MMSE Analysis of Certain Large Isometric Random Precoded Systems

Mérouane Debbah[1,3], Walid Hachem[2], Philippe Loubaton[3], Marc de Courville[1]

Abstract

Linear precoding consists in multiplying by a $N \times K$ matrix a K-dimensional vector obtained by serial to parallel conversion of a symbol sequence to be transmitted. In this paper, new tools, borrowed from the so-called *free probability theory*, are introduced for the purpose of analyzing the performance of MMSE receivers for certain large random isometric precoded systems on fading channels. The isometric condition represents the case of precoding matrices with orthonormal columns. It is shown in this contribution that the Signal to Interference plus Noise Ratio at the equalizer output converges almost surely to a deterministic value depending on the probability distribution of the channel coefficients when $N \to +\infty$ and $K/N \to \alpha \leq 1$. These asymptotic results are used to analyze the impact of orthogonal spreading as well as to optimally balance the redundancy introduced between linear precoding versus classical convolutional coding, while preserving a simple MMSE equalization scheme at the receiver.

Keywords

Free Probability, Random Matrices, Signal Space Diversity, MMSE Receivers

[1] Motorola Labs - Paris, 91193, Gif-Sur-Yvette, France.
[2] Corresponding author – Supélec, Service Radioélectricité, 91192, Gif-Sur-Yvette, France.
[3] Université de Marne-La-Vallée, Equipe Signal pour les Communications and URA820 CNRS/ENST, 77454, Marne-La-Vallée, France.

Permission to publish this abstract separately is granted

I. Introduction

High data rate transmission over Rayleigh fading channels requires the use of appropriate diversity schemes. These schemes aim at transmitting various replicas of the emitted signal, which when appropriately combined by the receiver, allow to enhance the signal to noise ratio. Recently, Giraud and Belfiore [1], [2], and then Boutros and Viterbo [3] introduced an attractive new diversity scheme called **signal space diversity**. In particular, contribution [3] describes a modulation scheme depicted in figure 1, in which the input symbol stream is serial to parallel converted, then the resulting K-dimensional symbol vector $\mathbf{s}(n)$ (a white vector process with $E\left(\mathbf{s}(n)\mathbf{s}^H(n)\right) = \mathbf{I}_K$) is multiplied by an isometric $N \times K$ matrix $\mathbf{W}_N$ (i.e. $\mathbf{W}_N^H\mathbf{W}_N = \mathbf{I}_K$) where $N \geq K$. This N-dimensional vector $\mathbf{W}_N\mathbf{s}(n)$ is parallel to serial converted, and the corresponding generated data stream is sent across a non selective Rayleigh fading channel. After serial to parallel conversion, the N–dimensional received vector $\mathbf{y}(n)$ can be written as:

$$\mathbf{y}(n) \;=\; \mathbf{H}_N(n)\,\mathbf{W}_N\,\mathbf{s}(n) \;+\; \mathbf{n}(n) \tag{1}$$

where $\mathbf{n}(n)$ is an white additive Gaussian noise such that $E\left(\mathbf{n}(n)\mathbf{n}(n)^H\right) = \sigma^2\mathbf{I}_N$, and where $\mathbf{H}_N(n) = \mathrm{diag}([h_1(n), \ldots, h_N(n)])$ is the $N \times N$ diagonal complex matrix bearing on its diagonal the channel gains. The role of matrix $\mathbf{W}_N$ is to introduce diversity so that it allows to transmit each component of $\mathbf{s}(n)$ over a duration N times longer than if $\mathbf{W}_N$ were reduced to $\mathbf{I}_N$.

Note that the model proposed for describing the system is broad enough to capture a multiplicity of transmission schemes. These include:

- Multi-Carrier CDMA (MC-CDMA) downlink transmissions [4][5]. In this case, the elements of $\mathbf{s}(n)$ represent K different streams of symbols destined to K different users, N coincides with the number of sub-carriers, and each column of $\mathbf{W}_N$ represents the code allocated to each user. Vector $\mathbf{W}_N\mathbf{s}(n)$ is sent to an OFDM modulator, and equation (1) represents the signal received after guard interval suppression and Fourier transformation. In the frequency domain, the received $N \times 1$ vector signal can be seen as resulting from a transmission over N parallel flat fading channels. Diagonal entries of $\mathbf{H}_N(n)$ represent the frequency domain channel gains for data frame n.

- Precoded OFDM [6] or Spread OFDM [7] in a single user context. In this case, matrix $\mathbf{W}_N$ acts as a means for spreading each component of $\mathbf{s}(n)$ over all carriers. This increases the overall frequency diversity of the modulator, so that deeply attenuated carriers can still be recovered by taking advantage of the sub-bands enjoying a high Signal to Noise Ratio.

An important problem lies in the choice of the amount of redundancy introduced by linear precoding, i.e. the ratio K/N, and also in the choice of matrix $\mathbf{W}_N$. [1] and [3] considered the case where $K/N = 1$, i.e. $\mathbf{W}_N$ is unitary. They assumed the entries of $\mathbf{H}_N(n)$ independent and identically distributed, and proposed to derive an upper bound of the error probability for the Maximum Likelihood (ML) detector of $\mathbf{s}(n)$. They discovered that, at least for high Signal to Noise Ratios, $\mathbf{W}_N$ has to be chosen in such a way that the minimum so-called L–distance product of the constellation $\{\mathbf{W}_N\mathbf{s}_i, i \in I\}$, where $\{\mathbf{s}_i, i \in I\}$ denotes the set of possible values taken by vector $\mathbf{s}(n)$, be maximum. But the optimization of the coefficients of $\mathbf{W}_N$ is hardly trivial and entails the use of sophisticated mathematical tools from algebraic number theory. Moreover, the high computational cost of the ML detector prevents its use in practical contexts.

Actually, due to its lower complexity, MMSE detection is often preferred. In our context, the SINR (Signal to Interference plus Noise Ratio) is a natural figure for evaluating the MMSE detector performance. For reasons related to calculus, the analysis will be conducted in the asymptotic regime ($N \to \infty, K \to \infty, \frac{K}{N} \to \alpha \leq 1$). We note that this is especially relevant in the precoded OFDM case since in usual wireless OFDM systems, a high number of carriers (≥ 64) are involved.

The output of the MMSE detector is $\hat{\mathbf{s}}(n) = [\hat{s}_1(n), \ldots, \hat{s}_K(n)]^T$ and is given by ([8]) :

$$\begin{aligned}\hat{\mathbf{s}}(n) &= E\left(\mathbf{s}(n)\mathbf{y}^H(n)\right)\ \left(E(\mathbf{y}(n)\mathbf{y}^H(n))\right)^{-1}\mathbf{y}(n)\\ &= \mathbf{W}_N^H\mathbf{H}_N^H(n)\left(\mathbf{H}_N(n)\mathbf{W}_N\mathbf{W}_N^H\mathbf{H}_N^H(n) + \sigma^2\mathbf{I}_N\right)^{-1}\mathbf{y}(n)\ .\end{aligned}$$

Each component $\hat{s}_k(n)$ of $\hat{\mathbf{s}}(n)$ is corrupted by the effect of both the thermal noise and by the "multi-user interference" due to the contributions of the other users $\{s_l(n)\}_{l\neq k}$. It has been shown in [9], and recently in [10], that this additive noise can be considered as Gaussian when K and N are large enough. Therefore, the Signal to Interference plus Noise Ratio (SINR) at the output of each component of the MMSE detector characterizes entirely the performance of the modulation scheme equipped with a MMSE receiver. Several papers [11][12][13] have recently analyzed the behavior of the SINR at the output of the MMSE detector in the case where the entries of $\mathbf{W}_N$ are independent and identically distributed random variables (to be referred to in the sequel as the i.i.d. case) and the matrix $\mathbf{H}_N(n)$ is reduced to $\mathbf{I}_N$ and the various users can have different powers. In this case, it has been shown that this SINR converges almost surely toward a well defined deterministic value which does not depend on the particular realization of $\mathbf{W}_N$. These results have been used by several authors to better understand the performance of the chosen transmitter/receiver chain. In particular, Biglieri *et al.* [14] and Shamai *et al.* [15] showed how to use these results to find the optimum value of the parameter α via the analysis of the system throughput. We also note that [16] considered the case where the columns of $\mathbf{W}_N$ are independent and identically distributed N–dimensional random vectors uniformly distributed on the unit sphere of $\mathbb{C}^N$.

In this paper, we also study the behavior of the SINR at the output of the MMSE detector assuming that $\mathbf{W}_N$ is a random matrix and take benefit of the corresponding results to discuss the choice of α. However, we address the case where $\mathbf{W}_N$ is isometric. In the sequel, this will be called the isometric case. The choice of an isometric spreading matrix is usual in systems where synchronization is ensured like signal space diversity systems or precoded OFDM or downstream MC-CDMA systems, since it provides much better results than the choice of an i.i.d. matrix. Moreover, we stress on the fact that in our models, matrix $\mathbf{H}_N(n)$ is not reduced to identity. To our knowledge, the problem we address here has not been considered in previous works.

From a technical stand point, the i.i.d. case study of [13] leans on mathematical results that concern the "limiting distribution of eigenvalues" of some large random matrices with independent and identically distributed entries (see e.g. [17]). As for the isometric case, a considerably more involved material will be needed. The results given here rely on the so-called *free probability theory* initially developed by D. Voiculescu [18] in order to solve deep problems of operator algebras classification. At the end of the eighties, Voiculescu realized that this theory could also be used to analyze the eigenvalue distribution of sums or products of certain independent large random matrices. This will be the starting point of our analysis. Note that Evans and Tse already introduced free probability theory in [12].

This paper is structured as follows. In section II, we precise the way we generate the random isometric valued matrices under consideration. Then, we present our main result (theorem 1), which asserts that the SINR at the output of the MMSE detector converges almost surely to a deterministic value. At the end of this section, an asymptotic result for the i.i.d. case is also given for the purpose of comparison. In section III, we confirm by simulations the fact that our asymptotic analysis allows to predict the performance of the MMSE detector for relatively small values of N. Section IV starts with the evaluation of the whole system throughput (see [14]) with respect to α for a fixed allocated bandwidth. The purpose of this computation is to determine the optimal amount of redundancy that should be spent on the linear precoder in a system where linear precoding is combined with convolutional encoding. This analysis is sustained by a performance comparison between the linearly precoded system and a system equipped with a classical convolutional forward error correcting code in addition to linear precoding, the overall coding rate being the same for both systems. We shall confirm by practical examples the fact that the redundancy trade-off between linear precoding (through the choice of α) and coding (through the choice of the rate of the convolutional code) can be given à priori by the throughput analysis. This discussion was motivated by the recent paper [6]. Section V and section VI contain the most technical part of the paper. As we believe that free probability theory is an important and promising tool, we propose to give in section V a comprehensive introduction to its most important aspects concerning our purpose. In section VI, we show how to apply an important lemma of [19] (also used in [12]) to our particular situation, and finally use the almost sure asymptotic freeness result presented in the recent monograph [20] to prove theorem 1.

II. Main Result

A. Hypotheses and Preliminary Properties

In this subsection, we first precise the properties of matrix $\mathbf{W}_N$ and of the diagonal entries of $\mathbf{H}_N(n)$. Since the time index n is not relevant in the following, we simply omit it. Therefore, equation (1) can be re-written as:

$$\mathbf{y} = \mathbf{H}_N \mathbf{W}_N \mathbf{s} + \mathbf{n} \ . \tag{2}$$

In the following, we assume that

A1: $\mathbf{H}_N = \mathrm{diag}([h_1, \ldots, h_N])$ has identically distributed centered random diagonal entries. $|h_i|^2$ is supposed to have a probability density $p(t)$ with finite moments of all orders.

We set $E(|h_i|^2) = 1$, so that σ^2 defined by $E(\mathbf{n}\mathbf{n}^H) = \sigma^2 \mathbf{I}_N$ represents the inverse of the SNR at the receiver input. As a typical example, if the coefficients $\{h_i\}_{i=1,\ldots,N}$ are complex Gaussian[1] (resulting in a Rayleigh fading), then $p(t) = e^{-t}$ which corresponds to a χ^2 distribution with two degrees of freedom. It is important to notice that random variables $\{h_i\}_{i=1,N}$ are not assumed to be independent. However, we assume that the following assumption holds:

A2: for each $l \geq 1$,

$$\lim_{N \to +\infty} \frac{1}{N} \sum_{k=1}^{N} |h_k|^{2l} = E(|h_i|^{2l}) \text{ almost surely.} \tag{3}$$

Assumption **A2** implies some kind of asymptotic independence between the random variables h_i and h_j if $|i - j| \to \infty$. This hypothesis is quite realistic in the context of signal-space diversity schemes or in precoded OFDM systems if large size interleavers/deinterleavers are inserted in the scheme represented in figure 1. Interleavers are needed in precoded OFDM systems because without interleaving, the coefficients $(h_k)_{k=1,\ldots,N}$ coincide with the values taken by the transfer function $h(e^{2i\pi f})$ of a multi-path Rayleigh fading channel at frequencies $\frac{(k-1)}{N}, k = 1, \ldots, N$. If the number of paths remains fixed when $N \to \infty$, the sequence $(h_k)_{k=1,\ldots,N}$ depend on a finite number of independent random variables, and the hypothesis **A2** cannot of course be fulfilled. However, in such a context, the tools developed in this paper can still be applied, but the main results have a different form. The reader is referred to [21] for more details.

We now explain how the random matrix $\mathbf{W}_N$ is generated. For this purpose, some notations and definitions need to be introduced. Denote by $\mathcal{U}$ the multiplicative group

[1]By a complex Gaussian random variable, we mean a complex random variable whose real and imaginary parts are independent Gaussian random variables having same variances.

of $N \times N$ unitary matrices, and by $\mathbf{\Theta}$ a random $N \times N$ unitary matrix. $\mathbf{\Theta}$ is said to be Haar distributed if the probability distribution of $\mathbf{\Theta}$ is invariant by left multiplication by constant unitary matrices [2]. Since the group $\mathcal{U}$ is compact, this condition is known to be equivalent to the invariance of the probability distribution of $\mathbf{\Theta}$ by right multiplication by constant unitary matrices. In order to generate Haar distributed unitary random matrices, let $\mathbf{X} = [x_{i,j}]_{1 \leq i,j \leq N}$ be a $N \times N$ random matrix with independent complex Gaussian centered unit variance entries. The unitary matrix $\mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1/2}$ is Haar distributed. To see this, notice that for each constant unitary matrix $\mathbf{U}$,

$$\mathbf{UX}(\mathbf{X}^H\mathbf{X})^{-1/2} = \mathbf{UX}((\mathbf{UX})^H\mathbf{UX})^{-1/2} .$$

Since the probability distribution of $\mathbf{X}$ and $\mathbf{UX}$ coincide, matrices $\mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1/2}$ and $\mathbf{UX}((\mathbf{UX})^H\mathbf{UX})^{-1/2}$ have the same distribution. The above equality thus implies that $\mathbf{UX}(\mathbf{X}^H\mathbf{X})^{-1/2}$ and $\mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1/2}$ are identically distributed.

There is another way for building Haar distributed unitary matrices that will be useful to our purpose. Instead of multiplying $\mathbf{X}$ by the inverse of the Hermitian square root of $\mathbf{X}^H\mathbf{X}$, one can introduce the uniquely defined upper triangular matrix with positive diagonal elements $\mathbf{Q}(\mathbf{X})$ defined by

$$\mathbf{X}^H\mathbf{X} = \mathbf{Q}(\mathbf{X})^H\mathbf{Q}(\mathbf{X}) .$$

The unitary matrix $\mathbf{V}(\mathbf{X})$ defined by

$$\mathbf{V}(\mathbf{X}) = \mathbf{XQ}(\mathbf{X})^{-1} \tag{4}$$

is also Haar distributed. To see this, we first remark that for each constant unitary matrix $\mathbf{U}$, the probability distribution of $\mathbf{V}(\mathbf{X})$ and of $\mathbf{V}(\mathbf{UX})$ coincide. But, it is obvious that $\mathbf{Q}(\mathbf{UX}) = \mathbf{Q}(\mathbf{X})$, so that $\mathbf{UV}(\mathbf{X}) = \mathbf{V}(\mathbf{UX})$. Therefore, the probability distribution of $\mathbf{V}(\mathbf{X})$ and of $\mathbf{UV}(\mathbf{X})$ coincide. Remark that the columns of $\mathbf{V}(\mathbf{X})$ are obtained by a Gram-Schmidt orthogonalization of the columns of $\mathbf{X}$.

Finally, we state an interesting property of Haar distributed unitary random matrices. $\mathbf{\Theta}$ being one such matrix, its probability distribution is also invariant under right multiplication by unitary matrices, hence this distribution coincides with the distribution of $\mathbf{\Theta P}$

[2]In other words, the probability distribution of $\mathbf{\Theta}$ coincides with the so-called Haar measure on $\mathcal{U}$.

for any permutation matrix $\mathbf{P}$. This shows that the $N \times K$ isometric matrices obtained by extracting *any* subset of K columns from $\mathbf{\Theta}$ have the same probability distribution. In the following, it will be assumed that

A3 : matrix $\mathbf{W}_N$ is generated by extracting K columns from a $N \times N$ Haar unitary random matrix $\mathbf{\Theta}_N$ independent of $\mathbf{H}_N$.

B. Statement of the Main Result.

Let us first recall the expression of the SINR at one of the K outputs of the MMSE detector. Let $\mathbf{w}_N$ be the column of $\mathbf{W}_N$ associated to some element of $\mathbf{s}$, and $\mathbf{U}_N$ the $N \times (K-1)$ isometric matrix which remains after extracting $\mathbf{w}_N$ from $\mathbf{W}_N$. The SINR $\beta_{\mathbf{w}_N}$ at the output of the MMSE detector is easily shown to express as (see e.g. [13]):

$$\beta_{\mathbf{w}_N} = \frac{\eta_{\mathbf{w}_N}}{1 - \eta_{\mathbf{w}_N}} \tag{5}$$

where

$$\eta_{\mathbf{w}_N} = \mathbf{w}_N^H \mathbf{H}_N^H \left(\mathbf{H}_N \mathbf{W}_N \mathbf{W}_N^H \mathbf{H}_N^H + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{H}_N \mathbf{w}_N \ . \tag{6}$$

Writing $\mathbf{H}_N \mathbf{W}_N \mathbf{W}_N^H \mathbf{H}_N^H = \mathbf{H}_N \mathbf{U}_N \mathbf{U}_N^H \mathbf{H}_N^H + \mathbf{H}_N \mathbf{w}_N \mathbf{w}_N^H \mathbf{H}_N^H$ and invoking the matrix inversion lemma, we get after some simple algebra another useful expression for this SINR:

$$\beta_{\mathbf{w}_N} = \mathbf{w}_N^H \mathbf{H}_N^H \left(\mathbf{H}_N \mathbf{U}_N \mathbf{U}_N^H \mathbf{H}_N^H + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{H}_N \mathbf{w}_N \ . \tag{7}$$

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

Theorem 1: *Assume that matrices $\mathbf{W}_N$ and $\mathbf{H}_N$ are chosen according to assumptions* **A1** *to* **A3**, *and moreover, that*

A4: *the probability density $p(t)$ of the random variables $(|h_i|^2)_{i \in \mathbb{N}}$ has a compact support included in the interval $[0, c]$, which implies that $\sup_{i \in \mathbb{N}} |h_i|^2 \leq c < \infty$ almost surely.*

When N grows toward infinity and $K/N \to \alpha \leq 1$, the SINR $\beta_{\mathbf{w}_N}$ at the output of a MMSE equalizer converges almost surely to a value $\overline{\beta}$ that is the unique solution of the equation

$$\int_0^\infty \frac{t}{\alpha t + \sigma^2 (1-\alpha)\overline{\beta} + \sigma^2} \ p(t) \ dt = \frac{\overline{\beta}}{\overline{\beta} + 1} \ . \tag{8}$$

This theorem is proved in section VI. In order to give some insights to the reader, we just briefly justify the result when $K = N$ for each N, which of course implies $\alpha = 1$. The case $\alpha = 1$ is easy to handle because $\eta_{\mathbf{w}_N} = \frac{\beta_{\mathbf{w}_N}}{1+\beta_{\mathbf{w}_N}}$ has a simple expression. Indeed, when $K = N$, matrix $\mathbf{W}_N$ is unitary, so $\mathbf{W}_N \mathbf{W}_N^H = \mathbf{I}_N$. Therefore, $\eta_{\mathbf{w}_N}$ can be written as

$$\eta_{\mathbf{w}_N} = \sum_{k=1}^{N} |w_{N,k}|^2 \frac{|h_k|^2}{|h_k|^2 + \sigma^2} \tag{9}$$

where $w_{N,k}$ denotes the k^{th} component of vector $\mathbf{w}_N$. Using the fact that $E|w_{N,k}|^2 = \frac{1}{N}$ and $E(|w_{N,k}|^2 |w_{N,l}|^2) = \frac{1}{N(N+1)}$ if $k \neq l$ and $\frac{2}{N(N+1)}$ if $k = l$ (see e.g. [20]), it is rather straightforward to show that $\eta_{\mathbf{w}_N}$ converges in the mean square sense to the quantity $\overline{\eta}$ defined by

$$\overline{\eta} = E(\frac{|h_k|^2}{|h_k|^2 + \sigma^2}) = \int_0^\infty \frac{t}{t + \sigma^2}\, p(t)\, dt$$

This shows that $\beta_{\mathbf{w}_N}$ converges in probability to the value $\overline{\beta}$ defined by $\overline{\beta} = \frac{\overline{\eta}}{1-\overline{\eta}}$ as expected. The proof of the almost sure convergence for $\alpha = 1$ is a little bit more complicated, but is a consequence of an important lemma of [19] already used in [12] (See section VI for more details).

Before proceeding, let us give some additional remarks on theorem 1 :

Remark 1: It is asserted that equation (8) has a unique solution. As a matter of fact, the left hand side of (8) is a positive decreasing function of $\overline{\beta}$ which converges to 0 when $\overline{\beta} \to +\infty$, while the right hand side of (8) is an increasing function of $\overline{\beta}$ which is 0 at $\overline{\beta} = 0$.

Remark 2: When the measure $p(t)dt$ is replaced by the Dirac measure $\delta(1)$ at point $t = 1$, matrix $\mathbf{H}_N$ is reduced to $\mathbf{I}_N$. By direct computation, the receiver output SINR in this situation (called the Gaussian channel situation) is easily shown to be $1/\sigma^2$. This value is also given by equation (8) when $p(t)dt$ is replaced by $\delta(1)$.

Remark 3: In the statement of theorem 1, $p(t)$ is assumed compactly supported. This hypothesis is important on a technical point of view because the most powerful results of free probability theory (in particular the asymptotic freeness of independent large random matrices, see below) require compactly supported measures. Although the usual probability distributions of the coefficients $(h_i)_{i \in \mathbb{N}}$, like the Rayleigh or the Rice distributions, do not meet this requirement, this restriction is of course not very important in practice. In particular, the use of formula (8) with $p(t) = e^{-t}$ (for the Rayleigh channel case) allows to predict quite well the performance of our precoded modulation scheme using MMSE detection.

Remark 4: One key information provided by theorem 1 is that the SINR does not depend on the particular realization of the isometric matrix sequence $(\mathbf{W}_N)_{N\in\mathbb{N}}$ if $N \to +\infty$ and $K/N \to \alpha$, provided it is extracted from a Haar distributed random unitary matrix. However, usual precoded OFDM or MC-CDMA systems use quite different precoding matrices, e.g. Walsh-Hadamard matrices, which of course do not coincide with realizations of Haar distributed random matrices. It is therefore important to check if, in practice, the most common precoding matrices provide the same asymptotic performance than realizations of Haar distributed random unitary matrices. When $\alpha = 1$, Equation (9) immediately shows that if matrix $\mathbf{W}_N$ is replaced by a deterministic isometric matrix whose entries have the same modulus $\frac{1}{\sqrt{N}}$ (the Walsh-Hadamard matrices, FFT matrices satisfy of course this condition), then $\eta_{\mathbf{w}_N}$, and thus $\beta_{\mathbf{w}_N}$ converge to the value predicted by Theorem 1. The case $\alpha < 1$ is studied by simulations in section III.

We now address the i.i.d case. In the next section, the performance of systems having large isometric precoder matrices will be compared to the performance of systems with i.i.d. matrices, and the impact of the precoder column orthogonality will be quantified.

Theorem 2: *Assume that the entries of* $\mathbf{W}_N$ *are centered i.i.d. random variables of variance* $1/N$*, that assumptions* **A1** *and* **A4** *hold, and that* **A2** *is replaced by* **A2'**
A2': *for each bounded continuous function* φ*,*

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} \varphi\left(\frac{1}{|h_i|^2}\right) = E\left(\varphi\left(\frac{1}{|h|^2}\right)\right) = \int \varphi\left(\frac{1}{t}\right) p(t)\ dt$$

almost surely.
Then, when N *grows toward infinity and* $K/N \to \alpha$*, the SINR* $\beta_{\mathbf{w}_N}$ *at the output of a MMSE equalizer converges almost surely to a value* $\overline{\beta}_1$ *that is the unique solution of the equation*

$$\int_0^\infty \frac{t}{\alpha t + \sigma^2\overline{\beta_1} + \sigma^2}\ p(t)\ dt = \frac{\overline{\beta}_1}{\overline{\beta}_1 + 1}\ . \tag{10}$$

As this paper is essentially devoted to the isometric case, this result will be justified briefly in the appendix. The proof is a direct consequence of corollary 1 in [12] and of the main result of [17]. Note that α in this theorem is not restricted to $[0,1]$ but belongs to $[0,+\infty[$.

Remark 5: Comparison with the Tse-Hanly formula. The so-called Tse-Hanly formula for the MMSE receiver [13, formula 4] gives the asymptotic SINR value β_1^* for CDMA systems with random i.i.d. codes in a flat fading channel, which in our context amounts to $h_n = h$ for each n, or equivalently $\mathbf{H}_N = h\mathbf{I}_N$. As the powers allocated to the various components of vector $\mathbf{s}$ all coincide in our context, this equation is written

$$\beta_1^* = \frac{1}{\frac{\alpha}{\beta_1^*+1} + \frac{\sigma^2}{|h|^2}} \ . \tag{11}$$

Recall that Tse and Hanly interpreted in [13] the factor $\frac{1}{\beta_1^*+1}$ as the effective interference of component i of $\mathbf{s}$ on the desired component k at the desired target SINR β_1^*. The term $\frac{\alpha}{\beta_1^*+1} \simeq \frac{1}{N}\frac{K}{\beta_1^*+1}$ thus represents the total amount of multi-user interference at the output of the MMSE receiver (the term $\frac{1}{\beta_1^*+1}$ is multiplied by K the number of users, while the coefficient $\frac{1}{N}$ is due to the spreading gain provided by the precoder).

We remark that equation (10) of theorem 2 can be rewritten

$$\overline{\beta}_1 = \int_0^{\infty} \frac{1}{\frac{\alpha}{\overline{\beta}_1+1} + \frac{\sigma^2}{t}} \ p(t) \ dt \tag{12}$$

It is interesting to note that the right-hand-side of (12) coincides with an averaged version (on the square of the amplitude of the channel coefficients) of the inverse of the sum of the *multi-user effective interference term* $\frac{\alpha}{\overline{\beta}_1+1}$ and of the term $\frac{\sigma^2}{|h|^2}$ which represents the contribution of a thermal noise of variance $\frac{\sigma^2}{|h|^2}$ in a flat fading channel of complex gain h. This shows first that the diversity provided by the precoder is of course due to the averaging on the values taken by $|h|^2$ in (12). More importantly, (12) also indicates that the important concept of multi-user effective interference introduced in [13] is still relevant if $\mathbf{H}_N$ is not a multiple of $\mathbf{I}_N$.

Remark 6: It is clear that for each $\beta > 0$,

$$\int_0^{\infty} \frac{t}{\alpha t + \sigma^2\beta + \sigma^2} \ p(t) \ dt \ \leq \ \int_0^{\infty} \frac{t}{\alpha t + \sigma^2(1-\alpha)\beta + \sigma^2} \ p(t) \ dt \quad .$$

This implies that for a fixed value of α, the SINR in the i.i.d case is always less than the SINR in the isometric case. Moreover, the performance gain induced by the use of isometric codes instead of i.i.d. ones grows when α grows toward 1. Conversely, the SINR in the i.i.d case is nearly equal to the SINR in the orthogonal case if α is close to 0. It is also interesting to note that the second term of the right-hand-side of (10) and (8) are similar. The multi-user interference term $\frac{\alpha}{\overline{\beta}+1}$ appears in both formulas, while the term $\frac{\sigma^2}{|h|^2}$, representing the effect of the thermal noise in the i.i.d. case, is multiplied in the isometric case by $1-\alpha\frac{\overline{\beta}}{1+\overline{\beta}}$, which is of course less than 1. In other words, for a given target SINR of $\overline{\beta}$, an isometric precoded system corrupted by a thermal noise of variance σ^2 provides the same performance as an i.i.d. precoded one corrupted by a thermal noise of variance $(1-\alpha\frac{\overline{\beta}}{1+\overline{\beta}})\sigma^2$.

Remark 7: Case of non equal powers. In this paper, we just consider the case where the components of $\mathbf{s}$ have the same power. This is because we are mainly motivated by the study of single user precoded systems, i.e. all the components of $\mathbf{s}$ are to be sent to the same user. The non equal power case is nevertheless quite relevant in the context multi-user systems. However, the approach used in the present paper cannot be generalized to this context because the calculation of $\overline{\beta}$ relies on the equal power assumption (see section VI). The non equal power case requires the use of more sophisticated tools. The interested reader is referred to [22] for more details.

Remark 8: Case where $\alpha > 1$. Theorem 2 remains valid when $\alpha > 1$. Although not intuitive, the use of a $N \times K$ precoder with $K > N$ may improve the performance in the i.i.d. case for low values of the signal to noise ratio (see e.g. [15] in the context of a frequency flat fading channel). Of course, the case $\alpha > 1$ does not make sense in the isometric case. However, instead of using a precoding matrix which columns are orthogonal, one may use if $\alpha > 1$ a matrix $\mathbf{W}_{N,K}$ whose **rows** are orthogonal. In this context, we model the precoding matrix $\mathbf{W}_{N,K}$ as a random matrix obtained first by extracting N rows from a Haar distributed unitary $K \times K$ matrix, and second by multiplying the resulting matrix by the scaling factor $\alpha^{1/2}$. Therefore, $\mathbf{W}_{N,K}$ satisfies

$$\mathbf{W}_{N,K}\mathbf{W}_{N,K}^H = \alpha\mathbf{I}_N \tag{13}$$

The scaling factor α in (13) normalizes the power allocated to each component of $\mathbf{s}$. Derivation of the MMSE output SINR in the asymptotic regime is similar to the case where $\alpha = 1$. To be precise, as $\mathbf{W}_{N,K}\mathbf{W}_{N,K}^H = \alpha \mathbf{I}_N$, $\eta_{\mathbf{w}_N}$ is given by

$$\eta_{\mathbf{w}_N} = \sum_{k=1}^{N} |w_{N,k}|^2 \frac{|h_k|^2}{\alpha |h_k|^2 + \sigma^2} \tag{14}$$

As in the case $\alpha = 1$, it is easy to check that $\eta_{\mathbf{w}_N}$ converges to $\overline{\eta}$ defined by

$$\overline{\eta} = E(\frac{|h_k|^2}{\alpha |h_k|^2 + \sigma^2}) = \int_0^\infty \frac{t}{\alpha t + \sigma^2}\ p(t)\ dt$$

Therefore, the SINR $\overline{\beta}$ converges toward the unique solution of the equation

$$\frac{\overline{\beta}}{\overline{\beta} + 1} = \int_0^\infty \frac{t}{\alpha t + \sigma^2}\ p(t)\ dt$$

In the section IV devoted to the choice of α, we show that in contrast with the i.i.d. case, it is not relevant to use fat precoding matrices in the Haar distributed case, even for low signal to noise ratios.

III. Numerical Illustration

In this section, we first study in a more precise manner the influence of α on the theoretical asymptotic SINR as well as on the bit error rate (BER) in a scenario where no convolutional encoding is implemented. These results will then be confirmed by simulation. Symbols for all users have their values in a QPSK constellation. The channel is assumed to be a Rayleigh fading channel, in other words $p(t) = e^{-t}$. Although this hypothesis does not meet the technical assumption **A4**, we shall nevertheless make use of formulas (8) and (10) to predict the asymptotic performances of our precoded systems. Except for figure 7, diagonal entries of $\mathbf{H}_N$ are independent, this assumption being justified when large size interleaver and de-interleaver are inserted at the transmitter side and receiver side respectively.

Let

$$E_i(x) = \int_x^{+\infty} \frac{e^{-u}}{u} du$$

be the so-called exponential integral function. For a Rayleigh fading channel, equations (8) and (10) will have the following forms:

Isometric case. Put $r = (\sigma^2(1-\alpha)\overline{\beta} + \sigma^2)/\alpha$. The SINR $\overline{\beta}$ is solution of :

$$1 - r\, e^r\, E_i(r) = \frac{\alpha\overline{\beta}}{\overline{\beta}+1}\ . \tag{15}$$

This is deduced from equation (8): writing $E_i(x) = e^{-x}\int_0^{+\infty} \frac{e^{-u}}{u+x} du$, we have:

$$\begin{aligned}
\int \frac{t}{\alpha t + \sigma^2(1-\alpha)\overline{\beta} + \sigma^2}\ p(t)\ dt &= \int_0^\infty \frac{t}{\alpha t + \sigma^2(1-\alpha)\overline{\beta} + \sigma^2}\ e^{-t}\ dt \\
&= \frac{1}{\alpha}\left(1 - \int_0^\infty \frac{e^{-t}}{\frac{\alpha t}{\sigma^2(1-\alpha)\overline{\beta}+\sigma^2} + 1}\ dt\right) \\
&= \frac{1}{\alpha}\left(1 - \frac{\sigma^2(1-\alpha)\overline{\beta}+\sigma^2}{\alpha} e^{-\frac{\sigma^2(1-\alpha)\overline{\beta}+\sigma^2}{\alpha}} E_i(\frac{\sigma^2(1-\alpha)\overline{\beta}+\sigma^2}{\alpha})\right)
\end{aligned}$$

Equation (15) follows.

i.i.d. case. Put $r = (\sigma^2(\overline{\beta_1} + 1))/\alpha$. After a computation similar to that of the isometric case, the resulting SINR $\overline{\beta_1}$ is solution of :

$$1 - r\, e^r\, E_i(r) = \frac{\alpha\overline{\beta_1}}{\overline{\beta_1}+1}\ . \tag{16}$$

As the contribution of the noise and of the multi-user interference at the output of large MMSE detectors can be considered as Gaussian [9], it is standard to associate to each asymptotic SINR the asymptotic BER given by $Q(\sqrt{\text{SINR}})$. We recall that $Q(x) = (1/\sqrt{2\pi})\int_x^\infty e^{-t^2/2} dt$.

Figures 2 and 3 show the BER in the isometric case and in the i.i.d. case respectively for $\alpha = 1, \frac{1}{2}$, and $\frac{1}{4}$. The BER in the Gaussian channel case (i.e. $\mathbf{H}_N = \mathbf{I}_N$) is also represented since, as can easily be shown (see equation (8)), it represents the asymptotic performance when $\alpha \to 0$. We notice that there is an important performance gap between systems with $\alpha = 1$ and systems with $\alpha = \frac{1}{2}$. This gap is clearly reduced when we pass from $\alpha = \frac{1}{2}$ to $\alpha = \frac{1}{4}$.

In order to verify the practical relevance of these evaluations, we also represent experimental results obtained by numerical simulations for $N = 256$ in the context of a single user precoded system : by single user system, we mean that all the components of $\mathbf{s}$ are to be sent to the same user. In this context, the empirical BER is of course obtained by averaging the errors over the K components of $\mathbf{s}$. Although the chosen matrix size N is not extremely large, we observe that the theoretical curves match quite well the experimental ones for both isometric and i.i.d. cases.

If we now compare figure 2 to figure 3, we notice that isometric precoding outperforms significantly i.i.d. precoding. Moreover, the results are in accordance with remark 6 in the sense that the performance gap between these two types of precoding is significant for $\alpha = 1$, while for $\alpha = \frac{1}{4}$, it is less than 1 dB. In order to compare more precisely isometric to i.i.d. precoding, we represent in figure 4 the SINR loss ($\frac{\text{SINR}_{\text{iid}}}{\text{SINR}_{\text{ortho}}}$ in dB) of i.i.d. precoders with respect to isometric ones for various values of α.

As pointed out earlier, we plotted in figure 6 the BER of the MMSE receiver when a Walsh-Hadamard precoder is used. Here $N = 256$ and $K = 64, 128$ and 256, and asymptotic theoretical plots are also given for comparison. A relatively close match between the two types of curves is observed. One must not conclude however that theoretical results obtained with Haar isometric matrices predict the performance of the Walsh-Hadamard codes. We simply notice that isometric Haar precoders can do as well as standard Walsh Hadamard codes in terms of BER.

The influence of N on the system's performance is shown in figure 5. It can be noticed here that for $N = 128$, asymptotic analysis is fairly precise. It is interesting to evaluate also the pertinence of this analysis in situations where channel gains at different frequencies are correlated. Regarding this point, only the mild condition **A2** need be satisfied. However, when channel gains are correlated, one can expect that larger values of N are needed to attain the asymptotic regime. In figure 7, a correlated channel model is generated by filtering in the frequency domain a Gaussian i.i.d. sequence with a first order transfer function $1/(1 - az^{-1})$. When $a = 0.9$, channel gains are highly correlated and figure 7 shows a performance degradation for $N = 128$. When $a = 0.6$, there is nearly no difference in performance with the case where channel gains are independent.

IV. Choice of α

In this section, we tackle the problem of finding the value of α that maximizes the system's throughput. As said above, the purpose of this analysis is to determine the amount of redundancy that should be spent on the linear precoder. The need for an optimum α can be justified intuitively. On the one hand, at fixed bandwidth, choosing α

close to 1 or even greater than 1 (see remark 8) allows to use small rate error correcting codes but the effect of the "multi-user interference" significantly decreases the SINR at the output of the MMSE detector. On the other hand, a small value of α increases the SINR at the output of the MMSE detector, but higher rate error correcting codes are then needed. Biglieri *et al.* [14] and Shamai *et al.* [15] have already used the spectral efficiency as a tool for answering the coding versus spreading issue in uplink CDMA with random i.i.d codes. They have shown in particular that a non-negligible amount of spreading should be spent when using MMSE receivers if $\frac{E_b}{N_0}$ is large enough while α should be chosen greater than 1 for $\frac{E_b}{N_0} < 3dB$. We look for an extension of such a result to the case where random Haar matrices are used. $\alpha \leq 1$ as well as $\alpha > 1$ (see again remark 8) will be considered.

In the context of this paper, the throughput is the total number of bit/s/Hz that can be reliably transmitted with our precoded system equipped with a MMSE receiver. Our throughput analysis will be confirmed by simulations. Practical convolutional coding schemes of rates $R(\alpha)$ are considered with rates satisfying $\alpha R(\alpha) = 1/2$ for different values of α.

A. *Throughput analysis.*

The throughput $\gamma\left(\alpha, \sigma^2\right)$ is defined by

$$\gamma\left(\alpha, \sigma^2\right) = \alpha\, C\left(\alpha, \sigma^2\right)$$

where $C\left(\alpha, \sigma^2\right)$ is solution of the equation

$$C\left(\alpha, \sigma^2\right) = \log_2\left(1 + \mathrm{SINR}(\alpha, \sigma^2)\right) \tag{17}$$

It is understood that $E_b/N_0 = (C\sigma^2)^{-1}$, see [8] for more details.

We compare the throughput of the current systems to the maximum of the throughput in the Gaussian (non-fading) channel case, which is of course reached for $\alpha = 1$. This upper bound will be called the Gaussian channel bound in the following. It obviously coincides with the capacity of the standard Gaussian channel. Capacity of the Rayleigh fading channel is also plotted.

In figure 9, E_b/N_0 is fixed to 10 dB and we plot the throughput versus α in the Haar distributed and in the i.i.d. code matrices cases. We confirm that Haar precoding outperforms i.i.d. precoding. The performance gap becomes clear when $\alpha > 0.3$ [3].

In figure 10, E_b/N_0 is fixed to 2 dB and we plot the throughput versus α for Haar distributed and for i.i.d. code matrices. In contrast with the case where $E_\mathrm{b}/N_0 = 10$ dB, the optimum value of α in the i.i.d. case is greater than 1. Note however that the gain between $\alpha = 1$ and the optimum value of α is not significant. In contrast, the throughput in the Haar distributed case is optimum for $\alpha = 1$.

This observation is confirmed in Figure 8 which shows the behavior of the optimum value of α (i.e. for which the throughput is maximum) with respect to E_b/N_0 for both Haar distributed ($\alpha \leq 1$ and $\alpha > 1$) and i.i.d. cases. In the Haar distributed case, figure 8 shows that nearly no redundancy should be spent on the precoder to maximize the throughput. In contrast, in the i.i.d case, a significant amount of redundancy is needed when $\frac{E_\mathrm{b}}{N_0} > 4$dB. For instance, at $\frac{E_\mathrm{b}}{N_0} = 10dB$, the optimum value of α is 0.63.

In figure 11, we represent the optimum (w.r.t. α) throughput versus E_b/N_0 in the isometric and i.i.d case. At $\frac{E_\mathrm{b}}{N_0} = 15dB$, 5.3bit/s/Hz can be reliably transmitted using a random isometric spreading matrix ($\alpha = 0.95$) while only 3.2bit/s/Hz can be transmitted using a random i.i.d spreading matrix ($\alpha = 0.68$). Notice that the Rayleigh channel capacity for this value of $\frac{E_\mathrm{b}}{N_0}$ is 7 bit/s/Hz.

We also represent in figure 12 the throughput for different values of α $(1, \frac{8}{10}, \frac{3}{4}, \frac{2}{3})$ in the isometric case. The Gaussian non-fading channel capacity is also represented. One can notice that for $\frac{3}{4} < \alpha < 1$, little can be gained by optimizing α.

B. Performance of practical coding schemes.

Our throughput analysis is sustained in this part by simulations. The input bit stream is first serial to parallel converted to produce K sub-streams (see figure 1). Each sub-stream is convolutionnally encoded with a code of rate R and time interleaved. It is assumed that the same code is used for each sub-stream. The resulting bits of sub-stream k are then mapped onto a QPSK constellation to produce component k of symbol vector

[3] As it can be easily shown, the throughput of i.i.d and Haar precoders converges as $\alpha \to \infty$ to the same value. It is solution of the equation $E_{|h|^2}\left[\frac{|h|^2}{\gamma|h|^2 + \frac{1}{\frac{Eb}{No}}}\right] = log_2(e)$. This result can be obtained by noting that $\frac{1}{\sigma^2} = \frac{\gamma Eb}{\alpha No}$. Expanding the term (17) and considering the first term of the Taylor series of (17) together with equation (10) yields the result.

$\mathbf{s}$ to be processed by the precoding matrix $\mathbf{W}_N$. It is assumed that a soft-output Viterbi algorithm is used at the receiver to decode the transmitted bits. The same decoder is applied on each component of vector $\mathbf{s}$ and processes the real and imaginary part of each component of the de-interleaved output $\hat{\mathbf{s}}(n)$ of the MMSE detector. We note that we encode each bit sub-stream by the same code independently in order to implement a reasonably simple Viterbi decoder (to be in accordance with the low computational cost of a MMSE detector). Otherwise the metrics calculation could not be processed on a per component basis due to the inter-component noise correlations introduced by the application of the MMSE equalizer on the received samples. This would exponentially increase the number of states of the Viterbi algorithm trellis. Note that Schramm *et .al* [23] conducted a similar analysis in the case of uplink CDMA with i.i.d random spreading.
In our simulations, different values of α $(1, \frac{8}{10}, \frac{3}{4}, \frac{2}{3})$ are considered and each of them is associated to a convolutional code of rate $R(\alpha)$. In order to compare the corresponding systems at fixed spectral efficiency, we assume that $\alpha R(\alpha) = \frac{1}{2}$ for each α. Figure 13 illustrates the performance of the various coded schemes in the isometric case. Our throughput analysis is in accordance with figure 12 in the sense that the best results are obtained for $\alpha \approx 1$. Figure 14 compares the performance of the i.i.d case for the same values of α. The value $\frac{2}{3}$ provides significantly better performance in accordance with the throughput analysis of figures 9 and 8 in the i.i.d. case.

In summary, we have observed in this section that :

- In a system designed with Haar distributed precoding matrices, the optimum trade-off between the redundancy of the linear precoder and the rate of the convolutional encoder favors values of α close to 1.
- On the other hand, in the i.i.d. case, non negligible spreading redundancy is needed.
- Random isometric precoders optimized w.r.t. α outperform optimized i.i.d. precoders significantly in terms of throughput.

V. Background on Free Probability Theory.

This section aims at introducing some useful notions relative to free probability theory. The interested reader is referred to the comprehensive introduction to this theory in [24].

A more thorough development is given in [18] and in the nice monograph [20].

A. Algebraic Context

Definition 1: A *noncommutative probability space* is a couple (A, ϕ) where A is a non commutative unital algebra (i.e., an algebra having a unit denoted by 1) over $\mathbb{C}$ and $\phi : A \to \mathbb{C}$ is a linear functional such that $\phi(1) = 1$.
When ϕ satisfies $\phi(ab) = \phi(ba)$ it is called a trace. As it will appear below, the role of ϕ can be compared to that of expectation in classical probability theory.

Definition 2: Let (A, ϕ) be a noncommutative probability space. In the context of free probability, a *random variable* is an element a of A. The *distribution* of a is the linear functional ρ_a on $\mathbb{C}[X]$, the algebra of complex polynomials in one variable, defined by $\rho_a(P) = \phi\left(P(a)\right)$.
In particular, the distribution of a non commutative random variable a is characterized by its moments, i.e. by the sequence $(\phi(a^k))_{k \in \mathbb{N}}$. We note that in certain practical cases, the distribution of a non commutative random variable is associated to a real probability measure μ_a (see e.g. the example below) in the sense that $\phi(a^k) = \int_{\mathbb{R}} t^k d\mu_a(t)$ for each $k \in \mathbb{N}$. In this case, the moments of all orders of μ_a are of course finite.

B. An Example of a Noncommutative Probability Space

We shall consider $N \times N$ random matrices whose entries are defined on some common probability space (meant in the classical sense) and have all their moments finite. The noncommutative probability space is obtained by associating to the algebra of these matrices the functional

$$\tau_N(\mathbf{X}) = \frac{1}{N} E\left(\mathrm{tr}(\mathbf{X})\right) = \frac{1}{N} \sum_{i=1}^{N} E(x_{ii}) \tag{18}$$

which is obviously a trace. This space will be denoted by (A_N, τ_N). Suppose $\mathbf{X}$ is a random matrix with real (random) eigenvalues $\lambda_1, \ldots, \lambda_N$. The real random measure

$$\mu = \frac{1}{N} \sum_{i=1}^{N} \delta\left(\lambda_i\right) \tag{19}$$

is called *empirical eigenvalue distribution* of $\mathbf{X}$. The k^{th} moment of this probability measure is $\frac{1}{N}\mathrm{tr}(\mathbf{X}^k) = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^k$. The distribution $\rho_{\mathbf{X}}$ of $\mathbf{X}$ is defined by the fact that its action on each monomial X^k of $\mathbb{C}[X]$ is given by

$$\rho_{\mathbf{X}}(X^k) = \tau_N(\mathbf{X}^k) = \frac{1}{N}(\sum_{i=1}^{N} E(\lambda_i^k))$$

This distribution is of course associated to the probability measure $\mu_{\mathbf{X}}$ defined by

$$\int f(t) d\mu_{\mathbf{X}}(t) = \frac{1}{N}(\sum_{i=1}^{N} E(f(\lambda_i^k))$$

for each bounded continuous function.

C. The Joint Distribution

The notion of distribution introduced in definition 2 can be generalized to the case of multiple random variables. Let a_1 and a_2 be two random variables in a noncommutative probability space (A, ϕ). Consider *noncommutative monomials* in two indeterminates, of the form $X_{i_1}^{k_1} X_{i_2}^{k_2} \dots X_{i_n}^{k_n}$, where for all j, $i_j \in \{1, 2\}$, $k_j \geq 1$ and $i_j \neq i_{j+1}$. The algebra $\mathbb{C}\langle X_1, X_2\rangle$ of noncommutative polynomials with two indeterminates will be the linear span of 1 and these noncommutative monomials. The *joint distribution* of a_1 and a_2 is the linear functional on $\mathbb{C}\langle X_1, X_2\rangle$ satisfying

$$\begin{array}{cccc} \rho & : & \mathbb{C}\langle X_1, X_2\rangle & \longrightarrow & \mathbb{C} \\ & & X_{i_1}^{k_1} X_{i_2}^{k_2} \dots X_{i_n}^{k_n} & \longmapsto & \rho(X_{i_1}^{k_1} X_{i_2}^{k_2} \dots X_{i_n}^{k_n}) = \phi(a_{i_1}^{k_1} a_{i_2}^{k_2} \dots a_{i_n}^{k_n}) \ . \end{array}$$

More generally, denote by $\mathbb{C}\langle X_i \,|\, i \in \{1, \dots, I\}\rangle$ the algebra of non commutative polynomials in I variables, which is the linear span of 1 and noncommutative monomials of the form $X_{i_1}^{k_1} X_{i_2}^{k_2} \dots X_{i_n}^{k_n}$ where $k_j \geq 1$ and $i_1 \neq i_2,\ i_2 \neq i_3,\ \dots,\ i_{n-1} \neq i_n$ are less than or equal to I. The joint distribution of the random variables $\{a_i\}_{i \in \{1, \dots, I\}}$ in (A, ϕ) is the linear functional:

$$\begin{array}{cccc} \rho & : & \mathbb{C}\langle X_i \,|\, i \in \{1, \dots, I\}\rangle & \longrightarrow & \mathbb{C} \\ & & X_{i_1}^{k_1} X_{i_2}^{k_2} \dots X_{i_n}^{k_n} & \longmapsto & \rho(X_{i_1}^{k_1} X_{i_2}^{k_2} \dots X_{i_n}^{k_n}) = \phi(a_{i_1}^{k_1} a_{i_2}^{k_2} \dots a_{i_n}^{k_n}) \ . \end{array} \tag{20}$$

In short, the joint distribution of the noncommutative random variables $\{a_i\}_{i \in \{1, \dots, I\}}$ is completely specified by their joint moments.

D. Freeness

Definition 3: Let (A, ϕ) be a noncommutative probability space. A family $\{A_i\}_{i\in\{1,\ldots,I\}}$ of unital subalgebras of A is called *free* if $\phi(a_1 a_2 \ldots a_n) = 0$ for all n-uples $(a_1, \ldots, a_n)$ verifying:

(i) $a_j \in A_{i_j}$ for some $i_j \leq I$ and $i_1 \neq i_2,\ i_2 \neq i_3,\ \ldots,\ i_{n-1} \neq i_n$.

(ii) $\phi(a_j) = 0$ for all $j = 1, \ldots, n$.

A family of subsets of A is free if the family of unital subalgebras generated by each one of them is free. Random variables $a_1, \ldots, a_n$ are free if the family of subsets $\{\{a_1\}, \ldots, \{a_n\}\}$ is free.

Notice that in the statement of condition (i), only two successive random variables in the argument of $\phi(a_1 a_2 \ldots a_n)$ belong to two different subalgebras. This condition does not forbid the fact that, for instance, $i_1 = i_3$. Note in particular that if a_1 and a_2 belong to two different free algebras, then $\phi(a_1 a_2 a_1 a_2) = 0$ whenever $\phi(a_1) = \phi(a_2) = 0$. This relation cannot of course hold if a_1 and a_2 are two real valued independent random variables (in the classical sense) and if ϕ coincides with the classical mathematical expectation operator. Therefore, freeness cannot be considered as a noncommutative generalization of independence because algebras generated by independent random variables in the classical sense are not free.

Let us make a simple computation involving freeness. A_1 and A_2 being two free subalgebras in A, any two elements a_1 and a_2 of A_1 and A_2 respectively can be written as $a_i = \phi(a_i)1 + a_i'$, so $\phi(a_i') = 0$. Now

$$\phi(a_1 a_2) = \phi\left(\left(\phi(a_1)1 + a_1'\right)\left(\phi(a_2)1 + a_2'\right)\right) = \phi(a_1)\phi(a_2)$$

in other words, the expectations of two free random variables factorize. By decomposing a random variable a_i into $\phi(a_i)1 + a_i'$, the principle of this computation can be generalized to the case of more than two random variables and/or to the case of higher order moments, and one can check that noncommutativity plays a central role there.

Proposition 1 ([24]) *Let $\{A_i\}_{i\in\{1,\ldots,I\}}$ be free subalgebras in (A, ϕ) and let $\{a_1, \ldots, a_n\} \subset A$ be such that for all $j = 1, \ldots, n,$ one has $a_j \in A_{i_j},\ i_j \leq I$. Let Π be the partition of*

$\{1,\ldots,n\}$ associated to the equivalence relation $j \equiv k \Leftrightarrow i_j = i_k$ (i.e. the r.vs. a_j are gathered together according to the free algebras to whom they belong) . For each partition π of $\{1,\ldots,n\}$, let $\phi_\pi = \prod_{\substack{\{j_1,\ldots,j_r\}\in\pi \\ j_1<\ldots<j_r}} \phi(a_{j_1}\ldots a_{j_r})$. There exists universal coefficients $c(\pi,\Pi)$ such that

$$\phi(a_1\ldots a_n) = \sum_{\pi\leq\Pi} c(\pi,\Pi)\phi_\pi$$

where "$\pi \leq \Pi$" stands for "π is finer than Π".

One consequence of this proposition is that given a family of free algebras $\{A_i\}_{i\in\{1,\ldots,I\}}$ in A, only restrictions of ϕ to the algebras A_i are needed to compute $\phi(a_1,\ldots,a_n)$ for any $a_1,\ldots,a_n \in A$ such that for all $j = 1,\ldots,n$, one has $a_j \in A_{i_j}$, $i_j \leq I$. The problem of computing explicitly the universal coefficients $c(\pi,\Pi)$ has been solved using a combinatorial approach.

E. Free Multiplication

Let μ and ν be two compactly supported probability measures on $[0,\infty]$. Then [20], it always exists two free random variables a_1 and a_2 in some noncommutative probability space (A,ϕ) having their distributions associated to μ and ν respectively. One can see that the distribution of the random variable a_1a_2 depends only on μ and on ν. The reason for this is the following: definition 2 says that distribution of a_1a_2 is fully characterized by the moments $\phi\left((a_1a_2)^n\right)$. To compute these moments, we would just need the restriction of ϕ to the algebras generated by $\{a_1\}$ and $\{a_2\}$, in other words, $\phi\left((a_1a_2)^n\right)$ depends on the moments of a_1 and a_2 only. It can be shown that the distribution of a_1a_2 is associated to a probability measure called *free multiplicative convolution* of the distributions μ and ν of these variables. It is denoted by $\mu \boxtimes \nu$ and is compactly supported on $[0,\infty]$ ([18, p. 30]) . Multiplicative free convolution is commutative, and moments of $\mu \boxtimes \nu$ are related in a universal manner to the moments of μ and to those of ν. It happens that direct computation of the moments of a multiplicative free convolution is hardly practicable. It is feasible on the other hand using tools of analytic function theory. By the use of the so-called S-transform introduced in [25], multiplicative free convolution is converted into a mere multiplication of power series:

Proposition 2: *Given a probability measure μ on $\mathbb{R}$ with compact support, let $\psi_\mu(z)$*

be the formal power series defined by

$$\psi_\mu(z) = \sum_{k\geq 1} z^k \int t^k \; d\mu(t) = \int \frac{zt}{1-zt} \; d\mu(t) \; . \tag{21}$$

Let χ_μ *be the unique function analytic in a neighborhood of 0, satisfying*

$$\chi_\mu(\psi_\mu(z)) = z \tag{22}$$

for $|z|$ *small enough. Let*

$$S_\mu(z) = \chi_\mu(z) \; \frac{1+z}{z} \; . \tag{23}$$

S_μ *is called the* S-transform of μ, *and the S-transform* $S_{\mu\boxtimes\nu}$ *of* $\mu \boxtimes \nu$ *is given by*

$$S_{\mu\boxtimes\nu} = S_\mu S_\nu \; .$$

There is also a result in the same vein for additive free convolution. It will not be needed in this paper but the interested reader is referred to [18].

F. Free Probability and Random Matrices

Voiculescu discovered very important relations between free probability theory and random matrix theory. Random matrices are typical noncommutative random variables as can be seen in the example in paragraph V-B. In [26], it is shown that certain independent matrix models exhibit asymptotic free relations.

Definition 4: Let $\{\mathbf{X}_{N,i}\}_{i\in\{1,\ldots,I\}}$ be a family of random $N \times N$ matrices that belong to the noncommutative probability space (A_N, τ_N) defined in paragraph V-B. Their joint distribution is said to have a limit distribution ρ on $\mathbb{C}\langle X_i \,|\, i \in \{1, \ldots, I\}\rangle$ as $N \to \infty$ if

$$\rho(X_{i_1}^{k_1} \ldots X_{i_n}^{k_n}) = \lim_{N\to\infty} \tau_N(\mathbf{X}_{N,i_1}^{k_1} \ldots \mathbf{X}_{N,i_n}^{k_n})$$

for any noncommutative monomial in $\mathbb{C}\langle X_i \,|\, i \in \{1, \ldots, I\}\rangle$.
Consider the particular case where $I = 1$ (we denote $\mathbf{X}_{N,1}$ by $\mathbf{X}_N$ to simplify the notations), and assume $\mathbf{X}_N$ has real eigenvalues and that the distribution of $\mathbf{X}_N$ has a limit distribution ρ. Then, for each $k \geq 0$,

$$\rho(X^k) = \lim_{N\to+\infty} \int t^k d\mu_{\mathbf{X}_N}(t) \tag{24}$$

where $\mu_{\mathbf{X}_N}$ is the measure associated to the distribution of $\mathbf{X}_N$.

Remark 9: If ρ is associated to a compactly supported probability measure μ, the convergence of the moments of $\mu_{\mathbf{X}_N}$ to the moments of μ expressed by (24) implies that the sequence $(\mu_{\mathbf{X}_N})_{N\in\mathbb{N}}$ converges weakly to μ, i.e.

$$\int f(t)d\mu(t) = \lim_{N\to+\infty} \int f(t)d\mu_{\mathbf{X}_N}(t) \tag{25}$$

for each continuous bounded function $f(t)$ (see e.g. [27]). [4]

Definition 5: The family $\{\mathbf{X}_{N,i}\}_{i\in\{1,\ldots,I\}}$ of random matrices in (A_N, τ_N) is said to be *asymptotically free* if the two following conditions are satisfied:

(i) For every integer $i \in \{1, \ldots, I\}$, $\mathbf{X}_{N,i}$ has a limit distribution on $\mathbb{C}[X_i]$.

(ii) For every family of integers $\{i_1, \ldots, i_n\}$ in $\{1, \ldots, I\}$ verifying $i_1 \neq i_2, \ldots, i_{n-1} \neq i_n$, and for every family of polynomials $\{P_1, \ldots, P_n\}$ in one indeterminate verifying

$$\lim_{N\to\infty} \tau_N \left(P_j(\mathbf{X}_{N,i_j})\right) = 0 \text{ for } j = 1, \ldots, n \tag{26}$$

we have

$$\lim_{N\to\infty} \tau_N \left(\prod_{j=1}^{n} P_j(\mathbf{X}_{N,i_j})\right) = 0 \ . \tag{27}$$

(i) and *(ii)* are together equivalent to the two following conditions: the family $\{\mathbf{X}_{N,i}\}$ has a joint limit distribution that we denote by ρ on $\mathbb{C}\langle X_i \,|\, i \in \{1, \ldots, I\}\rangle$, and the family of algebras $\{\mathbb{C}[X_i]\}_{i\in\{1,\ldots,I\}}$ is free in the noncommutative probability space $(\mathbb{C}\langle X_i \,|\, i \in \{1, \ldots, I\}\rangle, \rho)$.

The kind of asymptotic freeness introduced by Hiai and Petz is more useful for our purpose because it deals with almost sure convergence under the normalized classical matrix traces instead of convergence under the functionals τ_N. Following [20], the family $\{\mathbf{X}_{N,i}\}_{i\in\{1,\ldots,I\}}$ in (A_N, τ_N) is said to have a (non random) limit ρ almost everywhere if

$$\rho(X_{i_1}^{k_1} \ldots X_{i_n}^{k_n}) = \lim_{N\to\infty} \frac{1}{N} tr(\mathbf{X}_{N,i_1}^{k_1} \ldots \mathbf{X}_{N,i_n}^{k_n}) \ \text{ a.s.}$$

for any noncommutative monomial in $\mathbb{C}\langle X_i \,|\, i \in \{1, \ldots, I\}\rangle$. In the case where $N = 1$ and $\mathbf{X}_N$ has real eigenvalues, if the almost sure limit distribution of $\mathbf{X}_N$ is associated to a

[4] Convergence of moments implies weak convergence if the function $z \to \int e^{itz} d\mu(t)$ is analytic in a neighborhood of 0, a condition which is clearly met is μ is compactly supported.

compactly supported probability measure μ, this condition means that

$$\int f(t)d\mu(t) = \lim_{N\to+\infty} \frac{1}{N}\sum_{i=1}^{N} f(\lambda_{i,N}) \quad \text{a.s.}$$

for each continuous bounded function $f(t)$. In other words, the empirical eigenvalue distribution of $\mathbf{X}_N$ converges almost surely in distribution to the measure μ.

The family $\{\mathbf{X}_{N,i}\}_{i\in\{1,\ldots,I\}}$ in (A_N, τ_N) is said to be *asymptotically free almost everywhere* if for every $i \in \{1,\ldots,I\}$, $\mathbf{X}_{N,i}$ has a non random limit distribution on $\mathbb{C}[X_i]$ almost everywhere, and if the condition *(ii)* above is satisfied with $\tau_N()$ replaced by $\mathrm{tr}()/N$ in equations (26) and (27) and the limits there are understood as almost sure limits. These conditions imply in particular that $\{\mathbf{X}_{N,i}\}$ has a non random limit distribution almost everywhere on $\mathbb{C}\langle X_i \,|\, i \in \{1,\ldots,I\}\rangle$.
The first concrete random matrix models exhibiting asymptotic freeness were given in [26]. We now give [20, proposition 4.3.9] a useful asymptotic freeness theorem

Theorem 3: *Let $\mathbf{D}_N$ and $\mathbf{E}_N$ be $N\times N$ Hermitian random matrices, and let $\mathbf{\Theta}_N$ be a Haar distributed unitary random matrix independent from $\mathbf{D}_N$ and $\mathbf{E}_N$. Assume that the empirical eigenvalue distributions of $\mathbf{D}_N$ and of $\mathbf{E}_N$ converge almost surely toward compactly supported probability distributions. Then, the family $\{\mathbf{D}_N, \mathbf{\Theta}_N\mathbf{E}_N\mathbf{\Theta}_N^H\}$ is asymptotically free almost everywhere as $N\to\infty$.*

G. An Application Example

Assume that $\mathbf{H}_N$ and $\mathbf{W}_N$ satisfy the assumptions **A1** to **A4** formulated in section II. Put $\mathbf{D}_N = \mathbf{H}_N^H\mathbf{H}_N$, and denote by $d\nu(t) = p(t)dt$ the compactly supported probability distribution of the diagonal entries of this matrix. As relation (3) holds, the empirical eigenvalue distribution of $\mathbf{D}_N$ converges weakly almost everywhere to the probability distribution ν.
$\mathbf{W}_N$ is obtained by extracting K columns from a $N\times N$ Haar distributed random unitary matrix $\mathbf{\Theta}_N$. Without loss of generality, this matrix can be written as $\mathbf{W}_N = \mathbf{\Theta}_N\mathbf{E}_N$, where the $N\times N$ matrix $\mathbf{E}_N$ has the structure $\mathbf{E}_N = \mathrm{diag}([1,\ldots,1,0,\ldots,0])$ with $\mathrm{tr}(\mathbf{E}_N) = K$. Assume that $K/N \to \alpha \le 1$ when $N\to\infty$. Then, the empirical eigenvalue distribution of

$\mathbf{E}_N$ converges to $\mu = \alpha\delta(1) + (1-\alpha)\delta(0)$. Conditions of theorem 3 are satisfied and then the family $\{\mathbf{H}_N^H\mathbf{H}_N\ ,\ \mathbf{W}_N\mathbf{W}_N^H = \mathbf{\Theta}_N\mathbf{E}_N\mathbf{\Theta}_N^H\}$ is asymptotically free.

This allows to derive the almost sure limit eigenvalue distribution of $\mathbf{W}_N\mathbf{W}_N^H\mathbf{H}_N^H\mathbf{H}_N$. To see this, denote by ρ the a.s. limit distribution of $\{\mathbf{W}_N\mathbf{W}_N^H, \mathbf{H}_N^H\mathbf{H}_N\}$ in $\mathbb{C}\langle X_1, X_2\rangle$. In particular, for each $k \in \mathbb{N}$,

$$\rho((X_1X_2)^k) = \lim_{N\to+\infty} \frac{1}{N} tr((\mathbf{W}_N\mathbf{W}_N^H\mathbf{H}_N^H\mathbf{H}_N)^k)\ . \tag{28}$$

Distributions of the monomials X_1 and X_2 are associated to the compactly supported measures μ and ν respectively. As X_1 and X_2 are free, the distribution of X_1X_2 is associated to the measure $\mu \boxtimes \nu$. As $\mu \boxtimes \nu$ is also compactly supported, (28) and remark 9 imply that the empirical eigenvalue distribution of $\mathbf{W}_N\mathbf{W}_N^H\mathbf{H}_N^H\mathbf{H}_N$ converges weakly to $\mu \boxtimes \nu$. The easiest way to evaluate $\mu \boxtimes \nu$ consists in using the fact that the S-transform of $\mu \boxtimes \nu$ is the product of the S-transform of μ by the S-transform of ν. Let us precise this using the same notations as in proposition 2. It is obvious that the function $\psi_\mu(z)$ associated to μ there is given by

$$\psi_\mu(z) = \frac{\alpha z}{1-z}$$

After a simple calculation, we get that its inverse $\chi_\mu(z)$ is given by

$$\chi_\mu(z) = \frac{z}{\alpha + z}\ .$$

From $S_{\mu\boxtimes\nu} = S_\mu S_\nu$ and equation (23), we have

$$\chi_{\mu\boxtimes\nu}(z) = \frac{1+z}{\alpha+z}\chi_\nu(z)\ .$$

We infer that

$$z = \chi_{\mu\boxtimes\nu}\left(\psi_{\mu\boxtimes\nu}(z)\right) = \frac{1+\psi_{\mu\boxtimes\nu}(z)}{\alpha+\psi_{\mu\boxtimes\nu}(z)}\chi_\nu\left(\psi_{\mu\boxtimes\nu}(z)\right)$$

and hence

$$\psi_{\mu\boxtimes\nu}(z) = \psi_\nu\left(z\,\frac{\alpha+\psi_{\mu\boxtimes\nu}(z)}{1+\psi_{\mu\boxtimes\nu}(z)}\right)\ . \tag{29}$$

Therefore, the function $\psi_{\mu\boxtimes\nu}(z)$, which in principle, characterizes the measure $\mu \boxtimes \nu$, is obtained by solving equation (29).

VI. Proof of Theorem 1.

In this section, we assume **A1** to **A4** and establish theorem 1. The easy case $\alpha = 1$ has already been considered in section II, so we just concentrate on the case where $\alpha < 1$. We first recall the following useful result of [19] and [12] :

Lemma 1: Let $\mathbf{z}_N$ be a $N \times 1$ random vector and $\mathbf{B}_N$ a $N \times N$ random matrix independent of $\mathbf{z}_N$. Assume that the elements of $\mathbf{z}_N$ are centered i.i.d. random variables with unit variance and a finite eighth order moment and that $\sup_{N \in \mathbb{N}} \|\mathbf{B}_N\| < +\infty$ where $\| \ \|$ denotes the spectral norm (this spectral norm is said uniformly bounded). Denote by ξ_N the random variable defined by

$$\xi_N = \frac{1}{N}\left(\mathbf{z}_N^H \mathbf{B}_N \mathbf{z}_N \ - \ \mathrm{tr}(\mathbf{B}_N)\right)$$

Then,

$$E\left(|\xi_N|^4\right) \leq C/N^2 \tag{30}$$

where C is independent of N.

Using lemma 1, $\beta_{\mathbf{w}_N}$ in the i.i.d. case can be evaluated immediately : $\mathbf{w}_N$ and

$$\mathbf{F}_N = \mathbf{H}_N^H \left(\mathbf{H}_N \mathbf{U}_N \mathbf{U}_N^H \mathbf{H}_N^H + \sigma^2 \mathbf{I}_N\right)^{-1} \mathbf{H}_N \tag{31}$$

are independent. Therefore, $\beta_{\mathbf{w}_N}$ and $\frac{1}{N}\mathrm{tr}\left(\mathbf{F}_N\right)$ have the same asymptotic behavior. This, in conjunction with the results of [17], allows to evaluate the limit of the SINR. Details are given in the appendix.

The main technical problem encountered in the case where $\mathbf{W}_N$ is isometric follows from the observation that, in contrast with the i.i.d. case, $\mathbf{w}_N$ and $\mathbf{U}_N$ are no more independent. Therefore, $\mathbf{w}_N$ and $\mathbf{H}_N^H \left(\mathbf{H}_N \mathbf{U}_N \mathbf{U}_N^H \mathbf{H}_N^H + \sigma^2 \mathbf{I}_N\right)^{-1} \mathbf{H}_N$ are not independent. It is however possible to show that $\beta_{\mathbf{w}_N}$ and

$$\frac{1}{N-K}\mathrm{tr}\left((\mathbf{I} - \mathbf{U}_N \mathbf{U}_N^H)\mathbf{H}_N^H \left(\mathbf{H}_N \mathbf{U}_N \mathbf{U}_N^H \mathbf{H}_N^H + \sigma^2 \mathbf{I}_N\right)^{-1} \mathbf{H}_N (\mathbf{I}_N - \mathbf{U}_N \mathbf{U}_N^H)\right)$$

have the same behavior. This result can be used to address the more general case where the components of $\mathbf{s}$ have different powers (see [22] for more details). The equal power context of the present paper permits actually to use a simpler method [5]. Instead of studying $\beta_{\mathbf{w}_N}$,

[5] We thank one of the reviewers for having suggested the idea of the present approach.

we rather consider the asymptotic behavior of $\eta_{\mathbf{w}_N}$ given by equation (6) and related to $\beta_{\mathbf{w}_N}$ by equation (5). The idea here is to remark that the $K \times K$ matrix $\mathbf{A}_N$ defined by

$$\mathbf{A}_N = \mathbf{W}_N^H \mathbf{H}_N^H \left(\mathbf{H}_N \mathbf{W}_N \mathbf{W}_N^H \mathbf{H}_N^H + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{H}_N \mathbf{W}_N \tag{32}$$

is unitarily invariant, thanks to the fact that $\mathbf{W}_N$ is extracted from a Haar distributed random unitary matrix. Let $\mathbf{\Omega}_K$ be a $K \times K$ Haar distributed unitary matrix independent of $\mathbf{W}_N$ and $\mathbf{H}_N$. Then, $\mathbf{A}_N$ and $\mathbf{\Omega}_K^H \mathbf{A}_N \mathbf{\Omega}_K$ have the same distribution. Let $\mathbf{e}_K = [0, \ldots, 0, 1, 0, \ldots, 0]^T$ be the $K \times 1$ column selection vector such that $\mathbf{w}_N = \mathbf{W}_N \mathbf{e}_K$, and denote by $\boldsymbol{\omega}_K$ the vector $\mathbf{\Omega}_K \mathbf{e}_K$. Then, because $\mathbf{A}_N$ and $\mathbf{\Omega}_K^H \mathbf{A}_N \mathbf{\Omega}_K$ have the same distribution, so do $\eta_{\mathbf{w}_N} = \mathbf{e}_K^T \mathbf{A}_N \mathbf{e}_K$ and $\boldsymbol{\omega}_K^H \mathbf{A}_N \boldsymbol{\omega}_K = \mathbf{e}_K^T \mathbf{\Omega}_K^H \mathbf{A}_N \mathbf{\Omega}_K \mathbf{e}_K$. In particular, the following identity holds :

$$E \left| \eta_{\mathbf{w}_N} - \frac{1}{K} \mathrm{tr}(\mathbf{A}_N) \right|^4 = E \left| \boldsymbol{\omega}_K^H \mathbf{A}_N \boldsymbol{\omega}_K - \frac{1}{K} \mathrm{tr}(\mathbf{\Omega}_K^H \mathbf{A}_N \mathbf{\Omega}_K) \right|^4 = E \left| \boldsymbol{\omega}_K^H \mathbf{A}_N \boldsymbol{\omega}_K - \frac{1}{K} \mathrm{tr}(\mathbf{A}_N) \right|^4 . \tag{33}$$

It is known ([20]) that vector $\boldsymbol{\omega}_K$ is uniformly distributed on the unit sphere of $\mathbb{C}^K$, and can therefore be written as $\boldsymbol{\omega}_K = \frac{\mathbf{x}_K}{\|\mathbf{x}_K\|}$ for some complex $\mathcal{N}(0, \mathbf{I}_K)$ K–dimensional random vector, where $\mathcal{N}$ stands for the normal distribution. Put $f_N = \boldsymbol{\omega}_K^H \mathbf{A}_N \boldsymbol{\omega}_K - \frac{1}{K} \mathrm{tr}(\mathbf{A}_N)$ and write $f_N = f_{1,N} + f_{2,N}$ where $f_{1,N}$ and $f_{2,N}$ are defined by

$$f_{1,N} = \frac{\mathbf{x}_K^H \mathbf{A}_N \mathbf{x}_K}{\|\mathbf{x}_K\|^2} - \frac{\mathbf{x}_K^H \mathbf{A}_N \mathbf{x}_K}{K}$$

and

$$f_{2,N} = \frac{1}{K} \left(\mathbf{x}_K^H \mathbf{A}_N \mathbf{x}_K - \mathrm{tr}(\mathbf{A}_N) \right)$$

Then, $E|f_N|^4 \leq 8(E|f_{1,N}|^4 + E|f_{2,N}|^4)$. We first use lemma 1 to show that $E|f_{2,N}|^4 = O(N^{-2})$. As $\mathbf{W}_N, \mathbf{H}_N$ and $\mathbf{x}_K$ are independent, $\mathbf{A}_N$ and $\mathbf{x}_K$ are also independent. Moreover, the spectral norm of $\left(\mathbf{H}_N \mathbf{W}_N \mathbf{W}_N^H \mathbf{H}_N^H + \sigma^2 \mathbf{I}_N \right)^{-1}$ is uniformly bounded by $1/\sigma^2$. As $\|\mathbf{H}_N\|$ is uniformly bounded, the inequality $\|\mathbf{CD}\| \leq \|\mathbf{C}\| \, \|\mathbf{D}\|$ implies that $\mathbf{A}_N$ is uniformly bounded. Furthermore, $\mathbf{x}_K$ meets the conditions of lemma 1 since it has Gaussian independent elements. This lemma and the fact that $\frac{K}{N} \to \alpha$ imply that $E|f_{2,N}|^4 = O(N^{-2})$ (replace N by K, $\mathbf{z}_N$ by $\mathbf{x}_K$ and $\mathbf{B}_N$ by $\mathbf{A}_N$). By the first Borel-Cantelli Lemma, $E|f_{2,N}|^4 = O(N^{-2})$ implies that $f_{2,N}$ converges to 0 almost surely.

Let us now study the behavior of $E|f_{1,N}|^4$. $f_{1,N}$ can be written as

$$f_{1,N} = \left(\frac{\mathbf{x}_K^H \mathbf{A}_N \mathbf{x}_K}{K}\right)\left(\frac{K}{\|\mathbf{x}_K\|^2} - 1\right) .$$

As $f_{2,N}$ converges to 0 almost surely,

$$\frac{\mathbf{x}_K^H \mathbf{A}_N \mathbf{x}_K}{K} < 2\frac{\mathrm{tr}(\mathbf{A}_N)}{K} \leq 2\|\mathbf{A}_N\| \quad \text{a.s.}$$

for N large enough. The last inequality comes from the facts that for a given matrix $\mathbf{X}$, $\mathrm{tr}(\mathbf{X}) \leq \|\mathbf{X}\| \,\mathrm{rank}(\mathbf{X})$, and that the rank of $\mathbf{A}_N$ does not exceed K. As $\sup_{N\in\mathbb{N}} \|\mathbf{A}_N\| < \infty$, $\left(\mathbf{x}_K^H \mathbf{A}_N \mathbf{x}_K\right)/K$ is bounded almost everywhere.

$\|\mathbf{x}_K\|^2$ is χ^2 distributed with $2K$ degrees of freedom. Its probability density is the function $\frac{t^K}{K!}e^{-t}$, and a straightforward computation shows that

$$E\left(\frac{K}{\|\mathbf{x}_K\|^2} - 1\right)^4 = O(K^{-2})$$

which is $O(N^{-2})$ if $K/N \to \alpha$.

As $\left(\mathbf{x}_K^H \mathbf{A}_N \mathbf{x}_K\right)/K$ is bounded almost everywhere, we get that $E|f_{1,N}|^4 = O(N^{-2})$. Putting all the pieces together, this implies that $E|f_N|^4 = O(N^{-2})$, and that f_N converges almost surely to 0.

Using equation (33), we get immediately the following result :

Proposition 3: *Assume that $K/N \to \alpha$ when $N \to +\infty$. Then,*

$$\lim_{N\to+\infty}\left(\eta_{\mathbf{w}_N} - \frac{\mathrm{tr}(\mathbf{A}_N)}{K}\right) = 0 \quad \text{a.s.} \tag{34}$$

It remains to study the asymptotic behavior of $\mathrm{tr}(\mathbf{A}_N)/K$, which coincides with the behavior of

$$\frac{1}{\alpha}\lim_{N\to\infty}\frac{1}{N}\mathrm{tr}\left((\mathbf{H}_N\mathbf{W}_N\mathbf{W}_N^H\mathbf{H}_N^H + \sigma^2\mathbf{I}_N)^{-1}\mathbf{H}_N\mathbf{W}_N\mathbf{W}_N^H\mathbf{H}_N^H\right) .$$

For this, we use the results and keep the notations of subsection V-G. It is shown there that the empirical eigenvalue distribution, denoted here $d\theta_N$, of $\mathbf{W}_N\mathbf{W}_N^H\mathbf{H}_N^H\mathbf{H}_N$ converges weakly almost surely to the compactly supported measure $\mu \boxtimes \nu$. But, the eigenvalues of $\mathbf{H}_N\mathbf{W}_N\mathbf{W}_N^H\mathbf{H}_N^H$ of course coincide with the eigenvalues of $\mathbf{W}_N\mathbf{W}_N^H\mathbf{H}_N^H\mathbf{H}_N$. Therefore,

$$\frac{1}{N}\mathrm{tr}\left((\mathbf{H}_N\mathbf{W}_N\mathbf{W}_N^H\mathbf{H}_N^H + \sigma^2\mathbf{I}_N)^{-1}\mathbf{H}_N\mathbf{W}_N\mathbf{W}_N^H\mathbf{H}_N^H\right) \; = \; \int \frac{t}{t+\sigma^2}\, d\theta_N(t) \; . \tag{35}$$

The function $\varphi(t) = t/(t+\sigma^2)$ is continuous and bounded for $t \geq 0$. Therefore, $\eta_{\mathbf{w}_N}$ converges almost surely toward $\overline{\eta}$ defined by

$$\overline{\eta} = \frac{1}{\alpha} \int \frac{t}{t+\sigma^2} \, d\mu \boxtimes \nu(t) \ .$$

This shows that $\beta_{\mathbf{w}_N} = \frac{\eta_{\mathbf{w}_N}}{1-\eta_{\mathbf{w}_N}}$ converges almost surely toward $\overline{\beta} = \overline{\eta}/(1-\overline{\eta})$.

Equation (29) gives an expression for $\psi_{\mu\boxtimes\nu}$, which is related to $\mu \boxtimes \nu$ by (21). It can be easily checked from this last equation that

$$\overline{\eta} = -\frac{1}{\alpha} \ \psi_{\mu\boxtimes\nu}(-\frac{1}{\sigma^2}) \ .$$

Replacing in (29), we have after some simple manipulations

$$-\frac{\alpha\overline{\beta}}{1+\overline{\beta}} = \psi_\nu \left(-\frac{\alpha}{\sigma^2 \left(1+\overline{\beta}(1-\alpha)\right)} \right) \ .$$

The result is obtained after developing ψ_ν according to (21) and replacing $d\nu(t)$ by $p(t)\ dt$.

Appendix

I. Sketch of the proof of theorem 2.

In this appendix, we briefly justify theorem 2. Recall that the SINR $\beta_{\mathbf{w}_N}$ is given by

$$\beta_{\mathbf{w}_N} = \mathbf{w}_N^H \mathbf{F}_N \mathbf{w}_N$$

where $\mathbf{F}_N$ is defined in (31). This matrix can also be written

$$\mathbf{F}_N = (\mathbf{U}_N\mathbf{U}_N^H + \sigma^2(\mathbf{H}_N\mathbf{H}_N^H)^{-1})^{-1} \ .$$

Using assumption **A4**, it is easily seen that $\sup_N ||\mathbf{F}_N|| < \infty$. As $\mathbf{w}_N$ and $\mathbf{F}_N$ are independent, lemma 1 implies that $\beta_{\mathbf{w}_N} - \frac{tr(\mathbf{F}_N)}{N}$ converges to 0 almost surely when $N \to \infty$.

In order to evaluate the behavior of $\frac{tr(\mathbf{F}_N)}{N}$, we use the results of [17]. We first recall that if μ is a probability measure, its Stieltjes transform is the function $m_\mu(z)$ defined by

$$m_\mu(z) = \int \frac{d\mu(t)}{t-z} \ .$$

Assumption **A2'** implies that the empirical eigenvalue distribution of $(\mathbf{H}_N\mathbf{H}_N^H)^{-1}$ converges weakly to the distribution of random variable $\frac{1}{|h|^2}$, i.e. the measure ν defined by $d\nu(t) =$

$\frac{p(1/t)}{t^2}dt$. The results of [17] immediately imply that the empirical eigenvalue distribution of $\mathbf{U}_N\mathbf{U}_N^H + \sigma^2(\mathbf{H}_N\mathbf{H}_N^H)^{-1}$ converges weakly almost surely to a probability measure γ whose Stieltjes transform $m_\gamma(z)$ is defined by the equation

$$m_\gamma(z) = m_\nu\left(z - \frac{\alpha}{(1+m_\gamma(z))}\right) \tag{36}$$

As moreover the eigenvalues of $\mathbf{U}_N\mathbf{U}_N^H + \sigma^2(\mathbf{H}_N\mathbf{H}_N^H)^{-1}$ are greater than $d = \frac{\sigma^2}{c} > 0$ (recall that c is the upper bound of the $|h_i|^2$), the measure γ is carried by $[d, \infty]$. The function $t \to 1/t$ is bounded and continuous on $[d, \infty]$. The weak convergence of the empirical eigenvalue distribution of $\mathbf{U}_N\mathbf{U}_N^H + \sigma^2(\mathbf{H}_N\mathbf{H}_N^H)^{-1}$ thus implies that

$$\lim_{N\to\infty} \frac{tr(\mathbf{F}_N)}{N} = \int_d^\infty \frac{1}{t}\, d\gamma(t)$$

almost surely. Therefore, the SINR converges almost surely to $\int_d^\infty \frac{1}{t}\, d\gamma(t)$, which coincides with $m_\gamma(0)$. Equation (10) follows directly from relation (36).

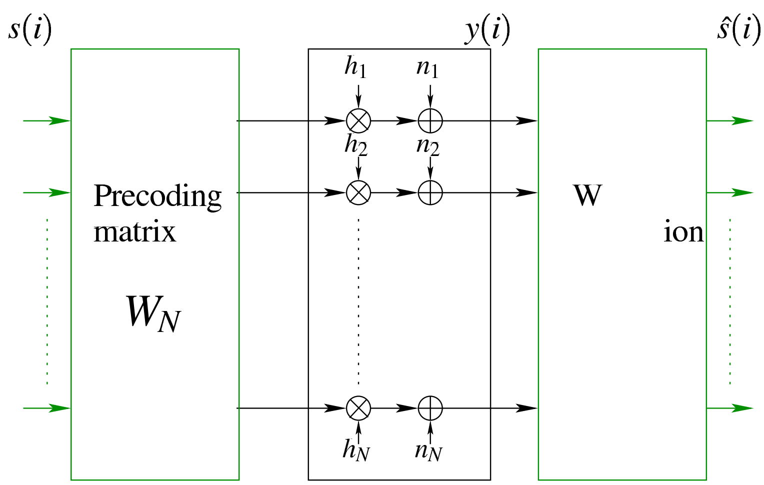

Fig. 1. System Model

References

[1] X. Giraud and J.C Belfiore, "Constellations Matched to the Rayleigh Fading Channel ," *IEEE Trans. on Information Theory*, pp. 106–115, Jan. 1996.

[2] J. Boutros, E. Viterbo, C. Rastello, and J.C. Belfiore, "Good lattice constellations for both Rayleigh fading and Gaussian channel," *IEEE Trans. on Information Theory*, pp. 502–518, Mar. 1996.

[3] J. Boutros and E. Viterbo, "Signal space diversity: a power and bandwidth efficient diversity technique for the Rayleigh fading channel," *IEEE Trans. on Information Theory*, pp. 1453–1467, July 1998.

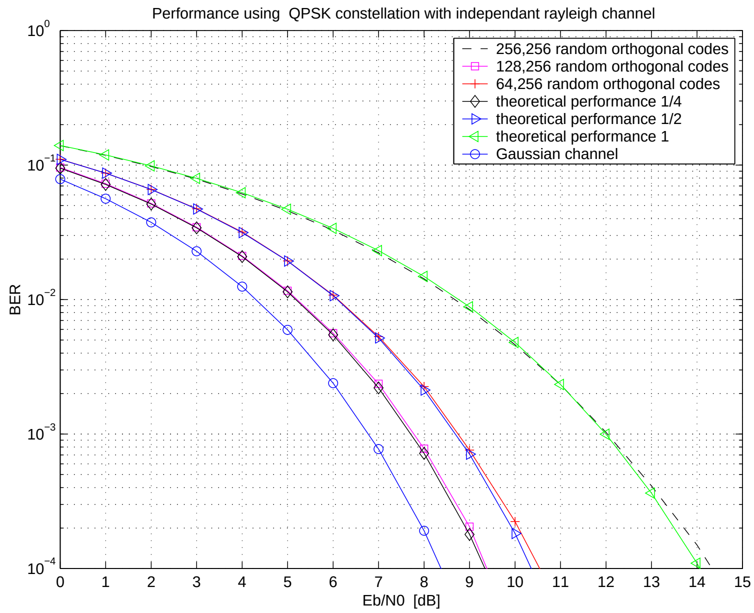

Fig. 2. Probability of error with random isometric precoding matrix

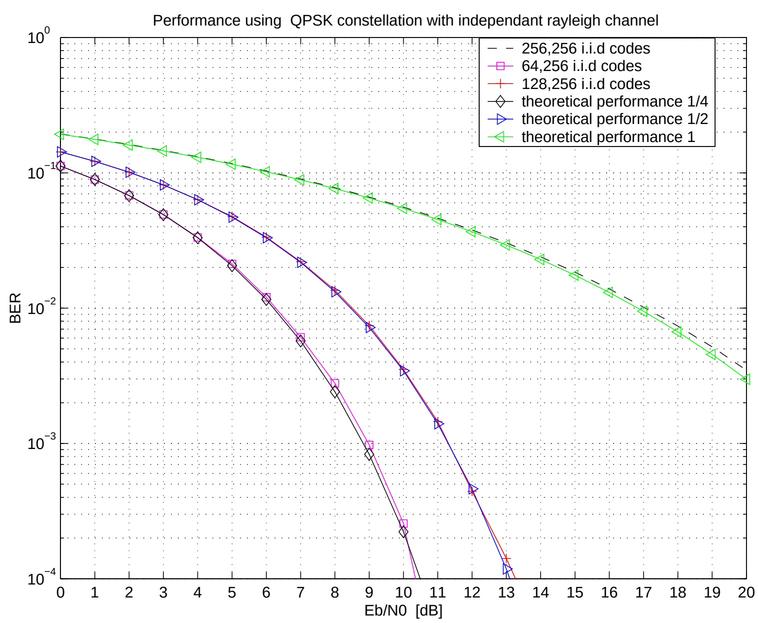

Fig. 3. Probability of error with i.i.d precoding matrix

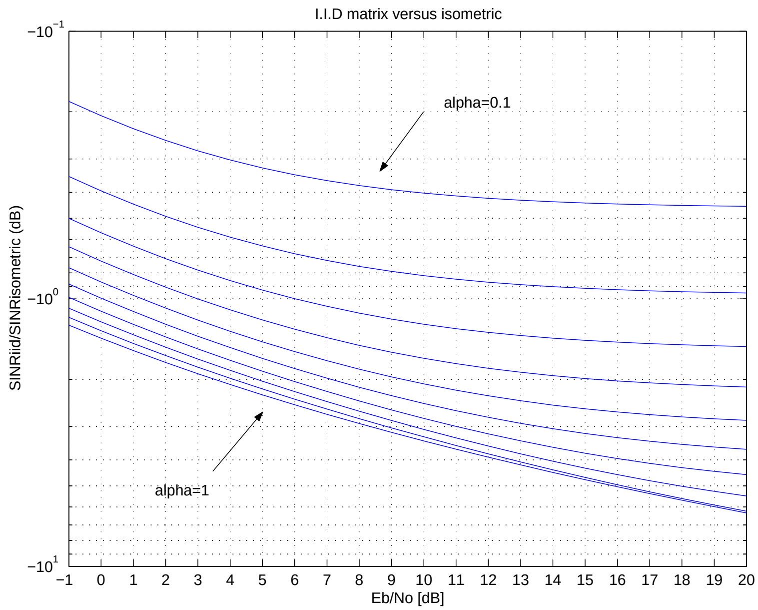

Fig. 4. SINR gain: isometric versus i.i.d

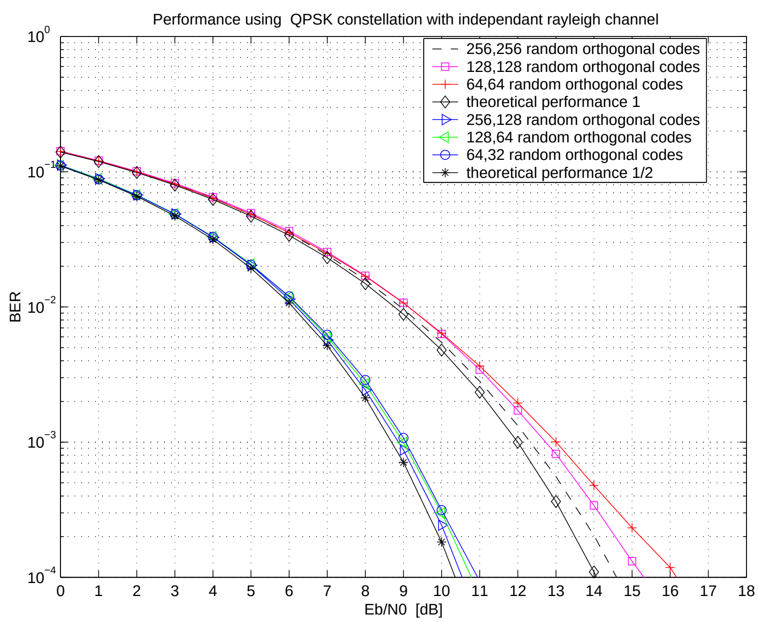

Fig. 5. Influence of the matrix size

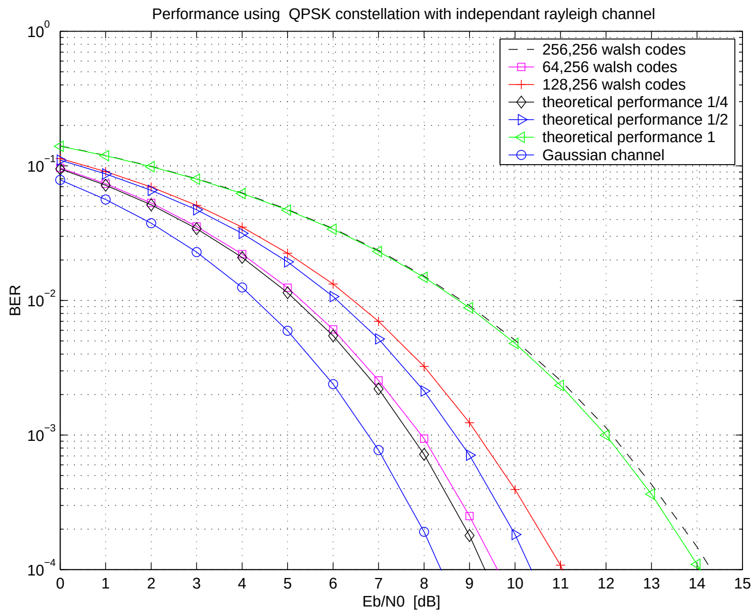

Fig. 6. Performance with Walsh Hadamard matrices

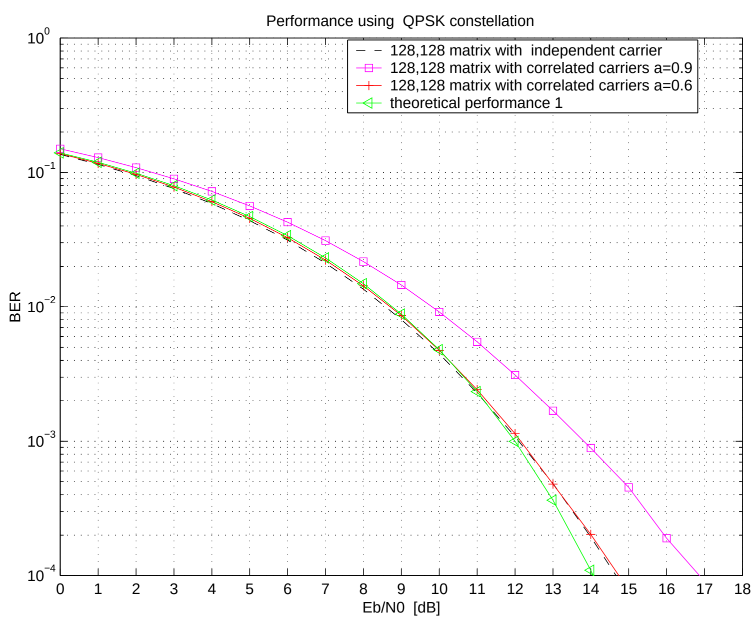

Fig. 7. Correlation Effect

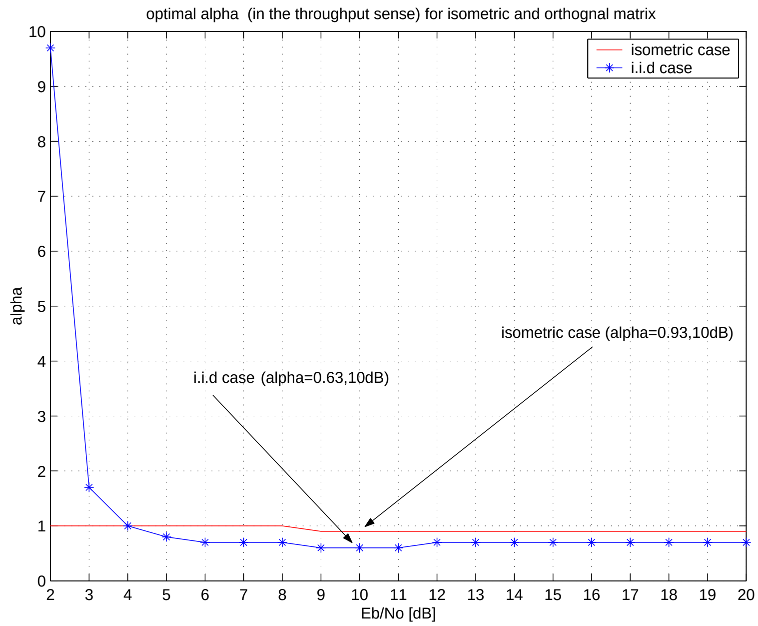

Fig. 8. Optimum α

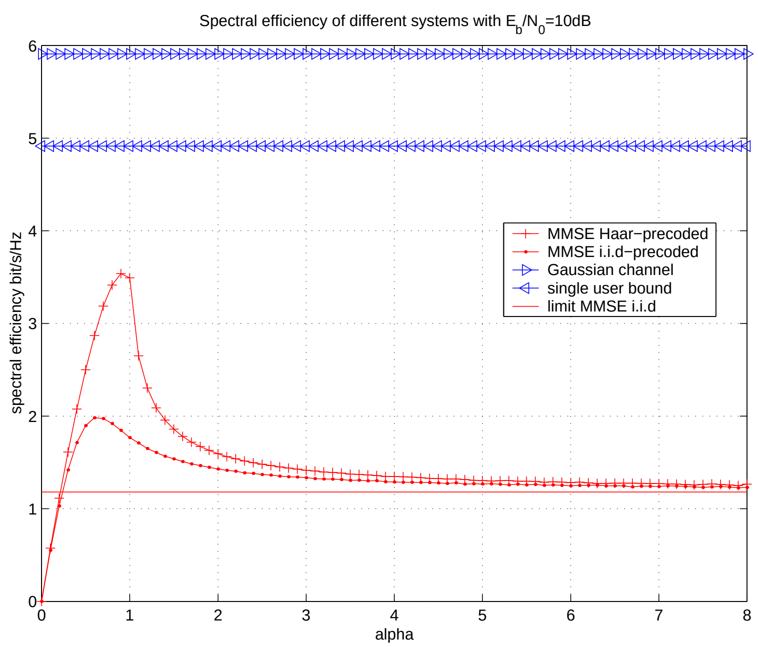

Fig. 9. Throughput at 10dB

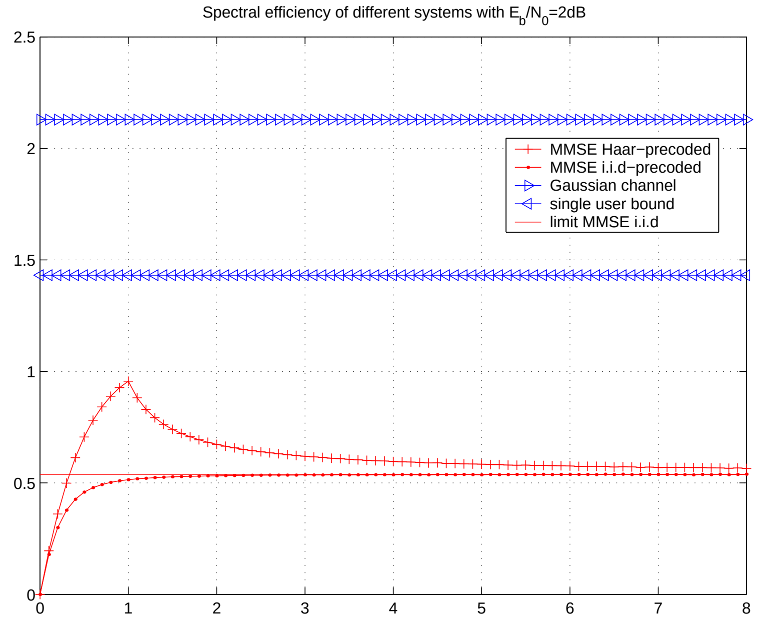

Fig. 10. Throughput at 2dB

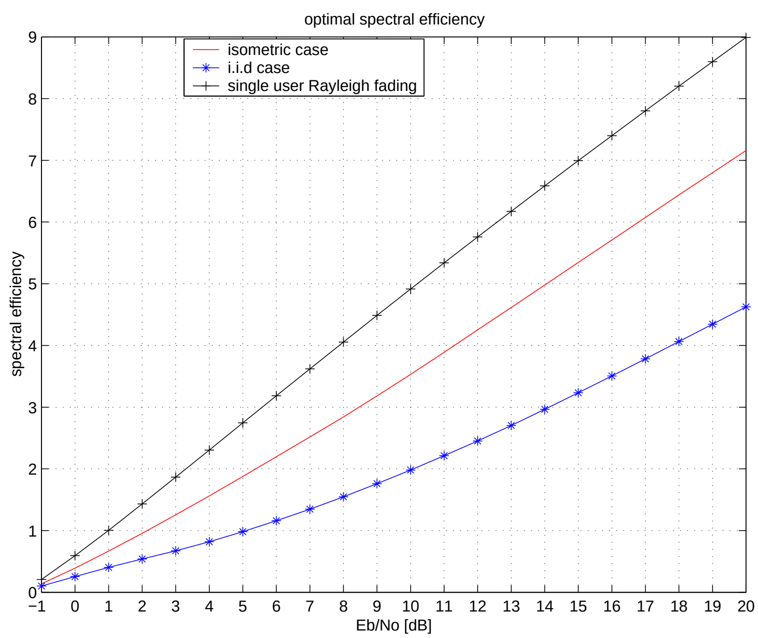

Fig. 11. Optimal capacity

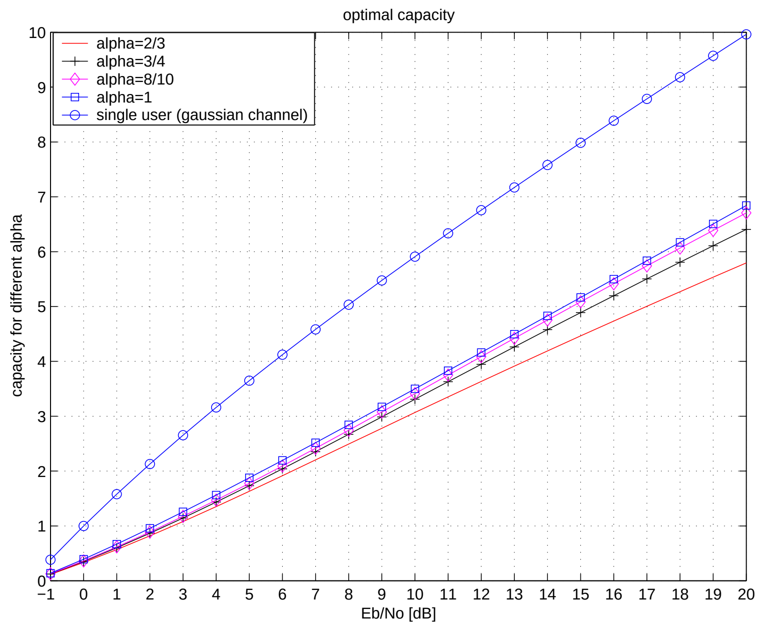

Fig. 12. Capacity versus $\frac{E_b}{N_0}$ for different α and for isometric precoding matrix

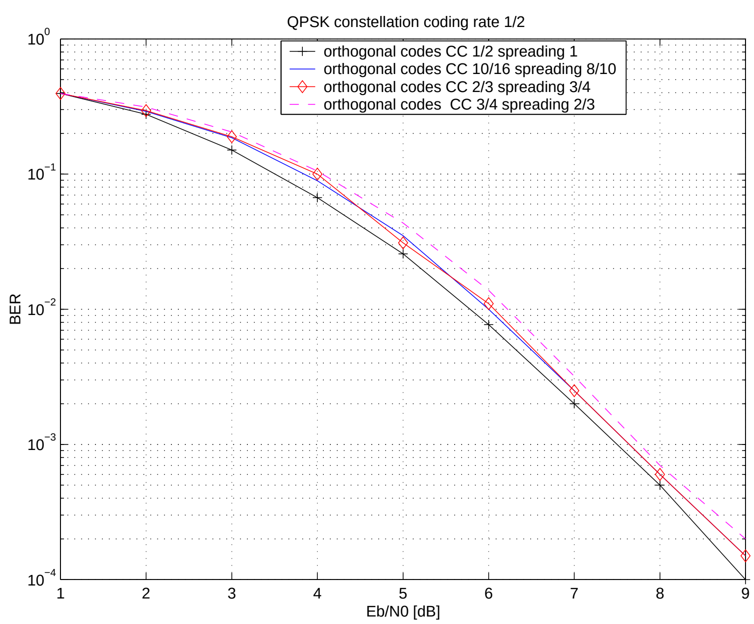

Fig. 13. Various convolutional coding rates for isometric precoding matrix

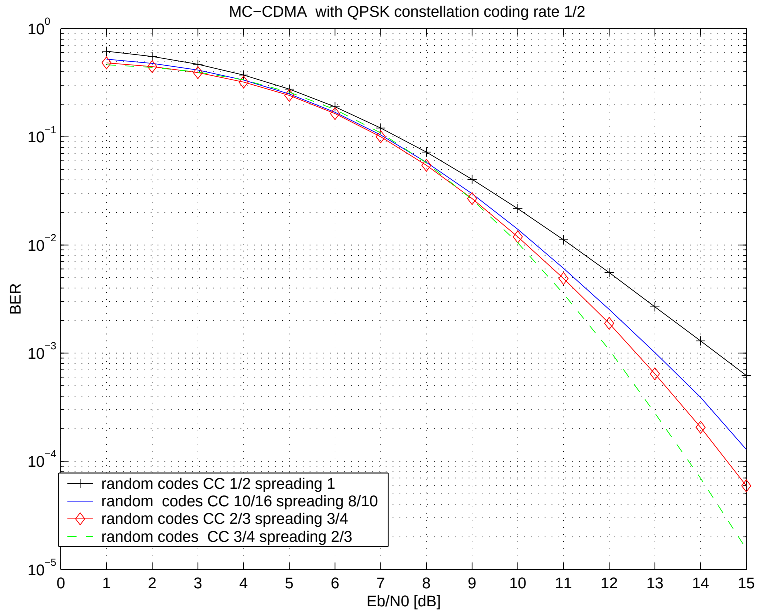

Fig. 14. Various convolutional coding rates for i.i.d precoding matrix

[4] J.A.C. Bingham, "Multicarrier Modulation for Data Transmission: An Idea Whose Time Has Come," *IEEE Communications Magazine*, vol. 28, no. 5, pp. 5–14, May 1990.

[5] K. Fazel, "Performance of CDMA/OFDM for mobile communication system," in *International Conference on Universal Personal Communications*, 1993, vol. 2, pp. 975–979.

[6] Z. Wang and G.B. Giannakis, "Linearly Precoded or Coded OFDM against Wireless Channel Fades ?," in *Proceedings of the 3rd Workshop on Signal Processing Advances in Wireless Communications*, Mar. 2001, pp. 267–270.

[7] M. Debbah, P. Loubaton, and M. de Courville, "Spread OFDM Performance with MMSE Equalization," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, USA, May 2001.

[8] J. G. Proakis, *Digital Communications*, Mc Graw Hill, New York, USA, 3rd ed., 1995.

[9] H. Poor and S. Verdú, "Probability of Error in MMSE Multiuser Detection," *IEEE Trans. on Information Theory*, vol. 43, no. 3, pp. 858–871, May 1997.

[10] J. Zhang, E.K.P. Chong, and D.N.C. Tse, "Output MAI Distributions of Linear MMSE Multiuser Receivers in DS-CDMA Systems," *IEEE Trans. on Information Theory*, vol. 47, no. 3, pp. 1128–1144, Mar. 2001.

[11] S. Verdu and S. Shamai, "Spectral Efficiency of CDMA with Random Spreading ," *IEEE Trans. on Information Theory*, pp. 622–640, Mar. 1999.

[12] J. Evans and D.N.C Tse, "Large System Performance of Linear Multiuser Receivers in Multipath Fading Channels," *IEEE Trans. on Information Theory*, pp. 2059–2078, Sept. 2000.

[13] D.N.C Tse and S. Hanly, "Linear Multi-User Receiver: Effective Interference, Effective Bandwidth and User Capacity," *IEEE Trans. on Information Theory*, vol. 45, no. 2, pp. 641–657, Mar. 1999.

[14] E. Biglieri, G. Caire, G. Taricco, and E. Viterbo, "How Fading Affects CDMA: An Asymptotic Analysis with Linear Receivers," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 2, pp. 191–201, Feb. 2001.

[15] S. Shamai and S. Verdu, "The impact of frequency-flat fading on the spectral efficiency of CDMA ," *IEEE Trans. on Information Theory*, pp. 1302–1327, May 2001.

[16] R. R. Müller, "Multiuser Receivers for Randomly Spread Signals: Fundamental Limits With and Without Decision-Feedback," *IEEE Trans. on Information Theory*, pp. 268–283, Jan. 2001.

[17] J.W. Silverstein and Z.D. Bai, "On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices," *J. Multivariate Anal.*, vol. 54, no. 2, pp. 175–192, 1995.

[18] D.V. Voiculescu, K.J. Dykema, and A. Nica, *Free Random Variables*, American Mathematical Society, CRM Monograph Series, Volume 1, Providence, Rhode Island, USA, 1992.

[19] Z.D. Bai and J.W. Silverstein, "No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Large Dimensional Sample Covariance Matrices," vol. 26, no. 1, pp. 316–345, 1998.

[20] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*, vol. 77 of *Mathematical Surveys and Monographs*, AMS, 2000.

[21] M. Debbah, P. Loubaton, and M. de Courville, "Linear Precoders for OFDM wireless communications with MMSE equalization: facts and results," in *Proceedings of EUSIPCO, European Sig. Proc. Conf.*, Toulouse, France, Sept. 2002.

[22] J.M. Chaufray, W. Hachem, and Ph. Loubaton, "Asymptotical Analysis of Optimum and Sub-Optimum CDMA Downlink MMSE Receivers," in *IEEE International Symposium on Information Theory*, Lausanne, Switzerland, July 2002, can be downloaded at `http://www-syscom.univ-mlv.fr/~loubaton/index.html`.

[23] P. Schramm and R. Muller, "Spectral Efficiency of CDMA Systems with Linear MMSE Interference Suppression," *IEEE Trans. on Communications*, vol. 47, pp. 722–731, May 1999.

[24] Ph. Biane, "Free Probability for Probabilists," *http://www.dma.ens.fr/~biane/*, 2000, download by choosing "Des notes d'introduction aux probabilités libres".

[25] D.V. Voiculescu, "Multiplication of Certain Non-Commuting Random Variables," *J. Operator Theory*, vol. 18, pp. 223–235, 1987.

[26] D.V. Voiculescu, "Limit Laws for Random Matrices and Free Products," *Inventiones mathematicae*, vol. 104, pp. 201–220, 1991.

[27] P. Billingsley, *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics. John Wiley, 3rd edition, 1995.