MIMO Channel Modelling and the Principle of Maximum Entropy

Mérouane Debbah[1] and Ralf Müller[2]

Abstract

In this paper[1], we devise theoretical grounds for constructing channel models for Multi-input Multi-output (MIMO) systems based on information theoretic tools. The paper provides a general method to derive a channel model which is consistent with one's state of knowledge. The framework we give here has already been fruitfully explored with success in the context of Bayesian spectrum analysis and parameter estimation. For each channel model, we conduct an asymptotic analysis (in the number of antennas) of the achievable transmission rate using tools from random matrix theory. A central limit theorem is provided on the asymptotic behavior of the mutual information and validated in the finite case by simulations. The results are both useful in terms of designing a system based on criteria such as quality of service and in optimizing transmissions in multiuser networks[2].

Keywords

MIMO, Entropy, Bayesian Probability Theory, Random Matrices, Channel Modelling, Antenna Arrays

[1]Mobile Communications Group, Institut Eurecom, 2229 Route des Cretes, B.P. 193, 06904 SOPHIA ANTIPOLIS CEDEX, debbah@eurecom.fr

[2]Forschungszentrum Telekommunikation Wien (FTW) Vienna, Austria, mueller@ftw.at

[1]This work is part of the European FLOWS (Flexible Convergence of Wireless Standards and Services) project and related documents can be downloaded at http://www.flows-ist.org

[2]This paper was presented in part at the 37th Annual Asilomar Conference on Signals, Systems and Computers, Pacific Grove, California, USA, 2003 and the IEEE International Symposium on Signal Processing and Information Technology, Darmstadt, Germany, 2003

Contents

I. Introduction

The problem of modelling channels is crucial for the efficient design of wireless systems [1]. The wireless channel suffers from constructive/destructive interference signaling [2], [3] and yields a randomized channel with certain statistics to be discovered. Recently ([4], [5]), the need to increase spectral efficiency has motivated the use of multiple antennas at both the transmitter and the receiver side. Hence, if the MIMO link is characterized by an $n_r \times n_t$ matrix with i.i.d Gaussian entries which are perfectly known to the receiver, it has been proved [6] that the ergodic capacity increase is $\min(n_r,n_t)$ bits per second per hertz for every 3dB increase at high signal-to-noise ratio (SNR). However, for realistic channel models, results are still unknown and may seriously put into doubt the MIMO hype. As a matter of fact, the actual design of efficient codes is tributary of the channel model available: the transmitter has to know in what environment the transmission occurs in order to provide the codes with the adequate properties: as a typical example, in Rayleigh fading channels, when coding is performed, hamming distance plays a central role whereas maximizing Euclidean distance is the commonly approved design criteria for Gaussian channels [7], [8].

As a consequence, channel modelling is the key in better understanding the limits of transmissions in wireless environments. Questions of the form: "what is the highest transmission rate on a propagation environment where we only know the mean of each path, the variance of each path and the directions of arrival?" are crucially important. Their answers will be decisive on the use of MIMO technologies for a given state of knowledge.

Before going further, let us first introduce the modelling constraints. We assume that the transmission takes place between a mobile transmitter and receiver. The transmitter has n_t antennas and the receiver has n_r antennas. Moreover, we assume that the transmitted signal propagates through a linear filter channel. Finally, we assume that the interfering noise is additive, white, and Gaussian distributed.

The transmitted signal and received signal are related as

$$\mathbf{y}(t) = \sqrt{\frac{\rho}{n_t}} \int \mathbf{H}_{n_r \times n_t}(\tau, t)\mathbf{x}(t-\tau)d\tau + \mathbf{n}(t) \tag{1}$$

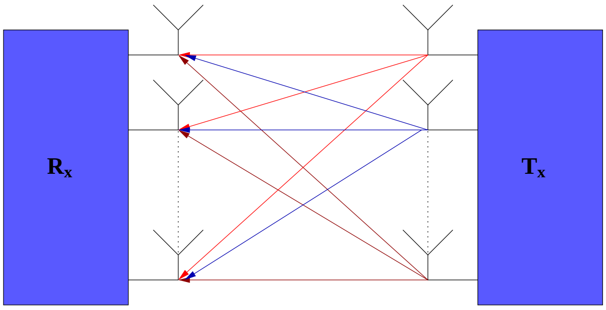

Fig. 1. MIMO channel representation.

with

$$\mathbf{H}_{n_r \times n_t}(\tau, t) = \int \mathbf{H}_{n_r \times n_t}(f, t) e^{j2\pi f\tau} df \quad (2)$$

where ρ is the received SNR, t, f and τ denote respectively time, frequency and delay, $\mathbf{y}(t)$ is the $n_r \times 1$ received vector, $\mathbf{x}(t)$ is the $n_t \times 1$ transmit vector, $\mathbf{n}(t)$ is an $n_r \times 1$ additive spatially and temporally white Gaussian noise vector with unit variance.

For the rest of the paper, we address the channel, without loss of generality, in its frequency domain representation. We provide some theoretical grounds to model the frequency response matrix $\mathbf{H}(f, t)$ based on a given state of knowledge. Knowing only certain properties of the channel such as directions of arrival (DoA), directions of departure (DoD), bandwidth, center frequency, number of transmitting and receiving antennas, number of chairs in the room, etc. we investigate how to attribute a joint probability distribution to the entries $h_{ij}(f, t)$ of the matrix $\mathbf{H}(f, t)$.

This problem can be answered in light of Bayesian probability theory. Bayesian probability theory has led to a profound theoretical understanding of various scientific areas [9], [10], [11], [12], [13], [14], [15], [16] and has shown the potential of entropy as a measure of our degree of knowledge when encountering a new problem. The principle of maximum entropy is at present the clearest theoretical justification in conducting scientific inference: we do not need a model, entropy maximization creates a model for us out of the information available [10], [11]. Choosing the distribution with greatest entropy avoids the arbitrary introduction or assumption of information that is not available.

In this contribution, we take the Bayesian viewpoint in which channel modelling represents our knowledge of reality [17]. We provide answers to the following question: what is the best model one can construct given some state of knowledge. This is admittedly a vague question since there is no strict definition of what is meant by best. In this contribution, our aim is to derive a model which reflects our state of knowledge. We need a measure of uncertainty which expresses the constraints of our knowledge and the desire to leave the unknown parameters to lie in an unconstrained space. To this end, many possibilities are offered to us to express our uncertainty. However, we need an information measure which is consistent—it complies to certain common sense desiderata as expressed in [18]—and is easy to manipulate. We need a simple general principle for translating information into probability assignment. Entropy is that measure of information that fulfills this criteria. Back in 1980, Shore et al. [18] proved that the principle of maximum entropy is the correct method of inference when given new information in terms of expected values. They proved that maximizing entropy is correct in the following sense: maximizing any function but entropy will lead to inconsistencies unless that function and entropy have the same maximum. Thus, aiming for consistency, we can maximize entropy without loss of generality. The consistency argument is at the heart of scientific inference and can be expressed through the following axiom:

Axiom 1: If the prior information $\mathbf{I}_1$ which the channel model $\mathbf{H}_1$ is based on can be equated to the prior information $\mathbf{I}_2$ of the channel model $\mathbf{H}_2$ then both models should be assigned the same probability distribution $P(\mathbf{H}) = P(\mathbf{H}_1) = P(\mathbf{H}_2)$.

Any other procedure would be inconsistent in the sense that, by changing indices 1 and 2, we could then generate a new problem in which our state of knowledge is the same but in which we are assigning different probabilities.

Moreover, the success over the years of the maximum entropy approach, see Boltzmann's kinetic gas law, [19] for the estimate of a single stationary sinusoidal frequency, [12] for estimating the spectrum density of a stochastic process subject to autocorrelation constraints, [20] for estimating parameters in the context of image reconstruction and restoration problems, [21] for applying the maximum entropy principle on solar proton event peak fluxes in order to determine the least biased distribution, has shown that this is the right tool to express our uncertainty. Recently, the maximum entropy principle has even been advocated to describe wave propagation. In [22], Franceschetti et al. show that the probability laws that describe electromagnetic waves are simply maximum entropy distributions with appropriate moment constraints.

It is noteworthy to say that if a prior distribution Q of the estimated distribution P is available in addition to the expected values constraints, then the principle of minimum cross-entropy which generalizes maximum entropy should be applied.

In this paper, we provide guidelines for creating models from an information theoretic point of view and therefore make extensive use of the principle of maximum entropy together with the principle of consistency. For various states of knowledge, such as DoA, DoD, the number of scatterers, the powers of the steering directions, a model is derived. In addition, the asymptotic mutual information for perfect channel knowledge at the receiver side is calculated. The general procedure is explained with the simplest example of no knowledge except for energy constraints on the path gains in Section III. Various degrees of knowledge on the DoA, the DoD, and the powers of steering directions are addressed in Section IV. Models for additional knowledge on path delay times, frequency-selectivity, and time-variance are given in Section V. In Section VI, channel models developed in literature on considerations different from the maximum entropy framework are linked to our models by determining which states of knowledge are needed to make these models be solutions of entropy maximization. In Section VII, we address some limitations of the maximum entropy approach when it comes to calculation of channel capacities of the modeled channel, before we draw some conclusions in Section VIII.

Throughout the paper, for sake of simplicity, we will often write $\mathbf{H}$ instead of $\mathbf{H}(f,t)$ without forgetting the dependency on frequency and time. In the following, upper and lower boldface symbols will be used for matrices and column vectors, respectively. $(.)^T$ will denote the transpose operator, $(.)^\star$ conjugation and $(.)^H = \left((.)^T\right)^\star$ hermitian transpose. ln is the natural logarithm such $\ln(e) = 1$. When this notation is used, the mutual information is given in nats/s. When the notation $\log_2(x) = \frac{\ln(x)}{\ln(2)}$ is used, the results are given in bits/s. The Stieltjes Transform $m(z)$ of a distribution F is defined as

$$m(z) = \int \frac{1}{\lambda - z} dF(\lambda) \tag{3}$$

$\delta(x)$ is the Dirac distribution whereas δ_{im} denotes the Kronecker product:

$$\delta_{im} = \begin{cases} 1 & \text{if } i = m \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Moreover, define

$$\mathbf{1}_{[0,2\pi]}(x) = \begin{cases} 1 & \text{if } x \in [0, 2\pi] \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

The operator $\text{vec}(\mathbf{H})$ stacks all the columns of matrix $\mathbf{H}$ into a single column.

II. Preliminaries

For almost all the models we construct based on maximum entropy and consistency arguments we will derive analytical expressions for the mutual information. For the convenience of the reader, we use this section to precise the notions of mutual informations which are used later on.

A. MIMO Considerations

Let us first review the pioneering work of Teletar[6] (later published as [23]) that triggered research in multi-antenna systems. In this paper, Telatar develops the channel capacity of a general MIMO channel. Assuming perfect knowledge of $\mathbf{H}$ at the receiver, the ergodic capacity of an $n_r \times n_t$ MIMO channel with input covariance matrix $\mathbf{Q} = \mathbb{E}(\mathbf{x}\mathbf{x}^H)$ is

$$\overline{C} = \max_{\mathbf{Q}} \mathbb{E}\left(C(\mathbf{Q})\right) \tag{6}$$

with

$$C(\mathbf{Q}) = \log_2 \det \left(\mathbf{I}_{n_r} + \frac{\rho}{n_t} \mathbf{H}\mathbf{Q}\mathbf{H}^H \right) \tag{7}$$

where the maximization is over the set of positive semi-definite hermitian matrices $\mathbf{Q}$ satisfying the power constraint $\text{tr}(\mathbf{Q}) \leq P$ and the expectation is with respect to the random channel matrix. In the original paper [6], Telatar exploits the isotropical property of Gaussian i.i.d $\mathbf{H}$ to show that in this case, ergodic capacity is achieved with $\mathbf{Q} = \mathbf{I}$.

In correlated fading, $C(\mathbf{I})$ is called the average mutual information with covariance $\mathbf{Q} = \mathbf{I}$. In general [24], [25], [26], [27], [28], [29], [30], capacity is not close to this mutual information except for certain particular cases, see [23], [29]. Often $C(\mathbf{I})$ underestimates the achievable rate: indeed, even though the channel realization is not known, the knowledge of the channel model statistics can be taken into account in order to optimize the coding scheme at the transmitter.

The dependency of the optimum $\mathbf{Q}$ on the distribution of $\mathbf{H}$ is one more motivation to study the probability distribution of the matrix $\mathbf{H}$. Such distributions can be very helpful for system design. One of the visions of future wireless communications the authors would like to advocate is the following: suppose that the environment (dense buildings, field, street, number of chairs, etc.), is provided to the user's terminal. This can be automated by downloading localization information from the base station. Based on that state of knowledge, a channel model is created online using the maximum entropy approach which incorporates only the available information and not more. The transmitted signal and the coding scheme is then online optimized for that specific scenario, e.g. by deriving new rank and determinant criteria for example. Such a service could be called "user customized channel model coding service". From a software defined radio perspective, this scenario is completely viable.

B. Outage Mutual Information

For a wireless content provider, the most important criteria is the quality of service to be delivered to customers. This quality of service can be quantified through measures such as outage capacity: if $q = 10^{-2}$ is the outage probability of having an outage capacity of R, then this means that the provider is able to ensure a rate of R in 99% of the cases. Since the channels are rarely ergodic, the derivations of ergodic capacities are of limited use for content providers.

If the channels are static, there is only one channel realization and an outage capacity defined as

$$C_q = \max_{\mathbf{Q}} \sup\{R : \Pr[C(\mathbf{Q}) < R] \leq q\} \tag{8}$$

is the measure of interest.

The covariance matrix $\mathbf{Q}$ which optimizes the ergodic capacity does not necessarily optimize the outage capacity. If the channel distribution is known, then the transmitter should optimize its signaling to this distribution even if the channel realization is unknown. Since this is not an obvious task, in general, in all the following we will derive the outage mutual information with Gaussian input covariance matrix $\mathbf{Q} = \mathbf{I}$. *In general, this is only a lower bound to the outage capacity.* Although not optimum, the mutual information with covariance $\mathbf{Q} = \mathbf{I}$ can be useful in the analysis of systems where the codebook can not be changed according to the wireless environment and therefore remains the same during the whole transmission. For further details on outage capacity the reader is referred to [31], [32], [33], [34], [35].

III. Gaussian i.i.d channel model

A. Model

In this section, we give a precise justification on why and when the Gaussian i.i.d model should be used. We recall the general model:

$$\mathbf{y} = \sqrt{\frac{\rho}{n_t}}\mathbf{H}\mathbf{x} + \mathbf{n}$$

Imagine now that the modeler is in a situation where he has no measurements and no knowledge where the transmission took place. The only thing the modeler knows is that the channel carries some energy E, in other words, $\frac{1}{n_r n_t}\mathbb{E}\left(\sum_{i=1}^{n_r}\sum_{j=1}^{n_t} \mid h_{ij} \mid^2\right) = E$. Knowing only this information, the modeler is faced with the following question: what is the consistent model one can make knowing only the energy E (but not the correlation even though it may exist)? In other words, based on the fact that:

$$\int d\mathbf{H} \sum_{i=1}^{n_r}\sum_{j=1}^{n_t} \mid h_{ij} \mid^2 P(\mathbf{H}) = n_t n_r E \qquad \text{(Finite energy)} \tag{9}$$

$$\int dP(\mathbf{H}) = 1 \qquad \text{(P}(\mathbf{H})\text{ is a probability distribution)} \tag{10}$$

What distribution $P(\mathbf{H})$[3] should the modeler assign to the channel? The modeler would like to derive the most general model complying with those constraints, in other

[3]It is important to note that we are concerned with $P(\mathbf{H} \mid I)$ where I represents the general background knowledge (here the variance) used to formulate the problem. However, for simplicity sake, $P(\mathbf{H} \mid I)$ will be denoted $P(\mathbf{H})$.

words the one which maximizes our uncertainty while being certain of the energy. This statement can simply be expressed if one tries to maximize the following expression using Lagrange multipliers with respect to P:

$$L(P) = -\int d\mathbf{H}P(\mathbf{H})\log P(\mathbf{H}) + \gamma\sum_{i=1}^{n_r}\sum_{j=1}^{n_t}[E - \int d\mathbf{H} \mid h_{ij} \mid^2 P(\mathbf{H})] \\ +\beta\left[1 - \int d\mathbf{H}P(\mathbf{H})\right]$$

If we derive $L(P)$ with respect to P, we get:

$$\frac{dL(P)}{dP} = -1 - \log P(\mathbf{H}) - \gamma\sum_{i=1}^{n_r}\sum_{j=1}^{n_t} \mid h_{ij} \mid^2 - \beta = 0$$

then this yields:

$$\begin{aligned} P(\mathbf{H}) &= e^{-(\beta+\gamma\sum_{i=1}^{n_r}\sum_{j=1}^{n_t}|h_{ij}|^2} \\ &= e^{-(\beta)}\prod_{i=1}^{n_r}\prod_{j=1}^{n_t}\exp-(\gamma \mid h_{ij} \mid^2) \\ &= \prod_{i=1}^{n_r}\prod_{j=1}^{n_t}P(h_{ij}) \end{aligned}$$

with

$$P(h_{ij}) = e^{-(\gamma|h_{ij}|^2+\frac{\beta+1}{n_r n_t})}.$$

One of the most important conclusions of the maximum entropy principle is that while we have only assumed the variance, this assumption imply independent entries since the joint probability distribution $P(\mathbf{H})$ simplifies into products of $P(h_{ij})$. Therefore, based on the previous state of knowledge, the only maximizer of the entropy is the i.i.d one. This does not mean that we have supposed independence in the model. In the generalized $L(P)$ expression, there is no constraint on the independence. Another surprising result is that the distribution achieved is Gaussian. Once again, gaussianity is not an assumption but a consequence of the fact that the channel has finite energy. The previous distribution is the least informative probability density function that is consistent with the previous state of knowledge. When only the energy of the channel is known (but not the frequency bandwidth, nor knowledge of how waves propagate, nor the fact that scatterers exist...) then the only consistent model one can make is the Gaussian i.i.d model. Hence, instead of saying that this model represents a rich scattering environment, it should be more correct to say that the model makes allowance for every case that could be present to happen since we have imposed no constraints besides the energy. The maximum entropy approach is appealing in the sense that if correlated scattering is given as a prior knowledge, then it can be immediately integrated in the channel modelling approach (as a constraint on the covariance matrix for example).

In order to fully derive $P(\mathbf{H})$, we need to calculate the coefficients β and γ. The coefficients are solutions of the following constraint equations:

$$\int d\mathbf{H} \sum_{i=1}^{n_r} \sum_{j=1}^{n_t} \mid h_{ij} \mid^2 P(\mathbf{H}) = n_t n_r E$$

$$\int d\mathbf{H} P(\mathbf{H}) = 1$$

Solving the previous equations yields the following probability distribution:

$$P(\mathbf{H}) = \frac{1}{(\pi E)^{n_r n_t}} \exp\{-\sum_{i=1}^{n_r} \sum_{j=1}^{n_t} \frac{\mid h_{ij} \mid^2}{E}\}$$

Of course, if one has any additional knowledge, then this information should be integrated in the $L(P)$ criteria and would lead to a different result.

As a typical example, suppose that the modeler knows that the frequency paths have different variances such as $\mathbb{E}(\mid h_{ij} \mid^2) = E_{ij}$. Using the same methodology, it can be shown that :

$$P(\mathbf{H}) = \prod_{i=1}^{n_r} \prod_{j=1}^{n_t} P(h_{ij})$$

with $P(h_{ij}) = \frac{1}{\pi E_{ij}} e^{-\frac{|h_{ij}|^2}{E_{ij}}}$. The principle of maximum entropy still attributes independent Gaussian entries to the channel matrix but with different variances.

Suppose now that the modeler knows that the path h_{pk} has a mean equal to $\mathbb{E}(h_{pk}) = m_{pk}$ and variance $\mathbb{E}(\mid h_{pk} - m_{pk} \mid^2) = E_{pk}$, all the other paths having different variances (but nothing is said about the mean). Using as before the same methodology, we show that:

$$P(\mathbf{H}) = \prod_{i=1}^{n_r} \prod_{j=1}^{n_t} P(h_{ij})$$

with for all $\{i,j,(i,j)\neq(p,k)\}$ $P(h_{ij})=\frac{1}{\pi E_{ij}}e^{-\frac{|h_{ij}|^2}{E_{ij}}}$ and $P(h_{pk})=\frac{1}{\pi E_{pk}}e^{-\frac{|h_{pk}-m_{pk}|^2}{E_{pk}}}$. Once again, different but still independent Gaussian distributions are attributed to the MIMO channel matrix.

The previous examples can be extended and applied whenever a modeler has some new source of information *in terms of expected values* on the propagation environment. The case where information is not given in terms of expected values is treated in Section IV. In the general case, if N constraints are given on the expected values of certain functions $\int g_i(\mathbf{H})P(\mathbf{H})d\mathbf{H}=\alpha_i$ for $i=1...N$, then the principle of maximum entropy attributes the distribution [36]

$$P(\mathbf{H})=e^{(-1+\lambda+\sum_{i=1}^{N}\lambda_i g_i(\mathbf{H}))} \tag{11}$$

where the values of λ and λ_i for $i\in\{1,\ldots,N\}$ can be obtained by solving the constraint equations.

B. Asymptotic Mutual Information

In [6], Telatar calculated the ergodic capacity for the i.i.d channel model when the channel is known at the receiver only. For the outage probability no closed form solution is known. However, in the asymptotic limit, i.e. letting the number of transmit antennas and the number of receive antennas grow large with fixed ratio, the following result was shown by Kamath et al. [31] making use of recent results in random matrix theory [37].

Theorem 1: With the Gaussian i.i.d model, as $n_t\to\infty$ with $n_r=\gamma n_t$, $C(n_t,n_r,\rho)-n_t\mu(\gamma,\rho)$ converges in distribution to a $N(0,\sigma^2(\gamma,\rho))$ random variable where

$$\mu_{\text{iid}}(\gamma,\rho) = \int_0^\infty \ln(1+\rho\lambda)dF_{\text{iid}}(\lambda) \tag{12}$$

$$= \gamma\ln(1+\rho-\rho\alpha_{\text{iid}}(\gamma,\rho))+\ln(1+\rho\gamma-\rho\alpha_{\text{iid}}(\gamma,\rho))-\alpha_{\text{iid}}(\gamma,\rho) \tag{13}$$

and

$${\sigma^2}_{\text{iid}}(\gamma,\rho)=-\ln\left[1-\frac{{\alpha^2}_{\text{iid}}(\gamma,\rho)}{\gamma}\right] \tag{14}$$

with

$$\alpha_{\text{iid}}(\gamma,\rho)=\frac{1}{2}\left[1+\gamma+\frac{1}{\rho}-\sqrt{\left(1+\gamma+\frac{1}{\rho}\right)^2-4\gamma}\right]. \tag{15}$$

It is noteworthy to precise that in this case, $I(n_t, n_r, \rho) = C(n_t, n_r, \rho)$. The theorem has been proved using a lemma in [37] (recalled in the Appendix as Lemma 1) which deals with linear spectral statistics of the form:

$$\frac{1}{n_t}\sum_{i=1}^{n_t} l(\lambda_i) = \int l(x)dF^{\mathbf{B_t}}(x)$$

where $(\lambda_1, ..., \lambda_{n_t})$ denotes the eigenvalues of matrix $\mathbf{B_t}$, $F^{\mathbf{B_t}}(\lambda) = \frac{1}{n_t} \mid \{j : \lambda_j \leq \lambda\} \mid$ and l is a function on $[0, \infty)$. Note that in the high SNR regime ($\rho \to \infty$), $C(n_t, n_r, \rho)$ converges in distribution to a Gaussian random variable:

$$n_t\mu_{\text{iid}} = \min(n_t, n_r)\ln(\rho)$$

$$\sigma^2{}_{\text{iid}} = \begin{cases} -\ln\left(1 - \frac{\min(n_t,n_r)}{\max(n_t,n_r)}\right) & \text{if } n_t \neq n_r \\ \frac{1}{2}\ln(\rho) & \text{if } n_t = n_r \end{cases} \tag{16}$$

C. How Far is Asymptotic?

Large random matrices were first proposed by Wigner in quantum mechanics to explain the measured energy levels of nuclei in terms of the eigenvalues of random matrices. With the works of Telatar [6], Grant & Alexander [38] in the context of code-division multiple access (CDMA), Tse & Hanly [39], and Verdú & Shamai [40], random matrix theory entered the field of telecommunications [4]. Since then, random matrix theory has become a standard tool for the analysis of CDMA in its various fashions and applications [42], [43], [44], [45], [46], [47]. All these results are striking in terms of closeness to simulations with reasonable matrix size and enable to derive performance measures of communication systems as a function of only few meaningful parameters. In the following, we will briefly demonstrate how many antennas are required for large system approximations to be reasonably tight.

The CDF of the capacity is given by Theorem 1 as

$$F(C) = 1 - Q\left(\frac{C - n_t\mu}{\sigma}\right) \tag{17}$$

In Figure 2, the CDF is plotted for a system with 1×1, 2×2 and 4×4 antennas for an SNR

[4]It should be noted that in the field of array processing, Silverstein used already in 1992 random matrix theory [41] for signal detection and estimation.

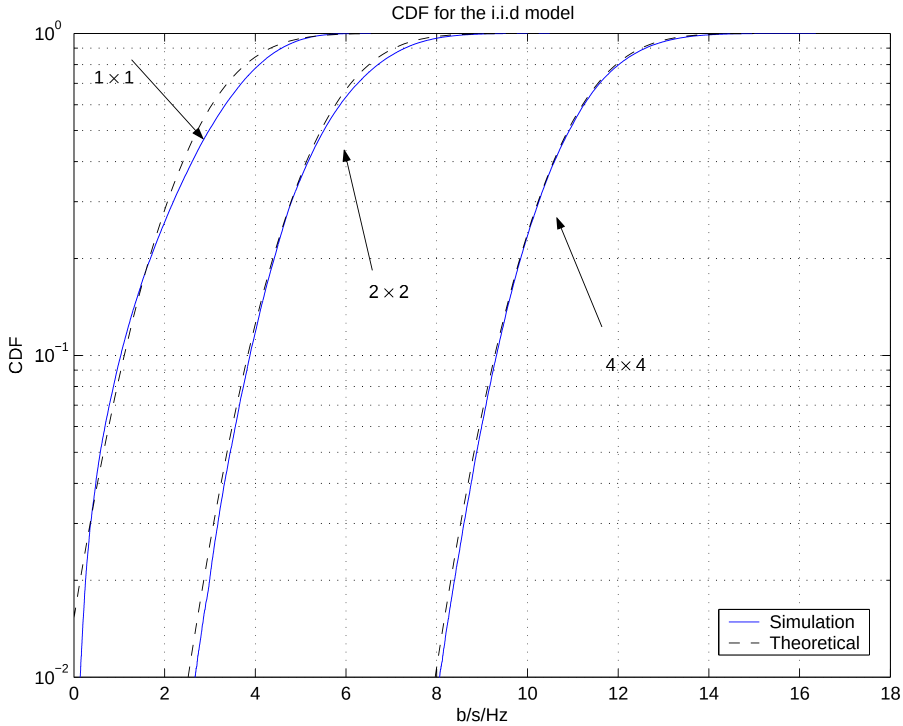

Fig. 2. CDF of mutual information for the i.i.d Gaussian model at 10 dB SNR. Dashed lines and solid lines show the true simulated CDF and the theoretical asymptotic limit, respectively.

of 10 dB. There is a quite realistic match between the asymptotical theoretical formulas and the finite size simulated system with a 4×4 system which shows the usefulness of the random matrix approach. We note that similar curves can be found in the work of Biglieri et al. [48] and Hochwald et al. [32]. To give further evidence on the closeness of asymptotic results with MIMO systems with only a few antenna elements, we note that, in a 6×6 and 3×3 MIMO system operating at 10 dB SNR, the asymptotic mean shows only 0.02% and 0.6% relative error, respectively, and the asymptotic variance has only 1% and 4% relative error, respectively.

As far as mutual information is concerned, infinity is only a couple of antennas and the results can be immediately used for designing future mobile systems. However, results are different for the SINR as shown in [49].

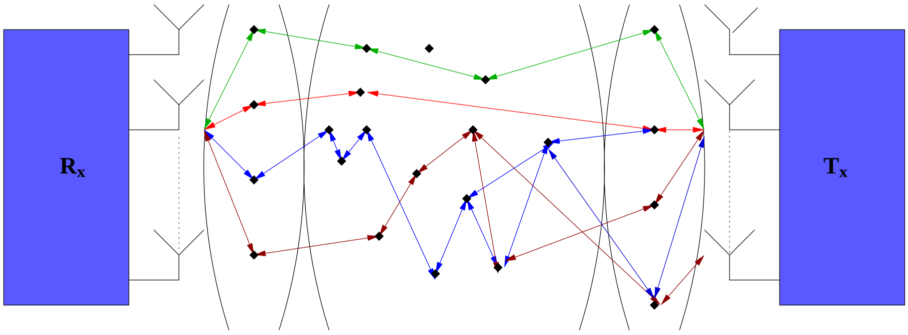

Fig. 3. Double directional based model.

IV. Knowledge of the directions of arrival and departure

In this section, we treat thoroughly the double directional model. Cases with fewer knowledge, e.g. single directional models will be handled as special cases of the double-directional model.

A. Model Construction

Imagine that the modeler is in a situation where he knows that the channel matrix $\mathbf{H}$ has a certain energy. There is no knowledge on the mean. The case where the paths have different non-zero means can be treated in the same way as for the i.i.d Gaussian model. The modeler is now interested in deriving a consistent double directional model, i.e. taking into account simultaneously the directions of arrival and the directions of departure. The motivation of such an approach lies in the fact that when a single bounce on a scatterer occurs, the directions of arrival and departure are deterministically related by Descartes's laws and, therefore, the distribution of the channel matrix depends on the joint DoA-DoD spectrum. The modeler is interested in modelling the channel over time scales over which the locations of scatterers do not change significantly relative to the transmitter or receiver. This is equivalent to considering time scales over which the channel statistics do not change significantly. However, the channel realizations do vary over such time scales. The modeler has knowledge of the directions of departure $\mathbf{\Psi}_{s_t \times n_t}$ from the transmitting antennas to a set of scatterers (A). He also knows the directions of arrival $\mathbf{\Phi}_{n_r \times s_r}$ from a set of scatterers (B) to the receiving antennas, see Figure 3. The modeller also knows the powers of the steering directions. However, the modeler has no knowledge of what happens between the two set of scatterers (A) and (B). In fact, the sets (A) and (B) may be equal, (A) may be included in (B) or there may be no relation between the two. The waves might bounce several times on other scatterers before arriving on the final scatterers (B) or they might directly propagate to them. Moreover, the modeller knows from electromagnetic theory that when a wave propagates from a scatterer to the receiving antennas, the signal can be written in an exponential form

$$\mathbf{s}(t,\mathbf{d}) = \mathbf{s_0}\, e^{j(\mathbf{k}^T\mathbf{d}-2\pi f t)} \tag{18}$$

which is the plane wave solution of the Maxwell equations in free non-dispersive space for wave vector $\mathbf{k} \in \mathbb{R}^{2\times 1}$ and location vector $\mathbf{d} \in \mathbb{R}^{2\times 1}$. The reader must note that other solutions to the Maxwell equations exist and therefore the modeler is making an important restriction. The direction of the vector $\mathbf{s_0}$ gives us knowledge on the polarization of the wave while the direction of the wave vector $\mathbf{k}$ gives us knowledge on the direction of propagation. The phase of the signal results in $\phi = \mathbf{k}^T\mathbf{d}$. The modeler knows (or considers for sake of simplicity) that the scatterers and the antennas lie in the same plane. The modeler makes use of the knowledge that the steering vector is known up to a multiplicative complex constant that is the same for all antennas.

Although correlation might exist between the scatterers, the modeler is not aware of that fact. Based on this state of knowledge, the modeler wants to derive a model which takes into account all the previous constraints while leaving as many degrees of freedom as possible to the other parameters to avoid the introduction of unjustified information.

Based on the fact that

$$\begin{aligned}\mathbf{H} &= \frac{1}{\sqrt{s_r s_t}}\begin{pmatrix} e^{j\phi_{1,1}} & \dots & e^{j\phi_{1,s_r}} \\ \vdots & \ddots & \vdots \\ e^{j\phi_{n_r,1}} & \dots & e^{j\phi_{n_r,s_r}} \end{pmatrix}\begin{pmatrix} \sqrt{P_1^r} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_r}^r} \end{pmatrix} \\ &\quad \mathbf{\Theta}_{s_r\times s_t}\begin{pmatrix} \sqrt{P_1^t} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_t}^t} \end{pmatrix}\begin{pmatrix} e^{j\psi_{1,1}} & \dots & e^{j\psi_{1,n_t}} \\ \vdots & \ddots & \vdots \\ e^{j\psi_{s_t,1}} & \dots & e^{j\psi_{s_t,n_t}} \end{pmatrix},\end{aligned} \tag{19}$$

the modeler must attribute a probability distribution to $\mathbf{\Theta}_{s_r\times s_t}$. The steering matrices

$\mathbf{\Phi}_{n_r \times s_r}$ and $\mathbf{\Psi}_{s_t \times n_t}$ represent the directions of arrival from scatterers (B) to the receiving antennas and the directions of departure from the transmitting antennas to scatterers (A), respectively, see also Figure 3. The powers of the steering directions are given by the diagonal matrices $\mathbf{P^r}$ and $\mathbf{P^t}$, respectively. The phases $\phi_{i,j} = \mathbf{k^r}^T \mathbf{d^r}_{i,j}$ and $\psi_{i,j} = \mathbf{k^t}^T \mathbf{d^t}_{i,j}$ are given as scalar products between the respective wave vectors and the respective locations of the scatterers.

Remark 1: In the introduction, we have recalled the work of Shore et al. [18] which shows that maximizing entropy leads to consistent solutions. However, incorporating information in the entropy criteria which is not given in terms of expected values is not an easy task. As a consequence, we will not maximize entropy based only on the information we have (expected values and the directions of arrival): we will maximize entropy based on the expected values and a structured form of the channel based on the product of five matrices. This is more than the information we have since the directions of arrival and departure are not constraint equations in the entropy criteria. This ad-hoc procedure is used because it is extremely difficult to incorporate knowledge on physical considerations (number of chairs, type of room...) in the entropy criteria. As a consequence, each time this ad-hoc procedure is used, we will verify that, although we maximize entropy under a structured constraint, we remain consistent. This will lead to a maximum entropy solution. With the maximum entropy approach, every new information on the environment should be incorporated in a consistent way: adding or retrieving information takes us one step forward or back but always in a consistent way. The models are somewhat like Russian dolls, imbricated one into the other.
The consistency argument, see Proposition 1, states that if the DoAs, the powers $\mathbf{P^t}$ and $\mathbf{P^r}$, and the DoDs are unknown, then the channel matrix $\mathbf{H}$ in (19) should be assigned an i.i.d zero mean Gaussian distribution, see Section III-A, since the modeler is in the same state of knowledge as before where he only knew the variance. Based on this consistency requirement, we can determine the distribution of $\mathbf{\Theta}_{s_r \times s_t}$.

The probability distribution of $P(\mathbf{H})$ is given by

$$\begin{aligned} P(\mathbf{H}) &= \int P(\mathbf{\Phi} {\mathbf{P^r}}^{\frac{1}{2}} \mathbf{\Theta} {\mathbf{P^t}}^{\frac{1}{2}} \mathbf{\Psi} \mid \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{P^r}, \mathbf{P^t}, s_r, s_t) \\ & \quad P(\mathbf{\Psi}, \mathbf{\Phi} \mid s_r, s_t) P(\mathbf{P^r}, \mathbf{P^t} \mid s_t, s_r) P(s_t, s_r) ds_r ds_t d\mathbf{P^r} d\mathbf{P^t} d\mathbf{\Psi} d\mathbf{\Phi} \qquad (20) \end{aligned}$$

- When $\mathbf{\Psi}, \mathbf{\Phi}, s_r, s_t, \mathbf{P^r}, \mathbf{P^t}$ are known: $P(\mathbf{\Phi\Psi} \mid s_r, s_t) = \delta(\mathbf{\Phi} - \mathbf{\Phi}_0)\delta(\mathbf{\Psi} - \Psi_0)$, $P(s_t, s_r) = \delta(s_r - a)\delta(s_t - b)$,$P(\mathbf{P^r}^{\frac{1}{2}}, \mathbf{P^t}^{\frac{1}{2}} \mid s_r, s_t) = \delta(\mathbf{P^r}^{\frac{1}{2}} - \mathbf{P_0}^{r\frac{1}{2}})\delta(\mathbf{P^t}^{\frac{1}{2}} - \mathbf{P_0}^{t\frac{1}{2}})$ and

$$P(\mathbf{H}) = P(\Phi_0 \mathbf{P_0}^{r\frac{1}{2}} \mathbf{\Theta} \mathbf{P_0}^{t\frac{1}{2}} \mathbf{\Psi}_0)$$

- Suppose now that $\mathbf{\Psi}, \mathbf{\Phi}, s_r, s_t$ are unknown, then each entry h_{ij} of $\mathbf{H}$ must have an i.i.d zero mean Gaussian distribution. In this case, the following result holds:

Proposition 1: $\mathbf{\Theta}_{s_r \times s_t}$ i.i.d zero mean Gaussian with variance 1 is solution of the consistency argument and maximizes entropy.

Proof: Let us show that $\mathbf{\Theta}_{s_r \times s_t}$ i.i.d zero mean Gaussian with variance 1 is solution of the consistency argument and maximizes entropy. Since $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are unknown, the principle of maximum entropy attributes i.i.d uniform distributed angles over 2π to the entries ϕ_{ij} and ψ_{ij}. In this case, if one chooses $\theta_{p,k}$ to be i.i.d zero mean Gaussian with variance 1 and knowing that $h_{ij} = \frac{1}{\sqrt{s_1 s}} \sum_{k=1}^{s_t} \sum_{p=1}^{s_r} \theta_{pk} \sqrt{P_k^{\ t}} \sqrt{P_p^{\ r}} e^{j\psi_{kj}} e^{j\phi_{ip}}$, then: $P\left(h_{ij} \mid \mathbf{\Psi}, \mathbf{\Phi}, s_r, s_t\right) = N(0, \frac{1}{s_t s_r} \sum_{p=1}^{s_r} \sum_{k=1}^{s_t} \mid \sqrt{P_p^{\ r}} e^{j\phi_{ip}} \sqrt{P_k^{\ t}} e^{j\psi_{kj}} \mid^2 = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}}$ (since $\frac{1}{s_r} \sum_{k=1}^{s_r} P_k^{\ r} = 1$ and $\frac{1}{s_t} \sum_{p=1}^{s_t} P_p^{\ t} = 1$ (due to power normalization)).

Therefore

$$\begin{aligned}
P(h_{ij}) &= \int \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}} P(\mathbf{\Phi}, \mathbf{\Psi} \mid s_t, s_r) P(\mathbf{P^r}, \mathbf{P^t} \mid s_t, s_r) P(s_t, s_r) P(s_t, s_r) d\mathbf{\Phi} d\mathbf{\Psi} \\
&\quad d\mathbf{P^r} d\mathbf{P^t} ds_t ds_r \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}} \int P(\mathbf{\Phi}, \mathbf{\Psi} \mid s_t, s_r) P(\mathbf{P^r}, \mathbf{P^t} \mid s_t, s_r) P(s_t, s_r) d\mathbf{\Phi} d\mathbf{\Psi} d\mathbf{P^r} d\mathbf{P^t} ds_t ds_r \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}}
\end{aligned}$$

Moreover, we have :

$$\begin{aligned}
\mathbb{E}_{\mathbf{\Phi},\mathbf{\Psi},\mathbf{\Theta}}(h_{ij}h_{mn}^*) &= \frac{1}{s_t s_r}\sum_{k=1}^{s_t}\sum_{p=1}^{s_r}\sum_{r=1}^{s_t}\sum_{l=1}^{s_r}\mathbb{E}_{\mathbf{\Theta}}(\theta_{pk}\theta_{lr}^*)\mathbb{E}_{\mathbf{\Psi}}(e^{-j\psi_{rn}+j\psi_{kj}})\mathbb{E}_{\mathbf{\Phi}}(e^{-j\phi_{ml}+j\phi_{ip}}) \\
&\quad \sqrt{P_k^t}\sqrt{P_r^t}\sqrt{P_p^r}\sqrt{P_l^r} \\
&= \frac{1}{s_t s_r}\sum_{k=1}^{s_t}\sum_{p=1}^{s_r}\sum_{r=1}^{s_t}\sum_{l=1}^{s_r}\delta_{pl}\delta_{kr}\mathbb{E}_{\mathbf{\Psi}}(e^{-j\psi_{rn}+j\psi_{kj}})\mathbb{E}_{\mathbf{\Phi}}(e^{-j\phi_{ml}+j\phi_{ip}}) \\
&\quad \sqrt{P_k^t}\sqrt{P_r^t}\sqrt{P_p^r}\sqrt{P_l^r} \\
&= \frac{1}{s_t s_r}\sum_{k=1}^{s_t}\sum_{p=1}^{s_r}\mathbb{E}_{\mathbf{\Psi}}(e^{-j\psi_{kn}+j\psi_{kj}})\mathbb{E}_{\mathbf{\Phi}}(e^{-j\phi_{mp}+j\phi_{ip}})P_k^t P_p^r \\
&= \delta_{im}\delta_{jn}\frac{1}{s_t s_r}\sum_{k=1}^{s_t}\sum_{p=1}^{s_r}P_k^t P_p^r \\
&= \delta_{im}\delta_{jn}
\end{aligned}$$

Which proves that $\mathbf{\Theta}_{s_r\times s_t}$ is solution of the consistency argument. One interesting point of the maximum entropy approach is that while we have not assumed uncorrelated scattering, the above methodology will automatically assign a model with uncorrelated scatterers in order to have as many degrees of freedom as possible. But this does not mean that correlation is not taken into account. The model in fact leaves free degrees for correlation to exist or not. Note that in this model, the entries of $\mathbf{H}$ are correlated, for general DoA's and DoD's.

B. General Case

In this part, we are interested in deriving the asymptotic mutual information per transmitting antenna. We denote: $\gamma = \frac{n_r}{s_r}, \xi = \frac{s_r}{n_t}, \gamma_1 = \frac{n_r}{s_t}, \xi_1 = \frac{s_t}{n_t}$. The asymptotic mutual information per transmitting antenna is given by:

$$\begin{aligned}\mu &= \frac{1}{n_t}\ln\left[\det(\mathbf{I}_{n_t} + \frac{\rho}{n_t}\mathbf{H}^H\mathbf{H})\right] \\ &= \frac{1}{n_t}\ln\left[\det(\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H)\right] \\ &= \frac{n_r}{n_t n_r}\ln\left[\det(\mathbf{I}_{n_r} + \frac{\rho n_r}{n_t n_r}\mathbf{H}\mathbf{H}^H)\right] \\ &= \gamma\xi\frac{1}{n_r}\sum_{i=1}^{n_r}\ln(1+\rho\gamma\xi\lambda_i) \\ &= \gamma\xi\int\ln(1+\rho\gamma\xi\lambda)dF_{n_r}(\lambda)\end{aligned}$$

where λ_i are the eigenvalues of matrix $\frac{1}{n_r}\mathbf{H}\mathbf{H}^H$ and $F_{n_r}(\lambda)$ is the empirical eigenvalue distribution of matrix $\frac{1}{n_r}\mathbf{H}\mathbf{H}^H$ defined by: $dF_{n_r}(\lambda) = \frac{1}{n_r}\sum_{i=1}^{n_r}\delta(\lambda-\lambda_i)$

In order to derive the asymptotic mutual information per transmitting antenna, we will show that the empirical eigenvalue distribution $F_{n_r}(\lambda)$ converges weakly to a non-random limiting distribution $F_{\mathbf{H}\mathbf{H}^H}(\lambda)$. More specifically, let

$$\frac{1}{s_r}\mathbf{P}^{\mathbf{r}H\frac{1}{2}}\mathbf{\Phi}^H\mathbf{\Phi}\mathbf{P}^{\mathbf{r}\frac{1}{2}} = \mathbf{V}_{\mathbf{\Phi}}\mathbf{\Lambda}_{\mathbf{\Phi}}\mathbf{V}_{\mathbf{\Phi}}^H$$

and

$$\frac{1}{s_t}\mathbf{P}^{\mathbf{t}\frac{1}{2}}\mathbf{\Psi}\mathbf{\Psi}^H\mathbf{P}^{\mathbf{t}H\frac{1}{2}} = \mathbf{V}_{\mathbf{\Psi}}\mathbf{\Lambda}_{\mathbf{\Psi}}\mathbf{V}_{\mathbf{\Psi}}^H$$

$\mathbf{V}_{\psi}$ and $\mathbf{V}_{\mathbf{\Phi}}$ are unitary matrices while $\mathbf{\Lambda}_{\mathbf{\Phi}}$ and $\mathbf{\Lambda}_{\psi}$ are diagonal matrices representing respectively the eigenvalues of matrices $\frac{1}{s_r}\mathbf{P}^{\mathbf{r}H\frac{1}{2}}\mathbf{\Phi}^H\mathbf{\Phi}\mathbf{P}^{\mathbf{r}\frac{1}{2}}$ and $\frac{1}{s_t}\mathbf{P}^{\mathbf{t}\frac{1}{2}}\mathbf{\Psi}\mathbf{\Psi}^H\mathbf{P}^{\mathbf{t}H\frac{1}{2}}$.

The non-zero eigenvalues of matrix $\frac{1}{n_r}\mathbf{H}\mathbf{H}^H = \frac{1}{s_r s_t}\mathbf{\Phi}\mathbf{P}^{\mathbf{r}\frac{1}{2}}\mathbf{\Theta}\mathbf{P}^{\mathbf{t}\frac{1}{2}}\mathbf{\Psi}\mathbf{\Psi}^H\mathbf{P}^{\mathbf{t}H\frac{1}{2}}\mathbf{\Theta}^H\mathbf{P}^{\mathbf{r}H\frac{1}{2}}\mathbf{\Phi}^H$ are the same as $\mathbf{\Theta}_1\mathbf{\Theta}_1^H = \frac{1}{n_r}[\mathbf{\Lambda}_{\mathbf{\Phi}}^{\frac{1}{2}}(\mathbf{V}_{\mathbf{\Phi}}{}^H\mathbf{\Theta}\mathbf{V}_{\mathbf{\Psi}})\mathbf{\Lambda}_{\mathbf{\Psi}}^{\frac{1}{2}}][\mathbf{\Lambda}_{\mathbf{\Psi}}^{\frac{1}{2}}(\mathbf{V}_{\mathbf{\Psi}}^H\mathbf{\Theta}^H\mathbf{V}_{\mathbf{\Phi}})\mathbf{\Lambda}_{\mathbf{\Phi}}^{\frac{1}{2}}]$. Without loss of generality, we will suppose that $s_r \leq n_r$. Therefore, the spectra of $\frac{1}{n_r}\mathbf{H}\mathbf{H}^H$ and $\mathbf{\Theta}_1\mathbf{\Theta}_1^H$ are related by:

$$f_{\mathbf{H}\mathbf{H}^H}(x) = (1-\frac{s_r}{n_r})\delta(x-0) + \frac{s_r}{n_r}f_{\mathbf{\Theta}_1\mathbf{\Theta}_1^H}(x)$$

and their Stieljes transforms are related as:

$$m_{\mathbf{H}\mathbf{H}^H}(z) = (\frac{1}{\gamma}-1)\frac{1}{z} + \frac{1}{\gamma}m_{\mathbf{\Theta}_1\mathbf{\Theta}_1^H}(z)$$

Matrix $\mathbf{V}_{\mathbf{\Phi}}\mathbf{\Theta}\mathbf{V}_{\mathbf{\Psi}}$ is an i.i.d zero mean Gaussian matrix with unit variance (only unitary transforms are applied). Therefore, matrix $\mathbf{\Theta}_1 = \frac{1}{\sqrt{n_r}}[\mathbf{\Lambda}_{\mathbf{\Phi}}^{\frac{1}{2}}(\mathbf{V}_{\mathbf{\Phi}}\mathbf{\Theta}\mathbf{V}_{\mathbf{\Psi}})\mathbf{\Lambda}_{\mathbf{\Psi}}^{\frac{1}{2}}]$ is a $s_r \times s_t$

random matrix composed of independent entries with zero mean and variances $\frac{1}{n_r}{\lambda_i}^{\mathbf{\Phi}}{\lambda_j}^{\mathbf{\Psi}} = \frac{1}{s_r}\frac{{\lambda_i}^{\mathbf{\Phi}}{\lambda_j}^{\mathbf{\Psi}}}{\gamma}$. The weak convergence of the empirical eigenvalue distribution of $\mathbf{\Theta}_1\mathbf{\Theta}_1^H$ to a limiting distribution holds and is an application of a theorem due to Girko [50] (the theorem is recalled in the appendix through Theorem 3). Note that this theorem was already used in [27] for deriving the mutual information of MIMO wireless systems under correlated fading and in [51] for analyzing CDMA Networks with Multiuser Receivers and Spatial Diversity.

Proposition 2: Suppose that the empirical joint distribution ${\lambda_i}^{\mathbf{\Phi}}{\lambda_j}^{\mathbf{\Psi}}$ converges to some joint limit distribution $f_{\lambda^{\mathbf{\Phi}}\lambda^{\mathbf{\Psi}}}$ as the size of the matrix $\mathbf{\Theta}_1$ grows large but $\gamma, \gamma_1, \xi, \xi_1$ remains fixed, the asymptotic mutual information per transmitting antenna is given by:

$$\mu = \int_0^\rho \xi \mathbb{E}_{\lambda^{\mathbf{\Phi}}}\left[\frac{\alpha^{\text{joint}}}{1+\rho\alpha^{\text{joint}}}\right] d\rho \tag{21}$$

with

$$\alpha^{\text{joint}} = \xi_1 \mathbb{E}_{\lambda^{\mathbf{\Psi}}}\left[\frac{\lambda^{\mathbf{\Phi}}\lambda^{\mathbf{\Psi}}}{1+\rho{\alpha_1}^{\text{joint}}}\right]$$

and

$${\alpha_1}^{\text{joint}} = \xi \mathbb{E}_{\lambda^{\mathbf{\Phi}}}\left[\frac{\lambda^{\mathbf{\Phi}}\lambda^{\mathbf{\Psi}}}{1+\rho\alpha^{\text{joint}}}\right]$$

Proof: the proof is provided in appendix.

Note that when taking the expectations, $\lambda^{\mathbf{\Phi}}$ depends on $\lambda^{\mathbf{\Psi}}$. In the case where the distribution of the angles of arrival are independent of the angles of departure, a simpler expression of equation (21) can be provided[5].

Proposition 3: Suppose that the empirical joint distribution ${\lambda_i}^{\mathbf{\Phi}}{\lambda_j}^{\mathbf{\Psi}}$ is separable and converges to a product of two limiting distributions, then as the size of the matrix $\mathbf{\Theta}_1$ grows large but $\gamma, \gamma_1, \xi, \xi_1$ remains fixed, the asymptotic mutual information per transmitting antenna is given by:

$$\mu = \xi_1 \mathbb{E}_{\lambda^{\mathbf{\Psi}}}(\ln(1+\rho\lambda^{\mathbf{\Psi}}\alpha_{\text{dod}})) + \xi\mathbb{E}_{\lambda^{\mathbf{\Phi}}}(\ln(1+\rho\lambda^{\mathbf{\Phi}}\alpha_{\text{doa}})) - \rho\alpha_{\text{doa}}\alpha_{\text{dod}}$$

with $\alpha_{\text{doa}} = \xi_1\mathbb{E}_{\lambda^{\mathbf{\Psi}}}\left[\frac{\lambda^{\mathbf{\Psi}}}{1+\rho\lambda^{\mathbf{\Psi}}\alpha_{\text{dod}}}\right]$ and $\alpha_{\text{dod}} = \xi\mathbb{E}_{\lambda^{\mathbf{\Phi}}}\left[\frac{\lambda^{\mathbf{\Phi}}}{1+\rho\lambda^{\mathbf{\Phi}}\alpha_{\text{doa}}}\right]$

Proof: the proof is provided in the appendix.

[5]Physical measurements have already indicated that the correlation between the directions of arrival and the directions of departure are negligible [52].

The formula is general enough to be applied for the i.i.d Gaussian case, the DoA based model and the DoD based model. Hence, for example, in the DoA case, one has:

$$\mathbb{E}_{\lambda^{\Psi}}(f) = \int f(\lambda)\delta(\lambda - 1)d\lambda$$

Proposition 3 is extremely useful as it shows that only the limiting eigenvalue distribution of the steering directions with powers matters: in other words, two antenna configuration can yield the same throughput as long as they give rise to the same eigenvalue distribution for the steering matrix. Based on this result, a future mobile scenario the authors would like to advocate is the following: imagine a set of reconfigurable antennas that can move on a grid. The antennas are at the beginning displayed with a Uniform Linear Array geometry. Once the transmission starts, the angles of arrival and the distances of the scatterers to the antennas are determined. The position of the antennas (for fixed scatterers) on the grid are then optimized in order to increase mutual information using the previous formulas. This is once more a viable scenario from a software defined radio perspective and gives means for future research in the field of antenna design. The antenna design problem can therefore be related to an eigenvalue optimization problem. What really governs the transmission limits of different scenarios are only the properties of the eigenvalues of the steering matrix.

Remark 2: Results of Free Probability Theory could also be used to prove Proposition 2. Indeed, one has to derive the limiting eigenvalue distribution of $\frac{1}{n_r}[\mathbf{\Lambda}_{\mathbf{\Phi}}^{\frac{1}{2}}\mathbf{\Theta}\mathbf{\Lambda}_{\mathbf{\Psi}}\mathbf{\Theta}^H\mathbf{\Lambda}_{\mathbf{\Phi}}^{\frac{1}{2}}]$ with $\mathbf{\Theta}$ i.i.d Gaussian and $\mathbf{\Lambda}_{\mathbf{\Phi}}$ and $\mathbf{\Lambda}_{\mathbf{\Psi}}$ diagonal. Since $\mathbf{\Theta}\mathbf{\Lambda}_{\mathbf{\Psi}}\mathbf{\Theta}^H$ is unitarily invariant[6] and asymptotically free from $\mathbf{\Lambda}_{\mathbf{\Phi}}$, one can obtain straightforwardly the law using the free multiplicative convolution of $\mathbf{\Theta}\mathbf{\Lambda}_{\mathbf{\Psi}}\mathbf{\Theta}^H$ and $\mathbf{\Lambda}_{\mathbf{\Phi}}$.

Remark 3: Although we have no formal proof on the uniqueness (the mean mutual information of proposition 3 has in fact multiple solutions. Therefore, only some physical arguments can be given to withdraw some solutions), one of the solutions of the mean mutual information for the double directional model can be shown to scale at high SNR as:

$$\mathbb{E}(I^M) = \min\left(n_t, n_r, s_t\mathbb{E}_{\lambda^{\Psi}}(\mathbf{1}_{[\lambda^{\Psi}>0]}), s_r\mathbb{E}_{\lambda^{\Phi}}(\mathbf{1}_{[\lambda^{\Phi}>0]})\right)\ln(\rho)$$

[6]A $N \times N$ self-adjoint random matrix is called unitarily invariant if the probability measure of A as a random matrix is equal to that of the matrix VAV^H for any unitary constant matrix V

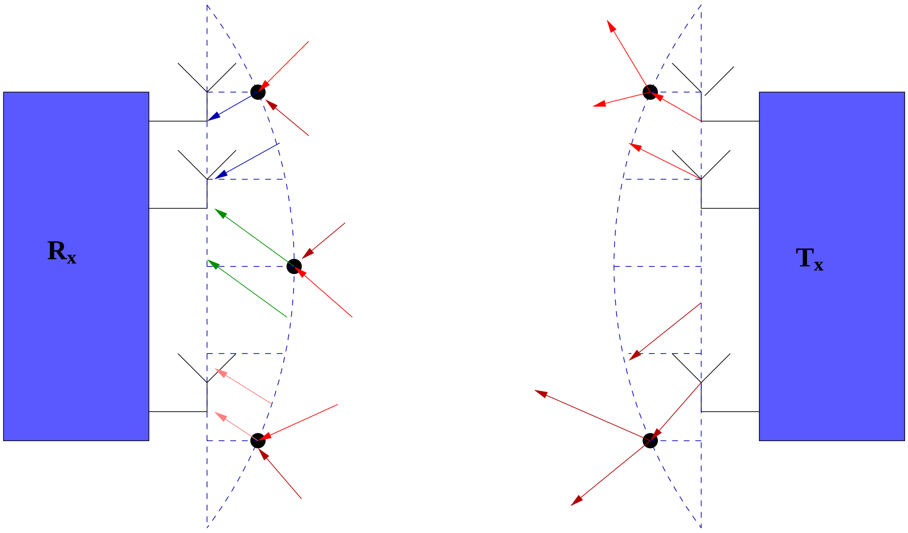

Fig. 4. Simple case: scatterers positioned on special directions.

The integral is on the support of non-zero eigenvalues and $\mathbb{E}_{\lambda^{\Phi}}(\mathbf{1}_{[\lambda^{\Phi}>0]})$ and $\mathbb{E}_{\lambda^{\Psi}}(\mathbf{1}_{[\lambda^{\Psi}>0]})$ express respectively the correlation factor of the s_r and s_t scatterers. Note that a similar result, based on a different approach, can be found in [53]

C. ULA and Fourier Directions Case

In this part, the modeler takes into account the geometry of the receiving and transmitting antenna (as he knows it) to derive the steering vectors: in the case of a uniform linear array, the steering DoA vector has the following form $[1, e^{-j2\pi\frac{d\sin(\phi)}{\lambda}}, ..., e^{-j2\pi\frac{d(n_r-1)\sin(\phi)}{\lambda}}]$ where d is the antenna spacing and ϕ is the direction of arrival.[7] The DoA ϕ of a source is defined as the angle between a line perpendicular to the incoming wave-front and a reference line through the array. The same holds for the directions of departure.

$$\mathbf{H} = \frac{1}{\sqrt{ss_1}}\begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ e^{j2\pi\frac{d(n_r-1)\sin(\phi_1)}{\lambda}} & \dots & e^{j2\pi\frac{d(n_r-1)\sin(\phi_{s_r})}{\lambda}} \end{pmatrix} \mathbf{P^r}^{\frac{1}{2}} \mathbf{\Theta}_{s_r \times s_t} \mathbf{P^t}^{\frac{1}{2}} \begin{pmatrix} 1 & \dots & e^{j2\pi\frac{d(n_t-1)\sin(\psi_1)}{\lambda}} \\ \vdots & \ddots & \vdots \\ 1 & \dots & e^{j2\pi\frac{d(n_t-1)\sin(\psi_{s_t})}{\lambda}} \end{pmatrix}$$

For simplicity sake, we will take $d = \frac{\lambda}{2}$. We will also suppose that $s_r \leq n_r$ and $s_t \leq n_t$. In order to have tractable explicit formulas, we will analyze the distribution of scatterers

[7]Note that the modeler is making a strong assumption based on the fact that the scatterers are far from the antenna. We assume in this case that the modeler has some evidence that he is not closely surrounded by obstacles.

in the case where for any i there exists a k such as $\sin(\phi_i) = \frac{2k}{n_r}$ (see Figure 4) and for any j there exists a l such as $\sin(\phi_j) = \frac{2l}{n_t}$. This case can be seen as an extreme case where all the scatterers are maximally distant from each other called here the Maxent Fourier model.

C.1 Equal Power case

We will suppose in this part that $\mathbf{P^r} = \mathbf{I}_{s_r}$ and $\mathbf{P^t} = \mathbf{I}_{s_t}$.

As a consequence, the DoA and DoD steering matrices have the following limiting eigenvalue distribution:

$$S_{\mathbf{\Phi}^H\mathbf{\Phi}}(\lambda) = \delta(\lambda - \gamma)$$

and

$$S_{\mathbf{\Psi}\mathbf{\Psi}^H}(\lambda) = \delta(\lambda - \frac{1}{\xi_1})$$

Proposition 4: The asymptotic mutual information per transmitting antenna for the double directional model in the equal power and Fourier directions case is given by:

$$\mu_{\text{double}} = \xi \ln(1 + \rho\gamma - \rho\gamma\alpha_{\text{double}}) + \xi_1 \ln(1 + \rho\gamma_1 - \rho\gamma\alpha_{\text{double}}) - \xi_1\alpha_{\text{double}}$$

with

$$\alpha_{\text{double}} = \frac{1}{2}\left[1 + \frac{\gamma_1}{\gamma} + \frac{1}{\rho\gamma} - \sqrt{(1 + \frac{\gamma_1}{\gamma} + \frac{1}{\rho\gamma})^2 - 4\frac{\gamma_1}{\gamma}}\right]$$

Proof: The result is an application of Proposition 3 and the proof is given in the appendix. ■

Note that in the equal power case, it is possible to derive the exact asymptotic distribution. In particular, the asymptotic variance can be derived.

Proposition 5: The asymptotic variance of the mutual information in the equal power and Fourier directions case for the double directional model is given by:

$$\sigma^2_{\text{double}} = -\ln\left[1 - \frac{\alpha^2_{\text{double}}\gamma}{\gamma_1}\right]$$

Proof: One can notice that:

$$\begin{aligned}
\mu &= \frac{1}{n_t}\ln\det(\mathbf{I}_{n_t} + \frac{\rho}{n_t}\mathbf{H}^H\mathbf{H}) \\
&= \frac{1}{n_t}\ln\det(\mathbf{I}_{n_t} + \frac{\rho\gamma}{n_t s_t}\mathbf{\Psi}^H\mathbf{\Theta}^H\mathbf{\Theta}\mathbf{\Psi}) \\
&= \frac{1}{n_t}\ln\det(\mathbf{I}_{s_t} + \frac{\rho\gamma}{n_t s_t}\mathbf{\Theta}^H\mathbf{\Theta}\mathbf{\Psi}\mathbf{\Psi}^H) \\
&= \frac{1}{n_t}\ln\det(\mathbf{I}_{s_t} + \frac{\rho\gamma n_t}{n_t}\frac{1}{s_t t}\mathbf{\Theta}^H\mathbf{\Theta}) \\
&= \frac{s_1}{n_t}\frac{1}{s_1}\ln\det(\mathbf{I}_{s_1} + \rho\gamma\frac{1}{s_1}\mathbf{\Theta}^H\mathbf{\Theta}) \\
&= \xi_1\frac{1}{s_1}\ln\det(\mathbf{I}_{s_1} + \rho\gamma\frac{1}{s_1}\mathbf{\Theta}^H\mathbf{\Theta})
\end{aligned}$$

Therefore, since $\mathbf{\Theta}$ is an i.i.d Gaussian matrix, results of section III-B can be applied. In particular, if one makes the variable change:

$$\rho \longrightarrow \rho\gamma$$

$$\gamma \longrightarrow \frac{\gamma_1}{\gamma}$$

in the formulas of theorem 1 then the result is proved. By doing so, one can notice that by this change of variable the same formula as in Proposition 4 (which was obtained with Girko's results) is obtained for the mean value. ■

At high SNR, it can be easily shown that:

$$n_t\mu_{\text{double}} = \min(s_t, s_r)\ln(\rho)$$

$$\sigma^2{}_{\text{double}} = \begin{cases} -\ln\left(1 - \frac{\min(s_t,s_r)}{\max(s_t,s_r)}\right) & \text{if } s_t \neq s_r \\ \frac{1}{2}\ln(\rho) & \text{if } s_t = s_r \end{cases} \tag{22}$$

Therefore, the limiting factor is only the number of scatterers at the transmitting and receiving side.

In Figure 5, simulations have been conducted with $n_r = n_t = 8$ antennas. Three cases have been plotted:

- $s_r = 8$ and $s_t = 8$
- $s_r = 4$ and $s_t = 4$
- $s_r = 4$ and $s_t = 8$

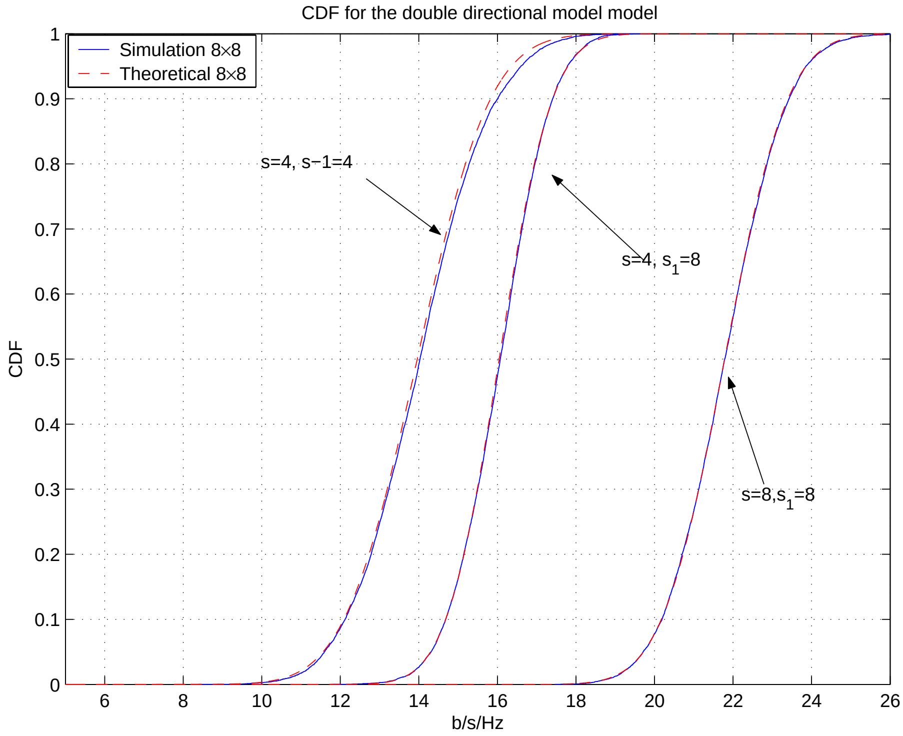

Fig. 5. Mutual information cumulative distribution in the case of the double directional model with equal power on Fourier directions.

In each case, a close match between the theoretical predictions and the simulations occurs. In order to determine the impact of the number of scatterers on the mutual information per receiving antennas, we have plotted in Figure 6 the mutual information versus $\xi = \frac{s_r}{n_t}$ and $\xi_1 = \frac{s_t}{n_t}$ for $n_r = n_t$. One can observe that due to the fact that $n_r = n_t$, the scatterers have the same effect on both the receiving and transmitting side. The maximum rate is achieved when $s_r = s_t = n_r = n_t$.

C.2 Non-Equal Power Case

We consider in this case that there is a finite set of K_r distinct amplitudes $\sqrt{P_i^r}$ of the receiving steering vectors with weight l_i^r (such as $\sum_{i=1}^{K_r} l_i^r = 1$) and K_t distinct amplitudes $\sqrt{P_i^t}$of the transmitting steering vectors with weight l_i^t (such as $\sum_{i=1}^{K_t} l_i^t = 1$). As a consequence, the limiting eigenvalue distribution S_{doa} of $\frac{1}{s_r}\mathbf{P^r}^{H\frac{1}{2}}\mathbf{\Phi}^H\mathbf{\Phi}\mathbf{P^r}^{\frac{1}{2}}$ has the

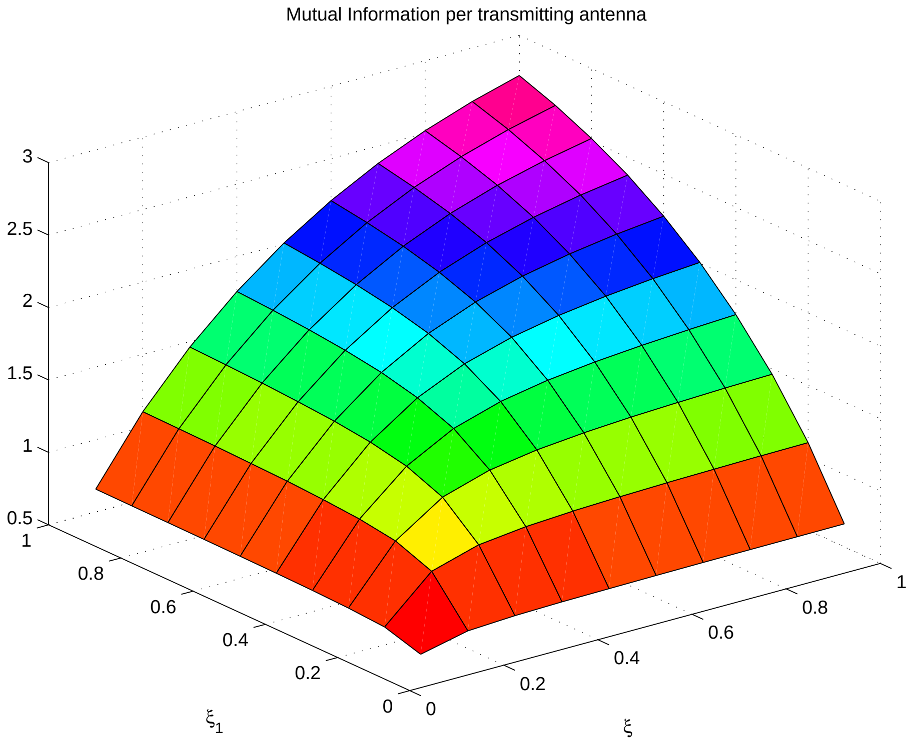

Fig. 6. Mutual Information per transmitting antenna versus ξ and ξ_1 for the double directional model with equal power on Fourier directions.

following expression:

$$S_{\text{doa}}(\lambda) = \sum_{i=1}^{K_r} {l_i}^r \delta(\lambda - \gamma {P_i}^r)$$

and the limiting eigenvalue distribution S_{dod} of $\frac{1}{s_t} \mathbf{P^t}^{\frac{1}{2}} \mathbf{\Psi}\mathbf{\Psi}^H \mathbf{P^t}^{H\frac{1}{2}}$ has the following expression:

$$S_{\text{dod}}(\lambda) = \sum_{i=1}^{K_t} {l_i}^t \delta(\lambda - \frac{{P_i}^t}{\xi_1})$$

Proposition 6: In this case, μ_{double} is equal to:

$$\mu_{\text{double}} = \xi_1 \sum_{i=1}^{K_t} {l_i}^t \ln(1 + \frac{\rho {P_i}^t \alpha_{\text{dod}}}{\xi_1}) + \xi \sum_{i=1}^{K_r} {l_i}^r \ln(1 + \rho {P_i}^r \gamma \alpha_{\text{doa}}) - \rho \alpha_{\text{doa}} \alpha_{\text{dod}}$$

with

$$\alpha_{\text{doa}} = \sum_{i=1}^{K_t} \frac{{l_i}^t {P_i}^t}{1 + \frac{\rho {P_i}^t \alpha_{\text{dod}}}{\xi_1}}$$

and

$$\alpha_{\text{dod}} = \xi \sum_{i=1}^{K_r} l_i{}^r \frac{P_i{}^r \gamma}{1 + \rho\gamma P_i{}^r \alpha_{\text{doa}}}$$

Proof: The proof is an application of the general Proposition 3 in the case of interest. ■

An important question concerns the power profile of the scatterers which optimizes the mean mutual information. The following proposition provides the optimum power profiles.

Proposition 7: The mean of the mutual information in the case of the double directional model with ULA and Fourier directions is maximized for $\mathbf{P^r} = \mathbf{I}_{s_r}$ and $\mathbf{P^t} = \mathbf{I}_{s_t}$.

Proof: The proof is provided in the appendix. ■

In Figure 7, simulations have been conducted in the two power case with $n_r = n_t = 8$ antennas. We have chosen $P_1{}^t = 2 - P_2{}^t$, $P_1{}^r = 2 - P_2{}^r$, , $l_1{}^r = l_2{}^r = l_1{}^t = l_2{}^t$, $s_r = 8$ and $s_t = 8$. In this case, we have ($\gamma = \frac{n_r}{s} = 1$,$\xi_1 = \frac{s_1}{n_t} = 1$,$\xi = \frac{s}{n_t} = 1$):

$$\alpha_{\text{doa}} = \frac{1}{2}\left(\frac{P_1{}^t}{1 + \rho P_1{}^t \alpha_{\text{dod}}} + \frac{2 - P_1{}^t}{1 + 2\rho\alpha_{\text{dod}} - \rho P_1{}^t \alpha_{\text{dod}}}\right)$$

$$\alpha_{\text{dod}} = \frac{1}{2}\left(\frac{P_1{}^r}{1 + \rho P_1{}^r \alpha_{\text{doa}}} + \frac{2 - P_1{}^r}{1 + 2\rho\alpha_{\text{doa}} - \rho P_1{}^r \alpha_{\text{doa}}}\right)$$

and

$$\begin{aligned}\mu_{\text{double}} &= \frac{1}{2}(\ln(1 + \rho P_1{}^t \alpha_{\text{dod}}) + \ln(1 + 2\rho\alpha_{\text{dod}} - \rho P_1{}^t \alpha_{\text{dod}})) \\ &+ \frac{1}{2}(\ln(1 + \rho P_1{}^r \alpha_{\text{doa}}) + \ln(1 + 2\rho\alpha_{\text{doa}} - \rho P_1{}^r \alpha_{\text{doa}})) - \rho\alpha_{\text{doa}}\alpha_{\text{dod}}\end{aligned}$$

The figure acknowledges the fact that the best throughput is obtained when all the steering directions have the same power on both sides.

D. Fourier versus Random Directions: Equal Power Case

In this section, we would like to quantify the impact of the steering matrix on the ergodic mutual information. The answer has a direct impact on the understanding and

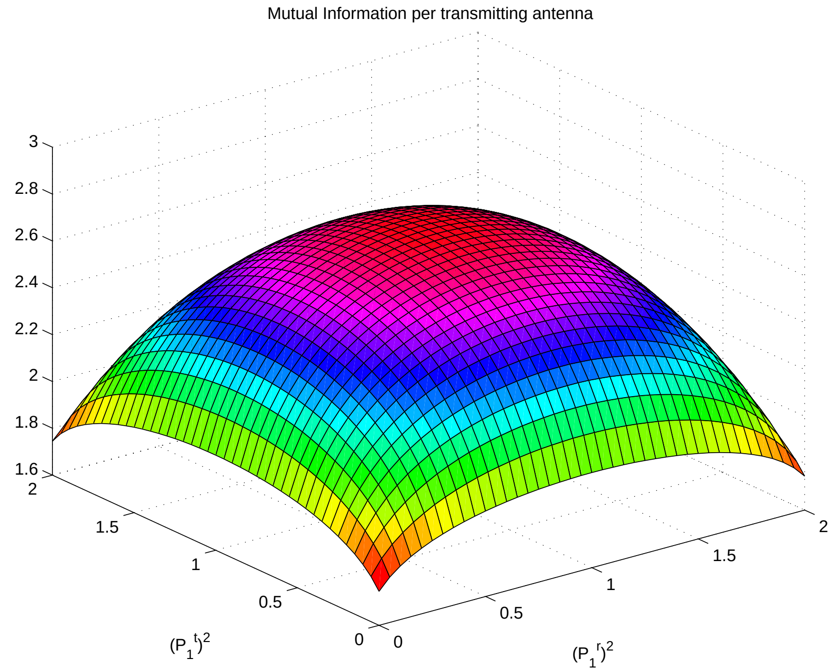

Fig. 7. Mean capacity per transmitting antenna versus ${P_1}^r$ and ${P_1}^t$ at 10dB.

the design of future mobile systems. In this respect, two extreme cases are compared, the Fourier and random directions case. For the random directions context, we will suppose that the entries of matrix $\mathbf{\Phi}$ and $\mathbf{\Psi}$:

1- Are a realization of independent and uniformly distributed exponential variables with zero mean and unit variance. This can be seen as a limiting case of near field scattering (all the rays, for a given scatterer do not come from the same direction). We agree on the fact that the near-field case is more complicated as the phases are not completely independent but linked through the geometry of the antenna. We mainly use the random approach in order to have tractable mutual information formulas. This case will be referred to random i.i.d directions.

2- Represent ULA antennas with the far field approximation where the scatterers are randomly located. In this case, the s_t and s_r phases of respectively matrices $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are uniformly distributed over $[0, 2\pi]$. This case will be referred to random directions with ULA.

For the random i.i.d directions, we can derive an explicit expression of the mean mutual information. The limiting eigenvalue distribution of $\frac{1}{s_r}\mathbf{\Phi}^H\mathbf{\Phi}$ and $\frac{1}{s_t}\mathbf{\Psi}\mathbf{\Psi}^H$ are well known in the literature ([54]) and Proposition 3 can be applied straightforwardly. However, we will take Müller's approach, as our framework is a particular case of [55] where he introduces an N fold scattering model as a product of N i.i.d random matrices $\mathbf{H} = \prod_{i=1}^{N} \mathbf{M}_i$. Using free probability theory[8], he proves the almost sure convergence of the limiting eigenvalue distribution of matrix $\mathbf{H}$ and gives an explicit form of its Stieljes transform. In the case considered here, $\mathbf{H} = \mathbf{\Phi\Theta\Psi}$ is the product of three random matrices. Using the results in [55], it can be easily shown that the Stieljes transform $m_{\mathbf{H^H H}}(x)$ is solution of the following equation:

$$\begin{aligned} & m_{\mathbf{H^H H}}(-x)\left(\frac{x m_{\mathbf{H^H H}}(-x) - 1 + \xi_1}{\xi_1}\frac{x m_{\mathbf{H^H H}}(-x) - 1 + \xi}{\xi}\frac{x m_{\mathbf{H^H H}}(-x) - 1 + \gamma\rho}{\gamma\rho}\right) \\ + \quad & x m_{\mathbf{H^H H}}(-x) = 1 \end{aligned}$$

Since $m_{\mathbf{H^H H}}(\frac{-1}{\rho}) = \rho(1 - \rho\frac{d\mu}{d\rho})$, the asymptotic mutual information per transmitting antenna can be obtained by solving the following equation:

$$\rho(1-\rho\frac{d\mu}{d\rho})\left[(1-\frac{\rho}{\xi_1}\frac{d\mu}{d\rho})(1-\frac{\rho}{\xi}\frac{d\mu}{d\rho})(1-\frac{\rho}{\gamma\xi}\frac{d\mu}{d\rho})+\frac{1}{\rho}\right]=1$$

and numerical integration of $\frac{d\mu}{d\rho}$ through:

$$\mu = \int \frac{d\mu}{d\rho} d\rho$$

with the boundary condition :$\lim_{\rho\to 0}\mu(\rho) = 0$

We have plotted in Figure 8 the theoretical asymptotic mean mutual information per receiving antenna of the random i.i.d directions scenario at 10 dB for various ratio of scatterers s_r ($\frac{s_r}{n_r}$ ranges from 0 to 1): as a matter of fact, since $n_r = n_t$, it does not matter whether one plots the mutual information with respect to $\frac{s_r}{n_r}$ or $\frac{s_t}{n_t}$. s_t has been chosen to

[8]Free Probability [56], [57] is a non-commutative probability theory, in which the concept of independence of classical probability is replaced by that of freeness. Voiculescu [58], [59], [60] discovered very important relations between free probability theory and random matrix theory. He showed in particular that random matrices can be considered as typical non-commutative random variables. To the authors knowledge, the first use of free probability in the telecommunication field was made by Evans and Tse in 1999 [61]. Since that date, it has been used for the performance analysis of several transmission schemes (CDMA [62], [63], OFDM [43], [46], [44] and MIMO [64], [65], [47]). Note that Free Probability is not only a prediction tool but has been proved by several authors to be very useful in the practical design of low-complex detectors [66], [67], [68] (Multi-stage detectors...)

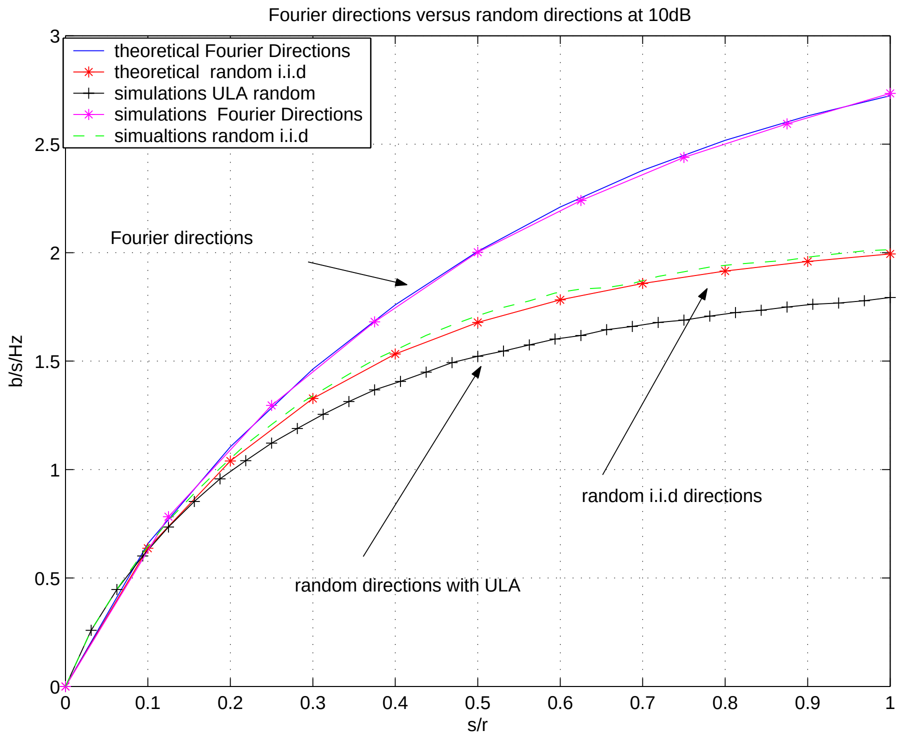

Fig. 8. Fourier versus random directions at 10dB.

be equal to n_t. We have also plotted a simulated curve with a system of 8×8 antennas. The angles of arrival were generated randomly according to a uniform distribution and kept fixed during all the trials. A close match between the theoretical formula and the simulations is obtained. We have also plotted the asymptotic mean mutual information of the far field ULA scenario where the scatterers are given by Fourier directions (see section.IV-C.1). One can observe that scatterers on Fourier directions yield better performance than scatterers on random i.i.d directions. In fact, in the Fourier direction case and in the case of $s_r = n_r = s_t = n_t$, the DoA matrix $\mathbf{\Phi}$ and DoD matrix $\mathbf{\Psi}$ are unitary Fourier matrices and have therefore no effect on $\mathbf{\Theta}_{s_r \times s_t}$. However, in the random i.i.d directions scenario, the non-unitary steering matrix $\mathbf{\Phi}$ and $\mathbf{\Psi}$ have a correlation effect on matrix $\mathbf{\Theta}_{s_r \times s_t}$. We have also plotted the case of random directions with ULA, which is outperformed by the random i.i.d directions case. When the scatterers are randomly located, this last example argues in favor of scatterers located near the antenna with antennas having no structured (ULA for example) geometry.

E. Some Considerations on the Directions of Arrival/Departure based Model

Without going into detailed calculus (the proof is a special case of Proposition 1), imagine that the modeler has knowledge of the directions of arrival and their respective power as well as the fact that the channel carries some energy. It can be shown that the resulting DoA model has the following form

$$\mathbf{H} = \frac{1}{\sqrt{s_r}}\mathbf{\Phi}\mathbf{P}^{\mathbf{r}\frac{1}{2}}\mathbf{\Theta}$$

and principle of maximum entropy will assign independent zero mean complex Gaussian entries to the matrix $\mathbf{\Theta}_{n_r \times s_t}$.

E.1 General Case

We are interested in the behavior of $I^M(n_t, n_r, s_r, \rho) = \log_2 \det\left(\mathbf{I}_{n_t} + \frac{\rho}{n_t}\mathbf{H}^H\mathbf{H}\right)$ and in particular the eigenvalue distribution of

$$\frac{1}{n_t}\mathbf{H}^H\mathbf{H} = \frac{1}{n_t s_r}\mathbf{\Theta}^H_{s_r \times n_t}\mathbf{P}^{\mathbf{r}\frac{1}{2}}\mathbf{\Phi}^H_{n_r \times s_r}\mathbf{\Phi}_{n_r \times s_r}\mathbf{P}^{\mathbf{r}\frac{1}{2}}\mathbf{\Theta}_{s_r \times n_t}$$

Let us first make some assumptions[9] on the matrix of the directions of arrival.

Assumption: The matrix size $\frac{1}{s_r}\mathbf{P}^{\mathbf{r}\frac{1}{2}}\mathbf{\Phi}^H_{n_r \times s_r}\mathbf{\Phi}_{n_r \times s_r}\mathbf{P}^{\mathbf{r}\frac{1}{2}}$ grows large with $\gamma = \frac{n_r}{s_r}$ remaining fixed such that the empirical eigenvalue distribution S_{s_r,n_r} of $\frac{1}{s_r}\mathbf{P}^{\mathbf{r}\frac{1}{2}}\mathbf{\Phi}^H_{n_r \times s_r}\mathbf{\Phi}_{n_r \times s_r}\mathbf{P}^{\mathbf{r}\frac{1}{2}}$ converges in distribution to a fixed distribution S_{doa}

$$S_{s_r,n_r}(\lambda) = \frac{1}{s_r} \mid \{j : \lambda_j \leq \lambda\} \mid \rightarrow S_{\text{doa}}(\lambda)$$

Theorem 2: As $n_t \to \infty$ with $s_r = \xi n_t$, $I^M{}_{\text{doa}}(n_t, n_r, s_r, \rho) - n_t \mu_{\text{doa}}(\xi, \gamma, \rho)$ converges in distribution to a $N(0, \sigma^2{}_{\text{doa}})$ random variable where:

$$\mu_{\text{doa}}(\xi, \gamma, \rho) = \int_0^\infty \ln(1 + \rho\lambda) dF_{\text{doa}}(\lambda)$$

$$m_{f_{\text{doa}}}(z) = \int \frac{dF_{\text{doa}}(\lambda)}{\lambda - z}$$

$$z = \frac{-1}{m_{f_{\text{doa}}}(z)} + \xi \int \frac{x}{1 + m_{f_{\text{doa}}}(z)x} dS_{\text{doa}}(x)$$

[9] Note that the assumption is here used in a mathematical meaning, not in a modelling perspective.

$$\sigma^2{}_{\text{doa}} = -\frac{1}{4\pi^2}\int_{C_x}\int_{C_y}\frac{\ln(1+\rho x)\ln(1+\rho y)}{(m_{f_{\text{doa}}}(x)-m_{f_{\text{doa}}}(y))^2}m'_{f_{\text{doa}}}(x)m'_{f_{\text{doa}}}(y)dxdy$$

C_x and C_y are any closed contour that enclose the support of F_{doa} but not $\frac{-1}{\rho}$.

f_{doa} is the limiting eigenvalue distribution of $\frac{1}{n_t}\mathbf{H}^H\mathbf{H}$ in the DoA based model while S_{doa} is the limiting eigenvalue distribution of $\frac{1}{s_r}\mathbf{P}^{\mathbf{r}\frac{1}{2}}\mathbf{\Phi}^H_{n_r\times s_r}\mathbf{\Phi}_{n_r\times s_r}\mathbf{P}^{\mathbf{r}\frac{1}{2}}$. This result is based on contribution [37]. Hence, if the directions of arrival and the powers can be estimated, one can completely determine the distribution of the mutual information by solving the previous equations. From a practical point of view, the receiver estimates the angles of arrival and determines the mean and the variance of the mutual information. This information is then sent back to the transmitter for scheduling the network [10]. One interesting point of the feedback mechanism is that asymptotically only two values (the mean and the variance) are needed. This reduces drastically the overhead of feedback transmissions.

Suppose that the DoA-distribution S_{doa} is given (using DoA channel estimation techniques for example). In this case, how does one derive μ_{doa} without explicitly knowing $F_{\text{doa}}(\lambda)$? One can first of all notice that:

$$\begin{aligned}\frac{d\mu_{\text{doa}}}{d\rho} &= \int_0^\infty \frac{\lambda}{1+\rho\lambda}dF_{\text{doa}}(\lambda)\\ &= \frac{1}{\rho}\int_0^\infty \frac{\rho\lambda+1-1}{1+\rho\lambda}dF_{\text{doa}}(\lambda)\\ &= \frac{1}{\rho}-\frac{1}{\rho^2}m_{f_{\text{doa}}}(-\frac{1}{\rho})\end{aligned}$$

Therefore, $m_{f_{\text{doa}}}(-\frac{1}{\rho}) = \rho\left(1-\rho\frac{d\mu_{\text{doa}}}{d\rho}\right)$ and based on the result of theorem 2, we have:

$$-\frac{1}{\rho} = \frac{-1}{\rho(1-\rho(\frac{d\mu_{\text{doa}}}{d\rho}))} + \xi\int\frac{x}{1+x\rho(1-\rho(\frac{d\mu_{\text{doa}}}{d\rho}))}dS_{\text{doa}}(x) \tag{23}$$

and

$$\mu_{\text{doa}}(\rho) = \int_0^\rho (\frac{d\mu_{\text{doa}}}{d\rho})d\rho$$

In the high SNR regime, the following result holds:

[10]Some results on the capacity of a MIMO multi-user network (where all the users have different angles of arrival) in the large system limit (high number of antennas) can be found in [69].

Proposition 8: In the high SNR regime, the mean mutual information of the DoA based model converges to:

$$\mathbb{E}(I^M) = \min\left(n_t, s_r \int_{\lambda>0} dS_{\text{doa}}(\lambda)\right) \ln(\rho)$$

Proof: Let $r = \rho\frac{d\mu_{\text{doa}}}{d\rho} \leq 1$ ($n_t r$ denotes in fact the multiplexing gain). According to equation (23), we have:

$$-\frac{1}{\rho} = \frac{-1}{\rho(1-r)} + \xi \int \frac{x}{1+x\rho(1-r)} dS_{\text{doa}}(x)$$

and at high SNR:

$$-1 = \frac{-1}{(1-r)} + \frac{\xi}{1-r}\int_{\lambda>0} dS_{\text{doa}}(\lambda)$$

which yields:

$$r = \begin{cases} \xi \int_{\lambda>0} dS_{\text{doa}}(\lambda) & \text{if } \xi \int_{\lambda>0} dS_{\text{doa}}(\lambda) \leq 1 \\ 1 & \text{otherwise} \end{cases} \tag{24}$$

and proves the result. ■

E.2 Non-Equal Power Case on Fourier Directions

We consider in this case that there is a finite set of K_r distinct amplitudes $\sqrt{P_i^r}$ with weight ${l_i}^r$ such as $\sum_{i=1}^{K_r} {l_i}^r = 1$. As a consequence, the limiting eigenvalue distribution S_{doa} of $\frac{1}{s_r}\mathbf{P^r}^{\frac{1}{2}}\mathbf{\Phi}_{n_r\times s_r}^H\mathbf{\Phi}_{n_r\times s_r}\mathbf{P^r}^{\frac{1}{2}}$ has the following expression:

$$S_{\text{doa}}(\lambda) = \sum_{i=1}^{K_r} {l_i}^r \delta(\lambda - \gamma {P_i}^r)$$

Proposition 9: In the non-equal power case with Fourier directions, $\mu_{\text{doa}}(\xi,\gamma,\rho)$ and $\sigma^2_{\text{doa}}(\xi,\gamma,\rho)$ are equal to:

$$\mu_{\text{doa}}(\xi,\gamma,\rho) = -\ln(\alpha_{\text{doa}}) + \xi\sum_{i=1}^{K_r} {l_i}^r \ln(1+\rho {P_i}^r\gamma\alpha_{\text{doa}}) - (1-\alpha_{\text{doa}})$$

and

$$\sigma^2_{\text{doa}}(\xi,\gamma,\rho) = -\ln\left[1 - \rho^2\xi{\alpha_{\text{doa}}}^2\sum_{i=1}^{K_r} {l_i}^r \frac{(\gamma {P_i}^r)^2}{(1+\rho\gamma {P_i}^r\alpha_{\text{doa}})^2}\right]$$

with

$$\sum_{i=1}^{K_r} \frac{l_i{}^r}{1+\rho\gamma P_i{}^r \alpha_{\text{doa}}} = \frac{\alpha_{\text{doa}}}{\xi} - \frac{1}{\xi} + 1$$

Proof: The proof is provided in the appendix. For the mean μ_{doa}, the proof is an application of the general Proposition 6 in the case of interest and is provided in the appendix. For the variance, results of [37] are used. ■

Note that α_{doa} is related to the Stieljes transform $m_{f_{\text{doa}}}$ of f_{doa} by:

$$m_{f_{\text{doa}}}(\frac{-1}{\rho}) = \rho(1-\alpha_{\text{doa}}).$$

In Figure 9 and Figure 10, simulations have been conducted in the two power case with $n_r = n_t = 8$ antennas. We impose $P_1{}^r = 2 - P_2{}^r$, $l_1{}^r = l_2{}^r = \frac{1}{2}$ and $s_r = 8$. In this case, we have ($\gamma = \frac{n_r}{s_r} = 1$ and $\xi = \frac{s_r}{n_t} = 1$):

$$\frac{1}{2}\left[\frac{1}{1+\rho P_1{}^r \alpha_{\text{doa}}} + \frac{1}{1+\rho(2-P_1{}^r)\alpha_{\text{doa}}}\right] = \alpha_{\text{doa}}$$

with

$$\mu_{\text{doa}} = -\ln(\alpha_{\text{doa}}) + \frac{1}{2}\left(\ln(1+\rho P_1{}^r \alpha_{\text{doa}}) + \ln(1+2\rho\alpha_{\text{doa}} - \rho P_1{}^r \alpha_{\text{doa}})\right) - (1-\alpha_{\text{doa}})$$

and

$$\sigma^2_{\text{doa}} = -\ln\left[1 - \frac{\rho^2\alpha^2_{\text{doa}}}{2}\left(\frac{(P_1{}^r)^2}{(1+\rho(P_1{}^r\alpha_{\text{doa}})^2} + \frac{(2-P_1{}^r)^2}{(1+\rho(2-P_1{}^r)\alpha_{\text{doa}})^2}\right)\right]$$

In Figure 9, the asymptotic mean mutual information has been plotted versus the amplitude $\sqrt{P_1{}^r}$. A close match between theoretical predictions and simulations is obtained for a low number of antennas (8×8 MIMO system). More importantly, one can observe that the best throughput is obtained when all the steering directions have equal power. The close match also pertains for the variance (see Figure 10) where the highest variance is obtained in the equal power case. In terms of outage mutual information, the equal power case is also the one which maximizes that criteria (see Figure 11 and proposition 7). Intuitively, one can easily understand this observation: any imbalances of power will reduce the effective number of scatterers and therefore the diversity generated by the environment.

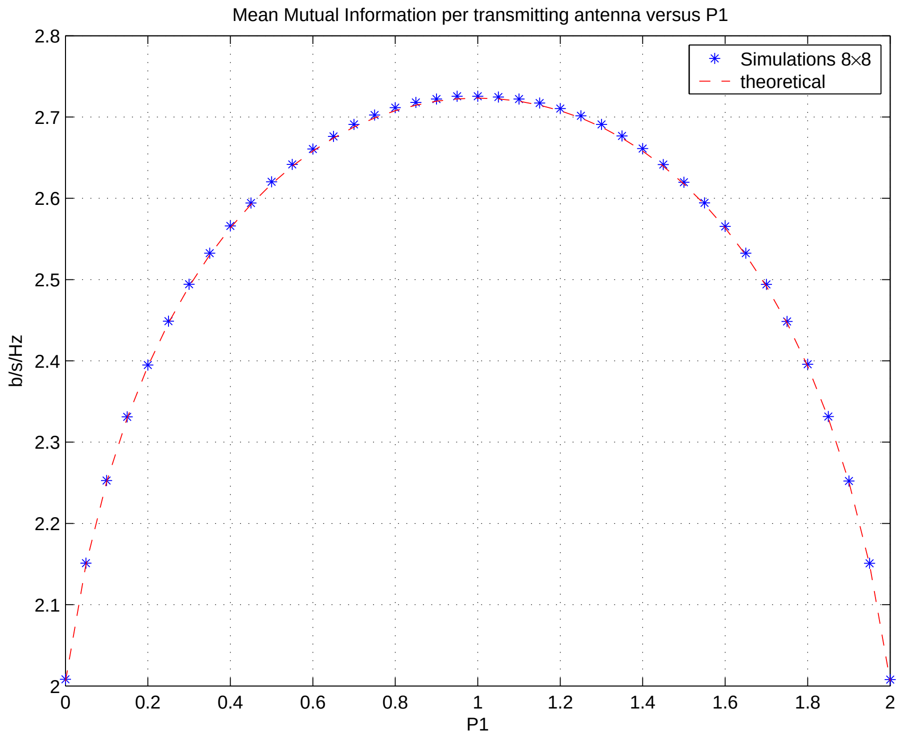

Fig. 9. Mean capacity per transmitting antenna versus ${P_1}^r$ at 10dB for an 8×8 DoA based model.

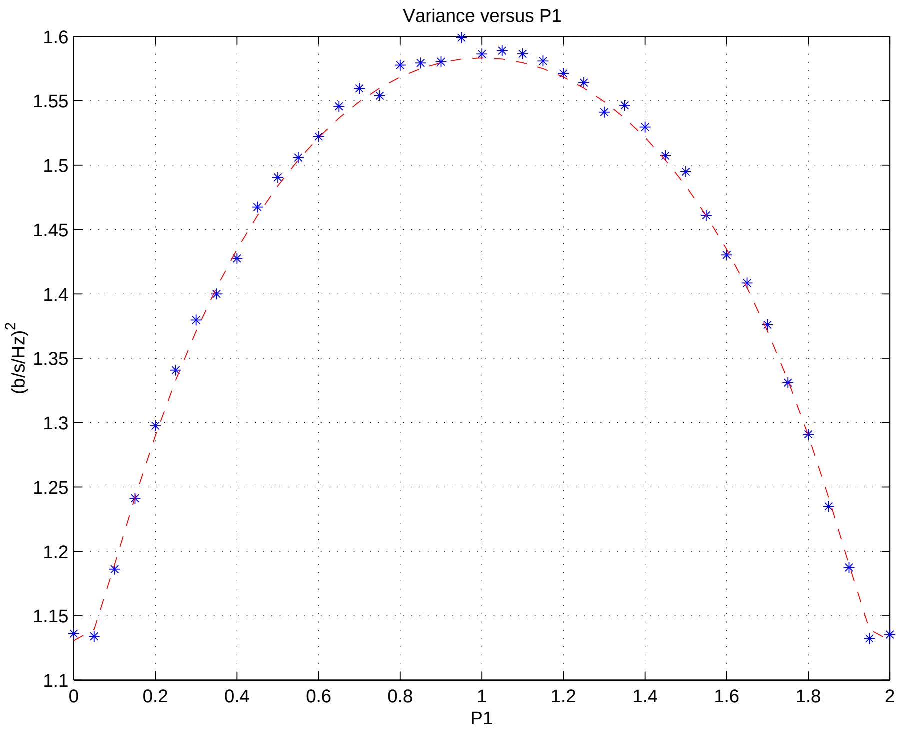

Fig. 10. Variance versus ${P_1}^r$ at 10dB for an 8×8 DoA based model.

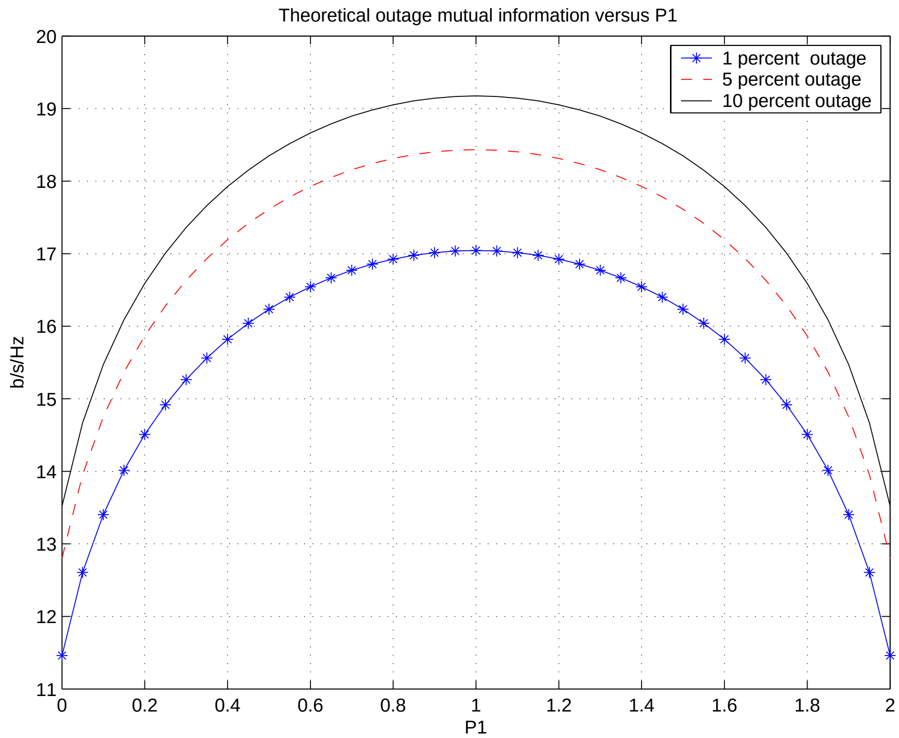

Fig. 11. Outage mutual information versus ${P_1}^r$ at 10dB for an 8×8 DoA based model.

E.3 Remarks on the Directions of Departure based Model

If only the directions of departure (with their respective power) and channel energy is known, the previous methodology yields the following DoD based model:

$$\mathbf{H} = \frac{1}{\sqrt{s_t}}\mathbf{\Theta}_{n_r \times s_t}\begin{pmatrix} \sqrt{P_1^t} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_t}^t} \end{pmatrix}\begin{pmatrix} e^{j\psi_{1,1}} & \dots & e^{j\psi_{1,n_t}} \\ \vdots & \ddots & \vdots \\ e^{j\psi_{s_t,1}} & \dots & e^{j\psi_{s_t,n_t}} \end{pmatrix}?$$

$\mathbf{\Psi}$ ($\mathbf{H} = \frac{1}{\sqrt{s_t}}\mathbf{\Theta}_{n_r \times s_t}\mathbf{P^t}^{\frac{1}{2}}\mathbf{\Psi}_{s_t \times n_t}$) is a $s_t \times n_t$ matrix (s_t is the number of scatterers) which represents the directions of departure from the transmitting antennas to randomly positioned scatterers with respective powers $\mathbf{P^t}$. $\mathbf{\Theta}_{n_r \times s_t}$ is an $n_r \times s_t$ matrix which represents the scattering environment between the receiving antennas and the scatterers.

For deriving the mutual information, it is straightforward to notice that the same result (due to the duality between the directions of arrival and departure based model) as the DoA based model is obtained if one:

- normalizes the mutual information with respect to the number of receive antennas.
- n_t, s_r, P^r are exchanged with n_r, s_t and P^t.
- the SNR ρ is replaced by $\frac{n_r}{n_t}\rho$.

In other words, the asymptotic Gaussian behavior remains valid and we have:

$$I^M{}_{\text{dod}}(n_t, n_r, s_t, \rho) = I^M{}_{\text{doa}}(n_r, n_t, s_r, \frac{n_r}{n_t}\rho)$$

V. Knowledge of the directions of arrival, departure, delay, bandwidth, power: frequency selective channel model with time variance

A. Model

The modeler wants to derive a consistent model taking into account the direction of arrivals and respective power profile, directions of departure and respective power profile, delay, Doppler effect. As a starting point, the modeler assumes that the position of the transmitter and receiver changes in time. However, the scattering environment (the buildings, trees,...) does not change and stays in the same position during the transmission. Let $\mathbf{v_t}$ and $\mathbf{v_r}$ be respectively the vector speed of the transmitter and the receiver with respect to a terrestrial reference (see Figure 12). Let $\mathbf{s}^t_{ij}$ be the signal between the transmitting antenna i and the first scatterer j. Assuming that the signal can be written in an exponential form (plane wave solution of the Maxwell equations) then:

$$\begin{aligned} \mathbf{s}^t_{ij}(t) &= \mathbf{s_0} e^{j(\mathbf{k}^t_{ij}{}^T(\mathbf{v_t}t + \mathbf{d^t}_{\mathbf{ij}}) + 2\pi f_c t)} \\ &= \mathbf{s_0} e^{j2\pi(\frac{f_c \mathbf{u}^t_{ij}\mathbf{v_t}}{c}t + f_c t)} e^{j\psi_{ij}} \end{aligned}$$

Here, f_c is the carrier frequency, $\mathbf{d^t}_{ij}$ is the initial vector distance between antenna i and scatterer j ($\psi_{ij} = \mathbf{k}^t_{ij}{}^T \mathbf{d^t}_{ij}$ is the scalar product between vector $\mathbf{k}^t_{ij}$ and vector $\mathbf{d^t}_{ij}$), $\mathbf{k}^t_{ij}$ is such as $\mathbf{k}^t_{ij} = \frac{2\pi}{\lambda}\mathbf{u^t_{ij}} = \frac{2\pi f_c}{c}\mathbf{u^t_{ij}}$. The quantity $\frac{1}{2\pi}\mathbf{k^t_{ij}}^T\mathbf{v_t}$ represents the Doppler effect.

In the same way, if we define $\mathbf{s}^r_{ij}(t)$ as the signal between the receiving antenna j and the scatterer i, then:

$$s^r_{ij}(t) = \mathbf{s_0} e^{j(2\pi(\frac{f_c \mathbf{v_r} \mathbf{u}^r_{ij}}{c}t + f_c t))} e^{j\phi_{ij}}$$

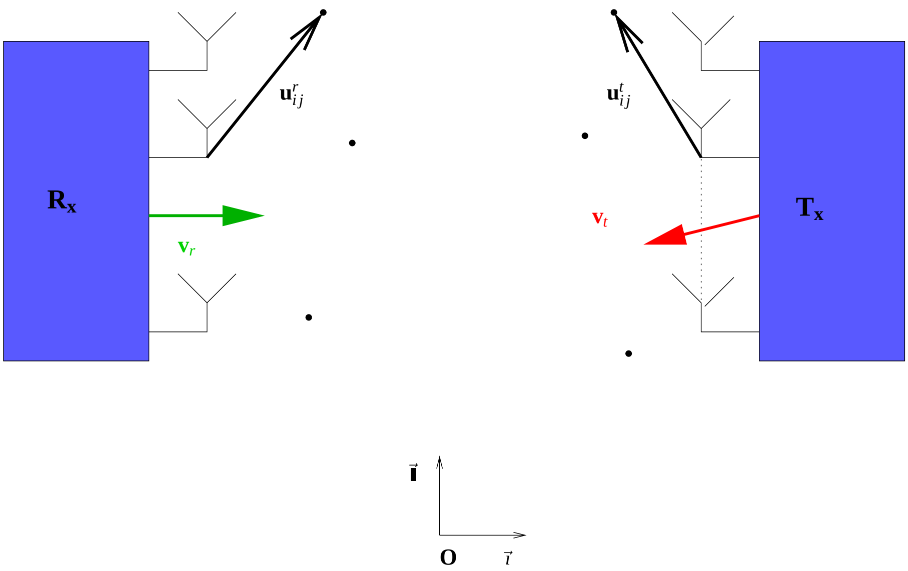

Fig. 12. Moving antennas.

In all the following, the modeler supposes as a state of knowledge the following parameters:

- speed $\mathbf{v_r}$.
- speed $\mathbf{v_t}$.
- the angle of departure from the transmitting antenna to the scatterers ψ_{ij} and power P^t_j.
- the angle of arrival from the scatterers to the receiving antenna ϕ_{ij} and power P^r_j.

The modeler has however no knowledge of what happens in between except the fact that a signal going from a steering vector of departure j to a steering vector of arrival i has a certain delay τ_{ij} due to possible single bounce or multiple bounces on different objects. The modeler also knows that objects do not move between the two sets of scatterers. The $s_r \times s_t$ delay matrix linking each DoA and DoD has the following structure:

$$\mathbf{D}_{s_r \times s_t}(f) = \begin{pmatrix} e^{-j2\pi f \tau_{1,1}} & \ldots & e^{-j2\pi f \tau_{1,s_t}} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi f \tau_{s_r,1}} & \ldots & e^{-j2\pi f \tau_{s_r,s_t}} \end{pmatrix}$$

The modeler also supposes as a given state of knowledge the fact that matrix $\mathbf{H}$ has a certain energy. Based on this state of knowledge, the modeler wants to model the $s_r \times s_t$

matrix $\mathbf{\Theta}_{s_r \times s_t}$ in the following representation:

$$\mathbf{H}(f,t) = \frac{1}{\sqrt{s_r s_t}} \begin{pmatrix} e^{j(\phi_{1,1}+2\pi\frac{f\mathbf{u}^r_{11}\mathbf{v_r}}{c}t)} & \dots & e^{j(\phi_{1,s_r}+2\pi\frac{f\mathbf{u}^r_{1s_r}\mathbf{v_r}}{c}t)} \\ \vdots & \ddots & \vdots \\ e^{j(\phi_{r,1}+2\pi\frac{f\mathbf{u_{nr1}}^r\mathbf{v_r}}{c}t)} & \dots & e^{j(\phi_{n_r,s_r}+2\pi\frac{f\mathbf{u}^r_{n_r s_r}\mathbf{v_r}}{c}t)} \end{pmatrix} \begin{pmatrix} \sqrt{P^r_1} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P^r_{s_r}} \end{pmatrix}$$
$$\mathbf{\Theta}_{s_r \times s_t} \odot \mathbf{D}_{s_r \times s_t}(f)$$
$$\begin{pmatrix} \sqrt{P^t_1} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P^t_{s_r}} \end{pmatrix} \begin{pmatrix} e^{j(\psi_{1,1}+2\pi\frac{f\mathbf{u}^t_{11}\mathbf{v_t}}{c}t)} & \dots & e^{j(\psi_{1,n_t}+2\pi\frac{f\mathbf{u}^t_{1n_t}\mathbf{v_t}}{c}t)} \\ \vdots & \ddots & \vdots \\ e^{j(\psi_{s_t,1}+2\pi\frac{f\mathbf{u}^t_{s_1 1}\mathbf{v_t}}{c}t)} & \dots & e^{j(\psi_{s_t,n_t}+2\pi\frac{f\mathbf{u}^t_{s_t n_t}\mathbf{v_t}}{c}t)} \end{pmatrix}$$

$\odot$ represents the Hadamard product defined as $c_{ij} = a_{ij}b_{ij}$ for a product matrix $\mathbf{C} = \mathbf{A} \odot \mathbf{B}$. As previously stated, one has to comply with the following constraints:

- Matrix $\mathbf{H}(f,t)$ has a certain energy.
- Consistency argument: if the DoA, DoD, powers, the delays, the Doppler effects are unknown then matrix $\mathbf{H}$ should be assigned an i.i.d Gaussian distribution.

Proposition 10: $\mathbf{\Theta}_{s_r \times s_t}$ i.i.d zero mean Gaussian with variance 1 is solution of the consistency argument and maximizes entropy.[11]

Proof: We will not go into the details but only provide the guidelines of the proof. First, remark that if $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are unknown, then the principle of maximum entropy attributes i.i.d uniform distribution to the angles ϕ_{ij} and ψ_{ij}. But what probability distribution should the modeler attribute to the delays and the Doppler effects when no information is available?

- **Delays**: The modeler knows that there is, due to measurements performed in the area, a maximum possible delay for the information to go from the transmitter to the receiver $\tau_{\max}$. The principle of maximum entropy attributes therefore a uniform distribution to all the delays τ_{ij} such as $P(\tau_{ij}) = \frac{1}{\tau_{\max}}$ with $\tau_{ij} \in [0, \tau_{\max}]$.
- **Doppler effect**: The modeler knows that the speed of the transmitter and receiver can not exceed a certain limit v_{limit} (in the least favorable case, v_{limit} would be equal

[11] Why does normality always appear in our models? Well, the answer is quite simple. In all this paper, we have always limited ourselves to the second moment of the channel. If more moments are available, then normal distributions would not appear in general.

to the speed of light) but if the transmission occurs in a city, the usual car speed limit can be taken as an upper bound. In this case, the speed v_t and v_r have also a uniform distribution such as $P(v_t) = P(v_r) = \frac{1}{v_{\text{limit}}}$. Moreover, if $\mathbf{v}_t = v_t \cos(\alpha_t)\tilde{\imath} + v_t \sin(\alpha_t)\tilde{\jmath}$, $\mathbf{v}_r = v_r \cos(\alpha_r)\tilde{\imath} + v_r \sin(\alpha_r)\tilde{\jmath}$, ${u^t}_{ij} = \cos({\beta^t}_{ij})\tilde{\imath} + \sin({\beta^t}_{ij})\tilde{\jmath}$ and ${u^r}_{ij} = \cos({\beta^r}_{ij})\tilde{\imath} + \sin({\beta^r}_{ij})\tilde{\jmath}$, the modeler will attribute a uniform distribution over 2π to the angles α_t, α_r,${\beta^t}_{ij}$ and ${\beta^r}_{ij}$.

With all these probability distributions derived and using the same methodology as in the narrowband (in terms of frequency selectivity) MIMO model proof, one can easily show that $\mathbf{\Theta}_{s_r \times s_t}$ i.i.d Gaussian is solution of the consistency argument and maximizes entropy.

Note that in the case $f = 0$, $\mathbf{v_t} = 0$ and $\mathbf{v_r} = 0$, the same model as the narrowband model is obtained. If more information is available on correlation or different variances of frequency paths, then this information can be incorporated in the matrix $\mathbf{D}_{s_r \times s_t}$, also known as the channel pattern mask [25]. Note that in the case of a ULA geometry and in the Fourier directions, we have $u^r_{ij} = u^r_j$ (any column of matrix $\mathbf{\Phi}$ has a given direction) and $u^t_{ij} = u^t_i$ (any line of matrix $\mathbf{\Psi}$ has a given direction). Therefore, the channel model simplifies to:

$$\mathbf{H}(f,t) = \frac{1}{\sqrt{s_r s_t}} \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ e^{j2\pi \frac{d(n_r-1)\sin(\phi_1)}{\lambda}} & \dots & e^{j2\pi \frac{d(n_r-1)\sin(\phi_{s_r})}{\lambda}} \end{pmatrix} \mathbf{\Theta}_{s_r \times s_t} \odot \mathbf{D}_{s_r \times s_t}(f,t) \begin{pmatrix} 1 & \dots & e^{j2\pi \frac{d(n_t-1)\sin(\psi_1)}{\lambda}} \\ \vdots & \ddots & \vdots \\ 1 & \dots & e^{j2\pi \frac{d(n_t-1)\sin(\psi_{s_t})}{\lambda}} \end{pmatrix}$$

In this case, the pattern mask $\mathbf{D}_{s_r \times s_t}$ has the following form:

$$\mathbf{D}_{s_r \times s_t}(f,t) = \begin{pmatrix} \sqrt{P^r_1}\sqrt{P^t_1} e^{-j2\pi f \tau_{1,1}} e^{j2\pi \frac{ft}{c}(\mathbf{u}^r_1 \mathbf{v}_r + \mathbf{u}^t_1 \mathbf{v}_t)} & \dots & \sqrt{P^r_1}\sqrt{P^t_{s_t}} e^{-j2\pi f \tau_{1,s_t}} e^{j2\pi \frac{ft}{c}(\mathbf{u}^r_1 \mathbf{v}_r + \mathbf{u}^t_{s_t} \mathbf{v}_t)} \\ \vdots & \ddots & \vdots \\ \sqrt{P^r_{s_r}}\sqrt{P^t_1} e^{-j2\pi f \tau_{s_r,1}} e^{j2\pi \frac{ft}{c}(\mathbf{u}^r_s \mathbf{v}_r + \mathbf{u}^t_1 \mathbf{v}_t)} & \dots & \sqrt{P^r_{s_r}}\sqrt{P^t_{s_t}} e^{-j2\pi f \tau_{s_r,s_t}} e^{j2\pi \frac{f\eta}{c}(\mathbf{u}^r_s \mathbf{v}_r + \mathbf{u}^t_{s_t} \mathbf{v}_t)} \end{pmatrix}$$

Although we take into account many parameters, the final model is quite simple. It is the product of three matrices: Matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$ taking into account the directions of arrival and departure; matrix $\mathbf{\Theta}_{s_r \times s_t} \odot \mathbf{D}_{s_r \times s_t}$ which is an independent Gaussian matrix with different variances. The frequency selectivity of the channel is therefore taken into account in the phase of each entry of the matrix $\mathbf{\Theta}_{s_r \times s_t} \odot \mathbf{D}_{s_r \times s_t}(f, t)$.

Remark: In the case of a one antenna system link ($n_r = 1$ and $n_t = 1$), we obtain:

$$
\begin{aligned}
\mathbf{H}(f,t) &= \frac{1}{\sqrt{s_r s_t}} \left[\begin{array}{ccc} e^{j(\phi_1 + 2\pi \frac{f \mathbf{u}_1^r \mathbf{v_r}}{c} t)} & \ldots & e^{j(\phi_{s_r} + 2\pi \frac{f \mathbf{u}_s^r \mathbf{v_r}}{c} t)} \end{array} \right] \begin{pmatrix} \sqrt{P_1^r} & 0 & \ldots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_r}^r} \end{pmatrix} \\
& \quad \mathbf{\Theta}_{s_r \times s_t} \odot \mathbf{D}_{s_r \times s_t}(f) \begin{pmatrix} \sqrt{P_1^t} & 0 & \ldots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_t}^t} \end{pmatrix} \left[\begin{array}{c} e^{j(\psi_1 + 2\pi \frac{f \mathbf{u}_1^t \mathbf{v_t}}{c} t)} \\ \vdots \\ e^{j(\psi_{s_t} + 2\pi \frac{f \mathbf{u}_{s_t}^t \mathbf{v_t}}{c} t)} \end{array} \right] \\
&= \frac{1}{\sqrt{s_r s_t}} \left[\begin{array}{ccc} \sum_{k=1}^{s_r} \theta_{k,1} P_k^r e^{j(\phi_k + 2\pi \frac{f \mathbf{u}_k^r \mathbf{v_r}}{c} t)} e^{-j2\pi f \tau_{k,1}} & \ldots & \sum_{k=1}^{s_r} \theta_{k,s_r} P_k^r e^{j(\phi_k + 2\pi \frac{f \mathbf{u}_k^r \mathbf{v_r}}{c} t)} e^{-j2\pi f \tau_{k,s_r}} \end{array} \right] \\
& \quad \begin{pmatrix} \sqrt{P_1^t} & 0 & \ldots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_t}^t} \end{pmatrix} \begin{pmatrix} e^{j(\psi_1 + 2\pi \frac{f \mathbf{u}_1^t \mathbf{v_t}}{c} t)} \\ \vdots \\ e^{j(\psi_{s_t} + 2\pi \frac{f \mathbf{u}_{s_t}^t \mathbf{v_t}}{c} t)} \end{pmatrix} \\
&= \sum_{l=1}^{s_t} \sum_{k=1}^{s_r} \rho_{k,l} e^{j2\pi \xi_{k,l} t} e^{-j2\pi f \tau_{k,l}}
\end{aligned}
$$

where $\rho_{k,l}$ ($\rho_{k,l} = \frac{1}{\sqrt{ss_1}} \theta_{k,l} \sqrt{P_k^r} \sqrt{P_l^t} e^{j(\phi_k + \psi_l)}$) are independent Gaussian variable with zero mean and variance $\mathbb{E}(\mid \rho_{k,l} \mid^2) = \frac{1}{ss_1} P_k^r P_l^t$, $\xi_{k,l} = \frac{f}{c}(\mathbf{u}_k^r \mathbf{v_r} - \mathbf{u}_l^t \mathbf{v_t})$ are the Doppler effect and $\tau_{k,l}$ are the delays. This previous result is a generalization of the SISO (Single Input Single Output) wireless model in the case of multifold scattering with the power profile taken into account.

B. Frequency Selectivity

In this section, we are interested in the ergodic mutual information of the frequency selective channel. The mutual information per transmitting antenna with input covariance $\mathbf{Q} = \mathbf{I}$ is given by:

$$
I^M(n_t, f)) = \frac{1}{n_t} \log_2 \det(\mathbf{I_{n_t}} + \frac{\rho}{n_t} \mathbf{H}^H(f,t) \mathbf{H}(f,t))
$$

Note that the mutual information depends on t due to the Doppler effect and has therefore no real meaning. Indeed, the perfect channel knowledge assumption at the receiver is not valid (since the channel varies) and a non-coherent mutual information should be calculated. This is not an easy task and an open problem even for simple channel models. A first step in this direction is the work of Marzetta and Hochwald [70] and the work of Zheng and Tsé [28]. An even more difficult problem concerns the practical schemes for achieving the non-coherent mutual information. Recently, in an interesting paper [71], Hassibi and Hochwald have shown that simple on the shelf training schemes can be optimal at high SNR (for the i.i.d Gaussian model) which circumvents therefore the need of using blind or semi-blind techniques in that regime.

Therefore, in the following, only the mutual information with no Doppler effect will be considered. In order to derive the mutual information, let us show that the spatial statistics of $\mathbf{H}(f)$ are independent of f. Since $\mathbf{H}(f)$ is Gaussian, all the statistics are described by the mean and the covariance matrix.

- **Mean:** Since the entries of matrix $\mathbf{\Theta}$ have zero mean,

$$\mathbb{E}_{\mathbf{\Theta}}(h_{ij}) = \frac{1}{\sqrt{ss_1}} \sum_{k=1}^{s_t} \sum_{p=1}^{s_r} \mathbb{E}(\theta_{pk}) \sqrt{P_k{}^t} \sqrt{P_p{}^r} e^{j2\pi f \tau_{pk}} e^{j\psi_{kj}} e^{j\phi_{ip}} = 0$$

for every i,j and is therefore independent of f.

- **Covariance:** Let us derive $Cov(i,j,m,n,f) = \mathbb{E}_{\mathbf{\Theta}}(h_{ij}(f)h^*_{mn}(f))$:

$$\begin{aligned} Cov(i,j,m,n,f) &= \frac{1}{s_r s_t} \sum_{k=1}^{s_t} \sum_{p=1}^{s_r} \sum_{q=1}^{s_t} \sum_{l=1}^{s_r} \mathbb{E}(\theta_{pk}\theta^*_{ql}) e^{j2\pi f(\tau_{pk}-\tau_{ql})} \\ & \sqrt{P^t_k}\sqrt{P^t_q}\sqrt{P^r_p}\sqrt{P^r_l} e^{j2\pi(\psi_{kj}-\psi_{qn})} e^{j2\pi(\phi_{ip}-\phi_{ml})} \end{aligned}$$

Since $\mathbb{E}(\theta_{pk}\theta^*_{ql}) = \delta_{pq}\delta_{kl}$, then :

$$Cov(i,j,m,n,f) = \frac{1}{s_r s_t} \sum_{k=1}^{s_t} \sum_{p=1}^{s_r} P^t_k P^r_p e^{j2\pi(\psi_{kj}-\psi_{kn})} e^{j2\pi(\phi_{ip}-\phi_{mlp})}$$

which is independent of f.

Since the statistics of $\mathbf{H}(f)$ are independent of f, the ergodic mutual information over the bandwidth W is given by:

$$I^M = \frac{W}{n_t} \mathbb{E}\left[\log_2 \det(\mathbf{I_{n_t}} + \frac{\rho}{n_t}\mathbf{H}^H(0)\mathbf{H}(0))\right]$$

One can observe that frequency selectivity does not affect the mutual information per transmitting antenna. Similar results have been reported in [25], [65]. In the wideband case with no Doppler effect, the ergodic mutual information is the same as in the narrowband case and all the results of chapter IV remain valid (ULA and Fourier directions, random directions approximation..).

VI. Other Models in View of the Maximum Entropy Framework

A. Müller's Model

In [65], Müller develops a channel model based on the product of two random matrices:

$$\mathbf{H} = \mathbf{\Phi}\mathbf{A}\mathbf{\Theta}$$

where $\mathbf{\Phi}$ and $\mathbf{\Theta}$ are two random matrices with zero mean unit variance i.i.d entries and $\mathbf{A}$ is a diagonal matrix (representing the attenuations). This model is intended to represent the fact that each signal bounces off a scattering object exactly once (also known as the single bounce model[12]). $\mathbf{\Phi}$ represents the steering directions from the scatterers to the receiving antennas while $\mathbf{\Theta}$ represents the steering directions from the transmitting antennas to the scatterers. Measurements in [65] confirmed the model quite accurately. Should we conclude that signals in day to day life bounce only once on the scattering objects?

With the maximum entropy approach developed in this contribution, new insights can be given on this model and explanations can be provided on why Müller's model works so well. In the maximum entropy framework, Müller's model can be seen as either:

- a DoA based model with random directions i.e matrix $\mathbf{\Phi}$ with different powers (represented by matrix $\mathbf{A}$) for each angle of arrival. In fact, the signal can bounce freely several times from the transmitting antennas to the final scatterers (matrix $\mathbf{\Theta}$). Contrary to past belief, this model takes into account multi-fold scattering and answers the following question from a maximum entropy standpoint: what is the consistent model when the state of knowledge is limited to:
  - Random directions scattering at the receiving side.

[12] Note that the terminology is misleading. Indeed, the modeler never assumed a single bounce but only a one to one mapping between DoA's and DoD's. It makes allowance for several bounces as long as each DoA is linked to one and only one DoD whatever happens in between.

– Each steering vector has a certain power.

– The channel carries a certain energy.

• a corresponding DoD based model with random directions i.e matrix $\mathbf{\Theta}$ with different powers (represented by matrix $\mathbf{A}$) for each angle of departure. The model permits also in this case the signal to bounce several times from the scatterers to the receiving antennas.

• DoA-DoD based model with random directions where the following question is answered: What is the consistent model when the state of knowledge is limited to:

– Random directions scattering at the receiving side.

– Random directions scattering at the transmitting side.

– Each angle of arrival is linked to one angle of departure.

As one can see, Müller's model is broad enough to include several maximum entropy directional models and this fact explains why the model complies so accurately with the measurements performed in [72]

B. Sayeed's Model

In [73], Sayeed proposes a virtual representation of the channel. The model is the following:

$$\mathbf{H} = \mathbf{A}_{n_r}\mathbf{S}\mathbf{A}_{n_t}{}^H$$

Matrices $\mathbf{A}_{n_r}$ and $\mathbf{A}_{n_t}$ are discrete Fourier matrices and $\mathbf{S}$ is a $n_r \times n_t$ matrix which represents the contribution of each of the fixed DoA's and DoD's. The representation is virtual in the sense that it does not represent the real directions but only the contribution of the channel to those fixed directions. The model is somewhat a projection of the real steering directions onto a Fourier basis. Sayeed's model is quite appealing in terms of simplicity and analysis. In this case, also, we can revisit Sayeed's model in light of our framework. We can show that every time, Sayeed's model answers a specific question based on a given assumption.

• Suppose matrix $\mathbf{S}$ has i.i.d zero mean Gaussian entries then Sayeed's model answers the following question: what is the consistent model for a ULA when the modeler assumes that the channel carries energy, the DoA and DoD are on Fourier directions but one does not know what happens in between.

• Suppose now that matrix $\mathbf{S}$ has a certain correlation structure then Sayeed's model answers the following question: what is the consistent model for a ULA when the modeler assumes that the channel carries energy, the DoA and DoD are on Fourier directions but assumes that the paths in between have a certain correlation.

As one can see, Sayeed's model has a simple interpretation in the maximum entropy framework: it considers a ULA geometry with Fourier directions each time. Although it may seem strange that Sayeed limits himself to Fourier directions, we do have an explanation for this fact. In his paper [25], Sayeed was mostly interested in the capacity scaling of MIMO channels and not the joint distribution of the elements. From that perspective, only the statistics of the uncorrelated scatterers is of interest since they are the ones which scale the capacity. The correlated scatterers have very small effect on the capacity. However, the entropy framework is not limited to the ULA case (for which the Fourier vector approach is valid) and can be used for any kind of antenna and field approximation. One of the great features of the maximum entropy approach is the quite simplicity for translating any additional physical information into probability assignment in the model. A one to one mapping between information and model representation is possible.

C. The "Kronecker" Model

In [27], Chuah et al. study the following Kronecker [13] model:

$$\mathbf{H} = \mathbf{R}_{n_r}{}^{\frac{1}{2}} \mathbf{\Theta} \mathbf{R}_{n_t}{}^{\frac{1}{2}}$$

Here, $\mathbf{\Theta}$ is an $n_r \times n_t$ i.i.d zero mean Gaussian matrix, $\mathbf{R}_{n_r}{}^{\frac{1}{2}}$ is an $n_r \times n_r$ receiving correlation matrix while $\mathbf{R}_{n_t}{}^{\frac{1}{2}}$ is a $n_t \times n_t$ transmitting correlation matrix. The correlation is supposed to decrease sufficiently fast so that $\mathbf{R}_{n_r}$ and $\mathbf{R}_{n_t}$ have a Toeplitz band structure. Using a software tool (Wireless System Engineering [76]), they demonstrate the validity of the model. Quite remarkably, although designed to take into account receiving and transmitting correlation, the model developed in the paper falls within the double direc-

[13] The model is called a Kronecker model because $\mathbb{E}(\text{vec}(\mathbf{H})^H \text{vec}(\mathbf{H})) = \mathbf{R}_{n_r} \bigotimes \mathbf{R}_{n_t}$ (the operator $\text{vec}(\mathbf{H})$ stacks all the columns of matrix $\mathbf{H}$ into a single column) is a Kronecker product. The justification of this approach relies on the fact that only immediate surroundings of the antenna array impose the correlation between array elements and have no impact on correlations observed between the elements of the array at the other end of the link. Some discussions can be found in [74], [75]

tional framework. Indeed, since $\mathbf{R}_{n_r}$ and $\mathbf{R}_{n_t}$ are band Toeplitz then these matrices are asymptotically diagonalized in a Fourier basis:

$$\mathbf{R}_{n_r} \sim \mathbf{F}_{n_r}\Lambda_{n_r}\mathbf{F}_{n_r}^{H}$$

and

$$\mathbf{R}_{n_t} \sim \mathbf{F}_{\mathbf{n_t}}\Lambda_{n_t}\mathbf{F}_{\mathbf{n_t}}^{\mathbf{H}}.$$

$\mathbf{F}_{n_r}$ and $\mathbf{F}_{n_t}$ are Fourier matrices while Λ_{n_r} and Λ_{n_t} represent the eigenvalue matrices of $\mathbf{R}_{n_r}$ and $\mathbf{R}_{n_t}$.

Therefore, matrix $\mathbf{H}$ can be rewritten as:

$$\begin{aligned} \mathbf{H} &= \mathbf{R}_{n_r}{}^{\frac{1}{2}}\mathbf{\Theta}\mathbf{R}_{n_t}{}^{\frac{1}{2}} \\ &= \mathbf{F}_{n_r}\left(\Lambda_{n_r}^{\frac{1}{2}}\mathbf{F}_{n_r}{}^{H}\mathbf{\Theta}\mathbf{F}_{n_t}\Lambda_{n_t}{}^{\frac{1}{2}}\right)\mathbf{F}_{n_t}{}^{H} \\ &= \mathbf{F}_{n_r}\left(\mathbf{\Theta_1} \odot \mathbf{D}_{n_r \times n_t}\right)\mathbf{F}_{n_t}{}^{H} \end{aligned}$$

$\mathbf{\Theta_1} = \mathbf{F}_{n_r}{}^{H}\mathbf{\Theta}\mathbf{F}_{n_t}$ is a $n_r \times n_t$ zero mean i.i.d Gaussian matrix and $\mathbf{D}_{n_r \times n_t}$ is a pattern mask matrix defined by:

$$\mathbf{D}_{s_r \times s_t} = \begin{pmatrix} \lambda_{1,n_t}^{\frac{1}{2}}\lambda_{1,n_r}^{\frac{1}{2}} & \dots & \lambda_{n_t,n_t}^{\frac{1}{2}}\lambda_{1,n_r}^{\frac{1}{2}} \\ \vdots & \ddots & \vdots \\ \lambda_{1,n_t}^{\frac{1}{2}}\lambda_{n_r,n_r}^{\frac{1}{2}} & \dots & \lambda_{n_t,n_t}^{\frac{1}{2}}\lambda_{n_r,n_r}^{\frac{1}{2}} \end{pmatrix}$$

Note that this connection with the double directional model has already been reported in [25]. Here again, the previous model can be reinterpreted in light of the maximum entropy approach. The model answers the following question: what is the consistent model one can make when the DoA are uncorrelated and have respective power λ_{i,n_r}, the DoD are uncorrelated and have respective power λ_{i,n_t}, each path has zero mean and a certain variance. The model therefore confirms the double directional assumption as well as Sayeed's approach and is a particular case of the maximum entropy approach. The comments and limitations made on Sayeed's model are also valid here.

D. "Keyhole" Model

In [77], Gesbert et al. show that low correlation[14] is not a guarantee of high capacity: cases where the channel is rank deficient can appear while having uncorrelated entries (for example when a screen with a small keyhole is placed in between the transmitting and receiving antennas). In [79], they propose the following model for a rank one channel:

$$\mathbf{H} = \mathbf{R}_{n_r}{}^{\frac{1}{2}} \mathbf{g}_r \mathbf{g}_t{}^H \mathbf{R}_{n_t}{}^{\frac{1}{2}} \tag{25}$$

Here, $\mathbf{R}_{n_r}{}^{\frac{1}{2}}$ is an $n_r \times n_r$ receiving correlation matrix while $\mathbf{R}_{n_t}{}^{\frac{1}{2}}$ is a $n_t \times n_t$ transmitting correlation matrix. $\mathbf{g}_r$ and $\mathbf{g}_t$ are two independent transmit and receiving Rayleigh fading vectors. Here again, this model has connections with the previous maximum entropy model:

$$\mathbf{H} = \frac{1}{\sqrt{s_r s_t}} \mathbf{\Phi}_{n_r \times s_r} \mathbf{\Theta}_{s_r \times s_t} \mathbf{\Psi}_{s_t \times n_t} \tag{26}$$

The Keyhole model can be either:

- A double direction model with $s_r = 1$ and $\mathbf{\Phi}_{n_r \times 1} = \mathbf{R}_{n_r}{}^{\frac{1}{2}} \mathbf{g}_r$. In this case, $\mathbf{g}_t{}^H \mathbf{R}_{n_t}{}^{\frac{1}{2}} = \mathbf{\Theta}_{1 \times s_t} \mathbf{\Psi}_{s_t \times n_t}$ where $\mathbf{\Theta}_{1 \times s_t}$ is zero mean i.i.d Gaussian.
- A double direction model with $s_t = 1$ and $\mathbf{\Psi}_{1 \times n_t} = \mathbf{g}_t{}^H \mathbf{R}_{n_t}{}^{\frac{1}{2}}$. In this case, $\mathbf{R}_{n_r}{}^{\frac{1}{2}} \mathbf{g}_r = \Phi_{n_r \times s_r} \mathbf{\Theta}_{s_r \times 1}$ where $\mathbf{\Theta}_{s_r \times 1}$ is zero mean i.i.d Gaussian.

As one can observe, the maximum entropy model can take into account rank deficient channels.

VII. Limitations of the Maximum Entropy Approach

In the previous paragraphs, the mutual information was derived based on the assumption that the channel model is adequate with reality. For example, knowing that the frequency paths are Gaussian i.i.d and the noise is additive white Gaussian, the transmitter will design codes to ensure a reliable transmission on such channels achieving that

[14]"keyhole" channels are MIMO channels with uncorrelated spatial fading at the transmitter and the receiver but have a reduced channel rank (also known as uncorrelated low rank models). They were shown to arise in roof-edge diffraction scenarios [78]

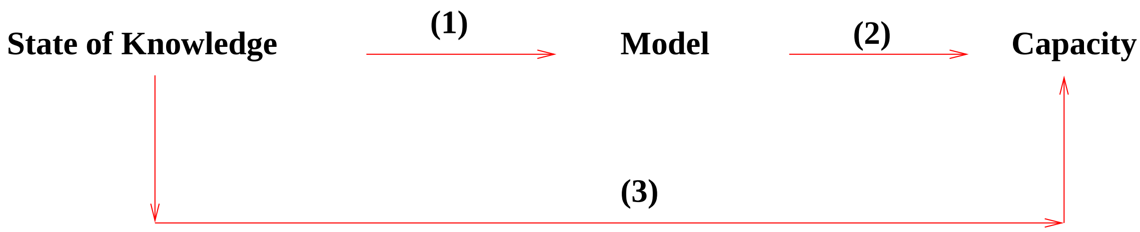

Fig. 13. Channel modelling approach and derivation of capacity.

rate. But whenever we are misrepresenting the channel with our state of knowledge, the formula

$$C(\mathbf{I}) = \log_2 \det\left(\mathbf{I}_{n_t} + \frac{\rho}{n_t}\mathbf{H}^H\mathbf{H}\right) \tag{27}$$

will misestimate the rate. Indeed, a surprising fact in our maximum entropy approach is that although it gives us a consistent model with our state of knowledge, it will also lead to mis-estimating the rate with formula (27). The problem is formulated in Figure 13.

- Transition (1): the modeler creates a model maximizing entropy.
- Transition (2): The modeler mis-estimates the real achievable rate because even though the model he creates is the best he can based on his state of knowledge, he derives the mutual information of the channel based on the assumption that the model is reality.
- Transition (3): A new measure of information rate should be derived based only on our state of knowledge, taking into account the fact that the model does not represent reality, but only our knowledge (which is scarce) of reality

As a matter of fact, for deriving the mutual information, a channel model is not required but only the state of knowledge. One can derive more useful information rate criteria which circumvent the need of a channel model such as the "worst case mean channel capacity":

$$I^M(\mathbf{H}) = \min_{P(\mathbf{H}):\mathbf{H}\in\Delta}\{\max_{\mathbf{Q}} \int \left(\log_2 \det\left(\mathbf{I}_{n_t} + \frac{\rho}{n_t}\mathbf{H}^H\mathbf{Q}\mathbf{H}\right)\right) P(\mathbf{H})d\mathbf{H}\}$$

Δ is the infinite set of matrices $\mathbf{H}$ with the same initial physical constraints (mean and variance for example). Of course, other measures of capacity performance can be derived. So, is there a contradiction in our maximum entropy modelling approach? No, as long as we understand the meaning of transition (2) in Figure 13. With the maximum entropy approach, we derive a channel model having as much degrees of freedom as possible (but still with the constraints of our state of knowledge) in order to cope with all the cases when they happen. We do this because we need a **unique** model consistent with our state of knowledge. Any other approach will add unjustified constraints. Suppose, for sake of simplicity, that each frequency path of the channel has a zero mean and a given variance (the mean and variance are here our state of knowledge). Transition (1)+(2) will give us a measure of the rate one can transmit on a "maximum entropy channel state knowledge". The problem stems from the fact that although models are consistent, functionals of the model are not. Indeed, consider the DoA based model: $\mathbf{H} = \mathbf{\Phi}\mathbf{\Theta}$, ($\mathbf{\Theta}$ is i.i.d Gaussian) then using Jensen's Inequality:

$$\mathbb{E}_{\mathbf{\Phi}}\left(\log\det(\mathbf{I}_{n_t} + \frac{\rho}{n_t}(\mathbf{\Theta}^H\mathbf{\Phi}^H\mathbf{\Phi}\mathbf{\Theta})\right) \leqslant \log\det(\mathbf{I}_{n_t} + \frac{\rho}{n_t}\mathbb{E}_{\mathbf{\Phi}}(\mathbf{\Theta}^H\mathbf{\Phi}^H\mathbf{\Phi}\mathbf{\Theta}))$$

For example, when the directions of arrival are unknown, the mutual information averaged across the unknown directions of arrival (here $\mathbb{E}_{\mathbf{\Phi}}\left(\log\det(\mathbf{I}_{n_t} + \frac{\rho}{n_t}(\mathbf{\Theta}^H\mathbf{\Phi}^H\mathbf{\Phi}\mathbf{\Theta})\right)$) does not yield the mutual information of the Gaussian i.i.d. model:

$$\log\det(\mathbf{I}_{n_t} + \frac{\rho}{n_t}\mathbb{E}_{\mathbf{\Phi}}(\mathbf{\Theta}^H\mathbf{\Phi}^H\mathbf{\Phi}\mathbf{\Theta})) = \log\det(\mathbf{I}_{n_t} + \frac{\rho}{n_t}(\mathbf{\Theta}^H\mathbf{\Theta}))$$

the model is consistent but not the functional. A remarkable feature of the previous result is that whenever we have more information (and therefore more constraints on the channel model), mutual information will be reduced as it constraints the degrees of freedom. This explains why, under the same initial constraints (as an example the mean and the variance of each path), correlated fading reduces the mutual information with respect to the completely i.i.d case. As an example, the fact that we take into account the DoA, mean, variance will reduce the mutual information compared with the case where only the same mean and same variance are taken into account. In fact, if one is interested only in one or some functions of the model, then he should construct a model which is consistent with those functionals and not in itself. A consistent model is for the case where we do not know which functions we (or others who we construct the model for) are interested in.

VIII. Conclusion

Where do we stand on channel modelling ?[15] This question is not simple to answer as many models have been proposed and each of them validated by measurements. Channel models are not getting better and better but they only answer different questions based on different states of knowledge. A generic method for creating models based on the principle of maximum entropy has been provided and proved to be theoretically sound. At every step, we create a model incorporating only our prior information and not more. The model achieved is broad as it complies as best it can with any case having more constraints (but at least includes the same prior constraints). The channel modelling method is summarized hereafter:

- $H(p) = \int -p\log p + \sum_i \lambda_i \{\text{prior information}\}_i$
- Argument of consistency

The consistency argument is extremely important as it shows that two channel modelling methods based on the same state of knowledge should lead to the same channel model. This fact has not always been fulfilled in the past. However, one must bear in mind that the less things are assumed as a priori information the greater are the chances that the model complies with any mismatched representation. Finally, note that recent campaign measures at 2.1 GHz and 5.2 GHz have shown that Maxent Fourier models are mutual information complying [81].

Appendix

A. Preliminaries

Lemma 1: Consider the $t \times t$ matrix (see Bai & Silverstein [37]):

$$B_t = \frac{1}{t}\mathbf{H}_{s\times t}{}^H \mathbf{A}_{s\times s}\mathbf{H}_{s\times t}$$

- $\mathbf{H}_{s\times t} = (h_{ij})$is an $s \times t$ matrix with i.i.d complex standardized entries having finite fourth moments, $\mathbb{E}({h_{ij}}^2) = 0$ and $\mathbb{E}(\mid h_{ij} \mid^4) = 2$ with $\lim_{t\to\infty} \frac{s}{t} = c$.
- $\mathbf{A}_{s\times s}$ is an $s \times s$ non-random Hermitian non-negative definite matrix, with empirical eigenvalue distribution that converges in distribution almost surely to a fixed G, and the

[15] This question has to be taken in light of a talk "Where do we stand on maximum entropy?" made by E.T. Jaynes in 1978 at MIT [80].

sequence of spectral norms $\|\mathbf{A}_{s\times s}\|$ is bounded.

• f is continuously differentiable with a bounded first derivative and analytic on an open interval containing $[(\max(0, 1-\sqrt{c}))^2 \liminf \lambda_{Amin}, (1+\sqrt{c})^2 \limsup \lambda_{Amax}]$ with λ_{Amin} and λ_{Amax} respectively the smallest and the largest eigenvalues of $A_{s\times s}$.

Then as $t \to \infty$ and $\frac{s}{t} \to c$,

$$t(f(B_t) - \mu_t) \to N(0, \sigma^2) \qquad \text{in distribution.}$$

In other words, the empirical spectral distribution of B_t is shown to have a Gaussian limit.

• $\mu = \int f(\lambda) dF(\lambda)$, F is the limiting distribution of F^{B_t}, solution of the implicit equation

$$z = -\frac{1}{m(z)} + c \int \frac{\tau}{1 + m(z)\tau} dG(\tau)$$

through its Stieljes Transform

$$m(z) = \int \frac{1}{\lambda - z} dF(\lambda)$$

• $N(0, \sigma^2)$ is a real valued, zero mean Gaussian random variable with asymptotic variance:

$$\sigma^2 = -\frac{1}{4\pi^2} \int_{C_x} \int_{C_y} \frac{f(x)f(y)}{(m(x) - m(y))^2} m'(x) m'(y) dx dy$$

and C_x and C_y are any closed positive contours that enclose the support of F.

Theorem 3: (see Girko [50]) Let the $N \times K$ random matrix $\mathbf{H}$ be composed of independent entries $(\mathbf{H})_{ij}$ with zero mean and variances $\frac{w_{ij}}{N}$ such as all w_{ij} are uniformly bounded from above. Assume that the empirical joint distribution of variances w: $[0, 1] \times [0, \beta] \to \mathbb{R}$ defined by $w(x, y) = w_{ij}$ for i, j satisfying:

$$\frac{i}{N} \leq x \leq \frac{i+1}{N}$$

and

$$\frac{j}{N} \leq y \leq \frac{j+1}{N}$$

converges to a bounded joint limit distribution $w(x, y)$ as $K = \beta N \to \infty$. Then, almost surely, the empirical eigenvalue distribution of $\mathbf{H}\mathbf{H}^H$ converges weakly to a limiting distribution whose Stieljes transform is given by:

$$m_{\mathbf{HH}^H}(s) = \int_0^1 u(x,s)dx$$

and $u(x,s)$ satisfies the fixed point equation:

$$u(x,s) = \left[-s + \int_0^\beta \frac{w(x,y)dy}{1+\int_0^1 u(x',s)w(x',y)dx'}\right]^{-1} \tag{28}$$

The solution to equation (28) exists and is unique in the class of functions $u(x,s) \geq 0$, analytic for $\mathrm{Im}(s) > 0$ and continuous on $x \in [0,1]$.

B. Proof of Proposition 2

Since $m_{\mathbf{HH}^H}(z) = (\frac{1}{\gamma}-1)\frac{1}{z} + \frac{1}{\gamma} m_{\mathbf{\Theta}_1\mathbf{\Theta}_1^H}(z)$ and $\mu = \gamma\xi\int \log_2(1+\rho\gamma\xi\lambda)dF(\lambda)$, it can easily be shown that

$$\frac{d\mu}{d\rho} = \frac{1}{\gamma}\left(\frac{\xi\gamma}{\rho} - \frac{1}{\rho^2} m_{\mathbf{\Theta}_1\mathbf{\Theta}_1^H}(-\frac{1}{\rho\gamma\xi})\right)$$

Using theorem 3, we have:

$$\begin{aligned} m_{\mathbf{\Theta}_1\mathbf{\Theta}_1^H}(-\frac{1}{\rho\gamma\xi}) &= \mathbb{E}_{\lambda^\phi}\left[\frac{1}{\frac{1}{\rho\gamma\xi} + \frac{\alpha^{\text{joint}}}{\gamma\xi}}\right] \\ &= \rho\gamma\xi\mathbb{E}_{\lambda^\phi}\left[\frac{1}{1+\rho\alpha^{\text{joint}}}\right] \end{aligned}$$

with (since $\gamma\xi = \gamma_1\xi_1$)

$$\alpha^{\text{joint}} = \frac{\gamma}{\gamma_1}\frac{\gamma\xi}{\gamma}\mathbb{E}_{\lambda^\psi}\left[\frac{\lambda^\phi\lambda^\psi}{1+\rho{\alpha_1}^{\text{joint}}}\right] = \xi_1\mathbb{E}_{\lambda^\psi}\left[\frac{\lambda^\phi\lambda^\psi}{1+\rho{\alpha_1}^{\text{joint}}}\right]$$

and

$${\alpha_1}^{\text{joint}} = \xi\mathbb{E}_{\lambda^\phi}\left[\frac{\lambda^\phi\lambda^\psi}{1+\rho\alpha^{\text{joint}}}\right].$$

Therefore,

$$\begin{aligned} \frac{d\mu}{d\rho} &= \frac{1}{\gamma}\left(\frac{\xi\gamma}{\rho} - \frac{1}{\rho^2}\rho\gamma\xi\mathbb{E}_{\lambda^\phi}\left[\frac{1}{1+\rho\alpha^{\text{joint}}}\right]\right) \\ &= \xi\mathbb{E}_{\lambda^\phi}\left[\frac{\alpha^{\text{joint}}}{1+\rho\alpha^{\text{joint}}}\right] \end{aligned}$$

which proves the result.

C. Proof of Proposition 3

In this case,

$$m_{\mathbf{\Theta}_1\mathbf{\Theta}_1^H}(-\frac{1}{\rho\gamma\xi}) = \rho\gamma\xi\mathbb{E}_{\lambda^\phi}\left[\frac{1}{1+\rho\lambda^\phi\alpha_{\text{doa}}}\right]$$

with $\alpha_{\text{doa}} = \xi_1\mathbb{E}_{\lambda^\psi}\left[\frac{\lambda^\psi}{1+\rho\lambda^\psi\alpha_{\text{dod}}}\right]$ and $\alpha_{\text{dod}} = \xi\mathbb{E}_{\lambda^\phi}\left[\frac{\lambda^\phi}{1+\rho\lambda^\phi\alpha_{\text{doa}}}\right]$
and

$$\begin{aligned}\frac{d\mu}{d\rho} &= \xi\mathbb{E}_{\lambda^\phi}\left[\frac{\lambda^\phi\alpha_{\text{doa}}}{1+\rho\lambda^\phi\alpha_{\text{doa}}}\right] \\ &= \alpha_{\text{doa}}\alpha_{\text{dod}}\end{aligned}$$

which integrates as:

$$\mu = \xi_1\mathbb{E}_{\lambda^\psi}(\ln(1+\rho\lambda^\psi\alpha_{\text{dod}})) + \xi\mathbb{E}_{\lambda^\phi}(\ln(1+\rho\lambda^\phi\alpha_{\text{doa}})) - \rho\alpha_{\text{doa}}\alpha_{\text{dod}}$$

D. Proof of Proposition 4

In the equal power case, let us apply the result of Proposition 3 when $S_{\lambda^\phi}(\lambda) = \delta(\lambda-\gamma)$ and $S_{\lambda^\psi}(\lambda) = \delta(\lambda-\frac{1}{\xi})$ (note that in all the following derivation, we will often use the fact that $\gamma\xi = \gamma_1\xi_1$). In this case,

$$\mu_{\text{double}} = \xi_1\ln(1+\frac{\rho\alpha_{\text{dod}}}{\xi_1}) + \xi\ln(1+\rho\gamma\alpha_{\text{doa}}) - \rho\alpha_{\text{doa}}\alpha_{\text{dod}} \tag{29}$$

with

$$\begin{aligned}\alpha_{\text{doa}} &= \frac{\xi_1\frac{1}{\xi_1}}{1+\frac{\rho\alpha_{\text{dod}}}{\xi_1}} \\ &= \frac{1}{1+\frac{\rho\alpha_{\text{dod}}}{\xi_1}}\end{aligned} \tag{30}$$

and

$$\alpha_{\text{dod}} = \frac{\gamma\xi}{1+\rho\gamma\alpha_{\text{doa}}} \tag{31}$$

Notice that in this case, we have:

$$\rho\alpha_{\text{doa}}\alpha_{\text{dod}} = \xi_1(1-\alpha_{\text{doa}}) \tag{32}$$

Using equation (30) and equation (31), α_{doa} is given by:

$$\alpha_{\text{doa}} = \frac{1}{1+\frac{\rho\xi\gamma}{\xi_1(1+\rho\gamma\alpha_{\text{doa}})}}$$
$$\iff \alpha_{\text{doa}}(1+\frac{\rho\gamma_1}{1+\rho\gamma\alpha_{\text{doa}}}) = 1$$

which yields:

$$\alpha_{\text{doa}}^2 + \alpha_{\text{doa}}(\frac{1}{\gamma\rho}+\frac{\gamma_1}{\gamma}-1) - \frac{1}{\rho\gamma} = 0$$

which has the following solution:

$$\alpha_{\text{doa}} = \frac{1}{2}\left[(1-\frac{\gamma_1}{\gamma}-\frac{1}{\gamma\rho}) + \sqrt{1-\frac{\gamma_1}{\gamma}-\frac{1}{\gamma\rho})^2+\frac{4}{\gamma\rho}}\right]$$

Since

$$(1-\frac{\gamma_1}{\gamma}-\frac{1}{\gamma\rho})^2+\frac{4}{\gamma\rho} = (1+\frac{\gamma_1}{\gamma}+\frac{1}{\gamma\rho})^2-\frac{4\gamma_1}{\gamma}$$

then $\alpha_{\text{doa}} = 1-\alpha_{\text{double}}$ where

$$\begin{aligned}\alpha_{\text{double}} &= 1-\alpha_{\text{doa}}\\ &= \frac{1}{2}\left[(1+\frac{\gamma_1}{\gamma}+\frac{1}{\gamma\rho}) - \sqrt{(1+\frac{\gamma_1}{\gamma}+\frac{1}{\gamma\rho})^2-\frac{4\gamma_1}{\gamma}}\right]\end{aligned}$$

Now let us derive α_{dod}:

We have

$$\alpha_{\text{dod}} = \frac{\gamma\xi}{1+\frac{\gamma\rho}{1+\rho\frac{\alpha_{\text{dod}}}{\xi_1}}}$$

and

$$\iff \alpha_{\text{dod}}(1+\frac{\gamma\rho}{1+\rho\frac{\alpha_{\text{dod}}}{\xi_1}}) = \gamma\xi$$

which yields:

$$\alpha_{\text{dod}}^2 + \alpha_{\text{dod}}(\frac{\xi_1}{\rho} + \gamma\xi_1 - \gamma\xi) - \frac{\gamma\xi\xi_1}{\rho} = 0$$

The solution to this equation is:

$$\alpha_{\text{dod}} = \frac{1}{2}\left[(\gamma_1\xi_1 - \gamma\xi_1 - \frac{\xi_1}{\rho}) + \sqrt{\gamma_1\xi_1 - \gamma\xi_1 - \frac{\xi_1}{\rho})^2 + \frac{4\gamma\xi\xi_1}{\rho}}\right]$$

It can be easily shown that:

$$\alpha_{\text{double}} = \frac{1}{\gamma\xi_1}(\gamma_1\xi_1 - \alpha_{\text{dod}})$$

We have therefore:

$$\alpha_{\text{doa}} = 1 - \alpha_{\text{double}} \qquad (33)$$

and

$$\alpha_{\text{dod}} = \gamma_1\xi_1 - \gamma\xi_1\alpha_{\text{double}} \qquad (34)$$

Using eq.(29), eq.(32), eq.(33) and eq.(34), one can show that:

$$\mu_{\text{double}} = \xi\ln(1 + \rho\gamma - \rho\gamma\alpha_{\text{double}}) + \xi_1\ln(1 + \rho\gamma_1 - \rho\gamma\alpha_{\text{double}}) - \xi_1\alpha_{\text{double}}$$

with

$$\alpha_{\text{double}} = \frac{1}{2}\left[1 + \frac{\gamma_1}{\gamma} + \frac{1}{\rho\gamma} - \sqrt{(1 + \frac{\gamma_1}{\gamma} + \frac{1}{\rho\gamma})^2 - 4\frac{\gamma_1}{\gamma}}\right]$$

E. Proof of Proposition 9

Let us first derive μ_{doa}. In the DoA based model, one can apply straightforwardly Proposition 6 if $\gamma = \frac{n_r}{s_r}$, $\xi = \frac{s_r}{n_t}$, $\gamma_1 = \frac{n_r}{n_t} = \gamma\xi$,$\xi_1 = \frac{s_t}{n_t} = 1$,$K_t = 1$,${P_i}^t = 1$. Therefore,

$$\mu_{\text{doa}} = \ln(1 + \rho\alpha_{\text{dod}}) + \xi\sum_{i=1}^{K_r} {l_i}^r\ln(1 + \rho {P_i}^r\gamma\alpha_{\text{doa}}) - \rho\alpha_{\text{doa}}\alpha_{\text{dod}}$$

with

$$\alpha_{\text{doa}} = \frac{1}{1 + \rho\alpha_{\text{dod}}}$$

and

$$\alpha_{\text{dod}} = \xi \sum_{i=1}^{K_r} \frac{{l_i}^r {P_i}^r \gamma}{1 + \rho\gamma {P_i}^r \alpha_{\text{doa}}}$$

Notice that

$$\alpha_{\text{doa}}(1 + \rho\alpha_{\text{dod}}) = 1$$

and therefore:

$$\alpha_{\text{dod}} = \frac{1}{\rho}(\frac{1}{\alpha_{\text{doa}}} - 1) \tag{35}$$

and

$$\rho\alpha_{\text{doa}}\alpha_{\text{dod}} = 1 - \alpha_{\text{doa}}$$

We can therefore rewrite μ_{doa} as:

$$\mu_{\text{doa}} = \ln(1 + \rho\frac{1}{\rho}(\frac{1}{\alpha_{\text{doa}}} - 1)) + \xi \sum_{i=1}^{K_r} {l_i}^r \ln(1 + \rho {P_i}^r \gamma\alpha_{\text{doa}}) - (1 - \alpha_{\text{doa}})$$

which yields:

$$\mu_{\text{doa}} = -\ln(\alpha_{\text{doa}}) + \xi \sum_{i=1}^{K_r} {l_i}^r \ln(1 + \rho {P_i}^r \gamma\alpha_{\text{doa}}) - (1 - \alpha_{\text{doa}})$$

We also have using eq.(35):

$$\frac{1}{\rho}(\frac{1}{\alpha_{\text{doa}}} - 1) = \xi \sum_{i=1}^{K_r} \frac{{l_i}^r {P_i}^r \gamma}{1 + \rho\gamma {P_i}^r \alpha_{\text{doa}}}$$

which can be simplified to:

$$\sum_{i=1}^{K_r} \frac{{l_i}^r}{1 + \rho\gamma {P_i}^r \alpha_{\text{doa}}} = \frac{\alpha_{\text{doa}}}{\xi} + 1 - \frac{1}{\xi}.$$

Let us now derive ${\boldsymbol{\sigma}_{\textbf{doa}}}^{\mathbf{2}}$: To this end, we will apply theorem 2. Since $S_{\text{doa}}(\lambda) = \sum_{i=1}^{K_r} {l_i}^r \delta(\lambda - \gamma {P_i}^r)$, we have:

$$z = \frac{-1}{m(z)} + \xi \sum_{i=1}^{K_r} \frac{{l_i}^r}{m(z) + \frac{1}{\gamma {P_i}^r}}$$

The asymptotic variance is therefore equal to:

$$\sigma_{\text{doa}}{}^2 = \frac{-1}{4\pi^2}\int_{C_{m_x}}\int_{C_{m_y}} \frac{\ln(1+\rho x(m_x))\ln(1+\rho y(m_y)))}{(m_x-m_y)^2} dm_x dm_y$$

For fixed m_y, let us calculate:

$$\begin{aligned}
& \frac{1}{j2\pi}\int \frac{\ln(1-\frac{\rho}{m}+\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{m+\frac{1}{\gamma P_i{}^r}})}{(m-m_y)^2} dm \\
= & \frac{1}{j2\pi}\int \frac{\frac{\rho}{m^2}-\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{(m+\frac{1}{\gamma P_i{}^r})^2}}{1-\frac{\rho}{m}+\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{m+\frac{1}{\gamma P_i{}^r}}} \frac{1}{m-m_y} dm \\
= & \frac{1}{j2\pi}\int \frac{\prod_{i=1}^{K_r}\rho(m+\frac{1}{\gamma P_i{}^r})^2-\rho\xi m^2\prod_{i=1}^{K_r}(m+\frac{1}{\gamma P_i{}^r})^2\sum_{i=1}^{K_r}\frac{l_i{}^r}{(m+\frac{1}{\gamma P_i{}^r})^2}}{m\prod_{i=1}^{K_r}(m+\frac{1}{\gamma P_i{}^r})\left(\prod_{i=1}^{K_r}(m+\frac{1}{\gamma P_i{}^r})(m-\rho+\rho\xi m\sum_{i=1}^{K_r}\frac{l_i{}^r}{m+\frac{1}{\gamma P_i{}^r}})\right)} \frac{1}{m-m_y} dm \\
= & \frac{1}{j2\pi}\int -\frac{1}{m(\frac{1}{\rho})} \frac{\prod_{i=1}^{K_r}\rho(m+\frac{1}{\gamma P_i{}^r})^2-\frac{\rho\xi m^2}{K_r}\prod_{i=1}^{K_r}(m+\frac{1}{\gamma P_i{}^r})^2\sum_{i=1}^{K_r}\frac{1}{(m+\frac{1}{\gamma P_i{}^r})^2}}{\prod_{i=1}^{K_r}(m+\frac{1}{\gamma P_i{}^r})\prod_{i=1}^{K_r}(m-m^i)(m-m_y)} \\
& \left(\frac{1}{m}-\frac{1}{m-m(-\frac{1}{\rho})}\right) dm \\
= & \frac{1}{m_y}-\frac{1}{m_y-m(-\frac{1}{\rho})}
\end{aligned}$$

The result stems from the fact that the contour C_{m_x} is chosen to include 0 and $m(-\frac{1}{\rho})$ but not $\frac{1}{\gamma P_i{}^r}$ and m^i for all i.

For notation sake, define $P(m)$ as::

$$\begin{aligned}
P(m) &= \prod_{i=1}^{K_r}(m+\frac{1}{\gamma P_i^r})(m-\rho+\frac{\rho m\xi}{K_r}\sum_{i=1}^{K_r}\frac{1}{m+\frac{1}{\gamma P_i{}^r}}) \\
&= (m-m(\frac{-1}{\rho}))\prod_{i=1}^{K_r}(m-m^i)
\end{aligned}$$

Notice that:

$$P'(m(\frac{-1}{\rho})) = \prod_{i=1}^{K_r}(m(-\frac{1}{\rho}) - m^i) \quad (36)$$

$$= \prod_{i=1}^{K_r}(m(-\frac{1}{\rho}) + \frac{1}{\gamma {P_i}^r})\left(1 + \frac{\rho\xi}{K_r}\sum_{i=1}^{K_r}\frac{1}{m(-\frac{1}{\rho}) + \frac{1}{\gamma {P_i}^r}} - \frac{\rho\xi}{K_r}\sum_{i=1}^{K_r}\frac{1}{(m(-\frac{1}{\rho}) + \frac{1}{\gamma {P_i}^r})^2}\right)$$

$$= \prod_{i=1}^{K_r}(m(-\frac{1}{\rho}) + \frac{1}{\gamma {P_i}^r})\left(\frac{\rho}{m(-\frac{1}{\rho})} - \rho\xi\sum_{i=1}^{K_r}\frac{{l_i}^r}{(m(-\frac{1}{\rho}) + \frac{1}{\gamma {P_i}^r})^2}\right) \quad (37)$$

The last equation comes from the fact that:

$$m(-\frac{1}{\rho})(1 + \rho\xi\sum_{i=1}^{K_r}\frac{{l_i}^r}{m(-\frac{1}{\rho}) + \frac{1}{\gamma {P_i}^r}}) = \rho$$

Therefore,

$$\begin{aligned}
{\sigma_{\text{doa}}}^2 &= \frac{1}{j2\pi}\int \ln(1 - \frac{\rho}{m} + \rho\xi\sum_{i=1}^{K_r}\frac{{l_i}^r}{m + \frac{1}{\gamma {P_i}^r}})\left(\frac{1}{m} - \frac{1}{m - m(-\frac{1}{\rho})}\right)dm \\
&= \frac{1}{j2\pi}\int \ln\left(\frac{(m - m(\frac{-1}{\rho}))\prod_{i=1}^{K_r}(m - m^i)}{m\prod_{i=1}^{K_r}(m + \frac{1}{\gamma {P_i}^r})}\right)\left(\frac{1}{m} - \frac{1}{m - m(-\frac{1}{\rho})}\right)dm \\
&= \frac{1}{j2\pi}\int \ln(\frac{m - m(-\frac{1}{\rho})}{m})\left(\frac{1}{m} - \frac{1}{m - m(-\frac{1}{\rho})}\right)dm \\
&+ \frac{1}{j2\pi}\int \ln(\frac{\prod_{i=1}^{K_r}(m - m^i)}{\prod_{i=1}^{K_r}(m + \frac{1}{\gamma {P_i}^r})})\left(\frac{1}{m} - \frac{1}{m - m(-\frac{1}{\rho})}\right)ddm
\end{aligned}$$

The first integral is zero since the integrand has a primitive :

$$-\frac{1}{2}\left[\ln(\frac{m - m(-\frac{1}{\rho})}{m})\right]^2$$

Therefore, the asymptotic variance is equal to:

$${\sigma_{\text{doa}}}^2 = \ln\left(\frac{\prod_{i=1}^{K_r} -m^i}{\prod_{i=1}^{K_r}\frac{1}{\gamma {P_i}^r}}\right) - \ln\left(\frac{\prod_{i=1}^{K_r}(m(-\frac{1}{\rho}) - m^i)}{\prod_{i=1}^{K_r}(m(-\frac{1}{\rho}) + \frac{1}{\gamma {P_i}^r})}\right)$$

Since $m(-\frac{1}{\rho})\prod_{i=1}^{K_r} -m^i = -\rho\prod_{i=1}^{K_r}\frac{1}{\gamma {P_i}^r}$ (product of the roots of polynomial $P(m)$ which is equal to $P(0)$) and

$$\frac{\prod_{i=1}^{K_r}(m(-\frac{1}{\rho}) - m^i)}{\prod_{i=1}^{K_r}(m(-\frac{1}{\rho}) + \frac{1}{\gamma m(-\frac{1}{\rho})})} = \frac{\rho}{m(-\frac{1}{\rho})} - \rho\xi\sum_{i=1}^{K_r}\frac{{l_i}^r}{(m(-\frac{1}{\rho}) + \frac{1}{\gamma {P_i}^r})^2}$$

The previous result comes from equation (36) and equation (37).

we have therefore:

$$\begin{aligned}
{\sigma_{\mathrm{doa}}}^2 &= -\ln\left(\left(\frac{\rho}{m(-\frac{1}{\rho})} - \rho\xi m(-\frac{1}{\rho}) \sum_{i=1}^{K_r} \frac{{l_i}^r}{(m(-\frac{1}{\rho}) + \frac{1}{\gamma {P_i}^r})^2}\right)\frac{m(-\frac{1}{\rho})}{\rho}\right) \\
&= -\ln\left(1 - \xi (m(-\frac{1}{\rho}))^2 \sum_{i=1}^{K_r} \frac{{l_i}^r}{(m(-\frac{1}{\rho}) + \frac{1}{\gamma {P_i}^r})^2}\right) \\
&= -\ln\left(1 - \rho^2 \xi {\alpha_{\mathrm{doa}}}^2 \sum_{i=1}^{K_r} \frac{{l_i}^r (\gamma {P_i}^r)^2}{(1 + \rho\gamma {P_i}^r \alpha_{\mathrm{doa}})^2}\right)
\end{aligned}$$

The last equation stems from the fact that $m(-\frac{1}{\rho}) = \rho\alpha_{\mathrm{doa}}$. [16]

F. Proof of Proposition 7

In this proof, we show that the optimal power profile which maximizes the mean mutual information in the case of the double directional model with ULA and Fourier directions is $\mathbf{P^r} = \mathbf{I}_{s_r}$ and $\mathbf{P^t} = \mathbf{I}_{s_t}$.

Let us maximize μ_{double} with respect to ${P_j}^t$ with the constraints $\sum_{i=1}^{K_t} {l_i}^t {P_i}^t = 1$ and $\sum_{i=1}^{K_r} {l_i}^r {P_i}^r = 1$

This corresponds to maximizing the following function:

$$\begin{aligned}
L &= \mu_{\mathrm{double}} + \lambda_1 \left(\sum_{i=1}^{K_t} {l_i}^t {P_i}^t - 1\right) + \lambda_2 \left(\sum_{i=1}^{K_r} {l_i}^r {P_i}^r - 1\right) \\
&= \xi_1 \sum_{i=1}^{K_t} {l_i}^t \ln(1 + \frac{\rho {P_i}^t \alpha_{\mathrm{dod}}}{\xi_1}) + \xi \sum_{i=1}^{K_r} {l_i}^r \ln(1 + \rho {P_i}^r \gamma \alpha_{\mathrm{doa}}) \\
&- \rho\alpha_{\mathrm{doa}}\alpha_{\mathrm{dod}} + \lambda_1 \left(\sum_{i=1}^{K_t} {l_i}^t {P_i}^t - 1\right) + \lambda_2 \left(\sum_{i=1}^{K_r} {l_i}^r {P_i}^r - 1\right)
\end{aligned}$$

Therefore,

[16] Note that in the i.i.d Gaussian channel case, $K_r = 1$, $n_r = s_r$ and therefore $\gamma = 1$ and $\xi = \frac{n_r}{n_t}$. Therefore, one can verify immediately that we obtain the same variance as in section III-A.

$$
\begin{aligned}
\frac{dL}{dP_j{}^t} &= \frac{d\alpha_{\text{dod}}}{dP_j{}^t}\rho\sum_{i=1}^{K_t}\frac{l_i{}^t P_i^t}{1+\frac{\rho P_i{}^t\alpha_{\text{dod}}}{\xi_1}} + \frac{d\alpha_{\text{doa}}}{dP_j{}^t}\rho\xi\sum_{i=1}^{K_r} l_i{}^r\frac{P_i{}^r\gamma}{1+\rho\gamma P_i{}^r\alpha_{\text{doa}}} \\
&- \rho\frac{d\alpha_{\text{dod}}}{dP_j{}^t}\alpha_{\text{doa}} - \rho\frac{d\alpha_{\text{doa}}}{dP_j{}^t}\alpha_{\text{dod}} + \frac{\xi_1 l_j{}^t\frac{\rho\alpha_{\text{dod}}}{\xi_1}}{1+\frac{\rho P_j{}^t\alpha_{\text{doa}}}{\xi_1}} + \lambda_1 l_j{}^t
\end{aligned}
$$

Since

$$
\alpha_{\text{doa}} = \sum_{i=1}^{K_t}\frac{l_i{}^t P_i^t}{1+\frac{\rho P_i{}^t\alpha_{\text{dod}}}{\xi_1}}
$$

and

$$
\alpha_{\text{dod}} = \xi\sum_{i=1}^{K_r} l_i{}^r\frac{P_i{}^r\gamma}{1+\rho\gamma P_i{}^r\alpha_{\text{doa}}}
$$

then:

$$
\begin{aligned}
\frac{dL}{dP_j{}^t} &= \rho\frac{d\alpha_{\text{dod}}}{dP_j{}^t}\alpha_{\text{doa}} + \rho\frac{d\alpha_{\text{doa}}}{dP_j{}^t}\alpha_{\text{dod}} \\
&- \rho\frac{d\alpha_{\text{dod}}}{dP_j{}^t}\alpha_{\text{doa}} - \rho\frac{d\alpha_{\text{doa}}}{dP_j{}^t}\alpha_{\text{dod}} + \frac{\xi_1 l_j{}^t\frac{\rho\alpha_{\text{dod}}}{\xi_1}}{1+\frac{\rho P_j{}^t\alpha_{\text{doa}}}{\xi_1}} + \lambda_1 l_j{}^t \\
&= \frac{l_j{}^t\rho\alpha_{\text{dod}}}{1+\frac{\rho P_j{}^t\alpha_{\text{doa}}}{\xi_1}} + \lambda_1 l_j{}^t
\end{aligned}
$$

Therefore,

$$
\frac{l_j{}^t\rho\alpha_{\text{dod}}}{1+\frac{\rho P_j{}^t\alpha_{\text{doa}}}{\xi_1}} + \lambda_1 l_j{}^t = 0
$$

and

$$
\frac{\rho\alpha_{\text{dod}}}{1+\frac{\rho P_j{}^t\alpha_{\text{doa}}}{\xi_1}} = -\lambda_1
$$

The last inequality holds for every j. Therefore, all $P_j{}^t$ are equal (to 1 due to the normalization constraint). The same proof holds for $P_j{}^r$ by taking the derivative with respect to $P_j{}^r$.

References

[1] G.J. Foschini and M.J. Gans, “On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas,” *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.

[2] E. Biglieri, J. Proakis, and S. Shamai(Shitz), “Fading channels: Information-Theoretic and Communications Aspects,” *IEEE Trans. on Information Theory*, pp. 2619–2692, Oct. 1998.

[3] S. N Diggavi, N. Al-Dhahir, A. Stamoulis, and A. R. Calderbank, “Great Expectations: The value of Spatial Diversity in Wireless Networks,” in *Proc. of the IEEE*, 219-270, Feb. 2004.

[4] G. Golden, C. Foschini, R. Valenzuela, and P. Wolniansky, “Detection Algorithm and Initial Laboratory Results using V-BLAST Space-Time Communication Architecture,” *Electronics Letters*, vol. 35, no. 1, pp. 14–16, Jan. 1999.

[5] P.W. Wolniansky, G.J. Foschini, G.D. Golden, and R.A. Valenzuela, “V-BLAST: An Architecture for Realizing Very High Data Rates Over the Rich-Scattering Wireless Channel,” in *International Symposium on Signals, Systems, and Electronics*, 1998, vol. 4, pp. 295–300.

[6] I.E. Telatar, “Capacity of Multi-Antenna Gaussian Channels,” Technical report, AT & T Bell Labs, 1995.

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

[8] 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.

[9] C. E. Shannon, “A Mathematical Theory of Communication,” *The Bell Labs Technical Journal*, pp. 379–457,623,656, July-October, vol. 27 1948.

[10] E. T. Jaynes, “Information Theory and Statistical Mechanics, Part 1,” *Phys. Rev*, vol. 106, pp. 620–630, 1957.

[11] E. T. Jaynes, “Information Theory and Statistical Mechanics, Part 2,” *Phys. Rev*, vol. 108, pp. 171–190, 1957.

[12] J. P. Burg, *Maximum Entropy Spectral Analysis*, Ph.D. thesis, Stanford University, 1975.

[13] A. Zellner, *An Introduction to Bayesian Inference in Econometrics*, J. Wiley and Sons, New York, 2nd edition, 1971.

[14] J. N Kapur, *Maximum Entropy Models in Science and Engineering*, John Wiley and Sons, Inc, New York, 1989.

[15] G. L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation*, Ph.D. thesis, Wahsington University, St. Louis, 1987.

[16] H. Jeffrey, *Theory of Probability*, Oxford University Press, London, 1939, later editions, 1948, 1961.

[17] E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge, The book will be released in June, 2003.

[18] J.E Shore and R.W Johnson, “Axiomatic Derivation of the Principle of Maximum Entropy and The Principle of Minimum Cross-Entropy,” *IEEE Trans. on Information Theory*, pp. 26–36, Jan. 1980.

[19] Bretthorst G. Larry, “An Introduction to Model Selection Using Probability Theory as Logic,” *in Maximum Entropy and Bayesian Methods, G. R. Heidbreder (ed), Kluwer Academic Publishers, Dordrecht the Netherlands*, pp. 1–42, 1996.

[20] A. Mohammad-Djafari and G. Demoment, “Utilisation de l’Entropie dans les Problèmes de Restauration et de Reconstruction d’Images,” *Traitement du Signal*, vol. 5, pp. 235–248, 1998.

[21] M.A. Xapsos, G.P. Summers, and E.A. Burke, “Probability Model for Peak Fluxes of Solar Proton Events,” *IEEE Transactions on Nuclear Science*, vol. 45, pp. 2948–2953, 1998.

[22] M. Franceschetti, S. Marano, and F. Palmieri, "The Role of Entropy in Wave Propagation," in *IEEE International Symposium on Information Theory*, Yokohama, Japan, July 2003.

[23] E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," *Eur. Trans. Telecomm. ETT*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[24] D. Shiu, G.J Foschini, M.J. Gans, and J. M. Khan, "Fading Correlation and its Effect on the Capacity of Multi-element Antenna Systems ," *IEEE Trans. on Communications*, pp. 502–513, Mar. 2000.

[25] K. Liu, V. Raghavan, and A. M. Sayeed, "Capacity Scaling and Spectral Efficiency in Wideband Correlated MIMO Channels," *IEEE Trans. on Information Theory*, pp. 2504 – 2526, Oct. 2003 2003.

[26] A. Medles, S. Visuri, and D.T.M. Slock, "On MIMO Capacity for Various Types of Partial Channel Knowledge at the Transmitter," IEEE Information Theory Workshop, Paris, France, 2003, pp. 99–102.

[27] C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, "Capacity Scaling in MIMO Wireless Systems under Correlated Fading," *IEEE Trans. on Information Theory*, pp. 637–650, March 2002.

[28] L. Zheng and D. Tse, "Diversity and Multiplexing: A fundamental Tradeoff in Multiple-Antenna Channels," *IEEE Trans. on Information Theory*, pp. 1073–1096, May 2003.

[29] A. Goldsmith, S.A Jafar, N. Jindal, and S. Vishwanath, "Capacity Limits of MIMO Channels," *IEEE Journal on Selected Areas in Communications*, pp. 684–702, vol. 21, no. 5 2003.

[30] A. M. Tulino and S. Verdu, *Random Matrices and Wireless Communications*, Foundations and Trends in Communications and Information Theory, NOW, The Essence of Knowledge, 1st edition, 2004.

[31] M.A Kamath, B.L Hughes, and Y. Xinying, "Gaussian approximations for the capacity of MIMO Rayleigh fading channels," *Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers*, Nov 2002.

[32] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multi-Antenna Channel-Hardening and its Implications for Rate Feedback and Scheduling," *IEEE Trans. on Information Theory*, submitted 2002.

[33] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO Capacity Through Correlated Channels in the Presence of Coorelated Interferers and Noise: A (not so) Large Analysis," Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974, May 2002, can be downloaded at *http://mars.bell-labs.com*.

[34] A. M. Sengupta and P.P. Mitra, "Capacity of Multivariate Channels with Multiplicative Noise: Random Matrix Techniques and Large-N Expansions for Full Transfer Matrices," Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974, Dec. 2002, can be downloaded at *http://mars.bell-labs.com*.

[35] A. Scaglione, "Statistical Analysis of the Capacity of MIMO Frequency Selective Rayleigh Fading Channels with Arbitrary Number of Inputs and Outputs," in *IEEE International Symposium on Information Theory*, Lausanne, Switzerland, July 2002.

[36] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, 1991.

[37] Z.D. Bai and J.W. Silverstein, "CLT of Linear Spectral Statistics of Large Dimensional Sample Covariance Matrices," *Accepted for publication in Annals of Probability*, 2003.

[38] A.J Grant and P.D Alexander, "Random Sequence Multisets for Synchronous Code-Division Multiple-Acess Channels," *IEEE Trans. on Information Theory*, vol. 44, no. 7, pp. 2832–2836, November 1998.

[39] D.N.C Tse and S. Hanly, "Linear multi-user receiver: Effective interference, effective bandwidth and user capacity," *IEEE Trans. on Information Theory*, pp. 641–657, Mar. 1999.

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

[41] J.W. Silverstein and P.L Combettes, "Large Dimensional Random Matrix Theory for Signal Detection and Estimation in Array Processing," *Statistical Signal and Array Processing*, pp. 276–279, 1992.

[42] S. Shamai and S. Verdu, "The impact of frequency-flat fading on the spectral efficiency of cdma," *IEEE Trans. on Information Theory*, pp. 1302–1326, May 2001.

[43] M. Debbah, W. Hachem, P. Loubaton, and M. de Courville, "MMSE Analysis of Certain Large Isometric Random Precoded Systems," *IEEE Transactions on Information Theory, Volume: 49 Issue: 5, Page(s): 1293 -1311*, May 2003.

[44] 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.

[45] R. Müller and S. Verdu, "Design and Analysis of Low-Complexity Interference Mitigation on Vector Channels," *IEEE Journal on Selected Areas in Communications*, pp. 1429–1441, Aug. 2001.

[46] M. Debbah, "Linear Precoders for OFDM Wireless Communications," *Ecole Normale Supérieure de Cachan, France, Phd. Thesis*, Oct. 2002.

[47] X. Mestre Pons, *Space-Time Processing and Channel Estimation: Performance Analysis and Asymptotic Results*, Ph.D. thesis, Universitat Politècnica de Catalunya, Nov. 2002.

[48] E. Biglieri, G. Taricco, and A. Tulino, "How Far is Infinity? Using Asymptotic Analysis in Multiple Antennas Systems," in *IEEE Spread Spectrum Techniques And Applications Proceedings*, 7th International Symposium, 2002, vol. 1, pp. 1–6.

[49] D.N.C Tse and O. Zeitouni, "Linear Multiuser Receivers in Random Environments," *IEEE Trans. on Information Theory*, pp. 171–188, Jan. 2000.

[50] V. L. Girko, "Theory of Random Determinants," *Kluwer Academic Publishers, Dordrecht, The Netherlands*, 1990.

[51] S. V. Hanly and D. N. C. Tse, "Ressource Pooling and Effective Bandwidths in CDMA Networks with Multiuser Receivers and Spatial diversity ," *IEEE Trans. on Information Theory*, pp. 1328–1351, May 2001.

[52] M. Steinbauer, D. Hampicke, G. Sommerkorn, A. Schneider, A. F. Molisch, R. Thoma, and E. Bonek, "Array Measurement of the Double Directional Mobile Radio Channel," in *In Proc. of IEEE VTC*, Tokyo, Japan, May 2000.

[53] A. Poon, D. Tse, and R. Brodersen, "Impact of scattering on the capacity, diversity and propagation range of multiple antenna channels," *IEEE Trans. on Information Theory*, p. submitted, Apr. 2004.

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

[55] R. Müller, "On The Asymptotic Eigenvalue Distribution of Concatenated Vector-valued Fading Channels ," *IEEE Trans. on Information Theory*, pp. 2086 –2091, Jul 2002.

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

[57] Ph. Biane, "Free probability for probabilists," `http://www.dma.ens.fr/~biane/`, 2000, download by choosing "Des notes d'introduction aux probabilités libres".

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

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

[60] 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.

[61] 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.

[62] A. Tulino and S. Verdu, "Asymptotic Analysis of Improved Linear Receivers for BPSK-CDMA Subject to Fading ," *IEEE Trans. on Information Theory*, pp. 1544–1555, August 2001.

[63] D. Tse, "Mutliuser Receivers, Random Matrices and Free Probability," in *Proceedings of the Allerton Conference*, 1999.

[64] E. Biglieri, G. Taricco, and A. Tulino, "Performance of Space-Time Codes for a Large Number of Antennas," *IEEE Trans. on Information Theory*, pp. 1794–1803, July 2002.

[65] R. Müller, "A Random Matrix Model of Communication via Antenna Arrays," *IEEE Trans. on Information Theory*, pp. 2495–2506, Sep 2002.

[66] W. Hachem, "Simple Polynomial MMSE Receivers for CDMA Transmissions on Frequency Selective Channels," *IEEE Trans. on Information Theory*, pp. 164 – 172, January 2004.

[67] P. Loubaton and W. Hachem, "Asymptotic Analysis of Reduced Rank Wiener Filters," IEEE Information Theory Workshop, Paris, France, 2003, pp. 328–331.

[68] L. Cottatellucci and R. Müller, "Asymptotic Design and Analysis of Multistage Detectors with Unequal Powers," IEEE Information Theory Workshop, Bangalore, India, 2002, pp. 167–170.

[69] M. Debbah, "Capacity of a MIMO Multi-user Network," *in*, preparation 2003.

[70] T.L. Marzetta and B.M. Hochwald, "Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading," *IEEE Trans. on Information Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[71] B. Hassibi and B.M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links," *IEEE Trans. on Information Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[72] R. Müller, "On the Accuracy of Modeling the Antenna Array Channel with Random Matrices ," in *International Symposium on Wireless Personal Multimedia Communications*, Aalborg, Denmark, 2001.

[73] A. M. Sayeed, "Deconstructing Multiantenna Fading Channels," *IEEE Trans. on Signal Processing*, pp. 2563–2579, Oct. 2002.

[74] J.P. Kermoal, L. Schumacher, K.I Pedersen, P.E. Mogensen, and F. Frederiken, "A Stochastic MIMO Radio Channel Model with Experimental Validation," *IEEE Journal on Selected Areas in Communications*, pp. 1211–1225, vol. 20, no. 6 2002.

[75] D. Chizhik, J. Lingand P.W. Wolnianski, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple-Input Multiple Output Measurements and Modeling in Manhattan," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3 2002.

[76] S.J. Fortune, D.H. Gay, B.W. Kernighan, O. Landron, R.A Valenzuela, and M. H. Wright, "WiSE Design of Indoor Wireless Systems: Practical Computation and Optimization," *IEEE Comput. Sci. Eng.*, vol. 2, pp. 58–68, Mar. 1995.

[77] D. Gesbert, H. Bölcskei, D. Gore, and A. Paulraj, "MIMO Wireless Channels: Capacity and Performance Prediction," *GLOBECOM conference records*, vol. 2, pp. 1083–1088, 2000.

[78] D. Chizhik, G.J Foschini, M. J. Gans, and R. A. Valenzuela, "Keyholes, Correlations and Capacities of Multielement Transmit and Receive Antennas," *IEEE Trans. on Wireless Communications*, vol. 1, no. 2, pp. 361–368, Apr. 2002.

[79] D. Gesbert, H. Bolcskei, D.A Gore, and A.J Paulraj, "Outdoor MIMO Wireless Channels: Models and Performance Prediction," *IEEE Trans. on Communications*, vol. 50, no. 12, pp. 1926–1934, Dec. 2002.

[80] E. T. Jaynes, "Where Do We Stand on Maximum Entropy?," *in The Maximum Entropy Formalism,R. D. Levine and M. Tribus (eds.), M. I. T. Press, Cambridge, MA,*, p. 15, 1978.

[81] M. Debbah, R. Müller, H. Hofstetter, and P. Lehne, "Validation of mutual information complying mimo models," *submitted to IEEE Transactions on Wireless Communications*, can be downloaded at http://www.eurecom.fr/∼debbah, 2004.