On the Fluctuations of the Mutual Information of Large Dimensional MIMO Channels

Walid Hachem
CNRS / ENST (UMR 5141)
46 rue Barrault
75634 Paris Cedex 13
hachem@enst.fr

Philippe Loubaton
Université de Marne la Vallée
IGM LabInfo, UMR 8049
77454 Marne la Valle Cedex 2
loubaton@univ-mlv.fr

Jamal Najim
CNRS / ENST (UMR 5141)
46 rue Barrault,
75634 Paris Cedex 13
najim@enst.fr

***Abstract*— In this article, large random matrix theory is used to study Shannon's mutual information of a general class of Multiple Input Multiple Output radio channels with random correlated gains. In the literature, there exists an approximation of this mutual information in the asymptotic regime where the number of transmitting antennas and the number of receiving antennas grow toward infinity at the same pace. This contribution is devoted to the study of the mutual information's fluctuations around this deterministic approximation under the form of a Central Limit Theorem (CLT). In particular, this CLT provides an approximation of the Outage Probability, which represents a fundamental performance index for communications on slow fading channels. The proof of this CLT relies on martingale theory.**

I. INTRODUCTION

In the context of wireless communications in presence of a large number of reflectors and scatterers, spectral efficiency can be increased dramatically by the use of multiple antennas at the transmitter and at the receiver. In this situation, it is relevant to represent the elements of the Multiple Input Multiple Output (MIMO) channel as random variables. Assuming these random variables are Gaussian independent and identically distributed (i.i.d.) and assuming the total transmitted power is fixed, Telatar [15] and Foschini [5] realized indeed in the mid nineties that Shannon's capacity grows with the number of antennas at the rate of $\min(r,t)$ where r is the number of receiver antennas and t is the number of transmitter antennas. Let $\mathbf{H}$ be the $r \times t$ matrix that represents the MIMO channel considered in [15], and let $I(\rho) = \mathbb{E}\mathcal{I}(\rho)$ be Shannon's capacity of this channel per receiver antenna, where $\mathcal{I}(\rho)$ designates the random variable $\mathcal{I}(\rho) = \frac{1}{r}\log\det\left(\frac{1}{\rho t}\mathbf{H}\mathbf{H}^H + \mathbf{I}_r\right)$ and ρ is the variance of a sample of the additive white Gaussian noise. A classical result of Random Matrix Theory (RMT) says that when the elements of $\mathbf{H}$ are iid, the empirical eigenvalue distribution of $\mathbf{H}\mathbf{H}^H$ converges weakly to a deterministic law when $t \to \infty$ and r/t converges towards a constant $c > 0$ [11]. Considering the fact that the capacity per receiver antenna $I(\rho)$ is the integral of a $\log$ function with respect to the empirical eigenvalue distribution, it is then possible to establish the convergence of $I(\rho)$ towards a constant. This result makes clear the assertion related to the capacity increase rate of $\min(r,t)$.

The findings of [15] and [5] have been generalized afterwards to channel statistical models more realistic than the iid model, *i.e.*, models that take into account the correlations between the elements of $\mathbf{H}$ often observed in practice. Without being exhaustive, let us cite among these generalizations [7], [8], [4], [12], [16]. These contributions produce a (deterministic) approximation $V(\rho)$ (depending on (r,t)) of the mutual information in the sense that $I(\rho) - V(\rho) \to 0$ when $t \to \infty$ in such a way that $r/t \to c > 0$. Let us note that $V(\rho)$ has a closed form expression for a few statistical models only. In general, $V(\rho)$ is a function of the solution of a system of implicit equations, an example of which will be shown below. An essential question raised by the study of the mutual information by means of RMT consists in characterizing the fluctuations of the random variable $\mathcal{I}(\rho) - V(\rho)$. In order to answer this question, one seeks a Central Limit Theorem (CLT), *i.e.*, one looks for a result of the form

$$r^{\alpha}\frac{\mathcal{I}(\rho) - \mathbb{E}\mathcal{I}(\rho)}{\Theta_t} \to \mathcal{N}(0,1) \tag{1}$$

in distribution where $\mathcal{N}(0,1)$ is the standard Gaussian law, and

$$r^{\beta}(\mathbb{E}\mathcal{I}(\rho) - V(\rho)) - B_t \to 0 \tag{2}$$

where Θ_t^2 is a variance term, B_t is a bias term, and the coefficients α and β characterize convergence speeds. Apart from the importance of this result in assessing the validity of the approximation $V(\rho)$ for a finite number of antennas, it provides an approximation of the so called Outage Probability $\mathbb{P}\left[r\mathcal{I}(\rho) \leq R\right]$ where R is a given data rate. Recall that the pertinent performance index is the outage probability instead of the mutual information $I = \mathbb{E}\mathcal{I}$ when the channel fading is slow.

In this article, we consider a channel model where $\mathcal{I}(\rho)$ is invariant when matrix $\mathbf{H}$ with correlated elements is replaced with a matrix $\mathbf{Y}$ with independent centered Gaussian elements having in general different variances. This model includes most MIMO models that can be found in the wireless communications literature in the centered case. When applied to our channel model, the existing mathematical results (let us cite [1], [2]) answer the problem of the fluctuations of $\mathcal{I}(\rho) - V(\rho)$ in some particular cases only. In this sense, the results shown here represent a new contribution to RMT independently of

their applications to wireless communications.
In Section II, after introducing the channel model, we shall characterize the approximation $V(\rho)$ of $\mathbb{E}\mathcal{I}(\rho)$ that we shall refer to as the first order approximation of the mutual information. We shall then study the fluctuations of $\mathcal{I}(\rho) - V(\rho)$ (Section III) and specify the content of expressions (1) and (2). This CLT will be valid in the asymptotic regime where $t \to \infty$ in such a way that $\liminf r/t > 0$ and $\limsup r/t < \infty$. This asymptotic regime, which is more general than the usual $r/t \to c > 0$, will be referred to in the sequel as "$t \to \infty$". The general principle of the proof of this CLT is described in Section IV. The detailed proof of this result can be found in [9] where the Gaussian assumption on the elements of $\mathbf{Y}$ is furthermore relaxed.

II. Channel model and expression of $V(\rho)$

Consider a communication channel with t antennas at the transmitter and r antennas at the receiver; let $\mathbf{H}$ be the $r \times t$ matrix representing the complex gains between the emitting and the receiving antennas. At a given moment, the received vector $\mathbf{y}$ of dimension r is described by the equation:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$$

where $\mathbf{z}$ is a random complex Gaussian vector with covariance matrix given by $\rho\mathbf{I}$ and $\mathbf{x}$ represents the transmitted vector of dimension t. Assuming that $\mathbf{x}$ is a Gaussian centered vector with covariance matrix given by $\mathbb{E}\mathbf{x}\mathbf{x}^H = \frac{1}{t}\mathbf{I}_t$, Shannon's mutual information of the channel per receiver antenna is given by $I(\rho) = \mathbb{E}\mathcal{I}(\rho)$ where

$$\mathcal{I}(\rho) = \frac{1}{r}\log\det\left(\frac{1}{\rho t}\mathbf{H}\mathbf{H}^H + \mathbf{I}_r\right) . \tag{3}$$

It is of great interest to evaluate $I(\rho)$ for various statistical models for the channel $\mathbf{H}$. These models take into account the correlations between the gains of the channel, mainly due the proximity of the emitting or the receiving antennas.

In this work, we shall consider the situation where $\mathbf{H}\mathbf{H}^H$ is unitarily equivalent to a Gram matrix $\mathbf{Y}\mathbf{Y}^H$ where $\mathbf{Y}$ is a random $r \times t$ matrix which (k,ℓ)-element is given by $Y_{k\ell} = \sigma_{k\ell}X_{k\ell}$, the random variables $X_{k\ell}$ are independent standard complex Gaussian circular and the $\sigma_{k\ell}$'s are real numbers. Here the family $(\sigma^2_{k\ell})_{k,\ell=1}^{r,t}$ is called a variance profile due to the fact that $\mathbb{E}|Y_{k\ell}|^2 = \sigma^2_{k\ell}$.
Thanks to this unitary equivalence, $\mathcal{I}(\rho)$ is also given by

$$\mathcal{I}(\rho) = \frac{1}{r}\log\det\left(\frac{1}{\rho t}\mathbf{Y}\mathbf{Y}^H + \mathbf{I}_r\right) . \tag{4}$$

The most popular centered channel models considered in the literature satisfy this property. Let us cite

- The model introduced by Sayeed in [13], according to which the channel matrix is written as $\mathbf{H} = \mathbf{F}_r\mathbf{Y}\mathbf{F}_t$ where $\mathbf{Y}$ is the matrix described above, and for a given integer n, $\mathbf{F}_n$ is the $n \times n$ Fourier matrix, *i.e.*, the matrix which (k,ℓ)-element is $\exp(2\imath\pi(k-1)(\ell-1)/n)/\sqrt{n}$ for $1 \le k,\ell \le n$. Otherwise stated, $\mathbf{H}$ is the 2-D Fourier transform of a matrix with independent elements with variances depending on the position of the element in the matrix. According to this model, the columns of matrix $\mathbf{F}_r$ (resp. $\mathbf{F}_t = \mathbf{F}_t^T$) can be considered as direction of arrival (resp. departure) elementary vectors. The element $Y_{k\ell}$ of matrix $\mathbf{Y}$ represents the path strength seen at the angles of arrival and departure represented by respectively column k of $\mathbf{F}_r$ and column ℓ of $\mathbf{F}_t$.
As Fourier matrices are unitary *i.e.* $\mathbf{F}_n\mathbf{F}_n^H = \mathbf{I}_n$, it is clear that $\log\det\left(\frac{1}{\rho t}\mathbf{H}\mathbf{H}^H + \mathbf{I}_r\right) = \log\det\left(\frac{1}{\rho t}\mathbf{Y}\mathbf{Y}^H + \mathbf{I}_r\right)$.
- The so-called Kronecker model $\mathbf{H} = \mathbf{\Psi}_{\mathrm{R}}\mathbf{W}\mathbf{\Psi}_{\mathrm{T}}$ where $\mathbf{W}$ is a $r \times t$ matrix with Gaussian centered i.i.d. entries, and $\mathbf{\Psi}_{\mathrm{R}}$ and $\mathbf{\Psi}_{\mathrm{T}}$ are $r \times r$ and $t \times t$ matrices that capture the path correlations at the receiver and at the transmitter sides respectively (see for instance [14]). By resorting to a spectral decomposition of the matrices $\mathbf{\Psi}_{\mathrm{R}}$ and $\mathbf{\Psi}_{\mathrm{T}}$, one can easily see that (3) can be replaced with (4) with $\sigma_{k\ell} = \sigma_{\mathrm{R},k}\sigma_{\mathrm{T},\ell}$ where $(\sigma_{\mathrm{R},k})_{1\le k\le r}$ and $(\sigma_{\mathrm{T},\ell})_{1\le \ell\le t}$ are the singular values of $\mathbf{\Psi}_{\mathrm{R}}$ and $\mathbf{\Psi}_{\mathrm{T}}$ respectively. With respect to mutual information calculations, the Kronecker model shows to be a particular case (often termed "the separable variance model") of Sayeed's model.

In short, our problem consists in evaluating $I(\rho) = \mathbb{E}\mathcal{I}(\rho)$ where $\mathcal{I}(\rho)$ is given by (4). A direct computation of $I(\rho)$ relies on massive Monte-Carlo simulations. In order to circumvent this difficulty, one is lead to search a more handy approximation of $I(\rho)$. Such an approximation $V(\rho)$ follows from the RMT and satisfies:

$$I(\rho) - V(\rho) \to 0 \text{ as } t \to \infty .$$

In the sequel, we shall need the following technical assumption:

A1 : There exists a nonnegative real number $\sigma_{\max}$ such that:

$$\sup_{t\ge 1}\max_{\substack{1\le k\le r\\ 1\le \ell\le t}} \sigma^2_{k\ell} < \sigma^2_{\max} .$$

We shall say that a complex function $g(z)$ belongs to the class $\mathcal{S}$ if $g(z)$ is analytic over the set $\mathbb{C}_+ = \{z \in \mathbb{C};\, \mathrm{im}(z) > 0\}$, if $g(z) \in \mathbb{C}_+$ for every $z \in \mathbb{C}_+$ and if $\mathrm{im}(z)|g(z)|$ is bounded in $\mathbb{C}_+$. Before providing the expression $V(\rho)$, we need the following result [6], [8]:

Proposition 1: Assume that **A1** holds true. Then the system of r functional equations

$$g_k(z) = \frac{1}{-z + \frac{1}{t}\sum_{j=1}^{t}\frac{\sigma^2_{kj}}{1+\frac{1}{t}\sum_{\ell=1}^{r}\sigma^2_{\ell j}g_\ell(z)}}, \quad 1 \le k \le r \tag{5}$$

admits a unique solution $(g_1(z), \cdots, g_r(z))$ satisfying $g_k(z) \in \mathcal{S}$. The functions $g_k(z)$ can be analytically extended over the set $\mathbb{C} - [0,\infty)$.
The deterministic approximation is given by the following theorem:

Theorem 1: Assume that **A1** holds true. Let

$$\begin{aligned} V(\rho) = & -\frac{1}{r}\sum_{i=1}^{r}\log(\rho g_i(-\rho)) \\ & + \frac{1}{r}\sum_{j=1}^{t}\log\left(1+\frac{1}{t}\sum_{\ell=1}^{r}\sigma_{\ell j}^2 g_\ell(-\rho)\right) \\ & - \frac{1}{rt}\sum_{i=1:r,j=1:t}\frac{\sigma_{ij}^2 g_i(-\rho)}{1+\frac{1}{t}\sum_{\ell=1}^{r}\sigma_{\ell j}^2 g_\ell(-\rho)} \end{aligned}$$

where functions $g_i(z)$ are defined in Proposition 1. Then

$$I(\rho) - V(\rho) \to 0 \quad \text{as} \quad t \to \infty$$

for every $\rho > 0$.
This result can be found in [8] and is based [6] on the study of the resolvent of matrix $t^{-1}\mathbf{Y}\mathbf{Y}^H$.

III. Fluctuations of $\mathcal{I}(\rho) - V(\rho)$

The quantity $rI(\rho) = r\mathbb{E}\mathcal{I}(\rho)$, sometimes called "ergodic mutual information" represents the maximum of the rate that the channel $\mathbf{H}$ can theoretically provide in the case where the length of the codeword is much larger than the coherence time of this channel. On the other extreme, in the case where the channel is invariant at the time scale of a codeword, the relevant performance indicator is no longer $rI(\rho)$ but rather the outage probability $\mathbb{P}\left[r\mathcal{I}(\rho) < R\right]$, where R is the target data rate.

The theory of random matrices have given us so far a deterministic approximation $V(\rho)$ of $I(\rho)$. In order to obtain an approximation of the outage probability, we shall establish a CLT on the random quantity $r(\mathcal{I}(\rho) - V(\rho))$ as $t \to \infty$. Indeed, Theorem 2 below says that $r(\mathcal{I}(\rho) - V(\rho))$ converges towards a centered Gaussian random variable which variance will immediatly yield an approximation of $\mathbb{P}\left(r\mathcal{I}(\rho) < R\right)$. Write

$$X_t = r(\mathcal{I}(\rho) - V(\rho)) = Z_t + b_t \text{ where } \begin{cases} Z_t = r(\mathcal{I}(\rho) - I(\rho)) \\ b_t = r(I(\rho) - V(\rho)) \end{cases}$$

Notice that Z_t is random and accounts for the fluctuations of $r\mathcal{I}(\rho)$ around its expectation while b_t is deterministic and represents a bias. In order to state the CLT, we first introduce the following slight assumption:

$$\mathbf{A2}: \quad \max\left(\liminf_{t\geq 1}\min_{1\leq \ell\leq t}\frac{1}{t}\sum_{k=1}^{r}\sigma_{k\ell}^2\ , \right. \\ \left. \liminf_{t\geq 1}\min_{1\leq k\leq r}\frac{1}{t}\sum_{\ell=1}^{t}\sigma_{k\ell}^2\right) > 0\ .$$

We are now in position to state the main contribution of the paper:

Theorem 2: Let $\mathbf{Y} = [Y_{k\ell}]$ be a $r \times t$ matrix where $Y_{k\ell} = \sigma_{k\ell}X_{k\ell}$, the random variables $\{X_{k\ell}\}_{k,l=1}^{r,t}$ being independent with distribution $\mathcal{CN}(0,1)$. Let $\mathbf{G}$ be the $r\times r$ diagonal matrix defined by:

$$\mathbf{G} = \mathrm{diag}\left(g_k;\ 1 \leq k \leq r\right)$$

where functions g_k are defined in Proposition (1); define by $\mathbf{C}_\ell$ the $r \times r$ diagonal matrices:

$$\mathbf{C}_\ell = \mathrm{diag}\left(\sigma_{k\ell}^2;\ 1 \leq k \leq r\right) \text{ for } 1 \leq \ell \leq t\ .$$

Let $\mathbf{A}$ be the $t \times t$ matrix defined by

$$\mathbf{A} = \left[\frac{1}{t}\frac{\frac{1}{t}\mathrm{tr}(\mathbf{C}_l\mathbf{C}_m\mathbf{G}^2)}{\left(1+\frac{1}{t}\mathrm{tr}\left(\mathbf{C}_l\mathbf{G}\right)\right)^2}\right]_{l,m=1}^{t}\ .$$

Assume that **A1** and **A2** are fulfilled, then the following results hold true:

1) The real number $\Theta_t > 0$ defined by $\Theta_t^2 = -\log\det\left(\mathbf{I}_t - \mathbf{A}\right)$ is well-defined and satisfies:
$$0 < \liminf_t \Theta_t^2 \leq \limsup_t \Theta_t^2 < \infty\ .$$
2) The sequence of random variables $Z_t = r(\mathcal{I}(\rho) - \mathbb{E}\mathcal{I}(\rho))$ satisfies:
$$\frac{Z_t}{\Theta_t} \xrightarrow[t\to\infty]{} \mathcal{N}(0,1) \quad \text{in distribution}\ .$$
3) The bias $b_t = r(\mathbb{E}\mathcal{I}(\rho) - V(\rho))$ where $V(\rho)$ is given in Theorem 1 satisfies:
$$b_t \xrightarrow[t\to\infty]{} 0\ .$$

Loosely speaking, this theorem states that the random variable $r(\mathcal{I}(\rho) - V(\rho))$ behaves as a Gaussian random variable $\mathcal{N}(0,\Theta_t)$ in the large dimension regime ($t \to \infty, \liminf r/t > 0, \limsup r/t < \infty$).

Remarks:

- As one can notice in Theorem 2, the assumptions related to the variance profile are very light. There exist in the literature results related to the fluctuations of $r(\mathcal{I}(\rho) - V(\rho))$ in the Kronecker (or separable) case where $\sigma_{k\ell}^2 = \sigma_{\mathrm{R},k}^2\sigma_{\mathrm{T},\ell}^2$. See for instance [12] and [7]. In the separable case, the system of r equations (5) shrinks to a system of two equations and the variance writes $\Theta_t^2 = -\log(1 - \xi(\rho))$ where $\xi(\rho)$ is a scalar quantity easy to compute.
- Theorem 2 can be generalized to the case where the elements of $\mathbf{Y}$ are not necessarily Gaussian [9]. In this case, one still has $\alpha = 1$ in (1) however a bias appears ($b_t \not\to 0$ in general) together with an extra term in the variance Θ_t^2, both proportional to the fourth cumulant $\kappa = \mathbb{E}|X_{k\ell}|^4 - 2$ of the $X_{k\ell}$'s. We do not provide details here as this case is not very relevant in the context of wireless communications we are interested in in the present contribution.
- In the case elements of $\mathbf{Y}$ are Gaussian, by exploiting the mathematical tools used in [7], it is possible to prove that the bias term b_t behaves as $b_t = \mathcal{O}(1/t)$, *i.e.*, $\beta = 2$ and $B_t = \mathcal{O}(1)$ in (2). Hence, the approximation $V(\rho)$ of the ergodic mutual information remains relevant even for a small number of antennas. Simulations can be found in [10].

IV. OUTLINE OF THE PROOF OF THEOREM 2

In this section, we outline the proof of the second item of Theorem 2. Let $Z_t = r(\mathcal{I}(\rho) - \mathbb{E}\mathcal{I}(\rho))$, denote by $\mathbf{y}_\ell$ the ℓth column of matrix $\mathbf{Y}$ and by $\mathbb{E}_\ell$ the conditional expectation $\mathbb{E}_\ell[\cdot] = \mathbb{E}[\cdot \| \mathbf{y}_\ell, \ldots, \mathbf{y}_t]$. We can write

$$Z_t = \sum_{\ell=1}^{t} (\mathbb{E}_\ell - \mathbb{E}_{\ell+1}) \log\det\left(\frac{1}{\rho t}\mathbf{Y}\mathbf{Y}^H + \mathbf{I}_r\right) \stackrel{d}{=} \sum_{\ell=1}^{t} W_\ell$$

where $\mathbb{E}_{t+1} = \mathbb{E}$. It is straightforward to check that the random variables $(W_t, W_{t-1}, \ldots, W_1)$ is a sequence of increments of martingales with respect to the σ-algebra $\sigma(\mathbf{y}_t), \ldots, \sigma(\mathbf{y}_t, \ldots, \mathbf{y}_1)$ (*cf.* [3]). A CLT for $Z_t = \sum_\ell W_\ell$ can be established with the help of general results of CLTs for martingales (see for instance [3, Ch. 35]). We begin by working out the expressions of the W_ℓ's. Denote by $\mathbf{Y}_\ell$ the matrix obtained from $\mathbf{Y}$ after deleting column $\mathbf{y}_l$. We have:

$$(\mathbb{E}_\ell - \mathbb{E}_{\ell+1})\left[\log\det\left(\frac{1}{\rho t}\mathbf{Y}_\ell^H\mathbf{Y}_\ell + \mathbf{I}_{t-1}\right)\right] = 0$$

Consequently, as $\det(\mathbf{B}\mathbf{B}^H + \mathbf{I}) = \det(\mathbf{B}^H\mathbf{B} + \mathbf{I})$, W_ℓ writes

$$\begin{aligned} W_\ell &= (\mathbb{E}_\ell - \mathbb{E}_{\ell+1}) \log\left(\frac{\det\left(\frac{1}{\rho t}\mathbf{Y}^H\mathbf{Y} + \mathbf{I}_t\right)}{\det\left(\frac{1}{\rho t}\mathbf{Y}_\ell^H\mathbf{Y}_\ell + \mathbf{I}_{t-1}\right)}\right) \\ &\stackrel{d}{=} (\mathbb{E}_\ell - \mathbb{E}_{\ell+1}) \log\left(\frac{\det \mathbf{\Xi}}{\det \mathbf{\Xi}_\ell}\right) . \end{aligned}$$

Recall that $\det\begin{bmatrix} a & \mathbf{b}^H \\ \mathbf{b} & \mathbf{B} \end{bmatrix} = (a - \mathbf{b}^H\mathbf{B}^{-1}\mathbf{b})\det\mathbf{B}$, thus

$$\begin{aligned} \det\mathbf{\Xi} = (\det\mathbf{\Xi}_\ell)\Bigg(&\frac{\|\mathbf{y}_\ell\|^2}{\rho t} + 1 \\ &- \frac{\mathbf{y}_\ell^H\mathbf{Y}_\ell}{\rho t}\left(\frac{1}{\rho t}\mathbf{Y}_\ell^H\mathbf{Y}_\ell + \mathbf{I}\right)^{-1}\frac{\mathbf{Y}_\ell^H\mathbf{y}_\ell}{\rho t}\Bigg) . \end{aligned}$$

Using the relation $\mathbf{I} - \mathbf{B}\left(\mathbf{B}^H\mathbf{B} + \mathbf{I}\right)^{-1}\mathbf{B}^H = (\mathbf{B}\mathbf{B}^H + \mathbf{I})^{-1}$, we get:

$$W_\ell = (\mathbb{E}_\ell - \mathbb{E}_{\ell+1})\log\left(1 + \frac{1}{t}\mathbf{y}_\ell^H\mathbf{Q}_\ell\mathbf{y}_\ell\right) \tag{6}$$

where $\mathbf{Q}_\ell$ is the resolvent matrix $\mathbf{Q}_\ell = \left(\frac{1}{t}\mathbf{Y}_\ell\mathbf{Y}_\ell^H + \rho\mathbf{I}_r\right)^{-1}$. A fundamental result that goes back to [11] (see also [2], [6], [8], [9]) states that if $\mathbf{x}$ is a random $r \times 1$ vector whose elements are i.i.d. with variance 1, and if $\mathbf{B}$ is a hermitian $r \times r$ independent of $\mathbf{x}$ with bounded specral norm, then $\frac{1}{t}(\mathbf{x}^H\mathbf{B}\mathbf{x} - \mathrm{tr}(\mathbf{B})) \to 0$ as $t \to \infty$. By definition, $\mathbf{y}_\ell$ writes $\mathbf{y}_\ell = \mathbf{C}_\ell^{1/2}\mathbf{x}_\ell$ where matrix $\mathbf{C}_\ell$ is defined in Theorem 2, $\mathbf{x}_\ell$ is a vector with centered i.i.d. elements and $\mathbf{x}_l$ and $\mathbf{Q}_l$ are independent. As a consequence, $\frac{1}{t}\mathbf{y}_\ell^H\mathbf{Q}_\ell\mathbf{y}_\ell$ is very close to $\frac{1}{t}\mathrm{tr}(\mathbf{C}_\ell\mathbf{Q}_\ell)$ for large t. Now, since

$$(\mathbb{E}_\ell - \mathbb{E}_{\ell+1})\log\left(1 + \frac{1}{t}\mathrm{tr}(\mathbf{C}_\ell\mathbf{Q}_\ell)\right) = 0 ,$$

Eq. (6) writes $W_\ell = (\mathbb{E}_\ell - \mathbb{E}_{l+1})\log(1 + \gamma_\ell)$ where

$$\gamma_\ell = \frac{\frac{1}{t}\left(\mathbf{y}_\ell^H\mathbf{Q}_\ell\mathbf{y}_\ell - \mathrm{tr}(\mathbf{C}_\ell\mathbf{Q}_\ell)\right)}{1 + \frac{1}{t}\mathrm{tr}(\mathbf{C}_\ell\mathbf{Q}_\ell)}$$

is small for large t. Using the approximation $\log(1+\gamma_\ell) \approx \gamma_\ell$ and noticing that $\mathbb{E}_{\ell+1}\gamma_\ell = 0$, we finally get:

$$W_\ell \approx \mathbb{E}_\ell\gamma_\ell .$$

In order to establish the CLT, we work out the sum of martingale increments $\sum_{\ell=1}^t \mathbb{E}_\ell\gamma_\ell$. The properties of the resolvents $\mathbf{Q}_\ell$ together with their links with matrix $\mathbf{G}$ defined in Theorem 2 play a fundamental role in this analysis.

The details to complete the proof and also to establish the first item of the theorem can be found in [9]. The third item can be proved by relying on Gaussian tools as developed in [7].

The study of the fluctuations of functionals of random matrices with the help of martingales has been initiated by Girko (see also [2]).

REFERENCES

[1] G.W. Anderson and O. Zeitouni, "A CLT for a band matrix model", *PTRF*, 134 (2), pp 283–338, 2006.

[2] Z.D. Bai and J.W. Silverstein, "CLT for Linear Spectral Statistics of Large-Dimensional Sample Covariance Matrices", *Ann. Prob.*, 32 (1A), pp 553–605, 2004.

[3] P. Billingsley, "Probability and Measure", Wiley, 1995.

[4] C.N. Chuah, D.N.C. Tse, J.M. Kahn, and R.A. Valenzuela, "Capacity Scaling in MIMO Wireless Systems Under Correlated Fading", *IEEE Trans. on IT*, 48 (3), pp 637–650, Mar. 2002.

[5] G.J. Foschini, "Layered Space-Time Architecture for Wireless Communication in a Fading Environment When Using Multiple Antennas", Bell Labs Technical Journal, 1 (2), pp 41–59, 1996.

[6] V.L. Girko, "Theory of Random Determinants", Volume 45 of Mathematics and its Applications (Soviet Series), Kluwer Academic Publishers Group, 1990.

[7] W. Hachem, O. Khorunzhiy, Ph. Loubaton, J. Najim and L. Pastur, "A New Approach for Capacity Analysis of Large Dimensional Multi-Antenna Channels", to be published in*IEEE Trans. on IT*, Nov. 2006, available online at `arXiv`.

[8] W. Hachem, Ph. Loubaton, J. Najim, "Deterministic equivalents for certain functionals of large random matrices", *Annals of Applied Probability*, 17 (3), 2007.

[9] W. Hachem, Ph. Loubaton, J. Najim, "A CLT for Information Theoretic Statistics of Gram Random Matrices with a Given Variance Profile", June 2007. To be published in *Annals of Applied Probability*. Available online at `HAL` and `arXiV`.

[10] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton and J; Najim, "On the capacity achieving covariance matrix for Rician MIMO channels: an asymptotic approach", october 2007, submitted to *IEEE Trans. Inf. Th.*.

[11] V.A. Marchenko and L.A. Pastur, "Distribution of Eigenvalues for Some Sets of Random Matrices", *Math. USSR – Sbornik*, 1 (4), pp 457–483, 1967.

[12] A.L. Moustakas, S.H. Simon and A.M. Sengupta, "MIMO Capacity Through Correlated Channels in the Presence of Correlated Interferers and Noise: A (Not So) Large N Analysis" *IEEE Trans. on IT*, 49 (10), pp 2545–2561, Oct 2003.

[13] A.M. Sayeed, "Deconstructing multiantenna fading channels", *IEEE Trans. on SP*, pp. 2563-2579, Oct 2002.

[14] D.S. Shiu and G.J. Foschini and M.J. Gans and J.M. Kahn, "Fading Correlation and its Effect on the Capacity of Multielement Antenna Systems", *IEEE Trans. Com.*, 48 (3), pp 502–513, March 2000.

[15] I.E. Telatar, "Capacity of Multi-Antenna Gaussian Channels", Technical Memorandum, Bell Laboratories, Lucent Technologies, October 1995, published in *ETT*, 10 (6), pp. 585–595, Nov/Dec 1999.

[16] A.M. Tulino, S. Verdu, "Random Matrix Theory and Wireless Communications", in Foundations and Trends in Communications and Information Theory, vol. 1, pp. 1-182, Now Publishers, June 2004.