A Deterministic Equivalent for the Capacity Analysis of Correlated Multi-User MIMO Channels

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

Abstract

This article provides capacity expressions for multi-user and multi-cell wireless communication schemes when the transmitters and receivers are equipped with multiple antennas and when the transmission channel has a certain correlation profile. In mathematical terms, this contribution provides novel deterministic equivalents for the Stieltjes and Shannon transforms of a class of large dimensional random matrices. These results are of practical relevance to evaluate the rate performance of communication channels with multiple users, multiple cells and with transmit and receive correlation at all communication pairs. In particular, we analyse the per-antenna achievable rates for these communication systems which, for practical purposes, is a relevant measure of the trade-off between rate performance and operating cost of every antenna. We study specifically the per-antenna rate regions of (i) multi-antenna multiple access channels and broadcast channels, as well as the capacity of (ii) multi-antenna multi-cell communications with inter-cell interference. Theoretical expressions of the per-antenna mutual information are obtained for these models, which extend previous results on multi-user multi-antenna performance without channel correlation to the more realistic Kronecker channel model. From an information theoretic viewpoint, this article provides, for scenario (i), a deterministic approximation of the per-antenna rate achieved in every point of the MAC and BC rate regions, a deterministic approximation of the ergodic per-antenna capacity with optimal precoding matrices in the uplink MAC and an iterative water-filling algorithm to compute the optimal precoders, while, for scenario (ii), this contribution provides deterministic approximations for the mutual information of single-user decoders and the capacity of minimum mean square error (MMSE) decoders. An original feature of this work is that the deterministic equivalents are proven asymptotically exact, as the system dimensions increase, even for strong correlation at both communication sides. The above results are validated by Monte Carlo simulations.

I. Introduction

When mobile networks were expected to run out of power and frequency resources while being simultaneously subject to a demand for higher transmission rates, Foschini [1] introduced the idea of multiple input multiple output (MIMO) systems and Telatar [2] predicted a growth of the capacity performance by a factor $\min(N, n)$,

[†] 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

compared to single-antenna schemes, for communications between an n-antenna transmitter and an N-antenna receiver. This capacity gain stands when the propagation channel matrix model is formed of independent and identically distributed (i.i.d.) Gaussian entries. In practical systems though, this linear multiplexing gain can only be achieved for large signal-to-interference plus noise ratios (SINR) and for uncorrelated transmit and receive antenna arrays at both communication sides. Today, in spite of this remark, the scarcity of available frequency resources has led to a widespread incentive for MIMO communications. Mobile terminal designers now embed more and more antennas in small devices. Due to space limitations mainly, this inevitably spawns non-negligible channel correlation and, thus, non-negligible effects on the achievable transmission rates. Since MIMO systems come along with a tremendous increase in signal processing requirements and, therefore, an even larger increase in power consumption, both infrastructure and mobile terminal manufacturers need to accurately assess the exact cost of increasing the achievable bit rates by adding more antennas on volume limited devices. The analysis of the exact throughput gain incurred by extra antennas is therefore paramount to evaluate the energy efficiency of multi-antenna devices. The first purpose of the present article is to evaluate the per-antenna achievable rate, which we further refer to as the *antenna efficiency*, for different communication models involving multiple users or multiple cells. The antenna efficiency criterion comes in line with the current incentive for energy-efficient communications, that are foreseen to predominate future telecommunication research interest.

Multi-cell and multi-user systems are among the scenarios of main interest to cellular service providers. Although alternative communication models could be treated, the present article investigates the following two wireless communication systems:

1) multiple access channels (MAC) in which K 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 terminal users. While the major scientific breakthroughs in multi-antenna broadcast channels are quite recent, e.g. [19], the practical applications are foreseen to arise in a near future, with e.g. the 3GPP long term evolution standard [3]
2) single-user decoding and minimum mean square error (MMSE) decoding [7] 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. Under these conditions, the achievable rates for every user in a cell, assuming no intra-cell interference, are the capacity of the single-user decoding scheme in which interfering signals are treated as Gaussian noise with a known variance. However, single-user decoders are difficult to implement and are often replaced in practical applications by suboptimal linear decoders, such as linear MMSE decoders. These decoders are attractive as they are known to maximize the signal-to-interference plus noise ratio (SINR) experienced at the receiver.

The achievable rate region of the multi-antenna MAC and BC have been known since the successive contributions [19]-[20], which established an important duality link between the MAC rate regions and the BC rate regions in both single antenna and MIMO channels, when the instantaneous channel realizations are assumed to be perfectly known at both communication sides. To achieve perfect channel state information at both communication ends, the channel must be somewhat static during a sufficient long period and is often referred to as a *block-fading channel*. Overall, communication channels can often be modelled as particular realizations of a stochastic process, in which case it is convenient to identify the parameters in the stochastic model that account for the communication rate performance. The mathematical field of large dimensional random matrices is particularly suited to this end, as it can provide approximations of achievable rates as a function of the relevant channel parameters only, e.g. as a function of the long-term transmit and receive channel covariance matrices in the present situation, or as a function of the deterministic line of sight components in Rician models. The earliest notable result in line with the present study is due to Tulino et al. [4], who provide an expression of the asymptotic mutual information of point-to-point MIMO communications when the random channel matrix is composed of i.i.d. Gaussian entries. The authors also provide an expression of the ergodic capacity-achieving power allocation policy at the base station. In [38], Hochwald et al. derive a central limit result of the asymptotic capacity result obtained in [4], providing therefore an asymptotic expression of the outage capacity of large MIMO uncorrelated channels. In [5], Peacock et al. extend the result from [4] in the direction of multi-user communications by considering the sum of K Gram matrices $\mathbf{H}_k\mathbf{H}_k^{\mathsf{H}}$, $k \in \{1, \dots, K\}$, of channels $\mathbf{H}_k$ with independent Gaussian entries and separable variance profile. The asymptotic eigenvalue distribution of this matrix model is derived (which is in fact a consequence of an earlier result from Girko [39]), but neither any explicit expression of the sum rate is provided as in [4], nor any ergodic capacity maximizing policy is derived. In [23], Soysal et al. derive the sum rate maximizing power allocation policy for a finite number of antennas at all transmit and receive devices in the case of K users whose channels $\mathbf{H}_k$, $1 \leq k \leq K$, are perfectly known at the transmitters and are modelled as *Kronecker channels*. We recall that Kronecker channels are made of a matrix with i.i.d. Gaussian entries multiplied both on the left and on the right by deterministic Hermitian matrices, hereafter referred to as the (left and right) *correlation matrices*. Those are more general than matrices of independent Gaussian entries with a separable variance profile, which can be seen as Gaussian i.i.d. matrices multiplied on the left and on the right by diagonal matrices. Contrary to [4], [23] does not provide a theoretical large dimensional analysis of the resulting capacity, and makes the strong assumption that all receive correlation matrices are equal. When the receive correlation matrices do not have the same eigenspace, determining the channel capacity, both in the finite and asymptotic regimes, is more complex and requires different mathematical tools. Those tools allow us in the present work to obtain a *deterministic equivalent* for every point in the per-antenna rate region of the MAC and BC. That is, for every deterministic precoding policy of the transmitters, we provide a deterministic approximation of the per-antenna achievable rate. This approximated value is more and more accurate as the system dimensions grow large. This is a consequence of our main result, stated in Theorem 2. We mention that the final formula of Theorem 2 is already found and used by Chen et al., Equation (32) in [10]. However, the latter is provided without proof, nor any rigorous hypotheses on the considered matrices, and stems in effect from a flawed usage of the previous Equation (6) in [10], which is only valid when all receive correlation matrices have the same eigenspace. Chen also provides the iterative water-filling algorithm which we shall introduce in the course of this article to derive the boundary of the ergodic rate region of MAC (see Table

II). The convergence of this algorithm to the correct capacity, which we will partly prove in the current article, is not provided in [10].

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 matrices. In [12], the authors carry out the performance analysis of TDMA-based networks with inter-cell interference. In [13], a random matrix approach is used to study large CDMA-based networks with inter-cell interference, basing their work on the Wyner model introduced in [12]. In our particular MIMO context, it is important to mention the work from Moustakas et al. [9] who propose an analytic solution to the single-user decoding problem with channel correlation and a single source of interference, using the replica method [11]. In the present article, we will extend the results from [9] to more interfering sources. Moreover, since the replica method is to this day not proven to be mathematically correct, we provide here a proof of the results in [9], under different hypotheses on the matrix model.

In real channels, each transmitter and each receiver is affected by different correlation patterns. Assuming those patterns mutually independent, independent of the propagation environment and known to both communication ends, the Kronecker channel is proven to be the most natural channel model [37]. The multi-user Kronecker channel model is more general than all previously described channel models: when all receive correlation matrices are equal, we fall back on the model in [23], when all correlation matrices are diagonal, we fall back on [5] and, when all correlation matrices are identity, we fall back on [4]. Nonetheless, the Kronecker model is only valid when no line-of-sight component is present in the channel, when a sufficiently large number of scatterers is found in the communication medium to justify the i.i.d. aspect of the inner Gaussian matrix, and when the channel is frequency flat on the transmission bandwidth. In their substantial contributions [25]-[27], Hachem et al. have extensively studied the point-to-point multi-antenna Rician channel model for which they provide a deterministic equivalent of the ergodic capacity [25], the corresponding ergodic capacity-achieving input covariance matrix [26] and a central limit theorem for the ergodic capacity [27]. We recall that Ricean channels are modelled as the sum of a deterministic line-of-sight matrix and a random matrix of independent entries with a variance profile. In [40], Moustakas et al. provide an expression of the mutual information in time varying frequency selective Rayleigh channels, using the replica method. This result has been recently proven by Dupuy et al. in a yet unpublished work. The same authors then derived the expression of the capacity maximizing precoding matrix for the frequency selective channel [41]. Part of the present study is inspired by the ideas in [41]. A more general frequency selective Rayleigh channel model with non-separable variance profile is studied in [28] by Rashidi et al. using alternative tools from free probability theory. The requirements from free probability theory on the studied matrices are more stringent, though, since Gaussian distribution must be assumed for the entries of the random matrices, while deterministic matrices in the model must have an eigenvalue distribution that converges weakly to a compactly supported distribution. Of practical interest is also the theoretical work of Tse [8] 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 are two theorems, contributing to the field of random matrix theory and enabling the evaluation of the per-antenna rate achieved at every point in the MAC and BC rate regions, as well as an iterative water-filling algorithm enabling the description of the boundaries of the *ergodic* rate region of the MAC channel, when all channels are modelled according to the Kronecker model.

The remainder of this paper is structured as follows: in Section II, we provide a short summary of our results and how they apply to multi-user and multi-cellular wireless communications. In Section III, our main two theorems are introduced. The complete proofs of both theorems are left to the appendices. In Section IV, the rate region of MAC and BC and the capacity of single-user decoding and MMSE decoding with inter-cell interference are studied. In this section, we also introduce our third main result: an iterative water-filling algorithm to describe the boundary of the ergodic rate region of the MAC. In Section V, we provide simulation results of the previously derived theoretical formulas. Finally, in Section VI, we give our conclusions.

Notation: In the following, boldface lower-case characters represent vectors, capital boldface characters denote matrices ($\mathbf{I}_N$ is the $N \times N$ identity matrix). X_{ij} denotes the (i,j) entry of $\mathbf{X}$. The Hermitian transpose is denoted $(\cdot)^{\mathsf{H}}$. The operators $\operatorname{tr}\mathbf{X}$, $|\mathbf{X}|$ and $\|\mathbf{X}\|$ represent the trace, determinant and spectral norm of matrix $\mathbf{X}$, respectively. The symbol $\mathbf{E}[\cdot]$ 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. The notation $x_n \xrightarrow{\text{a.s.}} x$ denotes the almost sure convergence of the sequence x_n to x.

II. Scope and Summary of Main Results

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

A. General Model

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

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

where $\mathbf{R}_k^{\frac{1}{2}} \in \mathbb{C}^{N \times N}$ and $\mathbf{T}_k^{\frac{1}{2}} \in \mathbb{C}^{n_k \times n_k}$ are the nonnegative Hermitian square roots of the Hermitian nonnegative matrices $\mathbf{R}_k$ and $\mathbf{T}_k$, respectively, and $\mathbf{X}_k \in \mathbb{C}^{N \times n_k}$ is a realization of a random matrix with Gaussian i.i.d. entries. The matrices $\mathbf{T}_k$ and $\mathbf{R}_k$ in this scenario model the correlation present in the channel at transmitter k and at the receiver, respectively. It is important to stress that those correlation patterns emerge both from the inter-antenna spacings on the volume limited devices and from the solid angles of *useful* transmitted and received signal energy; that is, even though the transmit antennas emit signals in an isotropic manner, only a limited solid angle of emission is effectively received, and the same holds for the receiver which captures signal energy from a limited solid angle. Without this second factor, it would make sense that all $\mathbf{R}_k$ matrices are equal at the receiver. This would mean that signals are received isotropically at the receiver, which is often too strong an assumption to characterize practical communication channels. This being said, one can assume physically identical and interchangeable *antennas* on each device. We therefore assume that the diagonal entries of $\mathbf{R}_k$ and $\mathbf{T}_k$, i.e. the variance of the channel fading on every antenna, are identical and, up to a scaling factor, equal to one. As a consequence, $\operatorname{tr}\mathbf{R}_k = N$ and $\operatorname{tr}\mathbf{T}_k = n_k$. We will see that under these trace constraints, the hypotheses made in Theorem 1, used to characterize the capacity of MMSE precoders, are always satisfied, therefore making Theorem 1 valid for all possible figures of correlation, including strongly correlated patterns. The hypotheses of Theorem 2, used to characterize the rate region of MAC and BC, require additional mild assumptions, making Theorem 2 valid for all but some unrealistic correlation matrices $\mathbf{R}_k$ and $\mathbf{T}_k$. These statements are of major importance and rather new since, in alternative contributions, e.g. [25], [26], it is usually assumed that the correlation matrices have uniformly bounded spectral norms (for all N). This physically means that only low correlation patterns are allowed; short distances between antennas ans small solid angles of energy propagation are therefore excluded. In the present work, this restriction is not needed. The counterpart of this interesting property is a reduction of the convergence rates of the derived deterministic equivalents, compared to those proposed in [25] and [26].

As will be evidenced in Sections IV-A and IV-B, most multi-cell or multi-user capacity performance rely on the so-called Stieltjes transform and Shannon transform of matrices $\mathbf{B}_N$ of the type

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

We study these matrices, using tools from the field of large dimensional random matrix theory [34]. The Stieltjes transform $m_N(z)$ of the Hermitian nonnegative definite matrix $\mathbf{B}_N \in \mathbb{C}^{N\times N}$ is defined, for $z \in \mathbb{C} \setminus \mathbb{R}^+$ as

$$m_N(z) = \int \frac{1}{\lambda - z} dF^{\mathbf{B}_N}(\lambda) \tag{3}$$

$$= \frac{1}{N} \operatorname{tr} (\mathbf{B}_N - z\mathbf{I}_N)^{-1} \tag{4}$$

where $F^{\mathbf{B}_N}$ denotes the distribution function of the eigenvalues of $\mathbf{B}_N$. The Stieltjes transform was originally used to characterize the asymptotic distribution of the eigenvalues of large dimensional random matrices. From a wireless communications point of view, it can be directly used to characterize the signal-to-interference plus noise ratio (SINR) of certain communication models. In the present work, the Stieltjes transform of $\mathbf{B}_N$ matrices defined in (2) will be used to approximate the SINR of MMSE decoders in single-user communications with inter-cell interference.

Then, there exists a link from the Stieltjes transform to the so-called Shannon transform $\mathcal{V}_N(x)$ of $\mathbf{B}_N$, defined for $x > 0$ as

$$\mathcal{V}_N(x) = \frac{1}{N} \log\det\left(\mathbf{I}_N + \frac{1}{x}\mathbf{B}_N\right) \tag{5}$$

$$= \int_0^{+\infty} \log\left(1 + \frac{\lambda}{x}\right) dF^{\mathbf{B}_N}(\lambda) \tag{6}$$

$$= \int_x^{+\infty} \left(\frac{1}{w} - m_N(-w)\right) dw. \tag{7}$$

The Shannon transform, named after Claude Shannon, is commonly used to provide approximations of capacity expressions in large dimensional systems. In the present work, the Shannon transform of $\mathbf{B}_N$ matrices will be used to provide a deterministic approximation of the achievable per-antenna rate for different communication models.

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

B. Main results

The main results of this work come as follows.

- We first introduce Theorem 1, which provides a deterministic equivalent $m_N^\circ(z)$ for the Stieltjes transform $m_N(z)$ of $\mathbf{B}_N$, under the assumption that N and n_k grow large but at the same rate and the distribution functions $\{F^{\mathbf{T}_k}\}_{n_k}$ and $\{F^{\mathbf{R}_k}\}_N$ form tight sequences [35]. This is, we provide an approximation $m_N^\circ(z)$ of $m_N(z)$ which does not depend on the realization of the $\mathbf{X}_k$ matrices and which is almost surely asymptotically exact when $N \to \infty$. The tightness hypothesis is the key assumption that allows degenerated $\mathbf{R}_k$ and $\mathbf{T}_k$ matrices to be valid in our framework, and that therefore allows us to study strongly correlated channel models.
- We then provide in Theorem 2 a deterministic equivalent $\mathcal{V}_N^\circ(x)$ for the Shannon transform $\mathcal{V}_N(x)$ of $\mathbf{B}_N$. For this theorem, the assumptions on the $\mathbf{R}_k$ and $\mathbf{T}_k$ matrices are only slightly more constraining and of marginal importance for practical purposes. Our results theoretically allow the largest eigenvalues of $\mathbf{T}_k$ or $\mathbf{R}_k$ to grow linearly with N, as the number of antennas increases, as long as the number of these large eigenvalues is of order $o(N)$ (Theorem 1 does not require this condition).

The major practical interest of Theorems 1 and 2 lies in the possibility to analyze mutual information expressions for multi-user multi-antenna channels, no longer as stochastic variables depending on the matrices $\mathbf{X}_k$ but as approximated deterministic quantities. The study of those quantities is in general simpler than the study of the stochastic expressions, even though the deterministic results are not closed-form expressions but solutions of implicit equations (see Section III). In particular, remember that we study here the trade-off 'throughput gain versus cost' of adding more antennas to the transmit or receive communication ends. For this reason, the typical figures of performance sought for are antenna efficiencies, i.e. the per-transmit antenna normalized capacity, sum rate or rate region. Those performance figures are related to the Stieltjes and Shannon transform of $\mathbf{B}_N$-like matrices. We do not provide in this study total rate expressions, i.e. N times the Shannon transform, for which asymptotic accuracy of the deterministic equivalents cannot be verified. Using different techniques and under more constraining channel conditions, alternative works have shown though that deterministic equivalents of the Shannon transform converge as $O(1/N^2)$, see e.g. [26] for the case of Rician channels.

In practical applications, such as the determination of the rate region of MAC and BC, Shannon transform expressions (5) are needed to evaluate the achievable rate for all points in the rate region, i.e. for all deterministic precoders. When all users have perfect channel state information, there exists a duality between MAC and BC rate regions. To determine the BC rate region, one therefore simply needs to treat the dual MAC uplink problem; see Section IV-A. In order to account for the effect of the transmit precoders in the channel model, the correlation matrices at the MAC transmitters will be replaced by the product of the channel correlation matrix and the precoding matrix. Nonetheless, the determination of an explicit form for the optimal precoding matrices which maximize (5) for block-fading channels is a very difficult problem both for finite N and in the asymptotic regime. To the authors' current knowledge, this has not been solved. In the ergodic sense, when the transmitters only have statistical state information about the time varying channel, it is usually possible to determine the precoders that reach the boundaries of the *ergodic* rate region. However, for partial channel state information in the system, MAC-BC duality no longer holds, so that the boundary of the *ergodic* rate region of BC cannot be determined.

In this article, we will provide (i) a deterministic equivalent for every point in the MAC and BC rate regions for all deterministic precoders in block-fading channels and (ii) the precoding matrices which maximize the deterministic equivalent $\mathcal{V}_N^{\circ}$ of the ergodic Shannon transform $\mathrm{E}\mathcal{V}_N$ of (5) in fast varying MAC channels when statistical channel state information is available at the transmitters. The reason why Theorem 1 and Theorem 2 are not able to determine the precoders that correspond to the MAC rate region boundary in the block-fading case is explained hereafter. When deterministic precoders are used, every point in the rate regions can be *estimated* by a deterministic equivalent, even on the boundary, for all finite system dimensions. This deterministic equivalent is (almost surely) asymptotically accurate if, as the system dimensions grow (in some predefined manner), the sequences of precoding matrices of growing dimensions satisfy some mild assumptions. One of these assumptions is that the precoding matrices, for growing dimensions, are chosen independently of the $\mathbf{X}_k$ matrices. The precoders that reach the rate region boundary of MAC block-fading channels are however explicitly built upon the $\mathbf{X}_k$ matrices. Such precoders do not satisfy the assumption of independence and our results therefore do not hold, i.e. the deterministic equivalents exist but are not asymptotically accurate in this case. In the ergodic sense though, optimal precoders are independent of the realizations of the random $\mathbf{X}_k$ matrices and can therefore be considered deterministic. Our results will therefore hold in this scenario. The precoders that maximize the deterministic equivalent of the ergodic sum rate, and which we characterize in this article, have the following interesting properties,

- their eigenspaces coincide respectively with the eigenspaces of the correlation matrices at the transmitters,
- their eigenvalues are solution of a classical optimization problem,
- we provide an iterative water-filling algorithm to determine these eigenvalues, which, upon convergence, is proved to converge to the correct solution.

Note that those precoders are not claimed to be the true sum-rate maximizing precoders, but only the matrices which maximize the *deterministic equivalent* of the ergodic sum rate. From this fact, it is easy to show that the difference between the rates achieved by the deterministic equivalent using those precoders and the true ergodic rates achieved using optimal precoders asymptotically goes to zero almost surely.

III. Mathematical Preliminaries

In this section, we first introduce Theorem 1, which provides a deterministic equivalent for the Stieltjes transform of matrices $\mathbf{B}_N$ defined in (2). The Shannon transform of $\mathbf{B}_N$ is then provided in Theorem 2, under slightly tighter assumptions on the matrices $\mathbf{R}_k$ and $\mathbf{T}_k$.

Theorem 1: Let K 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) $\mathbf{X}_k = \left(\frac{1}{\sqrt{n_k}} X_{ij}^k \right) \in \mathbb{C}^{N \times n_k}$ with $X_{ij}^k \in \mathbb{C}$ i.i.d. for all N, k, i, j, and $\mathrm{E}|X_{11}^k - \mathrm{E}X_{11}^k|^2 = 1$,
2) $\mathbf{R}_k^{\frac{1}{2}} \in \mathbb{C}^{N \times N}$ is the Hermitian nonnegative definite square root of the nonnegative definite Hermitian matrix $\mathbf{R}_k$,
3) $\mathbf{T}_k = \mathrm{diag}(\tau_1, \ldots, \tau_{n_k})$ with $\tau_i \geq 0$ for all i,
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 $M > 0$ such that $F^{\mathbf{T}_k}(M) > 1 - \varepsilon$ and $F^{\mathbf{R}_k}(M) > 1 - \varepsilon$ for all n_k, N,
5) $\mathbf{S} \in \mathbb{C}^{N \times N}$ is Hermitian nonnegative definite,
6) there exist $b > a > 0$ 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 $z \in \mathbb{C} \setminus \mathbb{R}^+$, $m_N(z) = \int \frac{1}{\lambda - z} dF^{\mathbf{B}_N}(\lambda)$, the Stieltjes transform of $\mathbf{B}_N$. Then, as all n_k and N grow large, with ratio c_k,

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

where

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

and the set of functions $\{e_i(z)\}$, $i \in \{1, \ldots, K\}$, forms 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 $\mathrm{sgn}(\Im[e_i(z)]) = \mathrm{sgn}(\Im[z])$ when $\Im[z] \neq 0$ and such that $e_i(z) > 0$ when z is real and negative.

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

$$\mathbb{C} \setminus (\{z : |z| \leq \varepsilon\} \cup \{z = x + iv : x > 0, |v| \leq \varepsilon\}) \,. \tag{13}$$

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

$$F^{\mathbf{B}_N} - F_N^\circ \Rightarrow 0 \tag{14}$$

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

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

Remark 1: In her PhD dissertation [42], Zhang derives an expression of the limiting eigenvalue distribution for the simpler case where $K = 1$ and $\mathbf{S} = 0$ but $\mathbf{T}_1$ is not constrained to be diagonal. Her work also uses a method based on the Stieltjes transform. Based on [42], it seems to the authors that Theorem 1 could well be extended to non-diagonal $\mathbf{T}_k$. However, proving so requires involved calculus, which we did not perform here. Similar conclusions can be drawn from the work of Rashidi et al. [28], based on operator-valued free probabilistic tools, which is a simpler method but which requires that the eigenvalue distributions of $\mathbf{T}_k$, $\mathbf{R}_K$ and $\mathbf{X}_k$ have finite support. The latter is too strong an assumption for our present application purposes. Also, in [6], using the same techniques as in the proof provided in Appendix A, Silverstein et al. do not assume that the matrices $\mathbf{T}_k$ are nonnegative definite. Our result could be extended to this less stringent requirement on the central $\mathbf{T}_k$ matrices, although in this case Theorem 1 does not hold for z real negative. For application purposes, it is fundamental here that the Stieltjes transform of $\mathbf{B}_N$ exist for $z \in \mathbb{R}^-$, for which it is sufficient that $\mathbf{T}_K \geq 0$ for all k.

We now claim that, under proper initialization, for $z \in \mathbb{C} \setminus \mathbb{R}^+$, a classical fixed-point algorithm converges surely to the solution of (12). This result is largely inspired by the original work of Dupuy et al. [41], used in the context of frequency selective channel models, and unfolds as follows

Proposition 1: For $z \in \mathbb{C} \setminus \mathbb{R}^+$, the fixed-point algorithm described in Table I converges surely to the unique solution $\{e_1(z), \ldots, e_K(z)\}$ of (12), such that $\mathrm{sgn}(\Im[e_i(z)]) = \mathrm{sgn}(\Im[z])$ when $\Im[z] \neq 0$ and such that $e_i(z) > 0$ when $z < 0$, for all i.

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

We shall see in Section IV-A that every point in the rate regions of MAC and BC can be described in terms of the solutions of (12), for $z = -\sigma^2 < 0$, where σ^2 is the additive Gaussian noise variance in the channel model. As such, Proposition 1 is of interest to the practical evaluation of all points in the rate regions. Note that alternative techniques are often used that produce faster convergence than the fixed-point algorithm described in Table I, such as the algorithm known as Newton's method.

Looser hypotheses will be used in the applications of Theorem 1 provided in Section IV. We will specifically need the corollary hereafter, Define $\varepsilon > 0$, the convergence threshold and $n \geq 0$, the iteration step. For all $k \in \{1, \dots, K\}$, set $e_j^0 = 1/z$ and $e_j^{-1} = \infty$.

while $\max_j\{|e_j^n - e_j^{n-1}|\} > \varepsilon$ **do**

  for $k \in \{1, \dots, K\}$ **do**

    Compute

$$e_k^{n+1} = \frac{1}{N} \operatorname{tr} \mathbf{R}_k \left(\mathbf{S} + \sum_{j=1}^{K} \int \frac{\tau_j dF^{\mathbf{T}_j}(\tau_j)}{1 + c_j \tau_j e_j^n} \mathbf{R}_j - z\mathbf{I}_N \right)^{-1} \tag{15}$$

  end for

  assign $n \leftarrow n + 1$

end while

TABLE I

FIXED-POINT ALGORITHM CONVERGING TO THE SOLUTION OF (12)

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

$$\mathbf{B}_N = \sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k \mathbf{X}_k^{\mathsf{H}} \mathbf{R}_k^{\frac{1}{2}} \tag{16}$$

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

1) $\mathbf{X}_k = \left(\frac{1}{\sqrt{n_k}} X_{ij}^k \right) \in \mathbb{C}^{N \times n_k}$, where the X_{ij}^k are i.i.d. Gaussian with zero mean and unit variance, for each i, j, N, k.
2) $\mathbf{R}_k^{\frac{1}{2}} \in \mathbb{C}^{N \times N}$ is the Hermitian nonnegative definite square root of the nonnegative definite Hermitian matrix $\mathbf{R}_k$,
3) $\mathbf{T}_k \in \mathbb{C}^{n_k \times n_k}$ is a nonnegative definite Hermitian matrix,
4) $\{F^{\mathbf{T}_k}\}_{n_k \geq 1}$ and $\{F^{\mathbf{R}_k}\}_{N \geq 1}$ form tight sequences,
5) there exist $b > a > 0$ for which

$$a \leq \liminf_N c_k \leq \limsup_N c_N \leq b \tag{17}$$

with $c_k = N/n_k$.

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

$$m_N(-x) - m_N^\circ(-x) \xrightarrow{\text{a.s.}} 0 \tag{18}$$

where

$$m_N^\circ(-x) = \frac{1}{N} \operatorname{tr} \left(\sum_{k=1}^{K} \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k \tau_k e_k(-x)} \mathbf{R}_k + x\mathbf{I}_N \right)^{-1} \tag{19}$$

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

$$e_i(-x) = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(\sum_{k=1}^{K} \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k \tau_k e_k(-x)} \mathbf{R}_k + x\mathbf{I}_N \right)^{-1} \tag{20}$$

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

Proof: Since the $\mathbf{X}_k$ are Gaussian, the joint distribution of the entries of $\mathbf{X}_k\mathbf{U}$ coincides with that of $\mathbf{X}_k$, for $\mathbf{U}$ any $n_k \times n_k$ unitary matrix. Therefore, $\mathbf{X}_k\mathbf{T}_k\mathbf{X}_k^{\mathsf{H}}$ in Theorem 1 can be substituted by $\mathbf{X}_k(\mathbf{U}\mathbf{T}_k\mathbf{U}^{\mathsf{H}})\mathbf{X}_k^{\mathsf{H}}$ without compromising the final result. As a consequence, the $\mathbf{T}_k$ can be taken non diagonal nonnegative definite Hermitian and the result of Theorem 1 holds. ■

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

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

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

$$\max_{1\leq k\leq K} \max(\lambda_{r_N+1}^{\mathbf{T}_k}, \lambda_{r_N+1}^{\mathbf{R}_k}) \leq \alpha \tag{21}$$

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 $\mathbf{T}_k$ and $\mathbf{R}_k$, $k \in \{1, \ldots, K\}$, and β some real constant such that $\beta > K(b/a)(1+\sqrt{a})^2$, $a_N = b_N^2\beta$ satisfies

$$r_N \log(1 + a_N/x) = o(N). \tag{22}$$

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

$$\mathcal{V}_N(x) - \mathcal{V}_N^{\circ}(x) \xrightarrow{\text{a.s.}} 0 \tag{23}$$

where

$$\begin{aligned}\mathcal{V}_N^{\circ}(x) =& \frac{1}{N}\log\det\left(\mathbf{I}_N + \frac{1}{x}\sum_{k=1}^{K}\mathbf{R}_k \int \frac{\tau_k}{1 + c_k e_k(-x)\tau_k} dF^{\mathbf{T}_k}(\tau_k)\right) \\ &+ \sum_{k=1}^{K} \frac{1}{c_k}\int \log\left(1 + c_k e_k(-x)\tau_k\right) dF^{\mathbf{T}_k}(\tau_k) \\ &+ x \cdot m_N^{\circ}(-x) - 1.\end{aligned} \tag{24}$$

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

Note that this last result is consistent both with [4] when the transmission channels are i.i.d. Gaussian and with [9] when $K = 2$. This result is also similar in nature to the expressions obtained in [25] for the multi-antenna Rician channel model and with [40] in the case of frequency selective channels. We point out that the expressions obtained in [40], [41] and [26], when the entries of the $\mathbf{X}_k$ matrices are Gaussian distributed, suggest a faster convergence rate of the deterministic equivalent of the Stieltjes and Shannon transforms than the one obtained here. Indeed, while we show here a convergence of order $o(1)$ (which is in fact refined to $o(\log^k N)$ for any k in Appendix A), in these works, the convergence is proven to be of order $O(1/N^2)$.

Contrary to these contributions though, we allow the $\mathbf{R}_k$ and $\mathbf{T}_k$ matrices to be more general than uniformly bounded in spectral norm. First, Theorem 1 and Corollary 1 require $\{F^{\mathbf{R}_k}\}$ and $\{F^{\mathbf{T}_k}\}$ to form tight sequences.

Remark now that, because of the trace constraint $\frac{1}{N}\operatorname{tr}\mathbf{R}_k = 1$, all sequences $\{F^{\mathbf{R}_k}\}$ are necessarily tight. Indeed, given $\varepsilon > 0$, take $M = 2/\varepsilon$; $N[1 - F^{\mathbf{R}_k}(M)]$ is the number of eigenvalues in $\mathbf{R}_k$ larger than $2/\varepsilon$, which is necessarily less than or equal to $N\varepsilon/2$ from the trace constraint, leading to $1 - F^{\mathbf{R}_k}(M) \leq \varepsilon/2$ and then $F^{\mathbf{R}_k}(M) \geq 1 - \varepsilon/2 > 1 - \varepsilon$. The same naturally holds for the $\mathbf{T}_k$ matrices. Observe now that Condition 2 in Theorem 2 requires a stronger assumption on the correlation matrices. Under the trace constraint, this requires that there exists $\alpha > 0$, such that the number of eigenvalues in $\mathbf{R}_k$ greater than α is of order $o(N/\log N)$. This may not always be the case, as we presently show with a counter-example. Consider the sequences of matrices $\mathbf{R}_k \in \mathbb{C}^{N\times N}$, N a power of 2, whose eigenvalue distribution is a mass in 0 of density $1 - 1/\log_2 N$ and a mass in $\log_2 N$ of density $1/\log_2 N$. Clearly this distribution satisfies the trace constraint and is unbounded, so that for all $\alpha > 0$, one can take N_0 a power of 2 such that $\log_2 N_0 > \alpha$; for $N > N_0$, $r_N = N/\log_2 N$ and then $r_N \log(1 + a_N/x)/N$ is away from 0 for all N large. This proves that the trace constraint is not enough to satisfies Condition 2 of Theorem 2. However, physically meaningful correlation matrices do not present this type of exceptional behaviour. Instead, low correlation tends to balance all eigenvalues around 1, in which case correlation matrices are uniformly bounded, while high correlation tends to bring a very few eigenvalues (much less than $N/\log_2 N$) to be large, the others being very small, in which case Condition 2 is satisfied. From now on, we claim that the conditions of Theorem 1 and Theorem 2 are satisfied for all *physically meaningful* correlation matrices.

IV. Applications

In this section, we provide two applications of Theorems 1 and 2 to the field of wireless communications. First, in Section IV-A, we derive an approximation of every point in the rate region of block-fading correlated multi-antenna MAC and BC, which is (almost surely) asymptotically accurate for all sequences of deterministic precoders, and an approximation of the boundary of the ergodic rate region of multiple access channels, which is (surely) asymptotically accurate. We then introduce an iterative power allocation algorithm to maximize the deterministic equivalent of the ergodic MAC rate region. Then, in Section IV-B, we provide an approximation of the capacity of the single user decoding and MMSE decoding in wireless MIMO networks with inter-cell interference. The latter are almost surely asymptotically accurate in the block-fading sense and surely asymptotically accurate in the ergodic sense.

Since in this section we study both uplink transmissions from mobile terminals to base stations and downlink transmissions from base stations to terminals, for notational consistency, $\mathbf{T}_k$ matrices will be used to model channel correlations at the base stations (be they transmitters or receivers) and $\mathbf{R}_k$ matrices will be used to model channel correlations at the mobile terminals (be they transmitters or receivers).

A. Rate Region of Multiple Access and 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.

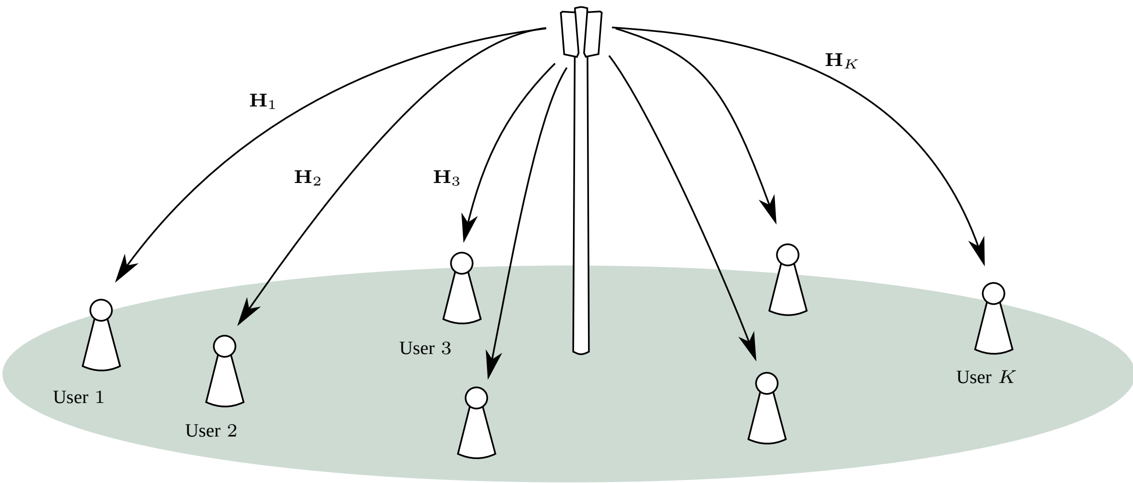


Fig. 1. Downlink scenario in multi-user broadcast channel

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 $\frac{1}{N} \operatorname{tr} \mathbf{P} \leq P$, $P > 0$, $\mathbf{y}_k \in \mathbb{C}^{n_k}$ 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.[1] The fading MIMO narrowband channel between the base station and user k is denoted $\mathbf{H}_k \in \mathbb{C}^{n_k \times N}$. Moreover, we assume that $\mathbf{H}_k$ follows the Kronecker model,

$$\mathbf{H}_k = \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k^{\frac{1}{2}} \tag{25}$$

with $\mathbf{R}_k \in \mathbb{C}^{n_k \times n_k}$ the (Hermitian) correlation matrix at terminal 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 with respect to $\mathbf{H}_k$ and $\mathbf{X}_k \in \mathbb{C}^{n_k \times N}$ a random matrix with Gaussian independent entries of variance $1/n_k$. In this model, $\mathbf{T}_k$ and $\mathbf{R}_k$ satisfy $\operatorname{tr} \mathbf{T}_k = N$ and $\operatorname{tr} \mathbf{R}_k = n_k$. We additionally constrain the eigenvalues of the matrices $\mathbf{R}_k$ and $\mathbf{T}_k$, $k \in \{1, \ldots, K\}$, to satisfy the mild Condition 2 of Theorem 2. From our previous remark, we mostly allow all but physically meaningless models of covariance matrices.

Under the above assumptions, the downlink communication model reads

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{s} + \mathbf{n}_k. \tag{26}$$

Denoting equivalently $\mathbf{s}_k$ the signal transmitted in the dual uplink (MAC) by user k, such that $\mathrm{E}[\mathbf{s}_k \mathbf{s}_k^\mathsf{H}] = \mathbf{P}_k$, $\frac{1}{n_k} \operatorname{tr} \mathbf{P}_k \leq P_k$, $\mathbf{y}$ and $\mathbf{n}$ the signal and the noise received by the base station, respectively, we have the dual uplink model

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{H}_k^\mathsf{H} \mathbf{s}_k + \mathbf{n}. \tag{27}$$

In the following, we will successively study the MAC and BC rate regions for block-fading channels by means of the MAC-BC duality [19]. We shall then consider the ergodic rate region for time varying MAC.

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

2) MAC and BC Rate Regions in Block Fading Channels: We start by assuming that the channels $\mathbf{H}_1, \ldots, \mathbf{H}_K$ are constant over the observation period. The per-receive antenna rate region $\mathcal{C}_{\mathrm{MAC}}(P_1, \ldots, P_K; \mathbf{H}^{\mathsf{H}})$, under respective transmit power constraints $P_1, \ldots, P_K$ for users 1 to K and channel $\mathbf{H}^{\mathsf{H}} = [\mathbf{H}_1^{\mathsf{H}} \ldots \mathbf{H}_K^{\mathsf{H}}]$, reads [21]

$$\begin{aligned}&\mathcal{C}_{\mathrm{MAC}}(P_1, \ldots, P_K; \mathbf{H}^{\mathsf{H}}) \\ &= \bigcup_{\substack{\frac{1}{1} n_i \operatorname{tr}(\mathbf{P}_i) \leq P_i \\ \mathbf{P}_i \geq 0 \\ i=1,\ldots,K}} \left\{ \{R_i, 1 \leq i \leq K\} : \sum_{i \in \mathcal{S}} R_i \leq \frac{1}{N} \log \left| \mathbf{I} + \frac{1}{\sigma^2} \sum_{i \in \mathcal{S}} \mathbf{H}_i^{\mathsf{H}} \mathbf{P}_i \mathbf{H}_i \right|, \forall \mathcal{S} \subset \{1, \ldots, K\} \right\}, \qquad (28)\end{aligned}$$

where $\mathbf{P}_i \geq 0$ stands for "$\mathbf{P}_i$ is nonnegative definite".

Since the entries of $\mathbf{X}_k$ have variance $1/n_k$, the power constraints on $\mathbf{P}_1, \ldots, \mathbf{P}_K$ necessarily scale with n_k. We might have alternatively assumed, as is often the case, that the $\mathbf{X}_k$ have entries of unit variance and therefore that the power constraints are independent of the channel dimensions. Before applying Corollary 1 and Theorem 2, we need to verify that $\{F^{\mathbf{R}_k \mathbf{P}_k}\}$ for growing n_k is necessarily tight and that Condition 2 of Theorem 2 is satisfied. From the argument given in Section III, both $\{F^{\mathbf{R}_k}\}$ and $\{F^{\mathbf{P}_k}\}$ are tight sequences. For $\varepsilon > 0$ such that $n_k \varepsilon \in \mathbb{N}$, we can therefore choose M such that $1 - F^{\mathbf{R}_k}(\sqrt{M}) < \varepsilon/2$ and $1 - F^{\mathbf{P}_k}(\sqrt{M}) < \varepsilon/2$ for all n_k; from Lemma 15, since the smallest $n_k \varepsilon/2 + 1$ eigenvalues of both $\mathbf{R}_k$ and $\mathbf{P}_k$ are less than $\sqrt{M}$, at least the smallest $n_k \varepsilon + 1$ eigenvalues of $\mathbf{R}_k \mathbf{P}_k$ are less than M, hence $1 - F^{\mathbf{R}_k \mathbf{P}_k}(M) < \varepsilon$ and $\{F^{\mathbf{R}_k \mathbf{P}_k}\}$ is tight. Once again, Condition 2 of Theorem 2 can be satisfied for all but meaningless $\mathbf{R}_k \mathbf{P}_k$ matrices, and we claim the latter of no relevance to the current investigation.

We can now apply Theorem 2, which presently states that, for any set $\mathcal{S} \subset \{1, \ldots, K\}$, we have for N, n_k large, almost surely

$$\begin{aligned}\frac{1}{N} \log \left| \mathbf{I}_N + \frac{1}{\sigma^2} \sum_{i \in \mathcal{S}} \mathbf{H}_i^{\mathsf{H}} \mathbf{P}_i \mathbf{H}_i \right| &= \frac{1}{N} \log \det \left(\mathbf{I}_N + \frac{1}{\sigma^2} \sum_{k \in \mathcal{S}} \mathbf{T}_k \int \frac{r_k}{1 + c_k e_k(-\sigma^2) r_k} dF^{\mathbf{R}_k \mathbf{P}_k}(r_k) \right) \\ &+ \sum_{k \in \mathcal{S}} \frac{1}{c_k} \int \log\left(1 + c_k e_k(-\sigma^2) r_k\right) dF^{\mathbf{R}_k \mathbf{P}_k}(r_k) \\ &+ \sigma^2 \cdot m_{\mathcal{S}}^{\circ}(-\sigma^2) - 1 + o(1), \qquad (29)\end{aligned}$$

where

$$m_{\mathcal{S}}^{\circ}(-\sigma^2) = \frac{1}{N} \operatorname{tr} \left(\sum_{k \in \mathcal{S}} \int \frac{r_k dF^{\mathbf{R}_k \mathbf{P}_k}(r_k)}{1 + c_k r_k e_k(-\sigma^2)} \mathbf{T}_k + \sigma^2 \mathbf{I}_N \right)^{-1} \qquad (30)$$

and $e_1(-\sigma^2), \ldots, e_K(-\sigma^2)$ are the unique positive solutions to

$$e_i(-\sigma^2) = \frac{1}{N} \operatorname{tr} \mathbf{T}_i \left(\sum_{k \in \mathcal{S}} \int \frac{r_k dF^{\mathbf{R}_k \mathbf{P}_k}(r_k)}{1 + c_k r_k e_k(-\sigma^2)} \mathbf{T}_k + \sigma^2 \mathbf{I}_N \right)^{-1}. \qquad (31)$$

From these equations, every point of the MAC rate region (28) can be deterministically approximated. Note that the convergence rates of (10) and (23) are dictated to some extent by the $F^{\mathbf{R}_k}$ and $F^{\mathbf{T}_k}$, so that the term $o(1)$ in (29) cannot necessarily be bounded for fixed N.

Now, we can similarly provide a deterministic equivalent to every point in the rate region of the block-fading broadcast channel. This rate region, name it $\mathcal{C}_{\mathrm{BC}}(P; \mathbf{H})$, has been recently shown [22] to be achieved by dirty paper coding (DPC). For a transmit power constraint P over the compound channel $\mathbf{H}$, it is shown by MAC-BC duality that [19]

$$\mathcal{C}_{\mathrm{BC}}(P;\mathbf{H}) = \bigcup_{\substack{P_1,\ldots,P_K \\ \sum_{k=1}^K P_k \leq P}} \mathcal{C}_{\mathrm{MAC}}(P_1,\ldots,P_K;\mathbf{H}^{\mathsf{H}}) \tag{32}$$

which is easily obtained from Equation (28).

To achieve the boundaries of the MAC and BC rate regions for the block-fading channel, the precoding matrices $\mathbf{P}_1,\ldots,\mathbf{P}_K$ need be tailored to the channels $\mathbf{H}_1,\ldots,\mathbf{H}_K$. To this day, no closed-form expression of these optimal precoders is available, although an iterative water-filling algorithm has been derived by Yu et al. in [43] to determine these precoding matrices. Theorem 2 cannot be used either, since the optimal precoders will be strongly dependent on the specific realization of the $\mathbf{H}_k$, and therefore dependent on the $\mathbf{X}_k$. If the $\mathbf{X}_k$ are not known to the transmitters though, the optimal precoding matrices obviously no longer depend on the $\mathbf{X}_k$, but certainly on the correlation matrices $\mathbf{R}_k$ and $\mathbf{T}_k$.

When the channel is varying too fast to allow reliable channel estimation, transmitters in a multiple access channel typically do not know the exact $\mathbf{H}_k$ matrices. On the contrary, the transmit and receive correlation matrices are in this case long-term channel variations that the transmitters can usually reliably estimate. We study this scenario in the next section.

3) MAC Rate Region in Fast Fading Channels: Suppose that the $\mathbf{H}_k$ channels are varying fast and that the transmitters in the MAC only have statistical channel state information, i.e. they only know their respective $\mathbf{T}_k$ and $\mathbf{R}_k$ matrices. In this case, the MAC rate region will be referred to as the *ergodic* rate region. The ergodic rate region $\mathcal{C}_{\mathrm{MAC}}^{(\mathrm{ergodic})}$ is in this case given by

$$\begin{aligned}&\mathcal{C}_{\mathrm{MAC}}^{(\mathrm{ergodic})}(P_1,\ldots,P_K;\mathbf{H}^{\mathsf{H}}) \\ &= \bigcup_{\substack{\frac{1}{n_i}\operatorname{tr}(\mathbf{P}_i)\leq P_i \\ \mathbf{P}_i\geq 0 \\ i=1,\ldots,K}} \left\{ \{R_i, 1\leq i\leq K\} : \sum_{i\in\mathcal{S}} R_i \leq \mathrm{E}\left[\frac{1}{N}\log\left|\mathbf{I}+\frac{1}{\sigma^2}\sum_{i\in\mathcal{S}}\mathbf{H}_i^{\mathsf{H}}\mathbf{P}_i\mathbf{H}_i\right|\right], \forall\mathcal{S}\subset\{1,\ldots,K\}\right\},\end{aligned} \tag{33}$$

where the expectation is taken over the joint random variable $(\mathbf{X}_1,\ldots,\mathbf{X}_K)$.

Now, Theorem 2 states that $\mathcal{V}_N(x) - \mathcal{V}_N^{\circ}(x) \to 0$, as N grows large, on a subset of measure 1 of the probability space Ω that engenders $(\mathbf{X}_1,\ldots,\mathbf{X}_K)$. Integrating this expression over Ω therefore leads to $\mathrm{E}\mathcal{V}_N(x) - \mathcal{V}_N^{\circ}(x) \to 0$. We can therefore apply Theorem 2 to determine the ergodic rate region $\mathcal{C}_{\mathrm{MAC}}^{(\mathrm{ergodic})}$ of the fast fading MAC. For fixed $\mathbf{P}_1,\ldots,\mathbf{P}_K$, we therefore have here

$$\begin{aligned}\mathrm{E}\left[\frac{1}{N}\log\left|\mathbf{I}+\frac{1}{\sigma^2}\sum_{i\in\mathcal{S}}\mathbf{H}_i^{\mathsf{H}}\mathbf{P}_i\mathbf{H}_i\right|\right] &= \frac{1}{N}\log\det\left(\mathbf{I}_N+\frac{1}{\sigma^2}\sum_{k\in\mathcal{S}}\mathbf{T}_k\int\frac{r_k}{1+c_ke_k(-\sigma^2)r_k}dF^{\mathbf{R}_k\mathbf{P}_k}(r_k)\right) \\ &+ \sum_{k\in\mathcal{S}}\frac{1}{c_k}\int\log\left(1+c_ke_k(-\sigma^2)r_k\right)dF^{\mathbf{R}_k\mathbf{P}_k}(r_k) \\ &+ \sigma^2\cdot m_{\mathcal{S}}^{\circ}(-\sigma^2) - 1 + o(1),\end{aligned} \tag{34}$$

with $m_{\mathcal{S}}^{\circ}(-\sigma^2)$ given by (30), and where the convergence with growing N is sure.

Define $\eta > 0$ the convergence threshold and $l \geq 0$ the iteration step. At step $l = 0$, for all $k \in \mathcal{S}$, set $p_k^\circ = P_k$.
while $\max_k\{|p_k^l - p_k^{l-1}|\} > \eta$ **do**
For $k \in \mathcal{S}$, define e_k^{l+1} as the solution of (12) for $z = -\sigma^2$, obtained from the fixed-point algorithm of Table I.
for $k \in \mathcal{S}$ **do**
for $i = 1 \dots, n_k$ **do**
Set $p_{k,i}^{l+1} = \left(\mu_k - \frac{1}{c_k e_k^{l+1} r_{ki}}\right)^+$, with μ_k such that $\frac{1}{n_k} \operatorname{tr} \mathbf{P}_k = P_k$.
end for
end for
assign $l \leftarrow l + 1$
end while

TABLE II

ITERATIVE WATER-FILLING ALGORITHM

The transmission policy that achieves the boundary of the ergodic rate region requires here to determine the rate optimal precoding matrices $\mathbf{P}_1, \dots, \mathbf{P}_{|\mathcal{S}|}$, for all $\mathcal{S} \subset \{1, \dots, K\}$, which are dependent only on the $\mathbf{T}_k$ and $\mathbf{R}_k$ correlation matrices. To this end, we first need the following result,

Proposition 2: If at least one of the correlation matrices $\mathbf{R}_k$, $k \in \mathcal{S}$ is invertible, then the right-hand side of (29) is a strictly concave function of $\mathbf{P}_1, \dots, \mathbf{P}_{|\mathcal{S}|}$.

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

From Proposition 2, we immediately prove that the $|\mathcal{S}|$-ary set of matrices $(\mathbf{P}_1^\star, \dots, \mathbf{P}_{|\mathcal{S}|}^\star)$, which maximize the deterministic equivalent of the ergodic sum rate over the set $\mathcal{S}$, is unique, provided that one of the $\mathbf{R}_k$ is invertible. In a very similar way as in [26], we then show that the matrices $\mathbf{P}_k^\star$, $k \in \{1, \dots, |\mathcal{S}|\}$, have the following properties: (i) their eigenspace of $\mathbf{P}_k^\star$ is the same as that of $\mathbf{R}_k$, (ii) the eigenvalues of $\mathbf{P}_k^\star$ are the solutions of a classical water-filling problem.

Proposition 3: 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 $\mathbf{D}_k = \operatorname{diag}(r_{k1}, \dots, r_{kn_k})$ diagonal. Then the covariance matrices $\mathbf{P}_1^\star, \dots, \mathbf{P}_{|\mathcal{S}|}^\star$ which maximize the right-hand side of (34) 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 $e_k^\star = e_k(-\sigma^2)$ when $\mathbf{P}_k = \mathbf{P}_k^\star$, for all k, the i^{th} diagonal entry $p_{ki}^\star$ of $\mathbf{Q}_k^\star$ satisfies

$$p_{ki}^\star = \left(\mu_k - \frac{1}{c_k e_k^\star r_{ki}}\right)^+ \tag{35}$$

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

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

We then propose an iterative water-filling algorithm to obtain the power allocation policy which maximizes the right-hand side of (29). This is provided in Table II.

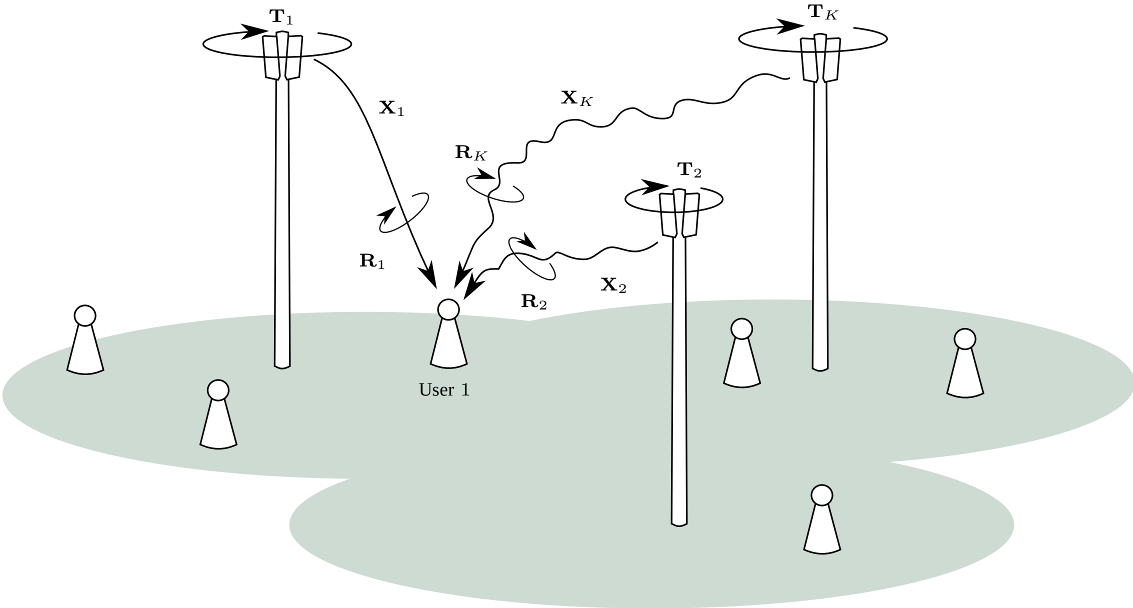


Fig. 2. Downlink multi-cell scenario

In [26], it is proven that the convergence of this algorithm implies its convergence towards the correct limit. The line of reasoning in [26] can be directly adapted to the current situation so that, if the iterative water-filling algorithm of Table II converges, then $\mathbf{P}_1, \ldots, \mathbf{P}_K$ converge to the capacity-achieving precoding matrices $\mathbf{P}_1^\star, \ldots, \mathbf{P}_K^\star$. However, similar to [26], it is difficult to prove the sure convergence of the water-filling algorithm. Nonetheless, extensive simulations suggested that convergence always happens.

B. Multi-User MIMO

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

- multi-cell uplink: consider a network of K cells. On a given time or frequency resource, the base station of the cell indexed by $i \in \{1, \ldots, K\}$ receives data from a unique terminal user of this cell 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 physical 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.

In the following, the downlink scheme is considered. Consider a wireless mobile network with $K \geq 1$ cells indexed from 1 to K, controlled by non-cooperative 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 antennas and hereafter referred to as *the user* or *the receiver*. Every base station $j \in \{1, \ldots, K\}$ is equipped with n_j antennas. Similar to previous sections, we denote $c_j = N/n_j$, although now this corresponds to the ratio of the number of antennas at the terminal user by the number of antennas at the base stations.

Denote $\mathbf{s}_j \in \mathbb{C}^{n_j}$ 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. We assume uniform power allocation across the base station antennas, so that $\mathbf{E}[\mathbf{s}_j \mathbf{s}_j^{\mathsf{H}}] = \mathbf{I}_{n_j}$. 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 a Kronecker channel, $\mathbf{H}_j = \mathbf{R}_j \mathbf{X}_j \mathbf{T}_j$, with $\mathbf{R}_j \in \mathbb{C}^{N \times N}$ and $\mathbf{T}_j \in \mathbb{C}^{n_j \times n_j}$. The communication model reads

$$\mathbf{y} = \mathbf{H}_1 \mathbf{s}_1 + \sum_{j=2}^{K} \mathbf{H}_j \mathbf{s}_j + \mathbf{n} \tag{36}$$

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: We assume block-fading channels and uniform power allocation across the base station antennas. If the receiver considers the signals from the $K-1$ interfering transmitters as Gaussian noise with a known variance pattern, 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

$$\mathcal{C}_{\mathrm{SU}}(\sigma^2) = \frac{1}{N} \log \left| \mathbf{I}_N + \frac{1}{\sigma^2} \sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} \right| - \frac{1}{N} \log \left| \mathbf{I}_N + \frac{1}{\sigma^2} \sum_{j=2}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} \right|. \tag{37}$$

Assume N and the n_i, $i \in \{1, \ldots, K\}$, are large. From Corollary 1, we define the functions $m^{i,\circ}(-\sigma^2)$ as the approximated Stieltjes transforms of $\sum_{j=i}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}}$, $i \in \{1, 2\}$, at point $-\sigma^2$,

$$m^{i,\circ}(-\sigma^2) = \frac{1}{N} \operatorname{tr} \left(\sum_{k=i}^{K} \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k \tau_k e_k^i(-\sigma^2)} \mathbf{R}_k + \sigma^2 \mathbf{I}_N \right)^{-1} \tag{38}$$

where the set of $e_j^i(-\sigma^2)$, $i \in \{1, 2\}$, $j \in \{1, \ldots, K\}$, forms the unique solution with positive entries of

$$e_j^i(-\sigma^2) = \frac{1}{N} \operatorname{tr} \mathbf{R}_j \left(\sum_{k=i}^{K} \int \frac{\tau_k dF^{\mathbf{T}_k}(\tau_k)}{1 + c_k \tau_k e_k^i(-\sigma^2)} \mathbf{R}_k + \sigma^2 \mathbf{I}_N \right)^{-1}. \tag{39}$$

From Theorem 2, we then have

$$\begin{aligned}
\mathcal{C}_{\mathrm{SU}}(\sigma^2) =& \frac{1}{N} \log \det \left(\mathbf{I}_N + \frac{1}{\sigma^2} \sum_{k=1}^{K} \mathbf{R}_k \int \frac{\tau_k}{1 + c_k e_k^1(-\sigma^2) \tau_k} dF^{\mathbf{T}_k}(\tau_k) \right) \\
& - \frac{1}{N} \log \det \left(\mathbf{I}_N + \frac{1}{\sigma^2} \sum_{k=2}^{K} \mathbf{R}_k \int \frac{\tau_k}{1 + c_k e_k^2(-\sigma^2) \tau_k} dF^{\mathbf{T}_k}(\tau_k) \right) \\
& + \sum_{k=1}^{K} \frac{1}{c_k} \int \log\left(1 + c_k e_k^1(-\sigma^2) \tau_k\right) dF^{\mathbf{T}_k}(\tau_k) \\
& - \sum_{k=2}^{K} \frac{1}{c_k} \int \log\left(1 + c_k e_k^2(-\sigma^2) \tau_k\right) dF^{\mathbf{T}_k}(\tau_k) \\
& + \sigma^2 \cdot [m^{1,\circ}(-\sigma^2) - m^{2,\circ}(-\sigma^2)] + o(1)
\end{aligned} \tag{40}$$

almost surely.

Similar to the previous section, the result naturally holds also in the ergodic sense, with almost sure convergence replaced here by sure convergence.

3) MMSE Decoding: Achieving rates close to $\mathcal{C}_{\mathrm{SU}}$ in practice requires to perform multi-stream decoding at once at the receiver. A suboptimal linear technique, the MMSE decoder, is often used instead as it allows to treat transmit data streams independently, while maximizing the SINR for each data stream at the receiver. In this section, we study the performances of the MMSE decoder for correlated transmission channels. We assume block-fading channels $\mathbf{H}_1, \ldots, \mathbf{H}_K$, which are supposed to be perfectly known at the receiver.

The communication model in this case reads

$$\mathbf{y} = \mathbf{H}_1^{\mathsf{H}} \left(\sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} + \sigma^2 \mathbf{I}_N \right)^{-1} \left(\sum_{j=1}^{K} \mathbf{H}_j \mathbf{s}_j + \mathbf{n} \right) \tag{41}$$

where $\mathbf{H}_1^{\mathsf{H}} \left(\sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} + \sigma^2 \mathbf{I}_N \right)^{-1}$ is the MMSE linear filter at the receiver.

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

$$\mathcal{C}_{\mathrm{MMSE}}(\sigma^2) = \frac{1}{N} \sum_{i=1}^{n_1} \log(1 + \gamma_i) \tag{42}$$

where, denoting $\mathbf{h}_j \in \mathbb{C}^{n_j}$ the j^{th} column of $\mathbf{H}_1$, $\mathbf{x}_j$ the j^{th} column of $\mathbf{X}_1$ and $t_1, \ldots, t_{n_1}$ the eigenvalues of $\mathbf{T}_1$, we have $\sqrt{t_j} \mathbf{R}_1^{\frac{1}{2}} \mathbf{x}_j = \mathbf{h}_j$ and the signal-to-interference plus noise ratio γ_i expresses as

$$\gamma_i = \frac{\mathbf{h}_i^{\mathsf{H}} \left(\sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{h}_i}{1 - \mathbf{h}_i^{\mathsf{H}} \left(\sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{h}_i} \tag{43}$$

$$= \mathbf{h}_i^{\mathsf{H}} \left(\sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} - \mathbf{h}_i \mathbf{h}_i^{\mathsf{H}} + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{h}_i \tag{44}$$

$$= t_i \mathbf{x}_i^{\mathsf{H}} \mathbf{R}_1^{\frac{1}{2}} \left(\sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} - \mathbf{h}_i \mathbf{h}_i^{\mathsf{H}} + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{R}_1^{\frac{1}{2}} \mathbf{x}_i, \tag{45}$$

where Equation (44) is a direct application of Lemma 4. The vectors $\mathbf{x}_i$ have i.i.d. complex Gaussian entries with variance $1/n_1$ and the inner matrix of the right-hand side of (45) 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, we have

$$\gamma_i = \frac{t_i}{n_1} \operatorname{tr} \mathbf{R}_1 \left(\sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} - \mathbf{h}_i \mathbf{h}_i^{\mathsf{H}} + \sigma^2 \mathbf{I}_N \right)^{-1} + o(1). \tag{46}$$

almost surely.

From Lemma 5, the rank 1 perturbation $(-\mathbf{h}_i \mathbf{h}_i^{\mathsf{H}})$ does not affect asymptotically the trace in (46). Therefore, in the large N limit, we have

$$\gamma_i = \frac{t_i}{n_1} \operatorname{tr} \mathbf{R}_1 \left(\sum_{j=1}^{K} \mathbf{H}_j \mathbf{H}_j^{\mathsf{H}} + \sigma^2 \mathbf{I}_N \right)^{-1} + o(1) \tag{47}$$

Correlation matrix	Eigenvalues							
$\mathbf{R}_1$	0.00	0.00	0.00	0.00	0.00	0.01	0.60	7.39
$\mathbf{R}_2$	0.00	0.00	0.00	0.00	0.09	0.97	3.03	3.91
$\mathbf{T}_1$	0.80	0.82	0.92	1.04	1.07	1.09	1.11	1.14
$\mathbf{T}_2$	0.48	0.52	0.56	0.63	0.79	1.18	1.47	2.37

TABLE III

EIGENVALUES OF CORRELATION MATRICES FOR $N = n_1 = n_2 = 8$, $d^{\mathrm{R}} = 0.5\lambda$ AND $d^{\mathrm{T}} = 10\lambda$.

almost surely.

In the appendix, Equation (91), we prove that the trace in (47) converges almost surely to $e_1^1(z)$, defined in (39). We therefore finally have the compact expression for $\mathcal{C}_{\mathrm{MMSE}}$,

$$\mathcal{C}_{\mathrm{MMSE}}(\sigma^2) = \frac{1}{N}\sum_{i=1}^{n_1} \log\left(1 + t_i c_1 e_1^1(-\sigma^2)\right) + o(1) \tag{48}$$

with $c_1 = N/n_1$ and where the convergence is with probability one.

Taking the expectation over all $\mathbf{X}_k$ matrices on the left-hand side of (48), we have that the same result holds for the ergodic MMSE decoding capacity.

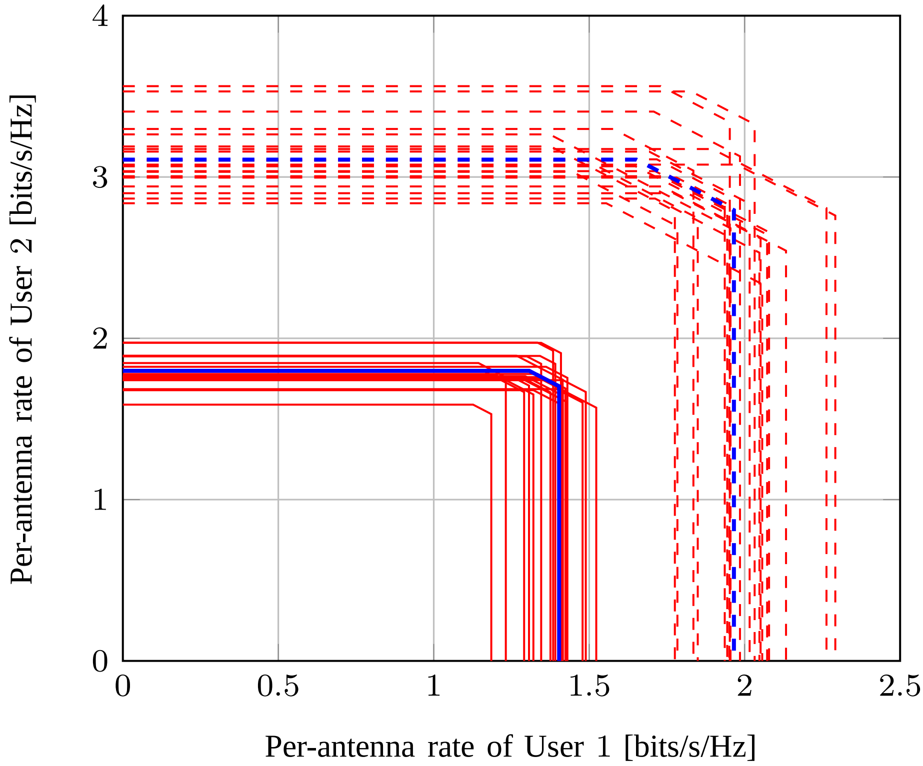
V. Simulations and Results

In the following, we apply the results obtained in Sections IV-A and IV-B to determine the rate region of block-fading and time varying multiple acess channels, as well as the capacity of multi-user MIMO with inter-cell interference. This section is moreover dedicated to the analysis of the antenna-efficiency of the aforementioned communication schemes.

A. Block Fading MAC and BC Rate Regions

First, we provide simulation results in the context of a two-user multi-access channel, with N antennas at the base station and $n_1 = n_2$ antennas at the user terminals. The antennas are placed in linear arrays. We further assume that both user terminals are physically identical. To model the transmit and receive correlation matrices, we consider both the effects of the distance d^{R} (resp. d^{T}) between adjacent antennas at the user terminals (resp. at the base station) and of the solid angles of effective energy transmission and reception. We assume a channel model where signals are transmitted and received isotropically in the vertical direction, but transmitted and received under a small angle $\pi/6$ in the horizontal direction. This simulates the situation where a strong propagation path exists in a given direction, while the other paths are strongly attenuated. We then model the entries of the correlation matrices from a natural extension of Jakes model with privileged direction of signal departure and arrival. Denoting λ the transmit signal wavelength, the entry (a, b) of, say matrix $\mathbf{T}_1$, is

$$T_{1_{ab}} = \int_{\theta_{\mathrm{min}}^{(\mathbf{T}_1)}}^{\theta_{\mathrm{max},}^{(\mathbf{T}_1)}} \exp\left(2\pi i |a-b| \frac{d^{\mathrm{T}}}{\lambda} \cos(\theta)\right) d\theta \tag{49}$$



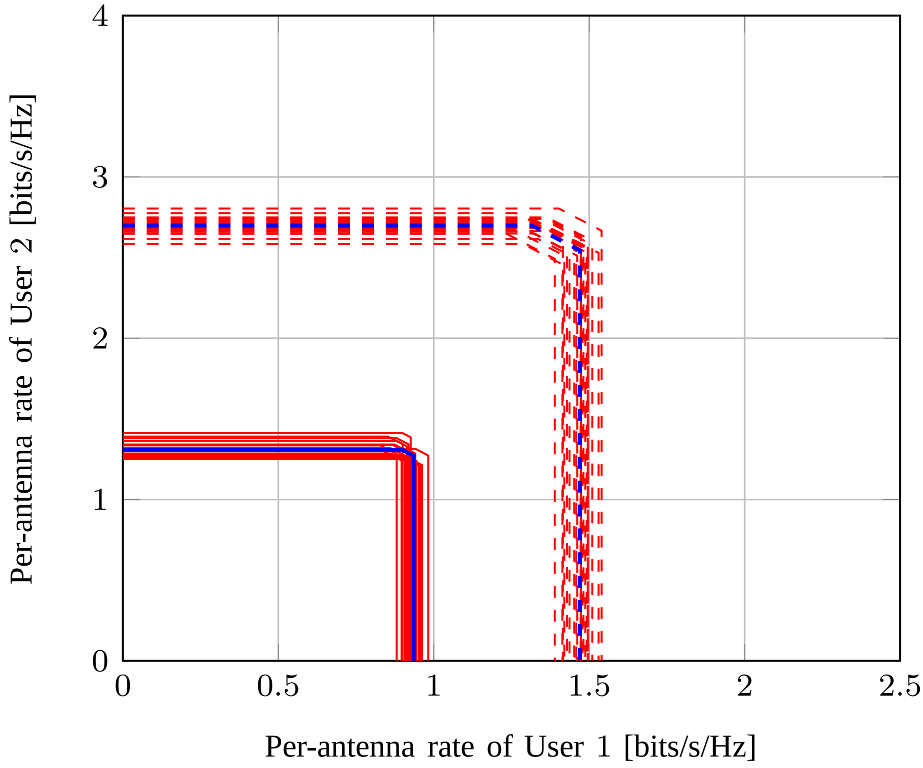


Fig. 3. Rate region of two-user MAC, equal power allocation, for $N = 8$ (top), $N = 16$ (bottom), $n_1 = n_2 = N$, ULA model, antenna spacing $\frac{d^{\mathrm{R}}}{\lambda} = 0.5$ (dashed) and $\frac{d^{\mathrm{R}}}{\lambda} = 0.1$ (solid), $\mathrm{SNR} = 20$ dB. Deterministic equivalents are given in thick lines.

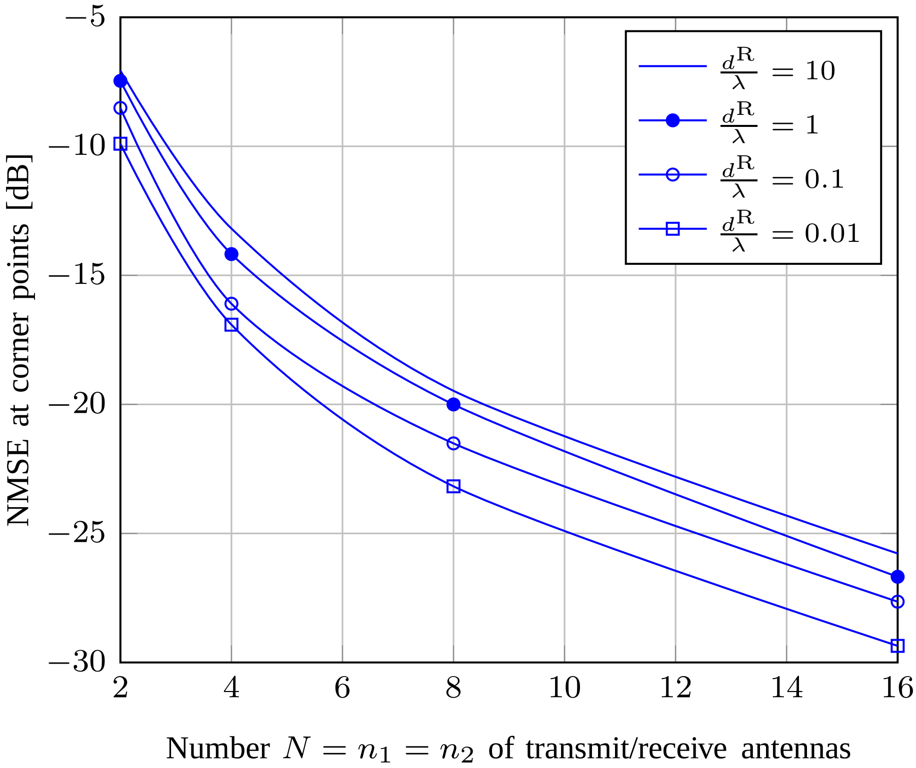


Fig. 4. Normalized mean square error of the corner points in the two-user MAC, equal power allocation, for $N = n_1 = n_2$ varying from 2 to 16, ULA model, antenna spacing $\frac{d^{\rm R}}{\lambda}$ varying from 0.01 to 10.

with $[\theta_{\rm min}^{(\mathbf{T}_1)}, \theta_{\rm max,}^{(\mathbf{T}_1)}]$ the effective horizontal directions of signal propagation. In our case, we consider $\theta_{\rm min}^{(\mathbf{T}_1)} = 2\pi/3$, $\theta_{\rm max}^{(\mathbf{T}_1)} = 5\pi/6$, $\theta_{\rm min}^{(\mathbf{T}_2)} = \pi$, $\theta_{\rm max}^{(\mathbf{T}_2)} = 7\pi/6$, $\theta_{\rm min}^{(\mathbf{R}_1)} = 0$, $\theta_{\rm max,}^{(\mathbf{R}_1)} = \pi/6$, $\theta_{\rm min}^{(\mathbf{R}_2)} = \pi/3$ and $\theta_{\rm max}^{(\mathbf{R}_2)} = 2\pi/3$, where $\theta_{\rm min}^{(\mathbf{Y})}$ and $\theta_{\rm max}^{(\mathbf{Y})}$ are the minimum and maximum angles of transmit or receive energy for the correlation matrix $\mathbf{Y}$.

We first assume the multi-access block-fading channel with the two users described above. We consider $N = n_1 = n_2$, SNR $= 20$ dB, $d^{\rm T} = 10\lambda$ and uniform power allocation. In Figure 3, we compare simulation results obtained from $1,000$ Monte Carlo simulations to the deterministic equivalent obtained in (29), when $N = 8$ (top) and $N = 16$ (bottom), and for $d^{\rm R} = 0.5\lambda$ or $d^{\rm R} = 0.1\lambda$. We first observe that the empirical rate regions show a large variance for $N = 8$ compared to $N = 16$. Nonetheless, the deterministic equivalent, even for $N = 8$, is an accurate estimate of any of the empirical pentagons. In terms of antenna efficiency, observe in this specific scenario that doubling the number of antennas at both communication sides reduces the achievable transmission rate of user 2 by 25% when $d^{\rm R} = 0.5\lambda$ (leading therefore to a 150% total throughput gain), and by 33% when $d^{\rm R} = 0.1\lambda$ (inducing a 133% total throughput gain). For high correlation, doubling the number of antennas therefore results in small rate increase. As for the accuracy of the deterministic equivalent, observe that even for strong transmit correlation, the deterministic equivalent is very precise, as claimed in Theorems 1 and 2.

To confirm the accuracy of the deterministic equivalents, even for strong correlation patterns, we depict in Figure

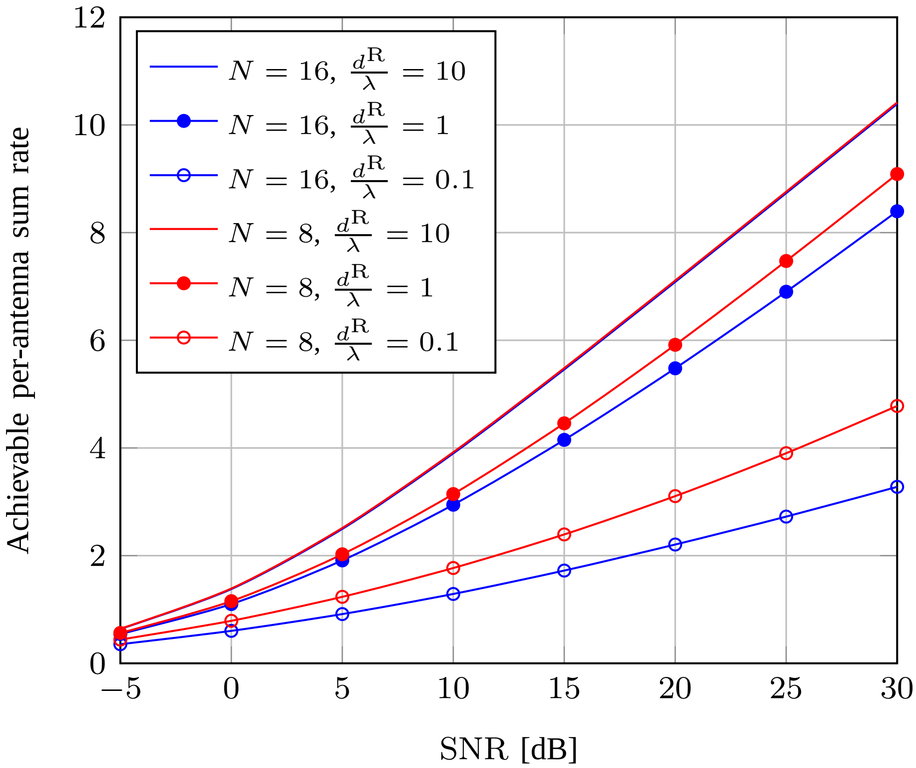


Fig. 5. Per-antenna achievable sum rate for SNR varying from -5 dB to 30 dB, for different values of $\frac{d^{\mathrm{R}}}{\lambda}$.

4 the normalized mean square error of the rate region corner points. That is, for $10,000$ Monte Carlo simulations, we take the average estimation error of the rates at the corner points provided by the deterministic equivalent, normalized by the empirical rates. Observe that, even for low d^{R}/λ ratios (corresponding to high correlations), the estimation error goes very fast to zero, as N increases. The fact that the normalized estimation error is even lower for higher correlation is only due to the intrinsic large variations of the empirical rate region observed in Figure 3 when d^{R}/λ is large. In Table III, the eigenvalues of the correlation matrices for $N = 8$, $d^{\mathrm{R}} = 0.5\lambda$ and $d^{\mathrm{T}} = 10\lambda$ are provided (they all sum up to 8). Those values confirm that it is possible to have some eigenvalues almost zero, while only a few eigenvalues are large, and still have consistent estimation of the per-antenna rate performance; this is in phase with the conditions of Theorem 1 and Theorem 2. Note, as a matter of fact, the importance of the angular direction of signal arrival or departure, which, for identical antenna spacings, can lead to very different correlation patterns.

It was observed in Figure 3 that doubling the number of transmit antennas seemed to be an interesting choice for low correlation as the antenna efficiency is not much impaired, while higher correlation seemed to reduce antenna efficiency as the number of antennas increases. This trend is verified in Figure 5, where the sum-rate of the MAC scenario under study is compared when $N = 8$, $N = 16$, for varying SNR and varying ratios d^{R}/λ. For very low correlated antennas, i.e. in the case of nearly i.i.d. channel entries, there is no loss in antenna efficiency by

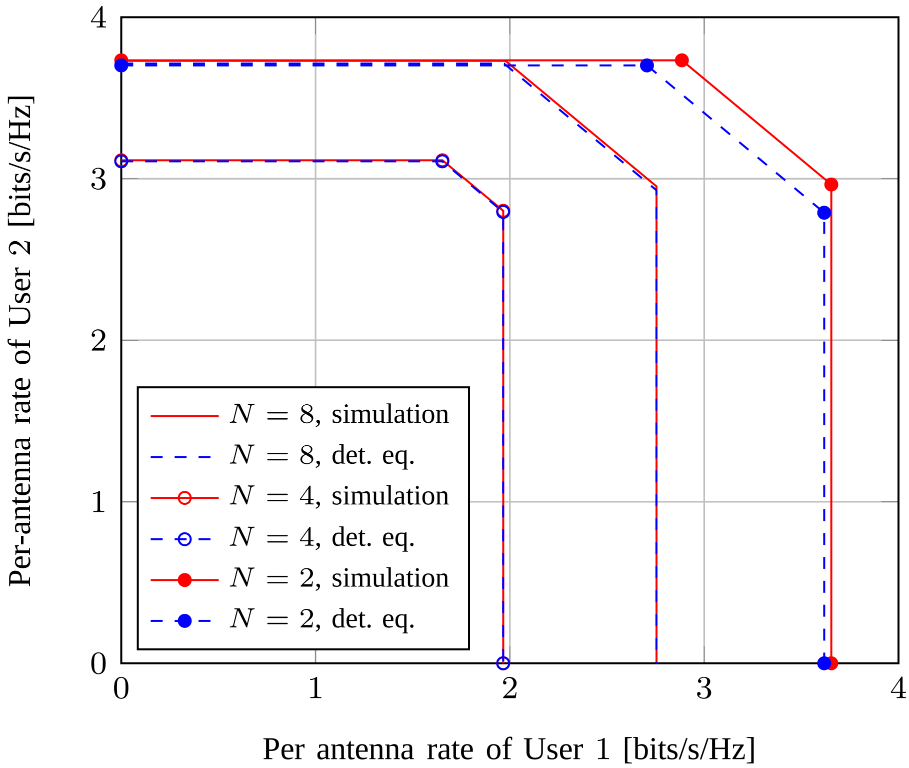


Fig. 6. Ergodic rate region of two-user MAC, uniform power allocation, for $N = 2$, $N = 4$ and $N = 8$, $n_1 = n_2 = N$, ULA model, antenna spacing $\frac{d^{\rm R}}{\lambda} = 0.5$. Comparison between simulations and deterministic equivalents (det. eq. in the figure).

doubling their number. On the opposite, when the correlation increases, doubling the number of antennas at both sides reduces the antenna efficiency. As we assume here equal power allocation across transmit antennas, in which case the precoding matrix is deterministic and independent of the channel realization, these conclusions are in phase with the conclusions of Goldsmith et al. in [45] and references therein. From Theorem 2, the exact description of this phenomenon can be thoroughly analyzed, as a function of the various system parameters involved (as long as a Kronecker channel model is assumed).

B. *Time Varying MAC Ergodic Rate Region*

We now move to the analysis of the ergodic rate region of time varying multiple access channels. In Figure 6, we provide a comparison between the simulated ergodic MAC rate region and the associated deterministic equivalents, for $N = n_1 = n_2$ varying from 2 to 8, $d^{\rm R}/\lambda = 0.5$ and all other parameters are as described above. Uniform power allocation is applied. It turns out that, although the $N = 2$ case is slightly mismatched, for $N \geq 4$, the deterministic equivalent of the ergodic rate region is very accurate. For $N = 8$, the deterministic equivalent is the same as that of Figure 3 (top); Figure 6 therefore indicates that the deterministic equivalent in the block-fading case is unbiased. In Figure 7, we consider $N = 8$ and provide both deterministic equivalents for $d^{\rm R}/\lambda = 0.5$,

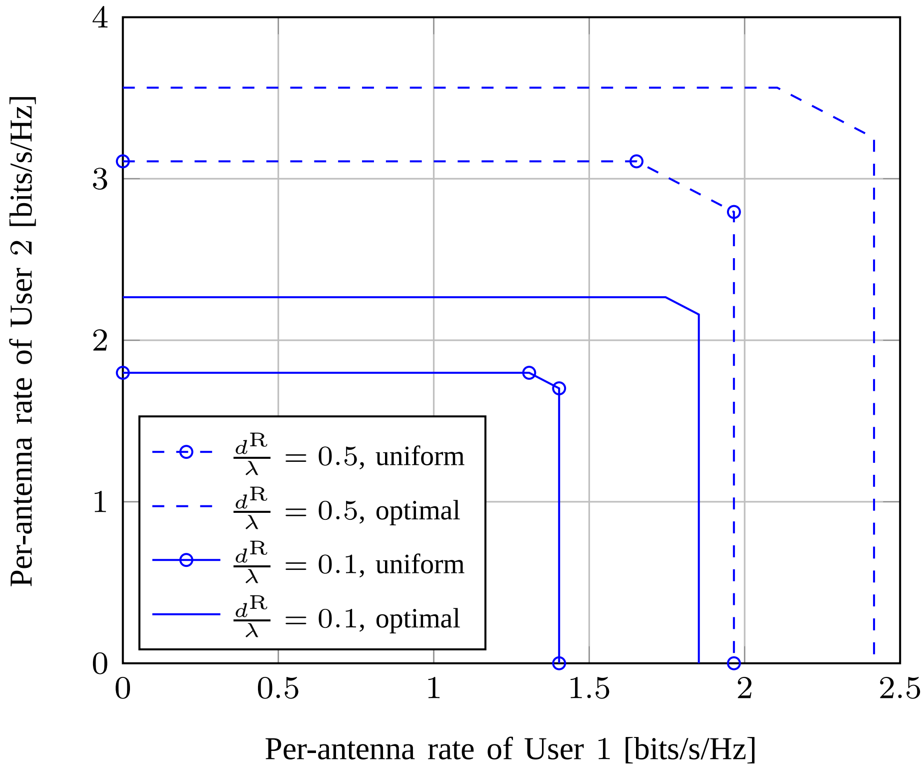


Fig. 7. Ergodic rate region of two-user MAC, equal power allocation (uniform) and rate maximizing power allocation (optimal), for $N = 8$, $n_1 = n_2 = N$, ULA model, antenna spacing $\frac{d^{\mathrm{R}}}{\lambda} = 0.5$ (dashed) and $\frac{d^{\mathrm{R}}}{\lambda} = 0.1$ (solid), $\mathrm{SNR} = 20$ dB.

$d^{\mathrm{R}}/\lambda = 0.1$, when optimal power allocation is applied or not. Again, both graphs with uniform power allocation correspond to the already presented graphs of Figure 3 (top). We did not provide Monte Carlo simulation results here, which were found to match exactly the theoretical curves. As expected [45], it turns out that the stronger the correlation patterns the higher the benefits of optimal power allocation. Under optimal power allocation though, it is less clear how antenna efficiency evolves as a function of d^{R}/λ. This is characterized in the following.

The antenna efficiency for the ergodic MAC sum rate is provided in Figure 8. When optimal power allocation is applied, the per-antenna rate loss incurred by the addition of extra antennas is similar to that observed with uniform power allocation policy. Compared to Figure 5 though, we observe that the antenna efficiency does not increase for low correlated antennas when optimal power allocation is applied, while the rate achieved when strong correlation is present increases significantly. Under the simulation conditions of Figure 8, we therefore conclude that doubling the number of antennas on a volume limited device has limited impact on the antenna efficiency whenever d^{R}/λ is of order 1 or more. We also observe the peculiar behaviour, already noticed in [45], that high correlated transmissions may lead to higher rates than low correlated transmission in the low SNR regime. The antenna efficiency is indeed shown here to be larger when $d^{\mathrm{R}}/\lambda = 0.1$ than when $d^{\mathrm{R}}/\lambda = 10$ below $\mathrm{SNR} = 0$ dB. Nonetheless, since strong correlation induces a large decrease in per-antenna efficiency as the number of antennas increases, the point at low

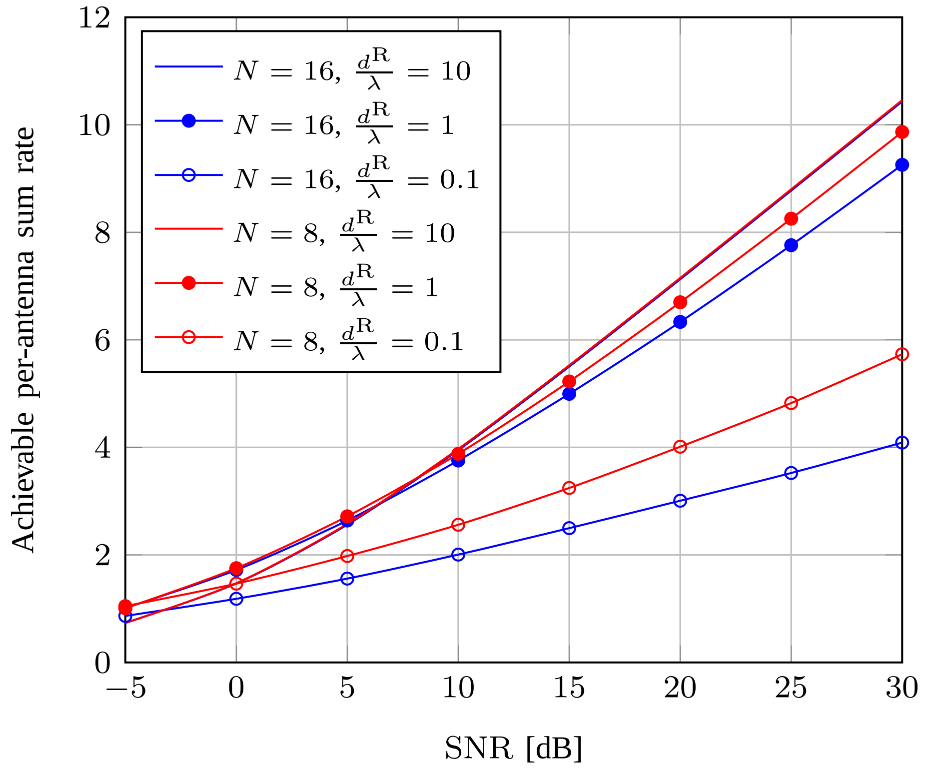


Fig. 8. Achievable per-antenna sum rate for SNR varying from -5 dB to 30 dB, for different values of $\frac{d^{\mathrm{R}}}{\lambda}$, time varying channels, sum rate maximizing power allocation.

SNR where the performance of strong correlation pattern takes over is pushed further down in the SNR range.

We therefore conclude that, for a fixed number of antennas, increasing channel correlation helps increasing communication rates in the low SNR regime, but that artificially enhancing correlation by adding more antennas does not further help. For practical applications in which high correlation and low SNR conditions may often arise, carrying a large number of antennas is therefore a choice of limited interest. In this case, a trade-off must be found between higher rates in occasional low correlated and high SNR scenarios and lower operating and manufacturing cost incurred when embedding a small number of antennas.

C. Multi-User MIMO

We now apply Equations (40) and (48) to the downlink of a two-cell network. The capacity analyzed here is the per-antenna ergodic achievable rate on the link between base station 1 and a given user, the latter of which is interfered by the transmissions from base station 2. The relative power of the interfering signal from base station 2 is on average Γ times that of base station 1. Base stations 1 and 2 are equipped with linear arrays of n_1 and n_2 antennas, respectively, and the user with a linear array of N antennas. The transmit and receive correlation matrices $\mathbf{T}_i$ and $\mathbf{R}_i$, $i \in \{1, 2\}$, are also modeled thanks to the generalized Jakes model, given in (49); the solid angles of

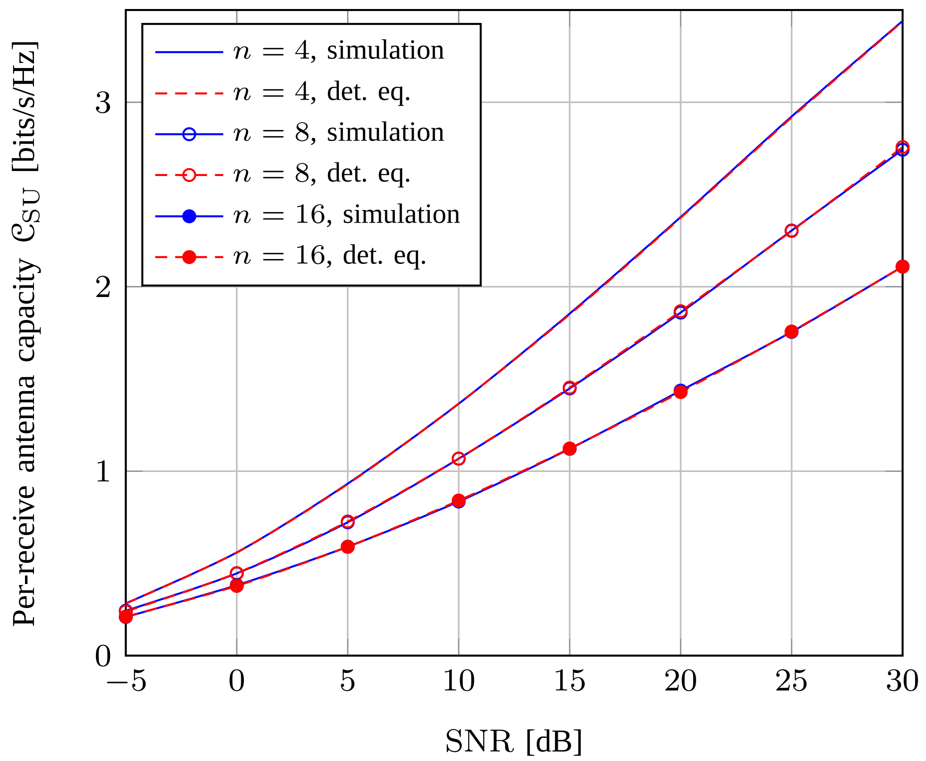


Fig. 9. Ergodic mutual information of point-to-point MIMO for the two-cell downlink scenario with single-user decoding, N varying from 4 to 16, $N = n_1 = n_2$, interfering cell with power $\Gamma = 25\%$.

effective energy transmission and reception are the same as in the previous section. Note however, in this downlink scenario, that now roles are interchanged as base stations are transmitters and no longer receivers.

In Figure 9, we consider the ergodic mutual information of single user decoding and take $N = n_1 = n_2 = 4$ to $N = n_1 = n_2 = 16$, $\Gamma = 0.25$. Uniform power allocation is applied. The distances between transmit antennas at the base stations (now transmitters) are $d^{\rm T} = 10\lambda$, while at the receive terminal, $d^{\rm R} = 0.5\lambda$. The SNR ranges from -5 dB to 30 dB. We observe here that Monte-Carlo simulations perfectly match the deterministic equivalent obtained in (40), already for $N = n = 4$. As it turns out, doubling the number of antennas in this scenario does not significantly reduce the antenna efficiency, even with strong correlation at the receiver. When performing single-user decoding, it is therefore an appropriate choice to increase the number of antennas at both communication ends, as long as the processing costs incurred are not dramatically increased.

In Figure 10, with the same assumptions as previously, we analyze the ergodic capacity of the MMSE decoding strategy. Here, the deterministic equivalent is only accurate for $N \geq 8$. We observe a significant difference in performance between the single-user and the suboptimal linear MMSE decoders, especially in the high SNR region, where the MMSE decoder performance no longer grows linearly with the SNR. Also, additional antennas bring marginal capacity gain, as their efficiency reduces rapidly with larger N. The comparison to Figure 9 suggests that additional antennas can be used much more efficiently by simultaneous stream decoding methods than by reinforcing

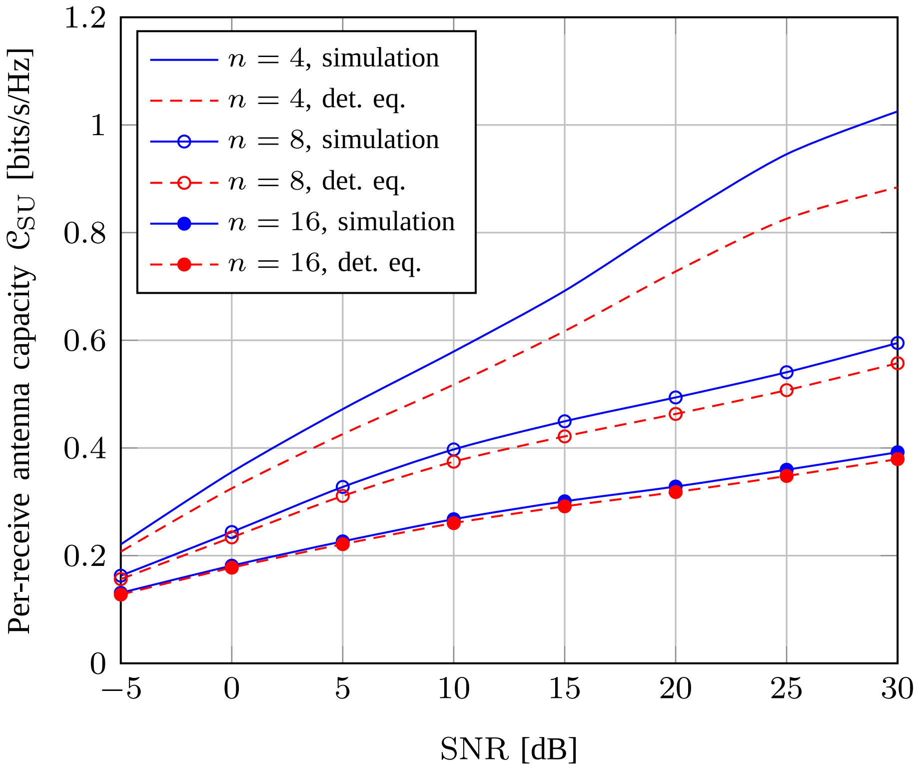


Fig. 10. Ergodic mutual information of point-to-point MIMO in two-cell downlink, MMSE decoding, N varying from 4 to 16, $N = n_1 = n_2$, interfering cell with power $\Gamma = 25\%$.

the MMSE decoding method with more antennas. Compromises in the decoding strategy might therefore be thought of when dealing with inter-cell interference. It is in particular known that single-user decoding can be achieved by performing successive MMSE decoding. If many antennas are available at the receiver side, several MMSE decoding steps are therefore expected to lead to strong performance increase compared to the single-step MMSE decoder.

Remark 2: It must be stressed that some scenarios show deterministic equivalent plots that do not converge as rapidly to the simulated plots as those presented in this section. The following misleading effect especially happens. As N and the n_k grow at the same rate, the per antenna rate performance usually decreases (as observed in all situations here), so that the Monte Carlo simulated capacity values decrease with large N. In parallel, the deterministic equivalents also decrease. Now, in the case of high correlation both at the transmit and receive sides, it often turns out that the convergence of the deterministic equivalent to the simulated capacity is very slow. The resulting effect is that, for moderately large N, the difference between the simulated performance and its corresponding deterministic equivalent decreases slowly while both curves decrease rapidly to zero; therefore, the approximation error *relative to the exact capacity value* increases with N, although the absolute error slowly decreases. For instance, if both simulation plots and deterministic equivalents decrease as $O(1/N^2)$, while their difference decreases as $O(1/N)$, then for moderate N the relative difference is of order $O(N)$. This effect is very unfortunate as it leads to plots where the simulated results are sometimes ten times larger than their deterministic equivalent, although this effect does not invalidated Theorem 1 and Theorem 2.

Note that the aforementioned problem is less accentuated when the correlation matrices have uniformly bounded spectral norms, as noticed in [26] for Rician channels, as the convergence of the deterministic equivalent is of order $O(1/N^2)$, while here the fastest convergence rate that we proved we could achieve is of order $o(1/\log^k N)$ for any k (see proof of Theorem 1 in Appendix A).

VI. Conclusion

In this contribution, we analyzed the per-antenna rate performance of a family of multi-antenna communication schemes including multiple cells and multiple users, while taking into account the correlation effects due to close antennas and reduced solid angles of energy transmission. We specifically studied the rate regions of block-fading MAC and BC channels, the rate region of the time varying MAC channel, as well as the uplink and downlink capacity of multi-cell networks with inter-cell interference. Our main results stem from novel deterministic equivalents of the Stieltjes transform and of the Shannon transform of a certain type of large dimensional random matrices. Based on these new tools, an accurate analysis of the effects of antenna correlation can be directly translated into the antenna efficiency of multi-user multi-cellular systems. It especially turned out that, for the same communication scheme, some decoding strategies suffer strongly from an increase in channel correlation, while others do not. This suggests that the trade-off between throughput gain of additional antennas and limited incurred processing cost strongly depends on the decoding strategy.

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 [6] and [18] (which provide more details in their appendices), that

$$|F^{\mathbf{B}_N} - F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}}\widetilde{\mathbf{X}}\mathbf{T}\widetilde{\mathbf{X}}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}}| \xrightarrow{\text{a.s.}} 0 \tag{50}$$

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 [6], 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 $\alpha > 0$

$$\mathbf{A}^{\alpha} = \mathbf{U}\,\mathrm{diag}(\lambda_1^{\mathbf{A}}\mathbf{1}_{\{\lambda_r^{\mathbf{A}}\leq\alpha\}}, \ldots, \lambda_r^{\mathbf{A}}\mathbf{1}_{\{\lambda_1\leq\alpha\}})\mathbf{U}^{\mathsf{H}} \tag{51}$$

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

$$\|F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}}\mathbf{Q}\mathbf{T}\mathbf{Q}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}} - F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}\alpha}Q\mathbf{T}^{\alpha}\mathbf{Q}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}\alpha}}\| \leq \frac{2}{N}\mathrm{rank}(\mathbf{R}^{\frac{1}{2}} - \mathbf{R}^{\frac{1}{2}\alpha}) + \frac{1}{N}\mathrm{rank}(\mathbf{T} - \mathbf{T}^{\alpha}) \tag{52}$$

$$= \frac{2}{N}\sum_{i=1}^{N}\mathbf{1}_{\{\lambda_i^{\mathbf{R}}>\alpha\}} + \frac{1}{N}\sum_{i=1}^{n}\mathbf{1}_{\{\lambda_i^{\mathbf{T}}>\alpha)\}} \tag{53}$$

$$= 2F^{\mathbf{R}}((\alpha,\infty)) + \frac{1}{c_N}F^{\mathbf{T}}((\alpha,\infty)) \tag{54}$$

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

$$\|F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}}\mathbf{Q}\mathbf{T}\mathbf{Q}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}}} - F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}\alpha_N}\mathbf{Q}\mathbf{T}^{\alpha_N}\mathbf{Q}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}\alpha_N}}\| \to 0 \tag{55}$$

as $N \to \infty$.

A metric D on probability measures defined on $\mathbb{R}$, which induces the topology of vague convergence, is introduced in [6] to handle the last truncation step. The matrices studied in [6] are essentially $\mathbf{B}_N$ with $\mathbf{R} = \mathbf{I}_N$. Following the steps beginning at (3.4) in [6], we see in our case that when α_N is chosen so that as $N \to \infty$, $\alpha_N \uparrow \infty$,

$$\alpha_N^8(\mathrm{E}|X_{11}^2 \mathbf{1}_{\{X_{11}|\geq \ln N\}} + N^{-1}) \to 0 \tag{56}$$

and

$$\sum_{N=1}^{\infty} \frac{\alpha_N^{16}}{N^2}\left(\mathrm{E}|X_{11}|^4 \mathbf{1}_{\{|X_{11}|<\sqrt{N})\}} + 1\right) < \infty \tag{57}$$

We will get

$$D(F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}\alpha_N}\widetilde{\mathbf{X}}\mathbf{T}^{\alpha_N}\widetilde{\mathbf{X}}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}\alpha_N}}, F^{\mathbf{S}+\mathbf{R}^{\frac{1}{2}\alpha_N}\overline{\mathbf{X}}\mathbf{T}^{\alpha_N}\overline{\mathbf{X}}^{\mathsf{H}}\mathbf{R}^{\frac{1}{2}\alpha_N}}) \xrightarrow{\text{a.s.}} 0 \tag{58}$$

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 (55) and (58) 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 (55) that we can also assume

$$\max(\|\mathbf{R}\|, \|\mathbf{T}\|) \leq \log N \tag{59}$$

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

$$\mathbf{B}_N = \mathbf{S} + \sum_{j=1}^{n} \tau_j \mathbf{y}_j \mathbf{y}_j^{\mathsf{H}} \tag{60}$$

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

$$e_N = e_N(z) = (1/N)\operatorname{tr} \mathbf{R}(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \tag{61}$$

and

$$p_N = -\frac{1}{nz}\sum_{j=1}^{n} \frac{\tau_j}{1 + c_N \tau_j e_N} = \int \frac{-\tau}{z(1 + c_N \tau e_N)} dF^{\mathbf{T}}(\tau) \tag{62}$$

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

$$e_N = (1/N)\operatorname{tr} \underline{\mathbf{R}}(\Lambda - z\mathbf{I}_N)^{-1} = (1/N)\sum_{i=1}^{N} \frac{\underline{R}_{ii}}{\lambda_i - z} \tag{63}$$

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 (59), it follows that

$$|e_N| \leq v^{-1} \log N \tag{64}$$

and

$$\left| \int \frac{\tau}{(1 + c_N \tau e_N)} dF^{\mathbf{T}}(\tau) \right| = |zp_N(z)| \leq |z| v^{-1} \log N \tag{65}$$

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

$$\frac{-\tau}{z(1 + m(z))} \tag{66}$$

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

$$\left| \frac{\tau}{1 + m(z)} \right| \leq \tau |z| v^{-1} \tag{67}$$

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

$$\mathbf{B}_N - z\mathbf{I}_N - \mathbf{D} = \sum_{j=1}^{n} \tau_j \mathbf{y}_j \mathbf{y}_j^{\mathsf{H}} + zp_N \mathbf{R} \tag{68}$$

Taking inverses and using Lemma 4 we have

$$\begin{aligned} (\mathbf{B}_N - z\mathbf{I}_N)^{-1} - \mathbf{D}^{-1} &= \sum_{j=1}^{n} \tau_j \mathbf{D}^{-1} \mathbf{y}_j \mathbf{y}_j^{\mathsf{H}} (\mathbf{B}_N - z\mathbf{I}_N)^{-1} + zp_N \mathbf{D}^{-1} \mathbf{R} (\mathbf{B}_N - z\mathbf{I}_N)^{-1} &(69)\\ &= \sum_{j=1}^{n} \tau_j \frac{\mathbf{D}^{-1} \mathbf{y}_j \mathbf{y}_j^{\mathsf{H}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1}}{1 + \tau_j \mathbf{y}_j^{\mathsf{H}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} + zp_N \mathbf{D}^{-1} \mathbf{R} (\mathbf{B}_N - z\mathbf{I}_N)^{-1} &(70)\end{aligned}$$

Taking traces and dividing by N, we have

$$\frac{1}{N} \operatorname{tr} \mathbf{D}^{-1} - m_N(z) = \frac{1}{n} \sum_{j=1}^{n} \tau_j d_j \equiv w_N^m \tag{71}$$

where

$$d_j = \frac{(1/N) \mathbf{x}_j^{\mathsf{H}} \mathbf{R}^{\frac{1}{2}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{D}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j}{1 + \tau_j \mathbf{y}_j^{\mathsf{H}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} - \frac{(1/N) \operatorname{tr} \mathbf{R} (\mathbf{B}_N - z\mathbf{I}_N)^{-1} \mathbf{D}^{-1}}{1 + c_N \tau_j e_N} \tag{72}$$

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

$$\frac{1}{N} \operatorname{tr} \mathbf{D}^{-1} \mathbf{R} - e_N(z) = \frac{1}{n} \sum_{j=1}^{n} \tau_j d_j^e \equiv w_N^e \tag{73}$$

where

$$d_j^e = \frac{(1/N) \mathbf{x}_j^{\mathsf{H}} \mathbf{R}^{\frac{1}{2}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{R} \mathbf{D}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j}{1 + \tau_j \mathbf{y}_j^{\mathsf{H}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} - \frac{(1/N) \operatorname{tr} \mathbf{R} (\mathbf{B}_N - z\mathbf{I}_N)^{-1} \mathbf{R} \mathbf{D}^{-1}}{1 + c_N \tau_j e_N} \tag{74}$$

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

$$\lim_{N \to \infty} (\log^k N) w_N^m = 0 \tag{75}$$

and

$$\lim_{n \to \infty} (\log^k N)\, w_N^e = 0 \tag{76}$$

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 (67) we have

$$\left| \frac{1}{1 + \tau_j \mathbf{y}_j^{\mathsf{H}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} \right| \leq \frac{|z|}{v}. \tag{77}$$

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

$$p_{(j)} = p_{(j)}(z) = \int \frac{-\tau}{z(1 + c_N \tau e_{(j)})} dF^{\mathbf{T}}(\tau) \tag{78}$$

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

Using Lemma 4, Equations (59) and (67), we have

$$|zp_N - zp_{(j)}| = |e_N - e_{(j)}| c_N \left| \int \frac{\tau^2}{(1 + c_N \tau e_N)(1 + c_N \tau e_{(j)})} dF^{\mathbf{T}}(\tau) \right| \leq \frac{c_N |z|^2 \log^3 N}{N v^3}. \tag{79}$$

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

$$d_j^1 = \frac{(1/N)\mathbf{x}_j^{\mathsf{H}} \mathbf{R}^{\frac{1}{2}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j}{1 + \tau_j \mathbf{y}_j^{\mathsf{H}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} - \frac{(1/N)\mathbf{x}_j^{\mathsf{H}} \mathbf{R}^{\frac{1}{2}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j}{1 + \tau_j \mathbf{y}_j^{\mathsf{H}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} \tag{80}$$

$$d_j^2 = \frac{(1/N)\mathbf{x}_j^{\mathsf{H}} \mathbf{R}^{\frac{1}{2}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j}{1 + \tau_j \mathbf{y}_j^{\mathsf{H}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} - \frac{(1/N) \operatorname{tr} \mathbf{R} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1}}{1 + \tau_j \mathbf{y}_j^{\mathsf{H}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} \tag{81}$$

$$d_j^3 = \frac{(1/N) \operatorname{tr} \mathbf{R} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1}}{1 + \tau_j \mathbf{y}_j^{\mathsf{H}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} - \frac{(1/N) \operatorname{tr} \mathbf{R} (\mathbf{B}_N - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}^{-1}}{1 + \tau_j \mathbf{y}_j^{\mathsf{H}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} \tag{82}$$

$$d_j^4 = \frac{(1/N) \operatorname{tr} \mathbf{R} (\mathbf{B}_N - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}^{-1}}{1 + \tau_j \mathbf{y}_j^{\mathsf{H}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{y}_j} - \frac{(1/N) \operatorname{tr} \mathbf{R} (\mathbf{B}_N - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}^{-1}}{1 + c_N \tau_j e_N} \tag{83}$$

From Lemma 4, Equations (59), (77) and (79), we have

$$\tau_j |d_j^1| \leq \frac{1}{N} \|\mathbf{x}_j\|^2 \frac{c_N \log^7 N |z|^3}{N v^7} \tag{84}$$

$$\tau_j |d_j^2| \leq |z| v^{-1} \frac{\log N}{N} \left| \mathbf{x}_j^{\mathsf{H}} \mathbf{R}^{\frac{1}{2}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j - \operatorname{tr} \mathbf{R} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1} \right| \tag{85}$$

$$\tau_j |d_j^3| \leq \frac{|z| \log^3 N}{v N} \left(\frac{1}{v^2} + \frac{c_N |z|^2 \log^3 N}{v^6} \right) \to 0, \text{ as } n \to \infty \tag{86}$$

$$\tau_j |d_j^4| \leq \frac{|z| c_N \log^4 N}{N v^3} \left(|\mathbf{x}_j^{\mathsf{H}} \mathbf{R}^{\frac{1}{2}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j - \operatorname{tr} \mathbf{R}^{\frac{1}{2}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{R}^{\frac{1}{2}} | + \frac{\log N}{v} \right) \tag{87}$$

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

$$\mathrm{E} \left| \frac{1}{N} \|\mathbf{x}_j\|^2 - 1 \right|^6 \leq K N^{-3} \log^{12} N \tag{88}$$

$$\mathrm{E} \frac{1}{N^6} |\mathbf{x}_j^{\mathsf{H}} \mathbf{R}^{\frac{1}{2}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j - \operatorname{tr} \mathbf{R} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{E} \mathbf{D}_{(j)}^{-1}|^6 \leq K N^{-3} v^{-12} \log^{24} N \tag{89}$$

$$\mathrm{E} \frac{1}{N^6} |\mathbf{x}_j^{\mathsf{H}} \mathbf{R}^{\frac{1}{2}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j - \operatorname{tr} \mathbf{R}^{\frac{1}{2}} (\mathbf{B}_{(j)} - z\mathbf{I}_N)^{-1} \mathbf{R}^{\frac{1}{2}}|^6 \leq K N^{-3} v^{-6} \log^{18} N \tag{90}$$

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 $k > 0$ $\log^k N \max_{j \leq n} \tau_j d_j \xrightarrow{\text{a.s.}} 0$ as $N \to \infty$. Hence Equations (75) and (76).

C. Existence and uniqueness of $m_N^\circ(z)$

We show now that for any N, n, $\mathbf{S}$, $\mathbf{R}$, $N \times N$ nonnegative definite and $\mathbf{T} = \mathrm{diag}(\tau_1, \ldots, \tau_N)$, $\tau_k \geq 0$ for all $1 \leq k \leq N$, there exists a unique e with positive imaginary part for which

$$e = \frac{1}{N} \operatorname{tr} \left(\mathbf{S} + \left[\int \frac{\tau}{1 + c_N \tau e} dF^{\mathbf{T}}(\tau) \right] \mathbf{R} - z \mathbf{I}_N \right)^{-1} \mathbf{R} \tag{91}$$

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

$$\mathbf{R}_{N_j} = \mathrm{diag}(\mathbf{R}, \mathbf{R}, \ldots, \mathbf{R}), \quad \mathbf{S}_{N_j} = \mathrm{diag}(\mathbf{S}, \mathbf{S}, \ldots, \mathbf{S}) \tag{92}$$

both $jN \times jN$ and

$$\mathbf{T}_{N_j} = \mathrm{diag}(\mathbf{T}, \mathbf{T}, \ldots, \mathbf{T}) \tag{93}$$

of size $jn \times jn$.

We see that $F^{\mathbf{T}_{N_j}} = F^{\mathbf{T}}$ and the right side of (91) remains unchanged for all N_j. Consider a realization where $w_{N_j}^e \to 0$ as $j \to \infty$. We have $|e_{N_j}(z)| = |(jN)^{-1} \operatorname{tr} \mathbf{R}(\mathbf{B}_{jN} - z\mathbf{I}_N)^{-1}| \leq v^{-1} \log N$, remaining bounded as $j \to \infty$. Consider then a subsequence for which e_{N_j} converges to, say, e. From (67), we see that

$$\left| \frac{\tau}{1 + c_N \tau e_{N_j}} \right| \leq \tau |z| v^{-1} \tag{94}$$

so that from the dominated convergence theorem we have

$$\int \frac{\tau}{1 + c_N \tau e_{N_j}(z)} dF^{\mathbf{T}}(\tau) \to \int \frac{\tau}{1 + c_N \tau e} dF^{\mathbf{T}}(\tau) \tag{95}$$

along this subsequence. Therefore e solves (91).

We now show uniqueness. Let e be a solution to (91) and let $e_2 = \Im[e]$. Recalling the definition of $\mathbf{D}$ we write

$$e = \frac{1}{N} \operatorname{tr} \left(\mathbf{D}^{-1} \mathbf{R} \mathbf{D}^{-\mathsf{H}} \left(\mathbf{S} + \left[\int \frac{\tau}{1 + c_N \tau e^*} dF^{\mathbf{T}}(\tau) \right] \mathbf{R} - z^* \mathbf{I} \right) \right) \tag{96}$$

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

$$e_2 = \frac{1}{N} \operatorname{tr} \left(\mathbf{D}^{-1} \mathbf{R} (\mathbf{D}^{\mathsf{H}})^{-1} \left(\left[\int \frac{c_N \tau^2 e_2}{|1 + c_N \tau e|^2} dF^{\mathbf{T}}(\tau) \right] \mathbf{R} + v \mathbf{I}_N \right) \right) = e_2 \alpha + v \beta \tag{97}$$

where we denoted

$$\alpha = \frac{1}{N} \operatorname{tr} \left(\mathbf{D}^{-1} \mathbf{R} (\mathbf{D}^{\mathsf{H}})^{-1} \left[\int \frac{c_N \tau^2}{|1 + c_N \tau e|^2} dF^{\mathbf{T}}(\tau) \right] \mathbf{R} \right) \tag{98}$$

$$\beta = \frac{1}{N} \operatorname{tr} \left(\mathbf{D}^{-1} \mathbf{R} (\mathbf{D}^{\mathsf{H}})^{-1} \right) \tag{99}$$

Let $\underline{e}$ be another solution to (91), 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

$$\gamma = \int \frac{c_N \tau^2}{(1 + c_N \tau e)(1 + c_N \tau \underline{e})} dF^{\mathbf{T}}(\tau) \frac{\operatorname{tr} \mathbf{D}^{-1} \mathbf{R} \underline{\mathbf{D}}^{-1} \mathbf{R}}{N} \tag{100}$$

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

$$|\gamma| \leq \left(\int \frac{c_N \tau^2}{|1 + c_N \tau e|^2} dF^{\mathbf{T}}(\tau) \frac{\operatorname{tr} \mathbf{D}^{-1} \mathbf{R} (\mathbf{D}^{\mathsf{H}})^{-1} \mathbf{R}}{N} \right)^{\frac{1}{2}} \left(\int \frac{c_N \tau^2}{|1 + c_N \tau \underline{e}|^2} dF^{\mathbf{T}}(\tau) \frac{\operatorname{tr} \underline{\mathbf{D}}^{-1} \mathbf{R} (\underline{\mathbf{D}}^{\mathsf{H}})^{-1} \mathbf{R}}{N} \right)^{\frac{1}{2}} \tag{101}$$

$$= \alpha^{\frac{1}{2}} \underline{\alpha}^{\frac{1}{2}} \tag{102}$$

$$= \left(\frac{e_2 \alpha}{e_2 \alpha + v \beta} \right)^{\frac{1}{2}} \left(\frac{\underline{e_2 \alpha}}{\underline{e_2 \alpha} + v \underline{\beta}} \right)^{\frac{1}{2}} \tag{103}$$

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

For $z < 0$ and $e > 0$, the same calculus can be performed, with γ remaining the same. The step (97) is changed by evaluating e, instead of e_2, using the same technique. We obtain the same α while β is replaced by another positive scalar. We therefore still have that $\gamma < 1$.

D. *Termination of the proof*

Let e_N° denote the solution to (91). We show now for any $\ell > 0$, almost surely

$$\lim_{N \to \infty} \log^\ell N (e_N - e_N^\circ) = 0 \tag{104}$$

Let $e_2^\circ = \Im[e_N^\circ]$, and $\alpha^\circ = \alpha_N^\circ$, $\beta^\circ = \beta_N^\circ$ be the values as above for which $e_2^\circ = e_2^\circ \alpha^\circ + v \beta^\circ$. We have, using (59) and (67),

$$e_2^\circ \alpha_N^\circ / \beta_N^\circ \leq e_2^\circ c_N \log N \int \frac{\tau^2}{|1 + c_N \tau e_N^\circ|^2} dF^{\mathbf{T}}(\tau) \tag{105}$$

$$= -\log N \Im \left[\int \frac{\tau}{1 + c_N \tau e_N^\circ} dF^{\mathbf{T}}(\tau) \right] \tag{106}$$

$$\leq \log^2 N |z| v^{-1} \tag{107}$$

Therefore

$$\alpha^\circ = \left(\frac{e_2^\circ \alpha^\circ}{e_2^\circ \alpha^\circ + v \beta^\circ} \right) \tag{108}$$

$$= \left(\frac{e_2^\circ \alpha^\circ / \beta^\circ}{v + e_2^\circ \alpha^\circ / \beta^\circ} \right) \tag{109}$$

$$\leq \left(\frac{\log^2 N |z|}{v^2 + \log^2 N |z|} \right) \tag{110}$$

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

$$e_N = \frac{1}{N} \operatorname{tr} \mathbf{D}^{-1} \mathbf{R} - w_N^e \tag{111}$$

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

$$e_2 = \frac{1}{N} \operatorname{tr} \left(\mathbf{D}^{-1} \mathbf{R} \mathbf{D}^{-\mathsf{H}} \left(\left[\int \frac{c_N \tau^2 e_2}{|1 + c_N \tau e_N|^2} dF^{\mathbf{T}}(\tau) \right] \mathbf{R} + v \mathbf{I}_N \right) \right) - \Im[w_N^e] \tag{112}$$

$$= e_2 \alpha + v \beta - \Im w_N^e \tag{113}$$

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

$$|\gamma| \leq \alpha^{\circ\frac{1}{2}}\alpha^{\frac{1}{2}} \tag{114}$$

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

$$|w_N^e| \leq \frac{v^3}{4c_N|z|^2\log^3 N} \tag{115}$$

Suppose $\beta \leq \frac{v^2}{4c_N|z|^2\log^3 N}$. Then by Equations (59) and (67) we get

$$\alpha \leq c_N v^{-2}|z|^2\log^3 N\beta \leq 1/4 \tag{116}$$

which implies $|\gamma| \leq 1/2$. Otherwise we get from (108) and (115)

$$|\gamma| \leq \alpha^{\circ\frac{1}{2}}\left(\frac{e_2\alpha}{e_2\alpha + v\beta - \Im[w_N^e]}\right)^{\frac{1}{2}} \tag{117}$$

$$\leq \left(\frac{\log N|z|}{v^2 + \log N|z|}\right)^{\frac{1}{2}} \tag{118}$$

Therefore for all N large

$$\log^\ell N|e_N - e_N^\circ| \leq \frac{(\log^\ell N)w_N^e}{1 - \left(\frac{\log^2 N|z|}{v^2+\log^2 N|z|}\right)^{\frac{1}{2}}} \tag{119}$$

$$\leq 2v^{-2}(v^2 + \log^2 N|z|)(\log^\ell N)w_N^e \tag{120}$$

$$\to 0 \tag{121}$$

as $n \to \infty$. Therefore (104) follows.

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

$$m_N - m_N^\circ \xrightarrow{\text{a.s.}} 0 \tag{122}$$

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

$$m_N - m_N^\circ = \gamma(e_N - e_N^\circ) - w_N^m \tag{123}$$

where now

$$\gamma = \int \frac{c_N\tau^2}{(1+c_N\tau e_N)(1+c_N\tau e_N^\circ)}dF^{\mathbf{T}}(\tau)\frac{\operatorname{tr}\mathbf{D}^{-1}\mathbf{R}\mathbf{D}^{\circ-1}}{N} \tag{124}$$

From (59) and (67) we get $|\gamma| \leq c_N|z|^2 v^{-4}\log^3 N$. Therefore, from (75) and (104), we get (122).

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 (122). Therefore, by Vitali's convergence theorem, page 168 of [15], for any $\varepsilon > 0$ we have with probability one $m_N(z) - m_N^\circ(z) \to 0$ uniformly in any region of $\mathbb{C}$ bounded by a contour interior to

$$\mathbb{C} \setminus (\{z : |z| \leq \epsilon\} \cup \{z = x + iv : x > 0, |v| \leq \varepsilon\}) \tag{125}$$

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

$$m_N^\circ(z) = \int \frac{1}{f(r) + r\int \frac{\tau}{1+c_N\tau e_N^\circ(z)} dF^{\mathbf{T}}(\tau) - z} dF^{\mathbf{R}}(r) \tag{126}$$

$$e_N^\circ(z) = \int \frac{r}{f(r) + r\int \frac{\tau}{1+c_N\tau e_N^\circ(z)} dF^{\mathbf{T}}(\tau) - z} dF^{\mathbf{R}}(r) \tag{127}$$

E. Extension to $K \geq 1$

Suppose now

$$\mathbf{B}_N = \mathbf{S} + \sum_{k=1}^{K} \mathbf{R}_k^{\frac{1}{2}} \mathbf{X}_k \mathbf{T}_k \mathbf{X}_k^{\mathsf{H}} \mathbf{R}_k^{\frac{1}{2}} \tag{128}$$

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

$$\mathbf{B}_N = \mathbf{S} + \sum_{k=1}^{K} \sum_{j=1}^{n_k} \tau_{k,j} \mathbf{y}_{k,j} \mathbf{y}_{k,j}^{\mathsf{H}} \tag{129}$$

Define

$$e_{N,k} = e_{N,k}(z) = (1/N)\operatorname{tr} \mathbf{R}_k(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \tag{130}$$

and

$$p_k = -\frac{1}{n_k z} \sum_{j=1}^{n_k} \frac{\tau_{k,j}}{1 + c_k \tau_{k,j} e_{N,k}} \tag{131}$$

$$= \int \frac{-\tau_k}{1 + c_k \tau_k e_{N,k}} dF^{\mathbf{T}_k}(\tau_k) \tag{132}$$

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

$$\mathbf{B}_N - z\mathbf{I}_N - \mathbf{D} = \sum_{k=1}^{K} \left(\sum_{j=1}^{n_k} \tau_{k,j} \mathbf{y}_{k,j} \mathbf{y}_{k,j}^{\mathsf{H}} + zp_k(z)\mathbf{R}_k \right) \tag{133}$$

Taking inverses and using Lemma 4, we have

$$\mathbf{D}^{-1} - (\mathbf{B}_N - z\mathbf{I}_N)^{-1} = \sum_{k=1}^{K} \left(\sum_{j=1}^{n_k} \tau_{k,j} \mathbf{D}^{-1} \mathbf{y}_{k,j} \mathbf{y}_{k,j}^{\mathsf{H}} (\mathbf{B}_N - z\mathbf{I}_N)^{-1} + zp_k \mathbf{D}^{-1} \mathbf{R}_k (\mathbf{B}_N - z\mathbf{I}_N)^{-1} \right) \tag{134}$$

$$= \sum_{k=1}^{K} \left(\sum_{j=1}^{n_k} \tau_{k,j} \frac{\mathbf{D}^{-1}\mathbf{y}_{k,j}\mathbf{y}_{k,j}^{\mathsf{H}}(\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1}}{1 + \tau_{k,j}\mathbf{y}_{k,j}^{\mathsf{H}}(\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_{k,j}} + zp_k \mathbf{D}^{-1}\mathbf{R}_k(\mathbf{B}_N - z\mathbf{I}_N)^{-1} \right) \tag{135}$$

Taking traces and dividing by N, we have

$$(1/N)\operatorname{tr}\mathbf{D}^{-1} - m_N(z) = \sum_{k=1}^{K} \frac{1}{n_k} \sum_{j=1}^{n_k} \tau_{k,j} d_{k,j} \equiv w_N^m \tag{136}$$

where

$$d_{k,j} = \frac{(1/N)\mathbf{x}_{k,j}^{\mathsf{H}}\mathbf{R}_k^{\frac{1}{2}}(\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}\mathbf{R}_k^{\frac{1}{2}}\mathbf{x}_{k,j}}{1 + \tau_{k,j}\mathbf{y}_{k,j}^{\mathsf{H}}(\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_{k,j}} - \frac{(1/N)\operatorname{tr}\mathbf{R}_k(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{D}^{-1}}{1 + c_k\tau_{k,j}e_{N,k}} \tag{137}$$

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

$$(1/N)\operatorname{tr}\mathbf{D}^{-1}\mathbf{R}_{\underline{k}} - e_{\underline{k}}(z) = \sum_{k=1}^{K}\frac{1}{n_k}\sum_{j=1}^{n_k}\tau_{k,j}d_{k\underline{k}j}^{e} \equiv w_{\underline{k}}^{e} \tag{138}$$

where

$$d_{k\underline{k}j}^{e} = \frac{(1/N)\mathbf{x}_{k,j}^{\mathsf{H}}\mathbf{R}_k^{\frac{1}{2}}(\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1}\mathbf{R}_{\underline{k}}\mathbf{D}^{-1}\mathbf{R}_k^{\frac{1}{2}}\mathbf{x}_{k,j}}{1 + \tau_{k,j}\mathbf{y}_{k,j}^{\mathsf{H}}(\mathbf{B}_{k,(j)} - z\mathbf{I}_N)^{-1}\mathbf{y}_{k,j}} - \frac{(1/N)\operatorname{tr}\mathbf{R}_k(\mathbf{B}_N - z\mathbf{I}_N)^{-1}\mathbf{R}_{\underline{k}}\mathbf{D}^{-1}}{1 + c_k\tau_{k,j}e_{N,k}} \tag{139}$$

In exactly the same way as in the case with $K = 1$ we find that for any nonnegative ℓ, $\log^{\ell} N w_N^m$ and the $\log^{\ell} w_{\underline{i}}^{e}$'s converge almost surely to zero. By considering block diagonal matrices as before with N, n_i's, $\mathbf{S}$, $\mathbf{R}_i$'s and $\mathbf{T}_i$'s all fixed we find that there exist $e_1^{\circ}, \dots, e_K^{\circ}$ with positive imaginary parts for which for each i

$$e_i^{\circ} = \frac{1}{N}\operatorname{tr}\mathbf{R}_i\left(\mathbf{S} + \sum_{k=1}^{K}\left[\int\frac{\tau}{1 + c_k\tau e_k^{\circ}}dF^{\mathbf{T}_k}(\tau)\right]\mathbf{R}_k - z\mathbf{I}_N\right)^{-1} \tag{140}$$

Let us verify uniqueness. Let $\mathbf{e}^{\circ} = (e_1^{\circ}, \dots, e_K^{\circ})^{\mathsf{T}}$, and let $\mathbf{D}^{\circ}$ denote the matrix in (140) whose inverse is taken (essentially $\mathbf{D}$ after the $e_{N,i}$'s are replaced by the e_i°'s). Let for each j, $e_{j,2}^{\circ} = \Im e_j^{\circ}$, and $\mathbf{e}_2^{\circ} = (e_{1,2}^{\circ}, \dots, e_{K,2}^{\circ})^{\mathsf{T}}$. Then, noticing that for each i, $\operatorname{tr}\mathbf{S}\mathbf{D}^{\circ-\mathsf{H}}\mathbf{R}_i\mathbf{D}^{\circ-1}$ is real and nonnegative (positive whenever $\mathbf{S} \neq 0$) and $\operatorname{tr}\mathbf{D}^{\circ-\mathsf{H}}\mathbf{R}_i\mathbf{D}^{\circ-1}$ and $\operatorname{tr}\mathbf{R}_j\mathbf{D}^{\circ-\mathsf{H}}\mathbf{R}_i\mathbf{D}^{\circ-1}$ are real and positive for all i, j, we have

$$e_{i,2}^{\circ} = \Im\left[\frac{1}{N}\operatorname{tr}\left(\mathbf{S} + \sum_{j=1}^{K}\left[\int\frac{\tau}{1 + c_j\tau\overline{e}_j^{\circ}}dF^{\mathbf{T}_j}(\tau)\right]\mathbf{R}_j - z^*\mathbf{I}\right)\mathbf{D}^{\circ-\mathsf{H}}\mathbf{R}_i\mathbf{D}^{\circ-1}\right] \tag{141}$$

$$= \sum_{j=1}^{K}e_{j,2}^{\circ}\frac{1}{N}\operatorname{tr}\mathbf{R}_j\mathbf{D}^{\circ-\mathsf{H}}\mathbf{R}_i\mathbf{D}^{\circ-1}c_j\int\frac{\tau^2}{|1 + c_j\tau e_j^{\circ}|^2}dF^{\mathbf{T}_j}(\tau) + \frac{v}{N}\operatorname{tr}\mathbf{D}^{\circ-\mathsf{H}}\mathbf{R}_i\mathbf{D}^{\circ-1} \tag{142}$$

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

$$c_{ij}^{\circ} = \frac{1}{N}\operatorname{tr}\mathbf{R}_j\mathbf{D}^{\circ-\mathsf{H}}\mathbf{R}_i\mathbf{D}^{\circ-1}c_j\int\frac{\tau^2}{|1 + c_j\tau e_j^{\circ}|^2}dF^{\mathbf{T}_j}(\tau) \tag{143}$$

and

$$b_i^{\circ} = \frac{1}{N}\operatorname{tr}\mathbf{D}^{\circ-\mathsf{H}}\mathbf{R}_i\mathbf{D}^{\circ-1} \tag{144}$$

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

$$\mathbf{e}_2^{\circ} = \mathbf{C}^{\circ}\mathbf{e}_2^{\circ} + v\mathbf{b}^{\circ} \tag{145}$$

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

Let $\underline{\mathbf{e}}^{\circ} = (\underline{e}_1^{\circ}, \dots, \underline{e}_K^{\circ})^{\mathsf{T}}$ be another solution to (140), with $\underline{\mathbf{e}}_2^{\circ}$, $\underline{\mathbf{D}}^{\circ}$, $\underline{\mathbf{C}}^{\circ} = (\underline{c}_{ij}^{\circ})$, $\underline{\mathbf{b}}^{\circ}$ defined analogously, so that (145) holds and $\rho(\underline{\mathbf{C}}^{\circ}) < 1$. We have for each i,

$$e_i^{\circ} - \underline{e}_i^{\circ} = \frac{1}{N}\operatorname{tr}\mathbf{R}_i\mathbf{D}^{\circ-1}\sum_{j=1}^{K}(e_j^{\circ} - \underline{e}_j^{\circ})c_j\int\frac{\tau^2}{(1 + c_j\tau e_j^{\circ})(1 + c_j\tau\underline{e}_j^{\circ})}dF^{\mathbf{T}_j}(\tau)\mathbf{R}_j\underline{\mathbf{D}}^{\circ-1} \tag{146}$$

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

$$a_{ij} = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \mathbf{D}^{\circ -1} \mathbf{R}_j \underline{\mathbf{D}}^{\circ -1} c_j \int \frac{\tau^2}{(1 + c_j \tau e_j^{\circ})(1 + c_j \tau \underline{e}_j^{\circ})} dF^{\mathbf{T}_j}(\tau) \tag{147}$$

we have

$$\mathbf{e}^{\circ} - \underline{\mathbf{e}}^{\circ} = \mathbf{A}(\mathbf{e}^{\circ} - \underline{\mathbf{e}}^{\circ}) \tag{148}$$

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

Applying Cauchy-Schwarz we have

$$|a_{ij}| \leq \left(\frac{1}{N} \mathbf{R}_i \mathbf{D}^{\circ -1} \mathbf{R}_j \mathbf{D}^{\circ -\mathsf{H}} \int \frac{\tau^2}{|1 + c_j \tau e_j^{\circ}|^2} dF^{\mathbf{T}_j}(\tau) \right)^{\frac{1}{2}} \left(\frac{1}{N} \mathbf{R}_i \underline{\mathbf{D}}^{\circ -1} \mathbf{R}_j \underline{\mathbf{D}}^{\circ -\mathsf{H}} \int \frac{\tau^2}{|1 + c_j \tau \underline{e}_j^{\circ}|^2} dF^{\mathbf{T}_j}(\tau) \right)^{\frac{1}{2}} \tag{149}$$

$$= {c_{ij}^{\circ}}^{1/2} {\underline{c}_{ij}^{\circ}}^{1/2} \tag{150}$$

Therefore from Lemmas 10 and 11 we get

$$\rho(\mathbf{A}) \leq \rho({c_{ij}^{\circ}}^{\frac{1}{2}} {\underline{c}_{ij}^{\circ}}^{\frac{1}{2}}) \leq \rho(\mathbf{C}^{\circ})^{\frac{1}{2}} \rho(\underline{\mathbf{C}}^{\circ})^{\frac{1}{2}} < 1 \tag{151}$$

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

The same reasoning can be applied to $z < 0$, with $e_i^{\circ} > 0$. In this case matrix $\mathbf{A}$ remains the same. The step (141) is now replaced by taking e_i°, instead of its imaginary part, using the same line of reasoning. This leads to the same matrix $\mathbf{C}^{\circ}$ with (145) remaining true with $\mathbf{b}^{\circ}$ replaced by another positive vector. The conclusion $\rho(\mathbf{A}) < 1$ therefore remains.

Let $\mathbf{e}_N = (e_{N,1}, \ldots, e_{N,K})^{\mathsf{T}}$ and $\mathbf{e}_N^{\circ} = (e_{N,1}^{\circ}, \ldots, e_{N,K}^{\circ})$ denote the vector solution to (140) for each N. We will show for any $\ell > 0$, almost surely

$$\lim_{N \to \infty} \log^{\ell} N(\mathbf{e}_N - \mathbf{e}_N^{\circ}) \to \mathbf{0} \tag{152}$$

We have

$$\mathbf{e}_N^{\circ} = (\frac{1}{N} \operatorname{tr} \mathbf{R}_1 \mathbf{D}^{\circ -1}, \ldots, \frac{1}{N} \operatorname{tr} \mathbf{R}_K \mathbf{D}^{\circ -1})^{\mathsf{T}} \tag{153}$$

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

$$\mathbf{e}_N = (\frac{1}{N} \operatorname{tr} \mathbf{R}_1 \mathbf{D}^{-1}, \ldots, \frac{1}{N} \operatorname{tr} \mathbf{R}_K \mathbf{D}^{-1})^{\mathsf{T}} + \mathbf{w}^e \tag{154}$$

Therefore

$$\mathbf{e}_N - \mathbf{e}_N^{\circ} = \mathbf{A}(N)(\mathbf{e}_N - \mathbf{e}_N^{\circ}) + \mathbf{w}^e \tag{155}$$

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

$$a_{ij}(N) = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \mathbf{D}^{-1} \mathbf{R}_j \mathbf{D}^{\circ -1} c_j \int \frac{\tau^2}{(1 + c_j \tau e_{N,j})(1 + c_j \tau e_{N,j}^{\circ})} dF^{\mathbf{T}_j}(\tau) \tag{156}$$

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

$$c_{ij}(N) = \frac{1}{N} \operatorname{tr} \mathbf{R}_j \mathbf{D}^{-\mathsf{H}} \mathbf{R}_i \mathbf{D}^{-1} c_j \int \frac{\tau^2}{|1 + c_j \tau e_{N,j}|^2} dF^{\mathbf{T}_j}(\tau) \tag{157}$$

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

$$b_{N,i} = \frac{1}{N} \operatorname{tr} \mathbf{D}^{-\mathsf{H}} \mathbf{R}_i \mathbf{D}^{-1} \tag{158}$$

Then, as above we find that

$$\mathbf{e}_{N,2} = \mathbf{C}(N) \mathbf{e}_{N,2} + v \mathbf{b}_N + \mathbf{w}_2^e \tag{159}$$

Using (59) and (67) we see there exists a constant $K_1 > 0$ for which

$$c_{ij}^\circ(N) \le K_1 \log^3 N b_{N,i}^\circ \tag{160}$$

and

$$c_{ij}(N) \le K_1 \log^3 N b_{N,i} \tag{161}$$

$$c_{ij}(N) \le K_1 \log^4 N \tag{162}$$

for each i, j. Therefore, from (145) we see there exists $\hat{K} > 0$ for which

$$e_{N,i}^\circ \le \hat{K} \log^4 N v b_{N,i}^\circ \tag{163}$$

Let $\mathbf{x}$ be such that $\mathbf{x}^\mathsf{T}$ is a left eigenvector of $\mathbf{C}^\circ(N)$ corresponding to eigenvalue $\rho(\mathbf{C}^\circ(N))$, guaranteed by Lemma 12. Then from (159) we have

$$\mathbf{x}^\mathsf{T} \mathbf{e}_{N,2}^\circ = \rho(\mathbf{C}^\circ(N)) \mathbf{x}^\mathsf{T} \mathbf{e}_{N,2}^\circ + v \mathbf{x}^\mathsf{T} \mathbf{b}_N^\circ \tag{164}$$

Using (164) we have

$$1 - \rho(\mathbf{C}^\circ(N)) = \frac{v \mathbf{x}^\mathsf{T} \mathbf{b}_N^\circ}{\mathbf{x}^\mathsf{T} \mathbf{e}_{N,2}^\circ} \ge (\hat{K} \log^4 N)^{-1} \tag{165}$$

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 $p \ge 12K - 7$. We will show for all N large

$$\rho(\mathbf{C}(N)) \le 1 + (\hat{K} \log^4 N)^{-1} \tag{166}$$

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

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11}(N) & \mathbf{C}_{12}(N) \\ \mathbf{C}_{21}(N) & \mathbf{C}_{22}(N) \end{pmatrix} \tag{167}$$

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 $v b_{N,i} + \mathbf{w}_{2,i}^e \le 0$, then necessarily $v b_{N,i} \le |\mathbf{w}_N^e| \le K_1 (\log n)^{-(3+p)}$, and so from (163) 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

$$\det(\mathbf{C}(N) - \lambda\mathbf{I}) = \det\left[\begin{pmatrix} \mathbf{I} & -\mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda\mathbf{I})^{-1} \\ 0 & \mathbf{I} \end{pmatrix}\begin{pmatrix} \mathbf{C}_{11}(N) - \lambda\mathbf{I} & \mathbf{C}_{12}(N) \\ \mathbf{C}_{21}(N) & \mathbf{C}_{22}(N) - \lambda\mathbf{I} \end{pmatrix}\right] \tag{168}$$

$$= \det\begin{pmatrix} \mathbf{C}_{11}(N) - \lambda\mathbf{I} - \mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda\mathbf{I})^{-1}\mathbf{C}_{21}(N) & 0 \\ \mathbf{C}_{21}(N) & \mathbf{C}_{22}(N) - \lambda\mathbf{I} \end{pmatrix} \tag{169}$$

$$= \det(\mathbf{C}_{11}(N) - \lambda\mathbf{I} - \mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda\mathbf{I})^{-1}\mathbf{C}_{21}(N))\det(\mathbf{C}_{22}(N) - \lambda\mathbf{I}) \tag{170}$$

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

$$\det(\mathbf{C}_{11}(N) - \lambda\mathbf{I} - \mathbf{C}_{12}(N)(\mathbf{C}_{22}(N) - \lambda\mathbf{I})^{-1}\mathbf{C}_{21}(N)) \tag{171}$$

must be zero.

Notice that from (163), 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 (163), we see that $\rho(\mathbf{C}(N)) \leq K_1 \log^4 N$. The determinant in (171) can be written as

$$\det(\mathbf{C}_{11}(N) - \lambda\mathbf{I}) + g(\lambda) \tag{172}$$

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 (163) 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

$$|\det(\mathbf{C}_{11}(N) - \lambda\mathbf{I})| = |\prod_{i=1}^{k_1}(\lambda_i - \lambda)| \tag{173}$$

$$> (\widehat{K}\log N)^{-4k_1} \tag{174}$$

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 (166).

As before we see that

$$|a_{ij}(N)| \leq c_{ij}^{1/2}(N) c_{ij}^{\circ\,1/2}(N) \tag{175}$$

Therefore, from (165), (166), and Lemmas 10 and 11, we have for all N large

$$\rho(\mathbf{A}(N)) \leq \left(\frac{\widehat{K}^2\log^8 N - 1}{\widehat{K}^2\log^8 N}\right)^{\frac{1}{2}} \tag{176}$$

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

$$\mathbf{e}_N - \mathbf{e}_N^0 = (\mathbf{I} - \mathbf{A}(N))^{-1}\mathbf{w}^e \tag{177}$$

By (59) and (67) we have the entries of $\mathbf{A}(N)$ bounded by $K_1 \log^4 N$. Notice also, from (176)

$$|\det(\mathbf{I}-\mathbf{A}(N))| \geq (1-\rho(\mathbf{A}(N))^K \geq \left(\widehat{K}^2 \log^8 N \left(1+\frac{\widehat{K}^2 \log^8 N - 1}{\widehat{K}^2 \log^8 N}\right)^{\frac{1}{2}}\right)^{-K} \geq (2\widehat{K}^2 \log^8 N)^{-K} \tag{178}$$

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

$$\frac{(K-1)! K_1 (\log N)^{4(K-1)}}{|\det(\mathbf{I}-\mathbf{A}(N))|} \leq K_3 (\log N)^{12K-4} \tag{179}$$

Therefore, since $p \geq 12K-7$ ($> 8k_1+4$), (152) follows on this realization, an event which occurs with probability one.

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

$$m_N - m_N^\circ = \vec{\gamma}^\mathsf{T}(\mathbf{e}_N - \mathbf{e}_N^\circ) \tag{180}$$

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

$$\gamma_j = \int \frac{c_N \tau^2}{(1+c_N \tau e_{N,j})(1+c_n \tau e_{N,j}^\circ)} dF^{\mathbf{T}_N}(\tau) \frac{\operatorname{tr} \mathbf{D}^{-1}\mathbf{R}_j \mathbf{D}^{\circ -1}}{N} \tag{181}$$

From (59) and (67) we get each $|\gamma_j| \leq c_N |z|^2 v^{-4} \log^3 N$. Therefore from (152) and the fact that $w_N^m \to 0$, we have

$$m_N - m_N^\circ \to 0 \tag{182}$$

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

This completes the proof. ■

APPENDIX B
PROOF OF THEOREM 2

We first prove that $\mathcal{V}_N^\circ(x)$ as defined in Equation (24) verifies

$$\mathcal{V}_N^\circ(x) = \int_x^\infty \left(\frac{1}{w} - m_N^\circ(-w)\right) dw \tag{183}$$

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

$$\mathcal{V}_N^\circ(x) - \mathcal{V}_N(x) \xrightarrow{\text{a.s.}} 0 \tag{184}$$

A. *Proof of* (183)

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

$$e_i(z) = \frac{1}{N} \operatorname{tr} \mathbf{R}_i \left(-z\left[\mathbf{I}_N + \sum_{k=1}^K \delta_k \mathbf{R}_k\right]\right)^{-1} \tag{185}$$

$$\delta_i(z) = \frac{1}{n_i} \operatorname{tr} \mathbf{T}_i \left(-z\left[\mathbf{I}_{n_i} + c_i e_i(z) \mathbf{T}_i\right]\right)^{-1} \tag{186}$$

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

$$m_N^\circ(z) = \frac{1}{N}\operatorname{tr}\left(-z\left[\mathbf{I}_N + \sum_{k=1}^{K}\delta_k\mathbf{R}_k\right]\right)^{-1} \tag{187}$$

Now, notice that

$$\frac{1}{z} - m_N^\circ(-z) = \frac{1}{N}\left((z\mathbf{I})^{-1} - \left(z\left[\mathbf{I}_N + \sum_{k=1}^{K}\delta_k\mathbf{R}_k\right]\right)^{-1}\right) \tag{188}$$

$$= \sum_{k=1}^{K}\delta_k(-z)\cdot e_k(-z) \tag{189}$$

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

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

$$\frac{d}{dz}\frac{1}{N}\log\det\left(\mathbf{I}_{n_k} + c_k e_k(-z)\mathbf{T}_k\right) = -z\cdot e_k'(-z)\cdot\delta_k(-z) \tag{190}$$

$$\frac{d}{dz}\left(z\sum_{k=1}^{K}\delta_k(-z)e_k(-z)\right) = \sum_{k=1}^{K}\delta_k(-z)e_k(-z) - z\sum_{k=1}^{K}\delta_k'(-z)\cdot e_k(-z) + \delta_k(-z)\cdot e_k'(-z) \tag{191}$$

Combining the last three lines, we have

$$\sum_{k=1}^{K}\delta_k(-z)e_k(-z) =$$
$$\frac{d}{dz}\left[-\frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k=1}^{K}\delta_k(-z)\mathbf{R}_k\right) - \sum_{k=1}^{K}\frac{1}{N}\log\det\left(\mathbf{I}_{n_k} + c_k e_k(-z)\mathbf{T}_k\right) + z\sum_{k=1}^{K}\delta_k(-z)e_k(-z)\right] \tag{192}$$

which after integration leads to

$$\int_z^{+\infty}\left(\frac{1}{w} - m_N^\circ(-w)\right)dw =$$
$$\frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k=1}^{K}\delta_k(-z)\mathbf{R}_k\right) + \sum_{k=1}^{K}\frac{1}{N}\log\det\left(\mathbf{I}_{n_k} + c_k e_k(-z)\mathbf{T}_k\right) - z\sum_{k=1}^{K}\delta_k(-z)e_k(-z) \tag{193}$$

which is exactly the right-hand side of (24).

B. *Proof of* (184)

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

$$\max_{i\leq K}\max(\lambda_{r_N+1}^{\mathbf{T}_i}, \lambda_{r_N+1}^{\mathbf{R}_i}) \leq \alpha \tag{194}$$

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

$$\lambda_{2r_N+1}^{\mathbf{R}_i^{\frac{1}{2}}\mathbf{X}_i\mathbf{T}_i\mathbf{X}_i^{\mathsf{H}}\mathbf{R}_i^{\frac{1}{2}}} = (s_{2r_N+1}^{\mathbf{R}_i^{\frac{1}{2}}\mathbf{X}_i\mathbf{T}_i^{\frac{1}{2}}})^2 \tag{195}$$

$$\leq \alpha^2\|\mathbf{X}_i\mathbf{X}_i^{\mathsf{H}}\| \tag{196}$$

and then we have, from Lemma 15,

$$\lambda_{2Kr_N+1}^{\mathbf{B}_N} \le \alpha^2(\|\mathbf{X}_1\mathbf{X}_1^{\mathsf{H}}\| + \cdots + \|\mathbf{X}_K\mathbf{X}_K^{\mathsf{H}}\|) \tag{197}$$

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. [34]),

$$\begin{aligned}\limsup_N \lambda_{2Kr_N+1}^{\mathbf{B}_N} &\le \limsup_N \alpha^2(\|\mathbf{X}_1'\mathbf{X}_1'^{\mathsf{H}}\| + \cdots + \|\mathbf{X}_K'\mathbf{X}_K'^{\mathsf{H}}\|) \\ &\le \alpha^2 K(b/a)(1+\sqrt{a})^2\end{aligned} \tag{198}$$

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

Since $\mathbf{S} = 0$ here, it follows as in [6] that $\{F^{\mathbf{B}_n}\}$ is almost surely tight. Let F_N° denote the distribution function having Stieltjes transform m_N°, and let f on $[0, \infty)$ be a continuous function. Then the function

$$f_{a^\circ}(x) = \begin{cases} f(x) & , \ x \le a^\circ \\ f(a^\circ) & , \ x > a^\circ \end{cases} \tag{199}$$

is bounded and continuous. Therefore, with probability 1,

$$\int f_{a^\circ}(x) dF^{\mathbf{B}_N}(x) - \int f_{a^\circ}(x) dF_N^\circ(x) \to 0 \tag{200}$$

as $N \to \infty$.

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

$$\frac{1}{N}\sum_{\lambda_i \le a^\circ} f(\lambda_i) - \int_{[0,a^\circ]} f(x) dF_N^\circ(x) \to 0 \tag{201}$$

as $N \to \infty$.

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

$$\|\mathbf{B}_n\| \le b_N^2(\|\mathbf{X}_1'\mathbf{X}_1'^{\mathsf{H}}\| + \cdots + \|\mathbf{X}_K'\mathbf{X}_K'^{\mathsf{H}}\|) \tag{202}$$

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

$$\int f(x) dF^{\mathbf{B}_n}(x) - \frac{1}{N}\sum_{\lambda_i \le a^\circ} f(\lambda_i) \to 0 \tag{203}$$

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

$$\int f(x) dF^{\mathbf{B}_N}(x) - \int_{[0,a^\circ]} f(x) dF_N^\circ(x) \to 0 \tag{204}$$

as $N \to \infty$.

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

By considering the bound on $\|\mathbf{B}_{n,j}\|$ analogous to (202), we must have $F_N^\circ(a_N) = 1$ for all N large.

Similar to (198) we see that, with probability 1

$$\limsup_j \lambda_{2Kjr_N+1}^{\mathbf{B}_{N,j}} \leq a^2((1+\sqrt{c_1})^2 + \cdots + (1+\sqrt{c_K})^2) \tag{205}$$

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

At this point we will use the fact that for probability measures P_N, P on $\mathbb{R}$ with P_N converging weakly to P, we have (see e.g. [36])

$$\liminf_N P_N(G) \geq P(G) \tag{206}$$

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

$$F_N^\circ((a^\circ, \infty)) = 1 - F_N^\circ(a^\circ) \leq \liminf_j F^{\mathbf{B}_{N,j}}((a^\circ, \infty)) \tag{207}$$

$$\leq 2Kr_N/N \tag{208}$$

Therefore, for all N large

$$\int_{(a^\circ,\infty)} f(x) dF_N^\circ(x) \leq f(a_N) 2Kr_N/N \to 0 \tag{209}$$

as $N \to \infty$.

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

$$\int f(x)dF^{\mathbf{B}_N}(x) - \int f(x)dF_N^\circ(x) \to 0 \tag{210}$$

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

APPENDIX C

PROOF OF PROPOSITION 2

The proof stems from the following result,

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

$$\phi(\lambda) = f\left(\lambda \mathbf{P}_{1_a} + (1-\lambda)\mathbf{P}_{1_b}, \ldots, \lambda \mathbf{P}_{K_a} + (1-\lambda)\mathbf{P}_{K_b}\right) \tag{211}$$

is strictly concave.

Let us use a similar notation as in (217) of the capacity,

$$\bar{I}(\lambda) = I(\lambda \mathbf{P}_{1_a} + (1-\lambda)\mathbf{P}_{1_b}, \ldots, \lambda \mathbf{P}_{|\mathcal{S}|_a} + (1-\lambda)\mathbf{P}_{|\mathcal{S}|_b}) \tag{212}$$

and consider a set $(\delta_k, e_k, \mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|})$ which satisfies the system of equations (217)-(219). Then, from remark (222) and (223),

$$\frac{dI}{d\lambda} = \sum_{k\in\mathcal{S}} \frac{\partial \bar{V}}{\partial \delta_k}\frac{\partial \delta_k}{\partial \lambda} + \frac{\partial \bar{V}}{\partial e_k}\frac{\partial e_k}{\partial \lambda} + \frac{\partial \bar{V}}{\partial \lambda} \tag{213}$$

$$= \frac{\partial \bar{V}}{\partial \lambda} \tag{214}$$

where

$$\bar{V} : (\delta_1, \ldots, \delta_{|\mathcal{S}|}, e_1, \ldots, e_{|\mathcal{S}|}, \lambda) \mapsto \bar{I}(\lambda) \tag{215}$$

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

$$\frac{\partial^2 \bar{V}}{\partial \lambda^2} = -\sum_{i\in\mathcal{S}} (c_i^2 e_i^2) \frac{1}{N} \operatorname{tr}\left(\mathbf{I} + c_i e_i \mathbf{R}_i \mathbf{P}_i\right)^{-2} \left(\mathbf{R}_i(\mathbf{P}_{i_a} - \mathbf{P}_{i_b})\right)^2 \tag{216}$$

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 4, the deterministic approximate on the right-hand side of (29) is strictly concave in $\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}$ if any of the $\mathbf{R}_i$ matrices is invertible.

APPENDIX D
PROOF OF PROPOSITION 3

The proof of Proposition 3 recalls the proof from [26], Proposition 2. 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 (29) along any $\mathbf{Q}_k$ is the same whether the $\delta_k^\star$ and the $e_k^\star$ are fixed or vary with $\mathbf{Q}_k$. In other words, using the form (193) for the capacity, let us define the functions

$$\begin{aligned} \mathcal{V}^\circ(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}) = \sum_{k\in\mathcal{S}} \frac{1}{N} \log\det\left(\mathbf{I}_{n_k} + c_k e_k \mathbf{R}_k \mathbf{P}_k\right) \\ + \frac{1}{N}\log\det\left(\mathbf{I}_N + \sum_{k\in\mathcal{S}} \delta_k \mathbf{T}_k\right) \\ - \sigma^2 \sum_{k=1}^{K} \delta_k(-\sigma^2) e_k(-\sigma^2) \end{aligned} \tag{217}$$

where

$$e_i = e_i(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}) = \frac{1}{N} \operatorname{tr} \mathbf{T}_i \left(\sigma^2 \left[\mathbf{I}_N + \sum_{k\in\mathcal{S}} \delta_k \mathbf{T}_k\right]\right)^{-1} \tag{218}$$

$$\delta_i = \delta_i(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}) = \frac{1}{n_i} \operatorname{tr} \mathbf{R}_i \mathbf{P}_i \left(\sigma^2\left[\mathbf{I}_{n_i} + c_i e_i(z) \mathbf{R}_i \mathbf{P}_i\right]\right)^{-1} \tag{219}$$

and $V : (\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}, \delta_1, \ldots, \delta_{|\mathcal{S}|}, e_1, \ldots, e_{|\mathcal{S}|}) \mapsto \mathcal{V}^\circ(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|})$. Then we need only prove that, for all $k \in \mathcal{S}$,

$$\frac{\partial V}{\partial \delta_k}(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}, \delta_1^\star, \ldots, \delta_{|\mathcal{S}|}^\star, e_1^\star, \ldots, e_{|\mathcal{S}|}^\star) = 0 \tag{220}$$

$$\frac{\partial V}{\partial e_k}(\mathbf{P}_1, \ldots, \mathbf{P}_{|\mathcal{S}|}, \delta_1^\star, \ldots, \delta_{|\mathcal{S}|}^\star, e_1^\star, \ldots, e_{|\mathcal{S}|}^\star) = 0 \tag{221}$$

Remark then that

$$\frac{\partial V}{\partial \delta_k}(\mathbf{P}_1,\dots,\mathbf{P}_{|\mathcal{S}|},\delta_1,\dots,\delta_{|\mathcal{S}|},e_1,\dots,e_{|\mathcal{S}|}) = \frac{1}{N}\operatorname{tr}\left[\left(\mathbf{I}+\sum_{i\in\mathcal{S}}\delta_i\mathbf{T}_i\right)^{-1}\mathbf{T}_k\right]-\sigma^2 e_k \tag{222}$$

$$\frac{\partial V}{\partial e_k}(\mathbf{P}_1,\dots,\mathbf{P}_{|\mathcal{S}|},\delta_1,\dots,\delta_{|\mathcal{S}|},e_1,\dots,e_{|\mathcal{S}|}) = c_k\frac{1}{N}\operatorname{tr}\left[(\mathbf{I}+c_k e_k\mathbf{R}_i\mathbf{P}_i)^{-1}\mathbf{R}_k\mathbf{P}_k\right]-\sigma^2\delta_k \tag{223}$$

both being null whenever, for all k, $e_k = e_k(-\sigma^2,\mathbf{P}_1,\dots,\mathbf{P}_{|\mathcal{S}|})$ and $\delta_k=\delta_k(-\sigma^2,\mathbf{P}_1,\dots,\mathbf{P}_{|\mathcal{S}|})$, which is true in particular for the unique power optimal solution $\mathbf{P}_1^\star,\dots,\mathbf{P}_{|\mathcal{S}|}^\star$ 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 $\mathbf{P}_k$ 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 $\mathbf{P}_k$. From the inequality, (see e.g. [2])

$$\det(\mathbf{I}_{n_k}+c_k e_k^\star\mathbf{R}_k\mathbf{P}_k)\leq\prod_{i=1}^{n_k}(\mathbf{I}_{n_k}+c_k e_k^\star\mathbf{R}_k\mathbf{P}_k)_{ii}) \tag{224}$$

where, only here, we denote $(\mathbf{X})_{ii}$ the entry (i,i) of matrix $\mathbf{X}$. The equality is obtained if and only if $\mathbf{I}_{n_k}+c_k e_k^\star\mathbf{R}_k\mathbf{P}_k$ 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 $\frac{1}{n_k}\operatorname{tr}(\mathbf{Q}_k)=P_k$ are solutions of the classical optimization problem under constraint,

$$\sup_{\substack{\mathbf{Q}_k\\ \frac{1}{n_k}\operatorname{tr}(\mathbf{Q}_k)\leq P_k}}\quad\log\det\left(\mathbf{I}_{n_k}+c_k e_k^\star\mathbf{Q}_k\mathbf{D}_k\right) \tag{225}$$

whose solution is given by the classical water-filling algorithm. Hence (35).

Appendix E
Proof of Proposition 1

The convergence of the fixed-point algorithm follows the same line of proof as the uniqueness in Section A-E. Instead of proving the convergence of the algorithm at $z=-\sigma^2$, we start by proving the convergence for $z\in\mathbb{C}^+$. If one considers the difference $\mathbf{e}^{n+1}-\mathbf{e}^n$, where $\mathbf{e}^n=(e_1^n,\dots,e_K^n)$, instead of $\mathbf{e}^\circ-\underline{\mathbf{e}}^\circ$, the same development as in Section A-E leads to

$$\mathbf{e}^{n+1}-\mathbf{e}^n=\mathbf{A}_n(\mathbf{e}^n-\mathbf{e}^{n-1}) \tag{226}$$

for $n\geq 1$, where $\mathbf{A}_n$ is defined, similarly as in (147), as $\mathbf{A}_n=(a_{ij}^n)$, with a_{ij}^n defined by

$$a_{ij}^n=\frac{1}{N}\operatorname{tr}\mathbf{R}_i{\mathbf{D}_{n-1}}^{-1}\mathbf{R}_j{\mathbf{D}_n}^{-1}c_j\int\frac{\tau^2}{(1+c_j\tau e_j^{n-1})(1+c_j\tau e_j^n)}dF^{\mathbf{T}_j}(\tau). \tag{227}$$

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

$$a_{ij}\leq\frac{|z|^2c_j}{v^4}\frac{\log N^4}{N} \tag{228}$$

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

$$\max_j\left(e_j^{n+1}-e_j^n\right)<K\frac{|z|^2c_0}{v^4}\leq\frac{\log N^4}{N}\max_j\left(e_j^n-e_j^{n-1}\right). \tag{229}$$

Let $0 < \varepsilon < 1$, and take now a countable set $\{z_1, z_2, \ldots\}$, $v_k = \Im[z_k]$, such that $K\frac{|z_k|^2 c_0}{v_k^4}\frac{\log N^4}{N} < 1 - \varepsilon$ for all z_k (this is possible by letting $v_k > 0$ be large enough). On this countable set, the sequences $\{\mathbf{e}^n\}$ are therefore Cauchy sequences on $\mathbb{C}^K$: they all converge. Since the e_j^n are holomorphic and bounded on every compact set included in $\mathbb{C} \setminus \mathbb{R}^+$, from Vitali's convergence theorem [15], the function $e_j^n(z)$ converges on such compact sets. Now, for $z = -\sigma^2$, from the fact that we forced the initialization step to be $e_j^0 = 1/\sigma^2$, e_j^0 is the Stieltjes transform of a distribution function at point $z = -\sigma^2$. It now suffices to verify that, if e_j^n is the Stieltjes transform of a distribution function at point z, then so is e_j^{n+1}. This requires to verify that $z \in \mathbb{C}^+,\ e_j^n \in \mathbb{C}^+$ implies $e_j^{n+1} \in \mathbb{C}^+$, $z \in \mathbb{C}^+,\ ze_j^n \in \mathbb{C}^+$ implies $ze_j^{n+1} \in \mathbb{C}^+$, and $\lim_{y\to\infty} -ye_n^j(iy) < \infty$ implies that $\lim_{y\to\infty} -ye_n^j(iy) < \infty$. This follows directly from the definition of e_j^n. From the dominated convergence theorem, we then also have that the limit of e_j^n is a Stieltjes transform that is solution to (12). From the uniqueness of the Stieltjes transform, solution to (12) (this follows from the pointwise uniqueness on $\mathbb{C}^+$ and the fact that the Stieltjes transform is holomorphic on all compact sets of $\mathbb{C} \setminus \mathbb{R}^+$), we then have that e_j^n converges for all j and $z \in \mathbb{C} \setminus \mathbb{R}^+$, if e_j^0 is initialized at a Stieltjes transform. The choice $z = -\sigma^2$, $e_j^0 = 1/\sigma^2$ follows this rule and the fixed-point algorithm converges to the correct solution.

APPENDIX F
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,

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

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 [6]) 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)

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

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 [6]

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

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

If $z < 0$, we also have

$$\left|\mathrm{tr}\left((\mathbf{B}-z\mathbf{I}_N)^{-1} - (\mathbf{B}+\mathbf{r}\mathbf{r}^{\mathsf{H}}-z\mathbf{I}_N)^{-1}\right)\mathbf{A}\right| \leq \frac{\|\mathbf{A}\|}{|z|} \tag{238}$$

From Lemma 2.2 of [16], and Theorems A.2, A.4, A.5 of [17], 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 [6], we need

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

$$\mathrm{E}|\mathbf{y}^{\mathsf{H}}\mathbf{A}\mathbf{y} - \mathrm{tr}\,\mathbf{A}|^6 \leq K\|\mathbf{A}\|^6 N^3 \log^{12} N \tag{239}$$

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 [29],

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

From Theorem 8.1.18 of [30],

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

$$\rho(\mathbf{A}) \leq \rho((|a_{ij}|)) \leq \rho(\mathbf{B}) \tag{240}$$

Also, from Lemma 5.7.9 of [31],

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

$$\rho((a_{ij}^{\frac{1}{2}} b_{ij}^{\frac{1}{2}})) \leq (\rho(\mathbf{A}))^{\frac{1}{2}} (\rho(\mathbf{B}))^{\frac{1}{2}} \tag{241}$$

And, Theorems 8.2.2 and 8.3.1 of [30],

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 [31], we also need Theorem 6.1.1,

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

Theorem 3.42 of [15],

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 [33]

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

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

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

$$s_{m+n+1}^{AB} \leq s_{m+1}^{A} s_{n+1}^{B} \tag{243}$$

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

$$s_{Kr+1}^{\mathbf{A}_1+\cdots+\mathbf{A}_K} \leq s_{r+1}^{\mathbf{A}_1} + \cdots + s_{r+1}^{\mathbf{A}_K} \tag{244}$$

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] 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.

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

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

[4] 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.

[5] 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.

[6] 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

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

[8] 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.

[9] 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.

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] E. C. Titchmarsh, “The Theory of Functions,” Oxford University Press, 1939.

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

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

[18] 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.

[19] 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.

[20] 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.

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

[22] 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.

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

[24] 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.

[25] 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.

[26] 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.

[27] W. Hachem, P. Loubaton and J. Najim, “A CLT for Information Theoretic Statistics of Gram Random Matrices with a Given Variance Profile,” Annals of Applied Probability, vol. 18, no. 6, pp. 2071-2130, 2008.

[28] R. Rashibi Far, T. Oraby, W. Bryc, R. Speicher, “On slow-fading MIMO systems with nonseparable correlation,” IEEE Transactions on Information Theory, vol. 54, no. 2, pp. 544-553, 2008.

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

[30] R.A. Horn and C.R. Johnson, “Matrix Analysis,” Cambridge University Press, 1985.

[31] R.A. Horn and C.R. Johnson, “Topics in Matrix Analysis,” Cambridge University Press, 1991.

[32] 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.

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

[34] Z. Bai and J. W. Silverstein, “Spectral Analysis of Large Dimensional Random Matrices,” 2nd Edition, Springer, 2009.

[35] P. Billingsley, “Probability and Measures,” 3rd Edition, Wiley Inter-Science, 1995.

[36] P. Billingsley, “Convergence of Probability Measures,” John Wiley & Sons, Second Revised Edition, 1999.

[37] 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.

[38] B. M. Hochwald, T. L. Marzetta, V. Tarokh, “Multi-antenna channel hardening and its implications for rate feedback and scheduling,” IEEE Trans. on Information Theory, vol. 50, no. 9, pp. 1893-1909, 2004.

[39] V. L. Girko, “Theory of Random Determinants,” Kluwer, Dordrecht, 1990.

[40] A. L. Moustakas, S. H. Simon, “On the outage capacity of correlated multiple-path MIMO channels,” IEEE Trans. on Information Theory, vol. 53, no. 11, pp. 3887, 2007.

[41] F. Dupuy, P. Loubaton, “On the Capacity Achieving Covariance Matrix for Frequency Selective MIMO Channels Using the Asymptotic Approach,” Arxiv preprint arXiv:1001.3102, 2010.

[42] L. Zhang, *Spectral Analysis of Large Dimensional Random Matrices*, PhD dissertation.

[43] W. Yu, W. Rhee, S. Boyd, J. M. Cioffi, “Iterative water-filling for Gaussian vector multiple-access channels,” IEEE Trans. on Information Theory, vol. 50, no. 1, pp. 145-152, 2004.

[44] M. Vu, A. Paulraj, “Capacity optimization for Rician correlated MIMO wireless channels,” Proceedings of Asilomar Conference, pp. 133-138, 2005.

[45] A. Goldsmith, S. A. Jafar, N. Jindal, S. Vishwanath, “Capacity limits of MIMO channels,” IEEE Journal on selected areas in Communications, vol. 21, no. 5, pp. 684-702, 2003.