Institut Eurécom
Department of Corporate Communications
2229, route des Crètes
B.P. 193
06904 Sophia-Antipolis
FRANCE

Research Report RR-05-149

An Information Theoretic Point of View to MIMO Channel Modelling

Mérouane Debbah

Tel : (+33) 4 93 00 26 86
Fax : (+33) 4 93 00 26 27
Email : {merouane.debbah}@eurecom.fr

Preface

The work presented in this document was initiated while the author was a senior researcher at the Forschungszentrum Telekommunikation Wien (FTW) from November 2002 until July 2003. Many chapters have been continuously added since while the was appointed assistant Professor at Institut Eurecom. The initial goal was to develop and analyze wireless MIMO models. However, when starting the bibliographical work, the author got rapidly confused considering the huge amount of papers dedicated to MIMO channel models[1]. Indeed, nearly all the models proposed are different and moreover validated by measurement campaigns! Who should we trust? This contribution aims at providing some partial answers to this question by combining the influence of three fields namely communication theory, random matrix theory and Bayesian probability theory.

The author was first introduced to Bayesian probability theory and particularly Jaynes work as a graduate student of Prof. Guy Demoment. Formalizing Laplace's statement that "Probability Theory is nothing but common sense reduced to calculation", Jaynes shows how to conduct scientific inference with incomplete information (a situation encountered by all of us everyday). Probability Theory is developed as a reasonable degree of belief and a procedure to translate information into probability assignment is given through the use of the principle of maximum entropy. From a practical standpoint, the author was quite astonished at that time by the number of fields where the Bayesian inference procedure was proved to be successful (spectrum estimation, image reconstruction...). Bayesian inference is still in its infancy and the author hopes that this contribution will be a new application field of the powerful feature of "probability theory as logic".

As far as random matrix theory is concerned, the author was introduced to this theory as a PhD student of Prof. Philippe Loubaton while conducting research on the use of Linear Pre-coders for wireless OFDM (Orthogonal Frequency Division Multiplexing) transmissions (a single user version of downlink Multi-Carrier Code Divison Multiple Access also known as MC-CDMA). The goal was to study the usefulness of precoding in wireless fading channels and in particular analyze the Signal to Interference plus Noise Ratio (SINR) at the output of different receiver structures[2]. As the SINR did not have an interpretable expression and in order to overcome this problem, results from random matrix theory were used where the precoder was modeled as a random matrix (upon realistic justification with constraints such as orthogonality...). One of the useful features of this approach is its ability to predict, under certain conditions, the behavior of the empirical eigenvalue distribution of products and sums of matrices. The results were striking in terms of closeness to simulations with reasonable matrix size and the theory was shown to be an efficient tool to forecast the behavior of wireless systems with only few meaningful parameters (see [1]). The author has now no more doubts on the usefulness of this tool for the engineering community and hopes again that this contribution will confirm his point

---

[1]A web search on "Wireless channel Modelling" dated November 2002 (when this work started) showed more than 5000 publications on channel modelling. At a reading rate of 10 papers per day, it would take 500 days to have a small overview of the field! Note that for the non-british community, the number of publications is even higher (as modelling takes only one l).

[2]The SINR is usually used as a performance measure in communication theory. In the case where the interference is Gaussian, the SINR can be easily related to measures such as bit error rate (BER) or capacity.

of view and push forward random matrix analysis in engineering programs.

Contents

Chapter 1

Introduction

The problem of modelling channels is crucial for the efficient design of wireless systems [2, 3, 4]. The wireless channel suffers from constructive/destructive interference signaling [5, 6]. This yields a randomized channel with certain statistics to be discovered. Recently ([7, 8]), the need to increase spectral efficiency has motivated the use of multiple antennas at both the transmitter and the receiver side. Hence, in the case of i.i.d Gaussian entries of the MIMO link and perfect channel knowledge at the receiver, it has been proved [9] that the ergodic capacity increase is $\min(n_r, n_t)$ bits per second per hertz for every 3dB increase (n_r is the number of receiving antennas and n_t is the number of transmitting antennas) at high Signal to Noise Ratio (SNR)[1]. However, for realistic[2] channel models, results are still unknown and may seriously put into doubt the MIMO hype. As a matter of fact, the actual design of efficient codes is tributary of the channel model available: the transmitter has to know in what environment the transmission occurs in order to provide the codes with the adequate properties: as a typical example, in Rayleigh fading channels, when coding is performed, the Hamming distance (also known as the number of distinct components of the multi-dimensional constellation) plays a central role whereas maximizing the Euclidean distance is the commonly approved design criteria for Gaussian channels (see Giraud and Belfiore [10] or Boutros and Viterbo [11]).

As a consequence, channel modelling is the key in better understanding the limits of transmissions in wireless and noisy environments. In particular, questions of the form: "what is the highest transmission rate on a propagation environment where we only know the mean of each path, the variance of each path and the directions of arrival?" are crucially important. It will justify the use (or not) of MIMO technologies for a given state of knowledge.

Let us first introduce the modelling constraints. We assume that the transmission takes place between a mobile transmitter and receiver. The transmitter has n_t antennas and the receiver has n_r antennas. Moreover, we assume that the input transmitted signal goes through a time variant linear filter channel. Finally, we assume that the interfering noise is additive white Gaussian.

The transmitted signal and received signal are related as:

---

[1]In the single antenna Additive White Gaussian Noise (AWGN) channel, 1 bit per second per hertz can be achieved with every 3dB increase at high SNR.

[2]By realistic, we mean models representing our state of knowledge of reality which might be different from reality. Chapter 2 will explain in detail this idea.
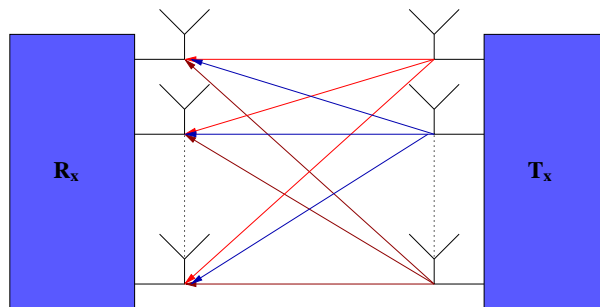
Figure 1.1: MIMO channel representation.

$$\mathbf{y}(t) = \sqrt{\frac{\rho}{n_t}} \int \mathbf{H}_{n_r \times n_t}(\tau, t) \mathbf{x}(t - \tau) d\tau + \mathbf{n}(t) \tag{1.1}$$

with

$$\mathbf{H}_{n_r \times n_t}(\tau, t) = \int \mathbf{H}_{n_r \times n_t}(f, t) e^{j2\pi f\tau} df \tag{1.2}$$

ρ is the received SNR (total transmit power per symbol versus total spectral density of the noise), t, f and τ denote respectively time, frequency and delay, $\mathbf{y}(t)$ is the $n_r \times 1$ received vector, $\mathbf{x}(t)$ is the $n_t \times 1$ transmit vector, $\mathbf{n}(t)$ is an $n_r \times 1$ additive standardized white Gaussian noise vector.

In the rest of the paper, we will only be interested in the frequency domain modelling (knowing that the impulse response matrix can be accessed through an inverse Fourier transform according to relation 1.2). We would like to provide some theoretical grounds to model the frequency response matrix $\mathbf{H}(f, t)$ based on a given state of knowledge. In other words, knowing only certain things related to the channel (Directions of Arrival (DoA), Directions of Departure (DoD), bandwidth, center frequency, number of transmitting and receiving antennas, number of chairs in the room...), how to attribute a joint probability distribution to the entries $h_{ij}(f, t)$ of the matrix:

$$\mathbf{H}_{n_r \times n_t}(f, t) = \begin{pmatrix} h_{11}(f, t) & \ldots & \ldots & h_{1n_t}(f, t) \\ \vdots & \ldots & \ldots & \vdots \\ \vdots & \ldots & \ldots & \vdots \\ h_{n_r 1}(f, t) & \ldots & \ldots & h_{n_r n_t}(f, t) \end{pmatrix} \tag{1.3}$$

This question can be answered in light of the Bayesian probability theory and the principle of maximum entropy. Bayesian probability theory has led to a profound theoretical understanding of various scientific areas [12, 13, 14, 15, 16, 17, 18, 19] 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[3] is at present the clearest theoretical justification in conducting scientific inference: we do not need a model, entropy maximization creates a model for us out of the information available. Choosing the distribution with greatest entropy avoids the arbitrary

---

[3]The principle of maximum entropy was first proposed by Jaynes [13, 14] as a general inference procedure although it was first used in Physics.

introduction or assumption of information that is not available[4]. Bayesian probability theory improves on maximum entropy by expressing some prior knowledge on the model and estimating the parameters of the model.

As we will emphasize all along this paper, channel modelling is not a science representing reality but only our knowledge of reality as thoroughly stated by Jaynes in [21]. It answers in particular the following question: based on a given state of knowledge (usually brought by raw data or prior information), what is the best model one can make? This is, of course, a vague question since there is no strict definition of what is meant by best. But what do we mean then by best? In this contribution, our aim is to derive a model which is adequate with 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 (complying to certain common sense desiderata, see [22] to express these desiderata and for the derivation of entropy) and easy to manipulate: we need a general principle for translating information into probability assignment. Entropy is the measure of information that fulfills this criteria. Hence, already in 1980, Shore et al. [22] 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[5]. The consistency argument is at the heart of scientific inference and can be expressed through the following axiom[6]:

Axiom 1 *If the prior information* $\mathbf{I_1}$ *on which is based the channel model* $\mathbf{H_1}$ *can be equated to the prior information* $\mathbf{I_2}$ *of the channel model* $\mathbf{H_2}$ *then both models should be assigned the same probability distribution* $P(\mathbf{H}) = P(\mathbf{H_1}) = P(\mathbf{H_2})$.

Any other procedure would be inconsistent in the sense that, by changing indices 1 and 2, we could then generate a new problem in which our state of knowledge is the same but in which we are assigning different probabilities. More precisely, Shore et al. [22] formalize the maximum entropy approach based on four consistency axioms stated as follows[7]:

- Uniqueness: If one solves the same problem twice the same way then the same answer should result both times.

- Invariance: If one solves the same problem in two different coordinate systems then the same answer should result both times.

---

[4]Keynes named it judiciously the principle of indifference [20] to express our indifference in attributing prior values when no information is available.

[5]Thus, aiming for consistency, we can maximize entropy without loss of generality.

[6]The consistency property is only one of the required properties for any good calculus of plausibility statement. In fact, R.T Cox in 1946 derived three requirements known as Cox's Theorem[23]:

- Divisibility and comparability: the plausibility of of a statement is a real number between 0 (for false) and 1 (for true) and is dependent on information we have related to the statement.

- Common sense: Plausibilities should vary with the assessment of plausibilities in the model.

- Consistency: If the plausibility of a statement can be derived in two ways, the two results should be equal.

[7]In all the rest of the document, the consistency argument will be referred to as Axiom 1.

- System independence: It should not matter whether one accounts for independent information about independent systems separately in terms of different densities or together in terms of a joint density.

- Subset independence: It should not matter whether one treats an independent subset of system states in terms of a separate conditional density or in terms of the full system density.

These axioms are based on the fundamental principle that if a problem can be solved in more than one way, the results should be consistent. Given this statement in mind, the rules of probability theory should lead every person to the same unique solution, provided each person bases his model on the same information.[8]

Moreover, the success over the years of the maximum entropy approach (see Boltzmann's kinetic gas law, [24] for the estimate of a single stationary sinusoidal frequency, [15] for estimating the spectrum density of a stochastic process subject to autocorrelation constraints, [25] for estimating parameters in the context of image reconstruction and restoration problems, [26] for applying the maximum entropy principle on solar proton event peak fluxes in order to determine the least biased distribution) has shown that this information tool is the right way so far to express our uncertainty.

Let us give an example in the context of spectral estimation of the powerful feature of the maximum entropy approach which has inspired this monograph. Suppose a stochastic process x_i for which $p + 1$ autocorrelation values are known i.e $\mathbb{E}(x_i x_{i+k}) = \tau_k, k = 0, ..., p$ for all i. What is the consistent model one can make of the stochastic process based only on that state of knowledge, in other words the model which makes the least assumption on the structure of the signal? The maximum entropy approach creates for us a model and shows that, based on the previous information, the stochastic process is a pth auto-regressive (AR) order model process of the form [15]:

$$x_i = -\sum_{k=1}^{p} a_k x_{i-k} + b_i$$

where the b_i are i.i.d zero mean Gaussian distributed with variance σ^2 and $a_1, a_2, .., a_p$ are chosen to satisfy the autocorrelation constraints (through Yule-Walker equations).

In this contribution, we would like to provide guidelines for creating models from an information theoretic point of view and therefore make extensive use of the principle of maximum entropy together with the principle of consistency. Once the models are created, we will test

---

[8]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. The principle states that, of the distribution P that satisfy the constraints, one should choose the one which minimizes the functional:

$$D(P, Q) = \int P(x)\log \ \frac{P(x)}{Q(x)} \ dx$$

Minimizing cross-entropy is equivalent to maximizing entropy when the prior Q is a uniform distribution. Intuitively, cross-entropy measures the amount of information necessary to change the prior Q into the posterior P. If measured data is available, Q can be estimated. However, one can only obtain a numerical form for P in this case (which is not always useful for optimization purposes). Moreover, this is not a easy task for multidimensional vectors such as vec($\mathbf{H}$). As a consequence, we will always assume a uniform prior and use therefore the principle of maximum entropy.

them upon some useful metrics for communications engineers. In particular, we will focus our analysis on the mutual information and MMSE SINR. To ease (in terms of complexity) the validation process, for each state of knowledge, the asymptotic mutual information and MMSE SINR distribution will be derived based only on the number of parameters at hand. Moreover, answers will be given to the following question:

- What is the impact of the number of scatterers, DoD, DoA, power of the steering directions on the mutual information and MMSE (Minimum Mean Square Error) SINR distribution?

In the rest of this contribution, for sake of simplicity, we will write $\mathbf{H}$ instead of $\mathbf{H}(f, t)$ (without forgetting the dependency on frequency and time). Supposing that the model is adequate with reality and that the channel is perfectly known at the receiver, we will prove the following conjecture.

Conjecture: Let

$$I^M(n_t, n_r, \rho) = \log_2 \det \left(\mathbf{I}_{n_t} + \frac{\rho}{n_t} \mathbf{H}^H \mathbf{H} \right)\ [9]$$

and

$$\text{SINR}^k(n_t, n_r, \rho) = \mathbf{h}_k^H \left(\mathbf{U}\mathbf{U}^H + \frac{n_t}{\rho} \mathbf{I}_{n_r} \right)^{-1} \mathbf{h}_k\ [10].$$

Then, for many channel models,

$$\exists \mu, \sigma \quad \lim_{n_t \to \infty, \frac{n_r}{n_t} = \gamma} I^M(n_t, n_r, \rho) - n_t \mu \to N(0, \sigma^2) \qquad (1.4)$$

$$\exists \mu, \sigma \quad \lim_{n_t \to \infty, \frac{n_r}{n_t} = \gamma} \sqrt{n_r} \left(\text{SINR}^k - \mu \right) \to N(0, \sigma^2) \qquad (1.5)$$

The convergence is in distribution[11]. Only the mean μ and the variance σ^2 are needed to fully characterize the distribution. Note that $\mu = \mu(\gamma, \rho)$ and $\sigma^2 = \sigma^2(\gamma, \rho)$ depend on $\gamma = \frac{n_r}{n_t}$ and ρ. When this conjecture cannot be proved, only the mean will be derived.

For proving the conjecture, results of random matrix theory will be used [27]. Random matrices

---

[9]Note that $I^M(n_t, n_r, \rho)$ is also equal to $\log_2 \det \left(\mathbf{I}_{n_r} + \frac{\rho}{n_t} \mathbf{H}\mathbf{H}^H \right)$ (this result stems from the determinant identity $\det(I + AB) = \det(I + BA)$).

[10]$\mathbf{h}_k$ is the kth column of $\mathbf{H}$ and $\mathbf{U}$ is the $n_r \times (n_t - 1)$ matrix which remains after extracting $\mathbf{h}_k$ from $\mathbf{H}$.

[11]Although the result is asymptotical, we will use it in the finite dimension case upon justification. However, we will not forget that in the latter case, it is only an approximation. Otherwise, many disturbing results may appear. For example, $I^M(n_t, n_r, \rho)$ is always a positive quantity while by making the Gaussian approximation in the finite case, we make allowances of negative values to appear. Fortunately, the probability of $I^M(n_t, n_r, \rho)$ to be negative tends to zero as the number of transmitting antennas n_t increases. Indeed,

$$
\begin{aligned}
P(I^M \leq 0) &= \int_{-\infty}^{0} \mathrm{d}I^M p(I^M) \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{0} \mathrm{d}I^M e^{-\frac{(I^M - n_t\mu)^2}{2\sigma^2}} \\
&= 1 - Q\left(\frac{-n_t\mu}{\sigma}\right) \to_{n_t \to \infty} 0
\end{aligned}
$$

μ and σ are positive constant whereas $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \mathrm{d}t e^{\frac{-t^2}{2}}$.

were first proposed by Wigner in quantum mechanics to explain the measured energy levels of nuclei in terms of the eigenvalues of random matrices. When Telatar [9] (in the context of antenna capacity analysis), Grant & Alexander[12](in the analysis of random sequences for code-division multiple access) [28], Rapajic & Popescu [30] and then simultaneously Tsé & Hanly [31] and Verdu & Shamai [32] (for the analysis of uplink unfaded CDMA equipped with certain receivers) introduced random matrices, the random matrix theory entered the field of telecommunications[13]. From that time, random matrix theory has been successively extended to other cases such as uplink CDMA fading channels [34], OFDM [1, 35], downlink CDMA [36], multi-user detection [37][14], advanced MIMO models [40, 41, 42, 43]. One of the useful features of random matrix theory is the ability to predict, under certain conditions, the behavior of the empirical eigenvalue distribution of products or sums of matrices. The results are striking in terms of closeness to simulations with reasonable matrix size and enable to derive linear spectral statistics for these matrices with only few meaningful parameters. Note that other powerful tools of statistical mechanics based on the replica method[15] could be used[45, 46, 47, 48] to derive similar asymptotic results; however, these results will not be used in this monograph.

Notations: In the following, upper and lower boldface symbols will be used for matrices and column vectors, respectively. $(.)^T$ will denote the transpose operator, $(.)^\star$ conjugation and $(.)^H = \left((.)^T\right)^\star$ hermitian transpose. ln is the natural logarithm such $\ln(e) = 1$. When this notation is used, the mutual information is given in nats/s. When the notation $\log_2(x) = \frac{\ln(x)}{\ln(2)}$ is used, the results are given in bits/s. The Stieltjes Transform $m(z)$ of a distribution F is defined as

$$m(z) = \int \frac{1}{\lambda - z} dF(\lambda) \tag{1.6}$$

We recall that, if the distribution function $F(\lambda)$ has a continuous derivative, it is related to its Stieltjes transform $m(z) = \int \frac{dF(\lambda)}{\lambda - z}$ by

$$\frac{\mathrm{d}F}{\mathrm{d}\lambda} = \frac{1}{\pi} \lim_{y \to 0^+} \mathrm{Im}(m(\lambda + iy)).$$

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

$$\delta_{im} = \begin{cases} 1 & \text{if } i = m \\ 0 & \text{otherwise} \end{cases} \tag{1.7}$$

Moreover,

$$\mathbb{1}_{[0,2\pi]}(x) = \begin{cases} 1 & \text{if } x \in [0, 2\pi] \\ 0 & \text{otherwise} \end{cases} \tag{1.8}$$

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

---

[12]In [28], it is shown that the use of random sequences incurs asymptotically no rate penalty compared to the use of the optimal spreading sequences determined by Rupf and Massey in 1994 [29] which satisfy the Welch's lower bound.

[13]It should be noted that in the field of array processing, Silverstein used already in 1992 random matrix theory [33] for signal detection and estimation.

[14]Note that in the context of multiuser detection, the simulated curves of spectral efficiency of CDMA systems in [38] (later published in English in 1999 [39]) triggered theoretical analysis based on random matrices.

[15]Tanaka was the first to introduce in 2001 the replica method [44] as an efficient tool to evaluate analytically the performance of synchronous CDMA multiuser detectors

Chapter 2

Some Considerations

2.1 Channel Modelling Methodology

In this contribution, we provide a methodology (already successfully used in Bayesian spectrum analysis [24, 18]) for inferring on channel models. The goal of the modelling methodology is twofold:

- to define a set of rules, called hereafter *consistency axioms*, where only our state of knowledge needs to be defined.

- to use a measure of uncertainty, called hereafter *entropy*, in order to avoid the arbitrary introduction or assumption of information that is not available.

In other words, if two published papers make the same assumptions in the abstract (concrete buildings in Oslo where one avenue...), then both papers should provide the same channel model.

To achieve this goal, in all this document, the following procedure will be applied: every time we have some information on the environment (*and not make assumptions on the model!*), we will ask a question based on that the information and provide a model taking into account that information and nothing more! The resulting model and its compliance with later test measurements will justify whether the information used for modelling was adequate to characterize the environment in sufficient details. Hence, when asked the question, "what is the consistent model one can make knowing the directions of arrival, the number of scatterers, the fact that each path has zero mean and a given variance?" we will suppose that the information provided by this question is unquestionable and true i.e the propagation environment depends on fixed steering vectors, each path has effectively zero mean and a given variance. We will suppose that effectively, when waves propagate, they bounce onto scatterers and that the receiving antenna sees these ending scatterers through steering directions. Once we assume this information to be true, we will construct the model based on Bayesian tools.[1].

To explain this point of view, the author recalls an experiment made by his teacher during a tutorial explanation on the duality behavior of light: photon or wave. The teacher took two students of the class, called here A and B for simplicity sake. To student A, he showed view (1') (see Figure 2.1) of a cylinder and to student B, he showed view (2') of the same cylinder. For A, the cylinder was a circle and for B, the cylinder was a rectangle. Who was wrong? Well, nobody.

---

[1]Note that in Bayesian inference, all probabilities are conditional on some hypothesis space (which is assumed to be true).
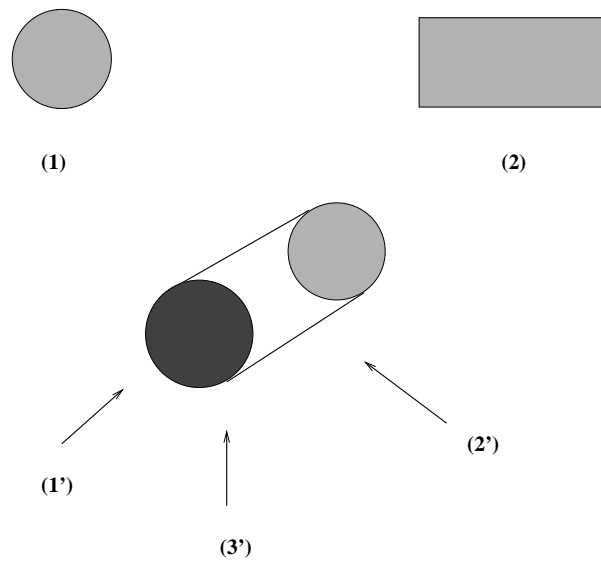
Figure 2.1: Duality wave-corpuscule?

Based on the state of knowledge (1'), representing the cylinder as a circle is the best one can do. Any other representation of the cylinder would have been made on unjustified assumptions (the same applies to view (2')). Unless we have another state of knowledge (view (3')), the true nature of the object will not be found.

Our channel modelling will not pretend to seek reality but only to represent view (1') or view (2') in the most accurate way (i.e if view (1') is available then our approach should lead into representing the cylinder as a circle and not as a triangle for example). If the model fails to comply with measurements, we will not put into doubt the model but conclude that the information we had at hand to create the model was insufficient. We will take into account the failure as a new source of information and refine/change our question in order to derive a new model based on the principle of maximum entropy which complies with the measurements. This procedure will be routinely applied until the right question (and therefore the right answer) is found. When performing scientific inference, every question asked, whether right or wrong, is important. Mistakes are eagerly welcomed as they lead the path to better understand the propagation environment. Note that the approach devised here is not new and has already been used by Jaynes [21] and Jeffrey [49]. We give hereafter a summary of the modelling approach:

1. **Question selection**: the modeler asks a question based on the information available.

2. **Construct the model**: the modeler uses the principle of maximum entropy (with the constraints of the question asked) to construct the model M_i.

3. **Test**: (When complexity is not an issue) The modeler computes the a posteriori probability of the model and ranks the model (see chapter.9)

4. **Return to 1.**: The outcome of the test is some "new information" evidence to keep/refine/change the question asked. Based on this information, the modeler can therefore make a new model selection.

This algorithm is iterated as many times as possible until better ranking is obtained. However, we have to alert the reader on one main point: the convergence of the previous algorithm is not at all proven. Does this mean that we have to reject the approach? we should not because our aim is to better understand the environment and by successive tests, we will discard some solutions and keep others.

We provide hereafter a brief historical example to highlight the methodology. In the context of spectrum estimation, the Schuster periodogram (also referred in the literature as the discrete Fourier transform power spectrum) is commonly used for the estimation of hidden frequencies in the data. The Schuster periodogram is defined as:

$$F(\omega) = \frac{1}{N} \mid \sum_{k=1}^{N} s_k e^{-j\omega t_k} \mid^2$$

s_k is the data of length N to be analyzed. In order to find the hidden frequencies in the data, the general procedure is to maximize $F(\omega)$ with respect to ω . But as in our case, one has to understand why/when to use the Schuster periodogram for frequency estimation. The Schuster periodogram answers a specific question based on a specific assumption (see the work of Bretthorst [18]). In fact, it answers the following question: "what is the optimal frequency estimator for a data set which contains a **single stationary sinusoidal frequency** in the presence of Gaussian white noise?" From the standpoint of Bayesian probability, the discrete Fourier Transform power spectrum answers a specific question about single (and not two or three....) stationary sinusoidal frequency estimation. Given this state of knowledge, the periodogram will consider everything in the data that cannot be fit to a single sinusoid to be noise and will therefore, if other frequencies are present, misestimate them. However, if the periodogram does not succeed in estimating multiple frequencies, the periodogram is not to blame but only the question asked! One has to devise a new model (a model maybe based on a two stationary sinusoidal frequencies?). This new model selection will lead to a new frequency estimator in order to take into account the structure of what was considered to be noise. This routine is repeated and each time, the models can be ranked to determine the right number of frequencies.

2.2 Information and Complexity

In the introduction, we have recalled the work of Shore et al. [22] 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. In particular, how does one incorporate information on the fact that the room has four walls and two chairs? In this case, we will not not maximize entropy based only on the information we have (expected values and number of chairs and walls): we will maximize entropy based on the expected values and a structured form of the channel matrix (which is more than the information we have since the chairs and walls are not constraint equations in the entropy criteria). This ad-hoc procedure will be used because it is extremely difficult to incorporate knowledge on physical considerations (number of chairs, type of room...) in the entropy criteria. Each time this ad-hoc procedure is used, we will verify that although we maximize entropy under a structured constraint, we remain consistent. Multiple iterations of this procedure will refine the structured form of the

channel until the modeler obtains a consistent structured models that maximizes entropy.

A question the reader could ask is whether we should take into account all the information provided, in other words, what information should is useful? We should of course consider all the available information but there is a compromise to be made in terms of model complexity. Each information added will not have the same effect on the channel model and might as well more complicate the model for nothing rather than bring useful insight on the behavior of the propagation environment. To assume further information by putting some additional structure would not lead to incorrect predictions: however, if the predictions achieved with or without the details are equivalent, then this means that the details may exist but are irrelevant for the understanding of our model[2]. As a typical example, when conducting iterative decoding analysis [50], Gaussian models of priors are often sufficient to represent our information. Inferring on other moments and deriving the true probabilities will only complicate the results and not yield a better understanding.

2.3 Some Metrics

In this section, we introduce two metrics of interest for which the models derived within this monograph should at least comply with.

2.3.1 Capacity Considerations

Before starting any discussions on MIMO capacity, let us first review the pioneering work of Telatar[9] (later published as [51]) that triggered research in multi-antenna systems[3]. In this paper, Telatar develops the channel capacity of a general MIMO channel. Assuming perfect knowledge of $\mathbf{H}$ at the receiver, the mutual information I^M between input and output is given by[4]:

$$
\begin{aligned}
I^M(\mathbf{x}; (\mathbf{y}, \mathbf{H})) &= I^M(\mathbf{x}; \mathbf{H}) + I^M(\mathbf{x}; \mathbf{y} \mid \mathbf{H})) \\
&= I^M(\mathbf{x}; \mathbf{y} \mid \mathbf{H}))^5 \\
&= \text{Entropy}(\mathbf{y} \mid \mathbf{H}) - \text{Entropy}(\mathbf{y} \mid \mathbf{x}, \mathbf{H})^6 \\
&= \text{Entropy}(\mathbf{y} \mid \mathbf{H}) - \text{Entropy}(\mathbf{n} \mid \mathbf{H})
\end{aligned}
$$

In the case where the entries have a covariance matrix $\mathbf{Q}$ ($\mathbf{Q} = \mathbb{E}(\mathbf{x}\mathbf{x}^H)$), we have, since $\mathbf{y} = \frac{\rho}{n_t}\mathbf{H}\mathbf{x} + \mathbf{n}$:

$$
\mathbb{E}(\mathbf{y}\mathbf{y}^H) = \mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}\mathbf{Q}\mathbf{H}^H
$$

$$
\mathbb{E}(\mathbf{n}\mathbf{n}^H) = \mathbf{I}_{n_r}
$$

---

[2]Limiting one's information is a general procedure that can be applied to many other fields. As a matter of fact, the principle "one can know less but understand more" seems the only reasonable way to still conduct research considering the huge amount of papers published each year.

[3]For contribution [51], Telatar received the 2001 Information Theory Society Paper Award. After the Shannon Award, the IT Society Paper Award is the highest recognition award by the IT society.

[4]Note that the channel is entirely described with input $\mathbf{x}$ and output $(\mathbf{y}, \mathbf{H}) = (\mathbf{H}\mathbf{x} + \mathbf{n}, \mathbf{H})$.

If the $x_i, i = 1...n_t$ are Gaussian[7]:

$$
\begin{aligned}
C(\mathbf{Q}) &= \mathbb{E}\left(\text{Entropy}(\mathbf{y} \mid \mathbf{H} = H) - \text{Entropy}(\mathbf{n} \mid \mathbf{H} = H)\right) \\
&= \mathbb{E}\left(\log_2 \det 2\pi e(\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}\mathbf{Q}\mathbf{H}^H) - \log_2 \det 2\pi e \mathbf{I}_{n_r}\right) \\
&= \mathbb{E}\left(\log_2 \det(\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}\mathbf{Q}\mathbf{H}^H)\right)
\end{aligned}
$$

As a consequence[8], the ergodic capacity of an $n_r \times n_t$ MIMO channel with Gaussian entries and covariance matrix $\mathbf{Q}$ ($\mathbf{Q} = \mathbb{E}(\mathbf{x}\mathbf{x}^H)$)[9] is :

$$\overline{C} = \max_{\mathbf{Q}} \mathbb{E}_{\mathbf{H}}\left(C(\mathbf{Q})\right) \tag{2.1}$$

where the maximization is over a set of positive semi-definite hermitian matrices $\mathbf{Q}$ satisfying the power constraint $\text{trace}(\mathbf{Q}) \leq P$, and the expectation is with respect to the random channel matrix. Therefore, the ergodic capacity is achieved for a particular choice of the matrix $\mathbf{Q}$. It is quite astonishing that usual channel measurements define the ergodic capacity as:

$$I^M = \frac{1}{N}\sum_{i=1}^{N} \log_2 \det\left(\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H_i}\mathbf{H_i}^H\right) \tag{2.2}$$

Where has the matrix $\mathbf{Q}$ disappeared? In the original paper [9], Telatar exploits the isotropical property of Gaussian i.i.d $\mathbf{H}$ to show that in this case, ergodic capacity is achieved with $\mathbf{Q} = \mathbf{I}$. However, this result has been proved only for a Gaussian i.i.d channel matrix and was not extended to other types of matrices. In correlated fading, I^M in (2.2) is called the average mutual information with covariance $\mathbf{Q} = \mathbf{I}$. It has never been proved that capacity was close to this mutual information except for certain particular cases (see [51, 52]). I^M in (2.2) underestimates the achievable rate[10]: indeed, even though the channel realization is not known, the knowledge of the channel model (is it i.i.d Rayleigh fading? is it i.i.d Rice fading? is it correlated Rayleigh fading with a certain covariance matrix?...) can be taken into account in order to optimize the coding scheme at the transmitter. There is no reason why one should transmit independent substreams on each antenna. It is as if one stated that the space-time codes designed through the rank and determinant criterion [53] were optimal for **all kinds of MIMO channels**.

The only explanation to this historical accident is the difficulty in deriving the optimum matrix $\mathbf{Q}$ when the channel is not Gaussian i.i.d. But at least, I^M in (2.2) should be called accordingly as an average mutual information with covariance matrix identity. In [54, 55, 56, 57, 58, 52]

---

[7]The differential entropy of a complex Gaussian vector $\mathbf{x}$ with covariance $\mathbf{Q}$ is given by $\log_2 \det(\pi e \mathbf{Q})$.

[8]We only have derived the mutual information with Gaussian entries and have not proved this achieves capacity. This stems from the fact that for a given covariance $\mathbf{Q}$, the entropy of $\mathbf{x}$ is always inferior to $\log_2 \det(\pi e \mathbf{Q})$ with equality if and only if $\mathbf{x}$ complex Gaussian

[9]In the general case where the noise is Gaussian with a covariance matrix $\mathbf{Z}$, the capacity is given by: $C(\mathbf{Q}, \mathbf{Z}) = \log_2 \frac{\det\ \mathbf{Z} + \frac{\rho}{n_t}\mathbf{H}\mathbf{Q}\mathbf{H}^H}{\det(\mathbf{Z})}$

[10]Note however that although not optimum, the mutual information with covariance $\mathbf{Q} = \mathbf{I}$ can be useful in the analysis of systems where the codebook can not be changed according to the wireless environment and therefore remains the same during the whole transmission.

this fact is well understood (only bounds on the ergodic capacity are derived for correlated channels); in many other papers, the mistake persists when considering ergodic capacity. This misconception is also a motivation for studying more precisely the probability distribution of a matrix $\mathbf{H}$ in order to derive the coherent ergodic capacity (receiver has perfect channel knowledge, the transmitter only knows the statistics). Such calculations do not have only an academic dissemination goal for deriving capacity issues but also a practical one. One of the visions of future wireless communications the author would like to advocate is the following: depending on the environment (dense or not, field, street, number of chairs,...), the user writes down that information on his mobile terminal (this can also be done by downloading localization information from the base station if the user wants an automatic process). Immediately, based on that state of knowledge, an online channel model is created using the maximum entropy approach (which incorporates only that information and not more!). The transmitted signal and the coding scheme is then online optimized for that specific scenario (by deriving new rank and determinant criteria for example). Such a service could be called "user customized channel model coding service". From a software defined radio perspective, this scenario is completely viable.

Another misconception concerns the design of MIMO systems. For a wireless content provider, the most important criteria is the quality of service to be delivered to customers. This quality of service can be quantified through measures such as outage capacity: if q=1% is the outage probability of having an outage capacity of R, then this means that the provider is able to ensure a rate of R in 99% of the cases. Since the channels are rarely ergodic, the derivations of ergodic capacities are of no use for content providers. We give hereafter the two definitions of ergodic and outage capacity:

- If the channels are ergodic, $C(\mathbf{Q})$ can be averaged over many channel realizations and the corresponding capacity is defined as : $\overline{C} = \max_{\mathbf{Q}} \mathbb{E}(C(\mathbf{Q}))$

- If the channels are static, there is only one channel realization and an outage probability for each positive rate of transmission can be defined:[11]

$$C_q = \max_{\mathbf{Q}} \sup\{R \geq 0 : Pr[C(\mathbf{Q}) < R] \leq q\}$$

Once again, when deriving the outage capacity, if the channel distribution is known, then it is possible at the transmitter side to optimize the covariance of the transmitting signal. However, this is not an obvious task[12] and in all the following we will therefore derive the outage mutual information with Gaussian input covariance matrix $\mathbf{Q} = \mathbf{I}$ (**and not the outage capacity!**).

Many results have already been derived on the ergodic capacity of channels based on different channel models taking into account correlation [59, 57, 60, 61] or not [9]. However, very few have been devoted to the outage capacity [62, 63, 64] or deriving the capacity distribution [65, 66]. In this respect, conjecture 1.4 is extremely useful. Indeed, let q denote the outage probability

---

[11]Note that the covariance matrix $\mathbf{Q}$ which optimizes the ergodic capacity does not necessarily optimize the outage capacity.

[12]Except in the case of i.i.d Gaussian entries where this mutual information is equal to capacity.

and $I^M_{\ q}$ the corresponding outage mutual information with covariance $\mathbf{Q} = \mathbf{I}$, then:

$$
\begin{aligned}
q &= \mathrm{P}\left(I^M \leq I^M_{\ q}\right) \\
&= \int_{-\infty}^{I^M_{\ q}} \mathrm{d}I^M p(I^M) \\
&\approx \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{I^M_{\ q}} \mathrm{d}I^M e^{-\frac{(I^M - n_t\mu)^2}{2\sigma^2}} \\
&\approx 1 - Q\left(\frac{I^M_{\ q} - n_t\mu}{\sigma}\right)
\end{aligned}
$$

$$
I^M_{\ q} \approx n_t\mu + \sigma Q^{-1}(1-q)
$$

We define[13]

$$
Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \mathrm{d}t e^{\frac{-t^2}{2}}
$$

Therefore, for deriving the outage mutual information, only knowledge of the mean and variance of the mutual information distribution is needed in large system limit.

2.3.2 MMSE SINR Considerations

As far as the MMSE SINR is concerned and based on the model 1.1, the output of the MMSE detector $\hat{\mathbf{x}} = [\hat{x}_1, \ldots, \hat{x}_{n_t}]^T$ is given by

$$
\begin{aligned}
\hat{\mathbf{x}} &= E\left(\mathbf{x}\mathbf{y}^H\right)\left[E(\mathbf{y}\mathbf{y}^H)\right]^{-1}\mathbf{y} & (2.3) \\
&= \sqrt{\frac{\rho}{n_t}}\mathbf{H}^H\left(\frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H + \mathbf{I}_{n_r}\right)^{-1}\mathbf{y} & (2.4) \\
&= \sqrt{\frac{\rho}{n_t}}\mathbf{H}^H\left(\mathbf{A}\right)^{-1}\mathbf{y}\ . & (2.5)
\end{aligned}
$$

with $\mathbf{A} = \frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H + \mathbf{I}_{n_r}$. Each component $\hat{x}_k$ of $\hat{\mathbf{x}}$ is corrupted by the effect of both the thermal noise and by the "multi-user interference" due to the contributions of the other symbols $\{x_l\}_{l\neq k}$. Let us now derive the expression of the SINR at one of the n_t outputs of the MMSE detector. Let $\mathbf{h}_k$ be the column of $\mathbf{H}$ associated to element x_k, and $\mathbf{U}$ the $n_r \times (n_t - 1)$ matrix which remains after extracting $\mathbf{h}_k$ from $\mathbf{H}$.

The component $\hat{x}_k$ after MMSE equalization has the following form:

$$
\hat{x}_k = \eta_{\mathbf{h}_k} x_k + \tau_k
$$

where

$$
\eta_{\mathbf{h}_k} = \frac{\rho}{n_t}\mathbf{h}_k^H\left(\mathbf{A}\right)^{-1}\mathbf{h}_k\ . \tag{2.6}
$$

---

[13]Usual programming tools use the complementary error function defined as $\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}dt$. In this case, $Q(x) = \frac{1}{2}\mathrm{erfc}(\frac{x}{\sqrt{2}})$.

and

$$\tau_k = \frac{\rho}{n_t}\mathbf{h}_k^H\left(\mathbf{A}\right)^{-1}\mathbf{H}[x_1,\ldots,x_{k-1},0,x_{k+1},\ldots,x_{n_t}]^T + \sqrt{\frac{\rho}{n_t}}\mathbf{h}_k^H\left(\mathbf{A}\right)^{-1}\mathbf{n}$$

The variance of τ_k is given by: $V = E(|\tau_k|^2|\mathbf{H})$. Knowing that $\mathbf{U}\mathbf{U}^H = \mathbf{H}\mathbf{H}^H - \mathbf{h}_k\mathbf{h}_k^H$, we get:

$$
\begin{aligned}
V &= \frac{\rho^2}{n_t^2}\mathbf{h}_k^H\left(\mathbf{A}\right)^{-1}\mathbf{U}\mathbf{U}^H\left(\mathbf{A}\right)^{-1}\mathbf{h}_k + \frac{\rho}{n_t}\mathbf{h}_k^H\left(\mathbf{A}\right)^{-1}\left(\mathbf{A}\right)^{-1}\mathbf{h}_k \\
&= \frac{\rho}{n_t}\left(\mathbf{h}_k^H\left(\mathbf{A}\right)^{-1}[\frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H - \frac{\rho}{n_t}\mathbf{h}_k\mathbf{h}_k^H + \mathbf{I}_{n_r}]\left(\mathbf{A}\right)^{-1}\mathbf{h}_k\right) \\
&= \frac{\rho}{n_t}\left(\eta_{\mathbf{h}_k} - \eta_{\mathbf{h}_k}^2\right) \\
&= \frac{\rho}{n_t}\eta_{\mathbf{h}_k}(1 - \eta_{\mathbf{h}_k})
\end{aligned}
$$

The Signal to Interference plus Noise Ratio SINR^k at the output k of the MMSE detector can thus be expressed as:

$$
\begin{aligned}
\text{SINR}^k &= \frac{E[|\eta_{\mathbf{h}_k}x_k|^2|\mathbf{H}]}{E[|\tau_k|^2|\mathbf{H}]} \\
&= \frac{(\eta_{\mathbf{h}_k})^2}{\eta_{\mathbf{h}_k}(1 - \eta_{\mathbf{h}_k})} \\
&= \frac{\eta_{\mathbf{h}_k}}{1 - \eta_{\mathbf{h}_k}}
\end{aligned}
$$

Writing $\mathbf{H}\mathbf{H}^H = \mathbf{U}\mathbf{U}^H + \mathbf{h}_k\mathbf{h}_k^H$ and invoking the matrix inversion lemma[14], we get after some simple algebra another useful expression for this SINR (see e.g. [31]):

$$\text{SINR}^k = \mathbf{h}_k^H\left(\mathbf{U}\mathbf{U}^H + \frac{n_t}{\rho}\mathbf{I}_{n_r}\right)^{-1}\mathbf{h}_k. \tag{2.7}$$

For a fixed n_r and n_t, it is extremely difficult to get insight on the performance of the MMSE receiver from the expressions (2.7) and (2.6). As a consequence, in order to obtain interpretable expressions, we will focus on an asymptotical analysis of the SINR. Moreover, it has been shown in [67] and [68] that the additive noise τ_k can be considered as Gaussian when n_t and n_r are large enough. In this case, τ_k is an asymptotically zero mean Gaussian noise of variance and one can easily derive performance measures such as BER or spectral efficiency with MMSE equalization. For example, the average probability or outage probability of error can be immediately derived. Hence, let q denote the outage probability and p_q the corresponding outage probability of error with QPSK uncoded constellations. In this respect, conjecture 1.5 is extremely useful and yields:

---

[14]The matrix inversion lemma states that for nay invertible matrix $\mathbf{F}$ and $\mathbf{E}$: $(\mathbf{D}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}^H)^{-1} = \mathbf{D} - \mathbf{D}\mathbf{F}(\mathbf{E} + \mathbf{F}^H\mathbf{D}\mathbf{F})^{-1}\mathbf{F}^H\mathbf{D}^H$

$$
\begin{aligned}
1 - q &= P\left(Q(\sqrt{\text{SINR}}) \le p_q\right) \\
&= P\left(\text{SINR} \le (Q^{-1}(p_q))^2\right) \\
&\approx \int_{-\infty}^{(Q^{-1}(p_q))^2} \sqrt{\frac{n_r}{2\pi\sigma^2}} e^{-\frac{n_r}{2\sigma^2}(\text{SINR}-\mu)^2} \, d\text{SINR} \\
&\approx 1 - Q\left(\frac{\sqrt{n_r}(Q^{-1}(p_q))^2 - \mu}{\sigma}\right)
\end{aligned}
$$

Therefore,

$$
p_q \approx Q\left(\sqrt{\frac{\sigma}{\sqrt{n_r}}} Q^{-1}(q) + \mu\right)
$$

The Gaussian behavior of the SINR is appealing as error-control codes which are optimal for the Gaussian channel will also be optimal for the MIMO channel when using the MMSE receiver.

2.3.3 SINR versus Capacity

Historical note: The MMSE receiver plays a central role in telecommunications. It seems that the MMSE estimator, which as its roots in Signal Processing, plays a key role in Information Theory. Nice discussions on the Shannon [12, 69] versus Wiener[70, 71] legacy are given by Forney [72] and Guo [73, 74]. It may seem strange that it took more than 50 years to discover quite fundamental relationships between the input output mutual information and the minimum mean square error of an estimate. Astonishingly, it is shown in [73] that the derivative of the mutual information (nats) with respect to the SNR is equal to half the MMSE. The MMSE receiver has also several attributes that makes it appealing for use. The MMSE receiver is known to generate a soft decision output which maximizes the output Signal to interference plus Noise ratio (see [75]) (whereas in a AWGN channel with no interference, the match filter maximizes the output SNR). This advantage combined with the low complex implementation of the receiver (due in part to its linearity) has triggered the search for other MMSE based receivers such such as the MMSE DFE. The MMSE DFE [76, 77] is at the heart of very famous schemes such as BLAST [7]. BLAST which stands for Bell Labs Layered Space Time, was invented by Foschini at Bell Labs in 1998. Note that spatial multiplexing was already introduced in a 1994 Standford University patent by A. Paulraj [78]: the idea is based on successive interference cancellation where each layer is decoded, re-encoded and subtracted from the transmitted signal. The fact that this approach is optimal with MMSE successive equalizations (with respect to the achievable spectral efficiency) dates back back to Varanasi and Guess[79][15]. The MMSE receiver is therefore at the heart of many schemes and studying the SINR distribution enables to design and understand the performance of many MMSE based receivers (MMSE DFE [76, 77, 80], MMSE Parallel Interference cancellation [81]) In the MIMO case, similar results can be proven [82].

Indeed, assuming that $\mathbf{x}$ and $\mathbf{n}$ are uncorrelated with one another, we have:

---

[15]The optimality follows in fact directly by a simple determinant identity.

$$\mathbf{R}_y = \mathbb{E}(\mathbf{y}\mathbf{y}^H) = \frac{\rho}{n_t}\mathbf{H}\mathbf{Q}\mathbf{H}^H + \mathbf{I} \tag{2.8}$$

$$\mathbf{R}_{xy} = \mathbb{E}(\mathbf{x}\mathbf{y}^H) = \sqrt{\frac{\rho}{n_t}}\mathbf{Q}\mathbf{H}^H \tag{2.9}$$

From the previous paragraph, the MMSE estimate of $\mathbf{x}$ and its covariance matrix are given by:

$$\hat{\mathbf{x}} = \mathbf{R}_{xy}\mathbf{R}_y^{-1}\mathbf{y} \tag{2.10}$$

and the covariance matrix of the MMSE receiver is:

$$
\begin{aligned}
\mathbf{R}_{\text{MMSE}} &= \mathbb{E}\left[(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})^H\right] \tag{2.11}\\
&= \mathbf{R}_{xy}\mathbf{R}_y^{-1}\mathbf{R}_{xy}^H - 2\mathbf{R}_{xy}\mathbf{R}_y^{-1}\mathbf{R}_{xy}^H + \mathbf{Q} \tag{2.12}\\
&= \mathbf{Q} - \mathbf{R}_{xy}\mathbf{R}_y^{-1}\mathbf{R}_{xy}^H \tag{2.13}
\end{aligned}
$$

It follows from the inversion lemma that:

$$
\begin{aligned}
\mathbf{R}_{\text{MMSE}}^{-1} &= \mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{R}_{xy}\left(\mathbf{R}_y - \mathbf{R}_{xy}^H\mathbf{Q}^{-1}\mathbf{R}_{xy}\right)^{-1}\mathbf{R}_{xy}^*\mathbf{Q}^{-1} \tag{2.14}\\
&= \mathbf{Q}^{-1} + \frac{\rho}{n_t}\mathbf{H}^H\left(\mathbf{R}_y - \mathbf{H}\mathbf{Q}\mathbf{H}^*\right)^{-1}\mathbf{H} \tag{2.15}\\
&= \mathbf{Q}^{-1} + \mathbf{H}^H\mathbf{H} \tag{2.16}
\end{aligned}
$$

Finally, the capacity is given by:

$$
\begin{aligned}
C &= \log_2\det\left(\mathbf{I} + \frac{\rho}{n_t}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right) \tag{2.17}\\
&= \log_2\det\left(\mathbf{I} + \frac{\rho}{n_t}\mathbf{H}^H\mathbf{H}\mathbf{Q}\right) \tag{2.18}\\
&= \log_2\det\left(\mathbf{Q}\right) + \log_2\det\left(\mathbf{Q}^{-1} + \mathbf{H}^H\mathbf{H}\right) \tag{2.19}
\end{aligned}
$$

Hence, the channel capacity can be rewritten:

$$C = \log_2\frac{\det\left(\mathbf{Q}\right)}{\det\left(\mathbf{R}_{\text{MMSE}}\right)} \tag{2.20}$$

The expression relates, in a simple manner, the channel capacity to the covariance matrix of the MMSE estimate of $\mathbf{x}$. In this form, the channel capacity formula has an intuitive appeal. In fact, the MMSE estimate $\hat{\mathbf{x}}$ lies (with high probability) in a "small cell" centered around the codeword $\mathbf{x}$. The volume of the cell is proportional to $\det(\mathbf{R}_{\text{MMSE}})$. The volume of the codebook space (in which $\mathbf{x}$ lies with high probability) is proportional to $\det(\mathbf{R_x})$. The ratio $\rho = \frac{\det(\mathbf{R_x})}{\det(\mathbf{R}_{\text{MMSE}})}$ gives the number of cells that can be packed into the codebook space without significant overlapping. The "center" of each such cell, the codeword, can be reliably detected, for instance, using $\hat{\mathbf{x}}$. As a consequence, one can communicate reliably using a codebook of size ρ, which contains $\log_2(\rho)$ information bits. This provides an intuitive motivation to the capacity formula, in the same vein as [83].

2.3.4   The Infamous SINR-Multiplexing Trade-Off

This section makes clear several notions used inappropriately by many authors. *****In October********* A scheme is said to achieve spatial multiplexing gain r if the data rate:

$$\lim_{\text{SNR}\to\infty} \frac{R(\text{SNR})}{\log \text{SNR}} = r \tag{2.21}$$

[84, 78]

A scheme is said to achieve diversity gain d if the data rate:

$$\lim_{\text{SNR}\to\infty} \frac{\log \text{P}_e(\text{SNR})}{\log \text{SNR}} = -d \tag{2.22}$$

*** ****** [85]

Historical note: Space-time codes have been introduced by Tarokh et al. [86, 87, 88] in order to provide the bandwidth efficiency predicted by MIMO information theory [9] for which he won the Waterman Prize in 2001. Note however that one of the first proposals to use multiple transmitters with time processing is due to Wittneben in 1991 in the context of multicast [89]: simultaneous transmission from several physically displaced network access points to a given terminal. The space-time coding scheme can be seen as a generalization of the work of Boutros and Viterbo [90, 91, 11, 92] dealing with multidimensional constellations for fading channels (note that unitary transformations for Multidimensional QAM constellations date back to Lang [93]). The real space-time coding revolution began in 1998 when Alamouti invented a two transmit antenna space-time block code [94] with maximum diversity and only linear processing at the receiver. The generalization of this scheme to more than two antennas is known as space-time block codes based on orthogonal design (STBC-OD)[95]. Note that STBC-OD (and also the well-known V-Blast scheme[96]) belong to a more general class of STBC known as linear dispersion space-time block codes [97]. LD codes use a linear modulation and the transmitted codeword is a linear combination over space and time of *dispersion matrices* with the transmitted symbols as combining coefficients. STBC-OD suffer a loss in performance compared to more advance coding schemes (space-time trellis codes,..). Note that recently, another scheme called Threaded Algebraic Space Time (TAST) coding [98, 99] has been proposed which guarantees full diversity and full rate with arbitrary number of transmit and receive antennas.

Chapter 3

Gaussian i.i.d Channel Model

3.1 Model

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

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

Imagine now that the modeler is in a situation where it 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, $\frac{1}{n_r n_t} \mathbb{E} \left(\sum_{i=1}^{n_r} \sum_{j=1}^{n_t} |h_{ij}|^2 \right) = E$. Knowing only this information, the modeler is faced with the following question: what is the consistent model one can make knowing only the energy E (but not the correlation even though it may exist) ? In other words, based on the fact that:

$$\int d\mathbf{H} \sum_{i=1}^{n_r} \sum_{j=1}^{n_t} |h_{ij}|^2 P(\mathbf{H}) = n_t n_r E \qquad \text{(Finite energy)} \qquad (3.1)$$

$$\int dP(\mathbf{H}) = 1 \qquad \text{(P}(\mathbf{H}) \text{ is a probability distribution)} \qquad (3.2)$$

What distribution $P(\mathbf{H})$[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 maximize the following expression using Lagrange multipliers with respect to P:

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

$$+ \beta \left[1 - \int d\mathbf{H} P(\mathbf{H}) \right]$$

---

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

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

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

then this yields:

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

with

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

One of the most important conclusions of the maximum entropy principle is that while we have only assumed the variance, these assumptions imply independent entries since the joint probability distribution $P(\mathbf{H})$ simplifies into products of $P(h_{ij})$. Therefore, based on the previous state of knowledge, the only maximizer of the entropy is the i.i.d one. This does not mean that we have supposed independence in the model. In the generalized $L(P)$ expression, there is no constraint on the independence. Another surprising result is that the distribution achieved is Gaussian. Once again, gaussianity is not an assumption but a consequence of the fact that the channel has finite energy. The previous distribution is the least informative probability density function that is consistent with the previous state of knowledge. When only the variance of the channel paths are 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.

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

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

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

Solving the previous equations yields the following probability distribution:

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

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

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

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

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

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

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

with for all $\{i,j,(i,j) \neq (p,k)\}$ $P(h_{ij}) = \frac{1}{\pi E_{ij}} e^{-\frac{\mid h_{ij} \mid^2}{E_{ij}}}$ and $P(h_{pk}) = \frac{1}{\pi E_{pk}} e^{-\frac{\mid h_{pk} - m_{pk} \mid^2}{E_{pk}}}$. 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[2]. In the general case, if N constraints are given on the expected values of certain functions $\int g_i(\mathbf{H})P(\mathbf{H})d\mathbf{H} = \alpha_i$ for $i = 1...N$, then the principle of maximum entropy attributes the following distribution [100]:

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

where the values of λ and λ_i (for $i = 1..N$) can be obtained by solving the constraint equations.

Although these conclusions are widely known in the Bayesian community, the author is surprised that many MIMO channel papers begin with: "let us assume a $n_r \times n_t$ matrix with Gaussian i.i.d entries...". No assumptions on the model should be made. Only the state of knowledge should be clearly stated at the beginning of each paper and the conclusion of the maximum entropy approach can be straightforwardly used.[3]

As a matter of fact, the Gaussian i.i.d model should not be "thrown" away but be extensively used whenever our information on the propagation conditions is scarce (we don't know in what environment we are transmitting our signal i.e the frequency, the bandwidth, WLAN scenario, we do not know what performance measure we target...)[4].

---

[2]The case where information is not given in terms of expected values is treated in chapter 5.

[3]"Normality is not an assumption of physical fact at all. It is a valid description of our state of information", Jaynes.

[4]In "The Role of Entropy in Wave Propagation" [101], Franceschetti et al. show that the probability laws that describe electromagnetic magnetic waves are simply maximum entropy distributions with appropriate moment constraints. They suggest that in the case of dense lattices, where the inter-obstacle hitting distance is small compared to the distance traveled, the relevant metric is non-Euclidean whereas in sparse lattices, the relevant metric becomes Euclidean as propagation is not constrained along the axis directions.

3.2 From Conjecture to Theorem:

3.2.1 Mutual Information

In this section, we recall some important results proved by Kamath et al.[62]:

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:*

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

and

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

with

$$
\alpha_{iid}(\gamma, \rho) = \frac{1}{2}[1 + \gamma + \frac{1}{\rho} - \sqrt{(1 + \gamma + \frac{1}{\rho})^2 - 4\gamma}]
$$

Note that α_{iid} is related to the Stieltjes transform $m_{f_{iid}}$ of the limiting eigenvalue distribution f_{iid} of $\frac{1}{n_t}\mathbf{H}^H\mathbf{H}$ through:

$$
\rho(1 - \alpha_{iid}) = m_{f_{iid}}(\frac{-1}{\rho}) = \int \frac{dF_{iid}(\lambda)}{\lambda + \frac{1}{\rho}}
$$

This theorem extends Telatar's result [9]. It is noteworthy to precise that in this case, $I(n_t, n_r, \rho) = C(n_t, n_r, \rho)$. The theorem has been proved using a lemma in [27] (recalled in the Appendix as Lemma 1) which deals with linear spectral statistics of the form:

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

where $(\lambda_1, ..., \lambda_{n_t})$ denotes the eigenvalues of matrix B_{n_t}, $F^{B_{n_t}}(\lambda) = \frac{1}{n_t} | \{j : \lambda_j \le \lambda\} |$ and l is a function on $[0, \infty[$. The proof of theorem 1 (provided in the appendix) follows the introductory example treated in [27] concerning the distribution of:

$$
T_N = \log(\det(\mathbf{S}_N))
$$

where

$$
\mathbf{S}_N = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_k \mathbf{x}_k^H \right)
$$

and $\mathbf{x}_k$ is a n dimensional mean zero random vector with i.i.d standard normal entries.

Note that in the high SNR regime ($\rho \to \infty$), $C(n_t, n_r, \rho)$ converges in distribution to a Gaussian random variable:

$$
n_t \mu = \min(n_t, n_r) \ln(\rho)
$$

$$
\sigma^2{}_{iid} = \begin{cases} -\ln\left(1 - \frac{\min(n_t, n_r)}{\max(n_t, n_r)}\right) & \text{if } n_t \ne n_r \\ \frac{1}{2}\ln(\rho) & \text{if } n_t = n_r \end{cases} \tag{3.3}
$$

The striking result that the mean value of the mutual information scales linearly with n_t or n_r by a factor of $\ln(\rho)$ is at the origin of the success story of MIMO systems (which dates back to 1995). In simple terms, suppose $n_t \le n_r$, it is as if there were n_t non interfering cables that connect the transmitting and receiving antennas (in a environment full of scatterers)!

3.2.2 MMSE SINR

In this section, we recall some important results proved by Tse et al. [102]:

Theorem 2 *With the Gaussian i.i.d model, as $n_t \to \infty$ with $n_r = \gamma n_t$*

$$\sqrt{n_t} \left(SINR^k - \mu_{iid}(\gamma, \rho) \right)$$

converges in distribution to a $N(0, \sigma^2(\gamma, \rho))$ random variable where:

$$\mu_{iid}(\gamma, \rho) = \frac{(\gamma - 1)}{2} \frac{\rho}{\gamma} - \frac{1}{2} + \sqrt{\frac{(1 - \gamma)^2}{4} \frac{\rho^2}{\gamma^2} + \frac{(1 + \gamma)}{2} \frac{\rho}{\gamma} + \frac{1}{4}}$$

and

$$\sigma^2_{iid}(\gamma, \rho) = \frac{2\mu_{iid}(1 + \mu_{iid})^2}{1 + \frac{\gamma(1 + \mu_{iid})^2}{\rho}} - 2\frac{\mu^2_{iid}}{\gamma}$$

There is a small problem with complex and real case.... This theorem extends results of Tse et al. [31][5] where it is shown that in a large system limit, the SINR of a user at the output of the MMSE receiver converges to a deterministic limit. The proof uses ideas of Silverstein [103]. In fact, for many linear receivers (MMSE, matched filter, decorrelator), asymptotic normality of the receiver output (in the general the ouput decision statistic) can be proven [104, 68].

3.3 How Far is Asymptotic?

How far is asymptotic, in other words to what extent can we apply the asymptotic formulas in the finite regime? The answer to this question depends on many factors such as SNR, ratio of the number of transmitting to receiving antennas but also the performance measure we target such as BER, mutual information. However, we would like to provide a rule of thumb on the minimum number of antennas which satisfy the asymptotic regime.

3.3.1 Mutual Information

In the following, we have plotted in Figure 3.1 and 3.2 respectively the mean and the variance of the mutual information for systems of 1×1, 2×2, 3×3, ...15×15 antennas at 10dB. One can observe that

- With 6 antennas, we are at 0.02% $\left(\frac{2.723 - 2.7225}{2.723} \right)$ of the asymptotic mean value while the variance is only at 1% $\left(\frac{1.6 - 1.58}{1.58} \right)$ of the asymptotic variance value.

- With 3 antennas, we are at 0.6% $\left(\frac{2.74 - 2.723}{2.723} \right)$ of the asymptotic mean value while the variance is only at 4% $\left(\frac{1.64 - 1.58}{1.58} \right)$ of the asymptotic variance value.

As far as MIMO mutual information is concerned, infinity is only a couple of antennas and the results can be immediately used for designing future mobile systems[6].

---

[5]For this contribution, Tse and Hanly received the IEEE Communications and Information Theory Society Joint Paper Award in 2001.

[6]Confirmation that "infinity" in multi-antenna systems can be met with realistic systems can also be found in [105]. In [106], the authors even joke on the fact that infinity "ain't what it used to be".
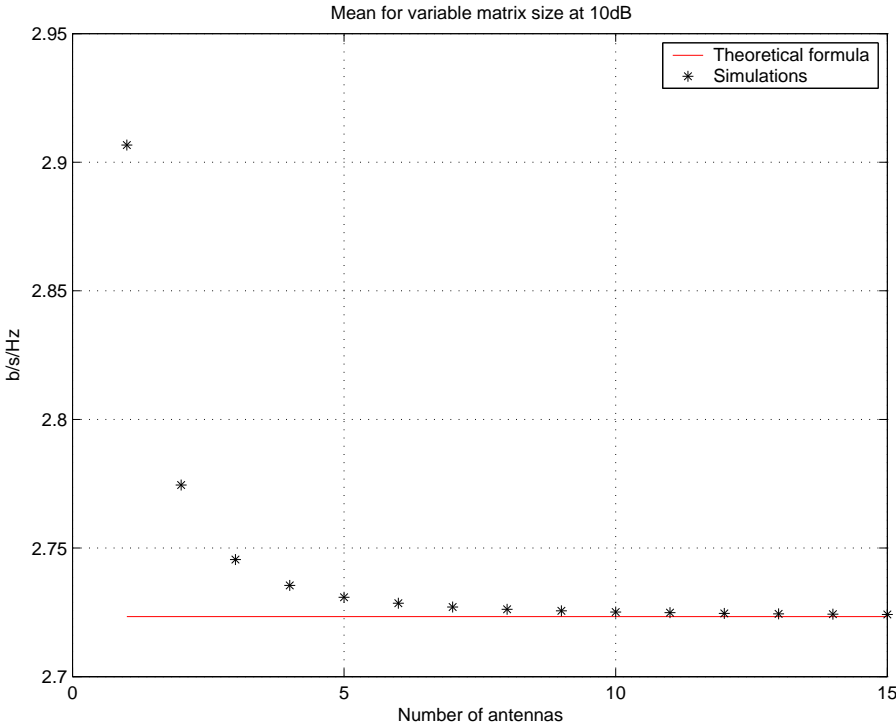
Figure 3.1: Empirical mean versus the number of antennas in the i.i.d Gaussian model.
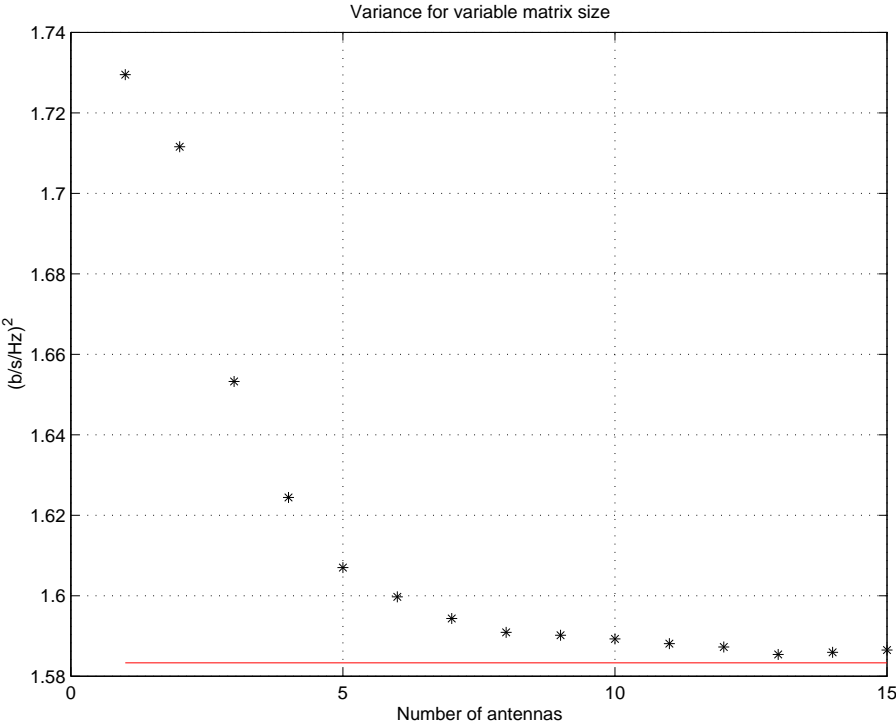
Figure 3.2: Empirical variance versus the number of antennas in the i.i.d Gaussian model.
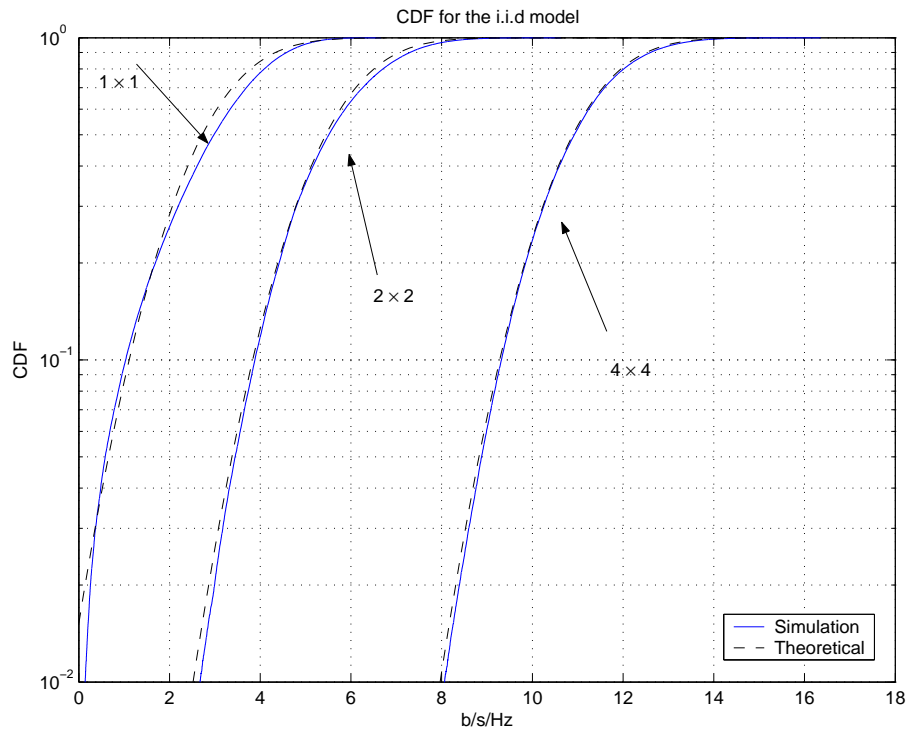
Figure 3.3: Mutual Information cumulative distribution function for the i.i.d Gaussian model

The asymptotic probability distribution of the capacity is given by:

$$P(C) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(C-n_t\mu)^2}{2\sigma^2}}$$

The CDF of the capacity is given by:

$$F(C) = 1 - Q(\frac{C - t\mu}{\sigma})$$

In Figure 3.3, the cumulative distribution function of the capacity is plotted for a system with 1×1, 2×2 and 4×4 antennas for a SNR of 10dB. There is a quite realistic match between the asymptotical theoretical formulas and the finite size simulated system for a 4×4 system which shows the usefulness of the random matrix approach. The reader must also note that similar curves can be found in the work of Biglieri et al. [105] and Hochwald et al. [63] (using results of central limit theorems involving $\log\left(\det(\mathbf{H}^H\mathbf{H})\right)$ [107])

3.3.2 MMSE SINR

In Figure 3.4 and Figure 3.5, we have plotted respectively the empirical mean and variance of the SINR versus the theoretical formulas of theorem 2. As one can observe, although the formulas are accurate, the asymptotic limit is reached for a number of antennas far greater than for the mutual information.

- With 6 antennas, we are at 22.3% ($\frac{3.4768-2.7016}{3.4768}$) of the asymptotic mean value while the variance is only at 42.57% ($\frac{28.9706-16.6379}{28.9706}$) of the asymptotic variance value.
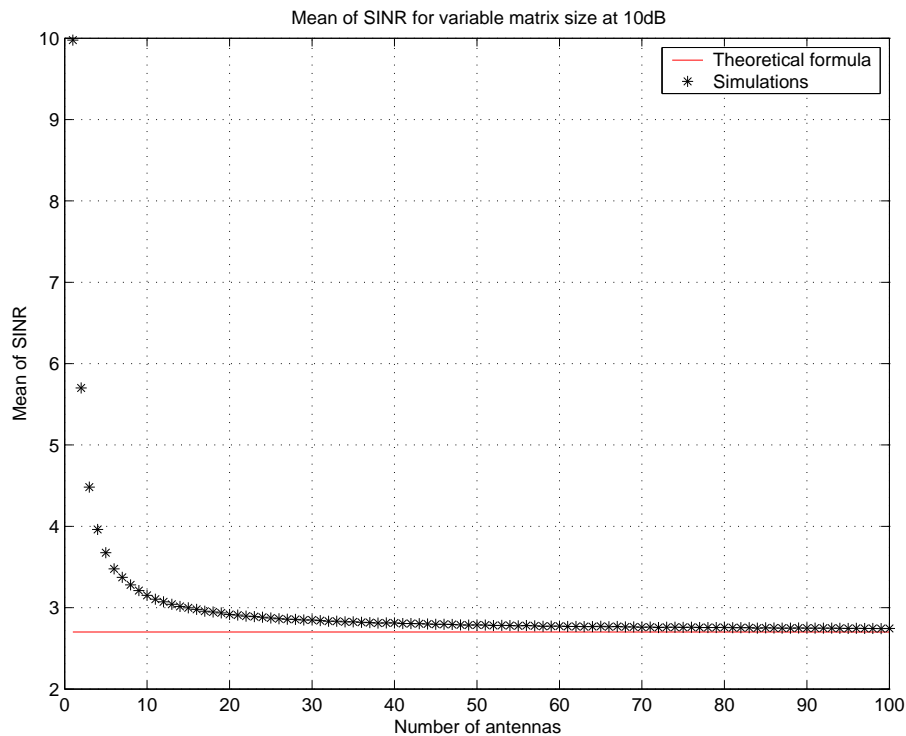
Figure 3.4: Empirical mean versus the number of antennas in the i.i.d Gaussian model

- With 12 antennas, we are at 12.08% $\left(\frac{3.0727-2.7016}{3.0727}\right)$ of the asymptotic mean value while the variance is only at 23.22% $\left(\frac{21.6695-16.6379}{21.6695}\right)$ of the asymptotic variance value.

This is mainly due to the convergence of order $\frac{1}{\sqrt{n_r}}$ whereas for the mutual information, the convergence was of the order $\frac{1}{n_r}$. As a consequence, from a practical standpoint, the random matrix approach is less appealing for the SINR amalysis. However, it gives us a hint on the number of antennas needed to obtain a Gaussian behavior. As a consequence, in all the following, we will focus on the mean behavior of the SINR and derive the mean.
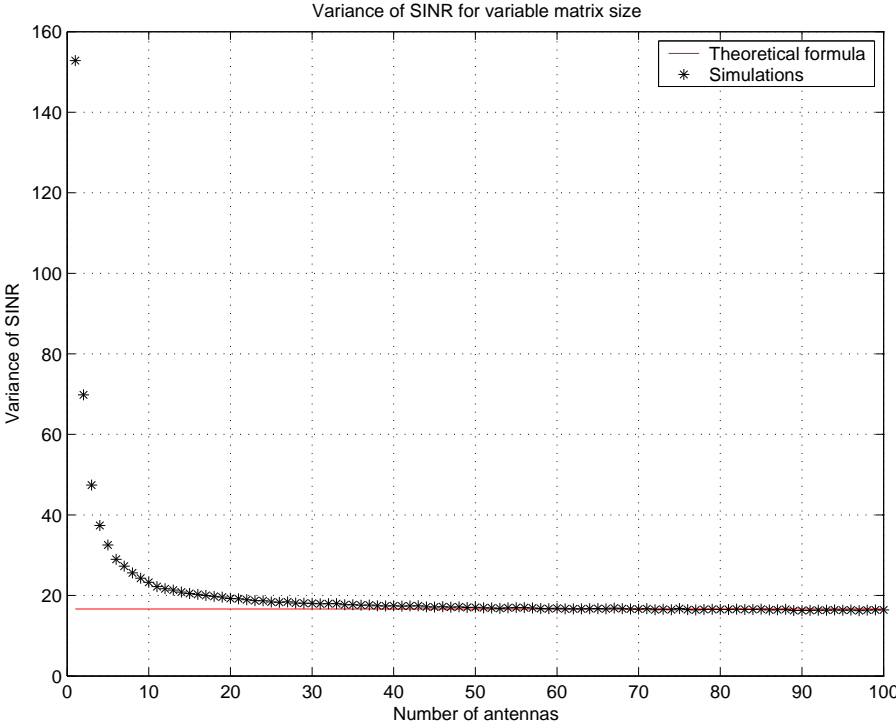
Figure 3.5: Empirical variance versus the number of antennas in the i.i.d Gaussian model

Chapter 4

More on Gaussian channels

4.1 Knowledge of the Existence of Correlation (but not the Value of the Correlation Matrix)

In many MIMO applications, one is aware of some correlation between the entries of the channel (due to reduced antenna spacing at the transmitter or/and the receiver for example) but does not know the value of the correlation matrix. However, as we will see afterward, such a knowledge is sufficient to determine a probability distribution for the MIMO matrix.

4.1.1 Energy unknown

Before considering the case of unknown correlation matrix, we will first consider a case similar to section 3.1 where the modeler is in a situation where it has no measurements and no knowledge where the transmission took place. The modeler does know that the channel carries some energy E but is not aware of its value.

In the case where the modeler knows the value of E, we have shown that:

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

In general, when E is unknown, the probability distribution is derived according to:

$$
\begin{aligned}
P(\mathbf{H}) &= \int P(\mathbf{H}, E) dE \\
&= \int P(\mathbf{H} \mid E) P(E) dE
\end{aligned}
$$

and is consistent with the case where E is known i.e $P(E) = \delta(E - E_0)$:

$$P(\mathbf{H}) = \frac{1}{(\pi E_0)^{n_r n_t}} \exp\{-\sum_{i=1}^{n_r} \sum_{j=1}^{n_t} \frac{\mid h_{ij} \mid^2}{E_0}\}$$

In the case were the energy E is unknown, one has to determine $P(E)$. E is a positive variance parameter and the channel can not carry more energy than what is transmitted (i.e

$E \leq E_{\max}$) . This is merely the sole knowledge the modeler has about E on which the modeler has to derive a prior distribution[1].

In this case, using maximum entropy arguments, one can derive $P(E)$:

$$P(E) = \frac{1}{E_{\max}} \qquad\qquad 0 \leq E \leq E_{\max}$$

As a consequence,

$$P(\mathbf{H}) = \int_0^{E_{\max}} \frac{1}{(\pi E)^{n_r n_t}} \exp\{-\sum_{i=1}^{n_r}\sum_{j=1}^{n_t} \frac{\mid h_{ij} \mid^2}{E}\}dE$$

With the change of variables $u = \frac{1}{E}$, we obtain:

$$P(\mathbf{H}) = \frac{1}{E_{\max}\pi^{n_r n_t}} \int_{\frac{1}{E_{\max}}}^{\infty} u^{n_r n_t - 2} e^{-\sum_{i=1}^{n_r}\sum_{j=1}^{n_t}|h_{ij}|^2 u} du$$

Note that the distribution is invariant by unitary transformations, is not Gaussian and moreover the entries are not independent when the modeler has no knowledge on the amount of energy carried by the channel. This point is critical and shows the effect of the lack of information on the exact energy[2].

In the case $n_t = 1$ and $n_r = 2$, we obtain:

$$P(\mathbf{H}) = \frac{1}{E_{\max}\pi^2 \sum_{i=1}^{2}\mid h_{i1}\mid^2} e^{-\frac{\sum_{i=1}^{2}|h_{i1}|^2}{E_{\max}}}$$

4.1.2 Correlation matrix unknown

Suppose now that the modeler knows that correlation exists between the entries of the channel matrix $\mathbf{H}$ but is not aware of the value of the correlation matrix $\mathbf{Q} = \mathbb{E}(\text{vec}(\mathbf{H})\text{vec}(\mathbf{H})^H)$. What consistent distribution should the modeler attribute to the channel based only on that knowledge?

To answer this question, suppose that the correlation matrix $\mathbf{Q} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ is known ($\mathbf{V} = [\mathbf{v}_1, ...\mathbf{v}_{n_r n_t}]$ is a $n_r n_t \times n_r n_t$ unitary matrix whereas $\mathbf{\Lambda}$ is a $n_r n_t \times n_r n_t$ diagonal matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, ..., \lambda_{n_r n_t})$ with $\lambda_i \geq 0$ for $1 \leq i \leq n_r n_t$).

Using the maximum entropy principle, one can easily show that:

$$P(\mathbf{H} \mid \mathbf{V}, \mathbf{\Lambda}) = \frac{1}{\prod_{i=1}^{n_r n_t} \pi\lambda_i} \exp\{\sum_{i=1}^{n_r n_t} \frac{\mid \mathbf{v_i}^H \text{vec}(\mathbf{H}) \mid^2}{\lambda_i}\}$$

---

[1]Jeffrey [49] already in 1939 proposed a way to handle this issue based on invariance properties and consistency axioms. He suggested that a proper way to express incomplete ignorance of a continuous variable known to be positive is to assign uniform prior probability to its logarithm, in other words: $P(E) \propto \frac{1}{E}$. However, the distribution is improper and one can not therefore marginalize with this distribution.

[2]In general, closed form solutions of the distributions do not exist. In this case, a powerful tool for approximate Bayesian inference that uses Markov Chain Monte Carlo to compute marginal posterior distributions of interest can be used through WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml.)

The channel distribution can be obtained:

$$
\begin{aligned}
P(\mathbf{H}) &= \int P(\mathbf{H}, \mathbf{V}, \mathbf{\Lambda}) d\mathbf{V} d\mathbf{\Lambda} \\
&= \int P(\mathbf{H} \mid \mathbf{V}, \mathbf{\Lambda}) P(\mathbf{V}, \mathbf{\Lambda}) d\mathbf{V} d\mathbf{\Lambda}
\end{aligned}
$$

If the correlation matrix is perfectly known, then $P(\mathbf{V}, \mathbf{\Lambda}) = \delta(\mathbf{V} - \mathbf{V}^0)\delta(\mathbf{\Lambda} - \mathbf{\Lambda}^0)$ and

$$
P(\mathbf{H}) = \frac{1}{\prod_{i=1}^{n_r n_t} \pi \lambda^0_i} \exp\{ \sum_{i=1}^{n_r n_t} \frac{\mid \mathbf{v^0_i}^H \mathrm{vec}(\mathbf{H}) \mid^2}{\lambda^0_i} \}
$$

In the case were the correlation matrix $\mathbf{Q}$ is unknown, one has to determine $P(\mathbf{V}, \mathbf{\Lambda}) = P(\mathbf{\Lambda} \mid \mathbf{V})P(\mathbf{V})$. This is the problem of constructing an ignorance prior corresponding to ignorance of both scale (up to some constraints proper to our problem) and rotation.

Let us first determine $P(\mathbf{\Lambda} \mid \mathbf{V})$: In the coordinate system of $\mathbf{V}$, each individual variance parameter λ_i is judged independent and positive. Moreover, each variance parameter λ_i is bounded by some extreme value $\lambda_{\max}$ (otherwise, the channel could carry infinite energy). This is merely the sole knowledge the modeler has. In this case, expressing complete ignorance yields:

$$
P(\mathbf{\Lambda} \mid \mathbf{V}) = P(\lambda_1)...P(\lambda_{n_r n_t})
$$

and

$$
P(\lambda_i) = \frac{1}{\lambda_{\max}} \qquad\qquad 0 \leq \lambda_i \leq \lambda_{\max}
$$

Let us now determine $P(\mathbf{V})$: $\mathbf{V}$ is a unitary matrix, in particular each column of the matrix has unit norm. Based solely on this state of knowledge and using the maximum entropy principle, $\mathbf{V}$ is unitary haar distributed (its distribution does not change by right or left multiplication by deterministic unitary matrices. The matrix $\mathbf{V}$ can be generated by Gramm Schmidt orthogonalization of an i.i.d Gaussian matrix). As a consequence,

$$
\begin{aligned}
P(\mathbf{H}) &= \int P(\mathbf{H}, \mathbf{V}, \mathbf{\Lambda}) d\mathbf{V} d\mathbf{\Lambda} \\
&= \int \frac{1}{\prod_{i=1}^{n_r n_t} \pi \lambda_{\max} \lambda_i} \exp\{ \sum_{i=1}^{n_r n_t} \frac{\mid \mathbf{v_i}^H \mathrm{vec}(\mathbf{H}) \mid^2}{\lambda^0_i} \} d\lambda_1....d\lambda_{n_r n_t} P(\mathbf{V}) d\mathbf{V} \\
&= \int \frac{1}{\pi^{n_t n_r} \lambda_{\max}^{n_t n_r}} \prod_{i=1}^{n_t n_r} \left(\mathrm{Ei}(\frac{\mid \mathbf{v_i}^H \mathrm{vec}(\mathbf{H}) \mid^2}{\lambda_{\max}}) \right) P(\mathbf{V}) d\mathbf{V}
\end{aligned}
$$

with $\mathrm{Ei}(x) = \int_x^\infty \frac{e^{-u}}{u} du$

In this case, the distribution is not invariant by unitary transform.

4.2 Multiple versus Single Antenna: is there a Contradiction?

4.2.1 Statement of the problem

[108, 109] In section 3.1, we have shown that if the modeler has knowledge of the line of sight component m_{ij} between the transmitting antenna i and receiving antenna j, then the modeler obtains the following distribution (also known as MIMO Rice distribution):

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

Astonishingly, although a Rice distribution is well known to enhance the performance with respect to the Rayleigh one in the SISO case, these results cannot be straightforwardly extended to the MIMO case. Indeed, consider the following example:

Example 1 *Suppose that the channel matrix is deterministic with equal entries 1 (this is the limit case of a Rice distribution with variance 0). The mutual information per transmitting antenna with input Gaussian entries and covariance matrix $\mathbb{E}(\mathbf{x}\mathbf{x}^H) = \mathbf{I}_{n_t}$ is:*

$$\begin{aligned} I^M &= \frac{1}{n_t} \log det(\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H). \\ &= \frac{1}{n_t}\sum_{i=1}^{n_r} \log(1 + \frac{\rho}{n_t}\lambda_i) \end{aligned}$$

In this case, since $\mathbf{H}\mathbf{H}^H$ is rank one, it has one single eigenvalue equal to $n_r n_t$ and the mutual information is given by:

$$\begin{aligned} I^M &= \frac{1}{n_t} \log(1 + \rho n_r) \\ &\to 0 \end{aligned}$$

when $n_r \to \infty$ and $\frac{n_t}{n_r} \to \gamma$.

This example shows that the line of sight component has a dramatic effect on the mutual information since it is well known that in the zero mean i.i.d Gaussian case, the mutual information per transmitting antenna (see section 3.1) is constant.

In fact, the analysis of Rice MIMO models is quite important as it determines the way antennas should be placed. In light of the previous result, one could conclude that it would be better to place the transmitting antenna not in line of sight of the receiving one but to effectively "hide" the transmitting antenna (under a table for example) in order to increase the scattering effect![3]

4.2.2 Some Considerations on MIMO Rice Channels

As a consequence, a more profound analysis should be conducted for determining the parameters governing the performance of the Rice distribution with respect to the i.i.d Gaussian case. Before going into detail, let us introduce more precisely the MIMO Rice model: The complex entries of $\mathbf{H}$ are independently Gaussian distributed with identical variance and mean $\mathbb{E}(h_{ij}) = m_{ij}$. Denoting by K the Rice factor of the channel, we rewrite the channel matrix $\mathbf{H}$ as

$$\mathbf{H} = \sqrt{\frac{K}{K+1}}\mathbf{H}^{\text{LOS}} + \sqrt{\frac{1}{K+1}}\mathbf{H}^{\text{NLOS}}$$

---

[3] I often joke in my presentations on this point where I tell the attendees that MIMO will not have only an impact on rate but also on the usual common sense of people!

in order to separate the random component of the channel from the deterministic part:

- $\mathbf{H}^{\mathrm{LOS}}$ represents the line of sight component of the channel such as $\|\mathbf{H}^{\mathrm{LOS}}\|_F^2 = n_t n_r$ with entries $h^{\mathrm{LOS}}{}_{ij} = \sqrt{\frac{K+1}{K}}\mu_{ij}$.

- $\mathbf{H}^{\mathrm{NLOS}}$ is the random component of the channel with Gaussian, independent and identically distributed entries. The complex element $h^{\mathrm{NLOS}}{}_{ij}$ is circularly symmetric, with zero mean and unit variance.

Note that the model is general enough to take into account line of sight (LOS) and non line of sight (NLOS) cases. Indeed, as $K \to \infty$, (4.1) models a deterministic channel, whereas for $K = 0$ it describes a Rayleigh fading channel.

We would like to predict the mutual information of a general Ricean MIMO channel using only a few meaningful parameters, namely the asymptotic eigenvalue distribution of the mean matrix $\mathbf{H}^{\mathrm{LOS}}$, the Ricean factor K, the SNR, and the number of receive antennas per transmit antenna $\gamma = \frac{n_r}{n_t}$. Results in this case are based on random matrix theory [110] and use the following assumption

Assumption: As $n_r, n_t \to \infty$ with constant ratio $\gamma = \frac{n_r}{n_t}$, the sequence of the empirical eigenvalue distribution of the matrix $\frac{\mathbf{H}^{\mathrm{LOS}}\mathbf{H}^{\mathrm{LOS}H}}{n_t}$ is assumed to converge in distribution to a deterministic limit function $F_{\frac{\mathbf{H}^{\mathrm{LOS}}}{\sqrt{n_t}}}$.

In this case, let us recall some important results of Cottatellucci et al. [111, 43]:

Theorem 3 *As $n_r = \gamma n_t \to \infty$ with $\frac{n_r}{n_t} \to \gamma$, the asymptotic mutual information per transmitting antenna with Gaussian input entries and covariance matrix $\mathbf{Q} = \mathbf{I}_{n_t}$ converges almost surely to a deterministic value:*

$$I^M = \frac{1}{\gamma} \int_0^\infty \log(1 + \rho\lambda) dF_{\frac{\mathbf{H}}{\sqrt{n_t}}}(\lambda)$$

$F_{\frac{\mathbf{H}^{LOS}}{\sqrt{n_t}}}(\lambda)$ is the limit distribution function of the eigenvalues of $\frac{\mathbf{H}\mathbf{H}^H}{n_t}$, whose Stieltjes transform $m_{\frac{\mathbf{H}}{\sqrt{n_t}}}(z)$ is the unique solution of the fixed point equation

$$m_{\frac{\mathbf{H}}{\sqrt{n_t}}}(z) = \int \frac{\mathrm{d}\, F_{\frac{\mathbf{H}^{LOS}}{\sqrt{n_t}}}(\lambda)}{\frac{K\lambda}{\gamma m_{\frac{\mathbf{H}}{\sqrt{n_t}}}(z)+K+1} - z\left(\frac{\gamma m_{\frac{\mathbf{H}}{\sqrt{n_t}}}(z)}{K+1} + 1\right) + \frac{1-\gamma}{K+1}} \tag{4.1}$$

such that $\mathrm{Im}(m_{\frac{\mathbf{H}}{\sqrt{n_t}}}(z)) > 0$ for $\mathrm{Im}(z) > 0$.

Remarkably, I^M is completely determined knowing only $F_{\frac{\mathbf{H}}{\sqrt{n_t}}}(\lambda)$, γ, ρ, and K and not the particular fluctuations of the fading.

Remark

- In the case $K \to \infty$, equation (4.1) simplifies to

$$m_{\frac{\mathbf{H}}{\sqrt{n_t}}}(z) = \int \frac{\mathrm{d}\, F_{\frac{\mathbf{H}}{\sqrt{n_t}}}(\lambda)}{\lambda - z}$$

which is nothing else then the Stieltjes transform of the distribution of the line of sight component.

- In the case $K \to 0$, equation (4.1) simplifies to

$$m_{\frac{\mathbf{H}}{\sqrt{n_t}}}(z) = \frac{1}{-z(\gamma m_{\frac{\mathbf{H}}{\sqrt{n_t}}}(z) + 1) + (1 - \gamma)}$$

which yields

$$\mathrm{d}\, F_{\frac{\mathbf{H}}{\sqrt{n_t}}}(\lambda) = \begin{cases} [1 - \frac{1}{\gamma}]^+ \delta(\lambda) + \frac{1}{\pi \gamma \lambda} \sqrt{\lambda - \frac{1}{4}(\lambda - 1 - \frac{1}{\gamma})^2} & \text{if}(\sqrt{\frac{1}{\gamma}} - 1)^2 \leq \lambda \leq (\sqrt{\frac{1}{\gamma}} + 1)^2 \\ 0 & \text{otherwise} \end{cases}$$

where $[z]^+ = \max(0, z)$. In this case, one obtains the solution of the i.i.d zero mean Gaussian channel of section 3.2.1.

4.2.3 The infamous diversity versus path loss trade-off

So should the user "hide under the table to communicate"? In fact, the result depends mostly on how the mean matrix (through its limiting singular value decomposition) is structured as revealed by equation 4.1. Even the example 1 is tricky: one cannot compare Rice and Rayleigh at the same SNR ρ. Indeed, in the case of line of sight, the path loss incurred by the Rice distribution is less dramatic than in the case of non-line of sight and as a consequence favors the Rice model. However, a question the reader could ask is whether there is a trade-off between the degrees of freedom of the channel (through the rank of $\mathbf{H}$) and the path loss factor (which influences ρ).
In order to answer this question, consider these two extreme cases:

- The $n \times n$ Rice matrix has rank one (with zero variance) and the path loss factor $\rho = \frac{\rho_{\max}}{r^2}$ (path loss model in free space, r is the distance). In this case, the total mutual information at high SNR is given by: $\ln(\frac{n\rho_{\max}}{r^2})$.

- The Rayleigh case has a path loss factor $\rho = \frac{\rho_{\max}}{r^l}$ ($l \geq 2$). In this case, the total mutual information at high SNR is given by: $n\ln(\frac{\rho_{\max}}{r^l})$.

At high SNR, the path loss factor influences the slope of the mutual information (with respect to the distance). The intersection of the two cases is given by:

$$n\ln(\rho_{\max}) - nl\ln(r) = \ln(\rho_{\max}) + \ln(n) - 2\ln(r)$$

which gives for a very high number of antennas: $r \sim \rho_{\max}^{\frac{1}{l}}$

As a consequence, there is a trade-off between diversity and the path loss factor which depends mainly on how far the receiving antenna is with respect to the transmitting one. However, in practical cases, this trade-off has small chances to be effective for two reasons:

- The Rice matrix is rarely rank one and the Rayleigh matrix is not full rank due to correlations.

- The previous examples assume that one is operating at very high SNR ($\rho_{\max} >> r^l$), a case that is not always fulfilled.

Chapter 5

Knowledge of the Directions of Arrival or Departure

5.1 Model

The modeler[1] is interested in modelling the channel over time scales over which the locations of scatterers do not not change significantly relative to the transmitter or receiver. This is equivalent to considering time scales over which the channel statistics do not change significantly. However, the channel realizations do vary over such time scales. Imagine that the modeler is in a situation where it knows the energy carried by the channel (nothing is known about the mean)[2]. Moreover, the modeller knows from electromagnetic theory that when a wave propagates from a scatterer to the receiving antennas, the signal can be written in an exponential form

$$\mathbf{s}(t, \mathbf{d}) = \mathbf{s_0}\, e^{j(\mathbf{k}^T \mathbf{d} - 2\pi f t)} \tag{5.1}$$

which is the plane wave solution of the Maxwell equations in free non-dispersive space for wave vector $\mathbf{k} \in \mathbb{R}^{2\times 1}$ and location vector $\mathbf{d} \in \mathbb{R}^{2\times 1}$. The reader must note that other solutions to the Maxwell equations exist and therefore the modeler is making an important restriction. The direction of the vector $\mathbf{s_0}$ gives us knowledge on the polarization of the wave while the direction of the wave vector $\mathbf{k}$ gives us knowledge on the direction of propagation. The phase of the signal results in $\phi = \mathbf{k}^T \mathbf{d}$. The modeler considers for simplicity sake 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 such a thing. 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 (since the modeler does not want to introduce unjustified information). In other words, based on the fact that:

---

[1]We treat in this section thoroughly the directions of arrival model and show how the directions of departure model can be easily obtained from the latter case.

[2]The case where the paths have different non-zero means can be treated the same way.

$$\mathbf{H} = \frac{1}{\sqrt{s_r}} \begin{pmatrix} e^{j\phi_{1,1}} & \cdots & e^{j\phi_{1,s_r}} \\ \vdots & \ddots & \vdots \\ e^{j\phi_{n_r,1}} & \cdots & e^{j\phi_{n_r,s_r}} \end{pmatrix} \mathbf{\Theta}_{s_r \times n_t}$$

what distribution should the modeler attribute to $\mathbf{\Theta}_{s_r \times n_t}$? $\mathbf{H}$ is equal to $\frac{1}{\sqrt{s_r}}\mathbf{\Phi\Theta}$, $\phi_{i,j} = \mathbf{k}.\mathbf{r}_{i,j}$ and $\mathbf{r}_{i,j}$ is the distance between the receiving antenna i and receiving scatterer j and $\mathbf{\Phi}$ is a $n_r \times s_r$ matrix (s_r is the number of scatterers) which represents the directions of arrival from randomly positioned scatterers to the receiving antennas. $\mathbf{\Theta}_{s_r \times n_t}$ is an $s_r \times n_t$ matrix which represents the scattering environment between the transmitting antennas and the scatterers (see Figure 5.1).

The consistency argument (see Proposition 1) states that if the DoA (Directions of Arrival) are unknown then $\mathbf{H} = \frac{1}{\sqrt{s_r}}\mathbf{\Phi}_{n_r \times s_r}\mathbf{\Theta}_{s_r \times n_t}$ should be assigned an i.i.d Gaussian distribution (see section 3.1) since the modeler is in the same state of knowledge as before where it only knew the variance.

Based on the previous remarks, let us now derive the distribution of $\mathbf{\Theta}_{s_r \times n_t}$. The probability distribution $P(\mathbf{H})$ is given by:

$$P(\mathbf{H}) = \int P(\mathbf{\Phi\Theta} \mid \mathbf{\Phi}, s_r)P(\mathbf{\Phi} \mid s_r)P(s_r)ds_r d\mathbf{\Phi}$$

- When $\mathbf{\Phi}$ and s_r are known, then $P(\mathbf{\Phi} \mid s_r) = \delta(\mathbf{\Phi} - \mathbf{\Phi}^0)$ and $P(s_r) = \delta(s_r - s_r{}^0)$. Therefore $P(\mathbf{H}) = P(\mathbf{\Phi}^0\mathbf{\Theta})$.

- When $\mathbf{\Phi}$ and s_r are unknown: the probability distribution of the frequency path h_{ij} is:

$$P(h_{ij}) = \int P\left(h_{ij} \mid \mathbf{\Phi}, s_r\right) P(\mathbf{\Phi} \mid s_r)P(s_r)d\mathbf{\Phi}ds_r \tag{5.2}$$

In the case when $P(\mathbf{\Phi} \mid s_r)$ and $P(s_r)$ are unknown, the consistency argument states that:

- The $\mathbf{\Theta}_{s_r \times n_t}$ matrix is such as each h_{ij} is zero mean Gaussian.
- The $\mathbf{\Theta}_{s_r \times n_t}$ matrix is such as $\mathbb{E}(h_{ij}h_{mn}{}^*) = \delta_{im}\delta_{jn}$ (since h_{ij} is Gaussian, decorrelation is equivalent to independence).

In this case, the following result holds:

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

Proof: Since $\mathbf{\Phi}$ is unknown, the principle of maximum entropy attributes independent uniformly distributed angles to each entry ϕ_{ij}:

$$P(\phi_{ij}) = \frac{1}{2\pi}\mathbf{1}_{[0,2\pi]}.$$

Let us show that $\mathbf{\Theta}_{s_r \times n_t}$ i.i.d zero mean with variance 1 is solution of the consistency argument.

Since $h_{ij} = \frac{1}{\sqrt{s_r}} \sum_{k=1}^{s_r} \theta_{kj} e^{j\phi_{ik}}$ then $P(h_{ij} \mid \mathbf{\Phi}, s_r) = N(0, \frac{1}{s_r} \sum_{k=1}^{s_r} \mid e^{j\phi_{ik}} \mid^2 = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}}$ and therefore h_{ij} is zero mean Gaussian since:

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

Moreover, we have:

$$
\begin{aligned}
\mathbb{E}(h_{ij} h_{mn}^*) &= \mathbb{E}_{\mathbf{\Theta}, \mathbf{\Phi}} \left(\frac{1}{\sqrt{s_r}} \sum_{k=1}^{s_r} \theta_{kj} e^{j\phi_{ik}} \frac{1}{\sqrt{s_r}} \sum_{l=1}^{s_r} \theta_{ln}^* e^{-j\phi_{ml}} \right) \\
&= \frac{1}{s_r} \sum_{k=1}^{s_r} \sum_{l=1}^{s_r} \mathbb{E}_{\mathbf{\Theta}}(\theta_{kj} \theta_{ln}^*) \mathbb{E}_{\mathbf{\Phi}}(e^{j\phi_{ik} - j\phi_{ml}}) \\
&= \frac{1}{s_r} \sum_{k=1}^{s_r} \sum_{l=1}^{s_r} \delta_{kl} \delta_{jn} \mathbb{E}_{\mathbf{\Phi}}(e^{j\phi_{ik} - j\phi_{ml}}) \\
&= \delta_{jn} \frac{1}{s_r} \sum_{k=1}^{s_r} \mathbb{E}_{\mathbf{\Phi}}(e^{j\phi_{ik} - j\phi_{mk}}) \\
&= \delta_{jn} \delta_{im}
\end{aligned}
$$

which proves that $\mathbf{H}$ is i.i.d Gaussian for unknown angles.

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. The maximum entropy approach is appealing in the sense that if correlated scattering is given as a prior knowledge, then it can be immediately integrated in the channel modelling approach (as a constraint on the covariance matrix for example). Note also that in this model, the entries of $\mathbf{H}$ are correlated for general DoA's.

Suppose now that the modeler assumes that the different steering vectors have different amplitudes $\sqrt{P_i^r}$. What distribution should the modeler attribute to the matrix $\Theta_{s_r \times n_t}$ in the following representation:

$$
\mathbf{H} = \frac{1}{\sqrt{s_r}} \begin{pmatrix} e^{j\phi_{1,1}} & \dots & e^{j\phi_{1,s_r}} \\ \vdots & \ddots & \vdots \\ e^{j\phi_{n_r,1}} & \dots & e^{j\phi_{n_r,s_r}} \end{pmatrix} \begin{pmatrix} \sqrt{P_1^r} & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_r}^r} \end{pmatrix} \Theta_{s_r \times n_t} ?
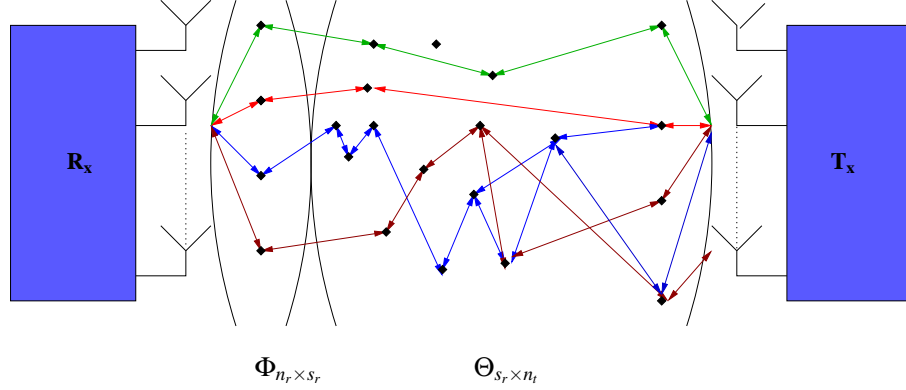$$

Figure 5.1: Directions of arrival based model.

Proposition 2 $\boldsymbol{\Theta}_{s_r \times n_t}$ *i.i.d Gaussian with variance 1 is solution of the consistency argument and maximizes entropy*

Proof: We will not go into the details as the proof is a particular case of the proof of Proposition 6.

5.2 Mutual Information

5.2.1 General Case

We recall hereafter the general DoA based model:

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

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

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

Let us first make some assumptions[3] on the matrix of the directions of arrival.

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

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

Theorem 4 *As $n_t \rightarrow \infty$ with $s_r = \xi n_t$, $I^M{}_{doa}(n_t, n_r, s_r, \rho) - n_t \mu_{doa}(\xi, \gamma, \rho)$ converges in distribution to a $N(0, \sigma^2{}_{doa})$ random variable where:*

---

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

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

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

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

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

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

f_{doa} is the limiting eigenvalue distribution of $\frac{1}{n_t}\mathbf{H}^H\mathbf{H}$ in the DoA based model while S_{doa} is the limiting eigenvalue distribution of $\frac{1}{s_r}\mathbf{P^r}^{\frac{1}{2}}\mathbf{\Phi}^H_{n_r \times s_r}\mathbf{\Phi}_{n_r \times s_r}\mathbf{P^r}^{\frac{1}{2}}$. This result is based on contribution [27]. Hence, if the directions of arrival and their powers can be estimated, one can completely determine the distribution of the mutual information by solving the previous equations. From a practical point of view, the receiver estimates the angles of arrival and determines the mean and the variance of the mutual information. This information is then sent back to the transmitter for scheduling the network [4]. One interesting point of the feedback mechanism is that asymptotically only two values (the mean and the variance) are needed. This reduces drastically the overhead of feedback transmissions.

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

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

Therefore, $m_{f_{\text{doa}}}\left(-\frac{1}{\rho}\right) = \rho\left(1 - \rho\frac{d\mu_{\text{doa}}}{d\rho}\right)$ and based on the result of theorem 4, we have:

$$-\frac{1}{\rho} = \frac{-1}{\rho(1 - \rho(\frac{d\mu_{\text{doa}}}{d\rho}))} + \xi \int \frac{x}{1 + x\rho(1 - \rho(\frac{d\mu_{\text{doa}}}{d\rho}))} dS_{\text{doa}}(x) \quad (5.3)$$

and

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

In the high SNR regime, the following result holds:

---

[4]Some results on the capacity of a MIMO multi-user network (where all the users have different angles of arrival) in the large system limit (high number of antennas) can be found in [112].

Proposition 3 *In the high SNR regime, the mean mutual information of the DoA based model converges to:*

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

The integral is on the support of non-zero eigenvalues and $\int_{\lambda > 0} dS_{\text{doa}}(\lambda))$ expresses the correlation factor of the s_r scatterers. Note that a similar result, based on a different approach, can be found in [113].

Let $r = \rho \frac{d\mu_{\text{doa}}}{d\rho} \leq 1$ ($n_t r$ denotes in fact the multiplexing gain).
According to equation (5.3), we have:

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

and at high SNR:

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

which yields:

$$r = \begin{cases} \xi \int_{\lambda > 0} dS_{\text{doa}}(\lambda) & \text{if } \xi \int_{\lambda > 0} dS_{\text{doa}}(\lambda) \leq 1 \\ 1 & \text{otherwise} \end{cases} \tag{5.4}$$

and proves the result.

5.2.2 ULA and Fourier Directions Case

In this part, the modeler takes into account the geometry of the receiving antenna (as the modeler knows it) to derive the steering vectors: in the case of a Uniform Linear Array (ULA), the steering vector has the following form $[1, e^{-j2\pi \frac{d \sin(\phi)}{\lambda}}, ..., e^{-j2\pi \frac{d(n_r-1) \sin(\phi)}{\lambda}}]$ where d is the antenna spacing and ϕ is the direction of arrival.[5] The DoA ϕ of a source is defined as the angle between a line perpendicular to the incoming wave-front and a reference line through the array (see Figure 5.1).

$$\mathbf{H} = \frac{1}{\sqrt{s_r}} \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ e^{j2\pi \frac{d(n_r-1) \sin(\phi_1)}{\lambda}} & \cdots & e^{j2\pi \frac{d(n_r-1) \sin(\phi_{s_r})}{\lambda}} \end{pmatrix} \begin{pmatrix} \sqrt{P_1^r} & 0 & \cdots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_r}^r} \end{pmatrix} \Theta_{s_r \times n_t}$$

For simplicity sake, we will take $d = \frac{\lambda}{2}$. We will also suppose that $s_r \leq n_r$. 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 as $\sin(\phi_i) = \frac{2k}{n_r}$ (see Figure 5.2). This case can be seen as an extreme case where all the scatterers are maximally distant from each other called here the Fourier direction case.

---

[5]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 the antennas are not closely surrounded by obstacles.
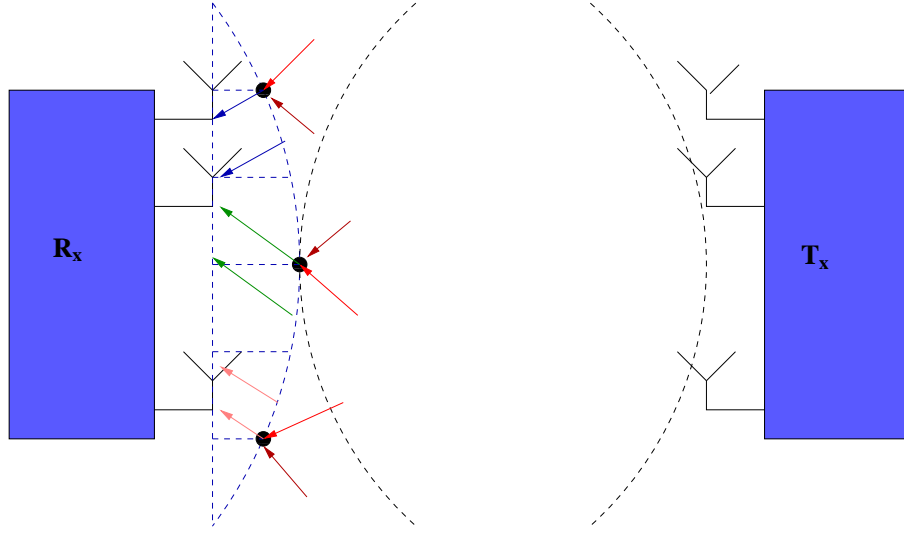
Figure 5.2: Simple case: Scatterers positioned on special directions.

Equal Power Case

In this case, we consider $\mathbf{P^r} = \mathbf{I}_{s_r \times s_r}$. As a consequence, the limiting eigenvalue distribution S_{doa} of $\frac{1}{s_r} \Phi^H_{n_r \times s_r} \Phi_{n_r \times s_r}$ has the following expression (since the column vectors of $\Phi_{n_r \times s_r}$ are orthogonal):

$$S_{\text{doa}}(\lambda) = \delta(\lambda - \gamma)$$

Proposition 4 *In the ULA and Fourier directions case, $\mu_{doa}(\xi, \gamma, \rho)$ and $\sigma^2_{doa}(\xi, \gamma, \rho)$ are equal to:*

$$\mu_{doa}(\xi, \gamma, \rho) = \xi \ln(1 + \rho\gamma - \rho\gamma\alpha_{doa}(\xi, \gamma, \rho)) + \ln(1 + \rho\xi\gamma - \rho\gamma\alpha_{doa}(\xi, \gamma, \rho)) - \alpha_{doa}(\xi, \gamma, \rho)$$

and

$$\sigma^2_{doa}(\xi, \gamma, \rho) = -\ln[1 - \frac{\alpha_{doa}{}^2(\xi, \gamma, \rho)}{\xi}]$$

with

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

Proof: One can notice that in this case, $\frac{1}{n_t} \mathbf{H}^H \mathbf{H} = \frac{1}{n_t s_r} \mathbf{\Theta}^H_{s_r \times n_t} \Phi^H_{n_r \times s_r} \Phi_{n_r \times s_r} \mathbf{\Theta}_{s_r \times n_t} = \frac{\gamma}{n_t} \mathbf{\Theta}^H_{s_r \times n_t} \mathbf{\Theta}_{s_r \times n_t}$. Therefore, the same result as the i.i.d (cf. theorem.1) can be applied if one does the following change:

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

$$\gamma \longrightarrow \xi$$

.

In the high SNR regime ($\rho \to \infty$), it can be easily shown that:
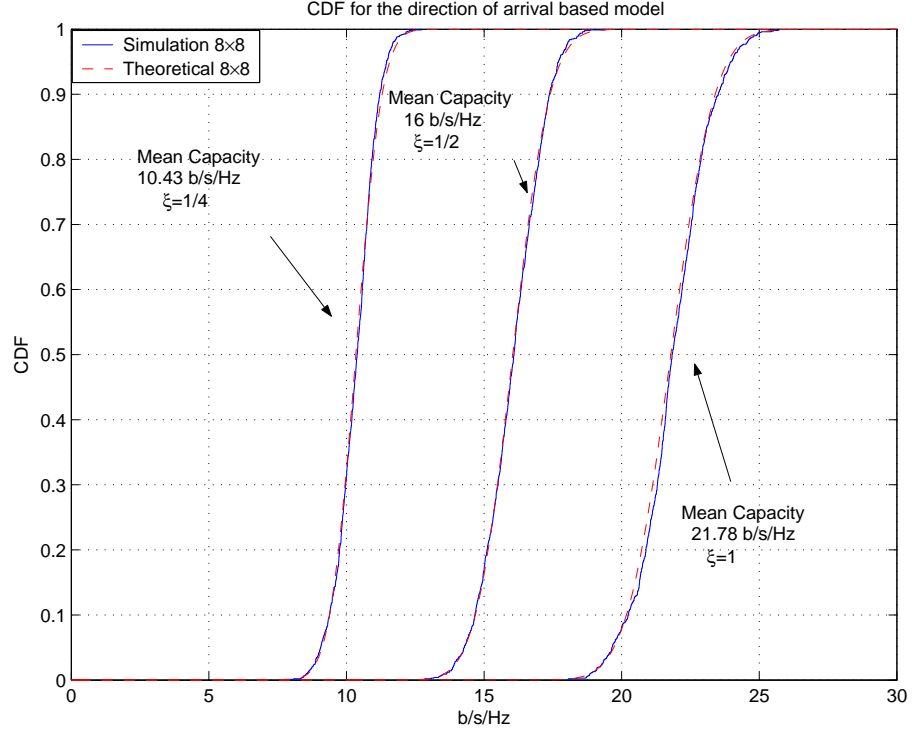
- $n_t \mu_{\text{doa}} = \min(s_r, n_t) \ln(\rho)$

Figure 5.3: Mutual information cumulative distribution for the DoA based model with an 8×8 MIMO system and different values of the ratio $\frac{s_r}{n_r}$.

-

$$\sigma^2{}_{\mathrm{doa}} = \begin{cases} -\ln\left(1 - \frac{\min(n_t, s_r)}{\max(n_t, s_r)}\right) & \text{if } n_t \neq s_r \\ \frac{1}{2}\ln(\rho) & \text{if } n_t = s_r \end{cases} \tag{5.5}$$

In Figure 5.3, simulations have been conducted with $n_r = n_t = 8$ antennas and an SNR of 10dB. In this case, $\frac{s_r}{n_r} = \xi = \frac{1}{\gamma}$. Three cases have been plotted $\xi = \frac{1}{4}$, (2 scatterers), $\xi = \frac{1}{2}$, (4 scatterers) and finally $\xi = 1$, (8 scatterers). A close match between the theoretical formulas and simulations is observed. We can also quantify the impact of the number of scatterers on the distribution of the mutual information. In Figure 5.4, the asymptotic variance of the mutual information is plotted versus $\xi = \frac{s_r}{n_t}$ for several values of SNR (SNR=5, 10 and 15 dB). For each SNR, the asymptotic variance has a maximum value. In Figures 5.5 and 5.6, the mean and the variance have been simulated versus $\frac{s_r}{n_t}$ for a system of 32×32 antennas and compared to the theoretical formula[6]. A close match between theory and simulations is also obtained in this case whatever the number of scatterers. In Figure 5.5, the asymptotic mean of the mutual information with respect to $\xi = \frac{s_r}{n_t}$ at 10dB shows that the number of scatterers does not yield a linear gain. The result also acknowledges the well known fact that the full transmission potential is obtained when the number of scatterers is equal to the number of antennas.

---

[6]The number of antennas has been chosen high (i.e 32) in order to have a wide range of scatterers ($s_r = 1..32$) and not because the result does not fit with 8 or 6 antennas.
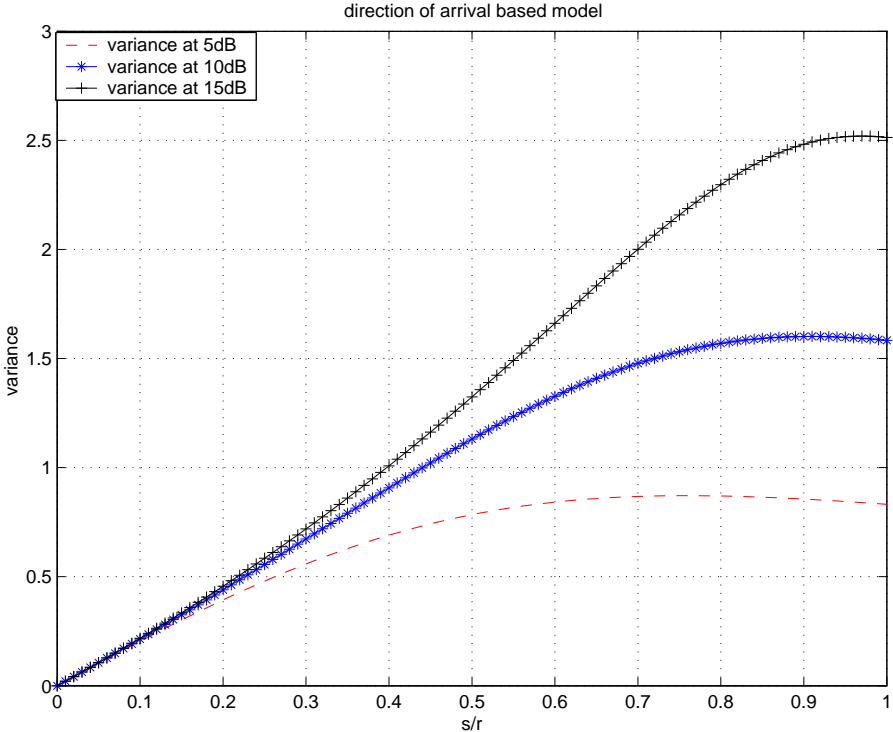
Figure 5.4: Variance of the mutual information versus ξ for the DoA based model with an 8×8 MIMO system.
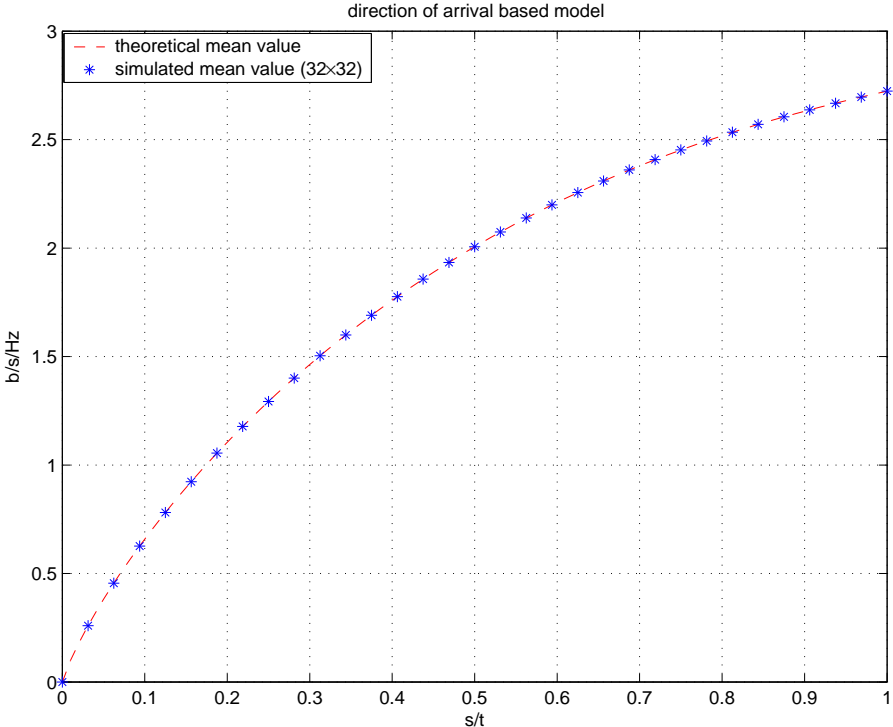
Figure 5.5: Theoretical versus simulated mean for a 32×32 system at 10dB for the DoA based model.
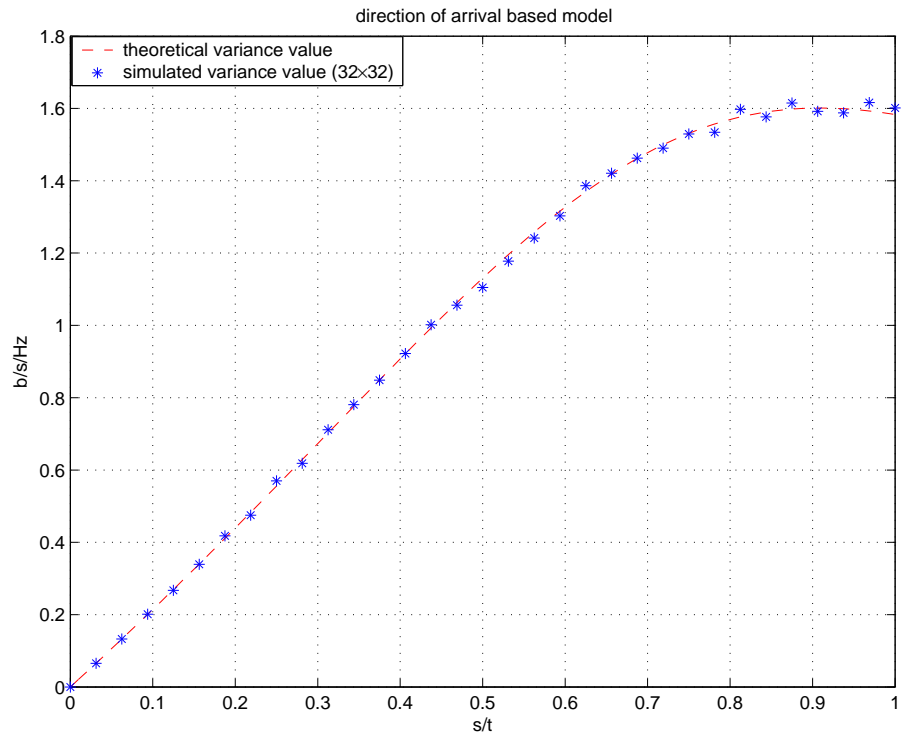
Figure 5.6: Theoretical versus simulated variance for a 32×32 system at 10dB for the DoA based model.

Non-equal Power Case

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 $\frac{1}{r}\mathbf{P^r}^{\frac{1}{2}}\mathbf{\Phi}_{n_r \times s_r}^H \mathbf{\Phi}_{n_r \times s_r}\mathbf{P^r}^{\frac{1}{2}}$ has the following expression:

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

Proposition 5 *In the non-equal power case with Fourier directions, $\mu_{doa}(\xi, \gamma, \rho)$ and $\sigma_{doa}^2(\xi, \gamma, \rho)$ are equal to:*

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

and

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

with

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

Proof The proof is provided in the appendix. For the mean μ_{doa}, the proof is an application of the general Proposition 11 in the case of interest and is provided in the appendix (before reading the proof, the reader is encouraged to read further the document until Proposition 11). For the variance, results of [27] are used.

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

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

In Figure 5.7 and Figure 5.8, 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 $\xi = \frac{s_r}{n_t} = 1$):

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

with

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

and

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

In Figure 5.7, the asymptotic mean mutual information has been plotted versus the amplitude 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
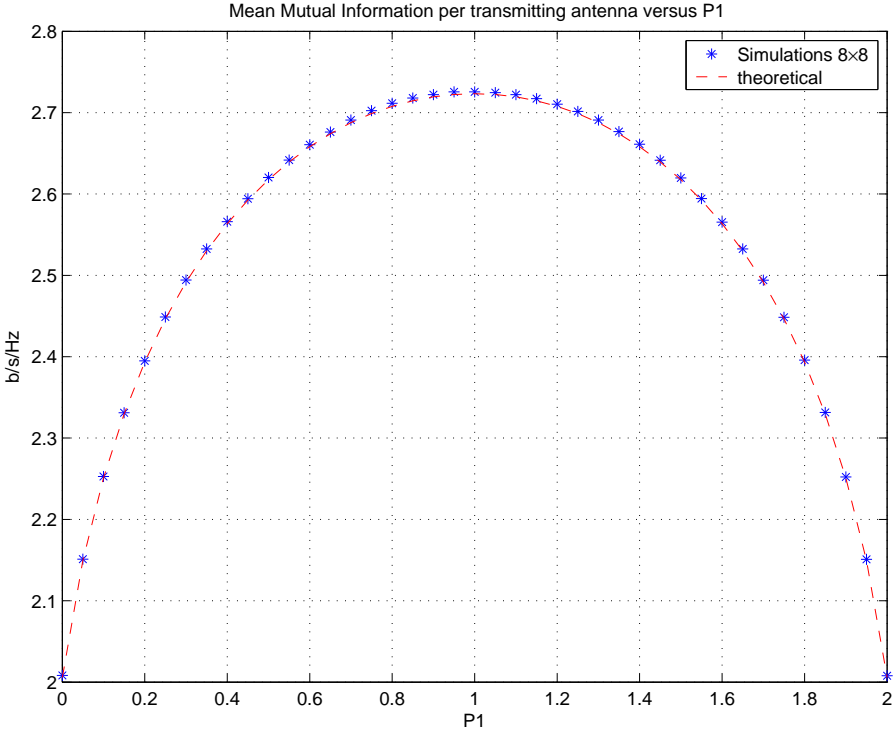
Figure 5.7: Mean capacity per transmitting antenna versus $P_1{}^r$ at 10dB for an 8×8 DoA based model.
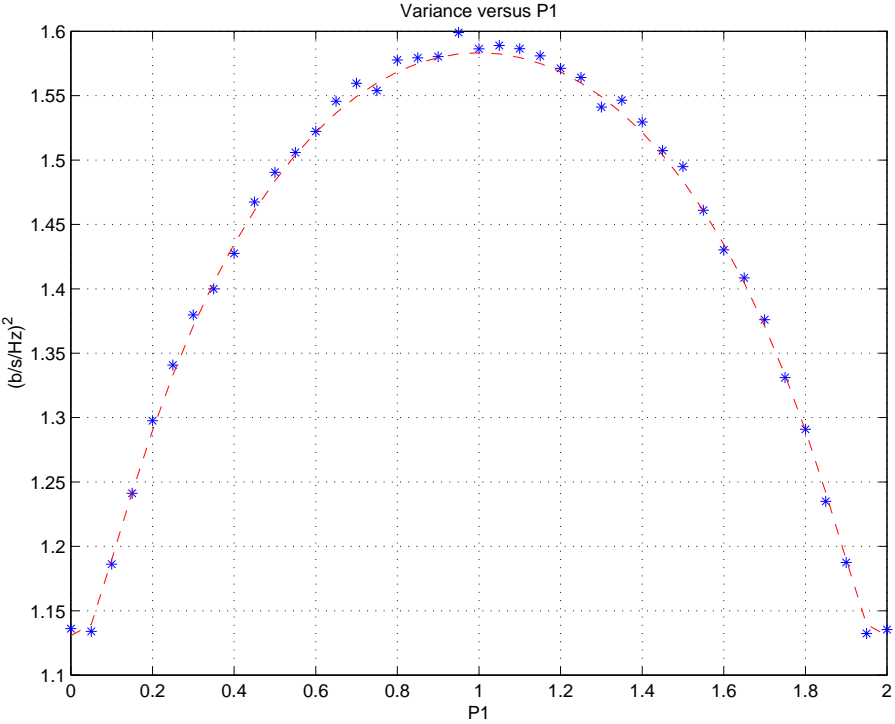
Figure 5.8: Variance versus $P_1{}^r$ at 10dB for an 8×8 DoA based model.
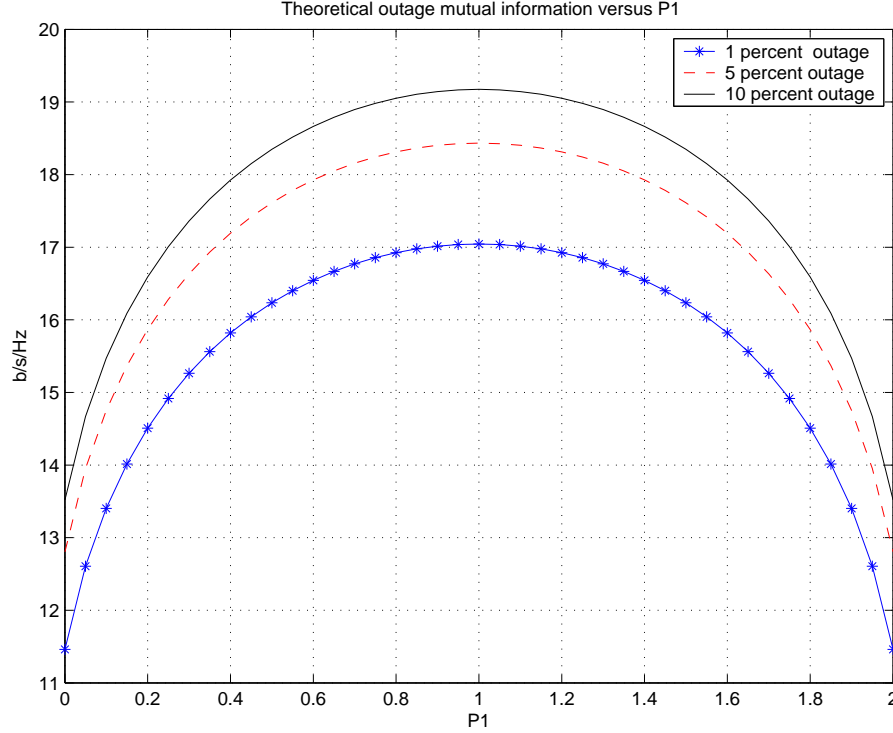
Figure 5.9: Outage mutual information versus $P_1{}^r$ at 10dB for an 8×8 DoA based model.

obtained when all the steering directions have equal power. The close match also pertains for the variance (see Figure 5.8) where the highest variance is obtained in the equal power case. In terms of outage mutual information, the equal power case is also the one which maximizes that criteria (see Figure 5.9 and proposition 12). Intuitively, one can easily understand this observation: any imbalances of power will reduce the effective number of scatterers and therefore the diversity generated by the environment

5.2.3   Fourier versus Random Directions: Equal Power Case

One important question is to know, for a given number of scatterers, the effect of the directions of departure on the mutual information. The answer has a direct impact on the understanding and the design of future mobile systems. We will consider here an extreme case where the exponential entries of matrix $\mathbf{\Phi}$ are independent and identically distributed random variables with zero mean and unit variance. The value of the angles do not change during the whole transmission. This is a limiting case of near field scattering (all the rays, for a given scatterer do not come from the same direction). In this case, the Stieltjes transform of the limiting eigenvalue distribution of $\frac{1}{s_r}\mathbf{\Phi}^H\mathbf{\Phi}$ is given by [114]:

$$m_{S_{\text{doa}}}(z) = \sqrt{\frac{(1-\gamma)^2}{4z^2} - \frac{(1+\gamma)}{2z} + \frac{1}{4}} - \frac{1}{2} - \frac{(1-\gamma)}{2z}$$

Using equation (5.3), one has to solve the following equation to determine the derivative of

the asymptotic mean mutual information:

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

which yields:

$$(-\rho\frac{d\mu}{d\rho})^3 + (\gamma\xi+\xi+1)(\rho\frac{d\mu}{d\rho})^2 + (\frac{\xi}{\rho}+\xi+\gamma\xi^2+\gamma\xi)(-\rho\frac{d\mu}{d\rho})^2 + \gamma\xi^2 = 0 \qquad (5.6)$$

The solution to this equation can be obtained explicitly and the asymptotic mean mutual information can be obtained through numerical integration with the boundary condition: $\lim_{\rho\to0}\mu(\rho) = 0$. Note that a similar result was obtained by Müller [59] (eq.(54)) using free probability theory. However, the result here is stronger as the complete distribution is provided and not only the asymptotic mean.

Historical Note: Free Probability [115, 116] is a non-commutative probability theory, in which the concept of independence of classical probability is replaced by that of freeness. Voiculescu [117, 118, 119] discovered very important relations between free probability theory and random matrix theory. He showed in particular that random matrices can be considered as typical non-commutative random variables. To the author's knowledge, the first use of free probability in the telecommunication field was made by Evans and Tse in 1999 [120]. Since that date, it has been used for the performance analysis of several transmission schemes (CDMA [121, 122], OFDM [1, 35, 36] and MIMO [106, 59, 123]). 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 [124, 125, 126, 127] (Multi-stage detectors...).

For simulations purpose, we consider the case $n_t = n_r$. In this case, $\xi = \frac{s_r}{n_t} = \frac{1}{\gamma}$ and equation (5.6) simplifies to:

$$(-\rho\frac{d\mu}{d\rho})^3 + (2+\xi)(\rho\frac{d\mu}{d\rho})^2 + (\frac{\xi}{\rho}+\xi+\xi+1)(-\rho\frac{d\mu}{d\rho}) + \xi = 0$$

and has the following solution:

$$\mu = \int_0^\rho \left(\frac{-1}{6\gamma\rho^2}\tau(\gamma,\rho) - \frac{2}{3\rho}\frac{\rho\gamma^2-2\gamma\rho-3\gamma+\rho}{\gamma\tau(\gamma,\rho)} + \frac{1}{3}\frac{1+2\gamma}{\gamma\rho}\right)d\rho \qquad (5.7)$$

with:

$$\begin{aligned}\tau(\gamma,\rho) &= ((8\rho\gamma^3 - 24\rho\gamma^2 + 24\rho\gamma + 72\gamma^2 + 36\gamma - 8\rho \\ &+ 12\sqrt{3}\sqrt{\frac{-4\rho^2+12\rho^2\gamma-\rho-12\rho^2\gamma^2+4\gamma^3\rho^2+8\rho\gamma^2+20\gamma\rho+4\gamma}{\rho}}\gamma)\rho^2)^{\frac{1}{3}}\end{aligned}$$

We have plotted in Figure 5.10 the theoretical ergodic mutual information per receiving antenna of the random directions scenario at 10 dB for various ratio of scatterers ($\frac{s_r}{n_r}$ ranges from 0 to 1). We have also plotted a simulated curve with a system of 8×8 antennas. The
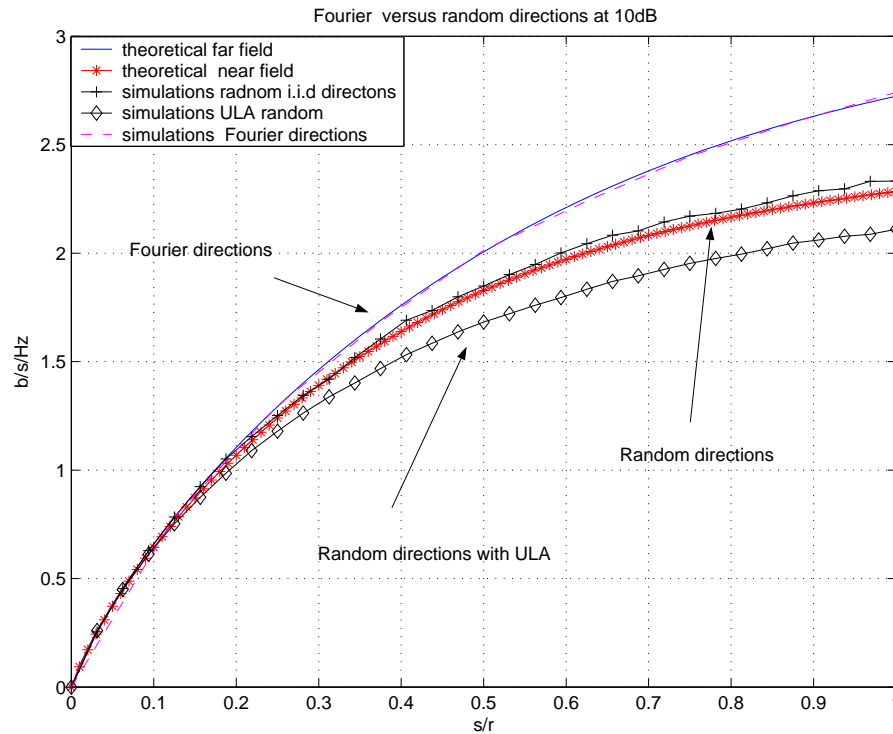
Figure 5.10: Fourier versus random directions at 10dB.

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 uncorrelated and given by Fourier directions (see section 5.2.2). One can observe that far field scattering on Fourier directions yields better performance than near field scattering. A simple explanation can be provided to this observation: in the Fourier direction scenario and in the case of $s_r = n_t$, the DoA matrix $\mathbf{\Phi}$ is a unitary Fourier matrix and has therefore no effect on $\mathbf{\Theta}_{s_r \times n_t}$. However, in the random directions scenario, the non-unitary steering matrix $\mathbf{\Phi}$ has a correlation effect on matrix $\mathbf{\Theta}_{s_r \times n_t}$. One of the conclusions of this observation is that a better transmission occurs when the mobile is far from the scatterers and the scatterers are located in distant positions.

5.3   SINR

*********** *************************

5.4   Remarks on the Directions of Departure based Model

Imagine now that the modeler is in a situation where it assumes (due to the fact that the antennas are directional for example) that the signal goes from the transmitting antennas to a certain number of randomly located scatterers. The waves then propagate from these scatterers to the

receiving antennas in an erratic manner: multiple bounces over other scatterers are possible as well as direct propagation to the receiving antennas. The modeler assumes that each steering direction has a certain power. As previously, the modeler is not aware of correlation between the scattering directions (especially the covariance structure). The modeler also assumes that the channel $\mathbf{H}$ carries some energy. Based on this state of knowledge, what is the consistent model the modeler can make knowing only the directions of departure, the number of scatterers, the zero mean and the variance i.e

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

Denoting $\mathbf{H} = \frac{1}{\sqrt{s_t}}\boldsymbol{\Theta}_{n_r \times s_t}\mathbf{P^t}^{\frac{1}{2}}\boldsymbol{\Psi}_{s_t \times n_t}$ where $\boldsymbol{\Psi}$ is a $s_t \times n_t$ matrix (s_t is the number of scatterers) which represents the directions of departure from the transmitting antennas to randomly positioned scatterers with respective powers $\mathbf{P^t}$. $\boldsymbol{\Theta}_{n_r \times s_t}$ is an $n_r \times s_t$ matrix which represents the scattering environment between the receiving antennas and the scatterers. It can be shown that the principle of maximum entropy will assign independent zero mean complex Gaussian entries to the matrix $\boldsymbol{\Theta}_{n_r \times s_t}$.

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

- normalizes the mutual information with respect to the number of receive antennas.

- n_t, s_r, P^r are exchanged with n_r, s_t and P^t.

- the SNR ρ is replaced by $\frac{n_r}{n_t}\rho$.

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

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

In the high SNR regime, it can be easily shown that the mean mutual information of the DoD based model converges to:

$$\lim_{\rho \to \infty} \mathbb{E}(I^M) = \min(n_r, s_t \int_{\lambda > 0} dS_{\mathrm{dod}}(\lambda))\ln(\rho)$$

In the case of the DoD based model on Fourier directions (with equal powers), it can also be shown that at high SNR:

$$
\begin{aligned}
n_t \mu_{\mathrm{dod}} &= \min(s_t, n_r)\ln(\rho) \\
\sigma^2{}_{\mathrm{dod}} &= -\ln\left(1 - \frac{\min(s_t, n_r)}{\max(s_t, n_r)}\right)
\end{aligned}
$$

Chapter 6

Knowledge of the Directions of Arrival and Departure

6.1 Model

The modeler is now interested in deriving a consistent double directional model i.e taking into account simultaneously the directions of arrival and the directions of departure. The motivation of such an approach lies in the fact that when a single bounce on a scatterer occurs, the direction of arrival and departure are deterministically related by Descartes laws and therefore the distribution of the channel matrix depends on the joint DoA-DoD spectrum. The modeler assumes as a state of knowledge the directions of departure from the transmitting antennas to the set of transmitting scatterers $(1...s_t)$. The modeler also assumes as a state of knowledge the directions of arrival from the set of receiving scatterers $(1...s_r)$ to the receiving antennas. The modeler also has some knowledge that the steering directions have different powers. However, the modeler has no knowledge of what happens in between. The set $(1...s_t)$ and $(1...s_r)$ may be equal, $(1...s_t)$ may be included in $(1...s_r)$ or there may be no relation between the two. The modeler also knows that the channel carries some energy. Based on this state of knowledge, what is the consistent model the modeler can make of $\mathbf{H}$

$$
\mathbf{H} = \frac{1}{\sqrt{s_r s_t}} \begin{pmatrix} e^{j\phi_{1,1}} & \ldots & e^{j\phi_{1,s_r}} \\ \vdots & \ddots & \vdots \\ e^{j\phi_{n_r,1}} & \ldots & e^{j\phi_{n_r,s_r}} \end{pmatrix} \begin{pmatrix} \sqrt{P_1^r} & 0 & \ldots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_r}^r} \end{pmatrix}
$$

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

In other words, how to model $\mathbf{\Theta}_{s_r \times s_t}$? As previously stated, the modeler must comply with the following constraints:

- The channel has a certain energy.

- Consistency argument: If the DoD and DoA are unknown then $\frac{1}{\sqrt{s_r s_t}} \mathbf{\Phi}_{n_r \times s_r} \mathbf{P^r}^{\frac{1}{2}} \mathbf{\Theta}_{s_r \times s_t} \mathbf{P^t}^{\frac{1}{2}} \mathbf{\Psi}_{s_t \times n_t}$ should be assigned an i.i.d zero mean Gaussian distribution.
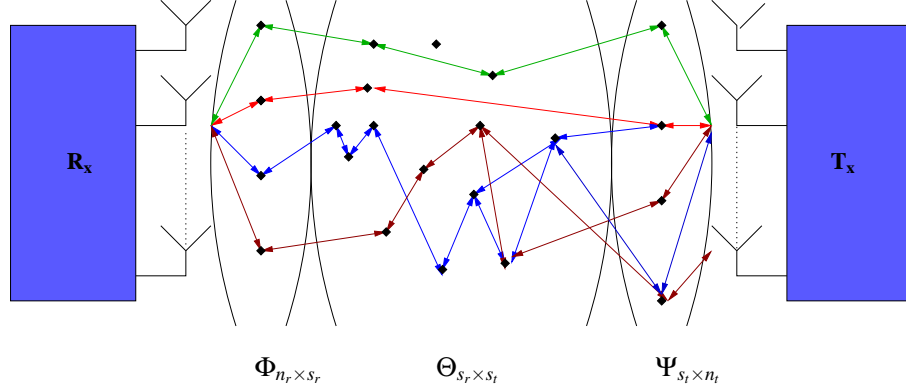
Figure 6.1: Double directional based model.

Let us now determine the distribution of $\mathbf{\Theta}_{s_r \times s_t}$. The probability distribution of $P(\mathbf{H})$ is given by:

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

- When $\mathbf{\Psi}, \mathbf{\Phi}, s_r, s_t, \mathbf{P^r}, \mathbf{P^t}$ are known: $P(\mathbf{\Phi}\mathbf{\Psi} \mid s_r, s_t) = \delta(\mathbf{\Phi} - \mathbf{\Phi}^0)\delta(\mathbf{\Psi} - \mathbf{\Psi}^0)$, $P(s_t, s_r) = \delta(s_r - s^0{}_r)\delta(s_t - s^0{}_t)$, $P(\mathbf{P^r}, \mathbf{P^t} \mid s_r, s_t) = \delta(\mathbf{P^r} - \mathbf{P^{0r}})\delta(\mathbf{P^t} - \mathbf{P^{0t}})$ and

$$
P(\mathbf{H}) = P(\mathbf{\Phi}^0 \mathbf{P^{0r}}^{\frac{1}{2}} \mathbf{\Theta} \mathbf{P^{0t}}^{\frac{1}{2}} \mathbf{\Psi}^0)
$$

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

Proposition 6 $\mathbf{\Theta}_{s_r \times s_t}$ *i.i.d zero mean Gaussian with variance* 1 *is solution of the consistency argument and maximizes entropy.*

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

$P(h_{ij} \mid \mathbf{\Psi}, \mathbf{\Phi}, s_r, s_t) = N(0, \frac{1}{s_t s_r} \sum_{p=1}^{s_r} \sum_{k=1}^{s_t} \mid \sqrt{P_p{}^r} e^{j\phi_{ip}} \sqrt{P_k{}^t} e^{j\psi_{kj}} \mid^2 = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{|h_{ij}|^2}{2}}$ (since $\frac{1}{s_r} \sum_{k=1}^{s_r} P_k{}^t = 1$ and $\frac{1}{s_t} \sum_{p=1}^{s_t} P_p{}^t = 1$ (due to power normalization as we assume the energy known).

Therefore

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

Moreover, we have :

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

which proves that $\boldsymbol{\Theta}_{s_r \times s_t}$ is solution of the consistency argument. Once again, 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.

6.2 Mutual Information

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

per transmitting antenna is given by:

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

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

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

$$
\frac{1}{s_r}\mathbf{P^r}^H\boldsymbol{\Phi}^H\boldsymbol{\Phi}\mathbf{P^r} = V_\phi\boldsymbol{\Lambda}_\phi V_\phi^H
$$

and

$$
\frac{1}{s_t}\mathbf{P^t}\boldsymbol{\Psi}\boldsymbol{\Psi}^H\mathbf{P^t}^H = V_\psi\boldsymbol{\Lambda}_\psi V_\psi^H
$$

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

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

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

and their Stieltjes transforms are related as:

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

Matrix $V_\phi\boldsymbol{\Theta}V_{\boldsymbol{\Psi}}$ is an i.i.d zero mean Gaussian matrix with unit variance (only unitary transforms are applied). Therefore, matrix $\boldsymbol{\Theta}_1 = \frac{1}{\sqrt{n_r}}[\boldsymbol{\Lambda}_{\boldsymbol{\Phi}}^{\frac{1}{2}}(V_{\boldsymbol{\Phi}}\boldsymbol{\Theta}V_{\boldsymbol{\Psi}})\boldsymbol{\Lambda}_{\boldsymbol{\Psi}}^{\frac{1}{2}}]$ is a $s_r \times s_t$ random matrix composed of independent entries with zero mean and variances $\frac{1}{n_r}\lambda_i^\phi\lambda_j^\psi = \frac{1}{s_r}\frac{\lambda_i^{\boldsymbol{\Phi}}\lambda_j^{\boldsymbol{\Psi}}}{\gamma}$. The weak convergence of the empirical eigenvalue distribution of $\boldsymbol{\Theta}_1\boldsymbol{\Theta}_1^H$ to a limiting distribution holds under certain assumptions and is an application of a theorem due to Girko [128] (the theorem is recalled in the appendix through Theorem 5). Note that this theorem was already used in [57] for deriving the mutual information of MIMO wireless systems under correlated fading and in [129] for analyzing CDMA Networks with Multiuser Receivers and Spatial Diversity.

Proposition 7 *Suppose that the empirical joint distribution $\lambda_i{}^{\boldsymbol{\Phi}}\lambda_j{}^{\boldsymbol{\Psi}}$ converges to some joint limit distribution $f_{\lambda^{\boldsymbol{\Phi}}\lambda^{\boldsymbol{\Psi}}}$ as the size of the matrix $\boldsymbol{\Theta}_1$ grows large but $\gamma, \gamma_1, \xi, \xi_1$ remains fixed, the asymptotic mutual information per transmitting antenna is given by:*

$$\mu = \int_0^\rho \xi \mathbb{E}_{\lambda^\phi} \left[\frac{\alpha^{joint}}{1 + \rho \alpha^{joint}} \right] d\rho \tag{6.1}$$

with

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

and

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

Proof: the proof is provided in Appendix[1].

Note that when taking the expectations, $\lambda^{\boldsymbol{\Phi}}$ depends on $\lambda^{\boldsymbol{\Psi}}$. In the case where the distribution of the angles of arrival are independent of the angles of departure, a simpler expression of equation (6.1) can be provided[2].

Proposition 8 *Suppose that the empirical joint distribution $\lambda_i{}^{\boldsymbol{\Phi}}\lambda_j{}^{\boldsymbol{\Psi}}$ is separable and converges to a product of two limiting distribution, then as the size of the matrix $\boldsymbol{\Theta}_1$ grows large but $\gamma, \gamma_1, \xi, \xi_1$ remains fixed, the asymptotic mutual information per transmitting antenna is given by:*

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

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

Proof: the proof is provided in the appendix.

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

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

Remark 1 *Results of Free Probability Theory could also be used to prove Proposition (7). Indeed, one has to derive the limiting eigenvalue distribution of $\frac{1}{n_r}[\boldsymbol{\Lambda}_{\boldsymbol{\Phi}}^{\frac{1}{2}} \boldsymbol{\Theta} \boldsymbol{\Lambda}_{\boldsymbol{\Psi}} \boldsymbol{\Theta}^H \boldsymbol{\Lambda}_{\boldsymbol{\Phi}}^{\frac{1}{2}}]$ with $\boldsymbol{\Theta}$ i.i.d Gaussian and $\boldsymbol{\Lambda}_{\boldsymbol{\Phi}}$ and $\boldsymbol{\Lambda}_{\boldsymbol{\Psi}}$ diagonal. Since $\boldsymbol{\Theta} \boldsymbol{\Lambda}_{\boldsymbol{\Psi}} \boldsymbol{\Theta}^H$ is unitarily invariant[3] and asymptotically free from $\boldsymbol{\Lambda}_{\boldsymbol{\Phi}}$, one can obtain straightforwardly the law using the free multiplicative convolution of $\boldsymbol{\Theta} \boldsymbol{\Lambda}_{\boldsymbol{\Psi}} \boldsymbol{\Theta}^H$ and $\boldsymbol{\Lambda}_{\boldsymbol{\Phi}}$. However, the result of Girko is more general as the entries of matrix $\boldsymbol{\Theta}$ need not be Gaussian.*

Remark 2 *Although we have no formal on the uniqueness (the mean mutual information of propostion 8 has in fact multiple solutions. Therefore, only some physical arguments can be given to withdraw some solutions), it can be shown that the mean mutual information for the double directional model scales at high SNR as:*

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

---

[1]Note that in the double directional model, the Gaussian behavior is still an open issue.

[2]Physical measurements have already indicated that the correlation between the directions of arrival and the directions of departure are negligible [130].

[3]A $N \times N$ self-adjoint random matrix is called unitarily invariant if the probability measure of A as a random matrix is equal to that of the matrix $V A V^H$ for any unitary constant matrix V.
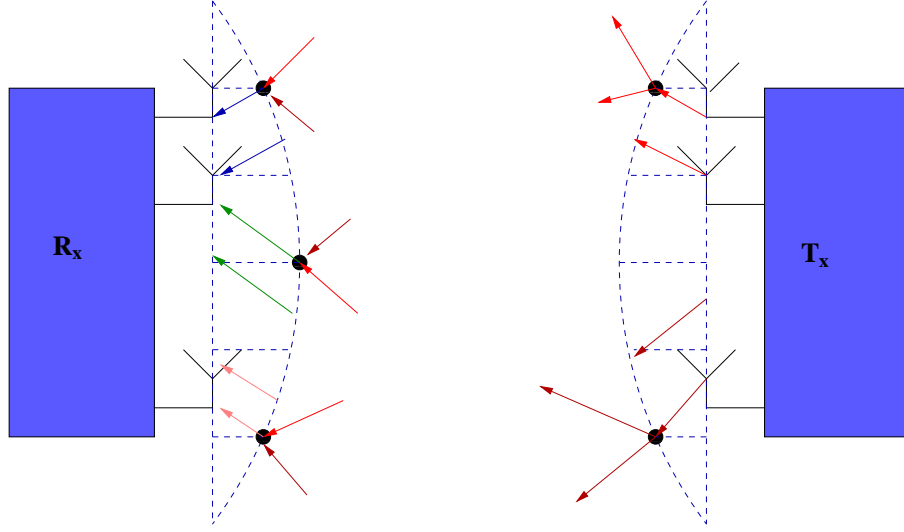
Figure 6.2: Simple case: scatterers positioned on special directions.

6.2.1 ULA and Fourier Directions Case

In this part, the modeler takes into account the geometry of the antennas to derive the steering vectors. In the case of a uniform linear array, the MIMO channel matrix has the following form:

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

In order to have a tractable formula as in section 5.2.2, we will suppose that the steering vectors are on Fourier directions.

Equal Power Case

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

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

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

and

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

Proposition 9 *The asymptotic mutual information per transmitting antenna for the double directional model in the equal power and Fourier directions case is given by:*

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

with

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

Proof 1 *The result is an application of Proposition 8 and the proof is given in the appendix.*

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

Proposition 10 *The asymptotic variance of the mutual information in the equal power and Fourier directions case for the double directional model is given by:*

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

Proof 2 *One can notice that:*

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

Therefore, since $\mathbf{\Theta}$ is an i.i.d Gaussian matrix, results of section 3.2.1 can be applied. In particular, if one makes the variable change:

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

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

in the formulas of theorem 1 then the result is proven. By doing so, one can notice that by this change of variable the same formula as in Proposition 9 (which was obtained with Girko's results) is obtained for the mean value.

At high SNR, it can be easilly shown that:

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

$$\sigma^2_{double} = \begin{cases} -\ln\left(1 - \frac{\min(s_t,s_r)}{\max(s_t,s_r)}\right) & \text{if } s_t \neq s_r \\ \frac{1}{2}\ln(\rho) & \text{if } s_t = s_r \end{cases} \tag{6.2}$$

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

In Figure 6.3, simulations have been conducted with $n_r = n_t = 8$ antennas. Three cases have been plotted:
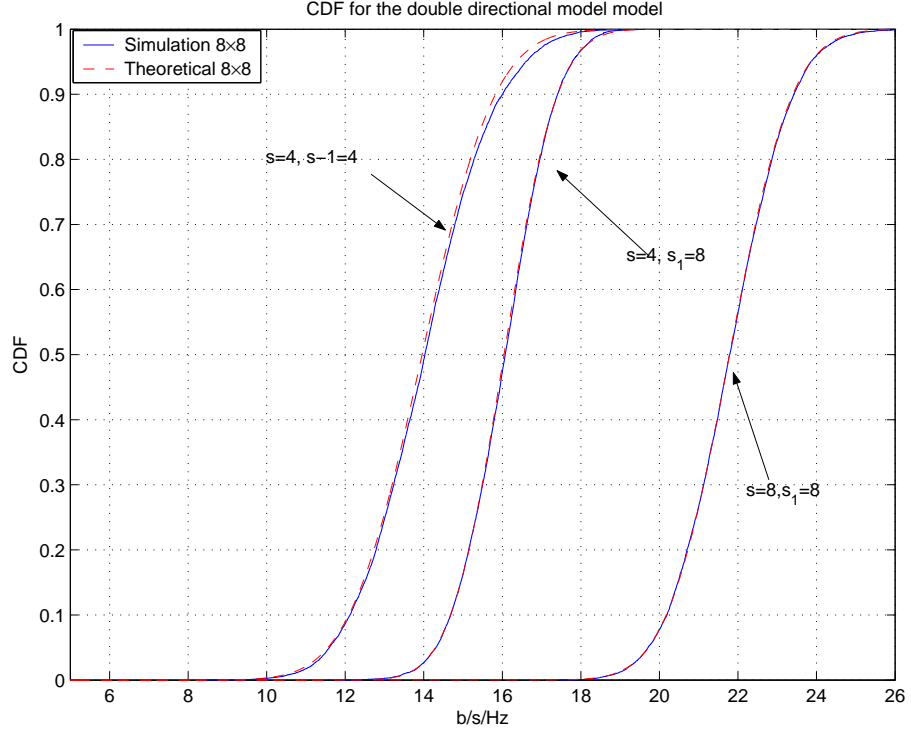
- $s_r = 8$ and $s_t = 8$

Figure 6.3: Mutual information cumulative distribution in the case of the double directional model with equal power on Fourier directions.

- $s_r = 4$ 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 receiving antennas, we have plotted in Figure 6.4 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 $s_r = s_t = n_r = n_t$.

Non-equal Power Case

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

$$S_{\mathrm{doa}}(\lambda) = \sum_{i=1}^{K_r} l_i^r \delta(\lambda - \gamma P_i^r)$$
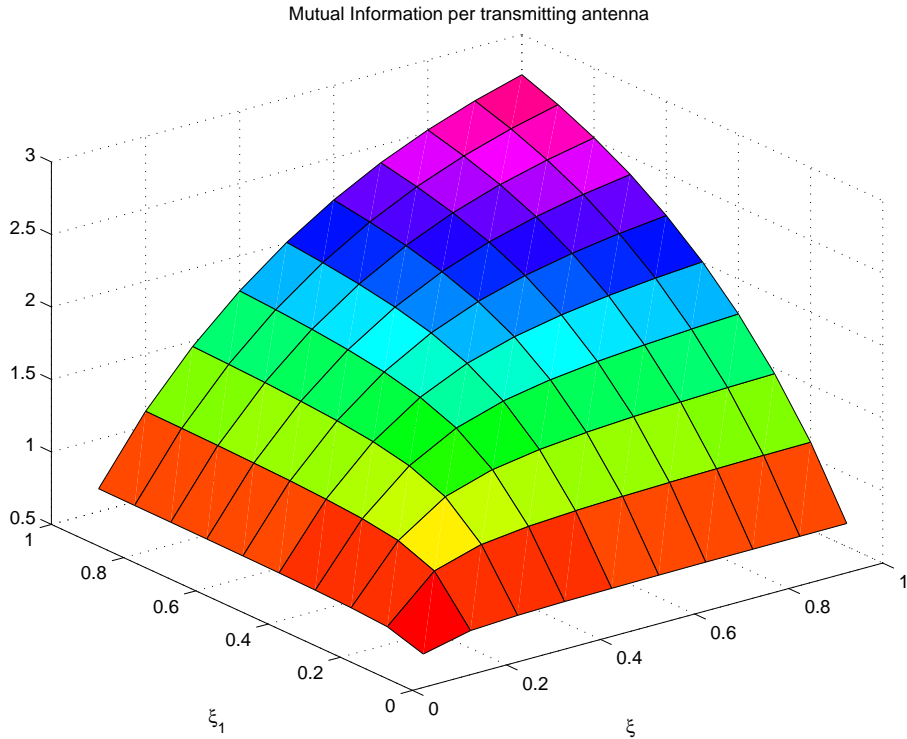
Figure 6.4: Mutual Information per transmitting antenna versus ξ and ξ_1 for the double directional model with equal power on Fourier directions.

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

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

Proposition 11 *In this case, μ_{double} is equal to:*

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

with

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

and

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

Proof 3 *The proof is an application of the general Proposition 8 in the case of interest.*

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

Proposition 12 *The mean of the mutual information in the case of the double directional model with ULA and Fourier directions is maximized for $\mathbf{P^r} = \mathbf{I}_{s_r}$ and $\mathbf{P^t} = \mathbf{I}_{s_t}$.*

Proof 4 *The proof is provided in the appendix.*

In Figure 6.5, simulations have been conducted in the two power case with $n_r = n_t = 8$ antennas. We have chosen $P_1{}^t = 2 - P_2{}^t$, $P_1{}^r = 2 - P_2{}^r$, , $l_1{}^r = l_2{}^r = l_1{}^t = l_2{}^t$, $s = 8$ and $s_1 = 8$. In this case, we have ($\gamma = \frac{n_r}{s_r} = 1, \xi_1 = \frac{s_t}{n_t} = 1, \xi = \frac{s_r}{n_t} = 1$):

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

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

and

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

The figure acknowledges the fact that the best throughput is obtained when all the steering directions have the same power on both sides.
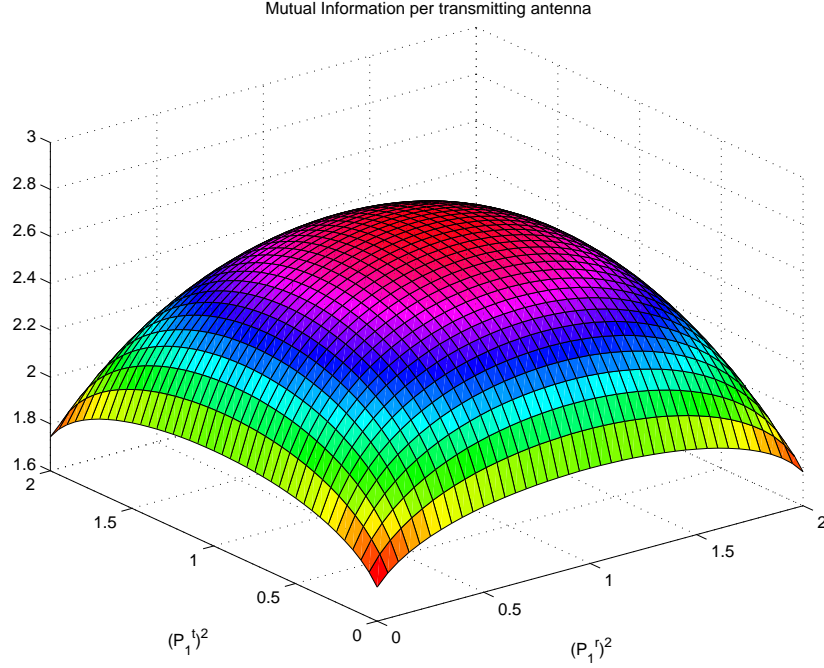
Figure 6.5: Mean capacity per transmitting antenna versus $P_1{}^r$ and $P_1{}^t$ at 10dB for the double directional model.

6.2.2 Fourier versus Random Directions: Equal Power Case

We would like to quantify the impact of random directions for the double directional based model. For the random directions context, we will suppose that the entries of matrix $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are a realization of independent and uniformly distributed exponential variables with zero mean and unit variance. The limiting eigenvalue distribution 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 ([131]) and Proposition 8 can be applied straightforwardly. However, we will take Müller's approach, as our framework is a particular case of [132]: in "On The Asymptotic Eigenvalue Distribution of Concatenated Vector-valued Fading Channels", Müller introduces an N fold scattering model as a product of N i.i.d random matrices $\mathbf{H} = \prod_{i=1}^{N}\mathbf{M}_i$. He proves the almost sure convergence of the limiting eigenvalue distribution of matrix $\mathbf{H}$ and gives an explicit form of its Stieltjes transform. In the case considered here, $\mathbf{H} = \boldsymbol{\Phi}\boldsymbol{\Theta}\boldsymbol{\Psi}$ is the product of three random matrices. Using the results in [132], it can be easily shown that the Stieltjes transform $m_{H^H H}(x)$ is solution of the following equation:

$$m_{H^H H}(-x)\left(\frac{xm_{H^H H}(-x)-1+\xi_1}{\xi_1}\frac{xm_{H^H H}(-x)-1+\xi}{\xi}\frac{xm_{H^H H}(-x)-1+\gamma\rho}{\gamma\rho}\right)$$
$$+\quad xm_{H^H H}(-x) = 1$$

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

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

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

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

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

We have plotted in Figure 6.6 the theoretical asymptotic mean mutual information per receiving antenna of the random 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 $\frac{s_t}{n_r}$. 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.6.2.1). One can observe that scatterers on Fourier directions yield better performance than scatterers on random directions. The same explanation as in the directional scenario can be provided: in the far field scenario with uncorrelated scattering and in the case of $s_r = n_t$, 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 directions scenario, the non-unitary steering matrix $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ have a correlation effect on matrix $\boldsymbol{\Theta}_{s_r \times s_t}$. One of the conclusions of this observation is that a better transmission occurs when the mobile is far from the scatterers and the scatterers are located in distant positions. Moreover, one can observe that the mutual information of the double directional model in the random directions scenario is less than the mutual information for the mono-directional models as shown in Figure 5.10.
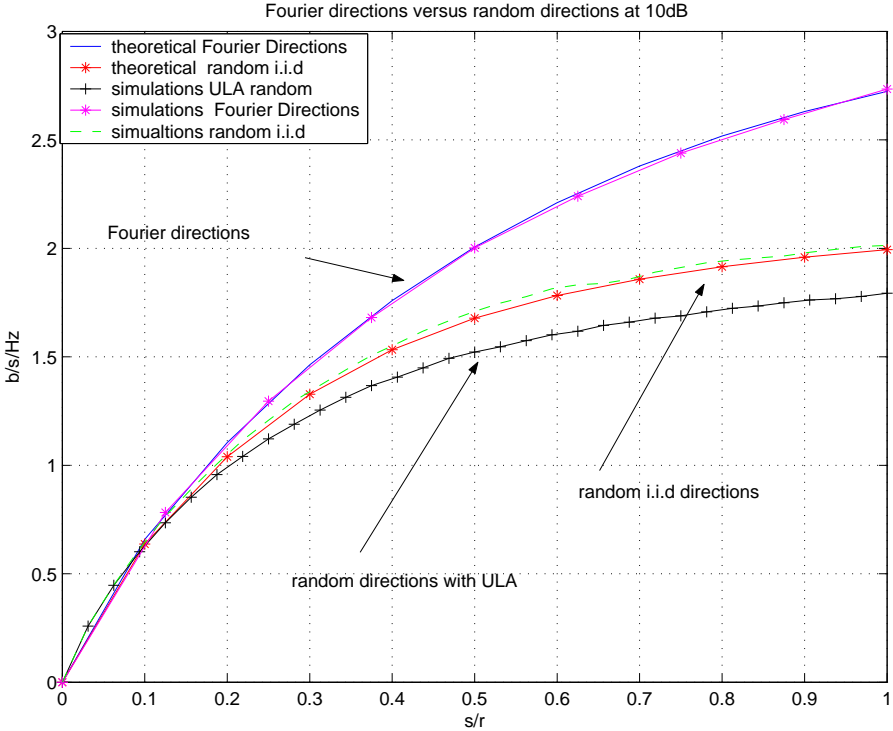
6.3 SINR

******** *****************************

Figure 6.6: Fourier versus random directions at 10dB.

Chapter 7

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

7.1 Model

The modeler wants to derive a consistent model taking into account the direction of arrivals and respective power profile, directions of departure and respective power profile, delay, Doppler effect. As a starting point, the modeler assumes that the position of the transmitter and receiver changes in time. However, the scattering environment (the buildings, trees,...) does not change and stays in the same position during the transmission. Let $\mathbf{v_t}$ and $\mathbf{v_r}$ be respectively the vector speed of the transmitter and the receiver with respect to a terrestrial reference (see Figure 7.1). Let $\mathbf{s}_{ij}^t$ 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:

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

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

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

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

In all the following, the modeler supposes as a state of knowledge the following parameters:
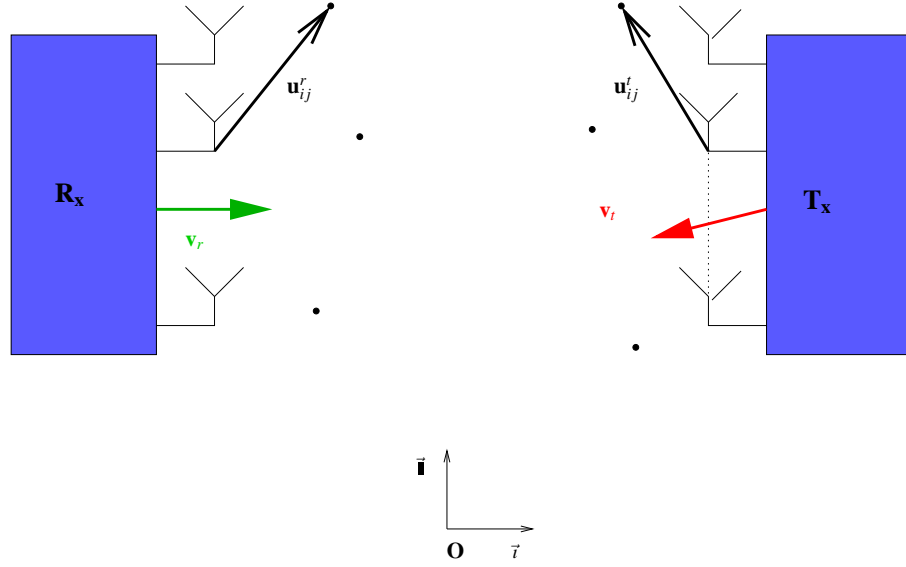
- speed $\mathbf{v_r}$.

Figure 7.1: Moving antennas.

- speed $\mathbf{v_t}$.

- 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 the fact that a signal going from a steering vector of departure j to a steering vector of arrival i has a certain delay τ_{ij} due to possible single bounce or multiple bounces on different objects. The modeler also knows that objects do not move between the two sets of scatterers. The $s_r \times s_t$ delay matrix linking each DoA and DoD has the following structure:

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

The modeler also supposes as a given state of knowledge the fact that each path h_{ij} of matrix $\mathbf{H}$ has a certain power. Based on this state of knowledge, the modeler wants to model the $s_r \times s_t$ matrix $\mathbf{\Theta}_{s_r \times s_t}$ in the following representation:

$$\mathbf{H}(f,t) = \frac{1}{\sqrt{s_r s_t}} \begin{pmatrix} e^{j(\phi_{1,1}+2\pi\frac{f\mathbf{u}_{11}^r\mathbf{v_r}}{c}t)} & \cdots & e^{j(\phi_{1,s}+2\pi\frac{f\mathbf{u}_{1s}^r\mathbf{v_r}}{c}t)} \\ \vdots & \ddots & \vdots \\ e^{j(\phi_{r,1}+2\pi\frac{f\mathbf{u_{r1}}^r\mathbf{v_r}}{c}t)} & \cdots & e^{j(\phi_{r,s}+2\pi\frac{f\mathbf{u}_{rs}^r\mathbf{v_r}}{c}t)} \end{pmatrix} \begin{pmatrix} \sqrt{P_1^r} & 0 & \cdots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_r}^r} \end{pmatrix}$$

$$\mathbf{\Theta}_{s_r \times s_t} \bigodot \mathbf{D}_{s_r \times s_t}(f)$$

$$\begin{pmatrix} \sqrt{P_1^t} & 0 & \cdots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_t}^t} \end{pmatrix} \begin{pmatrix} e^{j(\psi_{1,1}+2\pi\frac{f\mathbf{u}_{11}^t\mathbf{v_t}}{c}t)} & \cdots & e^{j(\psi_{1,n_t}+2\pi\frac{f\mathbf{u}_{1n_t}^t\mathbf{v_t}}{c}t)} \\ \vdots & \ddots & \vdots \\ e^{j(\psi_{s_1,1}+2\pi\frac{f\mathbf{u}_{s_11}^t\mathbf{v_t}}{c}t)} & \cdots & e^{j(\psi_{s_1,n_t}+2\pi\frac{f\mathbf{u}_{s_1n_t}^t\mathbf{v_t}}{c}t)} \end{pmatrix}$$

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

- Each entry of $\mathbf{H}(f,t)$ has a certain energy.

- Consistency argument: if the DoA, DoD, powers, the delays, the Doppler effects are unknown then matrix $\mathbf{H}$ should be assigned an i.i.d Gaussian distribution.

Proposition 13 $\mathbf{\Theta}_{s_r \times s_t}$ *i.i.d zero mean Gaussian with variance 1 is solution of the consistency argument and maximizes entropy.*[1]

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

- **Delays**: The modeler knows that there is, due to measurements performed in the area, a maximum possible delay for the information to go from the transmitter to the receiver $\tau_{\max}$. The principle of maximum entropy attributes therefore a uniform distribution to all the delays τ_{ij} such as $P(\tau_{ij}) = \frac{1}{\tau_{\max}}$ with $\tau_{ij} \in [0, \tau_{\max}]$

- **Doppler effect**: The modeler knows that the speed of the transmitter and receiver can not exceed a certain limit v_{limit} (in the least favorable case, v_{limit} would be equal 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 $P(v_t) = P(v_r) = \frac{1}{v_{\mathrm{limit}}}$. Moreover, if $\mathbf{v}_t = v_t \cos(\alpha_t)\tilde{\imath} + v_t \sin(\alpha_t)\tilde{\jmath}$, $\mathbf{v}_r = v_r \cos(\alpha_r)\tilde{\imath} + v_r \sin(\alpha_r)\tilde{\jmath}$, $\mathbf{u^t}_{ij} = \cos(\beta^t_{ij})\tilde{\imath} + \sin(\beta^t_{ij})\tilde{\jmath}$ and $\mathbf{u^r}_{ij} = \cos(\beta^r_{ij})\tilde{\imath} + \sin(\beta^r_{ij})\tilde{\jmath}$, the modeler will attribute a uniform distribution over 2π to the angles α_t, α_r, β^t_{ij} and β^r_{ij}.

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

---

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

Note that in the case $f = 0$, $\mathbf{v_t} = 0$ and $\mathbf{v_r} = 0$, the same model as the narrowband model is obtained. If more information is available on correlation or different variances of frequency paths, then this information can be incorporated in the matrix $\mathbf{D}_{s_r \times s_t}$, also known as the channel pattern mask [55]. Note that in the case of a ULA geometry and in the Fourier directions, we have $\mathbf{u}_{ij}^r = \mathbf{u}_j^r$ (any column of matrix $\boldsymbol{\Phi}$ has a given direction) and $\mathbf{u}_{ij}^t = \mathbf{u}_i^t$ (any line of matrix $\boldsymbol{\Psi}$ has a given direction). Therefore, the channel model simplifies to:

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

$$
\begin{pmatrix} 1 & \cdots & e^{j2\pi \frac{d(n_t-1)\sin(\psi_1)}{\lambda}} \\ \vdots & \ddots & \vdots \\ 1 & \cdots & e^{j2\pi \frac{d(n_t-1)\sin(\psi_{s_t})}{\lambda}} \end{pmatrix}
$$

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

$$
\mathbf{D}_{s_r \times s_t}(f,t) = \begin{pmatrix} \sqrt{P_1^r}\sqrt{P_1^t}e^{-j2\pi f\tau_{1,1}}e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r\mathbf{v}_r+\mathbf{u}_1^t\mathbf{v}_t)} & \cdots & \sqrt{P_1^r}\sqrt{P_{s_t}^t}e^{-j2\pi f\tau_{1,s_t}}e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r\mathbf{v}_r+\mathbf{u}_{s_t}^t\mathbf{v}_t)} \\ \vdots & \ddots & \vdots \\ \sqrt{P_{s_r}^r}\sqrt{P_1^t}e^{-j2\pi f\tau_{s_r,1}}e^{j2\pi \frac{ft}{c}(\mathbf{u}_{s_r}^r\mathbf{v}_r+\mathbf{u}_1^t\mathbf{v}_t)} & \cdots & \sqrt{P_{s_r}^r}\sqrt{P_{s_t}^t}e^{-j2\pi f\tau_{s_r,s_t}}e^{j2\pi \frac{ft}{c}(\mathbf{u}_s^r\mathbf{v}_r+\mathbf{u}_{s_t}^t\mathbf{v}_t)} \end{pmatrix}
$$

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

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

$$
\mathbf{H}(f,t) = \frac{1}{\sqrt{s_r s_t}} \begin{bmatrix} e^{j(\phi_1+2\pi \frac{f\mathbf{u}_1^r\mathbf{v_r}}{c}t)} & \cdots & e^{j(\phi_{s_r}+2\pi \frac{f\mathbf{u}_{s_r}^r\mathbf{v_r}}{c}t)} \end{bmatrix} \begin{pmatrix} \sqrt{P_1^r} & 0 & \cdots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_r}^r} \end{pmatrix}
$$

$$
\boldsymbol{\Theta}_{s_r \times s_t} \bigodot \mathbf{D}_{s_r \times s_t}(f) \begin{pmatrix} \sqrt{P_1^t} & 0 & \cdots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_t}^t} \end{pmatrix} \begin{bmatrix} e^{j(\psi_1+2\pi \frac{f\mathbf{u}_1^t\mathbf{v_t}}{c}t)} \\ \vdots \\ e^{j(\psi_{s_t}+2\pi \frac{f\mathbf{u}_{s_t}^t\mathbf{v_t}}{c}t)} \end{bmatrix}
$$

$$
= \frac{1}{\sqrt{s_r s_t}} \begin{bmatrix} \sum_{k=1}^{s_r} \theta_{k,1}\sqrt{P_k^r}e^{j(\phi_k+2\pi \frac{f\mathbf{u}_k^r\mathbf{v_r}}{c}t)}e^{-j2\pi f\tau_{k,1}} & \cdots & \sum_{k=1}^{s_R} \theta_{k,s_R}\sqrt{P_k^r}e^{j(\phi_k+2\pi \frac{f\mathbf{u}_k^r\mathbf{v_r}}{c}t)}e^{-j2\pi f\tau_{k,s_r}} \end{bmatrix}
$$

$$
\begin{pmatrix} \sqrt{P_1^t} & 0 & \cdots \\ 0 & \ddots & 0 \\ \vdots & 0 & \sqrt{P_{s_t}^t} \end{pmatrix} \begin{pmatrix} e^{j(\psi_1+2\pi \frac{f\mathbf{u}_1^t\mathbf{v_t}}{c}t)} \\ \vdots \\ e^{j(\psi_{s_t}+2\pi \frac{f\mathbf{u}_{s_t}^t\mathbf{v_t}}{c}t)} \end{pmatrix}
$$

$$
= \sum_{l=1}^{s_t}\sum_{k=1}^{s_r} \rho_{k,l}e^{j2\pi\xi_{k,l}t}e^{-j2\pi f\tau_{k,l}}
$$

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 $\mathbb{E}(|\rho_{k,l}|^2) = \frac{1}{s_r s_t}P_k^r P_l^t$, $\xi_{k,l} = \frac{f}{c}(\mathbf{u}_k^r\mathbf{v_r} - \mathbf{u}_l^t\mathbf{v_t})$ are the doppler effect and $\tau_{k,l}$

are the delays. This previous result is a generalization of the SISO (Single Input Single Output) wireless model in the case of multifold scattering with the power profile taken into account.

7.2   Some Remarks

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

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

Note that the mutual information depends on t due to the Doppler effect and has therefore no real meaning. Indeed, the perfect channel knowledge assumption at the receiver is not valid (since the channel varies) and a non-coherent mutual information should be calculated. This is not an easy task and an open problem even for simple channel models. A first step in this direction is the work of Marzetta and Hochwald [133] and the work of Zheng and Tsé [134][2]. The general focus is on an ideal setting where time is slotted in blocks of T symbols and within each block, fading is constant independent from block to block. It is shown [133] that capacity does not increase by having more antennas than T. The optimal strategy is to modulate orthogonal T-vectors at each antenna. An even more difficult problem concerns the practical schemes for achieving the non-coherent mutual information. In an interesting paper "How much training is needed in Multiple-antenna Wireless links" [135], Hassibi and Hochwald have shown that simple on the shelf training schemes can be optimal at high SNR (for the i.i.d Gaussian model) which circumvents therefore the need of using blind or semi-blind techniques in that regime[3].

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

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

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

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

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

$$Cov(i, j, m, n, f) = \frac{1}{s_r s_t} \sum_{k=1}^{s_t} \sum_{p=1}^{s_r} \sum_{q=1}^{s_t} \sum_{l=1}^{s_r} \mathbb{E}(\theta_{pk}\theta_{ql}^*) e^{j2\pi f(\tau_{pk}-\tau_{ql})}$$

$$\sqrt{P_k^t}\sqrt{P_q^t}\sqrt{P_p^r}\sqrt{P_l^r} e^{j2\pi(\psi_{kj}-\psi_{qn})} e^{j2\pi(\phi_{ip}-\phi_{ml})}$$

---

[2]For contribution [134], Zheng and Tsé received the 2003 IEEE Information Theory Society paper award.

[3]In fact, transmitting strategies differ quite dramatically with the state of knowledge at the transmitter and the receiver. There has been numerous papers since 2000 in this field [136]. For example, when the transmitter knows the channel matrix, then it can in principle adjust its transmitted power so that no power is wasted in signaling dimensions affected by severe fades[137]. Note that the optimum signaling in the low-power regime concentrates all its energy in the maximal-singular-value eigenspace of the channel matrix.

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

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

which is independent of f.

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

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

One can observe that frequency selectivity does not affect the mutual information per transmitting antenna. Similar results have been reported in [55, 59] and these conclusions will be assessed in chapter 11. In the wideband case with no Doppler effect, the ergodic mutual information is the same as in the narrowband case and all the results of chapter 6 remain valid (ULA and Fourier directions, random directions approximation..).

Chapter 8

Discussion

8.1 Müller's Model

In a paper "A Random Matrix Model of Communication via Antenna Arrays" [59], Müller develops a channel model based on the product of two random matrices:

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

where $\mathbf{\Phi}$ and $\mathbf{\Theta}$ are two random matrices with zero mean unit variance i.i.d entries and $\mathbf{A}$ is a diagonal matrix (representing the attenuations). This model is intended to represent the fact that each signal bounces off a scattering object exactly once. $\mathbf{\Phi}$ 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 [59] confirmed the model quite accurately. Should we conclude that signals in day to day life bounce only once on the scattering objects?

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

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

  - Random directions scattering at the receiving side.
  - Each steering vector at the receiving side has a certain power.
  - Each frequency path has a given variance.

- a corresponding DoD based model with random directions i.e matrix $\mathbf{\Theta}$ with different powers (represented by matrix $\mathbf{A}$) for each angle of departure. The model permits also in this case the signal to bounce several times from the scatterers to the receiving antennas. From a maximum entropy standpoint, the model answers the following question: what is the consistent model when the state of knowledge is limited to:

- Random directions scattering at the transmitting side.

- Each steering vector at the transmitting side has a certain power.

- Each frequency has zero mean and a certain variance.

- 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 [138]

8.2 Sayeed's Model

In a paper "Deconstructing Multi-antenna Fading Channels" [139], Sayeed proposes a virtual representation of the channel. The model is the following:

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

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

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

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

As one can see, Sayeed's model has a simple interpretation in the maximum entropy framework: it considers a ULA geometry with Fourier directions each time. Although it may seem strange that Sayeed limits himself to Fourier directions, we do have an explanation for this fact. In his paper [55], Sayeed was mostly interested in the capacity scaling of MIMO channels and not the joint distribution of the elements. From that perspective, only the statistics of the uncorrelated scatterers is of interest since they are the ones which scale the mutual information.

The correlated scatterers have very small effect on the information. In this respect, we must admit that Sayeed's intuition is quite impressive. 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 (which is not immediate in Sayeed's representation) approach is the quite simplicity for translating any additional physical information into probability assignment in the model. A one to one mapping between information and model representation is possible. With the maximum entropy approach, every new information on the environment can be straightforwardly incorporated and the models are consistent: adding or retrieving information takes us one step forward or back but always in a consistent way. The models are somewhat like Russian dolls, imbricated one into the other.

8.3 The "Kronecker" model

In a paper "Capacity Scaling in MIMO Wireless Systems Under Correlated fading", Chuah et al. study the following Kronecker [1] model:

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

Here, $\boldsymbol{\Theta}$ is an $n_r \times n_t$ i.i.d zero mean Gaussian matrix, $\mathbf{R}_{n_r}^{\frac{1}{2}}$ is an $n_r \times n_r$ receiving correlation matrix while $\mathbf{R}_{n_t}^{\frac{1}{2}}$ is a $n_t \times n_t$ transmitting correlation matrix. The correlation is supposed to decrease sufficiently fast so that $\mathbf{R}_{n_r}$ and $\mathbf{R}_{n_t}$ have a Toeplitz band structure. Using a software tool (Wireless System Engineering [142]), they demonstrate the validity of the model. Quite remarkably, although designed to take into account receiving and transmitting correlation, the model developed in the paper falls within the double directional framework. Indeed, since $\mathbf{R}_{n_r}$ and $\mathbf{R}_{n_t}$ are band Toeplitz then these matrices are asymptotically diagonalized in a Fourier basis

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

and

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

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

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

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

---

[1]The model is called a Kronecker model because $\mathbb{E}(\text{vec}(\mathbf{H})^H \text{vec}(\mathbf{H})) = \mathbf{R}_{n_r} \bigotimes \mathbf{R}_{n_t}$ is a Kronecker product. The justification of this approach relies on the fact that only immediate surroundings of the antenna array impose the correlation between array elements and have no impact on correlations observed between the elements of the array at the other end of the link. Some discussions can be found in [140, 141].

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

$$\mathbf{D}_{s \times s_1} = \begin{pmatrix} \lambda_{1,n_t}^{\frac{1}{2}} \lambda_{1,n_r}^{\frac{1}{2}} & \cdots & \lambda_{n_t,n_t}^{\frac{1}{2}} \lambda_{1,n_r}^{\frac{1}{2}} \\ \vdots & \ddots & \vdots \\ \lambda_{1,n_t}^{\frac{1}{2}} \lambda_{n_r,n_r}^{\frac{1}{2}} & \cdots & \lambda_{n_t,n_t}^{\frac{1}{2}} \lambda_{n_r,n_r}^{\frac{1}{2}} \end{pmatrix}$$

Note that this connection with the double directional model has already been reported in [55]. Here again, the previous model can be reinterpreted in light of the maximum entropy approach. The model answers the following question: what is the consistent model one can make when the DoA are uncorrelated and have respective power λ_{i,n_r}, the DoD are uncorrelated and have respective power λ_{i,n_t}, each path has zero mean and a certain variance. The model therefore confirms the double directional assumption as well as Sayeed's approach and is a particular case of the maximum entropy approach. The comments and limitations made on Sayeed's model are also valid here. **reference also [143, 144]**

8.4 The "Keyhole" Model

In [145], Gesbert et al. show that low correlation[2] is not a guarantee of high capacity: cases where the channel is rank deficient can appear while having uncorrelated entries (for example when a screen with a small keyhole is placed in between the transmitting and receiving antennas). In [147], they propose the following model for a rank one channel:

$$\mathbf{H} = \mathbf{R}_{n_r}{}^{\frac{1}{2}} \mathbf{g}_r \mathbf{g}_t{}^H \mathbf{R}_{n_t}{}^{\frac{1}{2}} \tag{8.1}$$

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

$$\mathbf{H} = \frac{1}{\sqrt{s_r s_t}} \boldsymbol{\Phi}_{n_r \times s_r} \boldsymbol{\Theta}_{s_r \times s_t} \boldsymbol{\Psi}_{s_t \times n_t} \tag{8.2}$$

The Keyhole model can be either:

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

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

---

[2]"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 [146].

8.5   Conclusion

After analyzing each of these models, we find that they all answer a specific question based on a given state of knowledge. All these models can be derived within the maximum entropy framework and have a simple interpretation. Moreover, each time the directional assumption appears which conjectures the correctness of the directional approach.

Chapter 9

Testing the Models

In all the previous chapters, we have developed several models based on different questions. But what is the right model, in other words how to choose between the set $\{M_0, M_1, ..., M_K\}$ of K models (note that M specifies only the type of model and not the parameters of the model)?

9.1 Conventional Methods

In the previous section, we have shown how probability theory can be used to rank the models. However, the integrals derived in equation (9.1) and equation (9.2) are not easy to compute, especially in the case of interest with a high number of antennas (8×8) since we have to marginalize our integrals across a great number of parameters. But however difficult the problem may be, it is not a reason to hide problems and the use of other methods should be clearly explained. The reader must now know that one can rank models and that there is an optimum number of parameters when representing information. The Bayesian framework gives us an answer by comparing the a posteriori probability ratios: $\frac{P(M|\mathbf{y},I)}{P(M_1|\mathbf{y},I)}$ **WHAT IS D?**. If one is to use other testing methods, then one has to clearly understand the limitations of these methods and justify the use of the criteria. In the following, we explain two procedures used by the channel modelling community and explain their limitations.

1- Parameter estimation methods

In this procedure, the data is cut into two parts, one for estimating the parameters, the other to validate the model incorporating the parameters.

- For estimating the parameters such as the angles of arrival, non-parametric methods such as the beamforming or the Capon method [148] can be used. In the case of parametric methods such as Music [149], Min-Norm [150] or Esprit method [151], they rely on properties of the structure of the covariance $\mathbf{R} = \mathbb{E}(\mathbf{y}\mathbf{y}^{\mathbf{H}}) = \mathbf{\Phi}\mathbf{K}\mathbf{\Phi}^{\mathbf{H}} + \sigma^{\mathbf{2}}\mathbf{I}$ of the output signal. In this case, one has to assume that matrix $\mathbf{K}$ ($\mathbf{K} = \mathbb{E}(\mathbf{\Theta}\mathbf{\Psi}\mathbf{x}\mathbf{x}^{H}\mathbf{\Psi}^{H}\mathbf{\Theta}^{H})$) has full rank.

- Once the parameters of the model have been estimated, the other set of the data is used to test the model. A mean square error is given. In general, a small mean square error is acknowledged to yield a good model and one seeks the smallest error possible.

If one is to use this procedure, one has to understand that in no way will it lead into judging the appropriateness of a model. Indeed, by adding more and more parameters to the model, one can always find a way of achieving a low mean square error by adjusting accordingly the parameters. This fact explains why some many models comply in the literature with the measurements. If the model minimizes the mean square error, then it is a **possible** candidate but the modeler can not conclude that it is a **good** candidate. Moreover, since the testing method has no real justification, many problems arise when using it.

- How does one cut the set of data? Do we use half the data to estimate the parameters and half the data to test the model? Why not using one quarter and three quarter? In the Bayesian viewpoint, this is not at all a problem as one takes into account all the data available and does not make any unjustified transformation on the data.

- If one is to use a Music or Esprit algorithm, $\mathbf{K}$ has to be full rank. This is obviously not the case for a double directional model where the steering DoD matrix $\mathbf{\Psi}$ is not always full rank since $\mathbf{K} = \mathbb{E}(\mathbf{\Theta}\mathbf{\Psi}\mathbf{x}\mathbf{x}^H\mathbf{\Psi}^H\mathbf{\Theta}^H)$.

2- Moment fitting:

Other authors [152] validate their model by finding the smallest error of a set of moments. They derive explicit theoretical formulas of the n_{th} moment $m_n(f)$ of the matrix $\mathbf{H}^H(f)\mathbf{H}(f)$ and find the optimal parameters in order to minimize:

$$\frac{1}{N} \sum_{n=1}^{N} | \frac{m_n(f)}{\hat{m}_n(f)} - 1 |$$

where

$$\hat{m}_n(f) = \frac{\text{Trace}(\mathbf{H}^H(f)\mathbf{H}(f))^n}{\text{Trace}(\mathbf{H}^H(f)\mathbf{H}(f))}$$

As previously stated, many models can minimize this criteria by adding more and more parameters and one cannot obviously conclude in this case if a model is better then the other or not. Moreover, how useful is it to have a channel that fits a certain amount of moments?[1].

The previous remarks show that when the abstract of a paper asserts: "This paper finds the theoretical predictions to accurately match data obtained in a recent measurement campaign", one has to be really cautious on the conclusions to be drawn.

---

[1]Note that if all the moments fit, then the criteria is sound in the sense that measures such as mutual information or SINR (which are of interest in communications) will behave similarly.

9.2 Bayesian Viewpoint

When judging the appropriateness of a model, Bayes[2] rules derives the posterior probability of the model. Bayes rule gives the posterior probability for the ith model according to: [3]

$$P(M_i \mid Y, I) = P(M_i \mid I) \frac{P(Y \mid M_i, I)}{P(Y \mid I)}$$

Y is the data (given by measurements), I is the prior information (ULA, far field scattering...). For comparing two models M and M_1, one has to compute the ratio:

$$\frac{P(M_1 \mid Y, I)}{P(M \mid Y, I)} = \frac{P(M_1 \mid I)}{P(M \mid I)} \frac{P(Y \mid M_1, I)}{P(Y \mid M, I)}$$

If $P(M_1 \mid Y, I) > P(M \mid Y, I)$, then one will conclude that model M_1 is better than model M. Let us now try to understand each term.

The first term, crucially important, is usually forgotten by the channel modelling community: $\frac{P(M_1 \mid I)}{P(M \mid I)}$. It favors one model or the other before the observation. As an example, suppose that the information $\{I = \text{The scatterers are near the antennas}\}$ is given. Then if one has to compare the model M (which considers ULA with far field scattering) and the model M_1 (assuming near field scattering) then one should consider $\frac{P(M_1 \mid I)}{P(M \mid I)} > 1.$ [4]

For understanding the second term, let us analyze and compare the following two specific models: the DoA based model M_{doa} and the double directional model M_{double}.

Model M_{doa}:

$$\mathbf{H}(f, t) = \frac{1}{\sqrt{s_r}} \Phi \left(\mathbf{\Theta} \bigodot D(t, f) \right)$$

with

$$D(t, f) = \begin{pmatrix} e^{-j2\pi f \tau_{1,1}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r \mathbf{v}_r)} & \cdots & e^{-j2\pi f \tau_{1,n_t}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r \mathbf{v}_r)} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi f \tau_{s_r,1}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_s^r \mathbf{v}_r)} & \cdots & e^{-j2\pi f \tau_{s_r,n_t}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_s^r \mathbf{v}_r)} \end{pmatrix}$$

deals with the DoA model taking into account the delays, Doppler effect (we suppose that the transmitting antenna does not move but only the receiving one) for a ULA (s is the number of scatterers). Let the information I on which is based the model be such that the powers of the steering directions are identical and that the transmitting antennas do not move. We recall that $\mathbf{u}_i^r \mathbf{v}_r = (\cos(\beta^r_i)\tilde{\mathbf{i}} + \sin(\beta^r_i)\tilde{\mathbf{j}})(v_r \cos(\alpha_r)\tilde{\mathbf{i}} + v_r \sin(\alpha_r)\tilde{\mathbf{j}}) = v_r \cos(\beta^r_i - \alpha_r)$

---

[2]This chapter is greatly inspired by the work of Jaynes and Bretthorst who have made the following ideas clear.

[3]We use here the notations and meanings of Jaynes [21] and Jeffrey [19]: $P(M_i \mid Y, I)$ is the "probability that the model M_i is true given that the data Y is equal to the true data y and that the information I on which is based the model is true". Every time, " (|" means conditional on the truth of the hypothesis I. In probability theory, all probabilities are conditional on some hypothesis space.

[4]The term $\frac{P(M_1 \mid I)}{P(M \mid I)}$ can be seen as the revenge of the measurement field scientist over the mathematician. It shows that modelling is both an experimental and theoretical science and that the experience of the field scientist (which attributes the values of the prior probabilities) does matter.

The set of parameters on which the model is defined is

$$p_{\text{doa}} = \{\mathbf{\Phi}, s_r, \tau, v_r, \mathbf{\Theta}, \alpha_r, \beta^r\}$$

and the parameters lie in a subspace $S_{p_{\text{doa}}}$. We recall here the DoA based model for a given frequency:

$$\mathbf{y}(t, f) = \frac{1}{\sqrt{s_r}} \mathbf{\Phi} \left(\mathbf{\Theta} \bigodot D(t, f) \right) \mathbf{x}(f) + \mathbf{n}(f)$$

The term of interest $P(\mathbf{y} \mid M_{\text{doa}}, I)$ can be derived the following way:

$$P(\mathbf{y} \mid M_{\text{doa}}, I) = \int P(\mathbf{y}, p_{\text{doa}} \mid M_{\text{doa}}, I) dp_{\text{doa}} = \int P(\mathbf{y} \mid p_{\text{doa}}, M_{\text{doa}}, I) P(p_{\text{doa}} \mid M_{\text{doa}}, I) dp_{\text{doa}}$$

Let us derive each probability distribution separately: $P(\mathbf{y} \mid p_{\text{doa}}, M_{\text{doa}}, I) =$

$$\frac{1}{\left(2\pi\sigma^2\right)^{\frac{N_1 N_r}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{N} \sum_{j=1}^{N_1} \left[y(t_j, f_i) - \frac{1}{\sqrt{s_r}} \mathbf{\Phi}(\mathbf{\Theta} \odot D(t_j, f_i)) x(f_i) \right]^H \left[y(t_j, f_i) - \frac{1}{\sqrt{s_r}} \mathbf{\Phi}(\mathbf{\Theta} \odot D(t_j, f_i)) x(f_i) \right]}$$

and

$$
\begin{aligned}
P(p_{\text{doa}} \mid M_{\text{doa}}, I) &= P(\mathbf{\Phi}, s_r, \tau, \beta^{\mathbf{r}}, v_r, \alpha_r, \mathbf{\Theta} \mid M_{\text{doa}}, I) \\
&= P(\mathbf{\Phi} \mid s_r, M_{\text{doa}}, I) P(s_r \mid M_{\text{doa}}, I) P(v_r \mid M_{\text{doa}}, I) P(\tau \mid M_{\text{doa}}, I) \\
&\quad P(\mathbf{\Theta} \mid M_{\text{doa}}, s_r, I) P(\alpha_r \mid M_{\text{doa}}, I) P(\beta^r \mid I, M_{\text{doa}})
\end{aligned}
$$

since all the priors are taken independent in the case of uninformative priors. The values of these priors have already been provided (the proof is given in chapter 7) and only the prior on $\mathbf{\Theta}$ and s_r remain to be given. We give these two priors now (and also the prior on the power although in the two models introduced for comparison, the power distribution is not needed):

- If only the mean and variance of each path is available then using maximum entropy arguments, one can show that:

$$
\begin{aligned}
P(\mathbf{\Theta} \mid s_r, M_{\text{doa}}, I) &= \frac{1}{(\sqrt{2\pi})^{n_t \times s_r}} e^{-\sum_{i=1}^{s_r} \sum_{j=1}^{n_t} |\theta_{i,j}|^2} \\
&= \frac{1}{(\sqrt{2\pi})^{n_t \times s_r}} e^{-\text{trace}(\mathbf{\Theta}\mathbf{\Theta}^H)}
\end{aligned}
$$

- How can we assign a prior probability $P(s_r \mid M_{\text{doa}}, I)$ for the unknown number of scatterers? The modeler has no knowledge if the measurements were taken in a dense area or not. The unknown number of scatterers could range from one (this prior only occurs in model that have a single bounce) up to a maximum. But what is the maximum value? There are $N \times N_1$ data values and if there were $N \times N_1$ scatterers, the data could be at most fit by placing a scatterer at each data value and adjusting the direction of arrivals. Because no additional information is available about the number of scatterers, $N \times N_1$ may be taken as an upper bound. Using the principle of maximum entropy, one obtains a uniform distribution for the number of scatterers $P(s_r \mid M_{\text{doa}}, I) = \frac{1}{N \times N_1}$.
  Note that in the general case, if one has precise available information then one has to take

it into account. But how can the modeler translate the prior on the scatterers due to the fact that the room has three chairs and a lamp in the corner? This is undoubtedly a difficult task and representing that information in terms of probabilities is not straightforward. But difficult is not impossible. The fact that there are several chairs (with respect to the case where there is no chairs) is a source of information and will lead to attributing in the latter case a peaky prior shifted around a higher number of scatterers.

- **Power**: The transmitter is limited in terms of transmit power to an upper bound value $P_{\max}^t$. Therefore, the principle of maximum entropy attributes a uniform distribution to the different amplitudes $P(P_i^t) = \frac{1}{P_{\max}^t}$, $P_i \in [0, P_{\max}^t]$. In the same vein, the receiver cannot, due to the amplifiers, process a receiving amplitude greater then $P_{\max}^r$. In this case, the principle of maximum entropy attributes a uniform distribution such as $P(P_i^r) = \frac{1}{P_{\max}^r}$, $P_i \in [0, P_{\max}^r]$

With all the previous priors given, one can therefore compute:

$$
\begin{aligned}
P(\mathbf{y} \mid M_{\mathrm{doa}}, I) &= \int \frac{1}{(2\pi\sigma^2)^{\frac{N_1 Nr}{2}}} \\
&e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{N} \sum_{j=1}^{N_1} \left\| y(t_j,f_i) - \frac{1}{\sqrt{sr}} \boldsymbol{\Phi}(\boldsymbol{\Theta} \odot D(t_j,f_i))x(f_i) \right\|^H \left\| y(t_j,f_i) - \frac{1}{\sqrt{sr}} \boldsymbol{\Phi}(\boldsymbol{\Theta} \odot D(t_j,f_i))x(f_i) \right\|} \\
&P(\boldsymbol{\Phi} \mid s_r, M_{\mathrm{doa}}, I)P(s_r \mid M_{\mathrm{doa}}, I)P(v_r \mid M_{\mathrm{doa}}, I)P(\alpha_r \mid M_{\mathrm{doa}}, I)P(\beta^r \mid M_{\mathrm{doa}}, I) \\
&P(\tau \mid M_{\mathrm{doa}}, I)P(\boldsymbol{\Theta} \mid M_{\mathrm{doa}}, I)d\boldsymbol{\Phi}d\boldsymbol{\Theta}ds_r d\tau dv_r d\alpha_r d\beta^{\mathbf{r}}
\end{aligned}
$$

which gives:

$$
\begin{aligned}
P(\mathbf{y} \mid M_{\mathrm{doa}}, I) &= \frac{1}{N \times N_1} \sum_{s_r=1}^{N \times N_1} \int_0^{2\pi} \int_0^{\infty} \int_0^{v_{\lim}} \int_0^{\tau_{\max}} \frac{1}{\left(2\pi\sigma^2\right)^{\frac{N_1 Nr}{2}}} \prod_{i=1}^{N} \prod_{j=1}^{N_1} \\
&e^{-\frac{1}{2\sigma^2} \left\| y(t_j,f_i) - \frac{1}{\sqrt{sr}} \boldsymbol{\Phi}(\boldsymbol{\Theta} \odot D(t_j,f_i))x(f_i) \right\|^H \left\| y(t_j,f_i) - \frac{1}{\sqrt{sr}} \boldsymbol{\Phi}(\boldsymbol{\Theta} \odot D(t_j,f_i))x(f_i) \right\|} \\
&(\frac{1}{\tau_{\max}})^{s_r \times n_t} \frac{1}{v_{\lim}}(\frac{1}{2\pi})^{n_r \times s_r} \frac{1}{2\pi}(\frac{1}{2\pi})^{s_r} \\
&d\phi_{11}...d\phi_{n_r s_r}d\theta_{11}...d\theta_{s_r n_t}d\tau_{11}...d\tau_{s_r n_t}dv_r d\alpha_r d\beta^r{}_1...d\beta^r{}_{s_r} \qquad (9.1)
\end{aligned}
$$

As one can see, the numerical integration is tedious but it is the only way to rank the models in an appropriate manner.

Model $M_{\mathbf{double}}$:

Let us now derive model M_{double}:

$$
\mathbf{H}(f,t) = \frac{1}{\sqrt{s_r s_t}} \boldsymbol{\Phi}\left(\boldsymbol{\Theta} \odot D(t,f)\right)\boldsymbol{\Psi}
$$

with

$$
D(t,f) = \begin{pmatrix} e^{-j2\pi f\tau_{1,1}}e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r \mathbf{v}_r)} & \cdots & e^{-j2\pi f\tau_{1,s_t}}e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r \mathbf{v}_r)} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi f\tau_{s_r,1}}e^{j2\pi \frac{ft}{c}(\mathbf{u}_{s_r}^r \mathbf{v}_r)} & \cdots & e^{-j2\pi f\tau_{s_r,s_t}}e^{j2\pi \frac{ft}{c}(\mathbf{u}_{s_r}^r \mathbf{v}_r)} \end{pmatrix}
$$

deals with the double directional model for which the set of parameters is

$$
p_{\mathrm{double}} = \{\boldsymbol{\Phi}, s_r, \boldsymbol{\Psi}, s_t, \tau, v_r,, \alpha_r, \beta_{\mathbf{r}}\boldsymbol{\Theta}\} = \{p_{\mathrm{doa}}, \boldsymbol{\Phi}, s_t\}
$$

by adding two new parameters $\boldsymbol{\Psi}$ and s_t and going to the new subspace $S_{p_{\text{double}}}$ in such a way that $\boldsymbol{\Psi} = \mathbf{F}_{n_t}$ $(n_t = s_t)$ represents model M_{doa}. Indeed, in this case, we have:

$$
\left(\boldsymbol{\Theta} \bigodot D(t,f)\right)\mathbf{F}_{n_t} = \begin{pmatrix} \sum_{i=1}^{n_t} \theta_{1i}e^{-j2\pi f\tau_{1,i}}e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r\mathbf{v}_r)} & \cdots & \sum_{i=1}^{n_t} \theta_{1i}e^{-j2\pi f\tau_{1,i}}e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r\mathbf{v}_r)}e^{j2\pi \frac{(n_t-1)i}{n_t}} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n_t} \theta_{s_r i}e^{-j2\pi f\tau_{s_r,i}}e^{j2\pi \frac{ft}{c}(\mathbf{u}_s^r\mathbf{v}_r)} & \cdots & \sum_{i=1}^{n_t} \theta_{s_r i}e^{-j2\pi f\tau_{s_r,i}}e^{j2\pi \frac{ft}{c}(\mathbf{u}_{s_r}^r\mathbf{v}_r)}e^{j2\pi \frac{(n_t-1)i}{n_t}} \end{pmatrix}
$$

$$
= \begin{pmatrix} \sum_{i=1}^{n_t} \theta_{1i}e^{-j2\pi f(\tau_{1,i}-\tau_{1,1})} & \cdots & \sum_{i=1}^{n_t} \theta_{1i}e^{-j2\pi f(\tau_{1,i}-\tau_{1,n_t})}e^{j2\pi \frac{(n_t-1)i}{n_t}} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n_t} \theta_{s_r i}e^{-j2\pi f(\tau_{s_r,i}-\tau_{s_r,1})} & \cdots & \sum_{i=1}^{n_t} \theta_{s_r i}e^{-j2\pi f(\tau_{s_r,i}-\tau_{s_r,n_t})}e^{j2\pi \frac{(n_t-1)i}{n_t}} \end{pmatrix} \bigodot D(t,f)
$$

$$
= \Theta_1 \bigodot D(t,f)
$$

Where Θ_1 is a matrix with i.i.d Gaussian entries.

We recall here the model for a given frequency:

$$
\mathbf{y}(f,t) = \frac{1}{\sqrt{s_r s_t}}\boldsymbol{\Phi}\left(\boldsymbol{\Theta} \bigodot D(t,f)\right)\boldsymbol{\Psi}\mathbf{x}(f) + \mathbf{n}(f)
$$

The same methodology applies and we have:

$$
P(\mathbf{y} \mid M_{\text{double}}, I) = \int \frac{1}{(2\pi\sigma^2)^{\frac{N_1 Nr}{2}}}
$$

$$
e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{N_1}\left|y(t_j,f_i)-\frac{1}{\sqrt{s_r s_t}}\boldsymbol{\Phi}(\boldsymbol{\Theta} \odot D(t_j,f_i))\boldsymbol{\Psi}x(f_i)\right|^{H}\left|y(t_j,f_i)-\frac{1}{\sqrt{s_r s_t}}\boldsymbol{\Phi}(\boldsymbol{\Theta} \odot D(t_j,f_i))\boldsymbol{\Psi}x(f_i)\right|}
$$

$$
P(\boldsymbol{\Phi} \mid s_r, M_{\text{double}}, I)P(s_r \mid M_{\text{double}}, I)P(\boldsymbol{\Psi} \mid s_t, M_{\text{double}}, I)P(s_t \mid M_{\text{double}}, I)
$$

$$
P(v_r \mid M_{\text{double}}, I)P(\alpha_r \mid M_{\text{double}}, I)P(\beta_{\mathbf{r}} \mid M_{\text{double}}, I)P(\tau \mid M_{\text{double}}, I)
$$

$$
P(\boldsymbol{\Theta} \mid M_{\text{double}}, I)d\boldsymbol{\Phi}d\boldsymbol{\Psi}d\boldsymbol{\Theta}ds_r ds_t d\tau dv_r d\alpha_r d\beta^{\mathbf{r}}
$$

and

$$
P(\mathbf{y} \mid M_{\text{double}}, I) = \left(\frac{2}{N \times N_1}\right)^2 \sum_{s_r=1}^{\frac{N \times N_1}{2}} \sum_{s_t=1}^{\frac{N \times N_1}{2}} \int_0^{2\pi} \int_0^{\infty} \int_0^{v_{\text{lim}}} \int_0^{\tau_{\text{max}}} \frac{1}{(2\pi\sigma^2)^{\frac{N_1 Nr}{2}}} \prod_{i=1}^{N}\prod_{j=1}^{N_1}
$$

$$
e^{-\frac{1}{2\sigma^2}\left|y(t_j,f_i)-\frac{1}{\sqrt{s_r s_t}}\boldsymbol{\Phi}(\boldsymbol{\Theta} \odot D(t_j,f_i))\boldsymbol{\Psi}x(f_i)\right|^{H}\left|y(t_j,f_i)-\frac{1}{\sqrt{s_r s_t}}\boldsymbol{\Phi}(\boldsymbol{\Theta} \odot D(t_j,f_i))\boldsymbol{\Psi}x(f_i)\right|}
$$

$$
\left(\frac{1}{\tau_{\text{max}}}\right)^{s_r \times s_t} \frac{1}{v_{\text{lim}}}\left(\frac{1}{2\pi}\right)^{n_r \times s_r}\left(\frac{1}{2\pi}\right)^{s_t \times n_t} \frac{1}{2\pi}\left(\frac{1}{2\pi}\right)^{s_r}
$$

$$
d\phi_{11}...d\phi_{n_r s_r}d\psi_{11}...d\psi_{1n_t}d\theta_{11}...d\theta_{s_r n_t}d\tau_{11}...d\tau_{s_r n_t}dv_r d\alpha_r d\beta^r{}_1...d\beta^r{}_{s_r} \quad (9.2)
$$

A common problem in the modelling process is the following: suppose, when testing the models with the data, that both models M and M_1 have the same maximum likelihood, in other words:

$$
P(\mathbf{y} \mid p_{\text{doa}}{}^{\text{max}}, M_{\text{doa}}, I) = P(\mathbf{y} \mid p_{\text{double}}{}^{\text{max}}, M_{\text{double}}, I)
$$

Which model should we choose? Hereafter, we give an example to show that Bayesian probability will choose the model with the smallest number of parameters.

First of all, we will suppose that the information I available does not give a preference to model before seeing the data: $P(M_{\text{double}} \mid I) = P(M_{\text{doa}} \mid I)$.

As previously shown,

$$
\begin{aligned}
P(\mathbf{y} \mid M_{\text{doa}}, I) &= \int P(\mathbf{y}, p_{\text{doa}} \mid M_{\text{doa}}, I) dp_{\text{doa}} \\
&= \int P(\mathbf{y} \mid p_{\text{doa}}, M_{\text{doa}}, I) P(p_{\text{doa}} \mid M_{\text{doa}}, I) dp_{\text{doa}}
\end{aligned}
$$

and

$$
\begin{aligned}
P(\mathbf{y} \mid M_{\text{double}}, I) &= \int P(\mathbf{y}, p_{\text{double}} \mid M_{\text{double}}, I) dp_{\text{double}} \tag{9.3} \\
&= \int P(\mathbf{y} \mid p_{\text{double}}, M_{\text{double}}, I) P(p_{\text{double}} \mid M_{\text{double}}, I) dp_{\text{double}} \tag{9.4}
\end{aligned}
$$

Since

$$
\begin{aligned}
P(p_{\text{double}} \mid M_{\text{double}}, I) &= P([p_{\text{doa}}, \mathbf{\Psi}, s_t] \mid M_{\text{double}}, I) \\
&= P(p_{\text{doa}} \mid \mathbf{\Psi}, s_t, M_{\text{double}}, I) P(\mathbf{\Psi}, s_t \mid M_{\text{double}}, I)
\end{aligned}
$$

From equation (9.3), we have:

$$
\begin{aligned}
P(\mathbf{y} \mid M_{\text{double}}, I) = \int \int & P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi}, s_t], M_{\text{double}}, I) P(p_{\text{doa}} \mid \mathbf{\Psi}, s_t, M_{\text{double}}, I) \\
& P(\mathbf{\Psi}, s_t \mid M_{\text{double}}, I) dp_{\text{doa}} d\mathbf{\Psi} ds_t
\end{aligned}
$$

In the following, we will suppose that the likelihood function $P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi}, s_t], M_{\text{double}}, I)$ is peaky around the maximum likelihood region and has near zero values elsewhere. Otherwise, the measurement data Y would be useless in the sense that the data does not provide any information. Suppose now that with model M_{double}, the maximum likelihood $P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi}, s_t] M_{\text{double}}, I)$ occurs at a point near $\mathbf{\Psi} = \mathbf{F}_t$ and $s_t = n_t$ for the parameters $\mathbf{\Psi}$ and s_t in other words $P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi}, s_t], M_{\text{double}}, I)$ is always null except for the value of $\mathbf{\Psi} = \mathbf{F}_t$ and $s_t = n_t$ then:

$$
\begin{aligned}
P(\mathbf{y} \mid M_{\text{double}}, I) = \int \int \int & P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi}, s_t], M_{\text{double}}, I) P(p_{\text{doa}} \mid \mathbf{\Psi}, s_t M_{\text{double}}, I) \\
& P(\mathbf{\Psi}, s_t \mid M_{\text{double}}, I) dp_{\text{doa}} d\mathbf{\Psi} ds_t \\
\approx \int & P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t], M_{\text{Double}}, I) \\
& P(p_{\text{doa}} \mid [\mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t] M_{\text{double}}, I) \tag{9.5} \\
& P([\mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t] \mid M_{\text{double}}, I) dp_{\text{doa}} \tag{9.6}
\end{aligned}
$$

One has to notice that $P(p_{\text{doa}} \mid [\mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t], M_{\text{double}}, I) = P(p_{\text{doa}} \mid M_{\text{doa}}, I)$ and $P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t], M_{\text{double}}, I) = P(\mathbf{y} \mid p_{\text{doa}}, M_{\text{doa}}, I)$ since both models are the same when $\mathbf{\Psi} = \mathbf{F}_{n_t}$ and $s_t = n_t$. We also have $P(\mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t \mid M_{\text{double}}, I) \leq 1$ (In fact, we can derive the exact value. Indeed, since we have no knowledge of the directions of arrival,

$P(\mathbf{\Phi} = \mathbf{F}_{n_t}, s_t = n_t \mid M_{\text{double}}, I) = \frac{1}{(2\pi)^{n_r \times s_t}})$. Using equation (9.5), Bayesian probability shows us that:

$$
\begin{aligned}
P(\mathbf{y} \mid M_{\text{double}}, I) &\leq \int P(\mathbf{y} \mid [p_{\text{doa}}, \mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t], M_{\text{Double}}, I) \\
&\qquad P(p_{\text{doa}} \mid [\mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t] M_{\text{double}}, I) P([\mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t] \mid M_{\text{double}}, I) dp_{\text{doa}} \\
&= \int P(\mathbf{y} \mid p_{\text{doa}}, M_{\text{doa}}, I) P(p_{\text{doa}} \mid M_{\text{doa}}, I) P([\mathbf{\Psi} = \mathbf{F}_{n_t}, s_t = n_t] \mid M_{\text{double}}, I) dp_{\text{doa}} \\
&\leq \int P(\mathbf{y} \mid p_{\text{doa}}, M_{\text{doa}}, I) P(p_{\text{doa}} \mid M_{\text{doa}}, I) dp_{\text{doa}} \\
&= \int P(\mathbf{y}, p_{\text{doa}} \mid M_{\text{doa}}, I) dp_{\text{doa}} \\
&\qquad P(\mathbf{y} \mid M_{\text{doa}}, I)
\end{aligned}
$$

Since M_{doa} has less parameters then M_{double}, Bayesian probability will favor the model M_{doa} with less parameters and therefore shows that "the best explanation is always the simplest"[5]. It is therefore wrong to think that by increasing the number of parameters one can always find a good model: one can indeed better fit the data to the model (expression $P(\mathbf{y} \mid p_{\text{doa}}, M_{\text{doa}}, I)$) but the prior probability $P(p_{\text{doa}} \mid M_{\text{doa}}, I)$ will spread over a larger space and assign as a consequence a lower value to $P(\mathbf{y} \mid M_{\text{doa}}, I)$.

But how does the a posteriori computation compare with the usual methodology of maximizing the likelihood $P(\mathbf{y} \mid p, M, I)$?

Following [21], let us expand $\log P(\mathbf{y} \mid p, M, I)$ around the maximum likelihood point $\hat{p} = \{p^1{}_{\text{max}}, ..., p^m{}_{\text{max}}\}$

$$
\log P(\mathbf{y} \mid p, M, I) = \log P(\mathbf{y} \mid p_{\text{max}}, M, I) + \frac{1}{2} \sum_{i,j=1}^{m} \frac{d^2 \log(P)}{dp^i dp^j} (p^i - p^i{}_{\text{max}})(p^j - p^j{}_{\text{max}}) + O()
$$

then near the peak a good approximation is a multivariate Gaussian such as:

$$
P(\mathbf{y} \mid p, M, I) = P(\mathbf{y} \mid p_{\text{max}}, M, I) e^{-\frac{1}{2}(p - p_{\text{max}}) \Delta^{-1} (p - p_{\text{max}})}
$$

with the inverse covariance matrix defined as:

$$
\Delta^{-1}{}_{ij} = \left(\frac{d^2 \log(P)}{dp^i dp^j} \right)_{\pi = \pi_{\text{max}}}
$$

Therefore,

$$
\begin{aligned}
P(\mathbf{y} \mid M, I) &= P(\mathbf{y} \mid p_{\text{max}}, M, I) \int e^{-\frac{1}{2}(p - p_{\text{max}}) \Delta^{-1} (p - p_{\text{max}})} P(p \mid M, I) dp \\
&= P(\mathbf{y} \mid p_{\text{max}}, M, I) G(M, I)
\end{aligned}
$$

---

[5] In statistical inference, this is known as Occam's razor. William of Occam was a theologian of the 14th century who wrote against the papacy in a series of treatise in which he tried to avoid many established pseudo explanations. In his terms, the logic of simplicity was stated in the following form "Causes shall not be multiplied beyond necessity" [100]. Note that Occam's razor has been extended to other fields such as metaphysics where it is interpreted as "nature prefers simplicity".

All the tools are now provided to better understand what is happening. Suppose we want to compare two models M and M_1. The a posteriori probability ratio for model M over M_1 is:

$$
\begin{aligned}
\frac{P(M \mid \mathbf{y}, I)}{P(M_1 \mid \mathbf{y}, I)} &= \frac{P(M \mid I)}{P(M_1 \mid I)} \frac{P(\mathbf{y} \mid M, I)}{P(\mathbf{y} \mid M_1, I)} \\
&= \frac{P(M \mid I)}{P(M_1 \mid I)} \frac{P(\mathbf{y} \mid p^{\max}, M, I)}{P(\mathbf{y} \mid p_1^{\max}, M_1, I)} \frac{G(M, I)}{G(M_1, I)}
\end{aligned}
$$

In the conventional methods, M is better than M_1 if $\frac{P(\mathbf{y}|p^{\max}M,I)}{P(\mathbf{y}|p_1^{\max}M_1,I)} > 1$ which is only one part of the three terms to be computed. In fact, in order to compare two models, three terms have to be calculated and

the mistake persists thinking that any model M_1 versus M is good as long as we increase the number of parameters: indeed, the fitting will get better and the ratio $\frac{P(\mathbf{y}|p^{\max},M,I)}{P(\mathbf{y}|p_{1\max},M_1,I)}$ will decrease but this is only looking at one part of the problem. First of all, one has to consider $\frac{P(M|I)}{P(M_1|I)}$ and moreover $\frac{G(M,I)}{G(M_1,I)}$. This last term depends on the prior information about the internal parameters and as the number of parameters increases this term decreases due to the fact that we add more and more uninformative priors.

The previous discussion shows that different methodologies can be used for model validation with various limitations (complexity, adequacy,...). In the following, we propose a metric based on mutual information and SINR compliance and discuss its features.

9.3   Mutual Information Complying Models

In usual communications systems, engineers are interested in models that fulfill a certain criteria. The fact that the model is adequate or not is not an issue as long as it reproduces in an accurate manner the same performance as measurements in the simulated chain. The criteria range from BER, Signal to interference ratio to mutual information. The mutual information is an interesting criteria from a network planning perspective.

In the previous chapters, we have derived the mutual information distribution of many models and shown that in many cases, the cumulative distribution function of the mutual information has the following form:

$$
F(I^M) = 1 - Q\left(\frac{I^M - n_t \mu}{\sigma}\right)
$$

In each case (i.i.d, DoA based, DoD based and double directional), expressions of μ and σ have been provided.

For a given frequency f, a model will be called mutual information complying if it minimizes:

$$
\int_0^{\infty} \mid F(I^M) - F_{\text{empirical}}(I^M, f) \mid^2 dI^M \tag{9.7}
$$

Here $F_{\text{empirical}}(I^M, f)$ is the empirical CDF given by measurements.
Note that we could also minimize the Kullback distance between the two distributions:

$$
D(P, P_{\text{empirical}}) = \int P(I^M) \log\left(\frac{P(I^M)}{P_{\text{empirical}}(I^M)}\right) dI^M \tag{9.8}
$$

where P and $P_{\text{empirical}}$ are respectively the theoretical and empirical probability distribution of the mutual information. From a practical point of view, expression (9.8) is quite easy to optimize. Indeed, let $\mu_{\text{empirical}}$ and $\sigma_{\text{empirical}}$ be respectively the empirical mean and empirical variance of $P_{\text{empirical}}(I^M)$. Writing $P_{\text{empirical}}(I^M) = \frac{1}{\sqrt{2\pi\sigma_{\text{empirical}}^2}} e^{-\frac{(I^M - \mu_{\text{empirical}})^2}{2\sigma_{\text{empirical}}^2}}$ (chapter 11 will show that $P_{\text{empirical}}$ is indeed Gaussian) , one has to find the parameters μ and σ which minimize the following expression:

$$
\begin{aligned}
D(P, P_{\text{empirical}}) &= -\frac{1}{2}\log(2\pi e\sigma^2) - \int P(I^M)\log\left(P_{\text{empirical}(I^M)}\right) dI^M \\
&= \log(\frac{\sigma_{\text{empirical}}}{\sigma}) + \frac{(\mu - \mu_{\text{empirical}})}{2\sigma_{\text{empirical}}^2}
\end{aligned}
$$

.

In general (except for the i.i.d Gaussian case where there is nothing to do), for minimizing the criteria in the directional cases, one has to optimize criteria (9.7) with respect to the eigenvalue distribution of the steering matrix (since μ and σ depend on the limiting eigenvalue distribution of the steering matrix. This is not an easy task. However, since we are interested in mutual information issues, only the non-correlated scatterers (called here the dominant scatterers) scale the mutual information and therefore we can use (as a first approximation) the Fourier models developed previously (which correspond to Sayeed's model since we are interested in capacities issues). In this case, let us review each model (ρ is the SNR):

I.I.D Gaussian model. There is no optimization to perform in this case and μ and σ are equal to:

$$
\mu(n_r, n_t, \rho) = \frac{n_r}{n_t}\ln(1 + \rho - \rho\alpha(n_r, n_t, \rho)) + \ln(1 + \rho\frac{n_r}{n_t} - \rho\alpha(n_r, n_t, \rho)) - \alpha(n_r, n_t, \rho)
$$

and

$$
\sigma^2(n_r, n_t, \rho) = -\ln[1 - \frac{n_t\alpha^2(n_r, n_t, \rho)}{n_r}]
$$

with

$$
\alpha(n_r, n_t, \rho) = \frac{1}{2}[1 + \frac{n_r}{n_t} + \frac{1}{\rho} - \sqrt{(1 + \frac{n_r}{n_t} + \frac{1}{\rho})^2 - 4\frac{n_r}{n_t}}]
$$

DoA based model. In this case, one has to optimize criteria (9.7) with respect to the number of scatterers s for which the expression of μ_{doa} and σ_{doa} are given by:

$$
\mu_{\text{doa}}(n_r, s_R, n_t, \rho) = \frac{s_r}{n_t}\ln(1 + \rho\frac{n_r}{s_r} - \rho\frac{n_r}{s_r}\alpha_{\text{doa}}(n_r, s_r, n_t, \rho)) + \ln(1 + \rho\frac{n_r}{n_t} - \rho\frac{n_r}{s_r}\alpha_{\text{doa}}(n_r, s_r, n_t, \rho)) - \alpha_{\text{doa}}(n_r
$$

and

$$
\sigma_{\text{doa}}^2(n_r, s_r, n_t, \rho) = -\ln[1 - \frac{n_t\alpha_{\text{doa}}^2(n_r, s_r, n_t, \rho)}{s_r}]
$$

with

$$\alpha_{\text{doa}}(n_r, s_r, n_t, \rho) = \frac{1}{2}[1 + \frac{s_r}{n_t} + \frac{1}{\rho\frac{n_r}{s_r}} - \sqrt{(1 + \frac{s_r}{n_t} + \frac{1}{\frac{n_r}{s_r}\rho})^2 - 4\frac{s_r}{n_t}}]$$

Since s_r will depend on the frequency, we will define as in [152] the richness spectrum as:
$R_{\text{doa}} = \frac{s_r(f)}{n_r}$

DoD based model. In this case, one has to optimize criteria (9.7) with respect to the number of scatterers s_t for which the expression of μ_{dod} and σ_{dod} are given by:

$$\mu_{\text{dod}}(n_r, s_t, n_t, \rho) = \frac{s_t}{n_t}\ln(1 + \rho\frac{n_r}{s_t} - \rho\alpha_{\text{dod}}(n_r, s_t, n_t, \rho)) + \frac{n_r}{n_t}\ln(1 + \rho - \rho\alpha_{\text{dod}}(n_r, s_t, n_t, \rho))$$
$$- \frac{s_t}{n_t}\alpha_{\text{dod}}(n_r, s_t, n_t, \rho)$$

and

$$\sigma_{\text{dod}}^2(n_r, s_t, n_t, \rho) = -\ln[1 - \frac{s_t}{n_r}\alpha_{\text{dod}}^2(n_r, s_t, n_t, \rho)]$$

with

$$\alpha_{\text{dod}}(n_r, s_t, n_t, \rho) = \frac{1}{2}[1 + \frac{n_r}{s_t} + \frac{1}{\rho} - \sqrt{(1 + \frac{n_r}{s_t} + \frac{1}{\rho})^2 - 4\frac{n_r}{s_t}}]$$

Since s_t will also depend on the frequency, we can also define the richness spectrum as:
$R_{\text{dod}} = \frac{s_t(f)}{n_t}$

Double Directional model. In this case, one has to optimize criteria (9.7) with respect to s_r (scatterers at the receiving side) and s_t (scatterers at the transmitting side) for which the expressions of μ_{double} and σ_{double} are given by:

$$\mu_{\text{double}}(n_r, s_r, s_t, n_t, \rho) = \frac{s_r}{n_t}\ln(1 + \rho\frac{n_r}{s_r} - \rho\frac{n_r}{s_r}\alpha_{\text{double}}(n_r, s_r, s_t, n_t, \rho))$$
$$+ \frac{s_t}{n_t}\ln(1 + \rho\frac{n_r}{s_t} - \rho\frac{n_r}{s_r}\alpha_{\text{double}}(n_r, s_r, s_t, n_t, \rho))$$
$$- \frac{s_t}{n_t}\alpha_{\text{double}}(n_r, s_r, s_t, n_t, \rho)$$

and

$$\sigma_{\text{double}}^2(n_r, s_r, s_t, n_t, \rho) = -\ln(1 - \frac{\alpha_{\text{double}}^2 s_t}{s_r})$$

with

$$\alpha_{\text{double}}(n_r, s_r, s_t, n_t, \rho) = \frac{1}{2}\left[1 + \frac{s_r}{s_t} + \frac{s_r}{\rho n_r} - \sqrt{(1 + \frac{s_r}{s_t} + \frac{s_r}{\rho n_r})^2 - 4\frac{s_r}{s_t}}\right]$$

As previously, the richness spectra can be defined independently on both sides. Note that having a model that gives the same mutual information as measurements does not validate at all the model but gives a model tool for simulating a capacity network. If the criteria changes,, the model may be completely inadequate even though it complies with mutual information measurements. As matter of fact, the mutual information criteria as several drawbacks as many models can give the same mutual imformation compliance. Many models can give the

same mutual information compliance. For instance, the DoA based model always gives the same mutual information compliance as the DoD based model, since mutual information is a symmetric measure with respect to channel input $\mathbf{x}$ and channel output $\mathbf{y}$, since

$$I(\mathbf{x}; \mathbf{y}) = I(\mathbf{y}; \mathbf{x}). \tag{9.9}$$

In other words: We can exchange transmitter and receiver without any change of mutual information. Of course, DoAs then become DoDs and vice versa. In fact, more crucially, if one is interested in mutual information compliance, then the double directional model will always give a better result than the i.i.d, DoA or DoD based. Why? The reason is simple to understand: the entropy framework was developed in this paper in order to provide consistent models. Retrieving or adding parameters takes us one step back or forward. As an example, suppose that the DoA based model complies quite accurately with the mutual information for a given parameter of scatterers. Then, immediately, the double directional model will also be a good candidate if one takes the same optimized parameter s_r and $s_t = n_t$. As one can see, the mutual information compliance criteria will never be able to discriminate between a well parameterized and an over parameterized model. This is mainly due to the fact that the models are consistent but not the criteria (see section 9.5). The mutual information compliance criteria is a necessary but not sufficient condition. The misunderstanding occurs as one uses a mono-dimensional measure to validate a multi-dimensional model. In other words, one is only validating a functional of the model and not the model itself. Therefore, many models can give the same mutual information compliance. Therefore, if the model gives accurate mutual information results, we will not pretend that the model is good (as many papers do) but only that it fulfills our mutual information criteria. Of course, other quality measures can be given.

Remark 3 *The measurement scientist, which is mostly interested in practical models, may have got bored, when reading the document, with all the previous chapters dedicated to derive theoretical formulas of the mutual information. Indeed, what are the theoretical formulas useful for since one can simulate the models derived and test them? There are many answers to give but the author will only focus here on one. As the reader can see from our formulas, the optimization is extremely easy to perform as we have clearly, thanks to the previous analysis, derived the parameters of interest. There is only a one dimensional search over s_r and s_t to perform and one can therefore, for each scenario, find the number of scatterers in a computational efficient way. The other aspects which are related to the information theoretical transmission limits have already been detailed previously.*

9.4   SINR complying Models

************** ***

9.5   Clearing up Mysteries

In the previous chapters, we have derived the mutual information 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
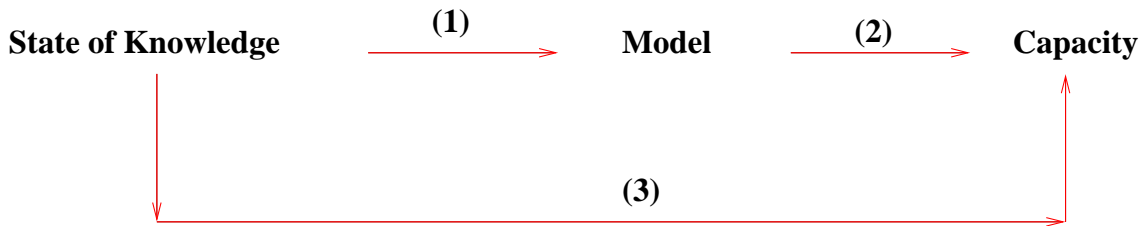
Figure 9.1: Channel modelling approach and derivation of capacity.

to ensure a reliable transmission on such channels. But whenever we are misrepresenting the channel with our state of knowledge, the formula

$$I^M = \log_2 \det\left(\mathbf{I}_{n_t} + \frac{\rho}{n_t}\mathbf{H}^H\mathbf{H}\right) \qquad (9.10)$$

will misestimate the rate. Indeed, a surprising fact in our maximum entropy approach is that although it gives us a consistent model with our state of knowledge, it will also lead to misestimating the rate with formula (9.10). The problem is formulated in Figure 9.1.

- Transition (1): the modeler creates a model maximizing entropy.

- Transition (2): The modeler misestimates the real achievable rate because even though the model he creates is the best he can base on his state of knowledge, he derives the mutual information of the channel based on the assumption that the model is reality.

- Transition (3): A new measure of information rate should be derived based only on our state of knowledge, taking into account the fact that the model does not represent reality, but only our knowledge (which is scarce) of reality

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

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

Δ is the infinite set of matrices $\mathbf{H}$ with the same initial physical constraints (mean and variance for example). Of course, other measures of capacity performance can be derived.

So, is there a contradiction in our maximum entropy modelling approach? No, as long as we understand the meaning of transition (2) in Figure 9.1. With the maximum entropy approach, we derive a channel model having as much degrees of freedom as possible (but still with the constraints of our state of knowledge) in order to cope with all the cases when they happen. We do this because we need a **unique** model consistent with our state of knowledge. Any other approach will add unjustified constraints. Suppose, for sake of simplicity, that each frequency path of the channel has a zero mean and a given variance (the mean and variance are here our state of knowledge). Transition (1)+(2) will give us a measure of the rate one can transmit on a "maximum entropy channel state knowledge".

The problem stems from the fact that although models are consistent, functionals of the model

are not. Indeed, consider the DoA based model: $\mathbf{H} = \mathbf{\Phi\Theta}$, ($\mathbf{\Theta}$ is i.i.d Gaussian) then using Jensen's Inequality:

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

For example, when the directions of arrival are unknown, the mutual information averaged across the unknown directions of arrival (here $\mathbb{E}_{\mathbf{\Phi}}\left(\log\det(\mathbf{I}_{n_t} + \frac{\rho}{n_t}(\mathbf{\Theta}^H\mathbf{\Phi}^H\mathbf{\Phi\Theta})\right)$) does not yield the mutual information of the Gaussian i.i.d. model ($\log\det(\mathbf{I}_{n_t} + \frac{\rho}{n_t}\mathbb{E}_{\mathbf{\Phi}}(\mathbf{\Theta}^H\mathbf{\Phi}^H\mathbf{\Phi\Theta}))$ which is equal to $\log\det(\mathbf{I}_{n_t} + \frac{\rho}{n_t}(\mathbf{\Theta}^H\mathbf{\Theta})))$): the model is consistent but not the functional. A remarkable feature of the previous result is that whenever we have more information (and therefore more constraints on the channel model), mutual information will be reduced as it constraints the degrees of freedom.

This explains why, under the same initial constraints (as an example the mean and the variance of each path), correlated fading reduces the mutual information with respect to the completely i.i.d case. As an example, the fact that we take into account the DoA, mean, variance will reduce the mutual information compared with the case where only the same mean and same variance are taken into account[6].

In fact, if one is interested only in one or some functions of the model, then the modler should construct a model which is consistent with those functionals and not in itself. A consistent model is for the case where we do not know which functions we (or others who we construct the model for) are interested in.

---

[6]A question to which we still did not find an answer was whether there is a relationship between entropy and the notion of diversity order and degrees of freedom as used in [58, 153, 85].

Chapter 10

Measurements Description

10.1 Measurement Set-up

In this section[1], we describe the wideband outdoor measurement campaign carried out in Oslo during summer 2002 and summer 2003 [156]. The measurements were performed at a center frequency of 2.1 GHz and 5.2 GHz with a bandwidth of 100 MHz in three different urban scenarios: a regular street grid scenario, an open city place and an indoor cell site. In each scenario, many routes have been measured: at 2.1 GHz, 150 routes have been measured (and in each case, many snapshots have been measured) whereas at 5.2 GHz, 79 routes have been measured. The measurements performed at 2.1 GHz are relevant for UMTS whereas those at 5.2 GHz are valuable for IEEE 802.11a. **Measuremens have also bee conducted at 5.2 Ghz [157]**

- The street grid scenario is in Oslo downtown and corresponds to Concrete/Brick buildings. The buildings are around 20-30 m high (see Figure 10.1). In this scenario, two different receiver positions were tested, one high position on a roof terrace and the other one, a low position on street level. The area is often referred as "Kvadraturen" (Urban Regular in the following).

  - For the low base station at 2.1 GHz, route Kvadraturen_01_01_15 has been studied.
  - For the high base station at 2.1 GHz, route Kvadraturen_02_01_14 has been studied.
  - For the low base station at 5.25 GHz[2], route Kvadraturen_03_05_09 has been studied.

- The urban open place is also in downtown Oslo and corresponds to an almost quadratic open market square of approximatively 100×100 meters. In Oslo, the square is called "Youngstorget". The square is partly filled with market stalls especially during the summer months (see Figure 10.2). The surrounding buildings are of variable size. In this scenario, the receiver was placed above some arcades.

  - At 2.1 GHz, route Youngstorget_01_02_02 has been studied.

---

[1]The author is grateful to Helmut Hofstetter and P.H. Lehne for useful discussions and providing the measurements. A comprehensive introduction to the measurement set-up can be found in [154] and [155]. For the different scenario routes, notations of [154] and [155] will be used.

[2]At 5.25 GHz, only the case where the receiver was at the street level was considered and not the high base station case.
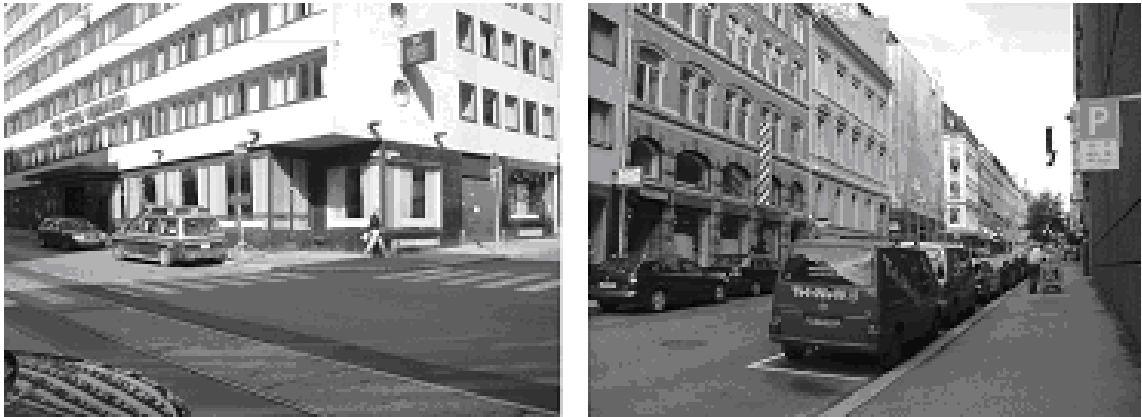
Figure 10.1: Urban Regular site.

Figure 10.2: Urban Open Place.

> – At 5.25 GHz, route Youngstorget_03_02_03 has been studied.

- For the indoor scenario, the measurements were performed in a modern office building with open indoor areas. The building (Telenor headquarters building at Fornebu) has a irregular structure and is mostly of glass and steel. Measurements were taken in two different parts of the building. Inside a work zone (see Figure 10.3 and 10.4) and in a common area called the "Atrium" (see Figure 10.5).

  – For the indoor scenario at 2.1 GHz, route Fornebu_In_01_01_19 has been studied.
  – For the "Atrium" scenario at 2.1 GHz, route Fornebu_Atrium_01_01_12 has been studied.
  – For the indoor scenario at 5.25 GHz, route FornebuIn_02_01_16 has been studied.
  – For the "Atrium" scenario at 5.25 GHz, route Fornebu_Atrium_02_01_15 has been studied.

In all the measurement set-up, a wideband channel sounder with synchronized switching between transmitter and receiver was used. The transmitter was placed arbitrarily and used as the mobile

Figure 10.3: Telenor Headquarters: work zone.

Figure 10.4: Telenor Headquarters: work zone.

Figure 10.5: Telenor Headquarters: "Atrium".
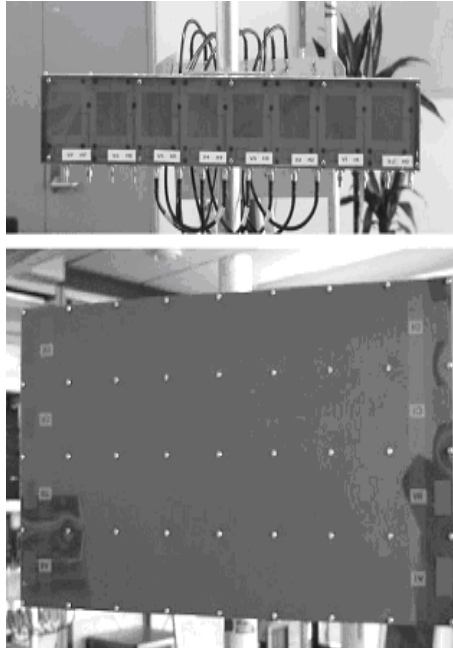
Figure 10.6: Antenna: The receiver is the 8×4 planer array while the transmitter is the 8 element uniform linear array

part, mounted on a trolley. Both transmitter and receiver antennas are broadband patch arrays with integrated switching networks.

10.2   2.1 GHz

10.2.1   Channel Sounder

The transmitter is an 8 element uniform linear array (ULA) while the receiver antenna is an 8×4 planar array, i.e two dimensional with 8 elements horizontally and 4 vertically, giving a total of 32 elements (see Figure 10.6). In all the cases, the receiver acted as the base station and only 8 elements were used. The transmitter antenna was connected using the 4 center elements with both polarizations. The multiplexing was the following:

$$H2 - V2 - H3 - V3 - H4 - V4 - H5 - V5$$

The main channel sounder specification are listed on the following Table (10.1). The sounder was manufactured by SINTEF Telecom and Informatics in Trondheim, Norway, on assignment from Telenor.

10.2.2   State of Knowledge

According to the specifications in table (10.1) and the measurement set-up, our state of knowledge is the following:

Measurement frequency	2.1 GHz
Measurement bandwidth	100 Mhz
Delay resolution	10ns
Sounding signal	linear frequency chirp
Transmitter antenna	8 element ULA
Element spacing	71.4 mm (0.5 λ)
Receiver antenna	32 (8 × 4) element
Element spacing	73.0 mm (0.51 λ)

Table 10.1: Channel sounder specification at 2.1 GHz.

- Uniform Linear Array.

- Far field Approximation.

- Antenna spacing 0.5λ.

- 4 dual-polarized transmit antenna elements.

- 8 single-polarized receive antenna elements.

Since there is no LOS between the transmitter and receiver, we can assume that the channel has zero mean and a certain energy (as recalled previously, this is at the heart of our inference methodology: if the model does not comply with the measurements, then we will retrieve that hypothesis and take the outcome of the measurement test as some new information evidence on how the model behaves when zero mean is assumed). We will also make the hypothesis that the models are directional. Moreover, since we are mainly concerned with mutual information issues, one can use the uncorrelated scattering model where the steering vectors are on Fourier directions. Indeed, as far as mutual information is concerned, only the uncorrelated scatterers will scale the throughput. We will also use as a first approximation the model with equal powers. Let us derive the mutual information when polarization is taken into account. The fact that polarization is used exclusively at the transmitter side, raises some concerns whether the symmetry of mutual information (9.9) still applies to our setup, since exchanging transmitter and receiver might influence the scattering environment if some scatterers scatter only one of the polarization modes. Let us, therefore, look at the mutual information when polarization is taken into account in greater detail.

In the general case, the received signal can be written as

$$\mathbf{y} = \sqrt{\frac{\rho}{n_t}} \mathbf{H}_h \mathbf{x_h} + \sqrt{\frac{\rho}{n_t}} \mathbf{H}_v \mathbf{x_v} + \mathbf{n}. \tag{10.1}$$

Here, $\mathbf{H}_h$[3] and $\mathbf{H}_v$ are $n_r \times \frac{n_t}{2}$ matrices whereas $\mathbf{x_h}$ and $\mathbf{x_v}$ are $\frac{n_t}{2} \times 1$ vectors. $n_r = 8$ and n_t, which represents the equivalent number of antennas, is equal to 8, as well.

---

[3]$\mathbf{H}_h$ stands for a matrix with horizontal polarization whereas $\mathbf{H}_v$ stands for a vertical polarization

- DoA case: In this case, the received signal can be written as

$$\mathbf{y} \;=\; \sqrt{\frac{\rho}{n_t s_r}}\left(\mathbf{\Phi}_{n_r \times s_r}\mathbf{\Theta}_h\mathbf{x_h} + \mathbf{\Phi}_{n_r \times s_r}\mathbf{\Theta}_v\mathbf{x_v}\right) + \mathbf{n} \tag{10.2}$$

$$=\; \sqrt{\frac{\rho}{n_t s_r}}\,\mathbf{\Phi}_{n_r \times s_r}[\mathbf{\Theta}_h\mathbf{\Theta}_v]\left[\begin{array}{c}\mathbf{x_h}\\ \mathbf{x_v}\end{array}\right] + \mathbf{n} \tag{10.3}$$

$$=\; \sqrt{\frac{\rho}{n_t s_r}}\,\mathbf{\Phi}_{n_r \times s_r}\mathbf{\Theta}_{h+v}\mathbf{x}_{h+v} + \mathbf{n}. \tag{10.4}$$

where $\mathbf{\Theta}_{h+v}$ is an $s_r \times n_t$ i.i.d Gaussian matrix and $\mathbf{x}_{h+v}$ is an $n_t \times 1$ vector. Therefore, all the results of section 9.3 can be used without any change. In the DoA based model, polarization at the transmitter does not change anything with respect to a system without polarization and a doubled number of transmitting antennas.

- DoD based model: In this case, the number of scatterers at the transmitter side might be different for the two polarization modes. We denote them by s_t^v and s_t^h, with $s_t^h + s_t^v = s_t$, and write the received signal in the form

$$\mathbf{y} = \sqrt{\frac{2\rho}{n_t s_t}}\left(\mathbf{\Theta}_h\mathbf{\Psi}_{s_t^h \times \frac{n_t}{2}}\mathbf{x_h} + \mathbf{\Theta}_v\mathbf{\Psi}_{s_t^v \times \frac{n_t}{2}}\mathbf{x_v}\right) + \mathbf{n} \tag{10.5}$$

where $\mathbf{\Theta}_h$ and $\mathbf{\Theta}_v$ are $n_r \times s_t^h$ and $n_r \times s_t^v$ i.i.d Gaussian matrices which represent the horizontal and the vertical polarization, respectively. Note that neither s_t^h nor s_t^v can be greater than $\frac{n_t}{2}$ in the Fourier framework.

The mutual information is given by

$$I^M(n_t, n_r, s_t, \rho) \;=\; \log_2\det\left(\mathbf{I}_{n_r} + \frac{2\rho}{n_t s_t}\left(\mathbf{\Theta}_h\mathbf{\Psi}\mathbf{\Psi^H}\mathbf{\Theta}_h{}^H + \mathbf{\Theta}_v\mathbf{\Psi}\mathbf{\Psi^H}\mathbf{\Theta}_v{}^H\right)\right) \tag{10.6}$$

$$=\; \log_2\det\left(\mathbf{I}_{n_r} + \frac{\rho}{s_t}\left(\mathbf{\Theta}_h\mathbf{\Theta}_h{}^H + \mathbf{\Theta}_v\mathbf{\Theta}_v{}^H\right)\right) \tag{10.7}$$

$$=\; \log_2\det\left(\mathbf{I}_{n_r} + \frac{\rho}{s_t}\,\mathbf{\Theta}_{h+v}\mathbf{\Theta}_{h+v}{}^H\right) \tag{10.8}$$

where $\mathbf{\Theta}_{h+v}$ is an $n_r \times s_t$ matrix. We observe that all the results of section 9.3 can be used without change. Note that one of the other drawbacks of the mutual information criteria is that the metric is not able to determine s_t^h and s_v^h but only there sum.

- In the double directional model, the received signal has the form

$$\mathbf{y} = \sqrt{\frac{2\rho}{n_t s_r s_t}}\left(\mathbf{\Phi}_{n_r \times s_r}\mathbf{\Theta}_h\mathbf{\Psi}_{s_t^h \times \frac{n_t}{2}}\mathbf{x_h} + \mathbf{\Phi}_{n_r \times s_r}\mathbf{\Theta}_v\mathbf{\Psi}_{s_t^v \times \frac{n_t}{2}}\mathbf{x_v}\right) + \mathbf{n}. \tag{10.9}$$
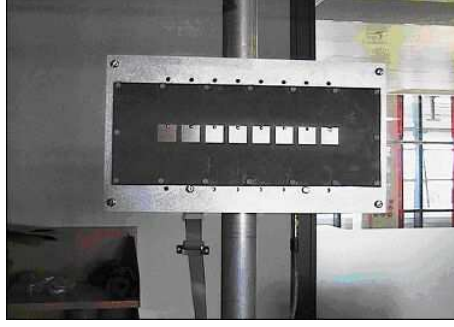
Figure 10.7: Antenna at 5.25 GHz.

The mutual information in the double directional case is given by:

$$
\begin{aligned}
I^M(n_t, n_r, s_r, s_t, \rho) &= \log_2 \det \left(\mathbf{I}_{n_r} + \frac{2\rho}{n_t s_r s_t} \left(\mathbf{\Phi}\mathbf{\Theta}_h \mathbf{\Psi}\mathbf{\Psi}^H \mathbf{\Theta}_h{}^H \mathbf{\Phi}^H + \mathbf{\Phi}\mathbf{\Theta}_v \mathbf{\Psi}\mathbf{\Psi}^H \mathbf{\Theta}_v{}^H \mathbf{\Phi}^H \right) \right) \\
&= \log_2 \det \left(\mathbf{I}_{n_r} + \frac{\rho}{s_r s_t} \left(\mathbf{\Phi}\mathbf{\Theta}_h \mathbf{\Theta}_h{}^H \mathbf{\Phi}^H + \mathbf{\Phi}\mathbf{\Theta}_v \mathbf{\Theta}_v{}^H \mathbf{\Phi}^H \right) \right) &\text{(10.10)} \\
&= \log_2 \det \left(\mathbf{I}_{n_r} + \frac{\rho}{s_r s_t} \mathbf{\Phi} \left(\mathbf{\Theta}_h \mathbf{\Theta}_h{}^H + \mathbf{\Theta}_v \mathbf{\Theta}_v{}^H \right) \mathbf{\Phi}^H \right) &\text{(10.11)} \\
&= \log_2 \det \left(\mathbf{I}_{n_r} + \frac{\rho}{s_r s_t} \mathbf{\Phi} \left(\mathbf{\Theta}_{h+v}\mathbf{\Theta}_{h+v}{}^H \right) \mathbf{\Phi}^H \right) &\text{(10.12)} \\
&= \log_2 \det \left(\mathbf{I}_{s_t} + \frac{\rho}{s_r s_t} \mathbf{\Theta}_{h+v}{}^H \mathbf{\Phi}^H \mathbf{\Phi}\mathbf{\Theta}_{h+v} \right) &\text{(10.13)} \\
&= \log_2 \det \left(\mathbf{I}_{s_t} + \frac{\rho n_r}{s_r s_t} \mathbf{\Theta}_{h+v}{}^H \mathbf{\Theta}_{h+v} \right) &\text{(10.14)}
\end{aligned}
$$

where again $\mathbf{\Theta}_{h+v}$ is an $s_r \times s_t$ i.i.d Gaussian matrix. Therefore, all the results of section 9.3 can be used without change.

In all the following figures, the SNR will be fixed at 10 dB.

10.3 5.25 GHz

10.3.1 Channel Sounder

In this case, broadband antennas, vertically patch arrays were used at the transmitter and the receiver. They were designed and manufactured by the Instituto Superior Técnico, Instituto of Telecommunicacões in Lisbon, Portugal on assignment by Telenor. Both antennas are 8 element Uniform Linear Arrays (ULA). The center frequency is 5.255 GHz. The element spacing is 28.54 mm corresponding to 0.5 λ at 5.255 GHz.

The main channel sounder specification are listed on the following Table (10.2).

Measurement frequency	5.255 GHz
Measurement bandwidth	100 Mhz
Sounding signal	linear frequency chirp
Transmitter antenna	8 element ULA
Element spacing	28.54 mm (0.5 λ)
Receiver antenna	8 element ULA

Table 10.2: Channel sounder specification at 5.255 GHz.

10.3.2 State of Knowledge

According to the specifications in table (10.2) and the measurement set-up, our state of knowledge is the following:

- Uniform Linear Array.

- Far field Approximation.

- Antenna spacing 0.5λ .

- 8×8 antenna system.

As previously, all the framework with mono-polarization developed in section.9.3 can be used if one takes the following values $n_r = 8$ and $n_t = 8$. Note that in all the figures, the SNR is be fixed at 10 dB.

Chapter 11

Validity of the Maximum Entropy Approach

11.1 2.1 GHz Results

11.1.1 Mutual Information Results

Are the Meaured Mutual Information Gaussian?

Before trying to see if the models derived within this paper are mutual information complying, one has to verify that the measured mutual information has a Gaussian behavior. In Figure 11.1, we have plotted respectively the measured mutual information for the scenarios of interest, namely the urban open place, the urban regular low antenna position, the urban regular high antenna, the indoor and the Atrium scenario. We have also plotted the Gaussian pdf of each scenario based on the first and second measured moment i.e if $\mu_{\text{empirical}}$ and $\sigma_{\text{empirical}}$ are respectively the measured mean and variance then for each scenario:

$$P(I^M) = \frac{1}{\sqrt{2\pi\sigma_{\text{empirical}}^2}} e^{-\frac{(I^M - \mu_{\text{empirical}})^2}{2\sigma_{\text{empirical}}^2}}$$

As one can see, the mutual information has a Gaussian behavior and therefore, the models derived in this paper can be considered as candidates for the mutual information compliance criteria [1]. In the following section, we will see how close are the measured capacities from the maximum entropy models.

What About Frequency Selectivity?

In the model derived in [158], we argued that frequency selectivity does not affect the mutual information. In Figure 11.2, we have plotted the mutual information for various frequencies (ranging from 2.05 to 2.15 GHz) in the urban open place scenario, the urban regular low antenna

---

[1]Actually, from all the 150 routes available in [154], only 8 routes (Kvadraturn_02_04_01, Kvadraturen_02_07_18, Kvadraturn_02_08_28, Fornebu_In_01_02_16, Fornebu_In_01_02_15, Fornebu_InOut_01_07_13, Fornebu_InOut_01_07_15, Fornebu_InOut_01_08_03) did not have a Gaussian behavior. We don't know if this is due to measurements errors or something else.
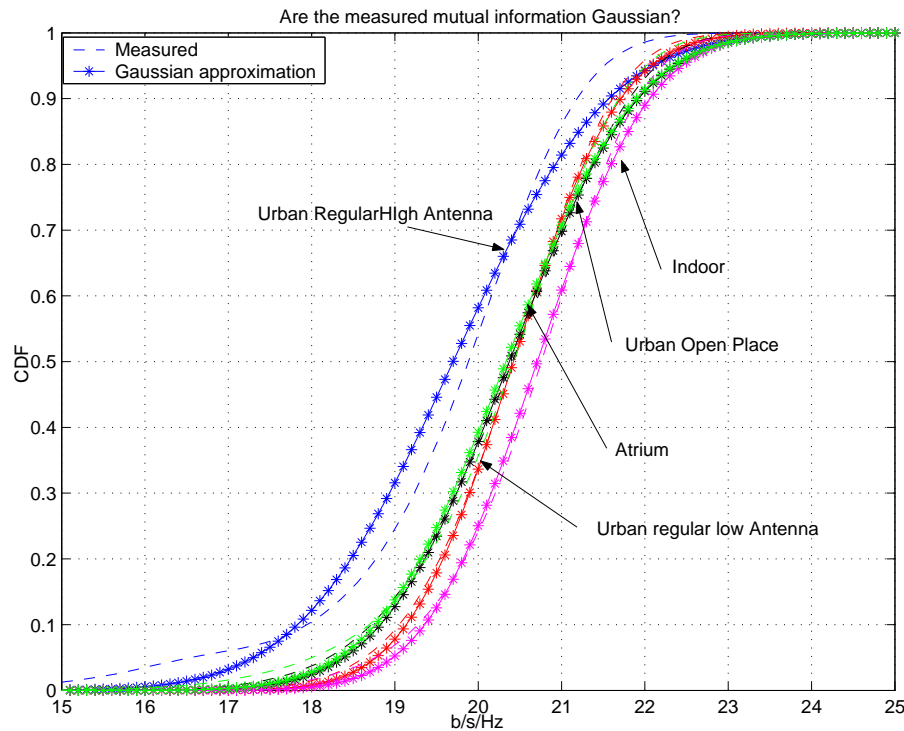
Figure 11.1: Are the measured mutual information Gaussian

position, the urban regular high antenna position, the indoor and Atrium scenario.[2]. As one can observe, for the different frequencies, the mutual information does not really vary which is adequate with our model structure: the highest variation occurs for the urban regular high antenna position and is about 0.3 b/s/Hz which makes a relative variation of ($\frac{19.85-19.55}{19.55} = 0.015$) around 1.5%.

One can observe that in all the cases, frequency selectivity does not affect mutual information. Note that the mutual information is smaller in the urban regular high antenna position than in the low antenna scenario. This maybe due to the fact that there is less scattering objects when the antenna is high. Note also that the mutual information in the indoor scenario is slightly higher than the outdoor case due to a possibly higher number of scattering objects.

Parameter Optimization

In Figure 11.3, we have plotted the measured CDF of the Urban Open place scenario with respect to the optimized DoA, DoD and double directional models. The curves were plotted at an average SNR of 10 dB.

- I.I.D model: The Gaussian i.i.d model is too optimistic and overestimates the achievable rate. An average gap with measurements of 3 b/s/Hz exists.

- DoA model: The DoA based model gives the same performance as the DoD one. The best fitting is obtained for a number of scatterers s_r equal to 6.

---

[2]Note that the CDF has been averaged over different time snapshots but at the same frequency.
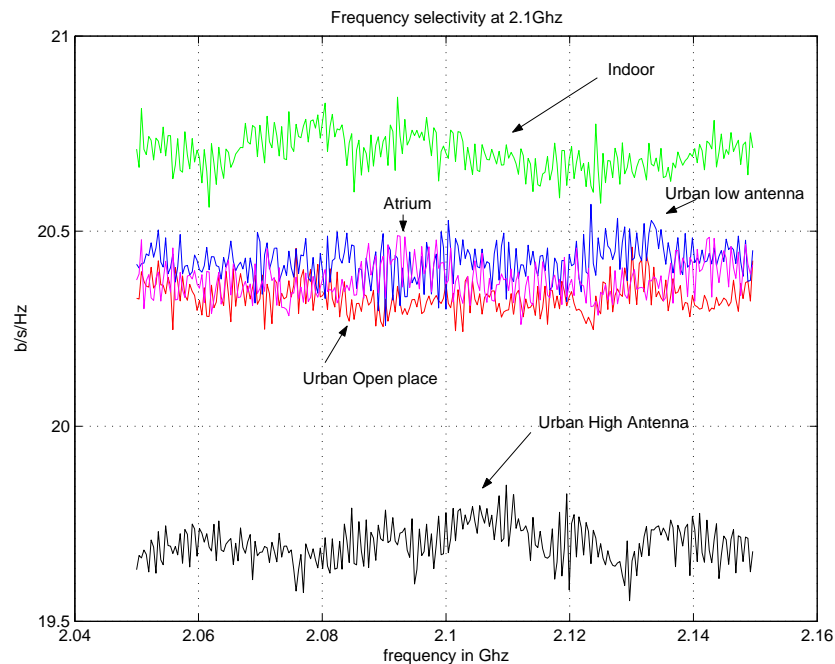
Figure 11.2: Frequency selectivity for many scenarios at 2.1 GHz

- Double directional model: The double directional model fits accurately the data with a number of scatterers equal to $s_r = 7$ and $s_t = 7$. It seems that the equal power case is sufficient to comply with the mutual information measurements. Therefore, the urban open place scenario can be fully described by a double directional model. One can observe that the number of scatterers is quite high. This may be explained by the fact that the open place site has quite a diverse building topography.

In Figure 11.4, we have plotted the measured CDF of the Urban Regular Low Antenna Position scenario with respect to the optimized DoA, DoD and double directional models.

- I.I.D model: The i.i.d Gaussian model does not at all represent this scenario and a gap of more than 4 b/s/Hz is revealed.

- DoA model: The DoA based model fits quite accurately the data with a number of scatterers equal to $s_r = 6$. However, the results are not so tight as for the Urban Open Place. But even when this mismatch, to our knowledge, no model was shown to fit so accurately the data.

- Double directional model: The double directional model gives similar results as the DoA based model with a number of scatterers equal to $s_r = 6$ and $s_t = 8$. Moreover, one can observe that there are more scatterers on the receiving side than the transmitting side ($s_r > s_t$). This is maybe due to the fact that the receiving antenna is low and therefore, many reflections occur at the receiving side. Note that one would get a better fitting curve if the power of the steering vectors are taken into account.
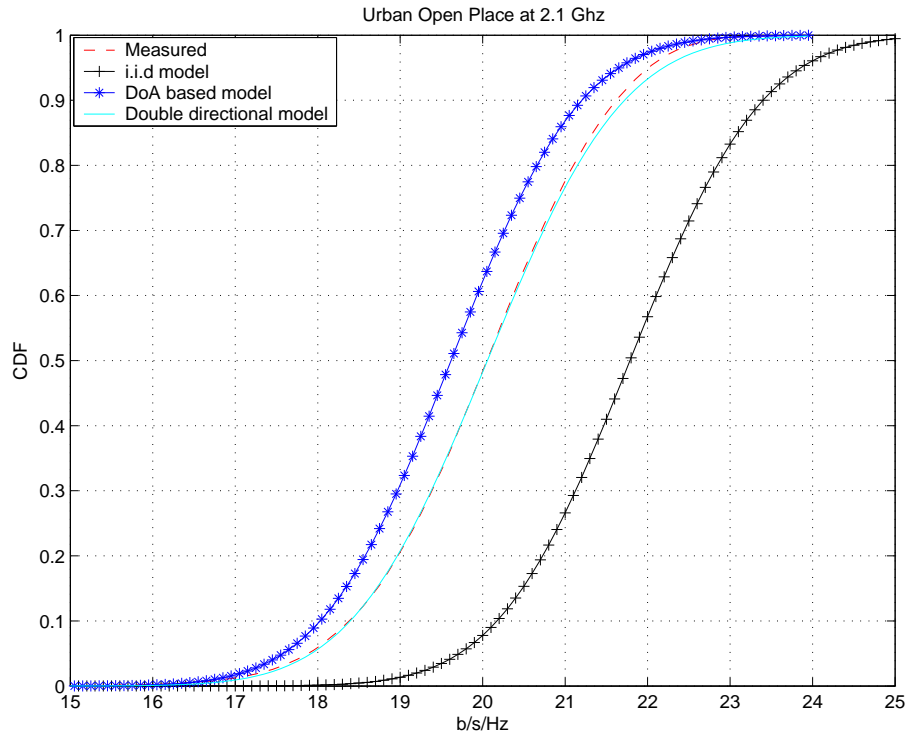
Figure 11.3: Urban Open Place at 2.1 GHz.
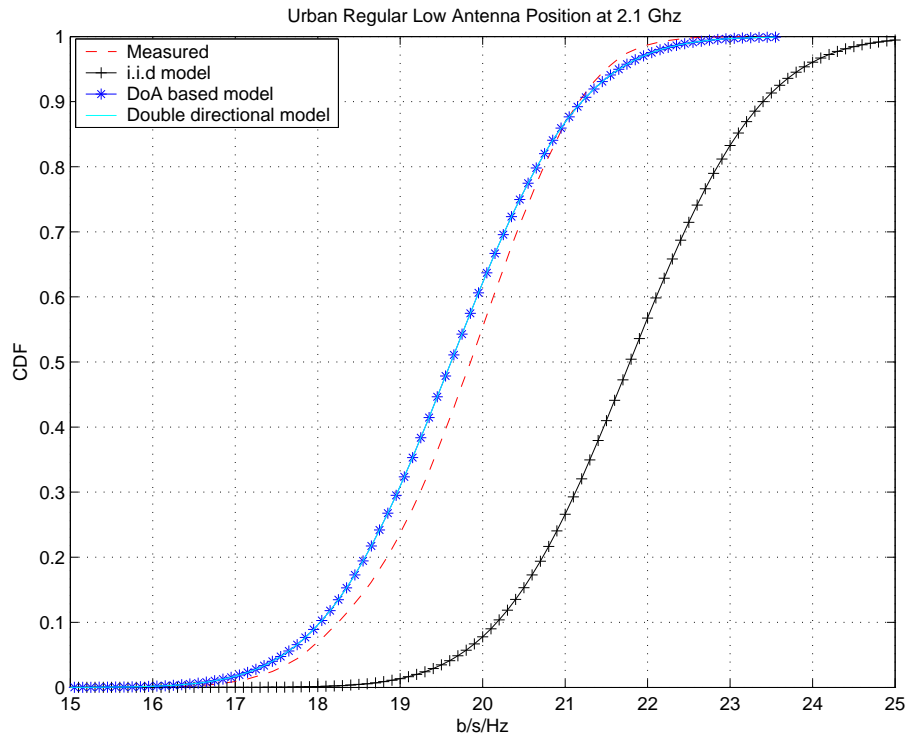
Figure 11.4: Urban Regular, Low Antenna Position at 2.1 GHz.
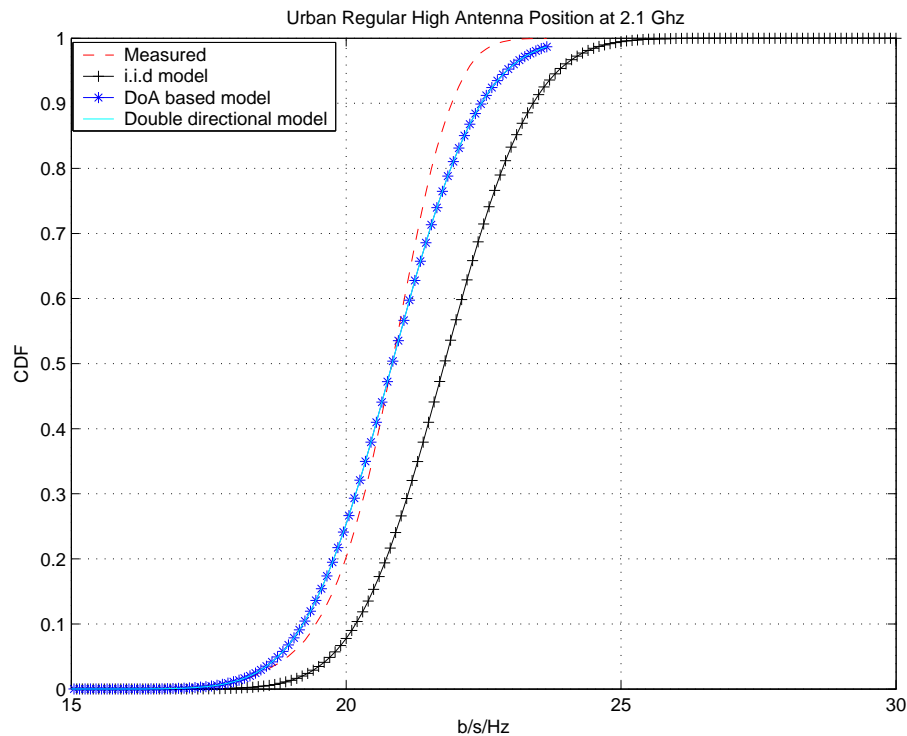
Figure 11.5: Urban Regular, High Antenna Position at 2.1 GHz.

In Figure 11.5, we have plotted the measured CDF of the Urban Regular High Antenna Position scenario with respect to the optimized DoA, DoD and double directional models.

- I.I.D model: In this case, a gap of 4 b/s/Hz appears.

- DoA model: In this case, the optimal number of scatterers is $s_r = 7$. A gap still appears at the lower tail of the curve.

- Double directional model: The double directional model fits with $s_r = 7$ and $s_t = 8$ but the power profile should be taken into account to have a more accurate fitting.

In Figure 11.6, we have plotted the measured CDF of the indoor scenario with respect to the optimized DoA, DoD and double directional models.

- I.I.D model: In this case, a gap of more than 1 b/s/Hz is revealed.

- DoA model: In the DoA case, $s_r = 7$ provides quite accurate results. However, better results could be obtained if the the power of the steering directions is to be taken into account.

- Double directional model: For $s_r = 7$ and $s_t = 8$, the same performance as the DoA model is achieved. As previously, the power of the steering vectors should be taken into account to achieve better results.
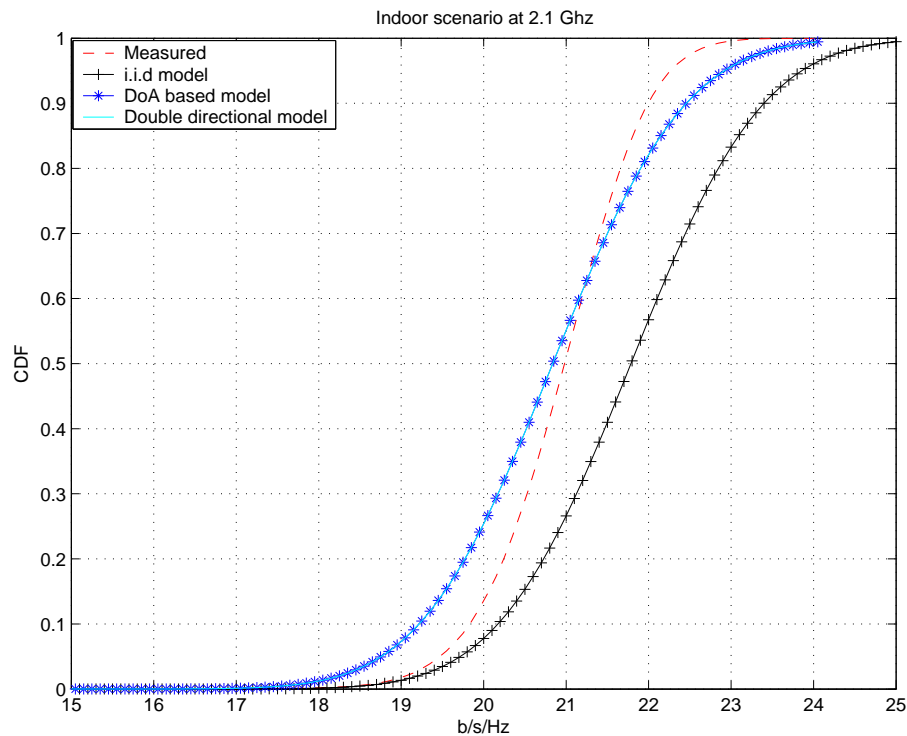
Figure 11.6: Indoor scenario at 2.1 GHz.

In Figure 11.7, we have plotted the measured CDF of the Atrium scenario with respect to the optimized DoA, DoD and double directional models.

- I.I.D model: In this case, a gap of more than 1 b/s/Hz is revealed with the measurements.

- DoA model: The best fit is obtained with $s_r = 7$ and the results are quite accurate. Only a little mismatch appears at the edges of the curve.

- DoD model: The DoD based model gives the same performance as the i.i.d one and a gap of more than 1 b/s/Hz appears. The best fitting is obtained for $s_t = 4$.

- Double directional model: The double directional model gives the same performance as the DoA based. The best fitting is obtained for $s_r = 7$ and $s_t = 4$.

  In summary, the maximum entropy Fourier directional models give accurate results in nearly all the cases. Note that for improving the results, one can take into account the following parameters:

  – The non-zero mean of the different paths

  – The power profile of the steering directions.

  – Retrieve the hypothesis of far field scattering and use the near field models derived.
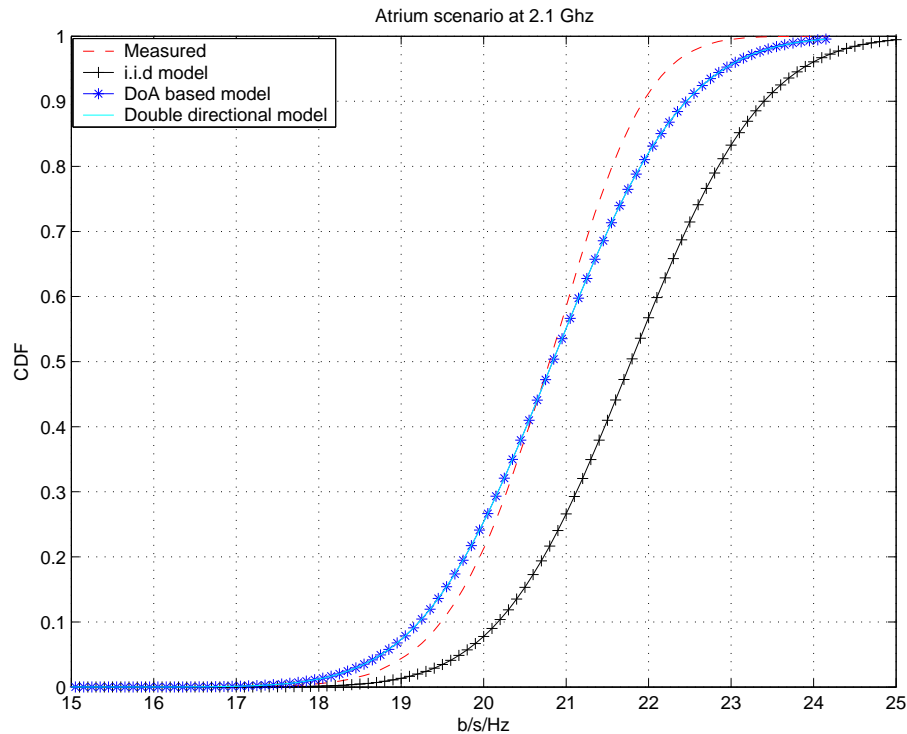
Figure 11.7: "Atrium" scenario at 2.1 GHz.

11.1.2 SINR Results

What About Frequency Selectivity

In the model derived in [158], we argued that frequency selectivity does not affect the mutual information. In Figure 11.8, we have plotted the SINR for various frequencies (ranging from 2.05 to 2.15 GHz) in the urban open place scenario, the urban regular low antenna position, the urban regular high antenna position, the indoor and Atrium scenario.. As one can observe, for the different frequencies, the SINR does not vary muc which is adequate with our model structure: the highest variation occurs for the Urban Open Place with relative variation of $\left(\frac{4.4881-3.9859}{3.9859} = 0.125\right)$ around 12.5%.

Parameter Optimization

In Figure **??**, we have plotted the measured CDF of the Indoor scenario scenario with respect to the optimized DoA, DoD and double directional models.

- I.I.D model:

- DoD model: The DoD model does not fit the data. The best result is obtained for $s_t = 8$.

- DoA model: In this case, the optimal number of scatterers is $s_r = 7$. A gap still appears at the lower tail of the curve.

- Double directional model: The double directional model fits with $s_r = 7$ and $s_t = 8$
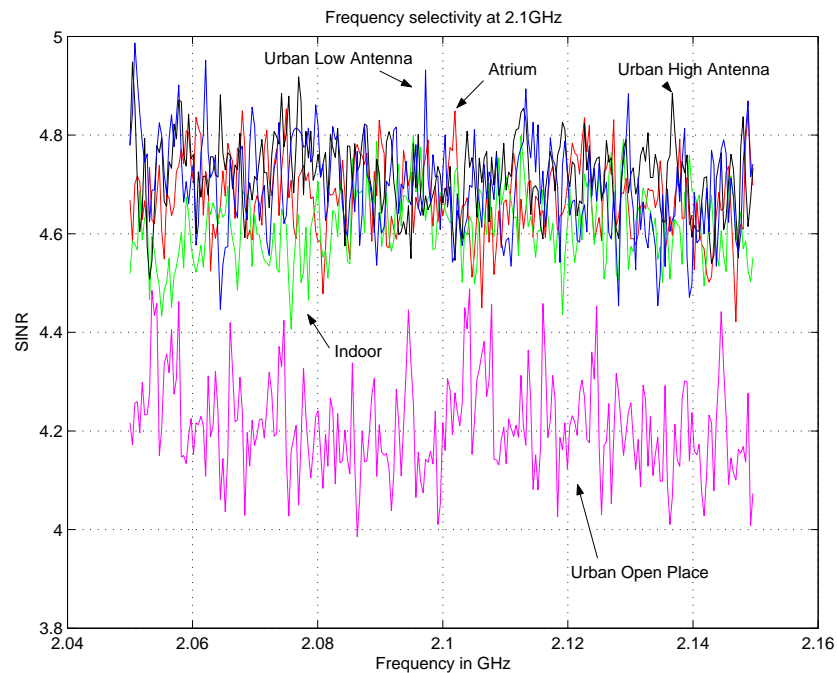
Figure 11.8: Frequency selectivity of the SINR of many scenarios at 2.1 GHz.

In Figure **??**, we have plotted the measured CDF of the Indoor scenario scenario with respect to the optimized DoA, DoD and double directional models.

- I.I.D model:

- DoD model: The DoD model does not fit the data. The best result is obtained for $s_t = 7$.

- DoA model: In this case, the optimal number of scatterers is $s_r = 7$. A gap still appears at the lower tail of the curve.

- Double directional model: The double directional model fits with $s_r = 7$ and $s_t = 8$

In Figure **??**, we have plotted the measured CDF of the Indoor scenario scenario with respect to the optimized DoA, DoD and double directional models.

- I.I.D model:

- DoD model: The DoD model does not fit the data. The best result is obtained for $s_t = 7$.

- DoA model: In this case, the optimal number of scatterers is $s_r = 7$. A gap still appears at the lower tail of the curve.

- Double directional model: The double directional model fits with $s_r = 7$ and $s_t = 8$

In Figure **??**, we have plotted the measured CDF of the Indoor scenario scenario with respect to the optimized DoA, DoD and double directional models.

- I.I.D model:

- DoD model: The DoD model does not fit the data. The best result is obtained for $s_t = 7$.

- DoA model: In this case, the optimal number of scatterers is $s_r = 7$. A gap still appears at the lower tail of the curve.

- Double directional model: The double directional model fits with $s_r = 7$ and $s_t = 8$

In Figure **??**, we have plotted the measured CDF of the Indoor scenario scenario with respect to the optimized DoA, DoD and double directional models.

- I.I.D model:

- DoD model: The DoD model does not fit the data. The best result is obtained for $s_t = 8$.

- DoA model: In this case, the optimal number of scatterers is $s_r = 7$. A gap still appears at the lower tail of the curve.

- Double directional model: The double directional model fits with $s_r = 7$ and $s_t = 8$

11.2   5.25 GHz Results

11.2.1   Mutual Information Results

Are the Measured Mutual Information Gaussian?

Before trying to see if the models derived within this paper are mutual information complying at 5.25 GHz, one has to verify as previously that the measured mutual information have a Gaussian behavior. In Figure 11.9, we have plotted respectively the measured mutual information for four scenarios of interest, namely the urban open place, the urban regular low antenna position, the indoor and the Atrium scenario. We have also plotted the Gaussian pdf of each scenario based on the first and second measured moment

As one can see, the mutual information has a Gaussian behavior (not as accurate as in the 2.1 GHz campaign) and therefore, the models derived in this paper can be considered as candidates for the mutual information compliance criteria [3]. In the following section, we will see how close are the maximum entropy models from the measurements.

What About Frequency Selectivity?

In the models derived in [158], we argued that frequency selectivity does not affect the mutual information. In Figure 11.10, we have plotted the mutual information for various frequencies (ranging from 5.205 to 5.305 GHz) in the urban open place scenario., the urban low antenna scenario, the indoor and the "Atrium" scenario. As one can observe, for the different frequencies,

---

[3] Actually, from all the 79 routes available in [155], 18 routes (Kvadraturn_03_03_01, Kvadraturn_03_04_05, Kvadraturn_03_06_03, Kvadraturn_03_13_07, Kvadraturn_03_03_01, Younstoget_03_02_02, Younstoget_03_08_16, Fornebu_In_02_03_04, Fornebu_In_02_04_06, Fornebu_In_02_05_10, Fornebu_In_02_05_11, Fornebu_In_02_05_12, Fornebu_In_02_07_15, Fornebu_In_02_08_03, Fornebu_Atrium_02_02_19, Fornebu_Atrium_02_04_06, Fornebu_Atrium_02_08_02, Fornebu_Atrium_01_08_03) did not have a Gaussian behavior. As previously, we don't know if this is due to measurements errors or something else.
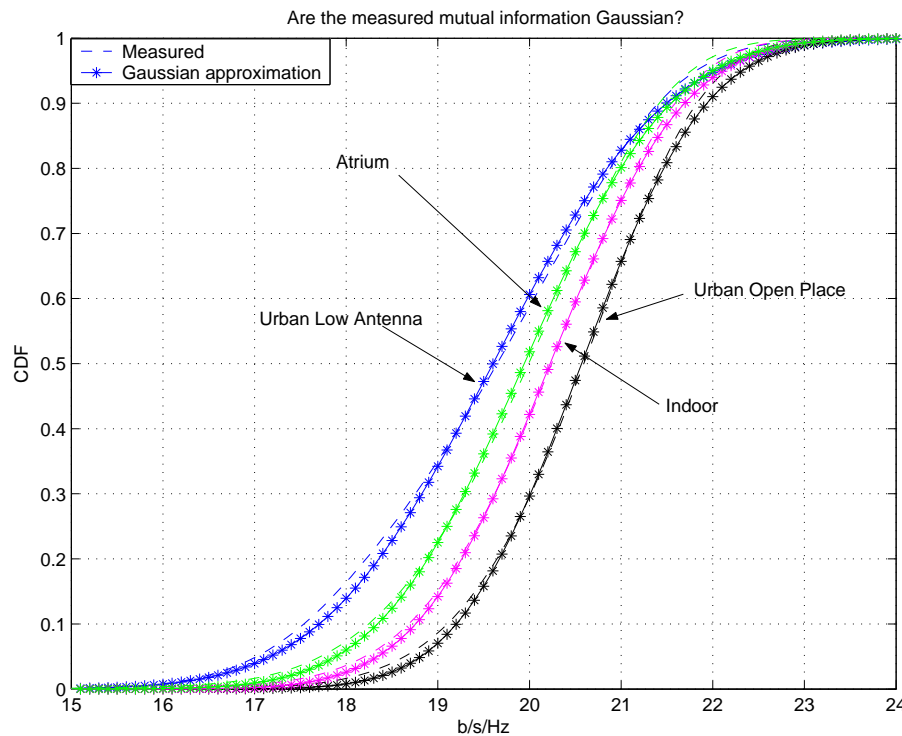
Figure 11.9: Are the measured mutual information Gaussian

the mutual information does not really change which is adequate with our model structure: the highest variation occurs in the indoor scenario and is about 0.16 b/s/Hz which makes a relative variation of ($\frac{20.3-20.14}{20.14} = 0.0079$) around 0.8%.

Parameter Optimization

In Figure 11.11, we have plotted the measured CDF of the urban open place scenario with respect to the optimized DoA, DoD and double directional models. The curves were plotted at an average SNR of 10 dB.

- I.I.D model: The Gaussian i.i.d model is too optimistic and overestimates the achievable rate. An average gap with measurements of 1 b/s/Hz exists at the edge of the curve.

- DoA model: The DoA based model does not fit accurately the data. The best fitting is obtained with $s_r = 8$ and gives the same performance (this is due to the fact that our models are consistent) as the i.i.d Gaussian case.

- Double directional model: The double directional model does not fit accurately the data with a number of scatterers equal to $s_r = 8$ and $s_t = 8$. It seems that the equal power case is not sufficient to comply with the mutual information measurements. One has to take into account the power of the steering directions or the non-zero mean of each path.

In Figure 11.12, we have plotted the measured CDF of the Urban Regular Low Antenna Position scenario with respect to the optimized DoA, DoD and double directional models.
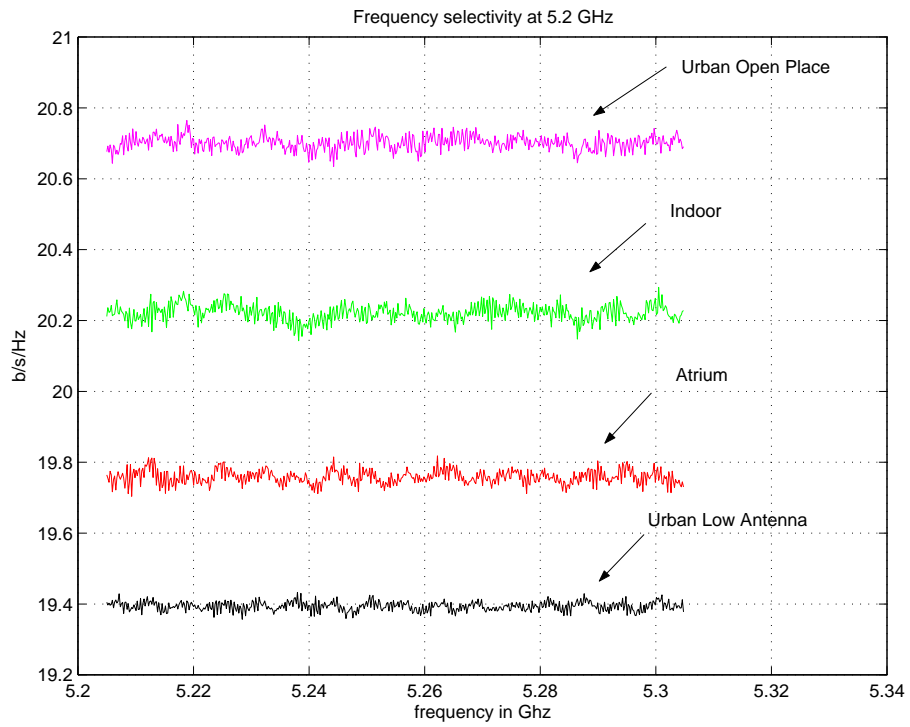
Figure 11.10: Frequency selectivity for various scenarios at 5.2 GHz
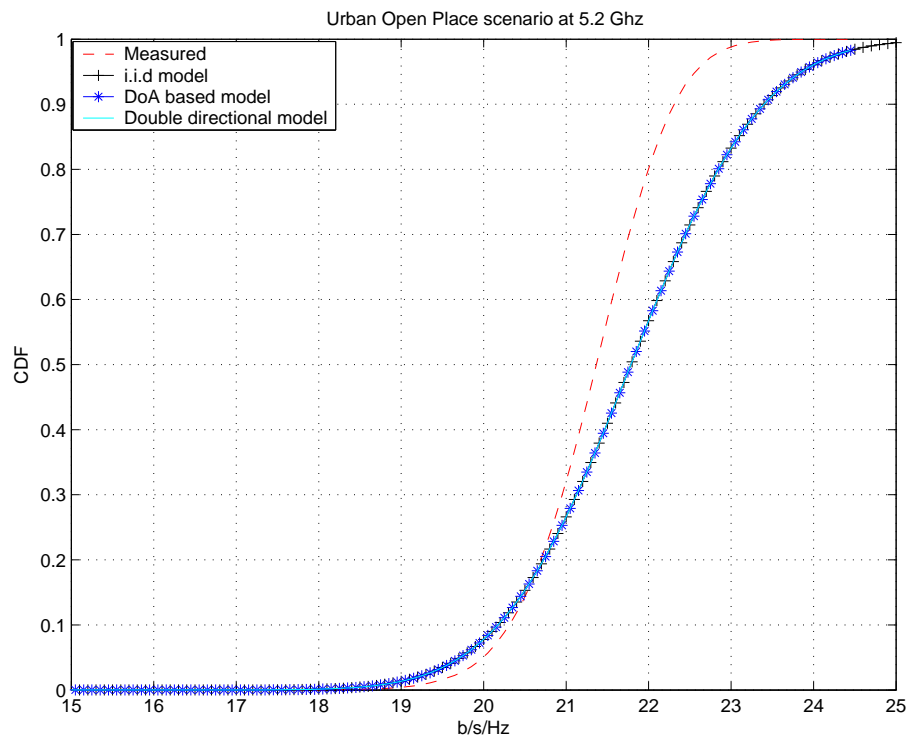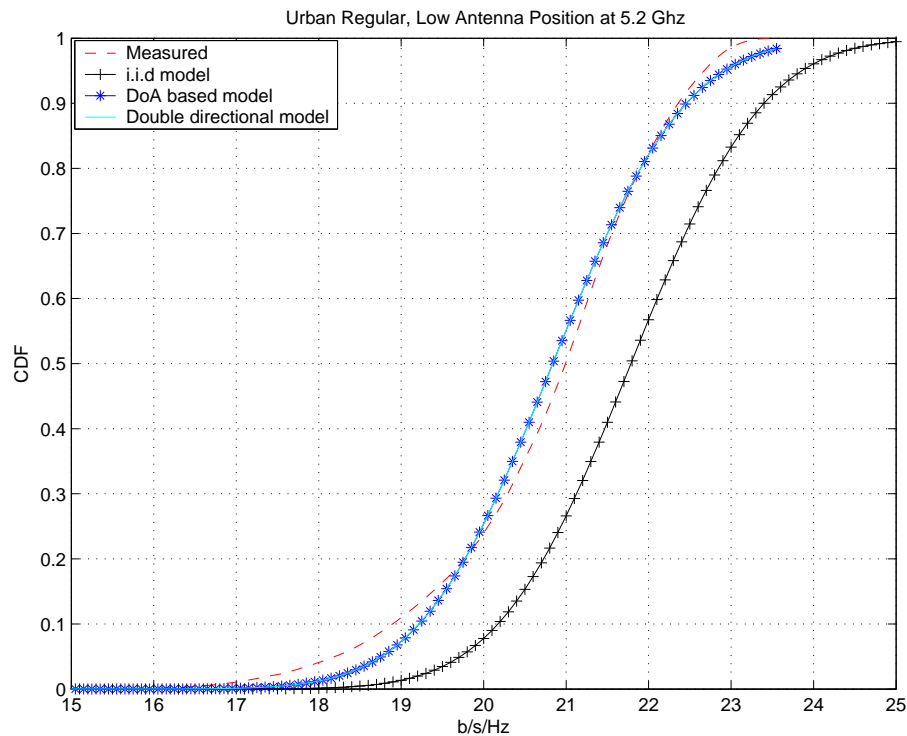

Figure 11.11: Urban Open Place at 5.2GHz.

Figure 11.12: Urban Regular, Low Antenna Position at 5.2 GHz.

- I.I.D model: The i.i.d Gaussian model does not at all represent this scenario and a gap of more than 1 b/s/Hz is revealed.

- DoA model: The DoA based model fits the data with a number of scatterers equal to $s_r = 7$.

- Double directional model: The double directional model gives accurate results with a number of scatterers equal to $s_r = 8$ and $s_t = 7$. One can also notice that the number of scatterers at 5.2 GHz is higher than at 2.1 GHz.

In Figure 11.13, we have plotted the measured CDF of the indoor scenario with respect to the optimized DoA, DoD and double directional models.

- I.I.D model: In this case, a gap of more than 2b/s/Hz is revealed.

- DoA model: The DoA based model fits the data with a number of scatterers equal to $s_r = 6$

- Double directional model: The double directional model gives accurate results with $s_r = 7$ and $s_t = 7$.

In Figure 11.14, we have plotted the measured CDF of the Atrium scenario with respect to the optimized DoA, DoD and double directional models.
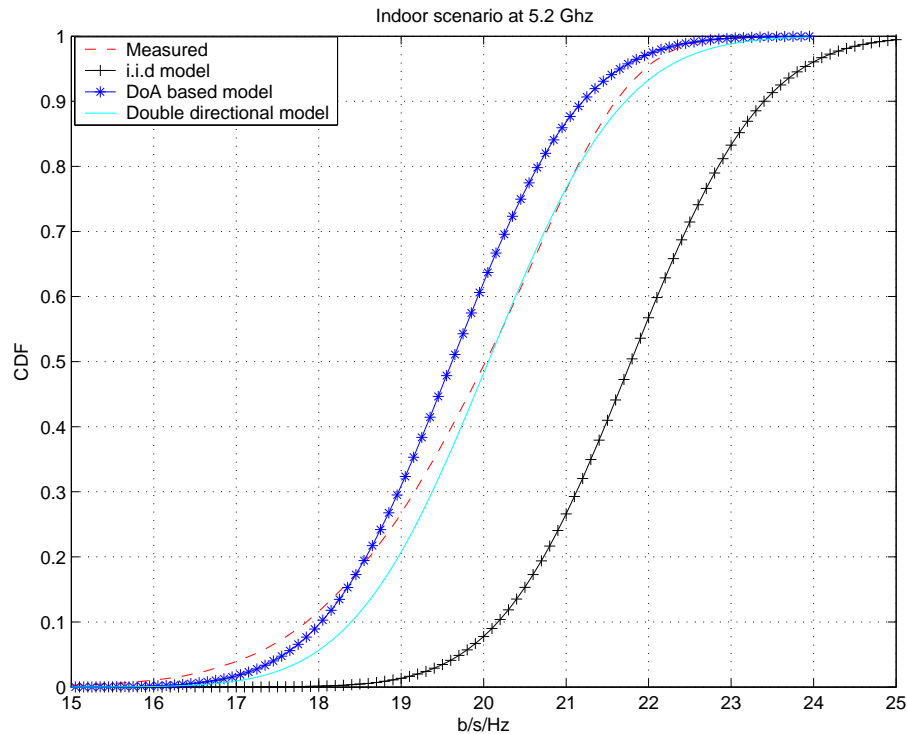
Figure 11.13: Indoor scenario at 5.2 GHz.

- I.I.D model: The i.i.d model does not fit at all the data and a gap of more than 2.5 b/s/Hz appears.

- DoA model: The DoA model fits the data for $s_r = 6$. However, a little mismatch appears at the lower edge of the curve.

- Double directional model: The double directional fits the data for $s_r = 6$ and $s_t = 8$ and gives the same performance as the DoA based model. Better matching could be obtained if one takes into account the power of the steering directions.

11.2.2   SINR Results

What About Frequency Selectivity

In the model derived in [158], we argued that frequency selectivity does not affect the mutual information. In Figure 11.15, we have plotted the SINR for various frequencies (ranging from 5.205 to 5.305 GHz) in the urban open place scenario, the urban regular low antenna position, the urban regular high antenna position, the indoor and Atrium scenario.. As one can observe, for the different frequencies, the SINR does not vary muc which is adequate with our model structure: the highest variation occurs for the Indoor scenario with relative variation of ($\frac{4.6494 - 4.3656}{4.3656} = 0.065$) around 6.5%.
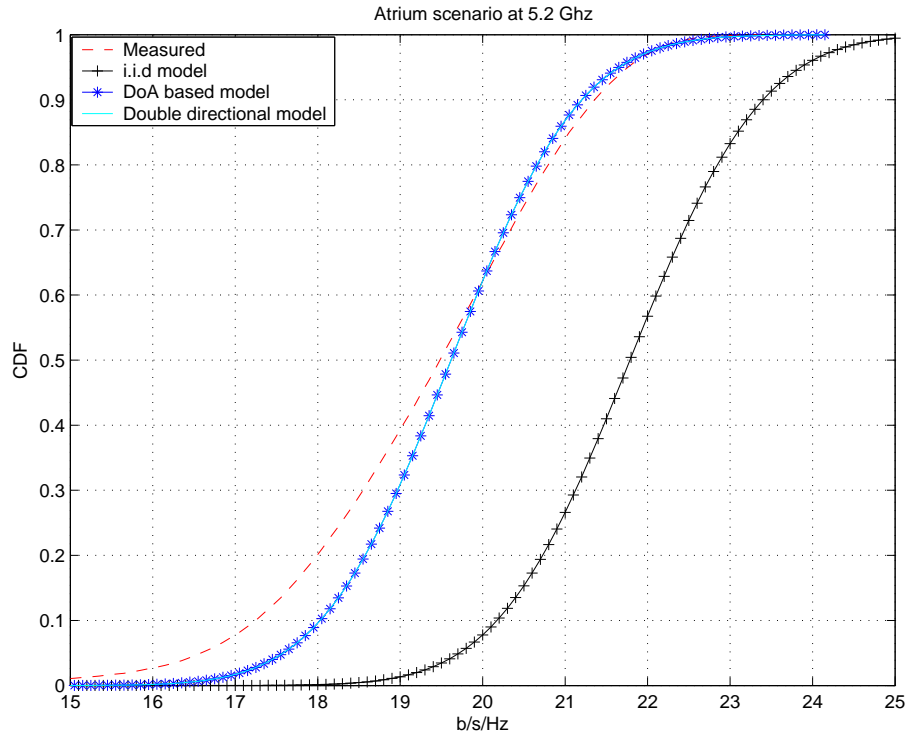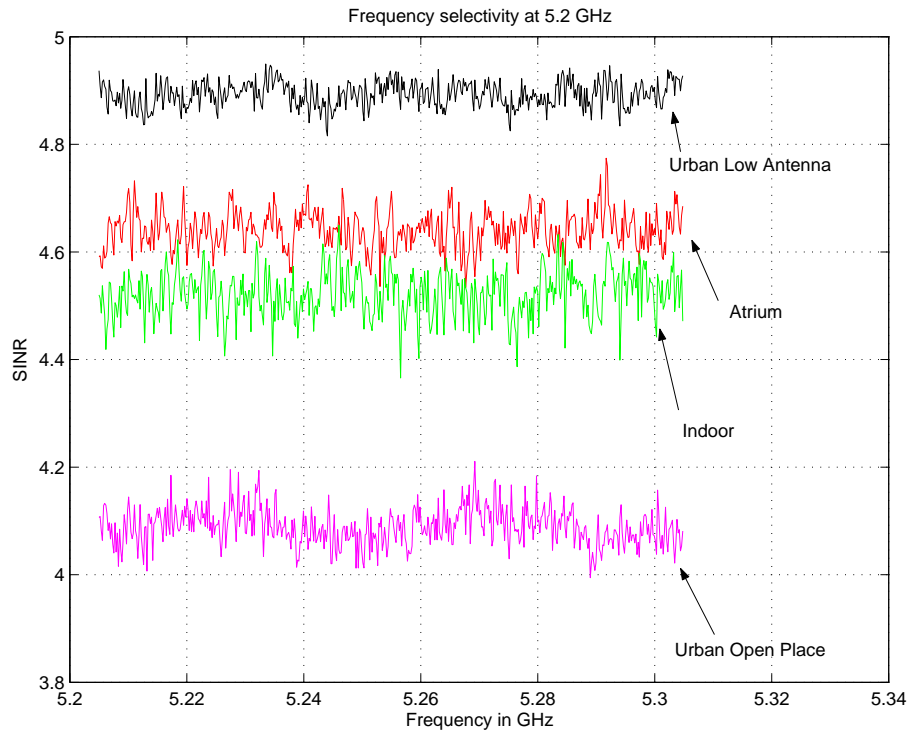
Figure 11.14: "Atrium" scenario at 5.2 GHz.

Figure 11.15: Frequency selectivity of the SINR of many scenarios at 5.2 GHz.

Parameter Optimization

In Figure **??**, we have plotted the measured CDF of the Indoor scenario scenario with respect to the optimized DoA, DoD and double directional models.

- I.I.D model:

- DoD model: The DoD model does not fit the data. The best result is obtained for $s_t = 8$.

- DoA model: In this case, the optimal number of scatterers is $s_r = 4$. A gap still appears at the lower tail of the curve.

- Double directional model: The double directional model fits with $s_r = 4$ and $s_t = 8$

In Figure **??**, we have plotted the measured CDF of the Indoor scenario scenario with respect to the optimized DoA, DoD and double directional models.

- I.I.D model:

- DoD model: The DoD model does not fit the data. The best result is obtained for $s_t = 7$.

- DoA model: In this case, the optimal number of scatterers is $s_r = 7$. A gap still appears at the lower tail of the curve.

- Double directional model: The double directional model fits with $s_r = 7$ and $s_t = 8$

In Figure **??**, we have plotted the measured CDF of the Indoor scenario scenario with respect to the optimized DoA, DoD and double directional models.

- I.I.D model:

- DoD model: The DoD model does not fit the data. The best result is obtained for $s_t = 7$.

- DoA model: In this case, the optimal number of scatterers is $s_r = 7$. A gap still appears at the lower tail of the curve.

- Double directional model: The double directional model fits with $s_r = 7$ and $s_t = 8$

In Figure **??**, we have plotted the measured CDF of the Indoor scenario scenario with respect to the optimized DoA, DoD and double directional models.
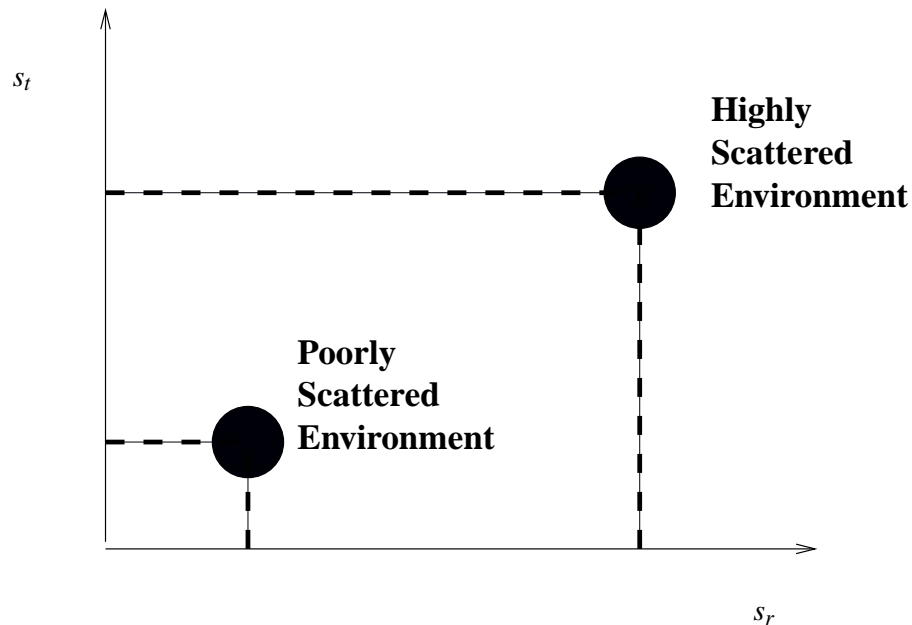
- I.I.D model:

- DoD model: The DoD model does not fit the data. The best result is obtained for $s_t = 7$.

- DoA model: In this case, the optimal number of scatterers is $s_r = 7$. A gap still appears at the lower tail of the curve.

- Double directional model: The double directional model fits with $s_r = 7$ and $s_t = 8$

In Figure **??**, we have plotted the measured CDF of the Indoor scenario scenario with respect to the optimized DoA, DoD and double directional models.

Figure 11.16: Characterization of the environment with s_r and s_t.

- I.I.D model:

- DoD model: The DoD model does not fit the data. The best result is obtained for $s_t = 7$.

- DoA model: In this case, the optimal number of scatterers is $s_r = 7$. A gap still appears at the lower tail of the curve.

- Double directional model: The double directional model fits with $s_r = 7$ and $s_t = 8$

11.3 Environment Classification

An important question raised by the model proposed within this contribution is whether the modelling environment is independent of the antennas or not: indeed, one would like to represent each environment independently of the antennas and only through the parameters (s_r, s_t): Depending on their value, these numbers would characterize entirely the environment as shown in Figure 11.16[4]. A quick look at the theoretical mutual information equations of section. 9.3 show that the number of scatterers depend on the number of transmitting and receiving antennas. Therefore, a question arises naturally: What are we modelling exactly? The answer depends on the type of simplification one does:

- With the general maximum entropy framework developed in contribution [158], the number of scatterers are defined as the number of distinct reflecting waves and are independent of the number of antennas. **We have decided** to take into account every object however

---

[4]Note that the notion of poorly or highly scattering environment depends on the number of antennas. Indeed, for a 2×2 MIMO system, 2 scatterers will be considered as a highly scattered environment whereas for a 8×8 system, the environment will be poorly scattered.

small the object may be (and this is also a justification of the asymptotic analysis). Within an object, there may be millions of waves reflected and therefore our model will consider millions of scatterers (in this case, the asymptotic analysis is even more appealing). In fact, the notion of scatterer is meaningless (how does one define a scatterer? The result will obviously depend on the resolution) and only the limiting power profile matters (which represents the distinct objects with their angular spread).

- With the maximum entropy framework on Fourier directions (which is a particular case of the virtual representation of Sayeed [139] where the inner matrix is i.i.d Gaussian), one assumes that the scatterers are on Fourier directions. The higher the number of antennas one has, the higher the resolution will be. Therefore, for the same environment the number of "virtual scatterers" depends crucially on the number of antennas. There is no relationship between the virtual scatterers and the real ones unless the number of antennas is high enough to capture with a good resolution all the environment. We conjecture that the number of virtual scatterers will depend on the number of antennas until a certain resolution where increasing the number of antennas will not yield an increase of scatterers. To confirm the variation of the number of scatterers with the number of antennas in the same environment, we have plotted in Figure 11.18 the optimal number of scatterers for the Urban Open Place scenario with 5×5, 6×6, 7×7 and 8×8 antennas (different antennas have been used for the same scenario) at 2.1 GHz. As one can observe, for the same environment, the number of scatterers with the Fourier representation changes. To confirm that the Fourier model is good for a high number of antennas[5] (for which all the scatterers can be captured), we have plotted in Figure (11.17) the CDF of the mutual information for 5×5, 6×6, 7×7 and 8×8 system. As one can see, as the number of antennas increases, the double directional model fits better and better the measurements. There are two explanations to this observation:

  - as the number of antennas increases, the resolution increases and one is able to capture all the scatterers. The Fourier representation is then similar to the general maximum entropy representation where the steering directions can be anywhere.
  - As the number of antennas increases, the Gaussian approximation becomes realistic.

Hence, although quite simple, the maximum entropy model on Fourier direction is a good model but for a high number of antennas (8 antennas seems enough). Otherwise, one should use the general maximum entropy model (where the directions are not Fourier ones).

11.4 Conclusion

The maximum entropy based model has been proved to be mutual information complying and is a good candidate to model the MIMO link based on other criteria such as BER. Although case depend, some general trends on the difference between the 5.2 GHz and 2.1 GHz measurements can be provided:

- In general, for each scenario, the number of scatterers is higher at 5.2 GHz than at 2.1 GHz.

---

[5]We insist on the fact that the maximum entropy Fourier model is good for high number of antennas. However, the general maximum entropy model is good for any number of antennas.
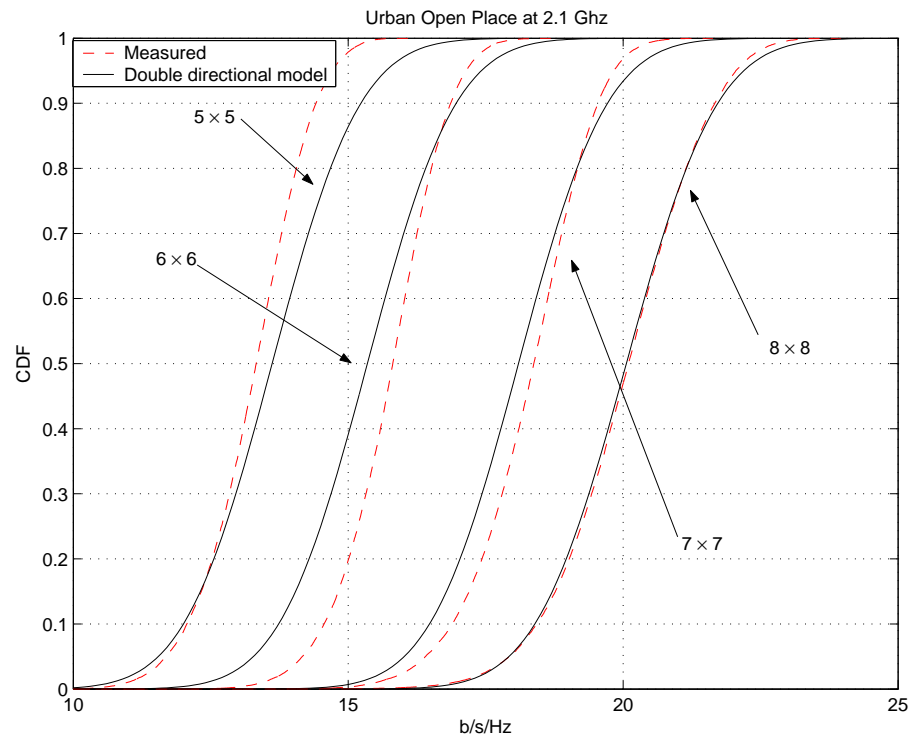
Figure 11.17: Various number of antennas for the Urban Open Place at 2.1 GHz.

- The mutual information seems to be higher at 5.2 GHz than at 2.1 GHz

- At 5.2 GHz and 2.1 GHz, frequency selectivity does not affect the mutual information.

- Indoor scenarios provide higher throughput than outdoor scenarios.

- When the receiving base station is low, the mutual information is higher than when the base station is high.

- The i.i.d model always overestimates the mutual information.

- At 2.1 GHz and 5.2 GHz, the measured mutual information has a Gaussian behavior. However, the curves fit better the Gaussian distribution at 2.1 GHz than at 5.2 GHz.

- The maximum entropy model with zero mean and equal power on the steering directions fits quite accurately the data at 2.1 GHz.

- At 5.2 GHz, the maximum entropy model with zero mean and equal power on the steering directions is not so accurate. The power of the steering directions should be taken into account. But even with this mismatch, to our knowledge, no model was shown to fit so accurately the data with only two free parameters.

As a conclusion, if one is interested in mutual information compliance, then the maximum entropy model with Fourier directions gives quite good results. The models are still being tested for other criteria.
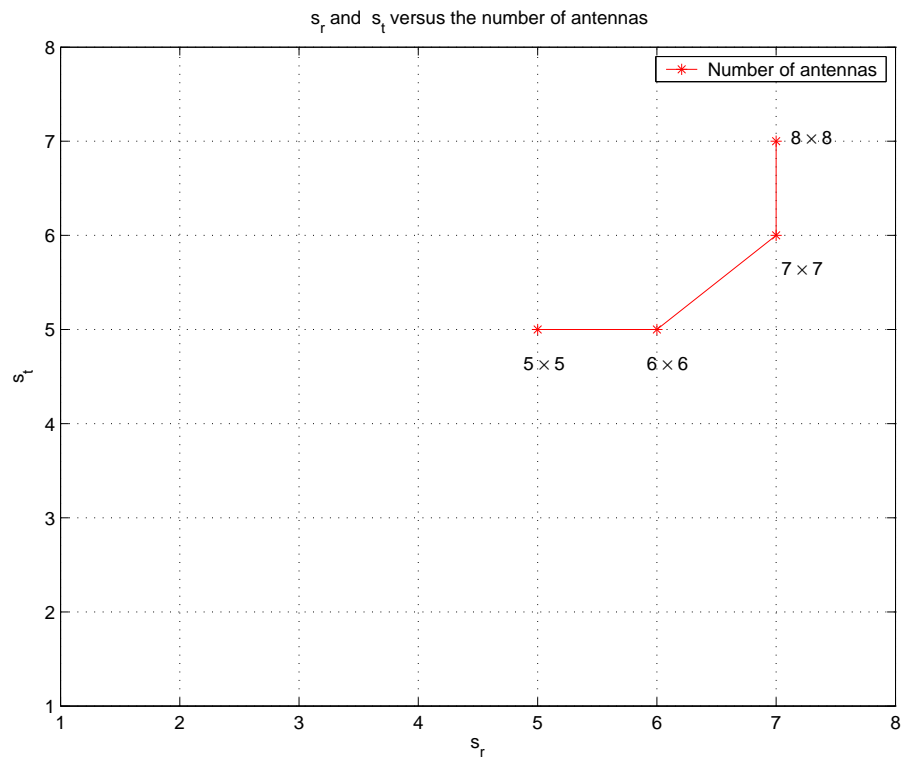
Figure 11.18: s_r and s_t versus the number of antennas.

Chapter 12

Antenna Design

********* In october ************

12.1 Antenna Allocation

A common question that arises in MIMO systems concernes the allocation of the antennas between transmitter and receiver, in other words suppose that one has a total number $n = n_t + n_r$ of antennas. What is the optimal proportion $\gamma = \frac{n_r}{n_t}$ that optimize the average mutual information for example [85]. Such considerations depend of course on the sate of knowledge at hand and a as a consequence on the type of model derived.

12.1.1 i.i.d Gaussian Model

Consider for example the case of the i.i.d Gaussian model. The mean mutual information, normalized to the total number of antennas is given by:

$$\mathbb{E}(\frac{I^M}{n})(\gamma, \rho) = \frac{1}{1 + \frac{1}{\gamma}}\ln(1 + \rho - \rho\alpha) + \frac{1}{1 + \gamma}\ln(1 + \rho\gamma - \rho\alpha) - \frac{1}{1 + \gamma}\alpha$$

with

$$\alpha = \frac{1}{2}[1 + \gamma + \frac{1}{\rho} - \sqrt{(1 + \gamma + \frac{1}{\rho})^2 - 4\gamma}]$$

$\mathbb{E}(\frac{I^M}{n})$ is a function of γ and ρ. In general, the previous equation as no analytical expression of for its maximum value. For each ρ, one has to search for the optimum value of this function. Note that in the case where $\rho \to \infty$, the optimum value is $\gamma = 1$. This mainly due to the fact that the mutual information scales as $\min(n_r, nr)\ln(\rho)$. For low values of the SNR, it can be easy easily shown that the

$$I^M = \frac{\gamma\rho}{1 + \gamma} + O(\rho)$$

For low values of the SNR, the optimum value is $\gamma \to \infty$. The shift toward the receiving antennas can be understood intuitively: in this regime, the mutual information scales linearly
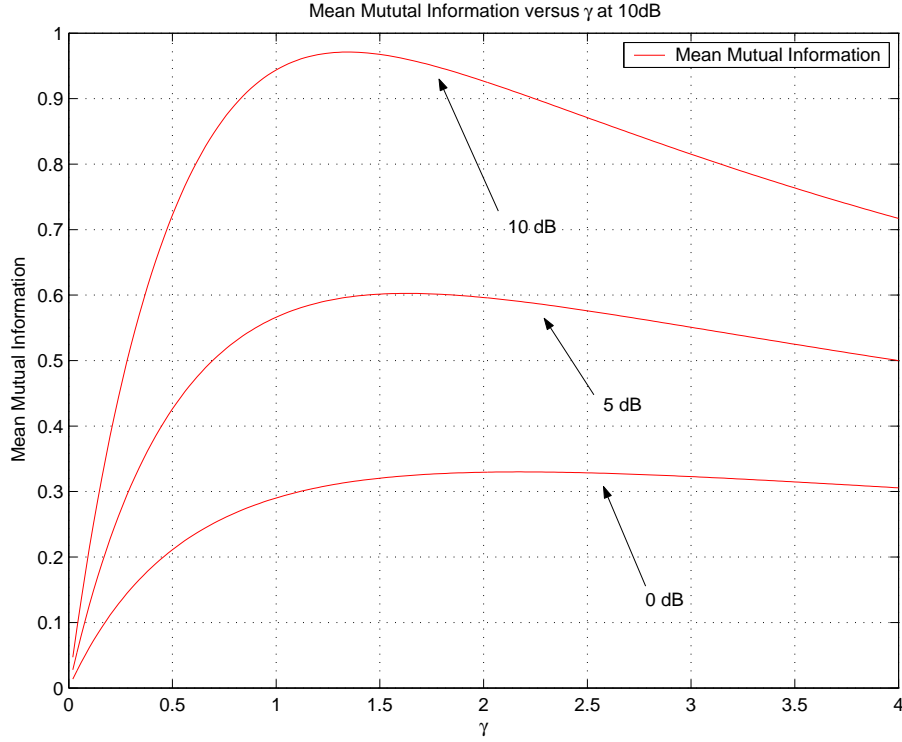
Figure 12.1: Mean Mutual Information versus γ for i.i.d Gaussian model.

with the SNR (and not with the degrees of freedom) and therefore one should focus on increasing the total received SNR (which can be done by increasing the number of receiving antennas).

For the general case, we have plotted in Figure 12.1 the mean mutual information versus γ for several values of the SNR ρ. In each case, a maximum value occurs which depends on ρ (The maximum value is between 1 and ∞) as shown in Figure 12.2.

12.1.2 Double directional model

Similarly to section 12.1.1, let us now consider the double directional model. In this case,

$$
\begin{aligned}
\mathbb{E}(\frac{I^M}{n})(\gamma,\rho) &= \frac{s_r}{n}\ln(1+\rho\frac{1}{1+\frac{1}{\gamma}}\frac{n}{s_r}-\rho\frac{1}{1+\frac{1}{\gamma}}\frac{n}{s_r}\alpha) \\
&+ \frac{s_t}{n}\ln(1+\rho\frac{n}{s_t}\frac{1}{1+\frac{1}{\gamma}}-\rho\frac{n}{s_r}\frac{1}{1+\frac{1}{\gamma}}\alpha) \\
&- \frac{s_t}{n}\alpha
\end{aligned}
$$

and

$$
\alpha = \frac{1}{2}\left[1+\frac{s_r}{s_t}+\frac{s_r}{\rho n}(1+\frac{1}{\gamma})-\sqrt{(1+\frac{s_r}{s_t}+\frac{s_r}{\rho n}(1+\frac{1}{\gamma}))^2-4\frac{s_r}{s_t}}\right]
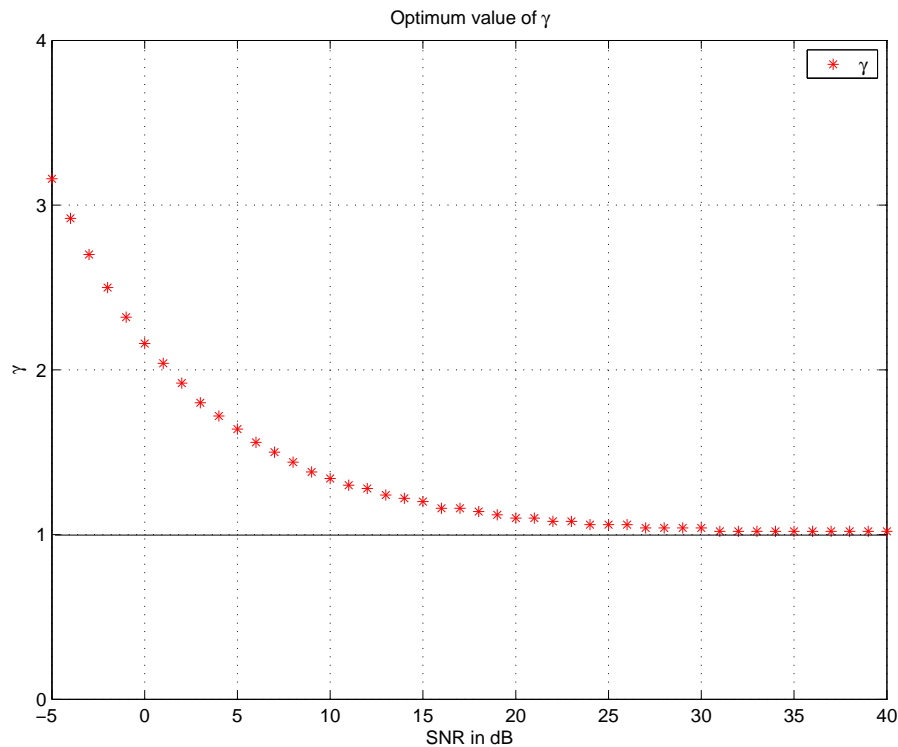$$

Figure 12.2: Optimum γ versus SNR for i.i.d Gaussian model.

It can also be shown that at very low SNR,

$$I^M = \frac{\gamma\rho}{1+\gamma} + O(\rho)$$

which favors a shift towards the number of receiving antenna in this regime. In the general case, there is no explicit expression for the optimum value of γ.

In Figure 12.3, we have plotted the optimum γ versus s_r and s_t at 10 dB. In fact, γ is only a function of s_r. This is due to the fact that $s_t \leq n_t$ and therefore n_t as no effect on the multiplexing gain (s_t being te limiting factor). As s_r decreases, γ increases to increase the received SNR (since the multiplexing gain is any case limited by $\min(s_r, s_t)$. Note that the values of γ are between 1 and ∞.

12.2 Antenna Geometry

Theorem **??** of section **??** (**a revoir pour le theoreme**) is extremely useful as it shows that only the limiting eigenvalue distribution of the steering directions and their respective powers matters: in other words, two antenna configuration can yield the same throughput as long as they give rise to the same eigenvalue distribution for the steering matrix. Based on this result, a future mobile scenario the author would like to advocate is the following: imagine a set of reconfigurable antennas that can move on a grid. The antennas are at the beginning displayed with a Uniform Linear Array geometry. Once the transmission starts, the angles of arrival and
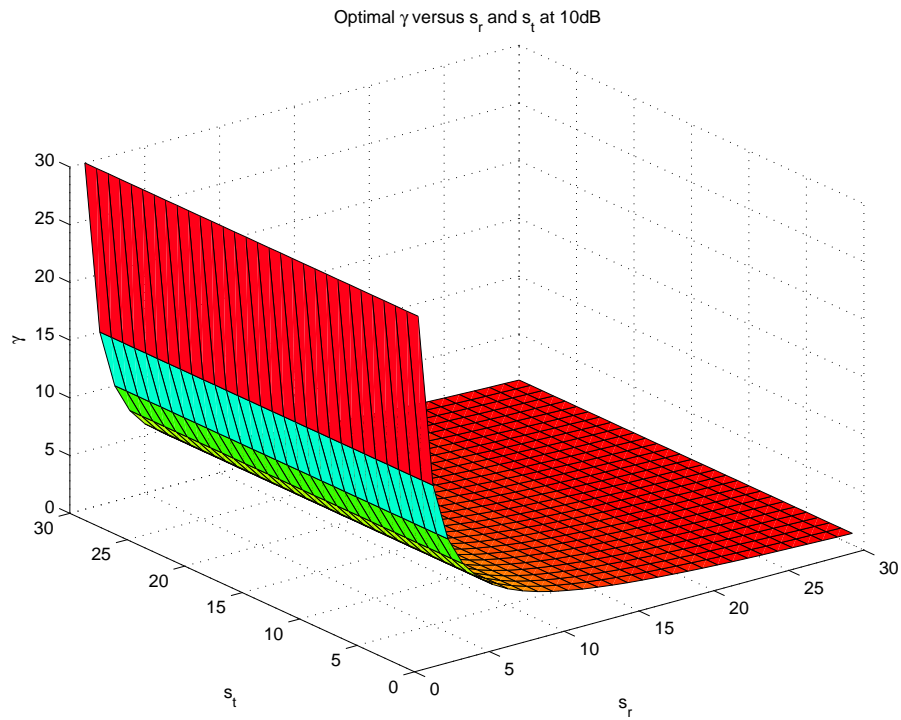
Figure 12.3: Optimum γ versus s_r and s_t for the double directional mdeol model.

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.

Chapter 13

Conclusion

Where do we stand on channel modelling ?[1] 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[2]. The crucial point is not creating a model but asking the right question based on a given state of knowledge (raw measurement data, prior information, are we in a urban area? is it a fixed network?..). A generic method for creating models based on the principle of maximum entropy has been provided and proved to be theoretically sound. At every step, we create a model incorporating only our prior information and not more! The model achieved is broad as it complies as best it can with any case having more constraints (but at least includes the same prior constraints). The channel modelling method is summarized hereafter:

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

- Argument of consistency

The consistency argument is extremely important as it shows that two channel modelling methods based on the same state of knowledge should lead to the same channel model. This fact has not always been fulfilled in the past. Our models are logical consequence of the use of the principle of maximum entropy and need not to be assumed without deeper justification. The models proposed may seem inadequate to reality for some readers: we argue as in [21] that the purpose of channel modelling is not to describe reality but only our information about reality. The model we achieve are consistent and any other representation is obviously unsound if based on the same state of knowledge. However, one must bear in mind that the less things are assumed as a priori information the greater are the chances that the model complies with any mismatched representation.

But what if the model fails to comply with measurements? The model is not to blame as it is a logic consequence of information theoretic tools [21]. With the methodology introduced, failure is greatly appreciated as it is a source of information and the maximum entropy approach

---

[1] This question has to be taken in light of a talk "Where do we stand on maximum entropy?" made by E.T. Jaynes in 1978 at MIT [159].

[2] This point of view is not new and the misconception persists in many other fields. Descartes, already in 1637, warned us when stating in the first lines of the French essay "Le discours de la méthode":" la diversité de nos opinions ne vient pas de ce que les uns sont plus raisonnables que les autres, mais seulement de ce que nous conduisons nos pensées par diverses voies, et ne considérons pas les mêmes choses".

is avid of information: the result of non-compliance is automatically taken into account as some new information evidence to be incorporated in the question. It only means that the question asked was not correct (double directional rather than directional for example) and should be adjusted accordingly in order to imply a new model (based on some new source of information); and as it is well known, finding the right question is almost finding the right answer.

Chapter 14

Appendix

14.0.1 Preliminaries

Lemma 1 *Consider the $t \times t$ matrix (see Bai & Silverstein [27]):*

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

- $\mathbf{H}_{s \times t} = (h_{ij})$ *is an $s \times t$ matrix with i.i.d complex standardized entries having finite fourth moments, $\mathbb{E}(h_{ij}{}^2) = 0$ and $\mathbb{E}(\mid h_{ij} \mid^4) = 2$ with $\lim_{t \to \infty} \frac{s}{t} = c$.*

- $\mathbf{A}_{s \times s}$ *is an $s \times s$ non-random Hermitian non-negative definite matrix, with empirical eigenvalue distribution that converges in distribution almost surely to a fixed G, and the sequence of spectral norms $\|\mathbf{A}_{s \times s}\|$ is bounded.*

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

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

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

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

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

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

  through its 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.

Theorem 5 *(see Girko [128]) Let the $N \times K$ random matrix $\mathbf{H}$ be composed of independent entries $(\mathbf{H})_{ij}$ with zero mean and variances $\frac{w_{ij}}{N}$ such as all w_{ij} are uniformly bounded from above. Assume that the empirical joint distribution of variances w: $[0,1] \times [0,\beta] \to \mathbb{R}$ defined by $w(x,y) = w_{ij}$ for i,j satisfying:*

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

and

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

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

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

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

$$u(x,s) = \left[-s + \int_0^\beta \frac{w(x,y)dy}{1 + \int_0^1 u(x',s)w(x',y)dx'} \right]^{-1} \tag{14.1}$$

The solution to equation (14.1) exists and is unique in the class of functions $u(x,s) \ge 0$, analytic for $\mathrm{Im}(s) > 0$ and continuous on $x \in [0,1]$.

14.0.2 Proof of Theorem 1

The proof can also be found in [62]. In the case of the i.i.d Gaussian channel, we have:

$$G_{\mathrm{iid}}(\tau) = \delta(\tau - 1).$$

Therefore, the Stieltjes transform $m_{f_{\mathrm{iid}}}(z)$ is solution of:

$$z = \frac{-1}{m_{f_{\mathrm{iid}}}(z)} + \frac{\gamma}{1 + m_{f_{\mathrm{iid}}}(z)}$$

which yields:

$$m_{f_{\mathrm{iid}}}(z) = \sqrt{(\frac{1}{2} + \frac{1-\gamma}{2z})^2 - \frac{1}{z}} - \frac{1}{2} - \frac{1-\gamma}{2z}$$

and the limiting distribution has the following density:

$$f_{\mathrm{iid}}(x) = (1 - \gamma)\delta(x) + \frac{1}{2\pi x}\sqrt{((1+\sqrt{\gamma})^2 - x)(x - (1-\sqrt{\gamma})^2)}$$

defined in the interval $[(1 - \sqrt{\gamma})^2, (1 + \sqrt{\gamma})^2]$.

Let us now derive μ_{iid}. The asymptotic mean value is therefore equal to:

$$\mu_{\text{iid}}(\gamma,\rho) = \int_{(1-\sqrt{\gamma})^2}^{(1+\sqrt{\gamma})^2} \ln(1+\rho\lambda)f(\lambda)d\lambda$$
$$= \gamma\ln(1+\rho-\rho\alpha(\gamma,\rho)) + \ln(1+\rho\gamma-\rho\alpha(\gamma,\rho)) - \alpha(\gamma,\rho)$$

with

$$\alpha(\gamma,\rho) = \frac{1}{2}[1+\gamma+\frac{1}{\rho} - \sqrt{(1+\gamma+\frac{1}{\rho})^2 - 4\gamma}]$$

For the explicit form of the integral, we have used the result of Proposition 8.

Let us now derive the asymptotic variance σ_{iid}^2 (the proof follows the same step as [27]):

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

If we apply the change of variable: $x(m) = \frac{-1}{m(x)} + \frac{\gamma}{1+m(x)}$ then:

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

Since C_x and C_y are positive contours that enclose the support of F, then we can choose them to cross the real axis in the intervals $(\frac{-1}{\rho},0)$ (the point $\frac{-1}{\rho}$ is due to the logarithm) and $((1+\sqrt{\gamma})^2, 0)$, the C_y contour encloses the C_x contour. Therefore, the contours C_{m_x} and C_{m_y} cross the real axis in the intervals $(m(\frac{-1}{\rho}),0)$ and $(m((1+\sqrt{\gamma})^2),0)$.

For a fixed m_y, let us calculate:

$$\frac{1}{j2\pi}\int\frac{\log(1+\rho x(m_x))}{(m_x - m_y)^2}dm_x = \frac{1}{j2\pi}\int\frac{\log(1 - \frac{\rho}{m_x} + \frac{\gamma\rho}{1+m_x})}{(m_x - m_y)^2}dm_x$$

$$= \frac{1}{j2\pi}\int\frac{\frac{\rho}{(m_x)^2} - \frac{\rho\gamma}{(1+m_x)^2}}{1 - \frac{\rho}{m_x} + \frac{\rho\gamma}{1+m_x}}\frac{1}{m_x - m_y}dm_x$$

$$= \frac{1}{j2\pi}\int\frac{(1+m_x)^2\rho - (m_x)^2\rho\gamma}{m_x(1+m_x)(m_x - m_y)(m_x - a)(m_x - b)}dm_x$$

$$= \frac{1}{j2\pi}\int\frac{-1}{a}\frac{(1+m_x)^2\rho - (m_x)^2\rho\gamma}{(1+m_x)(m_x - m_y)(m_x - b)}(\frac{1}{m_x} - \frac{1}{m_x - a})dm_x$$

$$= \frac{-\rho}{ab}\frac{1}{m_y} - \frac{1}{a}\frac{(1+a)^2\rho - a^2\rho\gamma}{(1+a)(a-b)}\frac{1}{m_y - a}$$

with

$$a = \frac{1}{2}[-1+\rho(1-\gamma) + \sqrt{(1-\rho(1-\gamma))^2 + \rho}] = m_{f_{\text{iid}}}(\frac{-1}{\rho})$$

and

$$b = \frac{1}{2}[-1+\rho(1-\gamma) - \sqrt{(1-\rho(1-\gamma))^2 + \rho}]$$

The contour C_x is chosen to enclose 0 and a but not -1 and b.

One can easily show that $\frac{-\rho}{ab} = 1$ and $\frac{1}{a}\frac{(1+a)^2\rho - a^2\rho\gamma}{(1+a)(a-b)} = 1$ Therefore, the asymptotic variance is equal to:

$$\begin{aligned}
\sigma_{\text{iid}}^2 &= \frac{1}{j2\pi}\int \log(1+\rho y(m))(\frac{1}{m}-\frac{1}{m-a})dm \\
&= \frac{1}{j2\pi}\int \log(1-\frac{\rho}{m}+\frac{\gamma\rho}{m+1})(\frac{1}{m}-\frac{1}{m-a})dm \\
&= \frac{1}{j2\pi}\int \log(\frac{(m-a)(m-b)}{m(m+1)})(\frac{1}{m}-\frac{1}{m-a})dm \\
&= \frac{1}{j2\pi}\int \log(\frac{m-a}{m})(\frac{1}{m}-\frac{1}{m-a})dm \\
&+ \frac{1}{j2\pi}\int \log(\frac{m-b}{m+1})(\frac{1}{m}-\frac{1}{m-a})dm \\
&= \log(-b)-\log(\frac{(a-b)}{1+a})
\end{aligned}$$

The first integral is zero since the integrand has primitive:

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

which is single valued along the contour.

Therefore, the asymptotic variance is equal to:

$$\begin{aligned}
\sigma_{\text{iid}}^2 &= \log(-b)-\log(\frac{(a-b)}{1+a}) \\
&- \log(1-(\frac{(a-b)}{(1+a)b}+1)) \\
&= -\log(1-(\frac{a(1+b)}{(1+a)b})) \\
&= -\log(1-\frac{a-\rho}{b-\rho})
\end{aligned}$$

The last equation comes from the fact that a and b are solution of:

$$m^2+m(1-\rho+\rho\gamma)-\rho=0$$

Therefore, $ab=-\rho$ and $a+b=-(1-\rho+\rho\gamma)$.
We have therefore:

$$\begin{aligned}
\sigma_{\text{iid}}^2 &= -\log(1-\frac{a-\rho}{b-\rho}) \\
&= \log(1-\frac{(a-\rho)(a-\rho)}{(b-\rho)(a-\rho)}) \\
&= \log(1-\frac{(a-\rho)^2}{\rho^2\gamma})
\end{aligned}$$

Since $\frac{a-\rho}{\rho}=-\frac{1}{2}[1+\gamma+\frac{1}{\rho}-\sqrt{(1+\gamma+\frac{1}{\rho})^2-4\gamma}]=-\alpha(\gamma,\rho)$, we have proved the result.

14.0.3 Proof of Proposition 7

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

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

Using theorem 5, we have:

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

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

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

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

and

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

Therefore,

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

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

which proves the result.

14.0.4 Proof of Proposition 8

In this case,

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

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

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

$$= \alpha_{\text{doa}}\alpha_{\text{dod}}$$

which integrates as:

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

14.0.5   Proof of Proposition 9

In the equal power case, let us apply the result of Proposition 8 when $G_{\lambda^\phi}(\lambda) = \delta(\lambda - \gamma)$ and $G_{\lambda^\psi}(\lambda) = \delta(\lambda - \frac{1}{\xi})$ (note that in all the following derivation, we will often use the fact that $\gamma\xi = \gamma_1\xi_1$). In this case,

$$\mu_{\text{double}} = \xi_1 \ln(1 + \frac{\rho\alpha_{\text{dod}}}{\xi_1}) + \xi \ln(1 + \rho\gamma\alpha_{\text{doa}}) - \rho\alpha_{\text{doa}}\alpha_{\text{dod}} \tag{14.2}$$

with

$$\begin{aligned}
\alpha_{\text{doa}} &= \frac{\xi_1 \frac{1}{\xi_1}}{1 + \frac{\rho\alpha_{\text{dod}}}{\xi_1}} \\
&= \frac{1}{1 + \frac{\rho\alpha_{\text{dod}}}{\xi_1}}
\end{aligned} \tag{14.3}$$

and

$$\alpha_{\text{dod}} = \frac{\gamma\xi}{1 + \rho\gamma\alpha_{\text{doa}}} \tag{14.4}$$

Notice that in this case, we have:

$$\rho\alpha_{\text{doa}}\alpha_{\text{dod}} = \xi_1(1 - \alpha_{\text{doa}}) \tag{14.5}$$

Using equation (14.3) and equation (14.4), α_{doa} is given by:

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

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

which yields:

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

which has the following solution:

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

Since

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

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

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

Now let us derive α_{dod}:

We have

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

and

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

which yields:

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

The solution to this equation is:

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

It can be easily shown that:

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

We have therefore:

$$\alpha_{\text{doa}} = 1 - \alpha_{\text{double}} \tag{14.6}$$

and

$$\alpha_{\text{dod}} = \gamma_1\xi_1 - \gamma\xi_1\alpha_{\text{double}} \tag{14.7}$$

Using eq.(14.2), eq.(14.5), eq.(14.6) and eq.(14.7), one can show that:

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

with

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

14.0.6   Proof of Proposition 5

Let us first derive μ_{doa}. In the DoA based model, one can apply straightforwardly Proposition 11 if $\gamma = \frac{n_r}{s_r}$, $\xi = \frac{s_r}{n_t}$, $\gamma_1 = \frac{n_r}{n_t} = \gamma\xi$, $\xi_1 = \frac{s_t}{n_t} = 1, K_t = 1, P_i^t = 1$. Therefore,

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

with

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

and

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

Notice that

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

and therefore:

$$\alpha_{\text{dod}} = \frac{1}{\rho}\left(\frac{1}{\alpha_{\text{doa}}} - 1\right) \qquad (14.8)$$

and

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

We can therefore rewrite μ_{doa} as:

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

which yields:

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

We also have using eq.(14.8):

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

which can be simplified to:

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

Let us now derive $\sigma_{\text{doa}}{}^2$: To this end, we will apply theorem 4. Since $S_{\text{doa}}(\lambda) = \sum_{i=1}^{K_r} l_i^{\,r}\delta(\lambda - \gamma P_i^{\,r})$, we have:

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

The asymptotic variance is therefore equal to:

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

For fixed m_y, let us calculate:

$$\frac{1}{j2\pi}\int \frac{\log(1-\frac{\rho}{m}+\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{m+\frac{1}{\gamma P_i{}^r}})}{(m-m_y)^2}dm$$

$$=\frac{1}{j2\pi}\int \frac{\frac{\rho}{m^2}-\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{(m+\frac{1}{\gamma P_i{}^r})^2}}{1-\frac{\rho}{m}+\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{m+\frac{1}{\gamma P_i{}^r}}}\frac{1}{m-m_y}dm$$

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

$$=\frac{1}{j2\pi}\int -\frac{1}{m(\frac{1}{\rho})}\frac{\prod_{i=1}^{K_r}\rho(m+\frac{1}{\gamma P_i{}^r})^2-\frac{\rho\xi m^2}{K_r}\prod_{i=1}^{K_r}(m+\frac{1}{\gamma P_i{}^r})^2\sum_{i=1}^{K_r}\frac{1}{(m+\frac{1}{\gamma P_i{}^r})^2}}{\prod_{i=1}^{K_r}(m+\frac{1}{\gamma P_i{}^r})\prod_{i=1}^{K_r}(m-m^i)(m-m_y)}$$

$$\left(\frac{1}{m}-\frac{1}{m-m(-\frac{1}{\rho})}\right)dm$$

$$=\frac{1}{m_y}-\frac{1}{m_y-m(-\frac{1}{\rho})}$$

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

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

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

Notice that:

$$P'(m(\frac{-1}{\rho})) = \prod_{i=1}^{K_r}(m(-\frac{1}{\rho})-m^i) \tag{14.9}$$

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

$$= \prod_{i=1}^{K_r}(m(-\frac{1}{\rho})+\frac{1}{\gamma P_i{}^r})\left(\frac{\rho}{m(-\frac{1}{\rho})}-\rho\xi\sum_{i=1}^{K_r}\frac{l_i{}^r}{(m(-\frac{1}{\rho})+\frac{1}{\gamma P_i{}^r})^2}\right) \tag{14.10}$$

The last equation comes from the fact that:

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

Therefore,

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

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

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

Therefore, the asymptotic variance is equal to:

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

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

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

The previous result comes from equation (14.9) and equation (14.10).
we have therefore:

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

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

---

[1]Note that in the i.i.d Gaussian channel case, $K_r=1$, $n_r=s_r$ and therefore $\gamma=1$ and $\xi=\frac{n_r}{n_t}$. Therefore, one can verify immediately that we obtain the same variance as in chapter 3.1.

14.0.7 Proof of Proposition 12

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

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

This corresponds to maximizing the following function:

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

Therefore,

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

Since

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

and

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

then:

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

Therefore,

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

and

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

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

Bibliography

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

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

[3] K. Yu and B. Ottersten, "Models for MIMO Propagation Channels: A review," *Wireless Communications and Mobile Computing*, vol. 2, pp. 653– 666, November 2002.

[4] H. Ozcelik, N. Czink, and E. Bonek, "What Makes a good MIMO Channel Model," in *Proceedings of the IEEE VTC conference*, 2005.

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

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

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

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

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

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

[11] J. Boutros and E. Viterbo, "Signal Space Diversity: a Power and Bandwidth Efficient Diversity Technique for the Rayleigh Fading Channel," *IEEE Trans. on Information Theory*, pp. 1453–1467, July 1998.

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

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

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

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

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

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

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

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

[20] J.M Keynes, *A Treatise on Probability*, MacMillan and Co., London, 1921.

[21] E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge, 2003.

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

[23] R. T Cox, *Probability, Frequency and Reasonable Expectation*, Am. Jour. Phys., 14:1-13 edition, 1946.

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

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

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

[27] Z.D. Bai and J.W. Silverstein, "CLT of Linear Spectral Statistics of Large Dimensional Sample Covariance Matrices," *Annals of Probability*, vol. 32(1A), pp. 553–605, 2004.

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

[29] M. Rupf and J.L. Massey, "Optimum sequence multisets for synchronous code-division multiple acess channels," *IEEE Trans. on Information Theory*, pp. 1261–1266, July 1994.

[30] P. Rapajic and D. Popescu, "Information Capacity of Random Signature Multiple-Input Multiple Output Channel," *IEEE Trans. on Communications*, vol. 48, pp. 1245–1248, Aug. 2000.

[31] D.N.C Tse and S. Hanly, "Linear Multi-user Receiver: Effective Interference, Effective Bandwidth and User Capacity," *IEEE Trans. on Information Theory*, pp. 641–657, Mar. 1999.

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

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

[34] S. Shamai and S. Verdu, "The Impact of Frequency-Flat Fading on the Spectral Efficiency of CDMA," *IEEE Trans. on Information Theory*, pp. 1302–1326, May 2001.

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

[36] J.M. Chaufray, W. Hachem, and Ph. Loubaton, "Asymptotical Analysis of Optimum and Sub-Optimum CDMA Downlink MMSE Receivers," in *IEEE International Symposium on Information Theory*, Lausanne, Switzerland, July 2002.

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

[38] R. Müller, *Demodulation von CDMA-Signalen bei Modellierung der Teilnehmerinterferenzen als farbiges Rauschen*, Ph.D. thesis, Universität Erlangen-Nüremberg, Erlangen, Germany, Master Thesis, Jan. 1996.

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

[40] A. M. Tulino, S. Verdu, and A. Lozano, "Capacity of Antenna Arrays with Space, Polarization and Pattern Diversity," Paris, France, 2003, pp. 324–327.

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

[42] A.M Tulino, A. Lozano, and S. Verdu, "Capacity of Multi-Antenna Channels with separable and non-separable correlations," Itd-03-44786f, AT & T Bell Labs, 2003.

[43] L. Cottatellucci and M. Debbah, "On the capacity of mimo rice channels," in *Proceedings of the 42nd Annual Allerton Conference on Communication, Control and Computing, oct*, 2004.

[44] T. Tanaka, "Statistical mechanics of cdma multiuser demodulation," in *European Physical Letter*, 2001, pp. 540–546.

[45] T. Tanaka, "A statistical-mechanics approach to large-system analysis of cdma multiuser detectors," *IEEE Trans. on Information Theory*, vol. 48, no. 11, pp. 2888 – 2910, Nov. 2002.

[46] R. Müller and W.H Gerstacker, "On the capacity loss due to separation of detection and decoding," *IEEE Trans. on Information Theory*, vol. 50, no. 8, pp. 1769 – 1778, Aug. 2004.

[47] D. Guo and S. Verdu, "Replica analysis of large-system cdma," in *IEEE International Symposium on Information Theory*, Paris, France, 2003, vol. mar, pp. 22–25.

[48] D. Guo and S. Verdu, *Multiuser Detection and Statistical Mechanics*, in Communications, Information and Network Security (V. Bhargava, H. V. Poor, V. Tarokh, and S. Yoon, eds.), ch. 13, pp. 229-277, Kluwer Academic Publishers, 2002.

[49] H. Jeffreys, *Theory of Probability*, Oxford University Press, London, later editions, 1948, 1961 edition, 1939.

[50] J. Boutros and G. Caire, "Iterative Multiuser Joint Decoding: Unified Framework and Asymptotic Analysis ," *IEEE Trans. on Information Theory*, pp. 1772–1793, July 2002.

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

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

[53] D. Gesbert, M. Shafi, D. Shiu, P.J Smith, and A. Naguib, "From Theory to Practice: an Overview of MIMO Space-Time Coded Wireless Systems," *IEEE Journal on Selected Areas in Communications*, pp. 281– 302, vol. 21, no. 3 2003.

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

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

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

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

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

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

[60] L. Hanlen and A. Grant, "Capacity analysis of correlated mimo channels," in *IEEE International Symposium on Information Theory*, Yokohama, Japan, July 2003.

[61] A. Grant, "Rayleigh fading multiantenna channels," in *Eurasip Journal on Applied Signal Processing, special issue on Space-Time coding*, 2002, vol. 3, pp. 316–329.

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

[63] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multi-Antenna Channel-Hardening and its Implications for Rate Feedback and Scheduling," *IEEE Trans. on Information Theory*, pp. vol. 50, no. 9, Sept. 2002.

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

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

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

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

[68] J. Zhang, E. Chong, and D. Tse, "Output MAI Distributions of Linear MMSE Multiuser Receivers in CDMA Systems," *IEEE Trans. on Information Theory*, pp. 1128–1144, Mar. 2001.

[69] C. E. Shannon, "Communication Theory of Secrecy Systems," *The Bell Labs Technical Journal*, pp. 656–715, May, vol. 28, No 4 1949.

[70] N. Wiener, *Cybernetics*, J. Wiley and Sons, Inc, New York, 1948.

[71] N. Wiener, *Extrapolation, Interpolation and Smooting of Stationary Time Series*, J. Wiley and Sons, Inc, New York, 1949.

[72] Jr G. D. Forney, "On the Role of the MMSE Estimation in Approaching the Information-Theoretic Limits of Linear Gaussian Channels: Shannon meets Wiener," in *41th Annual Allerton Conf. on Comm. Control and Comput., Monticello, Il*, Oct. 2-4, 2003.

[73] D. Guo, S. Shamai, and S. Verdu, "Mutual information and mmse in gaussian channels," in *IEEE International Symposium on Information Theory*, Chicago, IL, USA, 2004, p. 347.

[74] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in gaussian channels," *IEEE Trans. on Information Theory*, vol. 51, no. 4, pp. 1261 – 1282, Apr. 2005.

[75] U. Madhow and M. Honig, "MMSE Interference Supression for Direct-Sequence Spread Spectrum CDMA," *IEEE Trans. on Communications*, vol. 42, pp. 3178–3188, Dec. 1994.

[76] J.M. Cioffi, G.P. Dudevoir, M.V. Eyuboglu, and G.D. Forney, "MMSE Decision Feedback Equalizers and Coding- Part I: Equalization Results," *IEEE Trans. on Communications*, pp. 2582–2594, Oct. 1995.

[77] J.M. Cioffi, G.P. Dudevoir, M.V. Eyuboglu, and G.D. Forney, "MMSE Decision Feedback Equalizers and Coding- Part II: Coding Results," *IEEE Trans. on Communications*, pp. 2595–2604, Oct. 1995.

[78] A. J. Paulraj and T. Kailath, "Increasing Capacity in Wireless BroadcastSystems Using Distributed Transmission/Directional Reception," U.S Patent, No. 5345599, 1994.

[79] M. Varanasi and T. Guess, "Achieving vertices of the capacity region of the synchronous correlated-waveform multiple-access channel with decision-feedback receivers," in *IEEE International Symposium on Information Theory*, 1997, p. 270.

[80] N. Al-Dahir and J.M. Cioffi, "MMSE decision-feedback equalizers:finite-length results," *IEEE Trans. on Information Theory*, pp. 961–975, jul 1995.

[81] Albert M. Chan and Gregory W. Wornell, "A Class of Asymptotically Optimum Iterated-decision Multiuser Detectors," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, Utah, May 2001.

[82] P. Stoica, Y. Jiang, and J. Li, "On MIMO Channel Capacity: An Intuitive Discussion," pp. 83–84, May 2005.

[83] C. E. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE*, pp. 10–21, Jan, vol. 37, No 1 1949.

[84] L. Zheng and D. Tse, "Degrees of Freedom in Underspread MIMO Fading Channels," *submitted to IEEE Transactions on Information Theory*, 2003.

[85] O. Oyman, R. U. Nabar, H. Boelcskei, and A. J. Paulraj, "Characterizing the Statistical Properties of Mutual Information in MIMO Channels: Insights into Diversity-Multiplexing Tradeoff," *IEEE Trans. on Signal Processing*, vol. Vol. 51, p.2784 - 2795, Nov, 2003.

[86] V. Tarokh, N. Seshadri, and A.R. Calderbank, "Space-Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction," *IEEE Trans. on Information Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[87] V. Tarokh, H. Jafarkhani, and A.R. Calderbank, "Space-Time Block Coding for Wireless Communications: Performance Results," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, pp. 451–460, Mar. 1999.

[88] V. Tarokh, A. Naguib, N. Seshadri, and A.R. Calderbank, "Space-Time Codes for High Data Rate Wireless Communication: Performance Criteria in the Presence of Channel Estimation Errors, Mobility, and Multiple Paths," *IEEE Trans. on Communications*, vol. 47, no. 2, pp. 199–207, Feb. 1999.

[89] A. Wittneben, "Base Station Modulation Diversity for Digital SIMULCAST," in *Proceedings of the IEEE Vehicular Technology Conference*, 1991, pp. 848–853.

[90] J. Boutros, "Rotated Multidimensional QAM Constellations for Rayleigh fading channels," IEEE Information Theory Workshop, Haifa, Israel, June 1996, pp. 23–27.

[91] Joseph Boutros, *Lattice codes for Rayleigh Fading Channels*, Ph.D. thesis, École Nationale Supérieure desTélécommunications, June 1996.

[92] J. Boutros, E. Viterbo, C. Rastello, and J.C. Belfiore, "Good Lattice Constellations for Both Rayleigh Fading and Gaussian Channel," *IEEE Trans. on Information Theory*, pp. 502–518, Mar. 1996.

[93] G. R. Lang, "Rotational Transformation of Signals," *IEEE Trans. on Information Theory*, vol. 9, no. 3, pp. 191–198, July 1963.

[94] S.M. Alamouti, "A Simple Transmit Diversity Technique for Wireless Communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[95] V. Tarokh, H. Jafarkhani, and A.R. Calderbank, "Space-Time Block Codes from Orthogonal Designs," *IEEE Trans. on Information Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[96] G.J. Foschini, G.D. Golden, R.A. Valenzuela, and P.W. Wolniansky, "Simplified Processing for High Spectral Efficiency Wireless Communication Employing Multi-Element Arrays," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.

[97] B. Hassibi and B.M. Hochwald, "High Rate Codes that are Linear in Space and Time," *IEEE Trans. on Information Theory*, vol. 48, no. 2, pp. 1804–1824, July 2002.

[98] H. El Gamal and M. O. Damen, "Universal space-time coding," *IEEE Trans. on Information Theory*, pp. 1097–1119, May 2003.

[99] M. O. Damen, K. Abed-Meraim, and J.C Belfiore, "Diagonal algebraic space-time block codes," *IEEE Trans. on Information Theory*, p. vol 48, Mar 2002.

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

[101] M. Franceschetti, S. Marano, and F. Palmieri, "The role of entropy in wave propagation," in *IEEE International Symposium on Information Theory*, Yokohama, Japan, July 2003.

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

[103] J.W. Silverstein, "Weak Convergence of Random Functions defined by the Eigenvectors of Sample Covariance Matrices," *Annals of Probability*, vol. 18, pp. 1174–1194, 1990.

[104] D. Guo, S. Verdu, and L.K Rasmussen, "Asymptotic normality of linear multiuser receiver outputs," *IEEE Trans. on Information Theory*, vol. 48, no. 12, pp. 3080 – 3095, Dec. 2002.

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

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

[107] A. Gupta and D. Nagar, *Matrix Variate Distributions*, Chapman and Hall, Boca Raton, 2000.

[108] G. Lebrun, M. Faulkner, M. Shafi, and P.J. Smith, "MIMO Ricean Channel Capacity," in *Proceedings of the Int. Conf. on Communications*, 2004.

[109] A.L. Moustakas and S.H. Simon, "Optimizing Multiple-Input Single Output (MISO) Communication with General Gaussian Channels: Nontrivial Covariance and non-zero Mean," *IEEE Trans. on Information Theory*, vol. 49, no. 10, pp. 2770–2780, Oct 2003.

[110] V. L. Girko, "Theory of Stochastic Canonical Equations, Volumes I and II ," *Kluwer Academic Publishers, Dordrecht, The Netherlands*, 2001.

[111] L. Cottatellucci and M. Debbah, "The effect of line of sight on the asymptotic capacity of mimo systems," in *International Symposium on Information Theory, Chicago, USA, Pages: 542*.

[112] M. Debbah, "Capacity of a MIMO Multi-user Network," *in*, preparation 2005.

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

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

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

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

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

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

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

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

[121] A. Tulino and S. Verdu, "Asymptotic Analysis of Improved Linear Receivers for BPSK-CDMA Subject to Fading ," *IEEE Journal on Selected Areas in Communications*, pp. 1544–1555, August 2001.

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

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

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

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

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

[127] L. Cottatellucci, R.R Müller, and M. Debbah, "Asymptotic design and analysis of linear detectors for asynchronous cdma systems," in *International Symposium on Information Theory, Chicago, USA, Pages: 509*.

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

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

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

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

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

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

[134] L. Zheng and D. Tse, "Communicating on the Grassmann Manifold: a Geometric Approach to the Non-coherent Multiple Antenna Channel ," *IEEE Trans. on Information Theory*, pp. 359–383, February 2002.

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

[136] E. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multi-antenna systems with different types of channel state information," *IEEE Trans. on Signal Processing*, to appear, 2004.

[137] A.J Goldsmith and P.P Varaija, "Capacity of Fading Channels with Side Information," *IEEE Trans. on Information Theory*, , no. 5, pp. 1986–1992, Nov 1997.

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

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

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

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

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

[143] H. Ozcelik, M. Herdin, W. J. Weichselberg, and E. Bonek, "Deficiencies of "Kronecker" MIMO Radio Channel Model," *IEE Electronics Letters*, vol. 39, no. 16, pp. 1209–1210, Aug. 2003.

[144] T.S. Pollock, "Correlation Modelling in MIMO Systems: When can we Kronecker?," *Australian Communications Theory Workshop*, vol. Newcastle, NSW, no. Australia, pp. 149–153, 2004.

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

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

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

[148] J. Capon, "High-Resolution Frequency Wavenumber Spectrum Analysis," *Proceedings of the IEEE*, vol. 8, no. 57, pp. 1408–1418, 1969.

[149] R.O Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *Proceedings of RADC: Spectral Estimation Workshop, Rome*, pp. 243–258, 1979.

[150] R. Kumaresan and D. W Trufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Transactions on Aerospace and Electronic Systems*, pp. 134–139, 1983.

[151] R. Roy and T. Kailath, "ESPRIT-Estimation of Signal Parameters via Rotational invariance techniques," *IEEE Transactions on Signal Processing*, vol. 7, pp. 984–995, 1989.

[152] R. Müller and H. Hofstetter, "Confirmation of Random Matrix Model for the Antenna Array Channel by Indoor Measurements," in *IEEE Antennas and Propagation Society International Symposium, vol.1, pp.472-475*, Boston, Ma. USA, June 2001.

[153] R. Etkin and D. Tse, "Diversity and Multiplexing: a Fundamental Tradeoff in Multiple-Antenna Channels," *IEEE Trans. on Information Theory*, pp. 1073 –1096, May 2003.

[154] P. H. Lehne and H. Hofstetter, "Documentation of the Measurement Campaign-Part 1: 2.1 GHz, FLOWS deliverable No. D10(1)," Tech. Rep., IST Project Flows, IST-2001-32125, April, 2003.

[155] P. H. Lehne and H. Hofstetter, "Documentation of the Measurement Campaign-Part 2: 5.25 GHz, FLOWS deliverable No. D10(2)," Tech. Rep., IST Project Flows, IST-2001-32125, August, 2003.

[156] P. H. Lehne, H. Hofstetter, and M. Debbah, "Eigenvalue Distributions and Capacity Evaluations from Outdoor MIMO Measurements at 2.1 Ghz," in *IST Mobile Summit, Aveiro, Portugal*, June 2003.

[157] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlson, and M. Beach, "Second Order Statistics of NLOS Indoor MIMO Channels based on 5.2 GHz measurements," *IEEE Trans. on Signal Processing*, pp. 1002–1012, May 2001.

[158] M. Debbah and R. Müller, "Mimo channel modeling and the principle of maximum entropy," *IEEE Transactions on Information Theory*, Vol. 51 , May, No. 5 2005.

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