|z|}{v^2 + \log N|z|}\right)^{\frac{1}{2}} \tag{131}$$

Therefore for all N large

$$\log^{\ell} N|e_N - e_N^0| \leq \frac{(\log^{\ell} N)w_N^e}{1 - \left(\frac{\log^2 N|z|}{v^2 + \log^2 N|z|}\right)^{\frac{1}{2}}} \tag{132}$$

$$\leq 2v^{-2}(v^2 + \log^2 N|z|)(\log^{\ell} N)w_N^e \tag{133}$$

$$\to 0 \tag{134}$$

as $n \to \infty$. Therefore (117) follows.

Let $m_N^0 = N^{-1} \operatorname{tr} \mathbf{D}^0$. We finally show

$$m_N - m_N^0 \xrightarrow{\text{a.s.}} 0 \tag{135}$$

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

$$m_N - m_N^0 = \gamma(e_N - e_N^0) - w_N^m \tag{136}$$

where now

$$\gamma = \int \frac{c_N \tau^2}{(1 + c_N \tau e_N)(1 + c_N \tau e_N^0)} dF^{\mathbf{T}}(\tau) \frac{\operatorname{tr} \mathbf{D}^{-1} \mathbf{R} \mathbf{D}^{0^{-1}}}{N} \tag{137}$$

From (72) and (80) we get $|\gamma| \leq c_N |z|^2 v^{-4} \log^3 N$. Therefore, from (88) and (117), we get (135).

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 (135). Therefore, by Vitali's convergence theorem, page 168 of [16], for any $\varepsilon > 0$ we have with probability one $m_N(z) - m_N^0(z) \to 0$ uniformly in any region of $\mathbb{C}$ bounded by a contour interior to

$$\mathbb{C} \setminus (\{z : |z| \leq \epsilon\} \cup \{z = x + iv : x > 0, |v| \leq \varepsilon\}) \tag{138}$$

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

$$m_N^0(z) = \int \frac{1}{f(r) + r \int \frac{\tau}{1 + c_N \tau e_N^0(z)} dF^{\mathbf{T}}(\tau) - z} dF^{\mathbf{R}}(r) \tag{139}$$

$$e_N^0(z) = \int \frac{r}{f(r) + r \int \frac{\tau}{1 + c_N \tau e_N^0(z)} dF^{\mathbf{T}}(\tau) - z} dF^{\mathbf{R}}(r) \tag{140}$$

E. Extension to $K \geq 1$

Suppose now

$$\mathbf{B}_N = \mathbf{S} + \sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k \mathbf{X}_k^{\mathsf{H}} \mathbf{R}_k^{\frac{1}{2}} \tag{141}$$

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

$$\mathbf{B}_N = \mathbf{S} + \sum_{k=1}^{K} \sum_{j=1}^{n_k} \tau_{k,j} \mathbf{y}_{k,j} \mathbf{y}_{k,j}^{\mathsf{H}} \tag{142}$$

Define

$$e_{N,k} = e_{N,k}(z) = (1/N)\,\mathrm{tr}\,\mathbf{R}_k(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \tag{143}$$

and

$$p_k = -\frac{1}{n_k z}\sum_{j=1}^{n_k}\frac{\tau_{k,j}}{1 + c_k\tau_{k,j}e_{N,k}} \tag{144}$$

$$= \int \frac{-\tau_k}{1 + c_k\tau_k e_{N,k}}dF^{\mathbf{T}_k}(\tau_k) \tag{145}$$

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 $\mathbf{D} = -z\mathbf{I}_N + \mathbf{S} - \sum_{k=1}^{K} zp_k(z)\mathbf{R}_k$. We write

$$\mathbf{B}_N - z\mathbf{I}_N - \mathbf{D} = \sum_{k=1}^{K}\left(\sum_{j=1}^{n_k}\tau_{k,j}\mathbf{y}_{k,j}\mathbf{y}_{k,j}^{\mathsf{H}} + zp_k(z)\mathbf{R}_k\right) \tag{146}$$

Taking inverses and using Lemma 4, we have

$$\mathbf{D}^{-1} - (\mathbf{B}_N - z\mathbf{I}_N)^{-1} = \sum_{k=1}^{K}\left(\sum_{j=1}^{n_k}\tau_{k,j}\mathbf{D}^{-1}\mathbf{y}_{k,j}\mathbf{y}_{k,j}^{\mathsf{H}}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} + zp_k\mathbf{D}^{-1}\mathbf{R}_k(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\right) \tag{147}$$

$$= \sum_{k=1}^{K}\left(\sum_{j=1}^{n_k}\tau_{k,j}\frac{\mathbf{D}^{-1}\mathbf{y}_{k,j}\mathbf{y}_{k,j}^{\mathsf{H}}(\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1}}{1 + \tau_{k,j}\mathbf{y}_{k,j}^{\mathsf{H}}(\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_{k,j}} + zp_k\mathbf{D}^{-1}\mathbf{R}_k(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\right) \tag{148}$$

Taking traces and dividing by N, we have

$$(1/N)\,\mathrm{tr}\,\mathbf{D}^{-1} - m_N(z) = \sum_{k=1}^{K}\frac{1}{n_k}\sum_{j=1}^{n_k}\tau_{k,j}d_{k,j} \equiv w_N^m \tag{149}$$

where

$$d_{k,j} = \frac{(1/N)\mathbf{x}_{k,j}^{\mathsf{H}}\mathbf{R}_k^{\frac{1}{2}}(\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}\mathbf{R}_k^{\frac{1}{2}}\mathbf{x}_{k,j}}{1 + \tau_{k,j}\mathbf{y}_{k,j}^{\mathsf{H}}(\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_{k,j}} - \frac{(1/N)\,\mathrm{tr}\,\mathbf{R}_k(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}}{1 + c_k\tau_{k,j}e_{N,k}} \tag{150}$$

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

$$(1/N)\,\mathrm{tr}\,\mathbf{D}^{-1}\mathbf{R}_{\underline{k}} - e_{\underline{k}}(z) = \sum_{k=1}^{K}\frac{1}{n_k}\sum_{j=1}^{n_k}\tau_{k,j}d_{k\underline{k}j}^{e} \equiv w_{\underline{k}}^{e} \tag{151}$$

where

$$d_{k\underline{k}j}^{e} = \frac{(1/N)\mathbf{x}_{k,j}^{\mathsf{H}}\mathbf{R}_k^{\frac{1}{2}}(\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}_{\underline{k}}\mathbf{D}^{-1}\mathbf{R}_k^{\frac{1}{2}}\mathbf{x}_{k,j}}{1 + \tau_{k,j}\mathbf{y}_{k,j}^{\mathsf{H}}(\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_{k,j}} - \frac{(1/N)\,\mathrm{tr}\,\mathbf{R}_k(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}_{\underline{k}}\mathbf{D}^{-1}}{1 + c_k\tau_{k,j}e_{N,k}} \tag{152}$$

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 $e_1^0, \ldots, e_K^0$ with positive imaginary parts for which for each i

$$e_i^0 = \frac{1}{N}\,\mathrm{tr}\,\mathbf{R}_i\left(\mathbf{S} + \sum_{k=1}^{K}\left[\int\frac{\tau}{1 + c_k\tau e_k^0}dF^{\mathbf{T}_k}(\tau)\right]\mathbf{R}_k - z\mathbf{I}_N\right)^{-1} \tag{153}$$

Let us verify uniqueness. Let $\mathbf{e}^0 = (e_1^0, \ldots, e_K^0)^\mathsf{T}$, and let $\mathbf{D}^0$ denote the matrix in (153) whose inverse is taken (essentially $\mathbf{D}$ after the $e_{N,i}$'s are replaced by the e_i^0's). Let for each j, $e_{j,2}^0 = \Im e_j^0$, and $\mathbf{e}_2^0 = (e_{1,2}^0, \ldots, e_{K,2}^0)^\mathsf{T}$. Then, noticing that for each i, $\operatorname{tr} \mathbf{S}\mathbf{D}^{0^{-\mathsf{H}}}\mathbf{R}_i\mathbf{D}^{0^{-1}}$ is real and nonnegative (positive whenever $\mathbf{S} \neq 0$) and $\operatorname{tr} \mathbf{D}^{0^{-\mathsf{H}}}\mathbf{R}_i\mathbf{D}^{0^{-1}}$ and $\operatorname{tr} \mathbf{R}_j\mathbf{D}^{0^{-\mathsf{H}}}\mathbf{R}_i\mathbf{D}^{0^{-1}}$ are real and positive for all i, j, we have

$$e_{i,2}^0 = \Im\left[\frac{1}{N}\operatorname{tr}\left(\mathbf{S} + \sum_{j=1}^K \left[\int \frac{\tau}{1 + c_j\tau\overline{e}_j^0} dF^{\mathbf{T}_j}(\tau)\right]\mathbf{R}_j - z^*\mathbf{I}\right)\mathbf{D}^{0^{-\mathsf{H}}}\mathbf{R}_i\mathbf{D}^{0^{-1}}\right] \tag{154}$$

$$= \sum_{j=1}^K e_{j,2}^0 \frac{1}{N}\operatorname{tr} \mathbf{R}_j\mathbf{D}^{0^{-\mathsf{H}}}\mathbf{R}_i\mathbf{D}^{0^{-1}} c_j \int \frac{\tau^2}{|1 + c_j\tau e_j^0|^2} dF^{\mathbf{T}_j}(\tau) + \frac{v}{N}\operatorname{tr} \mathbf{D}^{0^{-\mathsf{H}}}\mathbf{R}_i\mathbf{D}^{0^{-1}} \tag{155}$$

Let $\mathbf{C}^0 = (c_{ij}^0)$, $\mathbf{b}^0 = (b_1^0, \ldots, b_N^0)^\mathsf{T}$, where

$$c_{ij}^0 = \frac{1}{N}\operatorname{tr} \mathbf{R}_j\mathbf{D}^{0^{-\mathsf{H}}}\mathbf{R}_i\mathbf{D}^{0^{-1}} c_j \int \frac{\tau^2}{|1 + c_j\tau e_j^0|^2} dF^{\mathbf{T}_j}(\tau) \tag{156}$$

and

$$b_i^0 = \frac{1}{N}\operatorname{tr} \mathbf{D}^{0^{-\mathsf{H}}}\mathbf{R}_i\mathbf{D}^{0^{-1}} \tag{157}$$

Therefore we have $\mathbf{e}_2^0$ satisfies

$$\mathbf{e}_2^0 = \mathbf{C}^0\mathbf{e}_2^0 + v\mathbf{b}^0 \tag{158}$$

We see that each $e_{j,2}^0$, c_{ij}^0, and b_j^0 are positive. Therefore, from Lemma 9 we have $\rho(\mathbf{C}^0) < 1$.

Let $\underline{\mathbf{e}}^0 = (\underline{e}_1^0, \ldots, \underline{e}_K^0)^\mathsf{T}$ be another solution to (153), with $\underline{\mathbf{e}}_2^0$, $\underline{\mathbf{D}}^0$, $\underline{\mathbf{C}}^0 = (\underline{c}_{ij}^0)$, $\underline{\mathbf{b}}^0$ defined analogously, so that (158) holds and $\rho(\underline{\mathbf{C}}^0) < 1$. We have for each i,

$$e_i^0 - \underline{e}_i^0 = \frac{1}{N}\operatorname{tr} \mathbf{R}_i\mathbf{D}^{0^{-1}} \sum_{j=1}^K (e_j^0 - \underline{e}_j^0)c_j \int \frac{\tau^2}{(1 + c_j\tau e_j^0)(1 + c_j\tau\underline{e}_j^0)} dF^{\mathbf{T}_j}(\tau)\mathbf{R}_j\underline{\mathbf{D}}^{0^{-1}} \tag{159}$$

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

$$a_{ij} = \frac{1}{N}\operatorname{tr} \mathbf{R}_i\mathbf{D}^{0^{-1}}\mathbf{R}_j\underline{\mathbf{D}}^{0^{-1}} c_j \int \frac{\tau^2}{(1 + c_j\tau e_j^0)(1 + c_j\tau\underline{e}_j^0)} dF^{\mathbf{T}_j}(\tau) \tag{160}$$

we have

$$\mathbf{e}^0 - \underline{\mathbf{e}}^0 = \mathbf{A}(\mathbf{e}^0 - \underline{\mathbf{e}}^0) \tag{161}$$

which means, if $\mathbf{e}^0 \neq \underline{\mathbf{e}}^0$, then $\mathbf{A}$ has an eigenvalue equal to 1.

Applying Cauchy-Schwarz we have

$$|a_{ij}| \leq \left(\frac{1}{N}\mathbf{R}_i\mathbf{D}^{0^{-1}}\mathbf{R}_j\mathbf{D}^{0^{-\mathsf{H}}} \int \frac{\tau^2}{|1 + c_j\tau e_j^0|^2} dF^{\mathbf{T}_j}(\tau)\right)^{\frac{1}{2}} \left(\frac{1}{N}\mathbf{R}_i\underline{\mathbf{D}}^{0^{-1}}\mathbf{R}_j\underline{\mathbf{D}}^{0^{-\mathsf{H}}} \int \frac{\tau^2}{|1 + c_j\tau\underline{e}_j^0|^2} dF^{\mathbf{T}_j}(\tau)\right)^{\frac{1}{2}} \tag{162}$$

$$= {c_{ij}^0}^{1/2} {\underline{c}_{ij}^0}^{1/2} \tag{163}$$

Therefore from Lemmas 10 and 11 we get

$$\rho(\mathbf{A}) \leq \rho({c_{ij}^0}^{\frac{1}{2}} {\underline{c}_{ij}^0}^{\frac{1}{2}}) \leq \rho(\mathbf{C}^0)^{\frac{1}{2}} \rho(\underline{\mathbf{C}}^0)^{\frac{1}{2}} < 1 \tag{164}$$

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

Let $\mathbf{e}_N = (e_{N,1}, \ldots, e_{N,K})^{\mathsf{T}}$ and $\mathbf{e}_N^0 = (e_{N,1}^0, \ldots, e_{N,K}^0)$ denote the vector solution to (153) for each N. We will show for any $\ell > 0$, almost surely

$$\lim_{N\to\infty} \log^{\ell} N(\mathbf{e}_N - \mathbf{e}_N^0) \to \mathbf{0} \tag{165}$$

We have

$$\mathbf{e}_N^0 = (\frac{1}{N}\operatorname{tr}\mathbf{R}_1\mathbf{D}^{0^{-1}}, \ldots, \frac{1}{N}\operatorname{tr}\mathbf{R}_K\mathbf{D}^{0^{-1}})^{\mathsf{T}} \tag{166}$$

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

$$\mathbf{e}_N = (\frac{1}{N}\operatorname{tr}\mathbf{R}_1\mathbf{D}^{-1}, \ldots, \frac{1}{N}\operatorname{tr}\mathbf{R}_K\mathbf{D}^{-1})^{\mathsf{T}} + \mathbf{w}^e \tag{167}$$

Therefore

$$\mathbf{e}_N - \mathbf{e}_N^0 = \mathbf{A}(N)(\mathbf{e}_N - \mathbf{e}_N^0) + \mathbf{w}^e \tag{168}$$

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

$$a_{ij}(N) = \frac{1}{N}\operatorname{tr}\mathbf{R}_i\mathbf{D}^{-1}\mathbf{R}_j\mathbf{D}^{0^{-1}}c_j\int\frac{\tau^2}{(1+c_j\tau e_{N,j})(1+c_j\tau e_{N,j}^0)}dF^{\mathbf{T}_j}(\tau) \tag{169}$$

We let $\mathbf{e}_{N,2}^0$, $b_{ij}^0(N)$, $\mathbf{C}^0(N)$, $b_{N,i}^0$, and $\mathbf{b}_N^0$, denote the quantities from above, reflecting now their dependence on N. Let $\mathbf{C}(N) = (c_{ij}(N))$ be $K\times K$ with

$$c_{ij}(N) = \frac{1}{N}\operatorname{tr}\mathbf{R}_j\mathbf{D}^{-\mathsf{H}}\mathbf{R}_i\mathbf{D}^{-1}c_j\int\frac{\tau^2}{|1+c_j\tau e_{N,j}|^2}dF^{\mathbf{T}_j}(\tau) \tag{170}$$

Let $\mathbf{e}_{N,2} = \Im[\mathbf{e}_N]$ and $\mathbf{w}_2^e = \Im[\mathbf{w}^e]$. Define $\mathbf{b}_N = (b_{N,1}, \ldots, b_{N,K})^{\mathsf{T}}$ with

$$b_{N,i} = \frac{1}{N}\operatorname{tr}\mathbf{D}^{-\mathsf{H}}\mathbf{R}_i\mathbf{D}^{-1} \tag{171}$$

Then, as above we find that

$$\mathbf{e}_{N,2} = \mathbf{C}(N)\mathbf{e}_{N,2} + v\mathbf{b}_N + \mathbf{w}_2^e \tag{172}$$

Using (72) and (80) we see there exists a constant $K_1 > 0$ for which

$$c_{ij}^0(N) \le K_1\log^3 N b_{N,i}^0 \tag{173}$$

and

$$c_{ij}(N) \le K_1\log^3 N b_{N,i} \tag{174}$$

$$c_{ij}(N) \le K_1\log^4 N \tag{175}$$

for each i, j. Therefore, from (158) we see there exists $\hat{K} > 0$ for which

$$e_{N,i}^0 \le \hat{K}\log^4 N v b_{N,i}^0 \tag{176}$$

Let $\mathbf{x}$ be such that $\mathbf{x}^{\mathsf{T}}$ is a left eigenvector of $\mathbf{C}^0(N)$ corresponding to eigenvalue $\rho(\mathbf{C}^0(N))$, guaranteed by Lemma 12. Then from (172) we have

$$\mathbf{x}^{\mathsf{T}}\mathbf{e}_{N,2}^0 = \rho(\mathbf{C}^0(N))\mathbf{x}^{\mathsf{T}}\mathbf{e}_{N,2}^0 + v\mathbf{x}^{\mathsf{T}}\mathbf{b}_N^0 \tag{177}$$

Using (177) we have

$$1 - \rho(\mathbf{C}^0(N)) = \frac{v\mathbf{x}^\mathsf{T}\mathbf{b}_N^0}{\mathbf{x}^\mathsf{T}\mathbf{e}_{N,2}^0} \geq (\hat{K}\log^4 N)^{-1} \tag{178}$$

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

$$\rho(\mathbf{C}(N)) \leq 1 + (\hat{K}\log^4 N)^{-1} \tag{179}$$

For each N we rearrange the entries of $\mathbf{e}_{N,2}$, $v\mathbf{b}_m + \mathbf{w}_2^e$, and $\mathbf{C}(n)$ depending on whether the i^{th} entry of $v\mathbf{b}_m + \mathbf{w}_2^e$ is greater than, or less than or equal to zero. We can therefore assume

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11}(N) & \mathbf{C}_{12}(N) \\ \mathbf{C}_{21}(N) & \mathbf{C}_{22}(N) \end{pmatrix} \tag{180}$$

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 $vb_{N,i} + \mathbf{w}_{2,i}^e \leq 0$, then necessarily $vb_{N,i} \leq |\mathbf{w}_N^e| \leq K_1(\log n)^{-(3+p)}$, and so from (176) 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 $|\lambda| \geq 1/2$, we have $(\mathbf{C}_{22}(N) - \lambda\mathbf{I}_N)^{-1}$ existing with entries uniformly bounded. We have

$$\det(\mathbf{C}(N) - \lambda\mathbf{I}) = \det\left[\begin{pmatrix} \mathbf{I} & -\mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda\mathbf{I})^{-1} \\ 0 & \mathbf{I} \end{pmatrix}\begin{pmatrix} \mathbf{C}_{11}(N) - \lambda\mathbf{I} & \mathbf{C}_{12}(N) \\ \mathbf{C}_{21}(N) & \mathbf{C}_{22}(N) - \lambda\mathbf{I} \end{pmatrix}\right] \tag{181}$$

$$= \det\begin{pmatrix} \mathbf{C}_{11}(N) - \lambda\mathbf{I} - \mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda\mathbf{I})^{-1}\mathbf{C}_{21}(N) & 0 \\ \mathbf{C}_{21}(N) & \mathbf{C}_{22}(N) - \lambda\mathbf{I} \end{pmatrix} \tag{182}$$

$$= \det(\mathbf{C}_{11}(N) - \lambda\mathbf{I} - \mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda\mathbf{I})^{-1}\mathbf{C}_{21}(N))\det(\mathbf{C}_{22}(N) - \lambda\mathbf{I}) \tag{183}$$

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

$$\det(\mathbf{C}_{11}(N) - \lambda\mathbf{I} - \mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda\mathbf{I})^{-1}\mathbf{C}_{21}(N)) \tag{184}$$

must be zero.

Notice that from (176), 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 (176), we see that $\rho(\mathbf{C}(N)) \leq K_1 \log^4 N$. The determinant in (184) can be written as

$$\det(\mathbf{C}_{11}(N) - \lambda\mathbf{I}) + g(\lambda) \tag{185}$$

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 (176) we see that for all $|\lambda| \geq 1/2$, $g(\lambda)$ can be made smaller than any negative power of $\log N$ by making p sufficiently large. Choose p so that $|g(\lambda)| < (\widehat{K}\log N)^{-4k_1}$ for these λ. It is clear that any $p > 8k_1 + 4$ will suffice. Let $\lambda_1, \ldots, \lambda_{k_1}$ 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

$$|\det(\mathbf{C}_{11}(N) - \lambda\mathbf{I})| = |\prod_{i=1}^{k_1}(\lambda_i - \lambda)| \tag{186}$$

$$> (\widehat{K}\log N)^{-4k_1} \tag{187}$$

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 (179).

As before we see that

$$|a_{ij}(N)| \leq c_{ij}^{1/2}(N) c_{ij}^{0\,1/2}(N) \tag{188}$$

Therefore, from (178), (179), and Lemmas 10 and 11, we have for all N large

$$\rho(\mathbf{A}(N)) \leq \left(\frac{\widehat{K}^2\log^8 N - 1}{\widehat{K}^2\log^8 N}\right)^{\frac{1}{2}} \tag{189}$$

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

$$\mathbf{e}_N - \mathbf{e}_N^0 = (\mathbf{I} - \mathbf{A}(N))^{-1}\mathbf{w}^e \tag{190}$$

By (72) and (80) we have the entries of $\mathbf{A}(N)$ bounded by $K_1\log^4 N$. Notice also, from (189)

$$|\det(\mathbf{I} - \mathbf{A}(N))| \geq (1 - \rho(\mathbf{A}(N))^K \geq \left(\widehat{K}^2\log^8 N\left(1 + \frac{\widehat{K}^2\log^8 N - 1}{\widehat{K}^2\log^8 N}\right)^{\frac{1}{2}}\right)^{-K} \geq (2\widehat{K}^2\log^8 N)^{-K} \tag{191}$$

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

$$\frac{(K-1)!K_1(\log N)^{4(K-1)}}{|\det(\mathbf{I} - \mathbf{A}(N))|} \leq K_3(\log N)^{12K-4} \tag{192}$$

Therefore, since $p \geq 12K - 7$ ($> 8k_1 + 4$), (165) follows on this realization, an event which occurs with probability one.

Letting $m_N^0 = \frac{1}{N}\operatorname{tr}\mathbf{D}^{0^{-1}}$, we have

$$m_N - m_N^0 = \vec{\gamma}^{\mathsf{T}}(\mathbf{e}_N - \mathbf{e}_N^0) \tag{193}$$

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

$$\gamma_j = \int \frac{c_N\tau^2}{(1 + c_N\tau e_{N,j})(1 + c_n\tau e_{N,j}^0)} dF^{\mathbf{T}_N}(\tau)\frac{\operatorname{tr}\mathbf{D}^{-1}\mathbf{R}_j\mathbf{D}^{0^{-1}}}{N} \tag{194}$$

From (72) and (80) we get each $|\gamma_j| \leq c_N|z|^2 v^{-4}\log^3 N$. Therefore from (165) and the fact that $w_N^m \to 0$, we have

$$m_N - m_N^0 \to 0 \tag{195}$$

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

This completes the proof. ■

Appendix B
Proof of Theorem 2

We first prove that $\mathcal{V}^{(0)}(x)$ as defined in Equation (22) verifies

$$\mathcal{V}^{(0)}(x) = \int_x^{\infty} \left(\frac{1}{w} - m_N^{(0)}(-w) \right) dw \tag{196}$$

and then we prove that, under the conditions of Theorem 2, $\mathcal{V}^{(0)}(x)$ defined as such verifies

$$\mathcal{V}^{(0)}(x) - \mathcal{V}(x) \xrightarrow{\text{a.s.}} 0 \tag{197}$$

A. *Proof of* (196)

First, observe that we can rewrite $e_i(z)$ under the symmetric form,

$$e_i(z) = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(-z \left[\mathbf{I}_N + \sum_{k=1}^{K} \delta_k \mathbf{R}_k \right] \right)^{-1} \tag{198}$$

$$\delta_i(z) = \frac{1}{n_i} \operatorname{tr} \mathbf{T}_i \left(-z \left[\mathbf{I}_{n_i} + c_i e_i(z) \mathbf{T}_i \right] \right)^{-1} \tag{199}$$

and then for $m_N^{(0)}(z)$,

$$m_N^{(0)}(z) = \frac{1}{N} \operatorname{tr} \left(-z \left[\mathbf{I}_N + \sum_{k=1}^{K} \delta_k \mathbf{R}_k \right] \right)^{-1} \tag{200}$$

Now, notice that

$$\frac{1}{z} - m_N^{(0)}(-z) = \frac{1}{N} \left((z\mathbf{I})^{-1} - \left(z \left[\mathbf{I}_N + \sum_{k=1}^{K} \delta_k \mathbf{R}_k \right] \right)^{-1} \right) \tag{201}$$

$$= \sum_{k=1}^{K} \delta_k(-z) \cdot e_k(-z) \tag{202}$$

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

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

$$\frac{d}{dz} \frac{1}{N} \log\det \left(\mathbf{I}_{n_k} + c_k e_k(-z) \mathbf{T}_k \right) = -z \cdot e_k'(-z) \cdot \delta_k(-z) \tag{203}$$

$$\frac{d}{dz} \left(z \sum_{k=1}^{K} \delta_k(-z) e_k(-z) \right) = \sum_{k=1}^{K} \delta_k(-z) e_k(-z) - z \sum_{k=1}^{K} \delta_k'(-z) \cdot e_k(-z) + \delta_k(-z) \cdot e_k'(-z) \tag{204}$$

Combining the last three lines, we have

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

$$\frac{d}{dz} \left[-\frac{1}{N} \log\det \left(\mathbf{I}_N + \sum_{k=1}^{K} \delta_k(-z) \mathbf{R}_k \right) - \sum_{k=1}^{K} \frac{1}{N} \log\det \left(\mathbf{I}_{n_k} + c_k e_k(-z) \mathbf{T}_k \right) + z \sum_{k=1}^{K} \delta_k(-z) e_k(-z) \right] \tag{205}$$

which after integration leads to

$$\int_z^{+\infty}\left(\frac{1}{w}-m_N^{(0)}(-w)\right)dw =$$
$$\frac{1}{N}\log\det\left(\mathbf{I}_N+\sum_{k=1}^K\delta_k(-z)\mathbf{R}_k\right)+\sum_{k=1}^K\frac{1}{N}\log\det\left(\mathbf{I}_{n_k}+c_ke_k(-z)\mathbf{T}_k\right)-z\sum_{k=1}^K\delta_k(-z)e_k(-z) \quad (206)$$

which is exactly the right-hand side of (22).

B. *Proof of* (197)

Consider now the existence of a nonrandom α and for each N a non-negative integer r_N for which

$$\max_{i\leq K}\max(\lambda_{r_N+1}^{\mathbf{T}_i},\lambda_{r_N+1}^{\mathbf{R}_i})\leq\alpha \quad (207)$$

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

$$\lambda_{2r_N+1}^{\mathbf{R}_i^{\frac{1}{2}}\mathbf{X}_i\mathbf{T}_i\mathbf{X}_i^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}} = (s_{2r_N+1}^{\mathbf{R}_i^{\frac{1}{2}}\mathbf{X}_i\mathbf{T}_i^{\frac{1}{2}}})^2 \quad (208)$$
$$\leq\alpha^2\|\mathbf{X}_i\mathbf{X}_i^{\mathsf{H}}\| \quad (209)$$

and then we have, from Lemma 15,

$$\lambda_{2Kr_N+1}^{\mathbf{B}_N}\leq\alpha^2(\|\mathbf{X}_1\mathbf{X}_1^{\mathsf{H}}\|+\cdots+\|\mathbf{X}_K\mathbf{X}_K^{\mathsf{H}}\|) \quad (210)$$

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. [33]),

$$\limsup_N\lambda_{2Kr_N+1}^{\mathbf{B}_N}\leq\limsup_N\alpha^2(\|\mathbf{X}_1'\mathbf{X}_1'^{\mathsf{H}}\|+\cdots+\|\mathbf{X}_K'\mathbf{X}_K'^{\mathsf{H}}\|)$$
$$\leq\alpha^2K(b/a)(1+\sqrt{a})^2 \quad (211)$$

Let a^0 be any real greater than $\alpha^2K(b/a)(1+\sqrt{a})^2$.

Since $\mathbf{S}=0$ here, it follows as in [4] that $\{F^{\mathbf{B}_n}\}$ is almost surely tight. Let F_N^0 denote the distribution function having Stieltjes transform $m_N^{(0)}$, and let f on $[0,\infty)$ be a continuous function. Then the function

$$f_{a^0}(x)=\begin{cases} f(x) & ,\ x\leq a^0 \\ f(a^0) & ,\ x>a^0\end{cases} \quad (212)$$

is bounded and continuous. Therefore, with probability 1,

$$\int f_{a^0}(x)dF^{\mathbf{B}_N}(x)-\int f_{a^0}(x)dF_N^0(x)\to 0 \quad (213)$$

as $N\to\infty$.

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

$$\frac{1}{N}\sum_{\lambda_i \le a^0} f(\lambda_i) - \int_{[0,a^0]} f(x)dF_N^0(x) \to 0 \tag{214}$$

as $N \to \infty$.

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

$$\|\mathbf{B}_n\| \le b_N^2(\|\mathbf{X}_1'\mathbf{X}_1'^{\mathsf{H}}\| + \cdots + \|\mathbf{X}_K'\mathbf{X}_K'^{\mathsf{H}}\|) \tag{215}$$

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

$$\int f(x)dF^{\mathbf{B}_n}(x) - \frac{1}{N}\sum_{\lambda_i \le a^0} f(\lambda_i) \to 0 \tag{216}$$

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

$$\int f(x)dF^{\mathbf{B}_N}(x) - \int_{[0,a^0]} f(x)dF_N^0(x) \to 0 \tag{217}$$

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^0 as $j \to \infty$. Properties on the eigenvalues of $\mathbf{B}_{N,j}$ will thus yield properties of F_N^0.

By considering the bound on $\|\mathbf{B}_{n,j}\|$ analogous to (215), we must have $F_N^0(a_N) = 1$ for all N large.

Similar to (211) we see that, with probability 1

$$\limsup_j \lambda_{2Kjr_N+1}^{\mathbf{B}_{N,j}} \le a^2((1+\sqrt{c_1})^2 + \cdots + (1+\sqrt{c_K})^2) \tag{218}$$

this latter number being less than a^0 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. [34])

$$\liminf_N P_N(G) \ge P(G) \tag{219}$$

for any open set G. Thus, with $G = (a^0, \infty)$ we see that, with probability 1, for all N large

$$F_N^0((a^0,\infty)) = 1 - F_N^0(a^0) \le \liminf_j F^{\mathbf{B}_{N,j}}((a^0,\infty)) \tag{220}$$

$$\le 2Kr_N/N \tag{221}$$

Therefore, for all N large

$$\int_{(a^0,\infty)} f(x)dF_N^0(x) \le f(a_N)2Kr_N/N \to 0 \tag{222}$$

as $N \to \infty$.

Therefore, we conclude that, $\int f(x)dF_N^0(x)$ is bounded, and with probability 1

$$\int f(x)dF^{\mathbf{B}_N}(x) - \int f(x)dF_N^0(x) \to 0 \tag{223}$$

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

APPENDIX C
PROOF OF PROPOSITION 1

The proof stems from the following result,

Proposition 3: $f(\mathbf{P}_1, \dots, \mathbf{P}_K)$ is a strictly concave matrix in the Hermitian nonnegative definite matrices $\mathbf{P}_1, \dots, \mathbf{P}_K$, if and only if, for any couples $(\mathbf{P}_{1_a}, \mathbf{P}_{1_b}), \dots, (\mathbf{P}_{K_a}, \mathbf{P}_{K_b})$ of Hermitian nonnegative definite matrices, the function

$$\phi(\lambda) = f\left(\lambda \mathbf{P}_{1_a} + (1-\lambda)\mathbf{P}_{1_b}, \dots, \lambda \mathbf{P}_{K_a} + (1-\lambda)\mathbf{P}_{K_b}\right) \tag{224}$$

is strictly concave.

Let us use a similar notation as in (31) of the capacity,

$$\bar{I}(\lambda) = I(\lambda \mathbf{P}_{1_a} + (1-\lambda)\mathbf{P}_{1_b}, \dots, \lambda \mathbf{P}_{|\mathcal{S}|_a} + (1-\lambda)\mathbf{P}_{|\mathcal{S}|_b}) \tag{225}$$

and consider a set $(\delta_k, e_k, \mathbf{P}_1, \dots, \mathbf{P}_{|\mathcal{S}|})$ which satisfies the system of equations (31)-(33). Then, from remark (36) and (37),

$$\frac{dI}{d\lambda} = \sum_{k \in \mathcal{S}} \frac{\partial \bar{V}}{\partial \delta_k}\frac{\partial \delta_k}{\partial \lambda} + \frac{\partial \bar{V}}{\partial e_k}\frac{\partial e_k}{\partial \lambda} + \frac{\partial \bar{V}}{\partial \lambda} \tag{226}$$

$$= \frac{\partial \bar{V}}{\partial \lambda} \tag{227}$$

where

$$\bar{V} : (\delta_1, \dots, \delta_{|\mathcal{S}|}, e_1, \dots, e_{|\mathcal{S}|}, \lambda) \mapsto \bar{I}(\lambda) \tag{228}$$

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

$$\frac{\partial^2 \bar{V}}{\partial \lambda^2} = -\sum_{i \in \mathcal{S}} (c_i^2 e_i^2) \frac{1}{N} \operatorname{tr}\left(\mathbf{I} + c_i e_i \mathbf{R}_i \mathbf{P}_i\right)^{-2} \left(\mathbf{R}_i(\mathbf{P}_{i_a} - \mathbf{P}_{i_b})\right)^2 \tag{229}$$

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 $\mathbf{P}_{i_a} \neq \mathbf{P}_{i_b}$, $\bar{I}'' < 0$. Then, from Proposition 3, the deterministic approximate on the right-hand side of (27) is strictly concave in $\mathbf{P}_1, \dots, \mathbf{P}_{|\mathcal{S}|}$ if any of the $\mathbf{R}_i$ matrices is invertible.

APPENDIX D
USEFUL LEMMAS

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

The statements in the following Lemma are well-known

Lemma 1: 1) For rectangular matrices $\mathbf{A}$, $\mathbf{B}$ of the same size,

$$\operatorname{rank}(\mathbf{A} + \mathbf{B}) \le \operatorname{rank}(\mathbf{A}) + \operatorname{rank}(\mathbf{B}) \tag{230}$$

2) For rectangular matrices $\mathbf{A}$, $\mathbf{B}$ for which $\mathbf{AB}$ is defined,

$$\operatorname{rank}(\mathbf{AB}) \leq \min(\operatorname{rank}(\mathbf{A}), \operatorname{rank}(\mathbf{B})) \tag{231}$$

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 [4]) For $N \times N$ Hermitian matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\|F^{\mathbf{A}} - F^{\mathbf{B}}\| \leq \frac{1}{N}\operatorname{rank}(\mathbf{A} - \mathbf{B}) \tag{232}$$

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)

$$\|F^{\mathbf{S}+\mathbf{AQBQ}^{\mathsf{H}}\mathbf{A}} - F^{\mathbf{S}+\mathbf{A}\overline{\mathbf{Q}}\mathbf{B}\overline{\mathbf{Q}}^{\mathsf{H}}\mathbf{A}}\| \leq \frac{2}{N}\operatorname{rank}(\mathbf{Q} - \overline{\mathbf{Q}}) \tag{233}$$

2)

$$\|F^{\mathbf{S}+\mathbf{AQBQ}^{\mathsf{H}}\mathbf{A}} - F^{\mathbf{S}+\overline{\mathbf{A}}\mathbf{QBQ}^{\mathsf{H}}\overline{\mathbf{A}}}\| \leq \frac{2}{N}\operatorname{rank}(\mathbf{A} - \overline{\mathbf{A}}) \tag{234}$$

and

3)

$$\|F^{\mathbf{S}+\mathbf{AQBQ}^{\mathsf{H}}\mathbf{A}} - F^{\mathbf{S}+\mathbf{AQ}\overline{\mathbf{B}}\mathbf{Q}^{\mathsf{H}}\mathbf{A}}\| \leq \frac{1}{N}\operatorname{rank}(\mathbf{B} - \overline{\mathbf{B}}) \tag{235}$$

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,

$$\mathbf{r}^{\mathsf{H}}(\mathbf{A} + \tau\mathbf{r}\mathbf{r}^{\mathsf{H}})^{-1} = \frac{1}{1 + \tau\mathbf{r}^{\mathsf{H}}\mathbf{A}^{-1}\mathbf{r}}\mathbf{r}^{\mathsf{H}}\mathbf{A}^{-1} \tag{236}$$

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 Lemma 2.6 of [4]

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

$$\left|\operatorname{tr}\left((\mathbf{B} - z\mathbf{I}_N)^{-1} - (\mathbf{B} + \mathbf{r}\mathbf{r}^{\mathsf{H}} - z\mathbf{I}_N)^{-1}\right)\mathbf{A}\right| = \left|\frac{\mathbf{r}^{\mathsf{H}}(\mathbf{B} - z\mathbf{I}_N)^{-1}\mathbf{A}(\mathbf{B} - z\mathbf{I}_N)^{-1}\mathbf{r}}{1 + \mathbf{r}^{\mathsf{H}}(\mathbf{B} - z\mathbf{I}_N)^{-1}\mathbf{r}}\right| \leq \frac{\|\mathbf{A}\|}{v}. \tag{237}$$

From Lemma 2.2 of [17], and Theorems A.2, A.4, A.5 of [18], 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 [4], 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, then

$$\mathrm{E}|\mathbf{y}^{\mathsf{H}}\mathbf{A}\mathbf{y}|^6 \leq K\|\mathbf{A}\|^6 N^3 \log^{12} N \tag{238}$$

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 [28],

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 $\mathbf{x}, \mathbf{b} \in^N$ 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 Theorem 8.1.18 of [29],

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

$$\rho(\mathbf{A}) \leq \rho((|a_{ij}|)) \leq \rho(\mathbf{B}) \tag{239}$$

Also, from Lemma 5.7.9 of [30],

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

$$\rho((a_{ij}^{\frac{1}{2}} b_{ij}^{\frac{1}{2}})) \leq (\rho(\mathbf{A}))^{\frac{1}{2}} (\rho(\mathbf{B}))^{\frac{1}{2}} \tag{240}$$

And, Theorems 8.2.2 and 8.3.1 of [29],

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 [30], we also need Theorem 6.1.1,

*Lemma 13: Ger***v***sgorin'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 [16],

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 [32]

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 $i >$ rank(A). Let m, n be arbitrary non-negative integers. Then for $\mathbf{A}$, $\mathbf{B}$ rectangular of the same size

$$s_{m+n+1}^{\mathbf{A}+\mathbf{B}} \leq s_{m+1}^{\mathbf{A}} + s_{n+1}^{\mathbf{B}} \tag{241}$$

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

$$s_{m+n+1}^{AB} \leq s_{m+1}^{A} s_{n+1}^{B} \tag{242}$$

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

$$s_{Kr+1}^{\mathbf{A}_1 + \cdots + \mathbf{A}_K} \leq s_{r+1}^{\mathbf{A}_1} + \cdots + s_{r+1}^{\mathbf{A}_K} \tag{243}$$

ACKNOWLEDGMENT

Silverstein's work is supported by the U.S. Army Research Office under Grant W911NF-05-1-0244. Debbah's work is partially supported by the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications NEWCOM++.

The authors would like to thank Walid Hachem for the fruitful discussions we shared concerning some results of this paper.

REFERENCES

[1] S. Sesia, I. Toufik and M. Baker, "LTE: The UMTS Long Term Evolution, From Theory to Practice," Wiley and Sons, 2009.

[2] A. M. Tulino, A. Lozano and S. Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," IEEE Trans. on Information Theory, vol. 51, no. 7, pp. 2491-2509, 2005.

[3] 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. on Information Theory, vol. 54, no. 5, pp. 2123-2138, 2008.

[4] J. Silverstein and Z. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, issue 2, pp. 175–192, 1995

[5] G. Taricco, E. Riegler, "On the Ergodic Capacity of the Asymptotic Separately-Correlated Rician Fading MIMO Channel with Interference," IEEE International Symposium on Information Theory, pp. 531-535, 2007

[6] G. Foschini and M. Gans, "On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas," Wireless Personal Communications, vol. 6, no. 3, pp. 311-335, 1998.

[7] E. Telatar, "Capacity of multi-antenna Gaussian channels," European transactions on telecommunications, vol. 10, no. 6, pp. 585-595, 1999.

[8] A. Paulraj, R. Nabar and D. Gore, "Introduction to Space-Time Wireless Communications," Cambridge University Press, 2003.

[9] C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, "Capacity Scaling in MIMO Wireless Systems under Correlated Fading," IEEE Trans. on Information Theory, pp. 637650, 2002.

[10] A. Moustakas, S. Simon and A. Sengupta, "MIMO Capacity Through Correlated Channels in the Presence of Correlated Interferers and Noise: A (Not So) Large N Analysis," IEEE Trans. on Information Theory, vol. 49, no. 10, 2003.

[11] C. K. Wen, Y. N. Lee, J. T. Chen, and P. Ting, "Asymptotic spectral efficiency of MIMO multiple-access wireless systems exploring only channel spatial correlations," IEEE Transactions on Signal Processing, vol. 53, no. 6, pp. 2059-2073, 2005.

[12] D. Guo and S. Verdú, "Multiuser Detection and Statistical Physics," Communications on Information and Network Security, Kluwer Academic Publishers, 2003.

[13] S. Shamai and A. Wyner, "Information-Theoretic Considerations for Symmetric, Cellular, Multiple-Access Fading Channels - Part I," IEEE Trans. on Information Theory, vol. 43, no. 6, 1997.

[14] B. Zaidel, S. Shamai and S. Verdú, "Multicell Uplink Spectral Efficiency of Coded DS-CDMA With Random Signatures," IEEE Journal on Selected Areas in Communications, vol. 19, no. 8, 2001.

[15] Z. Bai and J. Silverstein, "On the signal-to-interference-ratio of CDMA systems in wireless communications," Annals of Applied Probability vol. 17 no. 1, pp. 81-101, 2007.

[16] E. C. Titchmarsh, "The Theory of Functions," Oxford University Press, 1939.

[17] J. A. Shohat and J. D. Tamarkin, "The Problem of Moments," American Mathematical Society, Providence, RI, 1970.

[18] M. K. Krein and A. A. Nudelman, "The Markov Moment Problem and Extremal Problems," American Mathematical Society, Providence, RI, 1997.

[19] R. B. Dozier and J. W. Silverstein, "On the Empirical Distribution of Eigenvalues of Large Dimensional Information-Plus-Noise Type Matrices," Journal of Multivariate Analysis 98(4), pp. 678-694, 2007.

[20] S. Vishwanath, N. Jindal and A. Goldsmith, "Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels," IEEE Trans. on Information Theory, vol. 49, no. 10, 2003.

[21] P. Viswanath and D. Tse, "Sum Capacity of the Multiple Antenna Gaussian Broadcast Channel and Uplink-Downlink Duality," IEEE Trans. on Information Theory, vol. 49, no. 8, pp. 19121923, 2003.

[22] S. Verdù, "Multiple-access channels with memory with and without frame synchronism," IEEE Trans. on Information Theory, vol. 35, pp. 605-619, 1989.

[23] H. Weingarten, Y. Steinberg and S. Shamai, "The Capacity Region of the Gaussian Multiple-Input Multiple-Output Broadcast Channel," IEEE Trans. on Information Theory, vol. 52, no. 9, pp. 3936-3964, 2006.

[24] A. Soysal and S. Ulukus "Optimality of Beamforming in Fading MIMO Multiple Access Channels," *to be published*.

[25] A.M. Tulino, A. Lozano and S. Verdu, "Impact of antenna correlation on the capacity of multiantenna channels," IEEE Trans. on Information Theory, vol. 51, no. 7, pp. 2491-2509, 2005.

[26] W. Hachem, Ph. Loubaton and J. Najim, "Deterministic Equivalents for Certain Functionals of Large Ranfom Matrices," The Annals of Applied Probability, Vol. 17, No. 3, 875930, 2007.

[27] J. Dumont, S. Lasaulce, W. Hachem, Ph. Loubaton and J. Najim, "On the Capacity Achieving Covariance Matrix for Rician MIMO Channels: An Asymptotic Approach", submitted to IEEE Trans. on Information Theory, Octggober 2007, revised in October 2008.

[28] E. Seneta, "Non-negative Matrices and Markov Chains," Second Edition, Springer Verlag New York, 1981.

[29] R.A. Horn and C.R. Johnson, "Matrix Analysis," Cambridge University Press, 1985.

[30] R.A. Horn and C.R. Johnson, "Topics in Matrix Analysis," Cambridge University Press, 1991.

[31] Z. D. Bai and Y. Q. Yin, "Limit of the smallest eigenvalue of a large dimensional sample covariance matrix," The annals of Probability, pp. 1275-1294, Institute of Mathematical Statistics, 1993.

[32] K. Fan, "Maximum properties and inequalities for the eigenvalues of completely continuous operators," Proc. Nat. Acad. Sci. U.S.A. 37 760-766, 1951.

[33] Z. Bai and J. W. Silverstein, "Spectral Analysis of Large Dimensional Random Matrices," Science Press, Beijing, 2006.

[34] P. Billingsley, "Convergence of Probability Measures," John Wiley & Sons, Second Revised Edition, 1999.

[35] M. Debbah and R. R. Muller, "MIMO channel modeling and the principle of maximum entropy," IEEE Transactions on Information Theory, vol. 51, no. 5, pp. 1667-1690, 2005.