MIMO Channel Modeling and the Principle of Maximum Entropy

Mérouane Debbah, *Member, IEEE,* and Ralf R. Müller, *Member, IEEE*

***Abstract*—We devise theoretical grounds for constructing channel models for multiple-input multiple-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 useful both in terms of designing a system based on criteria such as quality of service and in optimizing transmissions in multiuser networks.**

***Index Terms*—Antenna arrays, Bayesian probability theory, channel modeling, entropy, multiple-input multiple-output (MIMO), random matrices.**

I. Introduction

THE problem of modeling 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 multiple-input multiple-output (MIMO) link is characterized (see Fig. 1) by an $n_r \times n_t$ matrix with independent and identically distributed (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 3-dB 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: the transmitter has to know in which environment the transmission occurs in order to provide codes with adequate properties: as a typical example, in Rayleigh-fading channels, when coding is performed,

Manuscript received December 18, 2003; revised September 4, 2004. This work was supported by the European FLOWS (Flexible Convergence of Wireless Standards and Services) project (http://www.flows-ist.org). The material in this paper was presented in part at the 37th Annual Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, 2003 and the IEEE International Symposium on Signal Processing and Information Technology, Darmstadt, Germany, 2003.

M. Debbah is with the Mobile Communications Group, Institut Eurecom, B.P. 193, 06904 Sophia-Antipolis Cedex (e-mail: debbah@eurecom.fr).

R. R. Müller is with the Department of Electronics and Telecommunications, the Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway (e-maul: mueller@iet.ntnu.no).

Communicated by M. Médard, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2005.846388

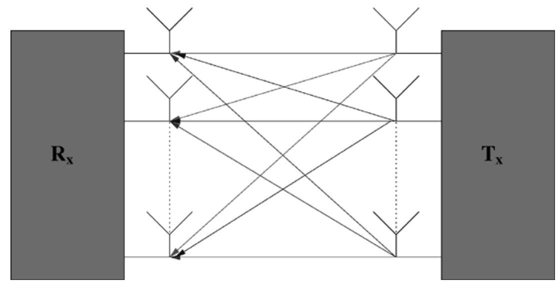

Fig. 1. MIMO channel representation.

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 modeling is the key for better understanding of the limits of wireless transmissions. Questions of the form: "what is the highest transmission rate on a propagation environment where we only know the mean of each path, the energy 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 modeling constraints. We assume that the transmission takes place between a mobile transmitter and a 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

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

with

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

where ρ is the received SNR, t, f, and τ denote, respectively, time, frequency, and delay, $\boldsymbol{y}(t)$ is the $n_r \times 1$ received vector, $\boldsymbol{x}(t)$ is the $n_t \times 1$ transmit vector, $\boldsymbol{n}(t)$ 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 $\boldsymbol{H}(f, t)$ 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, the number of transmit and receive antennas, the 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 $\boldsymbol{H}(f,t)$.

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]–[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 provides a 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 modeling 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 and Johnson [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 will be expressed through the following axiom.

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

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 [17]. 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, 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 [21], 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 consistency axiom. 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 Sectionn V. In Section VI, channel models developed in the 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 Sectionn VIII.

Throughout the paper, for the sake of simplicity, we will often write $\boldsymbol{H}$ instead of $\boldsymbol{H}(f,t)$ without forgetting the dependency on frequency and time. In the following, upper case and lower case boldface symbols will be used for matrices and column vectors, respectively. $(\cdot)^T$ will denote the transpose operator, $(\cdot)^\star$ conjugation, and $(\cdot)^H = \left((\cdot)^T\right)^\star$ Hermitian transpose. $\mathbb{E}$ denotes the expectation operator. $\ln$ is the natural logarithm such that $\ln(e) = 1$. When this notation is used, the mutual information is given in nats per second. When the notation $\log_2(x) = \frac{\ln(x)}{\ln(2)}$ is used, the results are given in bits per second. 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

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

The operator $\mathrm{vec}(\boldsymbol{H})$ stacks all the columns of matrix $\boldsymbol{H}$ 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 sharpen the notions of mutual informations which are used later.

A. MIMO Considerations

Let us first review the pioneering work of Telatar [6] (later published as [22]) that triggered research in multiple-antenna systems. In his paper, Telatar derives the channel capacity of a general MIMO channel. Assuming perfect knowledge of $\boldsymbol{H}$ at the receiver, the ergodic capacity of an $n_r \times n_t$ MIMO channel with input covariance matrix $\boldsymbol{Q} = \mathbb{E}(\boldsymbol{x}\boldsymbol{x}^H)$ is

$$\overline{C} = \max_{\boldsymbol{Q}} \mathbb{E}\left(C(\boldsymbol{Q})\right) \tag{5}$$

with

$$C(\boldsymbol{Q}) = \log_2 \det\left(\boldsymbol{I}_{n_r} + \frac{\rho}{n_t}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^H\right) \tag{6}$$

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

In correlated fading, $C(\boldsymbol{I})$ is called the average mutual information with covariance $\boldsymbol{Q} = \boldsymbol{I}$. In general [23]–[29], capacity is not close to this mutual information except for certain particular cases, see [22], [28]. Often $C(\boldsymbol{I})$ 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 $\boldsymbol{Q}$ on the distribution of $\boldsymbol{H}$ is one more motivation to study the probability distribution of the matrix $\boldsymbol{H}$. Such distributions can be very helpful for system design. One of the visions of future wireless communications the authors would like to advocate (for which the maximum-entropy framework is useful) is the following: suppose that the type of 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 on-line using the maximum entropy approach which incorporates only the available information and nothing more. The transmitted signal and the coding scheme is then (on-line) optimized for that specific scenario, e.g., by deriving a new rank and determinant criteria. 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 criterion 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 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.

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

$$C_q = \max_{\boldsymbol{Q}} \sup\{R : \Pr[C(\boldsymbol{Q}) < R] \leq q\} \tag{7}$$

is the measure of interest.

The covariance matrix $\boldsymbol{Q}$ 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 of the following we will derive the outage mutual information with Gaussian input covariance matrix $\boldsymbol{Q} = \boldsymbol{I}$. *In general, this is only a lower bound to the outage capacity*. Although not optimum, the mutual information with covariance $\boldsymbol{Q} = \boldsymbol{I}$ can be useful in the analysis of systems where the codebook cannot 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 [30]–[34].

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

$$\boldsymbol{y} = \sqrt{\frac{\rho}{n_t}}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}. \tag{8}$$

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

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

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

What distribution $P(\boldsymbol{H})$[1] should the modeler assign to the channel? The modeler would like to derive the most general model complying with those constraints, in other words, the one which maximizes our uncertainty while being certain of the energy. This statement can simply be expressed if one tries to max-

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

imize the following expression using Lagrange multipliers with respect to P:

$$L(P) = -\int d\boldsymbol{H} P(\boldsymbol{H}) \log P(\boldsymbol{H}) + \beta \left[1 - \int d\boldsymbol{H} P(\boldsymbol{H})\right] + \gamma \sum_{i=1}^{n_r} \sum_{j=1}^{n_t} [E - \int d\boldsymbol{H} \mid h_{ij} \mid^2 P(\boldsymbol{H})]. \quad (11)$$

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

$$\frac{dL(P)}{dP} = -1 - \log P(\boldsymbol{H}) - \gamma \sum_{i=1}^{n_r} \sum_{j=1}^{n_t} \mid h_{ij} \mid^2 - \beta = 0 \quad (12)$$

then this yields

$$P(\boldsymbol{H}) = e^{-(\beta + \gamma \sum_{i=1}^{n_r} \sum_{j=1}^{n_t} |h_{ij}|^2)} \quad (13)$$

$$= e^{-\beta} \prod_{i=1}^{n_r} \prod_{j=1}^{n_t} \exp -(\gamma \mid h_{ij} \mid^2) \quad (14)$$

$$= \prod_{i=1}^{n_r} \prod_{j=1}^{n_t} P(h_{ij}) \quad (15)$$

with

$$P(h_{ij}) = e^{-(\gamma |h_{ij}|^2 + \frac{\beta+1}{n_r n_t})}. \quad (16)$$

One of the most important conclusions of the maximum-entropy principle is that while we have only assumed the energy, this assumption implies independent entries since the joint probability distribution $P(\boldsymbol{H})$ 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 interesting 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. 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 modeling approach (as a constraint on the covariance matrix, for example).

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

$$\int d\boldsymbol{H} \sum_{i=1}^{n_r} \sum_{j=1}^{n_t} \mid h_{ij} \mid^2 P(\boldsymbol{H}) = n_t n_r E \quad (17)$$

$$\int d\boldsymbol{H} P(\boldsymbol{H}) = 1. \quad (18)$$

Solving the previous equations yields the following probability distribution:

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

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 energies such as $\mathbb{E}(\mid h_{ij} \mid^2) = E_{ij}$. Using the same methodology, it can be shown that

$$P(\boldsymbol{H}) = \prod_{i=1}^{n_r} \prod_{j=1}^{n_t} P(h_{ij}) \quad (19)$$

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 energy $\mathbb{E}(\mid h_{pk} - m_{pk} \mid^2) = E_{pk}$, all the other paths having different variances (but nothing is said about the mean). Using the same methodology as before, we show that

$$P(\boldsymbol{H}) = \prod_{i=1}^{n_r} \prod_{j=1}^{n_t} P(h_{ij}). \quad (20)$$

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 $\int g_i(\boldsymbol{H}) P(\boldsymbol{H}) d\boldsymbol{H} = \alpha_i$ for $i = 1, \ldots, N$, then the principle of maximum entropy attributes the distribution [35]

$$P(\boldsymbol{H}) = e^{(-1 + \lambda + \sum_{i=1}^{N} \lambda_i g_i(\boldsymbol{H}))} \quad (21)$$

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 derives the ergodic capacity for the i.i.d. channel model when the channel is known at the receiver only.

For the outage probability, no simple tractable closed-form solution is available. 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.* [30] which we recall for usefulness purposes.[2]

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

$$\begin{aligned}\mu_{\text{iid}}(\gamma,\rho) &= \int_0^\infty \ln(1+\rho\lambda) dF_{\text{iid}}(\lambda) \\ &= \gamma \ln(1+\rho-\rho\alpha_{\text{iid}}(\gamma,\rho)) - \alpha_{\text{iid}}(\gamma,\rho) \\ &\quad + \ln(1+\rho\gamma-\rho\alpha_{\text{iid}}(\gamma,\rho))\end{aligned} \tag{22}$$

and

$$\sigma^2{}_{\text{iid}}(\gamma,\rho) = -\ln\left[1 - \frac{\alpha^2{}_{\text{iid}}(\gamma,\rho)}{\gamma}\right] \tag{23}$$

with

$$\alpha_{\text{iid}}(\gamma,\rho) = \frac{1}{2}\left[1+\gamma+\frac{1}{\rho} - \sqrt{\left(1+\gamma+\frac{1}{\rho}\right)^2 - 4\gamma}\right]. \tag{24}$$

It is noteworthy to note that in this case, the capacity is achieved for $\boldsymbol{Q} = \boldsymbol{I}$. 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

$$\frac{1}{n_t}\sum_{i=1}^{n_t} l(\lambda_i) = \int l(x) dF^{\boldsymbol{B}_t}(x) \tag{25}$$

where $(\lambda_1, \ldots, \lambda_{n_t})$ denotes the eigenvalues of matrix $\boldsymbol{B}_t$

$$F^{\boldsymbol{B}_t}(\lambda) = \frac{1}{n_t} \mid \{j : \lambda_j \le \lambda\} \mid$$

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

$$n_t \mu_{\text{iid}} = \min(n_t, n_r) \ln(\rho) \tag{26}$$

$$\sigma^2{}_{\text{iid}} = \begin{cases} -\ln\left(1 - \frac{\min(n_t,n_r)}{\max(n_t,n_r)}\right), & \text{if } n_t \ne n_r \\ \frac{1}{2}\ln(\rho), & \text{if } n_t = n_r. \end{cases} \tag{27}$$

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 and Alexander [38], Tse and Hanly [39], Verdú and Shamai [40], and Rapajic and Popescu [41], random matrix theory entered the field of telecommunications.[3] Since then, random matrix theory has become a standard tool for the analysis of code-division multiple access (CDMA) in its various fashions and applications [43]–[48]. 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 a few meaningful parameters. In the following, we will briefly illustrate how many antennas are required for large system approximations to be reasonably tight.

The cumulative distribution function (CDF) of the capacity is given by Theorem 1 as

$$F(C) = 1 - Q\left(\frac{C - n_t\mu}{\sigma}\right). \tag{28}$$

In Fig. 2, the CDF is plotted for a system with 1×1, 2×2, and 4×4 antennas for an SNR of 10 dB. There is a quite good match between the asymptotic 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 arguments can be found in [49], [36], [31], [50], [51]. 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 [52].

IV. Knowledge of the Directions of Arrival and Departure

In this section, we address thoroughly the double-directional model. Cases with lesser 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 $\boldsymbol{H}$ has a certain energy. There is no knowledge on the mean. The case where the paths have different nonzero means can be treated in the same way. 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 channel statistics are supposed not to change during the modeling phase. However, the channel realizations do vary. The modeler has knowledge of the directions of departure $\boldsymbol{\Psi}_{s_t \times n_t}$ from the transmitting antennas to a set of scatterers (A). He also knows the directions of arrival $\boldsymbol{\Phi}_{n_r \times s_r}$ from a set of scatterers (B) to the receiving antennas, see Fig. 3. The modeler also knows the powers of the steering directions. However, the modeler has no knowledge of what happens between the two

[2]The mean and the variance of the mutual information were also derived using the replica method in [32], [33], [36].

[3]It should be noted that in the field of array processing, Silverstein used already in 1992 random matrix theory [42] for signal detection and estimation.

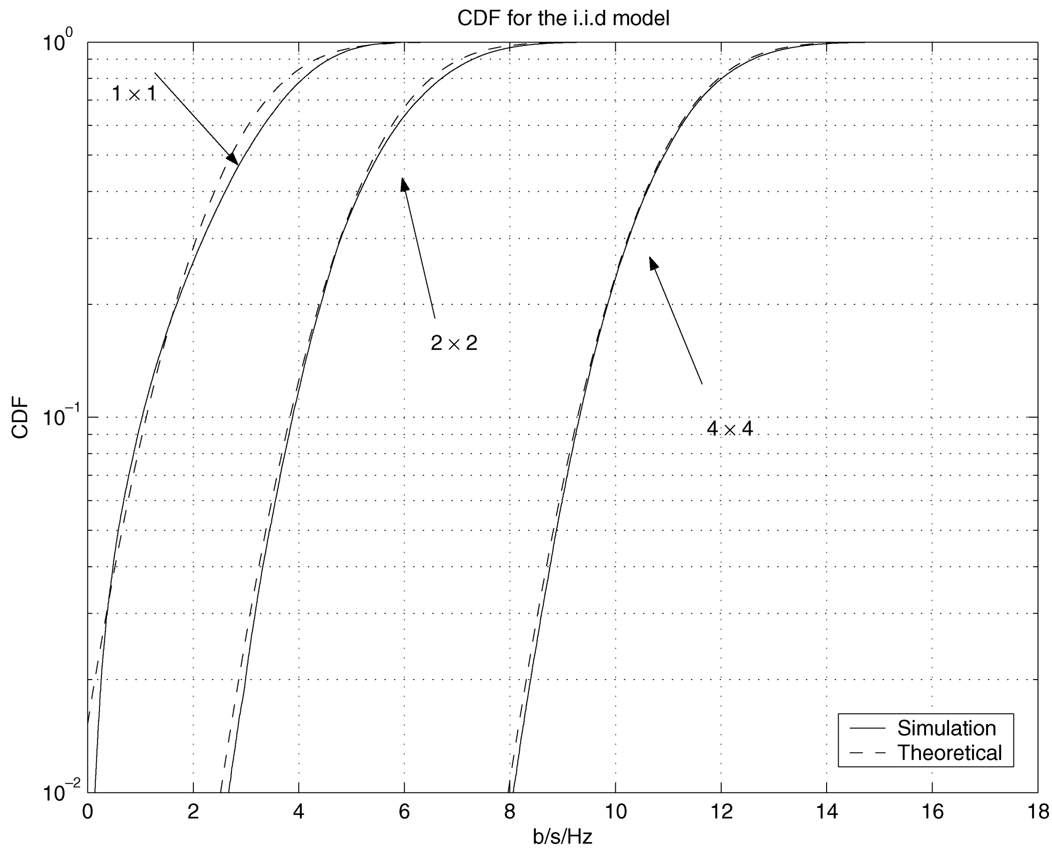

Fig. 2. CDF of mutual information for the i.i.d. Gaussian model at 10-dB SNR. Dashed lines and solid lines show the true simulated CDF and the theoretical asymptotic limit, respectively.

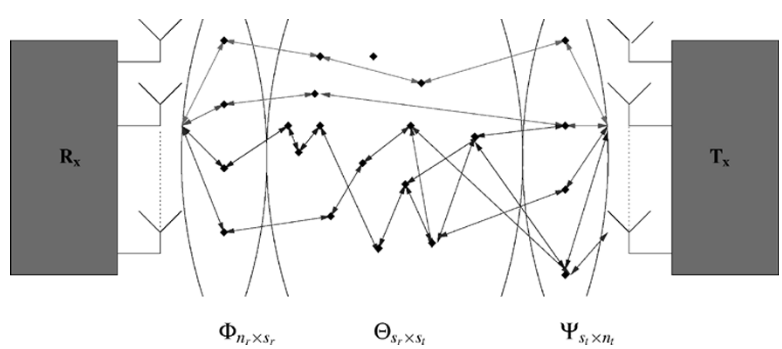

Fig. 3. Double-directional based model.

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 modeler 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

$$\boldsymbol{s}(t, \boldsymbol{d}) = \boldsymbol{s}_0 \, e^{j(\boldsymbol{k}^T \boldsymbol{d} - 2\pi f t)} \tag{29}$$

which is the plane-wave solution of the Maxwell equations in free nondispersive space for wave vector $\boldsymbol{k} \in \mathbb{R}^{2\times 1}$ and location vector $\boldsymbol{d} \in \mathbb{R}^{2\times 1}$. 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 $\boldsymbol{s}_0$ gives us knowledge on the polarization of the wave while the direction of the wave vector $\boldsymbol{k}$ gives us knowledge on the direction of propagation. The phase of the signal results in $\phi = \boldsymbol{k}^T \boldsymbol{d}$. 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

$$\boldsymbol{H} = \frac{1}{\sqrt{s_r s_t}} \boldsymbol{\Phi}_{n_r \times s_r} \boldsymbol{P}^{r\frac{1}{2}} \boldsymbol{\Theta}_{s_r \times s_t} \boldsymbol{P}^{t\frac{1}{2}} \boldsymbol{\Psi}_{s_t \times n_t} \tag{30}$$

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

$$\boldsymbol{\Phi}_{n_r \times s_r} = \begin{pmatrix} e^{j\phi_{1,1}} & \dots & e^{j\phi_{1,s_r}} \\ \vdots & \ddots & \vdots \\ e^{j\phi_{n_r,1}} & \dots & e^{j\phi_{n_r,s_r}} \end{pmatrix} \tag{31}$$

and

$$\boldsymbol{\Psi}_{s_t \times n_t} = \begin{pmatrix} e^{j\psi_{1,1}} & \dots & e^{j\psi_{1,n_t}} \\ \vdots & \ddots & \vdots \\ e^{j\psi_{s_t,1}} & \dots & e^{j\psi_{s_t,n_t}} \end{pmatrix} \tag{32}$$

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 Fig. 3. The phases $\phi_{i,j} = (\boldsymbol{k}^{r^T} \boldsymbol{d}^r)_{i,j}$ and $\psi_{i,j} = (\boldsymbol{k}^{t^T} \boldsymbol{d}^t)_{i,j}$ are given as scalar products between the respective wave vectors and the respective locations of the scatterers. The powers of the steering directions are given by the diagonal matrices $\boldsymbol{P}^r$ and $\boldsymbol{P}^t$

$$\boldsymbol{P}^{r\frac{1}{2}} = \begin{pmatrix} \sqrt{P_1^r} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_r}^r} \end{pmatrix} \tag{33}$$

$$\boldsymbol{P}^{t\frac{1}{2}} = \begin{pmatrix} \sqrt{P_1^t} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_t}^t} \end{pmatrix}. \tag{34}$$

Remark 1: In the Introduction, we have recalled the work of Shore and Johnson [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, nested one into the other.

The consistency argument, see Axiom 1, states that if the DoAs, the powers $\boldsymbol{P}^t$ and $\boldsymbol{P}^r$, and the DoDs are unknown, then the channel matrix $\boldsymbol{H}$ in (30) should be assigned an i.i.d. zero mean Gaussian distribution, see Sectionn III-A, since the modeler is in the same state of knowledge as before where only the energy was known. Based on this consistency requirement, we can determine the distribution of $\boldsymbol{\Theta}_{s_r \times s_t}$.

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

$$\begin{aligned} P(\boldsymbol{H}) = \int & P(\boldsymbol{\Phi}\boldsymbol{P}^{r\frac{1}{2}}\boldsymbol{\Theta}\boldsymbol{P}^{t\frac{1}{2}}\boldsymbol{\Psi} \mid \boldsymbol{\Phi}, \boldsymbol{\Psi}, \boldsymbol{P}^r, \boldsymbol{P}^t, s_r, s_t) \\ & P(\boldsymbol{\Psi}, \boldsymbol{\Phi} \mid s_r, s_t) P(\boldsymbol{P}^r, \boldsymbol{P}^t \mid s_t, s_r) \\ & P(s_t, s_r) ds_r ds_t d\boldsymbol{P}^r d\boldsymbol{P}^t d\boldsymbol{\Psi} d\boldsymbol{\Phi}. \end{aligned} \tag{35}$$

- When $\boldsymbol{\Psi}, \boldsymbol{\Phi}, s_r, s_t, \boldsymbol{P}^r, \boldsymbol{P}^t$ are known: $P(\boldsymbol{\Phi}\boldsymbol{\Psi} \mid s_r, s_t) = \delta(\boldsymbol{\Phi} - \boldsymbol{\Phi}_0)\delta(\boldsymbol{\Psi} - \Psi_0)$, $P(s_t, s_r) = \delta(s_{r_0} - a)\delta(s_{t_0} - b)$, $P(\boldsymbol{P}^{r\frac{1}{2}}, \boldsymbol{P}^{t\frac{1}{2}} \mid s_r, s_t) = \delta(\boldsymbol{P}^{r\frac{1}{2}} - \boldsymbol{P_0}^{r\frac{1}{2}})\delta(\boldsymbol{P}^{t\frac{1}{2}} - \boldsymbol{P_0}^{t\frac{1}{2}})$, and
$$P(\boldsymbol{H}) = P(\Phi_0 \boldsymbol{P_0}^{r\frac{1}{2}} \boldsymbol{\Theta} \boldsymbol{P_0}^{t\frac{1}{2}} \boldsymbol{\Psi}_0). \tag{36}$$
- Suppose now that $\boldsymbol{\Psi}, \boldsymbol{\Phi}, s_r, s_t$ are unknown, we must find (by solving (35)) the matrix $\boldsymbol{\Theta}_{s_r \times s_t}$ such as each entry h_{ij} of $\boldsymbol{H}$ has an i.i.d. zero mean Gaussian distribution. In this case, the following result holds.

Proposition 1: $\boldsymbol{\Theta}_{s_r \times s_t}$ i.i.d. zero-mean Gaussian with variance 1[4] is solution of the consistency argument and maximizes entropy.

Proof: Since $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are unknown, the principle of maximum entropy attributes i.i.d. uniformly distributed angles over $[0, 2\pi]$ 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

$$h_{ij} = \frac{1}{\sqrt{s_t s_r}} \sum_{k=1}^{s_t} \sum_{p=1}^{s_r} \theta_{pk} \sqrt{P_k}^t \sqrt{P_p}^r e^{j\psi_{kj}} e^{j\phi_{ip}}$$

then

$$\begin{aligned} & P(h_{ij} \mid \boldsymbol{\Psi}, \boldsymbol{\Phi}, s_r, s_t) \\ & = N(0, \frac{1}{s_t s_r} \sum_{p=1}^{s_r} \sum_{k=1}^{s_t} \mid \sqrt{P_p}^r e^{j\phi_{ip}} \sqrt{P_k}^t e^{j\psi_{kj}} \mid^2 = 1) \\ & = \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}} \end{aligned}$$

(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,

$$\begin{aligned} P(h_{ij}) = \int & \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}} P(\boldsymbol{\Phi}, \boldsymbol{\Psi} \mid s_t, s_r) \\ & P(\boldsymbol{P}^r, \boldsymbol{P}^t \mid s_t, s_r) P(s_t, s_r) d\boldsymbol{\Phi} d\boldsymbol{\Psi} \\ & d\boldsymbol{P}^r d\boldsymbol{P}^t ds_t ds_r \\ = & \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}} \int P(\boldsymbol{\Phi}, \boldsymbol{\Psi} \mid s_t, s_r) \\ & P(\boldsymbol{P}^r, \boldsymbol{P}^t \mid s_t, s_r) P(s_t, s_r) d\boldsymbol{\Phi} d\boldsymbol{\Psi} \\ & d\boldsymbol{P}^r d\boldsymbol{P}^t ds_t ds_r \\ = & \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}}. \end{aligned} \tag{37}$$

Moreover, we have

$$\begin{aligned} & \mathbb{E}_{\boldsymbol{\Phi},\boldsymbol{\Psi},\boldsymbol{\Theta}}(h_{ij} h_{mn}^*) \\ & \quad = \frac{1}{s_t s_r} \sum_{k=1}^{s_t} \sum_{p=1}^{s_r} \sum_{r=1}^{s_t} \sum_{l=1}^{s_r} \mathbb{E}_{\boldsymbol{\Theta}}(\theta_{pk} \theta_{lr}^*) \end{aligned} \tag{38}$$

[4]We suppose for simplicty sake that the energy $E = 1$.

$$\mathbb{E}_{\boldsymbol{\Psi}}(e^{-j\psi_{rn}+j\psi_{kj}})\mathbb{E}_{\boldsymbol{\Phi}}(e^{-j\phi_{ml}+j\phi_{ip}}) \sqrt{P_k^{\,t}}\sqrt{P_r^{\,t}}\sqrt{P_p^{\,r}}\sqrt{P_l^{\,r}}. \tag{39}$$

Therefore,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\Phi},\boldsymbol{\Psi},\boldsymbol{\Theta}}(h_{ij}h_{mn}^*) &= \frac{1}{s_t s_r}\sum_{k=1}^{s_t}\sum_{p=1}^{s_r}\sum_{r=1}^{s_t}\sum_{l=1}^{s_r}\delta_{pl}\delta_{kr} \\ &\quad \mathbb{E}_{\boldsymbol{\Psi}}(e^{-j\psi_{rn}+j\psi_{kj}})\mathbb{E}_{\boldsymbol{\Phi}}(e^{-j\phi_{ml}+j\phi_{ip}}) \\ &\quad \sqrt{P_k^{\,t}}\sqrt{P_r^{\,t}}\sqrt{P_p^{\,r}}\sqrt{P_l^{\,r}} \\ &= \frac{1}{s_t s_r}\sum_{k=1}^{s_t}\sum_{p=1}^{s_r}\mathbb{E}_{\boldsymbol{\Psi}}(e^{-j\psi_{kn}+j\psi_{kj}}) \quad (40)\\ &\quad \mathbb{E}_{\boldsymbol{\Phi}}(e^{-j\phi_{mp}+j\phi_{ip}})P_k^{\,t}P_p^{\,r} \quad (41)\\ &= \delta_{im}\delta_{jn}\frac{1}{s_t s_r}\sum_{k=1}^{s_t}\sum_{p=1}^{s_r}P_k^{\,t}P_p^{\,r} \quad (42)\\ &= \delta_{im}\delta_{jn} \quad (43)\end{aligned}$$

which proves that $\boldsymbol{\Theta}_{s_r\times s_t}$ is the 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 $\boldsymbol{H}$ are correlated, for general DoAs and DoDs.

B. Mutual Information: General Case

In this subsection, we are interested in the analysis of the scaling of the mutual information with respect to the numbers of scatterers. Denote $\gamma = \frac{n_r}{s_r}$, $\xi = \frac{s_r}{n_t}$, $\gamma_1 = \frac{n_r}{s_t}$, $\xi_1 = \frac{s_t}{n_t}$. Let us first make some assumptions[5] on the matrix of the directions of arrival and the matrix of directions of departure.

Assumption:

- When the matrix size $\frac{1}{s_r}\boldsymbol{P}^{r\,\frac{1}{2}}\boldsymbol{\Phi}_{n_r\times s_r}^H\boldsymbol{\Phi}_{n_r\times s_r}\boldsymbol{P}^{r\,\frac{1}{2}}$ grows large with $\gamma = \frac{n_r}{s_r}$ remaining fixed, the empirical eigenvalue distribution S_{s_r,n_r} of $\frac{1}{s_r}\boldsymbol{P}^{r\,\frac{1}{2}}\boldsymbol{\Phi}_{n_r\times s_r}^H\boldsymbol{\Phi}_{n_r\times s_r}\boldsymbol{P}^{r\,\frac{1}{2}}$ converges in distribution to a fixed distribution S_{doa}
$$S_{s_r,n_r}(\lambda) = \frac{1}{s_r}\mid\{j:\lambda_j\le\lambda\}\mid\rightarrow S_{\mathrm{doa}}(\lambda).$$
- When the matrix size $\frac{1}{s_t}\boldsymbol{P}^{t\,\frac{1}{2}}\boldsymbol{\Psi}_{s_t\times n_t}\boldsymbol{\Psi}_{s_t\times n_t}^H\boldsymbol{P}^{t\,\frac{1}{2}}$ grows large with $\xi_1 = \frac{s_t}{n_t}$ remaining fixed, the empirical eigenvalue distribution S_{s_t,n_t} of $\frac{1}{s_t}\boldsymbol{P}^{t\,\frac{1}{2}}\boldsymbol{\Psi}_{s_t\times n_t}\boldsymbol{\Psi}_{s_t\times n_t}^H\boldsymbol{P}^{t\,\frac{1}{2}}$ converges in distribution to a fixed distribution S_{dod}
$$S_{s_t,n_t}(\lambda) = \frac{1}{s_t}\mid\{j:\lambda_j\le\lambda\}\mid\rightarrow S_{\mathrm{dod}}(\lambda).$$

The asymptotic mutual information per transmitting antenna is given by

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

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

The asymptotic mutual information per transmitting antenna with two-sided correlation has been derived previously in [53], [29], [54], [32], [33][6] using either results of Girko [55], the replica method [56] or free probability theory[7] [60]. The results can be applied to model (30) and yield the following.

Proposition 2: As the size of the system grows large but γ, γ_1, ξ, ξ_1 remain fixed, then the asymptotic mutual information per transmitting antenna is given by

$$\mu = \xi_1\int\ln(1+\rho\lambda\alpha_{\mathrm{dod}})dS_{\mathrm{dod}}(\lambda) - \rho\alpha_{\mathrm{doa}}\alpha_{\mathrm{dod}} + \xi\int\ln(1+\rho\lambda\alpha_{\mathrm{doa}})dS_{\mathrm{doa}}(\lambda) \tag{44}$$

with

$$\alpha_{\mathrm{doa}} = \xi_1\int\frac{\lambda}{1+\rho\lambda\alpha_{\mathrm{dod}}}dS_{\mathrm{dod}}(\lambda) \tag{45}$$

and

$$\alpha_{\mathrm{dod}} = \xi\int\frac{\lambda}{1+\rho\lambda\alpha_{\mathrm{doa}}}dS_{\mathrm{doa}}(\lambda). \tag{46}$$

Proposition 2 is general enough to be applied for the i.i.d. Gaussian case, the DoA-based model, and the DoD-based model. The formula is extremely useful as it shows that only the limiting eigenvalue distribution of the steering directions with powers matters: in other words, two antenna configurations 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

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

[6]In contributions [32], [33], the second moment of the mutual information of MIMO correlated channels is also derived using the replica method. In this paper, for some particular cases (see the following sections), the distribution of the mutual information is proved to be asymptotically Gaussian and the variance is provided using random matrix theory.

[7]Free probability [57], [58] is a noncommutative probability theory, in which the concept of independence of classical probability is replaced by that of freeness. Voiculescu [59]–[61] discovered very important relations between the free probability theory and the random matrix theory. He showed in particular that random matrices can be considered as typical noncommutative random variables. To the authors' knowledge, the first use of free probability in the telecommunication field was made by Evans and Tse in 1999 [62]. Since that date, it has been used for the performance analysis of several transmission schemes (CDMA [63], [64], orthogonal frequency-division multiplexing (OFDM) [44], [47], [45], and MIMO [65], [66], [48]). 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 [67]–[69] (multistage detectors, etc.).

beginning disposed in a Uniform Linear Array (ULA) 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.

Although we have no formal proof on the uniqueness except in the case of the DoA-based model, see Section IV-E (the mean mutual information of Proposition 2 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 easily (see the example proof of Proposition 6 in the DoA case) shown to scale at high SNR as

$$\min\left(n_t, n_r, s_t \int_{\lambda>0} dS_{\mathrm{dod}}(\lambda), s_r \int_{\lambda>0} dS_{\mathrm{doa}}(\lambda)\right) \ln(\rho).$$

The integral is on the support of nonzero eigenvalues and $\int_{\lambda>0} dS_{\mathrm{doa}}(\lambda)$ and $\int_{\lambda>0} dS_{\mathrm{dod}}(\lambda)$ express, respectively, the correlation factor of the s_r and s_t scatterers. Hence, the previous result generalizes the multiplexing gain of i.i.d. MIMO systems [6] and gives an upper bound on the number of antennas to be used for a given scattering environment.

C. Mutual Information: ULA and Fourier Directions Case

In this subsection, 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: $[1, e^{-j2\pi \frac{d \sin(\phi)}{\lambda}}, \ldots, e^{-j2\pi \frac{d(n_r-1)\sin(\phi)}{\lambda}}]$. d is the antenna spacing and ϕ is the direction of arrival[8] which is defined as the angle between a line perpendicular to the incoming wavefront and a reference line through the array. The same holds for the directions of departure

$$\boldsymbol{\Phi}_{n_r \times s_r} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ e^{j2\pi \frac{d(n_r-1)\sin(\phi_1)}{\lambda}} & \cdots & e^{j2\pi \frac{d(n_r-1)\sin(\phi_{s_r})}{\lambda}} \end{pmatrix} \tag{47}$$

and

$$\boldsymbol{\Psi}_{s_t \times n_t} = \begin{pmatrix} 1 & \cdots & e^{j2\pi \frac{d(n_t-1)\sin(\psi_1)}{\lambda}} \\ \vdots & \ddots & \vdots \\ 1 & \cdots & e^{j2\pi \frac{d(n_t-1)\sin(\psi_{s_t})}{\lambda}} \end{pmatrix}. \tag{48}$$

For the sake of simplicity, 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 in the case where for any i there exists a k such that $\sin(\phi_i) = \frac{2k}{n_r}$

[8]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.

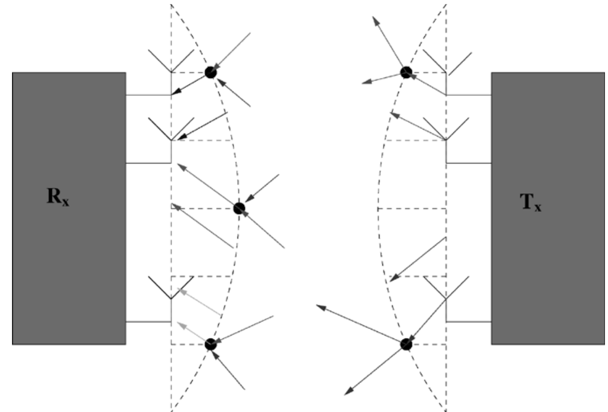

Fig. 4. Simple case: scatterers positioned on special directions.

(see Fig. 4) and for any j there exists a l such that $\sin(\psi_j) = \frac{2l}{n_t}$. 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 (related to the virtual representation, see Section VI-B).

1) Equal Power Case: We will assume in this part that $\boldsymbol{P}^r = \boldsymbol{I}_{s_r}$ and $\boldsymbol{P}^t = \boldsymbol{I}_{s_t}$. As a consequence, the DoA and DoD steering matrices have the following limiting eigenvalue distribution:

$$S_{\mathrm{doa}}(\lambda) = \delta(\lambda - \gamma) \tag{49}$$

and

$$S_{\mathrm{dod}}(\lambda) = \delta\left(\lambda - \frac{1}{\xi_1}\right). \tag{50}$$

Proposition 3: The asymptotic mutual information per transmitting antenna and the asymptotic variance of the mutual information for the double directional model in the equal power and Fourier directions case are, respectively, given by

$$\begin{aligned} \mu_{\mathrm{double}} = &\, \xi \ln(1 + \rho\gamma - \rho\gamma\alpha_{\mathrm{double}}) - \xi_1 \alpha_{\mathrm{double}} \\ &+ \xi_1 \ln(1 + \rho\gamma_1 - \rho\gamma\alpha_{\mathrm{double}}) \end{aligned} \tag{51}$$

$$\sigma^2{}_{\mathrm{double}} = -\ln\left[1 - \frac{\alpha_{\mathrm{double}}^2 \gamma}{\gamma_1}\right] \tag{52}$$

with

$$\alpha_{\mathrm{double}} = \frac{1}{2}\left[1 + \frac{\gamma_1}{\gamma} + \frac{1}{\rho\gamma} - \sqrt{(1 + \frac{\gamma_1}{\gamma} + \frac{1}{\rho\gamma})^2 - 4\frac{\gamma_1}{\gamma}}\right]. \tag{53}$$

Proof: One can notice that:

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

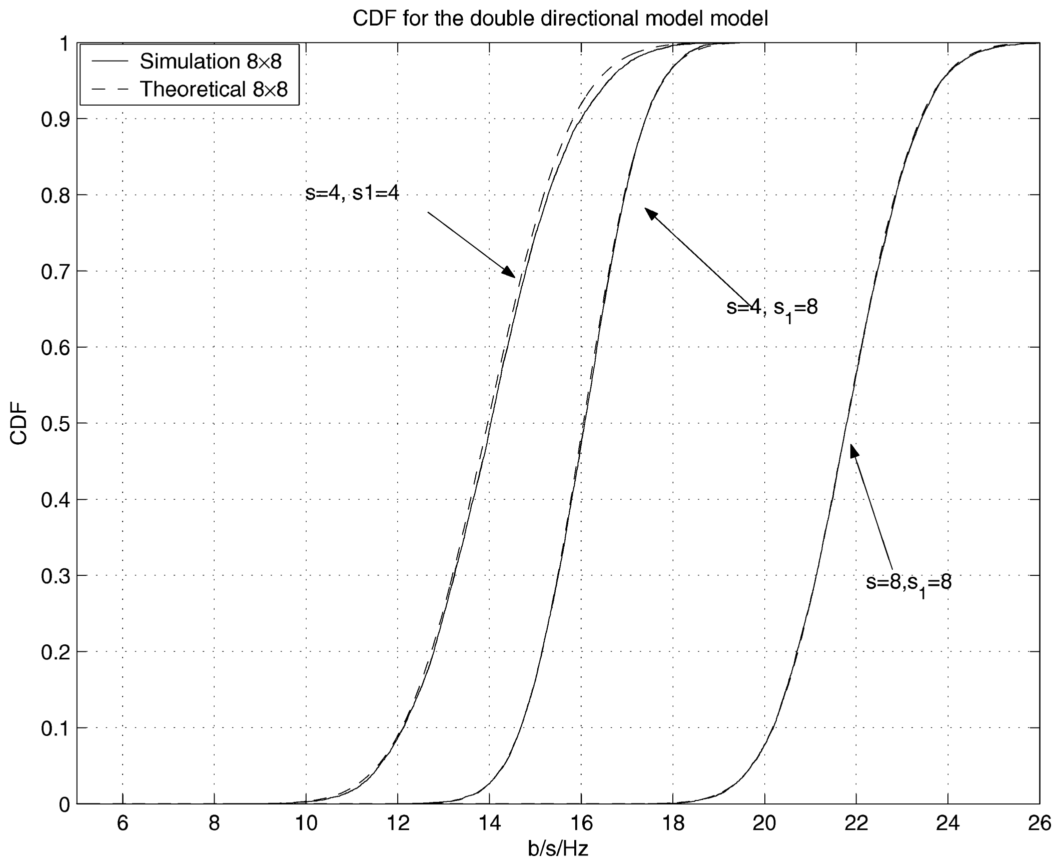

Fig. 5. Mutual information cumulative distribution in the case of the double-directional model with equal power on Fourier directions.

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

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

in the formulas of Theorem 1 then the result is proved. □

At high SNR, it can be easily shown that

$$n_t\mu_{\text{double}} = \min(s_t, s_r)\ln(\rho) \tag{54}$$

$$\sigma^2{}_{\text{double}} = \begin{cases} -\ln\left(1 - \frac{\min(s_t,s_r)}{\max(s_t,s_r)}\right), & \text{if } s_t \neq s_r \\ \frac{1}{2}\ln(\rho), & \text{if } s_t = s_r. \end{cases} \tag{55}$$

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

In Fig. 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 $s_t = 8$.

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 transmitting antennas, we have plotted in Fig. 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 $s_r = s_t = n_r = n_t$.

2) Nonequal Power Case: We consider the case where 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 that $\sum_{i=1}^{K_r} l_i{}^r = 1$) and K_t distinct amplitudes $\sqrt{P_i{}^t}$ of the transmitting steering vectors with weight $l_i{}^t$ (such that $\sum_{i=1}^{K_t} l_i{}^t = 1$). As a consequence, the limiting eigenvalue distribution S_{doa} of $\frac{1}{s_r}\boldsymbol{P}^{r\frac{1}{2}}\mathbf{\Phi}^H\mathbf{\Phi}\boldsymbol{P}^{r\frac{1}{2}}$ has the following expression:

$$S_{\text{doa}}(\lambda) = \sum_{i=1}^{K_r} l_i{}^r\delta(\lambda - \gamma P_i{}^r). \tag{56}$$

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

$$S_{\text{dod}}(\lambda) = \sum_{i=1}^{K_t} l_i{}^t\delta(\lambda - \frac{P_i{}^t}{\xi_1}). \tag{57}$$

Proposition 4: In this case, μ_{double} is equal to

$$\mu_{\text{double}} = \xi_1\sum_{i=1}^{K_t} l_i{}^t\ln\left(1 + \frac{\rho P_i{}^t\alpha_{\text{dod}}}{\xi_1}\right) - \rho\alpha_{\text{doa}}\alpha_{\text{dod}} + \xi\sum_{i=1}^{K_r} l_i{}^r\ln(1 + \rho P_i{}^r\gamma\alpha_{\text{doa}}) \tag{58}$$

with

$$\alpha_{\text{doa}} = \sum_{i=1}^{K_t}\frac{l_i{}^tP_i{}^t}{1 + \frac{\rho P_i{}^t\alpha_{\text{dod}}}{\xi_1}} \tag{59}$$

and

$$\alpha_{\rm dod} = \xi \sum_{i=1}^{K_r} l_i{}^r \frac{P_i{}^r \gamma}{1 + \rho\gamma P_i{}^r \alpha_{\rm doa}}. \tag{60}$$

Proof: The proof is an application of the general Proposition 2 to 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 profile.

Proposition 5: The mean of the mutual information in the case of the double directional model with ULA and Fourier directions is maximized for $\boldsymbol{P}^r = \boldsymbol{I}_{s_r}$ and $\boldsymbol{P}^t = \boldsymbol{I}_{s_t}$.

Proof: The proof is provided in Appendix B. □

The result acknowledges the fact that the best throughput is obtained when all the steering directions have the same power on both sides. Intuitively, one can easily understand this observation: any imbalance of power will reduce the effective number of scatterers and therefore the diversity generated by the environment.

D. Mutual Information: 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 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 $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$

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 $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ are uniformly distributed over $[0, 2\pi]$. This case will be referred to as 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 distributions of $\frac{1}{s_r}\boldsymbol{\Phi}^H\boldsymbol{\Phi}$ and $\frac{1}{s_t}\boldsymbol{\Psi}\boldsymbol{\Psi}^H$ are well known in the literature [70] and Proposition 2 can be applied straightforwardly. However, we will take Müller's approach, as our framework is a particular case of [71] where he introduces an N-fold scattering model as a product of N i.i.d. random matrices $\boldsymbol{H} = \prod_{i=1}^{N} \boldsymbol{M}_i$. Using free probability theory, he proves the almost-sure convergence of the limiting eigenvalue distribution of matrix $\boldsymbol{H}$ and gives an explicit form of its Stieltjes transform. In the case considered here, $\boldsymbol{H} = \boldsymbol{\Phi\Theta\Psi}$ is the product of three random matrices. Using the results in [71], it can be easily shown that the Stieltjes transform $m_{\boldsymbol{H}^H\boldsymbol{H}}(x)$ is a solution of the following equation:

$$(v(x,\xi_1)v(x,\xi)v(x,\gamma\rho))\, m_{\boldsymbol{H}}^{\boldsymbol{H}}\boldsymbol{H}(-x) + x m_{\boldsymbol{H}^H\boldsymbol{H}}(-x) = 1 \tag{61}$$

where $v(x,\alpha) = \frac{x m_{\boldsymbol{H}^H\boldsymbol{H}}(-x) - 1 + \alpha}{\alpha}$. Since $m_{\boldsymbol{H}^H\boldsymbol{H}}(\frac{-1}{\rho}) = \rho(1 - \rho\frac{d\mu}{d\rho})$ 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 Fig. 7 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 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-C1). 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 $\boldsymbol{\Phi}$, and DoD matrix $\boldsymbol{\Psi}$ are unitary Fourier matrices and have therefore no effect on $\boldsymbol{\Theta}_{s_r \times s_t}$. However, in the random i.i.d. directions scenario, the nonunitary steering matrix $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ have a correlation effect on matrix $\boldsymbol{\Theta}_{s_r \times s_t}$. 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 geometry.

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

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. Without going into detailed calculus (the proof is a special case of Proposition 1), it can be shown that the resulting model has the following form:

$$\boldsymbol{H} = \frac{1}{\sqrt{s_r}} \boldsymbol{\Phi} \boldsymbol{P}^{r\,\frac{1}{2}} \boldsymbol{\Theta}. \tag{62}$$

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

1) General Case: We are interested in the behavior of

$$I^M(n_t, n_r, s_r, \rho) = \log_2 \det\left(\boldsymbol{I}_{n_t} + \frac{\rho}{n_t}\boldsymbol{H}^H\boldsymbol{H}\right)$$

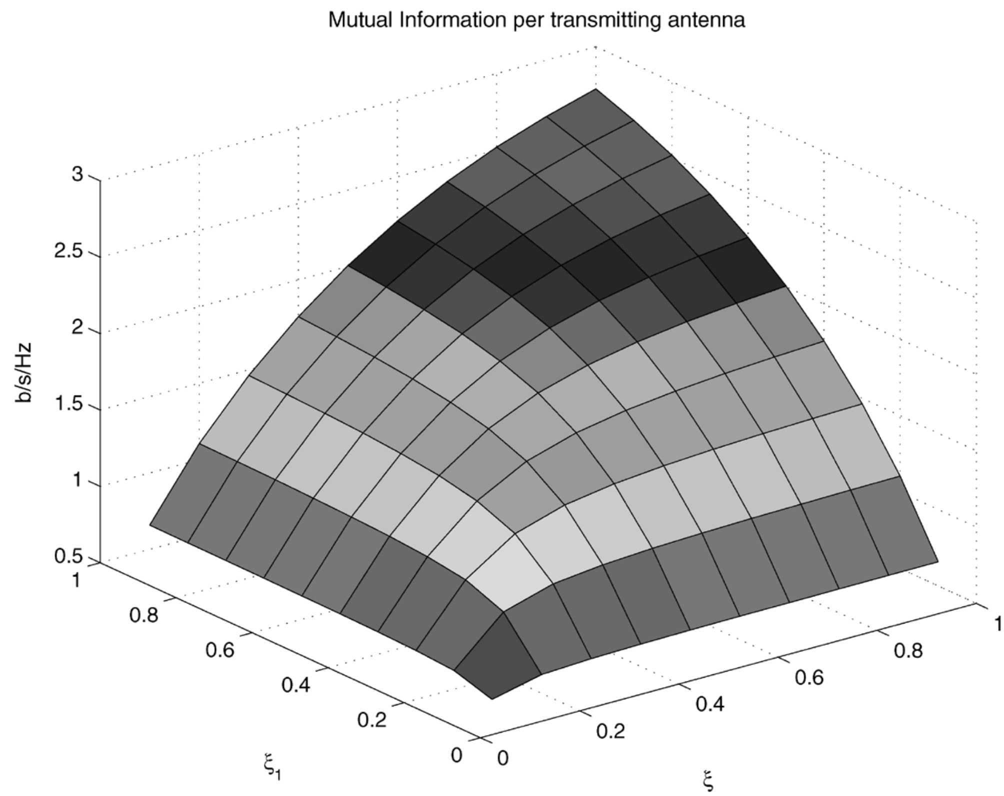

Fig. 6. Mutual Information per transmitting antenna versus ξ and ξ_1 for the double-directional model with equal power on Fourier directions.

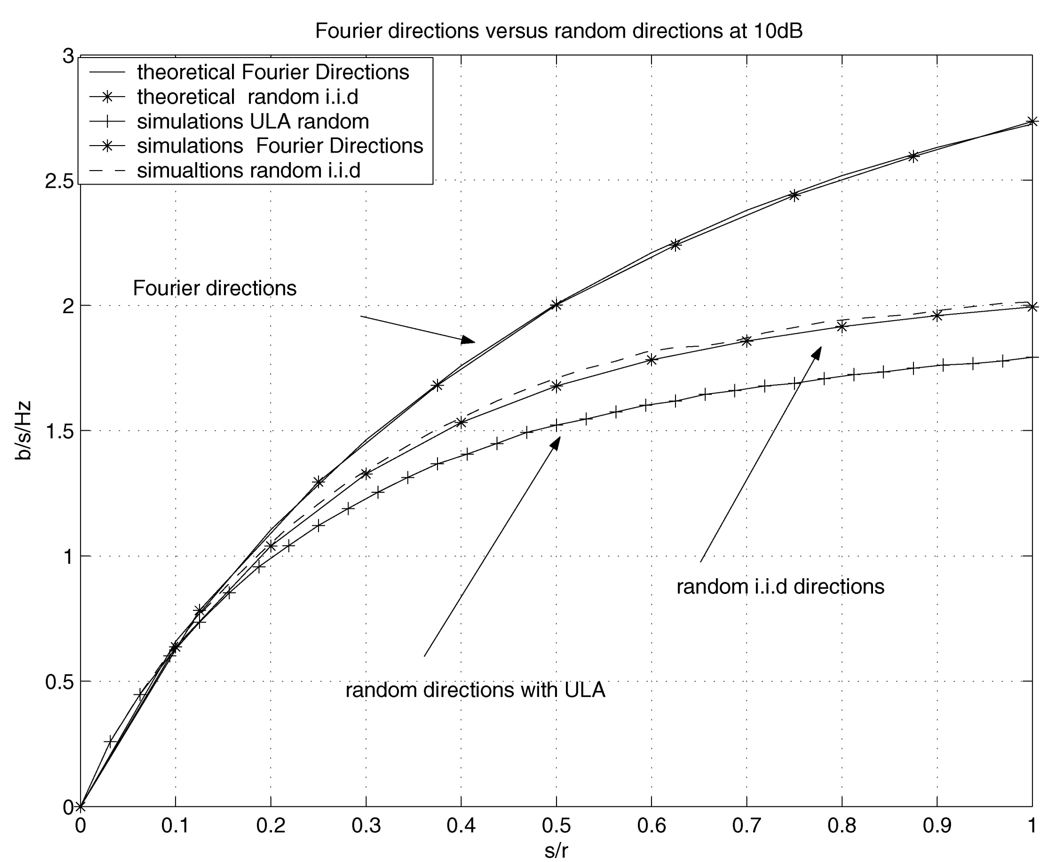

Fig. 7. Fourier versus random directions at 10 dB.

and, in particular, the eigenvalue distribution of

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

Theorem 2: With the DoA model, as $n_t\to\infty$ with $s_r=\xi n_t$

$$I^M{}_{\mathrm{doa}}(n_t,n_r,s_r,\rho) - n_t\mu_{\mathrm{doa}}(\xi,\gamma,\rho)$$

converges in distribution to a $N(0,\sigma^2{}_{\mathrm{doa}})$ random variable where

$$\mu_{\mathrm{doa}}(\xi,\gamma,\rho)=\int_0^\infty \ln(1+\rho\lambda)dF_{\mathrm{doa}}(\lambda) \quad (63)$$

$$m_{f_{\mathrm{doa}}}(z)=\int\frac{dF_{\mathrm{doa}}(\lambda)}{\lambda - z} \quad (64)$$

$$z=\frac{-1}{m_{f_{\mathrm{doa}}}(z)}+\xi\int\frac{x}{1+m_{f_{\mathrm{doa}}}(z)x}dS_{\mathrm{doa}}(x) \quad (65)$$

$$\sigma^2{}_{\mathrm{doa}} = -\frac{1}{4\pi^2}\int_{C_x}\int_{C_y}\frac{\ln(1+\rho x)\ln(1+\rho y)}{(m_{f_{\mathrm{doa}}}(x)-m_{f_{\mathrm{doa}}}(y))^2} m'_{f_{\mathrm{doa}}}(x)m'_{f_{\mathrm{doa}}}(y)dxdy. \quad (66)$$

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

f_{doa} is the limiting eigenvalue distribution of $\frac{1}{n_t}\boldsymbol{H}^H\boldsymbol{H}$ in the DoA-based model while S_{doa} is the limiting eigenvalue distribution of $\frac{1}{s_r}\boldsymbol{P}^{r\frac{1}{2}}\boldsymbol{\Phi}^H_{n_r\times s_r}\boldsymbol{\Phi}_{n_r\times s_r}\boldsymbol{P}^{r\frac{1}{2}}$. This result is based on [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 purposes.[9] 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 $F_{\mathrm{doa}}(\lambda)$? One can first of all notice that

$$\begin{aligned}\frac{d\mu_{\mathrm{doa}}}{d\rho} &= \int_0^\infty\frac{\lambda}{1+\rho\lambda}dF_{\mathrm{doa}}(\lambda)\\ &=\frac{1}{\rho}\int_0^\infty\frac{\rho\lambda+1-1}{1+\rho\lambda}dF_{\mathrm{doa}}(\lambda)\\ &=\frac{1}{\rho}-\frac{1}{\rho^2}m_{f_{\mathrm{doa}}}(-\frac{1}{\rho}).\end{aligned} \quad (67)$$

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

$$-\frac{1}{\rho}=\xi\int\frac{x}{1+x\rho(1-\rho(\frac{d\mu_{\mathrm{doa}}}{d\rho}))}dS_{\mathrm{doa}}(x) - \frac{1}{\rho(1-\rho(\frac{d\mu_{\mathrm{doa}}}{d\rho}))} \quad (68)$$

[9]Some results on the capacity of a MIMO multiuser network (where all the users have different angles of arrival) in the large system limit (high number of antennas) can be found in [72].

and

$$\mu_{\mathrm{doa}}(\rho)=\int_0^\rho(\frac{d\mu_{\mathrm{doa}}}{d\rho})d\rho. \quad (69)$$

In the high-SNR regime, the following result holds.

Proposition 6: In the high-SNR regime, the mean mutual information of the DoA-based model converges to

$$\min\left(n_t, s_r\int_{\lambda>0}dS_{\mathrm{doa}}(\lambda)\right)\ln(\rho). \quad (70)$$

Proof: Let $r=\rho\frac{d\mu_{\mathrm{doa}}}{d\rho}\leq 1$ ($n_t r$ denotes in fact the multiplexing gain).

According to (68), we have

$$-\frac{1}{\rho}=\frac{-1}{\rho(1-r)}+\xi\int\frac{x}{1+x\rho(1-r)}dS_{\mathrm{doa}}(x) \quad (71)$$

and at high SNR

$$-1=\frac{-1}{(1-r)}+\frac{\xi}{1-r}\int_{\lambda>0}dS_{\mathrm{doa}}(\lambda) \quad (72)$$

which yields

$$r=\begin{cases}\xi\int_{\lambda>0}dS_{\mathrm{doa}}(\lambda), & \text{if } \xi\int_{\lambda>0}dS_{\mathrm{doa}}(\lambda)\leq 1\\ 1, & \text{otherwise}\end{cases} \quad (73)$$

and proves the result. $\square$

2) Nonequal 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 $\frac{1}{s_r}\boldsymbol{P}^{r\frac{1}{2}}\boldsymbol{\Phi}^H_{n_r\times s_r}\boldsymbol{\Phi}_{n_r\times s_r}\boldsymbol{P}^{r\frac{1}{2}}$ has the following expression:

$$S_{\mathrm{doa}}(\lambda)=\sum_{i=1}^{K_r} l_i{}^r\delta(\lambda-\gamma P_i{}^r). \quad (74)$$

Proposition 7: In the nonequal power case with Fourier directions, $\mu_{\mathrm{doa}}(\xi,\gamma,\rho)$ and $\sigma^2{}_{\mathrm{doa}}(\xi,\gamma,\rho)$ are equal to

$$\mu_{\mathrm{doa}}(\xi,\gamma,\rho)=\xi\sum_{i=1}^{K_r} l_i{}^r\ln(1+\rho P_i{}^r\gamma\alpha_{\mathrm{doa}}) -(1-\alpha_{\mathrm{doa}})-\ln(\alpha_{\mathrm{doa}}) \quad (75)$$

and

$$\sigma^2{}_{\mathrm{doa}}(\xi,\gamma,\rho) = -\ln\left[1-\rho^2\xi\alpha_{\mathrm{doa}}{}^2\sum_{i=1}^{K_r} l_i{}^r\frac{(\gamma P_i{}^r)^2}{(1+\rho\gamma P_i{}^r\alpha_{\mathrm{doa}})^2}\right] \quad (76)$$

with

$$\sum_{i=1}^{K_r}\frac{l_i{}^r}{1+\rho\gamma P_i{}^r\alpha_{\mathrm{doa}}}=\frac{\alpha_{\mathrm{doa}}}{\xi}-\frac{1}{\xi}+1. \quad (77)$$

Proof: The proof is provided in Appendix C. For the mean μ_{doa}, the proof is an application of the general Proposition 4 in the case of interest and is provided in the Appendix. For the variance, results of [37] are used. $\square$

Note that α_{doa} is related to the Stieltjes transform $m_{f_{\mathrm{doa}}}$ of f_{doa} by

$$m_{f_{\mathrm{doa}}}(\frac{-1}{\rho})=\rho(1-\alpha_{\mathrm{doa}}). \quad (78)$$

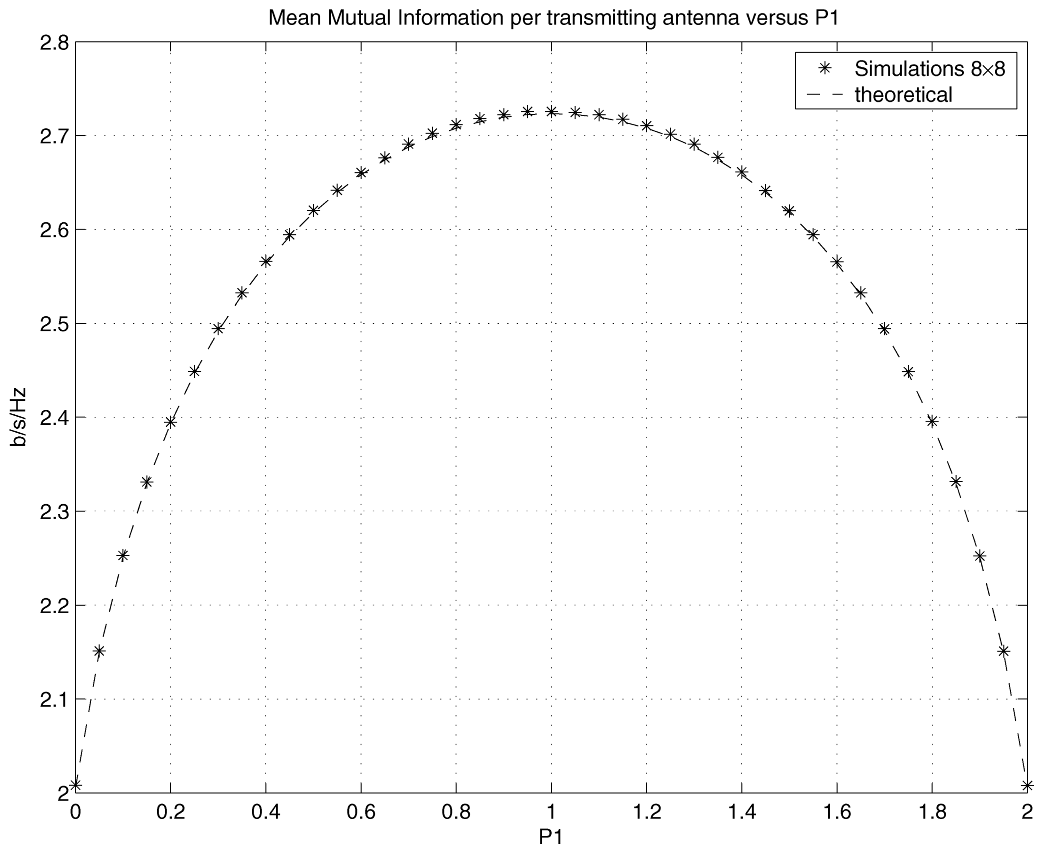

Fig. 8. Mean capacity per transmitting antenna versus ${P_1}^r$ at 10 dB for an 8×8 DoA-based model.

In Fig. 8, 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 $\xi = \frac{s_r}{n_t} = 1$)

$$\frac{1}{2}\left[\frac{1}{1+\rho {P_1}^r \alpha_{\mathrm{doa}}} + \frac{1}{1+\rho(2-{P_1}^r)\alpha_{\mathrm{doa}}}\right] = \alpha_{\mathrm{doa}} \quad (79)$$

with

$$\mu_{\mathrm{doa}} = -\ln(\alpha_{\mathrm{doa}}) + \frac{1}{2}\ln(1+\rho {P_1}^r \alpha_{\mathrm{doa}}) - (1-\alpha_{\mathrm{doa}}) + \frac{1}{2}\ln(1+2\rho\alpha_{\mathrm{doa}} - \frac{1}{2}\rho {P_1}^r \alpha_{\mathrm{doa}}) \quad (80)$$

and (81) at the bottom of the page.

In Fig. 8, 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. Note that the close match pertains also for the variance. In terms of outage mutual information, the equal-power case is also the one which maximizes that criteria (see Fig. 9).

3) Remarks on the Directions of Departure-Based Model: If only the directions of departure (with their respective power) and channel energy are known, the previous methodology yields the following DoD-based model:

$$\boldsymbol{H} = \frac{1}{\sqrt{s_t}} \boldsymbol{\Theta}_{n_r \times s_t} \boldsymbol{P}^{t\frac{1}{2}} \boldsymbol{\Psi}_{s_t \times n_t}. \quad (82)$$

$\boldsymbol{\Psi}_{s_t \times n_t}$ is an $s_t \times n_t$ 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 $\boldsymbol{P}^t$. $\boldsymbol{\Theta}_{n_r \times s_t}$ is an $n_r \times s_t$ i.i.d. zero-mean Gaussian matrix which represents the scattering environment between the receiving antennas and the scatterers.

In order to derive 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,
- exchanges n_t, s_r, P^r with n_r, s_t, and P^t
- replaces the SNR ρ by $\frac{n_r}{n_t}\rho$.

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

$$I^M{}_{\mathrm{dod}}(n_t, n_r, s_t, \rho) = I^M{}_{\mathrm{doa}}(n_r, n_t, s_r, \frac{n_r}{n_t}\rho). \quad (83)$$

Remark: The preceding expression shows that in the case where $n_t = n_r$, mutual information compliance **is not a good**

$$\sigma^2{}_{\mathrm{doa}} =, -\ln\left[1 - \frac{\rho^2\alpha_{\mathrm{doa}}^2}{2}\left(\frac{({P_1}^r)^2}{(1+\rho({P_1}^r\alpha_{\mathrm{doa}})^2} + \frac{(2-{P_1}^r)^2}{(1+\rho(2-{P_1}^r)\alpha_{\mathrm{doa}})^2}\right)\right]. \quad (81)$$

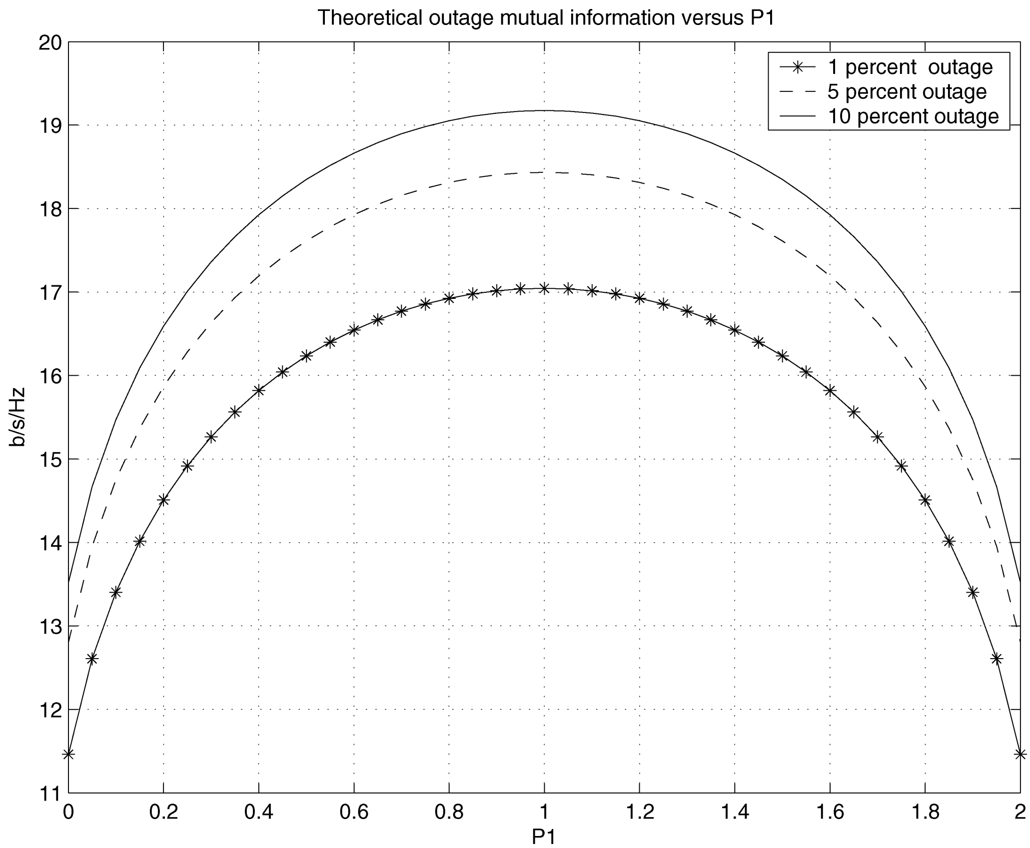

Fig. 9. Outage mutual information versus ${P_1}^r$ at 10 dB for an 8×8 DoA based model.

metric for model validation as the DoA- and DoD-based models give the same mutual information values. This point explains in particular why so many models comply with measurements in the literature and some discussions can be found in [73].

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 directions of arrival and respective power profile, directions of departure and respective power profile, delay, and Doppler effect. As a starting point, the modeler assumes that the positions of the transmitter and receiver change in time. However, the scattering environment (the buildings, trees, etc.) does not change and remains in the same position during the transmission. Let $\boldsymbol{v}_t$ and $\boldsymbol{v}_r$ be respectively the vector speed of the transmitter and the receiver with respect to a terrestrial reference (see Fig. 10). Let $\boldsymbol{s}_{ij}^t$ 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

$$\begin{aligned}\boldsymbol{s}_{ij}^t(t) &= \boldsymbol{s}_0 e^{j({\boldsymbol{k}_{ij}^t}^T(\boldsymbol{v}_t t + \boldsymbol{d^t}_{ij}) + 2\pi f_c t)} \\ &= \boldsymbol{s}_0 e^{j2\pi(\frac{f_c \boldsymbol{u}_{ij}^t \boldsymbol{v}_t}{c} t + f_c t)} e^{j\psi_{ij}}. \end{aligned} \tag{84}$$

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

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

$$\boldsymbol{s}_{ij}^r(t) = \boldsymbol{s}_0 e^{j2\pi(\frac{f_c \boldsymbol{v}_r \boldsymbol{u}_{ij}^r}{c} t + f_c t)} e^{j\phi_{ij}}. \tag{85}$$

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

- speed $\boldsymbol{v}_r$,
- speed $\boldsymbol{v}_t$,
- the angle of departure from the transmitting antenna to the scatterers ψ_{ij} and power P_j^t,
- the angle of arrival from the scatterers to the receiving antenna ϕ_{ij} and power P_j^r.

The modeler has however no knowledge of what happens in between except for 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:

$$\boldsymbol{D}_{s_r \times s_t}(f) = \begin{pmatrix} e^{-j2\pi f \tau_{1,1}} & \dots & e^{-j2\pi f \tau_{1,s_t}} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi f \tau_{s_r,1}} & \dots & e^{-j2\pi f \tau_{s_r,s_t}} \end{pmatrix}. \tag{86}$$

The modeler also supposes as a given state of knowledge the fact that matrix $\boldsymbol{H}$ has a certain energy. Based on this state

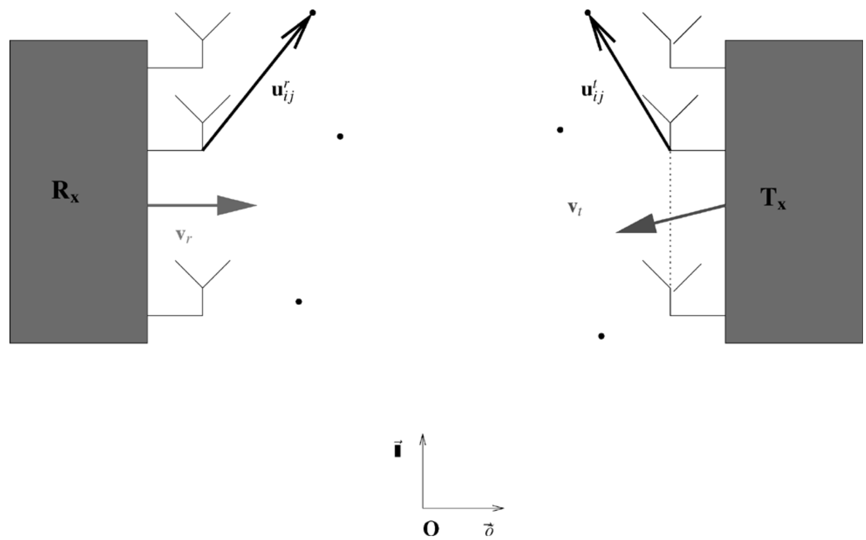

Fig. 10. Moving antennas.

of knowledge, the modeler wants to model the $s_r \times s_t$ matrix $\boldsymbol{\Theta}_{s_r \times s_t}$ in the following representation:

$$\begin{aligned} \boldsymbol{H}(f,t) = &\boldsymbol{\Phi}_{n_r \times s_r}(t)\boldsymbol{P}^{r\frac{1}{2}} \\ &\frac{1}{\sqrt{s_r s_t}}\left(\boldsymbol{\Theta}_{s_r \times s_t} \odot \boldsymbol{D}_{s_r \times s_t}(f)\right)\boldsymbol{P}^{t\frac{1}{2}}\boldsymbol{\Psi}_{s_t \times n_t}(t) \end{aligned} \tag{87}$$

with

$$\boldsymbol{\Phi}_{n_r \times s_r}(t) = \begin{pmatrix} e^{j(\phi_{1,1}+2\pi\frac{f\boldsymbol{u}^r_{11}\boldsymbol{v}_r}{c}t)} & \dots & e^{j(\phi_{1,s_r}+2\pi\frac{f\boldsymbol{u}^r_{1s_r}\boldsymbol{v}_r}{c}t)} \\ \vdots & \ddots & \vdots \\ e^{j(\phi_{r,1}+2\pi\frac{f\boldsymbol{u}_{n_r1}{}^r\boldsymbol{v}_r}{c}t)} & \dots & e^{j(\phi_{n_r,s_r}+2\pi\frac{f\boldsymbol{u}^r_{n_rs_r}\boldsymbol{v}_r}{c}t)} \end{pmatrix}$$

and

$$\boldsymbol{\Psi}_{s_t \times n_t}(t) = \begin{pmatrix} e^{j(\psi_{1,1}+2\pi\frac{f\boldsymbol{u}^t_{11}\boldsymbol{v}_t}{c}t)} & \dots & e^{j(\psi_{1,n_t}+2\pi\frac{f\boldsymbol{u}^t_{1n_t}\boldsymbol{v}_t}{c}t)} \\ \vdots & \ddots & \vdots \\ e^{j(\psi_{s_t,1}+2\pi\frac{f\boldsymbol{u}^t_{s_11}\boldsymbol{v}_t}{c}t)} & \dots & e^{j(\psi_{s_t,n_t}+2\pi\frac{f\boldsymbol{u}^t_{s_tn_t}\boldsymbol{v}_t}{c}t)} \end{pmatrix}.$$

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

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

Proposition 8: $\boldsymbol{\Theta}_{s_r \times s_t}$ i.i.d. zero-mean Gaussian with variance 1 is the solution of the consistency argument and maximizes entropy.[10]

Proof: We will not go into the details but only provide guidelines of the proof. First, note that if $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ 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 cannot exceed a certain limit v_{limit} (in the least favorable case, v_{limit} would be equal 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

$$\begin{aligned} \boldsymbol{v}_t &= v_t\cos(\alpha_t)\vec{\boldsymbol{\imath}} + v_t\sin(\alpha_t)\vec{\boldsymbol{\jmath}} \\ \boldsymbol{v}_r &= v_r\cos(\alpha_r)\vec{\boldsymbol{\imath}} + v_r\sin(\alpha_r)\vec{\boldsymbol{\jmath}} \\ {u^t}_{ij} &= \cos({\beta^t}_{ij})\vec{\boldsymbol{\imath}} + \sin({\beta^t}_{ij})\vec{\boldsymbol{\jmath}} \end{aligned}$$

and

$${u^r}_{ij} = \cos({\beta^r}_{ij})\vec{\boldsymbol{\imath}} + \sin({\beta^r}_{ij})\vec{\boldsymbol{\jmath}}$$

the modeler will attribute a uniform distribution over $[0; 2\pi]$ 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 narrow band (in terms of frequency selectivity) MIMO model proof, one can easily show that $\boldsymbol{\Theta}_{s_r \times s_t}$ i.i.d. Gaussian is the solution of the consistency argument and maximizes entropy.

Note that in the case $f = 0$, $\boldsymbol{v}_t = 0$, and $\boldsymbol{v}_r = 0$, the same model as the narrow-band 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 $\boldsymbol{D}_{s_r \times s_t}$, also known as the channel pattern mask [24]. 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 $\boldsymbol{\Phi}$ has a given

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

direction) and $u_{ij}^t = u_i^t$ (any line of matrix $\mathbf{\Psi}$ has a given direction). Therefore, the channel model simplifies to

$$\boldsymbol{H}(f,t) = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ e^{j2\pi\frac{d(n_r-1)\sin(\phi_1)}{\lambda}} & \cdots & e^{j2\pi\frac{d(n_r-1)\sin(\phi_{s_r})}{\lambda}} \end{pmatrix} \frac{1}{\sqrt{s_r s_t}}\left(\mathbf{\Theta}_{s_r\times s_t}\odot \boldsymbol{D}_{s_r\times s_t}(f,t)\right) \begin{pmatrix} 1 & \cdots & e^{j2\pi\frac{d(n_t-1)\sin(\psi_1)}{\lambda}} \\ \vdots & \ddots & \vdots \\ 1 & \cdots & e^{j2\pi\frac{d(n_t-1)\sin(\psi_{s_t})}{\lambda}} \end{pmatrix}. \tag{88}$$

In this case, the pattern mask $\boldsymbol{D}_{s_r\times s_t}(f,t)$ with elements $(d_{pk})_{p=1,\ldots,s_r}^{k=1,\ldots,s_t}$ has the following form:

$$d_{pk} = \sqrt{P_p^r}\sqrt{P_k^t}e^{-j2\pi f\tau_{p,k}}e^{j2\pi\frac{ft}{c}(\boldsymbol{u}_p\boldsymbol{v}_r+\boldsymbol{u}_k^t\boldsymbol{v}_t)}. \tag{89}$$

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 $\mathbf{\Theta}_{s_r\times s_t}\odot\boldsymbol{D}_{s_r\times s_t}$ 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 $\mathbf{\Theta}_{s_r\times s_t}\odot\boldsymbol{D}_{s_r\times s_t}(f,t)$.

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

$$\boldsymbol{H}(f,t) = \sum_{l=1}^{s_t}\sum_{k=1}^{s_r}\rho_{k,l}e^{j2\pi\xi_{k,l}t}e^{-j2\pi f\tau_{k,l}}. \tag{90}$$

where $\rho_{k,l}$ ($\rho_{k,l} = \frac{1}{\sqrt{s_r s_t}}\theta_{k,l}\sqrt{P_k^r}\sqrt{P_l^t}e^{j(\phi_k+\psi_l)}$) are independent Gaussian variable with zero mean and variance $\mathbb{E}(|\rho_{k,l}|^2) = \frac{1}{s_r s_t}P_k^r P_l^t$, $\xi_{k,l} = \frac{f}{c}(\boldsymbol{u}_k^r\boldsymbol{v}_r + \boldsymbol{u}_l^t\boldsymbol{v}_t)$ represents the Doppler effect, and $\tau_{k,l}$ are the delays. This previous result is a generalization of the single-input single-output (SISO) 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 $\boldsymbol{Q} = \boldsymbol{I}$ is given by

$$I^M(n_t,f,t) = \frac{1}{n_t}\log_2\det\left(\boldsymbol{I}_{n_t} + \frac{\rho}{n_t}\boldsymbol{H}^H(f,t)\boldsymbol{H}(f,t)\right).$$

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 noncoherent 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 [74], Zheng and Tse [27] for block-fading channels, and Liang and Veeravalli for more advanced time-varying models [75]. An even more difficult problem concerns the practical schemes for achieving the noncoherent mutual information. Recently, in [76], 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 therefore circumvents the need of using blind or semiblind 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 $\boldsymbol{H}(f)$ are independent of f. Since $\boldsymbol{H}(f)$ 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

$$\mathbb{E}_{\mathbf{\Theta}}(h_{ij}) = \frac{1}{\sqrt{ss_1}}\sum_{k=1}^{s_t}\sum_{p=1}^{s_r}\mathbb{E}(\theta_{pk})\sqrt{P_k}^t\sqrt{P_p}^r e^{j2\pi f\tau_{pk}}e^{j\psi_{kj}}e^{j\phi_{ip}} = 0 \tag{91}$$

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

- **Covariance:** Let us derive $\mathrm{Cov} = \mathbb{E}_{\mathbf{\Theta}}(h_{ij}(f)h_{mn}^*(f))$

$$\mathrm{Cov} = \frac{1}{s_r s_t}\sum_{k=1}^{s_t}\sum_{p=1}^{s_r}\sum_{q=1}^{s_t}\sum_{l=1}^{s_r}\mathbb{E}(\theta_{pk}\theta_{ql}^*) e^{j2\pi f(\tau_{pk}-\tau_{ql})}\sqrt{P_k^t}\sqrt{P_q^t}\sqrt{P_p^r}\sqrt{P_l^r} e^{j2\pi(\psi_{kj}-\psi_{qn})}e^{j2\pi(\phi_{ip}-\phi_{ml})}. \tag{92}$$

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

$$\mathrm{Cov} = \frac{1}{s_r s_t}\sum_{k=1}^{s_t}\sum_{p=1}^{s_r}P_k^t P_p^r e^{j2\pi(\psi_{kj}-\psi_{kn})}e^{j2\pi(\phi_{ip}-\phi_{mlp})} \tag{93}$$

which is independent of f.

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

$$I^M = \frac{W}{n_t}\mathbb{E}\left[\log_2\det\left(\boldsymbol{I}_{n_t} + \frac{\rho}{n_t}\boldsymbol{H}^H(0)\boldsymbol{H}(0)\right)\right]. \tag{94}$$

One can observe that frequency selectivity does not affect the ergodic mutual information per transmitting antenna. Similar results have been reported in [24], [66]. In the wide-band case with no Doppler effect, the ergodic mutual information is the same as in the narrowband case and all the results of Section IV remain valid.

VI. Other Models in View of the Maximum Entropy Framework

A. Müller's Model

In [66], Müller develops a channel model based on the product of two random matrices

$$\boldsymbol{H} = \mathbf{\Phi}\boldsymbol{A}\mathbf{\Theta} \tag{95}$$

where $\mathbf{\Phi}$ and $\mathbf{\Theta}$ are two random matrices with zero mean unit variance i.i.d. entries and $\boldsymbol{A}$ 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).[11] $\mathbf{\Phi}$ 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 [66] 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 one of three constructions.

- A DoA-based model with random directions, i.e., matrix $\mathbf{\Phi}$ with different powers (represented by matrix $\boldsymbol{A}$) 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 multifold 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,
  - 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 $\boldsymbol{A}$) 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.
- A 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 [77].

B. Virtual Representation Model

In [78], Sayeed proposes a virtual representation of the channel. The model is the following:

$$\boldsymbol{H} = \boldsymbol{A}_{n_r}\boldsymbol{S}\boldsymbol{A}_{n_t}{}^{H}. \tag{96}$$

Matrices $\boldsymbol{A}_{n_r}$ and $\boldsymbol{A}_{n_t}$ are discrete Fourier matrices and $\boldsymbol{S}$ is an $n_r \times n_t$ matrix which represents the contribution of each of the fixed DoAs and DoDs. 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. The virtual representation is quite appealing in terms of simplicity and analysis. In this case, also, we can revisit the virtual representation in light of our framework. We can show that in each case, the virtual representation answers a specific question based on a given assumption.

- Suppose matrix $\boldsymbol{S}$ has i.i.d. zero-mean Gaussian entries then the virtual representation 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 $\boldsymbol{S}$ has a certain correlation structure, then the virtual representation 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, the virtual representation has a simple interpretation in the maximum-entropy framework: it considers a ULA geometry with Fourier directions. Although it may seem strange to restrict oneself to this case, we do have an explanation for this fact. In the paper [24], the authors were 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 a very small effect on 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 simplicity of 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 [26], Chuah *et al.* study the following Kronecker[12] model

$$\boldsymbol{H} = \boldsymbol{R}_{n_r}{}^{\frac{1}{2}}\mathbf{\Theta}\boldsymbol{R}_{n_t}{}^{\frac{1}{2}}. \tag{97}$$

Here, $\mathbf{\Theta}$ is an $n_r \times n_t$ i.i.d. zero-mean Gaussian matrix, $\boldsymbol{R}_{n_r}{}^{\frac{1}{2}}$ is an $n_r \times n_r$ receiving correlation matrix, while $\boldsymbol{R}_{n_t}{}^{\frac{1}{2}}$ is an $n_t \times n_t$ transmitting correlation matrix. The correlation is supposed to decrease sufficiently fast for $\boldsymbol{R}_{n_r}$ and $\boldsymbol{R}_{n_t}$ to have a Toeplitz band structure. Using a software tool (Wireless System Engineering [81]), they demonstrate the validity of the model. Quite remarkably, although designed to take into account the receiving and the transmitting correlation, the Kronecker model falls within the double directional framework. As shown in [24], since $\boldsymbol{R}_{n_r}$ and $\boldsymbol{R}_{n_t}$ are band Toeplitz then these matrices are asymptotically diagonalized in a Fourier basis $\boldsymbol{F}$. Therefore, $\boldsymbol{H}$ can be rewritten as

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

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

[12]The model is called a Kronecker model because $\mathbb{E}(\mathrm{vec}(\boldsymbol{H})^H\mathrm{vec}(\boldsymbol{H})) = \boldsymbol{R}_{n_r} \bigotimes \boldsymbol{R}_{n_t}$ is a Kronecker product. The justification of this approach relies on the fact that only the immediate surroundings of the antenna array impose 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 [79], [80].

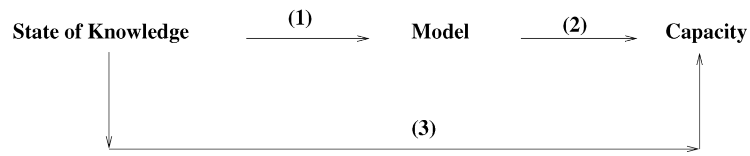

Fig. 11. Channel modeling approach and derivation of capacity.

$$= \boldsymbol{F}_{n_r} \left(\Lambda_{n_r}^{\frac{1}{2}} \boldsymbol{F}_{n_r}{}^{H} \boldsymbol{\Theta} \boldsymbol{F}_{n_t} \Lambda_{n_t}{}^{\frac{1}{2}} \right) \boldsymbol{F}_{n_t}{}^{H}$$
$$= \boldsymbol{F}_{n_r} \left(\boldsymbol{\Theta_1} \odot \boldsymbol{D}_{n_r \times n_t} \right) \boldsymbol{F}_{n_t}{}^{H} \tag{98}$$

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

$$\boldsymbol{D}_{s_r \times s_t} = \begin{pmatrix} \lambda_{1,n_t}^{\frac{1}{2}} \lambda_{1,n_r}^{\frac{1}{2}} & \dots & \lambda_{n_t,n_t}^{\frac{1}{2}} \lambda_{1,n_r}^{\frac{1}{2}} \\ \vdots & \ddots & \vdots \\ \lambda_{1,n_t}^{\frac{1}{2}} \lambda_{n_r,n_r}^{\frac{1}{2}} & \dots & \lambda_{n_t,n_t}^{\frac{1}{2}} \lambda_{n_r,n_r}^{\frac{1}{2}} \end{pmatrix}. \tag{99}$$

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 DoAs are on Fourier directions and have respective power λ_{i,n_r}, the DoDs are on Fourier directions and have respective power λ_{i,n_t}, each path has zero mean and a certain variance.

D. "Keyhole" Model

In [82], Gesbert *et al.* show that low-correlation[13] 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 between the transmitting and receiving antennas). In [84], they propose the following model for a rank one channel:

$$\boldsymbol{H} = \boldsymbol{R}_{n_r}{}^{\frac{1}{2}} \boldsymbol{g}_r \boldsymbol{g}_t{}^{H} \boldsymbol{R}_{n_t}{}^{\frac{1}{2}}. \tag{100}$$

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

$$\boldsymbol{H} = \frac{1}{\sqrt{s_r s_t}} \boldsymbol{\Phi}_{n_r \times s_r} \boldsymbol{\Theta}_{s_r \times s_t} \boldsymbol{\Psi}_{s_t \times n_t}. \tag{101}$$

The Keyhole model can be either

- a double direction model with $s_r = 1$ and
$$\boldsymbol{\Phi}_{n_r \times 1} = \boldsymbol{R}_{n_r}{}^{\frac{1}{2}} \boldsymbol{g}_r.$$
In this case, $\boldsymbol{g}_t{}^{H} \boldsymbol{R}_{n_t}{}^{\frac{1}{2}} = \boldsymbol{\Theta}_{1 \times s_t} \boldsymbol{\Psi}_{s_t \times n_t}$ where $\boldsymbol{\Theta}_{1 \times s_t}$ is zero mean i.i.d. Gaussian; or
- a double direction model with $s_t = 1$ and
$$\boldsymbol{\Psi}_{1 \times n_t} = \boldsymbol{g}_t{}^{H} \boldsymbol{R}_{n_t}{}^{\frac{1}{2}}.$$
In this case, $\boldsymbol{R}_{n_r}{}^{\frac{1}{2}} \boldsymbol{g}_r = \Phi_{n_r \times s_r} \boldsymbol{\Theta}_{s_r \times 1}$ where $\boldsymbol{\Theta}_{s_r \times 1}$ 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 rate. But whenever we are misrepresenting the channel with our state of knowledge, the formula

$$C(\boldsymbol{I}) = \log_2 \det \left(\boldsymbol{I}_{n_t} + \frac{\rho}{n_t} \boldsymbol{H}^H \boldsymbol{H} \right) \tag{102}$$

will mis-estimate 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 (102). The problem is formulated in Fig. 11.

- Transition 1: the modeler creates a model maximizing entropy.
- Transition 2: the modeler mis-estimates the real achievable rate because even though the created model is the best possible, based on the state of knowledge, it derives the mutual information of the channel based on the assumption that the model is reality.
- Transition 3: a new measure of the 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.[14]

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"

$$I^M = \min_{\{P(\boldsymbol{H})/\boldsymbol{H} \in \Delta\}} \{\max_{\boldsymbol{Q}} \int C(\boldsymbol{Q}) P(\boldsymbol{H}) d\boldsymbol{H}\}. \tag{103}$$

Δ is the infinite set of matrices $\boldsymbol{H}$ 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 modeling approach? No, as long as we understand the meaning of transition 2 in Fig. 11. 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)

[13]"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 [83].

[14]Quantifying the gap between transitions 1 + 2 and 3 is still an open issue.

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 the sake of simplicity, that each frequency path of the channel has 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 $\boldsymbol{H} = \boldsymbol{\Phi\Theta}$ ($\boldsymbol{\Theta}$ is i.i.d. Gaussian), then using Jensen's inequality

$$\mathbb{E}_{\boldsymbol{\Phi}}\left(\log\det(\boldsymbol{I}_{n_t} + \frac{\rho}{n_t}(\boldsymbol{\Theta}^H\boldsymbol{\Phi}^H\boldsymbol{\Phi\Theta})\right) \leqslant \log\det(\boldsymbol{I}_{n_t} + \frac{\rho}{n_t}\mathbb{E}_{\boldsymbol{\Phi}}(\boldsymbol{\Theta}^H\boldsymbol{\Phi}^H\boldsymbol{\Phi\Theta})). \quad (104)$$

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

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

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 constrains 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, and variance will reduce the mutual information compared with the case where only the same mean and the same variance are taken into account. In fact, if one is interested only in particular functions of the model, then he should construct a model which is consistent with those functionals and not with respect to axiom 1. 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 modeling ?[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 at its best with any case having more constraints (but at least includes the same prior constraints). The channel modeling method is summarized hereafter.

[15]This question has to be taken in light of the talk "Where do we stand on maximum entropy?" made by E. T. Jaynes in 1978 at MIT [85].

- $H(p) = \int -p\log p + \sum_i \lambda_i\{\text{prior information}\}_i$
- Consistency argument.

The consistency argument is extremely important as it shows that two channel modeling 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 fewer things are assumed as *a priori* information the greater are the chances that the model will comply with any mismatched representation. Finally, note that recent campaign measures at 2.1 and 5.2 GHz in [73] have shown that Maxent Fourier models are mutual information compliant.

Appendix A
Preliminaries

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

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

- $\boldsymbol{H}_{s\times t} = (h_{ij})$ 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 $\lim_{t\to\infty} \frac{s}{t} = c$.
- $\boldsymbol{A}_{s\times s}$ is an $s \times s$ nonrandom Hermitian nonnegative definite matrix, with empirical eigenvalue distribution that converges in distribution almost surely to a fixed G, and the sequence of spectral norms $\Gamma\boldsymbol{A}_{s\times s}\Gamma$ 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 $\frac{s}{t} \to c$

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

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}, the solution of the implicit equation
$$z = -\frac{1}{m(z)} + c\int \frac{\tau}{1+m(z)\tau}dG(\tau)$$
through its Stieltjes 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
$$\sigma^2 = -\frac{1}{4\pi^2}\int_{C_x}\int_{C_y} \frac{f(x)f(y)}{(m(x)-m(y))^2}m'(x)m'(y)dxdy$$
and C_x and C_y are any closed positive contours that enclose the support of F.

APPENDIX B
PROOF OF PROPOSITION 5

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 $\boldsymbol{P}^r = \boldsymbol{I}_{s_r}$ and $\boldsymbol{P}^t = \boldsymbol{I}_{s_t}$.

Let us maximize μ_{double} with respect to ${P_j}^t$ with the constraints

$$\sum_{i=1}^{K_t} {l_i}^t {P_i}^t = 1 \quad \text{and} \quad \sum_{i=1}^{K_r} {l_i}^r {P_i}^r = 1.$$

This corresponds to maximizing the following function:

$$\begin{aligned} L =& \mu_{\text{double}} + \lambda_1 \left(\sum_{i=1}^{K_t} {l_i}^t {P_i}^t - 1 \right) + \lambda_2 \left(\sum_{i=1}^{K_r} {l_i}^r {P_i}^r - 1 \right) \\ =& \xi_1 \sum_{i=1}^{K_t} {l_i}^t \ln(1 + \frac{\rho {P_i}^t \alpha_{\text{dod}}}{\xi_1}) + \xi \sum_{i=1}^{K_r} {l_i}^r \ln(1 + \rho {P_i}^r \gamma \alpha_{\text{doa}}) \\ & - \rho \alpha_{\text{doa}} \alpha_{\text{dod}} + \lambda_1 \left(\sum_{i=1}^{K_t} {l_i}^t {P_i}^t - 1 \right) \\ & + \lambda_2 \left(\sum_{i=1}^{K_r} {l_i}^r {P_i}^r - 1 \right). \end{aligned} \tag{105}$$

Therefore,

$$\begin{aligned} \frac{dL}{d{P_j}^t} =& \frac{d\alpha_{\text{dod}}}{d{P_j}^t} \rho \sum_{i=1}^{K_t} \frac{{l_i}^t {P_i}^t}{1 + \frac{\rho {P_i}^t \alpha_{\text{dod}}}{\xi_1}} \\ & + \frac{d\alpha_{\text{doa}}}{d{P_j}^t} \rho \xi \sum_{i=1}^{K_r} {l_i}^r \frac{{P_i}^r \gamma}{1 + \rho \gamma {P_i}^r \alpha_{\text{doa}}} - \rho \frac{d\alpha_{\text{dod}}}{d{P_j}^t} \alpha_{\text{doa}} \\ & - \rho \frac{d\alpha_{\text{doa}}}{d{P_j}^t} \alpha_{\text{dod}} + \frac{\xi_1 {l_j}^t \frac{\rho \alpha_{\text{dod}}}{\xi_1}}{1 + \frac{\rho {P_j}^t \alpha_{\text{doa}}}{\xi_1}} + \lambda_1 {l_j}^t. \end{aligned} \tag{106}$$

Since

$$\alpha_{\text{doa}} = \sum_{i=1}^{K_t} \frac{{l_i}^t {P_i}^t}{1 + \frac{\rho {P_i}^t \alpha_{\text{dod}}}{\xi_1}} \tag{107}$$

and

$$\alpha_{\text{dod}} = \xi \sum_{i=1}^{K_r} {l_i}^r \frac{{P_i}^r \gamma}{1 + \rho \gamma {P_i}^r \alpha_{\text{doa}}} \tag{108}$$

then

$$\begin{aligned} \frac{dL}{d{P_j}^t} =& \rho \frac{d\alpha_{\text{dod}}}{d{P_j}^t} \alpha_{\text{doa}} + \rho \frac{d\alpha_{\text{doa}}}{d{P_j}^t} \alpha_{\text{dod}} - \rho \frac{d\alpha_{\text{dod}}}{d{P_j}^t} \alpha_{\text{doa}} \\ & - \rho \frac{d\alpha_{\text{doa}}}{d{P_j}^t} \alpha_{\text{dod}} + \frac{\xi_1 {l_j}^t \frac{\rho \alpha_{\text{dod}}}{\xi_1}}{1 + \frac{\rho {P_j}^t \alpha_{\text{doa}}}{\xi_1}} + \lambda_1 {l_j}^t \\ =& \frac{{l_j}^t \rho \alpha_{\text{dod}}}{1 + \frac{\rho {P_j}^t \alpha_{\text{doa}}}{\xi_1}} + \lambda_1 {l_j}^t. \end{aligned} \tag{109}$$

Therefore,

$$\frac{{l_j}^t \rho \alpha_{\text{dod}}}{1 + \frac{\rho {P_j}^t \alpha_{\text{doa}}}{\xi_1}} + \lambda_1 {l_j}^t = 0 \tag{110}$$

and

$$\frac{\rho \alpha_{\text{dod}}}{1 + \frac{\rho {P_j}^t \alpha_{\text{doa}}}{\xi_1}} = -\lambda_1. \tag{111}$$

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 ${P_j}^r$.

APPENDIX C
PROOF OF PROPOSITION 7

Let us first derive μ_{doa}. In the DoA-based model, one can apply straightforwardly Proposition 4 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,

$$\begin{aligned} \mu_{\text{doa}} = \ln(1 + \rho \alpha_{\text{dod}}) + \xi \sum_{i=1}^{K_r} {l_i}^r \ln(1 + \rho {P_i}^r \gamma \alpha_{\text{doa}}) \\ - \rho \alpha_{\text{doa}} \alpha_{\text{dod}} \end{aligned} \tag{112}$$

with

$$\alpha_{\text{doa}} = \frac{1}{1 + \rho \alpha_{\text{dod}}} \tag{113}$$

and

$$\alpha_{\text{dod}} = \xi \sum_{i=1}^{K_r} \frac{{l_i}^r {P_i}^r \gamma}{1 + \rho \gamma {P_i}^r \alpha_{\text{doa}}}. \tag{114}$$

Notice that

$$\alpha_{\text{doa}}(1 + \rho \alpha_{\text{dod}}) = 1 \tag{115}$$

and therefore,

$$\alpha_{\text{dod}} = \frac{1}{\rho} \left(\frac{1}{\alpha_{\text{doa}}} - 1 \right) \tag{116}$$

which yields

$$\rho \alpha_{\text{doa}} \alpha_{\text{dod}} = 1 - \alpha_{\text{doa}}. \tag{117}$$

We can therefore rewrite μ_{doa} as

$$\begin{aligned} \mu_{\text{doa}} = \ln\left(1 + \rho \frac{1}{\rho} \left(\frac{1}{\alpha_{\text{doa}}} - 1 \right)\right) - (1 - \alpha_{\text{doa}}) \\ + \xi \sum_{i=1}^{K_r} {l_i}^r \ln(1 + \rho {P_i}^r \gamma \alpha_{\text{doa}}) \end{aligned} \tag{118}$$

which yields

$$\begin{aligned} \mu_{\text{doa}} = -\ln(\alpha_{\text{doa}}) + \xi \sum_{i=1}^{K_r} {l_i}^r \ln(1 + \rho {P_i}^r \gamma \alpha_{\text{doa}}) \\ - (1 - \alpha_{\text{doa}}). \end{aligned} \tag{119}$$

We also have using (116)

$$\frac{1}{\rho}\left(\frac{1}{\alpha_{\rm doa}}-1\right)=\xi\sum_{i=1}^{K_r}\frac{l_i{}^r P_i{}^r\gamma}{1+\rho\gamma P_i{}^r\alpha_{\rm doa}} \tag{120}$$

which can be simplified to

$$\sum_{i=1}^{K_r}\frac{l_i{}^r}{1+\rho\gamma P_i{}^r\alpha_{\rm doa}}=\frac{\alpha_{\rm doa}}{\xi}+1-\frac{1}{\xi}. \tag{121}$$

Let us now derive $\sigma^2{}_{\rm doa}$:

To this end, we will apply Theorem 2. Since

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

we have

$$z=\frac{-1}{m(z)}+\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{m(z)+\frac{1}{\gamma P_i{}^r}}. \tag{122}$$

The asymptotic variance is therefore equal to

$$\sigma^2{}_{\rm doa}=\frac{-1}{4\pi^2}\int_{C_{m_x}}\int_{C_{m_y}}\frac{\ln(1+\rho x(m_x))\ln(1+\rho y(m_y)))}{(m_x-m_y)^2}dm_x dm_y. \tag{123}$$

For fixed m_y, let us calculate over the contour C_{m_x} the following expression:

$$\begin{aligned}\frac{1}{j2\pi}&\int\frac{\ln(1-\frac{\rho}{m}+\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{m+\frac{1}{\gamma P_i{}^r}})}{(m-m_y)^2}dm\\&=\frac{1}{j2\pi}\int\frac{\frac{\rho}{m^2}-\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{(m+\frac{1}{\gamma P_i{}^r})^2}}{1-\frac{\rho}{m}+\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{m+\frac{1}{\gamma P_i{}^r}}}\frac{1}{m-m_y}dm\\&=\frac{1}{j2\pi}\int\frac{g_0(m)}{h_0(m)}\frac{1}{m-m_y}dm\end{aligned} \tag{124}$$

$$\begin{aligned}g_0(m)=&\prod_{i=1}^{K_r}\rho(m+\frac{1}{\gamma P_i{}^r})^2-\rho\xi m^2\\&\prod_{i=1}^{K_r}(m+\frac{1}{\gamma P_i{}^r})^2\sum_{i=1}^{K_r}\frac{l_i{}^r}{(m+\frac{1}{\gamma P_i{}^r})^2}\end{aligned} \tag{125}$$

$$h_0(m)=m\prod_{i=1}^{K_r}(m+\frac{1}{\gamma P_i{}^r})P(m) \tag{126}$$

where, for notational simplicity, we denote by $P(m)$ the following polynomial of degree K_r+1:

$$\begin{aligned}P(m)&=\prod_{i=1}^{K_r}\left(m+\frac{1}{\gamma P_i{}^r}\right)\left(m-\rho+\rho m\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{m+\frac{1}{\gamma P_i{}^r}}\right)\\&=\left(m-m\left(\frac{-1}{\rho}\right)\right)\prod_{i=1}^{K_r}(m-m^i).\end{aligned} \tag{127}$$

m^i and $m(\frac{-1}{\rho})$ are the roots of P (the fact that $m(\frac{-1}{\rho})$ is a root stems form (122)).

Therefore, if we define:

$$h_1(m)=\prod_{i=1}^{K_r}\left(m+\frac{1}{\gamma P_i{}^r}\right)\prod_{i=1}^{K_r}(m-m^i)(m-m_y) \tag{128}$$

then

$$\begin{aligned}\frac{1}{j2\pi}&\int\frac{\ln(1-\frac{\rho}{m}+\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{m+\frac{1}{\gamma P_i{}^r}})}{(m-m_y)^2}dm\\&=\frac{1}{j2\pi}\int-\frac{1}{m(\frac{-1}{\rho})}\frac{g_0(m)}{h_1(m)}\left(\frac{1}{m}-\frac{1}{m-m(-\frac{1}{\rho})}\right)dm\\&=\frac{1}{m_y}-\frac{1}{m_y-m(-\frac{1}{\rho})}.\end{aligned} \tag{129}$$

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.

Notice that

$$\begin{aligned}&P'\left(m\left(\frac{-1}{\rho}\right)\right)\\&=\prod_{i=1}^{K_r}\left(m\left(-\frac{1}{\rho}\right)-m^i\right)\\&=\prod_{i=1}^{K_r}(m(-\frac{1}{\rho})+\frac{1}{\gamma P_i{}^r})\\&\left(1+\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{m(-\frac{1}{\rho})+\frac{1}{\gamma P_i{}^r}}-\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{(m(-\frac{1}{\rho})+\frac{1}{\gamma P_i{}^r})^2}\right)\\&=\prod_{i=1}^{K_r}\left(m\left(-\frac{1}{\rho}\right)+\frac{1}{\gamma P_i{}^r}\right)\\&\left(\frac{\rho}{m(-\frac{1}{\rho})}-\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{(m(-\frac{1}{\rho})+\frac{1}{\gamma P_i{}^r})^2}\right).\end{aligned} \tag{130}$$

The last equation comes from the fact that (see (122))

$$m\left(-\frac{1}{\rho}\right)\left(1+\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{m(-\frac{1}{\rho})+\frac{1}{\gamma P_i{}^r}}\right)=\rho. \tag{131}$$

Therefore,

$$\begin{aligned}\sigma^2{}_{\rm doa}=&\frac{1}{j2\pi}\int\ln\left(1-\frac{\rho}{m}+\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{m+\frac{1}{\gamma P_i{}^r}}\right)\\&\left(\frac{1}{m}-\frac{1}{m-m(-\frac{1}{\rho})}\right)dm\\=&\frac{1}{j2\pi}\int\ln\left(\frac{(m-m(\frac{-1}{\rho}))\prod_{i=1}^{K_r}(m-m^i)}{m\prod_{i=1}^{K_r}(m+\frac{1}{\gamma P_i{}^r})}\right)\\&\left(\frac{1}{m}-\frac{1}{m-m(-\frac{1}{\rho})}\right)dm\\=&\frac{1}{j2\pi}\int\ln\left(\frac{m-m(-\frac{1}{\rho})}{m}\right)\left(\frac{1}{m}-\frac{1}{m-m(-\frac{1}{\rho})}\right)dm\\&+\frac{1}{j2\pi}\int\ln\left(\frac{\prod_{i=1}^{K_r}(m-m^i)}{\prod_{i=1}^{K_r}(m+\frac{1}{\gamma P_i{}^r})}\right)\\&\left(\frac{1}{m}-\frac{1}{m-m(-\frac{1}{\rho})}\right)dm.\end{aligned} \tag{132}$$

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

$$\sigma^2{}_{\text{doa}} = \ln\left(\frac{\prod_{i=1}^{K_r} -m^i}{\prod_{i=1}^{K_r}\frac{1}{\gamma P_i{}^r}}\right) - \ln\left(\frac{\prod_{i=1}^{K_r}(m(-\frac{1}{\rho}) - m^i)}{\prod_{i=1}^{K_r}(m(-\frac{1}{\rho}) + \frac{1}{\gamma P_i{}^r})}\right). \tag{133}$$

Since

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

(product of the roots of polynomial $P(m)$ which is equal to $P(0)$) and

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

(The previous result comes from (130).)

We have therefore,

$$\begin{aligned}
\sigma^2{}_{\text{doa}} &= -\ln\left(\frac{\rho}{m(-\frac{1}{\rho})}\right) \\
&\quad - \ln\left(\frac{\rho}{m(-\frac{1}{\rho})} - \rho\xi m(-\frac{1}{\rho})\sum_{i=1}^{K_r}\frac{l_i{}^r}{(m(-\frac{1}{\rho}) + \frac{1}{\gamma P_i{}^r})^2}\right) \\
&= -\ln\left(1 - \xi m(-\frac{1}{\rho})^2\sum_{i=1}^{K_r}\frac{l_i{}^r}{(m(-\frac{1}{\rho}) + \frac{1}{\gamma P_i{}^r})^2}\right) \\
&= -\ln\left(1 - \rho^2\xi\alpha_{\text{doa}}{}^2\sum_{i=1}^{K_r}\frac{l_i{}^r(\gamma P_i{}^r)^2}{(1+\rho\gamma P_i{}^r\alpha_{\text{doa}})^2}\right).
\end{aligned} \tag{134}$$

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

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, 1998.

[2] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, 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," *Proc. IEEE*, vol. 92, no. 4, pp. 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," *Electron. Lett.*, 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 *Proc. Int. Symp. Signals, Systems, and Electronics*, vol. 4, 1998, pp. 295–300.

[6] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," AT & T Bell Labs, Tech. Rep. , 1995.

[7] X. Giraud and J. C. Belfiore, "Constellations matched to the Rayleigh fading channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, 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. Inf. Theory*, vol. 44, no. 4, pp. 1453–1467, Jul. 1998.

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

[9] C. E. Shannon, "A mathematical theory of communication," *Bell Labs Tech. J.*, vol. 27, pp. 379–457, Jul./Oct. 1948.

[10] E. T. Jaynes, "Information theory and statistical mechanics, Part 1," *Phys. Rev.*, vol. 106, pp. 620–630, 1957.

[11] ——, "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. dissertation, Stanford Univ., Stanford, CA, 1975.

[13] A. Zellner, *An Introduction to Bayesian Inference in Econometrics*, 2nd ed. New York: Wiley, 1971.

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

[15] G. L. Bretthorst, "Bayesian spectrum analysis and parameter estimation," Ph.D. dissertation, Wahsington Univ., St. Louis, MO, 1987.

[16] H. Jeffrey, *Theory of Probability*. London, U.K.: Oxford Univ. Press, 1939, 1948, 1961.

[17] E. T. Jaynes, *Probability Theory: The Logic of Science*. Cambridge, U.K.: Cambridge Univ. Press, 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. Inf. Theory*, vol. IT-26, no. 1, pp. 26–36, Jan. 1980.

[19] B. G. Larry, "An introduction to model selection using probability theory as logic," in *Maximum Entropy and Bayesian Methods*, G. R. Heidbreder, Ed. Dordrecht, The Netherlands: Kluwer, 1996, pp. 1–42.

[20] A. Mohammad-Djafari and G. Demoment, "Utilization de l'entropie dans les problèmes de restauration et de reconstruction d'images," *Traitement du Signal*, vol. 5, pp. 235–248, 1998.

[21] M. Franceschetti, S. Marano, and F. Palmieri, "The role of entropy in wave propagation," in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, Jun./Jul. 2003.

[22] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm. ETT*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[23] 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. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[24] K. Liu, V. Raghavan, and A. M. Sayeed, "Capacity scaling and spectral efficiency in wideband correlated MIMO channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2504–2526, Oct. 2003.

[25] A. Medles, S. Visuri, and D. T. M. Slock, "On MIMO capacity for various types of partial channel knowledge at the transmitter," in *Proc. IEEE Information Theory Workshop*, Paris, France, 2003, pp. 99–102.

[26] C. Chuah, D. M. C. Tse, J. Kahn, and R. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

[27] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[28] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[29] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*, 1st ed. http://www.nowpublishers.com/: Now (an Internet-based publisher), 2004, The Essence of Knowledge: Foundations and Trends in Communications and Information Theory.

[30] M. A. Kamath, B. L. Hughes, and Y. Xinying, "Gaussian approximations for the capacity of MIMO Rayleigh fading channels," in *Conference Rec. 36th Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2002, pp. 614–618.

[31] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multi-antenna channel-hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inf. Theory*, submitted for publication.

[32] A. L. Moustakas, S. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large analysis," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.

[33] A. M. Sengupta and P. P. Mitra. (2002) 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. [Online]. Available: http://mars.bell-labs.com

[34] A. Scaglione, "Statistical analysis of the capacity of MIMO frequency selective Rayleigh fading channels with arbitrary number of inputs and outputs," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 278.

[35] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[36] E. Biglieri and G. Taricco, *Transmission and Reception with Multiple Antenna: Theoretical Foundations*, 1st ed. http://www.nowpublishers.com/: Now (an Internet-based publisher), 2004, The Essence of Knowledge: Foundations and Trends in Communications and Information Theory.

[37] Z. D. Bai and J. W. Silverstein, "CLT of linear spectral statistics of large dimensional sample covariance matrices," *Ann. Prob.*, vol. 32, no. 1A, pp. 553–605, 2004.

[38] A. J. Grant and P. D. Alexander, "Random sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2832–2836, Nov. 1998.

[39] D. N. C. Tse and S. Hanly, "Linear multi-user receiver: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inf. Theory*, no. 3, pp. 641–657, Mar. 1999.

[40] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 3, pp. 622–640, Mar. 1999.

[41] P. Rapajic and D. Popescu, "Information capacity of random signature multiple-input multiple output channel," *IEEE Trans. Commun.*, vol. 48, no. 8, pp. 1245–1248, Aug. 2000.

[42] J. W. Silverstein and P. L. Combettes, "Large dimensional random matrix theory for signal detection and estimation in array processing," *Statist. Signal Array Process.*, pp. 276–279, 1992.

[43] S. Shamai (Shitz) and S. Verdú, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1302–1326, May 2001.

[44] M. Debbah, W. Hachem, P. Loubaton, and M. de Courville, "MMSE analysis of certain large isometric random precoded systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1293–1311, May 2003.

[45] J. M. Chaufray, W. Hachem, and P. Loubaton, "Asymptotical analysis of optimum and sub-optimum CDMA downlink MMSE receivers," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 189.

[46] R. Müller and S. Verdú, "Design and analysis of low-complexity interference mitigation on vector channels," *IEEE J. Sel. Areas Commun.*, no. 8, pp. 1429–1441, Aug. 2001.

[47] M. Debbah, "Linear Precoders for OFDM Wireless Communications," Ph.D. dissertation, Ecole Normale Supérieure de Cachan, Cachan, France, 2002.

[48] X. M. Pons, "Space-Time Processing and Channel Estimation: Performance Analysis and Asymptotic Results," Ph.D. dissertation, Universitat Politècnica de Catalunya, Barcelona, Spain, 2002.

[49] E. Biglieri, G. Taricco, and A. Tulino, "How far is infinity? Using asymptotic analysis in multiple antennas systems," in *Proc. IEEE 7th Int. Symp. Spread Spectrum Techniques and Applications*, vol. 1, Prague, Czech Republic, Sep. 2002, pp. 1–6.

[50] P. J. Smith and M. Shafi, "On a Gaussian approximation to the capacity of wireless MIMO systems," in *Proc. IEEE Int. Conf. Communications*, New York, 2002, pp. 406–410.

[51] Z. Wang and G. B. Giannakis, "Outage mutual information of space-time MIMO channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 4, pp. 657–662, Apr. 2004.

[52] D. N. C. Tse and O. Zeitouni, "Linear multiuser receivers in random environments," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 171–188, Jan. 2000.

[53] A. M. Tulino, S. Verdú, and A. Lozano, "Capacity of antenna arrays with space, polarization and pattern diversity," in *Proc. IEEE Information Theory Workshop*, Paris, France, 2003, pp. 324–327.

[54] M. Debbah, "An Information Theoretic Point of View to MIMO Channel Modeling," Flows Tech. Rep., D13, 2nd part, 2003.

[55] V. L. Girko, *Theory of Stochastic Canonical Equations*. Dordrecht, The Netherlands: Kluwer, 2001, vol. I and II.

[56] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *J. Phys. F: Metal Phys.*, vol. 5, pp. 965–974, 1975.

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

[58] P. Biane. (2000) Free Probability for Probabilists. [Online]. Available: http://www.dma.ens.fr/~biane/

[59] D. V. Voiculescu, "Multiplication of certain noncommuting random variables," *J. Operator Theory*, vol. 18, pp. 223–235, 1987.

[60] ——, "Limit laws for random matrices and free products," *Inventiones Mathematicae*, vol. 104, pp. 201–220, 1991.

[61] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free Random Variables*. Providence, RI: AMS, 1992, vol. 1, CRM Monograph Series.

[62] J. Evans and D. N. C. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2059–2078, Sep. 2000.

[63] A. Tulino and S. Verdú, "Asymptotic analysis of improved linear receivers for BPSK-CDMA subject to fading," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1544–1555, Aug. 2001.

[64] D. Tse, "Multiuser receivers, random matrices and free probability," in *Proc. 37th Annu. Allerton Conf.*, Monticello, IL, Sep. 1999.

[65] E. Biglieri, G. Taricco, and A. Tulino, "Performance of space-time codes for a large number of antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1794–1803, Jul. 2002.

[66] R. Müller, "A random matrix model of communication via antenna arrays," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2495–2506, Sep. 2002.

[67] W. Hachem, "Simple polynomial MMSE receivers for CDMA transmissions on frequency selective channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 164–172, Jan. 2004.

[68] P. Loubaton and W. Hachem, "Asymptotic analysis of reduced rank Wiener filters," in *Proc. IEEE Information Theory Workshop*, Paris, France, 2003, pp. 328–331.

[69] L. Cottatellucci and R. Müller, "Asymptotic design and analysis of multistage detectors with unequal powers," in *Proc. IEEE Information Theory Workshop*, Bangalore, India, 2002, pp. 167–170.

[70] Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *Ann. Prob.*, vol. 26, no. 1, pp. 316–345, 1998.

[71] R. Müller, "On the asymptotic eigenvalue distribution of concatenated vector-valued fading channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 2086–2091, Jul. 2002.

[72] M. Debbah, "Capacity of a MIMO multi-user network," in preparation, 2005.

[73] M. Debbah, R. Müller, H. Hofstetter, and P. Lehne, "Validation of mutual information complying MIMO models," *IEEE Trans. Wireless Commun.*. Available at http://www.eurecom.fr/~debbah, submitted for publication.

[74] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[75] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3095–3109, Dec. 2004.

[76] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[77] R. Müller, "On the accuracy of modeling the antenna array channel with random matrices," in *Proc. Int. Symp. Wireless Personal Multimedia Communications*, Aalborg, Denmark, 2001.

[78] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[79] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiken, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1225, Aug. 2002.

[80] D. Chizhik, J. Ling, P. W. Wolnianski, R. A. Valenzuela, N. Costa, and K. Huber, "Multiple-input multiple output measurements and modeling in Manhattan," *IEEE J. Sel. Areas in Communications*, vol. 21, pp. 321–331, Apr. 2003.

[81] 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, no. 1, pp. 58–68, Spring 1995.

[82] D. Gesbert, H. Bölcskei, D. Gore, and A. Paulraj, "MIMO wireless channels: Capacity and performance prediction," in *GLOBECOM Conf. Rec.*, vol. 2, 2000, pp. 1083–1088.

[83] D. Chizhik, G. J. Foschini, M. J. Gans, and R. A. Valenzuela, "Keyholes, correlations and capacities of multielement transmit and receive antennas," *IEEE Trans. Wireless Commu.*, vol. 1, no. 2, pp. 361–368, Apr. 2002.

[84] D. Gesbert, H. Bölcskei, D. A. Gore, and A. J. Paulraj, "Outdoor MIMO wireless channels: Models and performance prediction," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 1926–1934, Dec. 2002.

[85] E. T. Jaynes, "Where do we stand on maximum entropy?," in *The Maximum Entropy Formalism*, R. D. Levine and M. Tribus, Eds. Cambridge, MA: MIT Press, 1978, p. 15.