MIMO Channel Modelling and the Principle of Maximum Entropy, Part II: Model Validation

Mérouane Debbah[1], Ralf Müller[2], Helmut Hofstetter[2] and Per Lehne[3]

Abstract

In [1][1], a unified framework for constructing MIMO models based on the principle of maximum entropy was provided. Several MIMO directional and double directional models were derived and in each case, the mutual information was shown to have asymptotically a Gaussian behavior. Interestingly, the results were shown to be accurate approximations, through simulations, in the finite regime with an astonishing small number of antennas. In this contribution, the models are analyzed and validated[2] based on a wideband outdoor measurement campaign carried out in Oslo during summer 2002 and summer 2003. 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. The measurements performed at 2.1 GHz are relevant for UMTS whereas those at 5.2 GHz are valuable for IEEE 802.11a.

Keywords

MIMO, Entropy, Measurements, Bayesian Probability Theory, Channel Modelling

[1]Mobile Communications Group, Institut Eurecom, 2229 Route des Cretes, B.P. 193, 06904 sophia antipolis cedex, debbah@eurecom.fr

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

[3]Telenor FoU/R&D N-1331 Fornebu Norway

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

[2]The notion of model validation will be defined in this contribution.

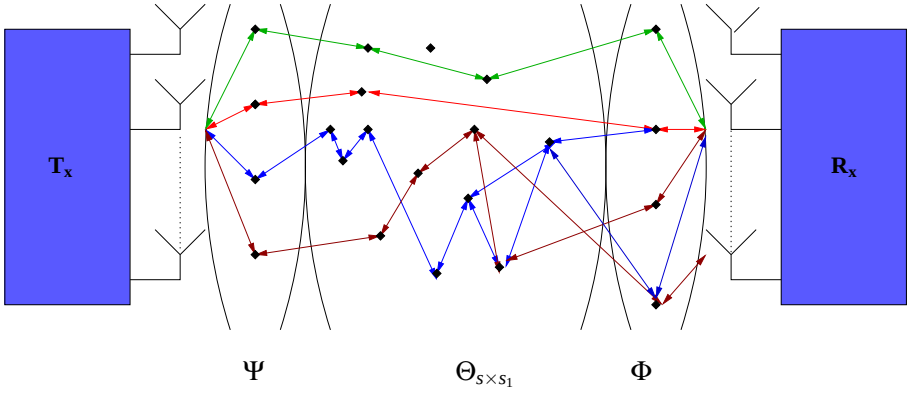

Fig. 1. Double directional based model.

I. Introduction

In [1], several models have been developed based on the principle of consistency. A general framework using the information available was provided to model the MIMO link in the following general frequency representation:

$$Y(f) = \sqrt{\frac{\rho}{n_t}}\mathbf{H}_{n_r \times n_t}(f,t)X(f) + N(f) \tag{1}$$

ρ is the received SNR, f, t, n_t and n_r represent respectively frequency, time, the number of transmitting and receiving antennas, $Y(f)$ is the $n_r \times 1$ received vector , $X(f)$ is the $n_t \times 1$ transmit vector , $N(f)$ is an $n_r \times 1$ additive standardized white Gaussian noise vector. The general double directional representation was shown to be given by (see figure 1):

$$\mathbf{H}(f,t) = \frac{1}{\sqrt{ss_1}}\Phi_{n_r \times s}\mathbf{P^r}\left(\mathbf{\Theta}_{s\times s_1} \bigodot D(f,t)\right)\mathbf{P^t}\Psi_{s_1 \times n_t} \tag{2}$$

Here, s and s_1 represent the scatterers respectively at the receiving and transmitting side. $\Phi_{n_r \times s}$ represents the matrix of steering directions of arrival vectors with respective powers $\mathbf{P^r}$ whereas $\Psi_{s_1 \times n_t}$ represents the matrix of steering directions of departure vectors with respective powers $\mathbf{P^t}$. $\bigodot$ represents the Hadamard product defined as $c_{ij} = a_{ij}b_{ij}$ for a product matrix $\mathbf{C} = \mathbf{A} \bigodot \mathbf{B}$. $\mathbf{\Theta}_{s\times s_1}$ is a $s \times s_1$ matrix with i.i.d Gaussian matrix and $D(f,t)$ is a *pattern mask* which describes the inner scattering environment according the information available. The model is consistent in the sense that this model incorporates

the i.i.d, DoA ($\Phi_{s_1 \times n_t} = F_{n_t}$ and $\mathbf{P}^t = \mathbf{I}_{n_t}$) and DoD based models ($\Psi_{n_r \times s} = F_{n_r}$ and $\mathbf{P}^r = \mathbf{I}_{n_r}$) as special cases from an information theoretic point of view. But how to validate the different models, 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, for example i.i.d, DoA, DoD or double directional model) when measurements are at hand? A common misconception affirms that we can always fit the model to the data as long as one takes enough parameters for the model. In the work of Jaynes [2] and Bretthorst [3], this idea is infirmed and it is explicitly shown how probability theory can be used to rank several models. The reader can find in the appendix a Bayesian ranking procedure. However, the integrals derived in the appendix (equation (6) and equation (7)) 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 [3]. But however difficult the problem may be, it is not a reason to hide problems and any validation method should be clearly explained. The reader must know that Bayesian probability gives means to rank model. In the Bayesian viewpoint, it can be proved that there is an optimum number of parameters when representing information (see [2]). The way to choose a model on behalf of the other is to compare the a posteriori probability ratios: $\frac{P(M|D,I)}{P(M_1|D,I)}$. In any case, whatever the method chosen, one has to clearly understand the limitations of his/her method and justify the use of the criteria. In the following, we explain two procedures used by the channel modelling community and explain their limitations.

A. Conventional methods

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 [4] can be used. In the case of parametric methods such as Music [5], Min-Norm [6] or Esprit method [7], they rely on properties of

[3]The authors are still in the process of numerically calculating the previous formulas through mathematical simplifications.

the structure of the covariance $\mathbf{R} = \mathbb{E}(\mathbf{Y}\mathbf{Y}^{\mathbf{H}}) = \mathbf{\Phi}\mathbf{P}\mathbf{\Phi}^{\mathbf{H}} + \sigma^{\mathbf{2}}$ of the output signal. In this case, one has to assume that matrix $\mathbf{P}$ has full rank ($\mathbf{P} = \mathbb{E}(\Theta\Psi X X^H \Psi^H \Theta^H)$).

• 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 somehow 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?

• If one is to use a Music or Esprit algorithm, $\mathbf{P}$ 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 $\mathbf{P} = \mathbb{E}(\Theta\Psi X X^H \Psi^H \Theta^H)$.

2- Moment fitting:

Other authors [8] 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:

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

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?[4].

B. 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 [1], the mutual information distribution of many models has been provided and in many cases, the cumulative distribution function of the mutual information was shown to have the following form:

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

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:

$$\int_0^\infty \mid F(I^M) - F_{\text{empirical}}(I^M, f) \mid^2 dI^M \tag{3}$$

Here $F_{\text{empirical}}(I^M, f)$ is the empirical CDF given by measurements[5]. 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 (3) with respect to the steering directions. 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

[4]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.

[5]Note that we could also minimize the Kullback distance between the two distributions:

$$D(P, P_{\text{empirical}}) = \int P(I^M) \log\left(\frac{P(I^M)}{P_{\text{empirical}}(I^M)}\right) dI^M \tag{4}$$

where P and $P_{\text{empirical}}$ are respectively the theoretical and empirical probability distribution of the mutual information.

information and therefore we can use (as a first approximation) Fourier directions. In this case, let us review each model (ρ is the SNR) given in [1]:

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 (3) with respect to the number of scatterers s for which the expression of $\mu_{\rm doa}$ and $\sigma_{\rm doa}$ are given by:

$$\begin{aligned}\mu_{\rm doa}(n_r,s,n_t,\rho) &= \frac{s}{n_t}\ln(1+\rho\frac{n_r}{s}-\rho\frac{n_r}{s}\alpha_{\rm doa}(n_r,s,n_t,\rho))+\ln(1+\rho\frac{n_r}{n_t}\\ &- \rho\frac{n_r}{s}\alpha_{\rm doa}(n_r,s,n_t,\rho))-\alpha_{\rm doa}(n_r,s,n_t,\rho)\end{aligned}$$

and

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

with

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

Since s will depend on the frequency, we will define as in [8] the richness spectrum as: $R_{\rm doa}=\frac{s(f)}{n_r}$

DoD based model. In this case, one has to optimize criteria (3) with respect to the number of scatterers s_1 for which the expression of $\mu_{\rm dod}$ and $\sigma_{\rm dod}$ are given by:

$$\begin{aligned}\mu_{\rm dod}(n_r,s_1,n_t,\rho) &= \frac{s_1}{n_t}\ln(1+\rho\frac{n_r}{s_1}-\rho\alpha_{\rm dod}(n_r,s_1,n_t,\rho))+\frac{n_r}{n_t}\ln(1+\rho-\rho\alpha_{\rm dod}(n_r,s_1,n_t,\rho))\\ &- \frac{s_1}{n_t}\alpha_{\rm dod}(n_r,s_1,n_t,\rho)\end{aligned}$$

and

$$\sigma^2_{\rm dod}(n_r, s_1, n_t, \rho) = -\ln[1 - \frac{s_1}{n_r}\alpha^2_{\rm dod}(n_r, s_1, n_t, \rho)]$$

with

$$\alpha_{\rm dod}(n_r, s_1, n_t, \rho) = \frac{1}{2}[1 + \frac{n_r}{s_1} + \frac{1}{\rho} - \sqrt{(1 + \frac{n_r}{s_1} + \frac{1}{\rho})^2 - 4\frac{n_r}{s_1}}]$$

Since s_1 will also depend on the frequency, we can also define the richness spectrum as: $R_{\rm dod} = \frac{s_1(f)}{n_t}$

Double Directional model. In this case, one has to optimize criteria (3) with respect to s (scatterers at the receiving side) and s_1 (scatterers at the transmitting side) for which the expressions of $\mu_{\rm double}$ and $\sigma_{\rm double}$ are given by:

$$\begin{aligned}\mu_{\rm double}(n_r, s, s_1, n_t, \rho) &= \frac{s}{n_t}\ln(1 + \rho\frac{n_r}{s} - \rho\frac{n_r}{s}\alpha_{\rm double}(n_r, s, s_1, n_t, \rho)) \\ &+ \frac{s_1}{n_t}\ln(1 + \rho\frac{n_r}{s_1} - \rho\frac{n_r}{s}\alpha_{\rm double}(n_r, s, s_1, n_t, \rho)) \\ &- \frac{s_1}{n_t}\alpha_{\rm double}(n_r, s, s_1, n_t, \rho)\end{aligned}$$

and

$$\sigma^2_{\rm double}(n_r, s, s_1, n_t, \rho) = -\ln(1 - \frac{\alpha^2_{\rm double}s_1}{s})$$

with

$$\alpha_{\rm double}(n_r, s, s_1, n_t, \rho) = \frac{1}{2}\left[1 + \frac{s}{s_1} + \frac{s}{\rho n_r} + \sqrt{(1 + \frac{s}{s_1} + \frac{s}{\rho n_r})^2 - 4\frac{s}{s_1}}\right]$$

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 and one focuses on BER, the model may be completely inadequate even though it complies with mutual information measurements. As matter of fact, 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 and $s_1 = n_t$. 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 (to be explained in section I-C). 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 but only that it fulfills our mutual information criteria. Of course, other quality measures can be given such as Signal to interference plus Noise ratio or Bit error rate.

C. Clearing up mysteries

In [1], the mutual information was derived based on the assumption that the channel model is adequate with reality. For example, knowing that the frequency paths are Gaussian i.i.d and the noise is additive white Gaussian, the transmitter will design codes to ensure a reliable transmission on such channels. But whenever we are misrepresenting the channel with our state of knowledge, the formula

$$C(\mathbf{I}) = \log_2 \det\left(\mathbf{I}_{n_t} + \frac{\rho}{n_t}\mathbf{H}^H\mathbf{H}\right) \tag{5}$$

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 (5). The problem is formulated in Figure 2.

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

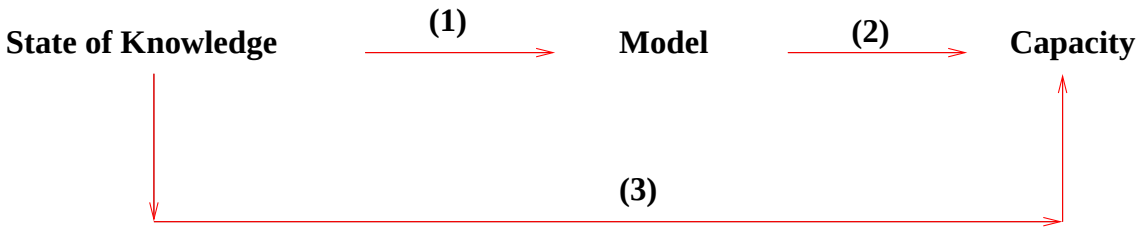

Fig. 2. Channel modelling approach and derivation of capacity.

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 2. 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: $\mathbf{H} = \mathbf{\Phi}\Theta$, ($\Theta$ is i.i.d Gaussian) then using Jensen's Inequality:

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

For example, when the directions of arrival are unknown, the mutual information averaged across the unknown directions of arrival (here $\mathbb{E}_{\mathbf{\Phi}}\left(\log\det(\mathbf{I}_{n_t} + \frac{\rho}{n_t}(\Theta^H\mathbf{\Phi}^H\mathbf{\Phi}\Theta)\right)$)

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

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

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

II. Measurement set-up

In this section[6], we describe the wideband outdoor measurement campaign carried out in Oslo during summer 2002 and summer 2003 [11]. 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 time snapshots) whereas at 5.2 GHz, 79 routes have 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.

- The street grid scenario is in Oslo downtown and corresponds to Concrete/Brick buildings. The buildings are around 20-30 m high (see Figure 3). 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".
  - For the low base station at 2.1 GHz, route Kvadraturen_01_01_15 has been studied.

[6] A comprehensive introduction to the measurement set-up can be found in [9] and [10]. For the different scenario routes, notations of [9] and [10] will be used.

Fig. 3. Urban Regular site.

Fig. 4. Urban Open Place.

– 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[7], 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 4). 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.

– 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

[7]At the 5.25 GHz, only the case where the receiver was at the street level was considered and not the high base station case.

Fig. 5. Telenor Headquarters: work zone.

Fig. 6. Telenor Headquarters: work zone.

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 5 and 6) and in a common area called the "Atrium" (see Figure 7).

– 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.

Fig. 7. Telenor Headquarters: "Atrium".

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

A. Channel sounder at 2.1 GHz

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 8). 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 (I). The

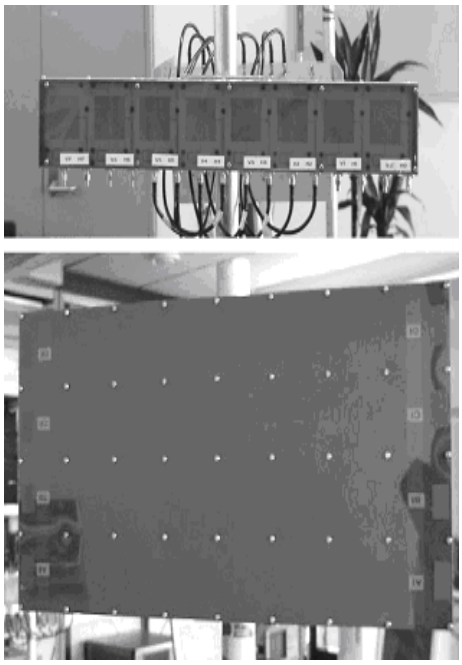

Fig. 8. Antenna: The receiver is the 8×4 planer array while the transmitter is the 8 element uniform linear array.

sounder was manufactured by SINTEF Telecom and Informatics in Trondheim, Norway, on assignment from Telenor.

B. Channel sounder at 5.25 GHz

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 (II).

III. 2.1 GHz

A. State of knowledge

According to the specifications in table (I) 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 I

CHANNEL SOUNDER SPECIFICATION AT 2.1 GHZ.

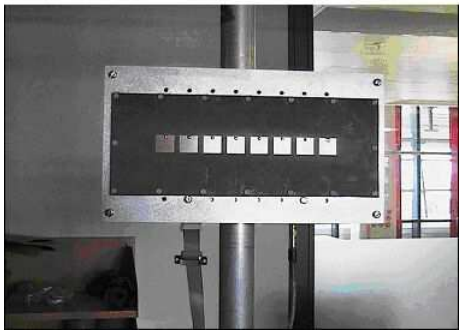

Fig. 9. Antenna at 5.25 GHz.

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 II

CHANNEL SOUNDER SPECIFICATION AT 5.255 GHZ.

- Uniform Linear Array.
- Far field Approximation.
- Antenna spacing 0.5λ .
- 8×4 antenna system.
- Dual polarization at the transmitter

Since there is no LOS between the transmitter and receiver assumed in the measurement set-up, we can assume that the maximum entropy model based on energy constraints (as recalled in [1], 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. In the general case, the received signal can be written:

$$Y = \sqrt{\frac{\rho}{n_t}}\mathbf{H}_h X_h + \sqrt{\frac{\rho}{n_t}}\mathbf{H}_v X_v + N$$

Here, $\mathbf{H}_h$[8] and $\mathbf{H}_v$ are $n_r \times \frac{n_t}{2}$ matrices whereas X_h and X_v are $\frac{n_t}{2} \times 1$ vectors[9].

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

$$\begin{aligned} Y &= \sqrt{\frac{\rho}{n_t s}}(\Phi_{n_r \times s}\Theta_h X_h + \Phi_{n_r \times s}\Theta_v X_v) + N \\ &= \sqrt{\frac{\rho}{n_t s}}\Phi_{n_r \times s}[\Theta_h \Theta_v] \begin{bmatrix} X_h \\ X_v \end{bmatrix} + N \\ &= \sqrt{\frac{\rho}{n_t s}}\Phi_{n_r \times s}\Theta_{h+v} X_{h+v} + N \end{aligned}$$

Θ_{h+v} is an $s \times n_t$ Gaussian i.i.d matrix and X_{h+v} is a $n_t \times 1$ vector. Therefore, all the results of section IV of [1] can be used without any change. In the DoA based model, the

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

[9] n_t, which represents the number of equivalent number of antennas, is equal to 8

polarization does not change anything with respect to a system without polarization and a double number of transmitting antennas.

- DoD based model: In this case, the received signal can be written under the form:

$$Y = \sqrt{\frac{\rho}{n_t s_1}}(\Theta_h \Psi_{s_1 \times \frac{n_t}{2}} X_h + \Theta_v \Psi_{s_1 \times \frac{n_t}{2}} X_v) + N$$

Θ_h and Θ_v are $n_r \times s_1$ i.i.d Gaussian matrices which represent the horizontal and the vertical polarization respectively. The mutual information is given by:

$$\begin{aligned} I^M(n_t, n_r, s_1, \rho) &= \log_2 \det \left(\mathbf{I}_{n_r} + \frac{\rho}{n_t s_1} (\Theta_h \Psi \Psi^H {\Theta_h}^H + \Theta_v \Psi \Psi^H {\Theta_v}^H) \right) \\ &= \log_2 \det \left(\mathbf{I}_{n_r} + \frac{\rho}{2 s_1} (\Theta_h {\Theta_h}^H + \Theta_v {\Theta_v}^H) \right) \\ &= \log_2 \det \left(\mathbf{I}_{n_r} + \frac{\rho}{2 s_1} \Theta_{h+v} {\Theta_{h+v}}^H \right) \end{aligned}$$

Θ_{h+v} is an $n_r \times 2s_1$ matrix. One can observe that all the results of section V of [1] can be used if one replaces s_1 with $2s_1$. Note that s_1 cannot be greater than $\frac{n_t}{2}$ in the Fourier framework.

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

$$Y = \sqrt{\frac{\rho}{n_t s s_1}}(\Phi_{n_r \times s} \Theta_h \Psi_{s_1 \times \frac{n_t}{2}} X_h + \Phi_{n_r \times s} \Theta_v \Psi_{s_1 \times \frac{n_t}{2}} X_v) + N$$

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

$$\begin{aligned} I^M(n_t, n_r, s, s_1, \rho) &= \log_2 \det \left(\mathbf{I}_{n_r} + \frac{\rho}{n_t s s_1} (\Phi \Theta_h \Psi \Psi^H {\Theta_h}^H \Phi^H + \Phi \Theta_v \Psi \Psi^H {\Theta_v}^H \Phi^H) \right) \\ &= \log_2 \det \left(\mathbf{I}_{n_r} + \frac{\rho}{2 s s_1} (\Phi \Theta_h {\Theta_h}^H \Phi^H + \Phi \Theta_v {\Theta_v}^H \Phi^H) \right) \\ &= \log_2 \det \left(\mathbf{I}_{n_r} + \frac{\rho}{2 s s_1} \Phi (\Theta_h {\Theta_h}^H + \Theta_v {\Theta_v}^H) \Phi^H \right) \\ &= \log_2 \det \left(\mathbf{I}_{n_r} + \frac{\rho}{2 s s_1} \Phi (\Theta_{h+v} {\Theta_{h+v}}^H) \Phi^H \right) \\ &= \log_2 \det \left(\mathbf{I}_{s_1} + \frac{\rho}{2 s s_1} {\Theta_{h+v}}^H \Phi^H \Phi \Theta_{h+v} \right) \\ &= \log_2 \det \left(\mathbf{I}_{s_1} + \frac{\rho n_r}{2 s s_1} {\Theta_{h+v}}^H \Theta_{h+v} \right) \end{aligned}$$

Θ_{h+v} is an $s \times 2s_1$ i.i.d Gaussian matrix. Therefore, all the results of section VI of [1] can be used if one replaces s_1 with $2s_1$. Note that s_1 cannot be greater than $\frac{n_t}{2}$ in the Fourier framework.

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

B. Are the measured 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 10, 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:

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

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 (except for the regular high antenna case) [10]. In the following section, we will see how close are the measured capacities from the maximum entropy models.

C. What about frequency selectivity?

In the model derived in [1], we argued that frequency selectivity does not affect the mutual information. In Figure 11, 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 position, the urban regular high antenna position, the indoor and Atrium scenario.[11]. As one can observe, for the different frequencies, the mutual informa-

[10] Actually, from all the 150 routes available in [9], 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.

[11] Note that the cdf has been averaged over different time snapshots but at the same frequency.

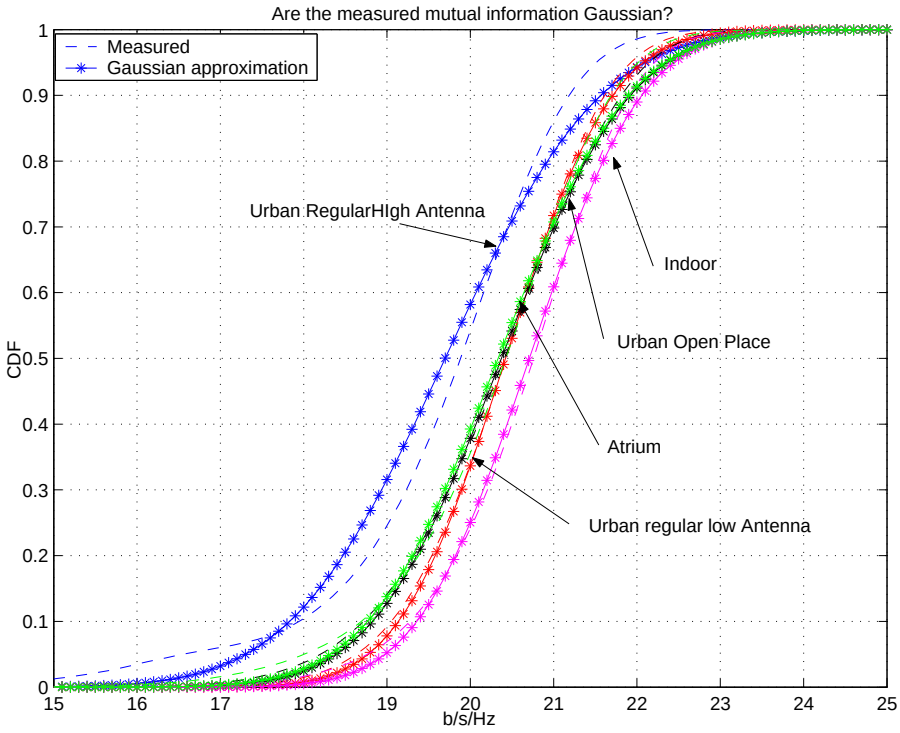

Fig. 10. Are the measured mutual information Gaussian

tion 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.

D. Results

In Figure 12, 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 achiev-

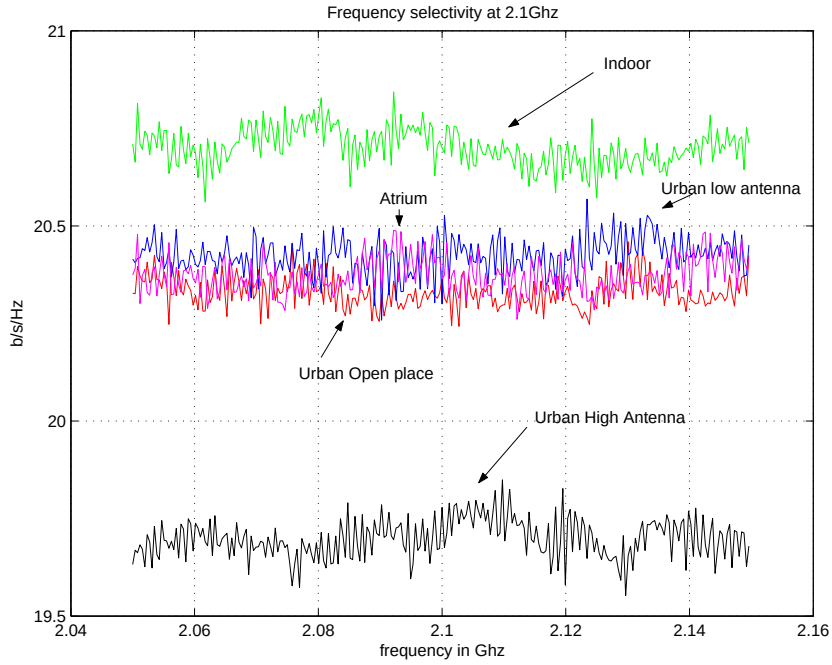

Fig. 11. Frequency selectivity for many scenarios at 2.1 GHz

able rate. An average gap with measurements of 3 b/s/Hz exists.

- DoD model: The DoD model gives a better performance than the i.i.d one and underestimates the rate by 1 b/s/Hz. The optimal number of scatterers s_1 is equal to 3.
- 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 equal to 6.
- Double directional model: The double directional model fits accurately the data with a number of scatterers equal to $s = 7$ and $s_1 = 3$. 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 13, 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.
- DoD model: The DoD based model gives better performance than the i.i.d model

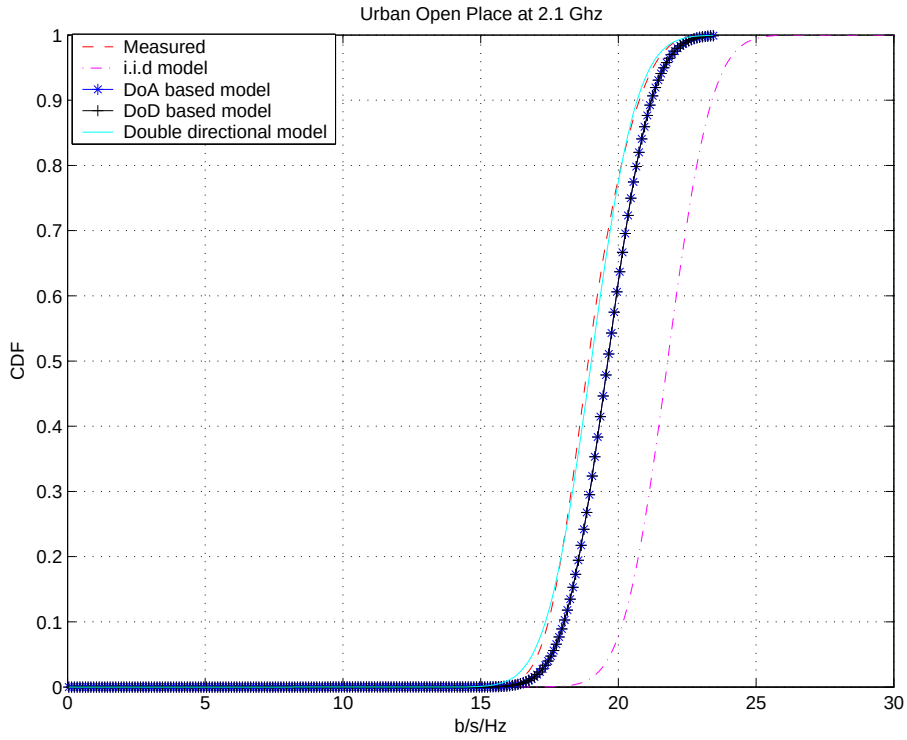

Fig. 12. Urban Open Place at 2.1 GHz.

but is still at 2 b/s/Hz from the measured data. The optimal number of scatterers is equal to $s_1 = 3$

- DoA model: The DoA based model fits quite accurately the data with a number of scatterers equal to $s = 5$. 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 $s = 5$ and $s_1 = 4$. Moreover, one can observe that there are more scatterers on the receiving side than the transmitting side ($s > s_1$). 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.

In Figure 14, 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.

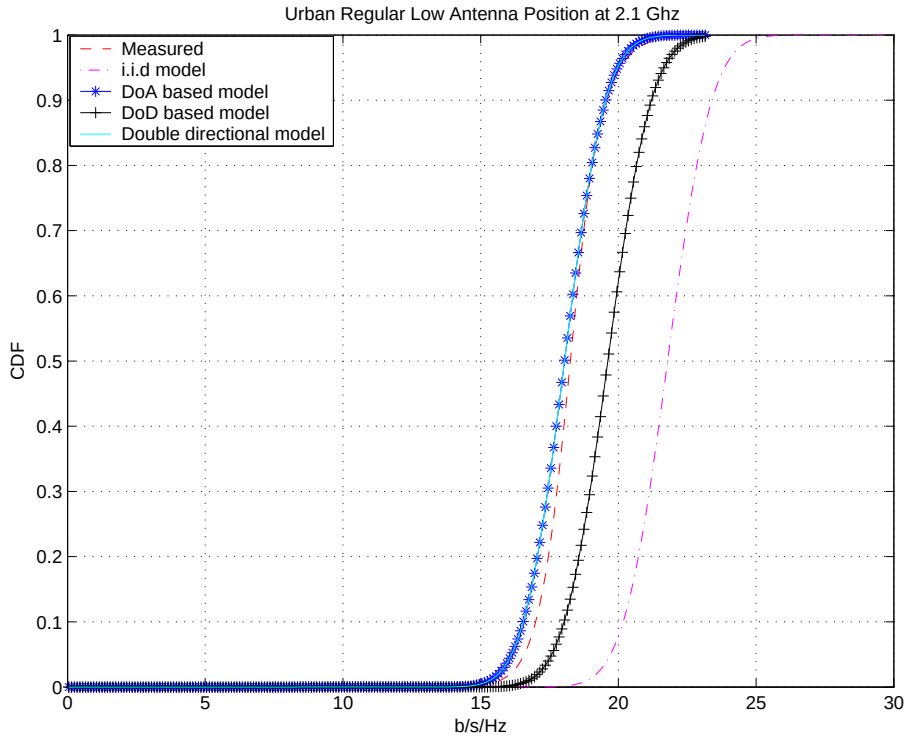

Fig. 13. Urban Regular, Low Antenna Position at 2.1 GHz.

• DoD model: The DoD model does not fit the data. The best result is obtained for $s_1 = 3$.

• DoA model: In this case, the optimal number of scatterers is $s = 5$. A gap still appears at the lower tail of the curve.

• Double directional model: The double directional model fits with $s = 6$ and $s_1 = 3$ but there is still a mismatch in this case. There may be a line of sight component that we should assume in this case.

In Figure 15, 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.

• DoD model: The DoD case yields the same performance as the i.i.d one. The best fit is obtained for a number of scatterers equal to $s_1 = 4$.

• DoA model: In the DoA case, $s = 7$ provides quite accurate results. However, better results could be obtained if the power of the steering directions is to be taken into account.

• Double directional model: For $s = 7$ and $s_1 = 4$, the same performance as the DoA

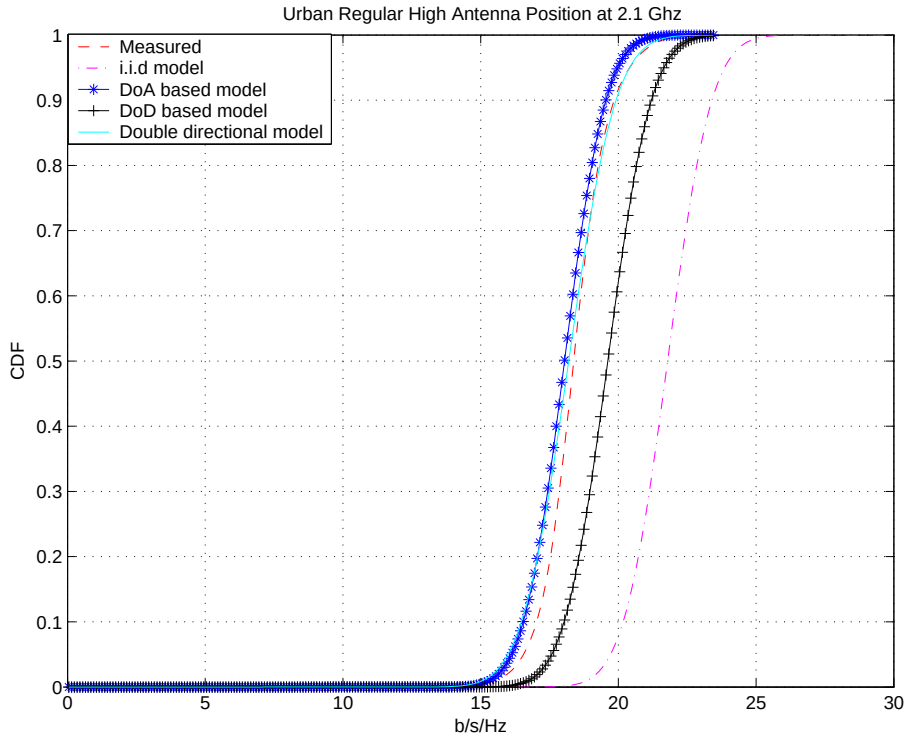

Fig. 14. Urban Regular, High Antenna Position at 2.1 GHz.

model is achieved. As previously, the power of the steering vectors should be taken into account to achieve better results.

In Figure 16, 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 $s = 7$ 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 $s_1 = 4$.

• Double directional model: The double directional model gives the same performance as the DoA based. The best fitting is obtained for $s = 7$ and $s_1 = 4$.

In summary, the maximum entropy Fourier directional models give accurate results. Note that for improving the results, one can take into account the following parameters:

– The non-zero mean of the different paths

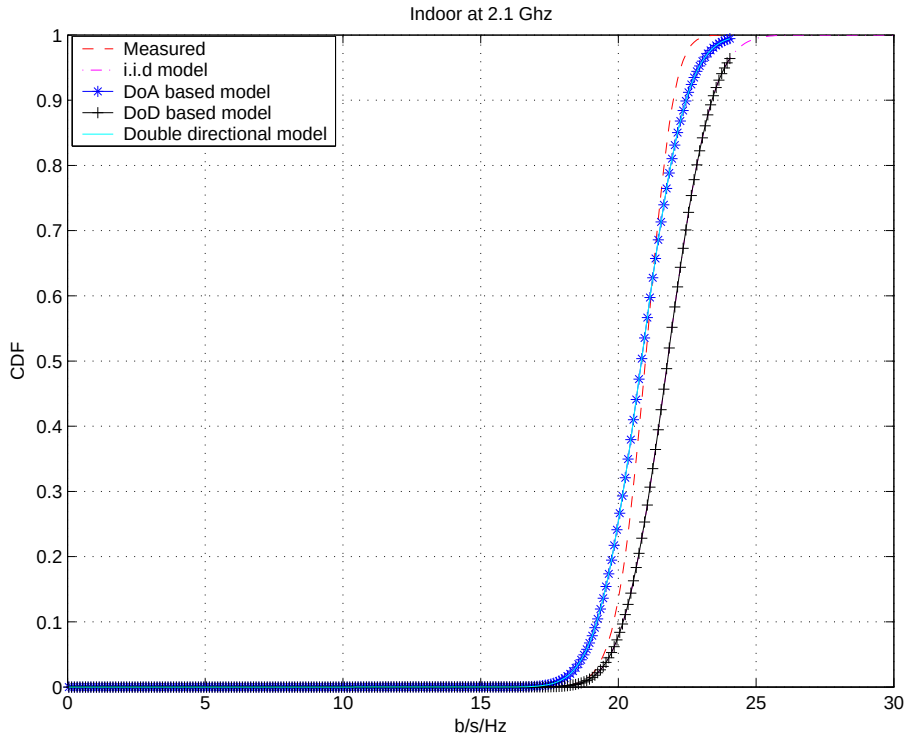

Fig. 15. Indoor scenario at 2.1 GHz.

– The power profile of the steering directions.

– Retrieve the hypothesis of far field scattering and use the near field assumption.

IV. 5.25 GHz

A. *State of knowledge*

According to the specifications in table (II) 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 [1] 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.

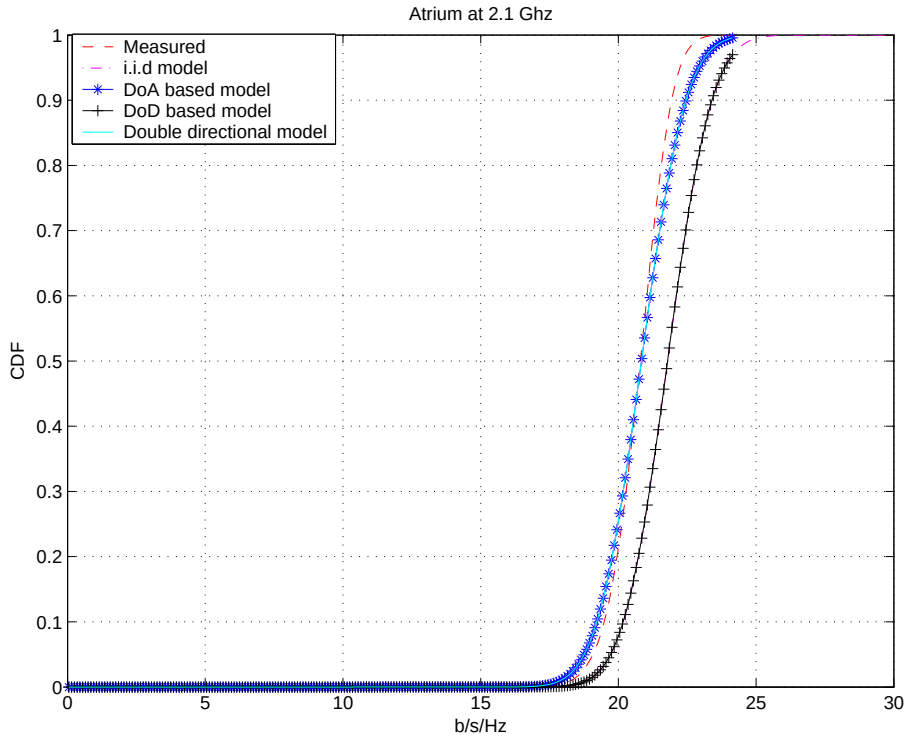

Fig. 16. "Atrium" scenario at 2.1 GHz.

B. 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 17, 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 [12]. In the following section,

[12]Actually, from all the 79 routes available in [10], 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.

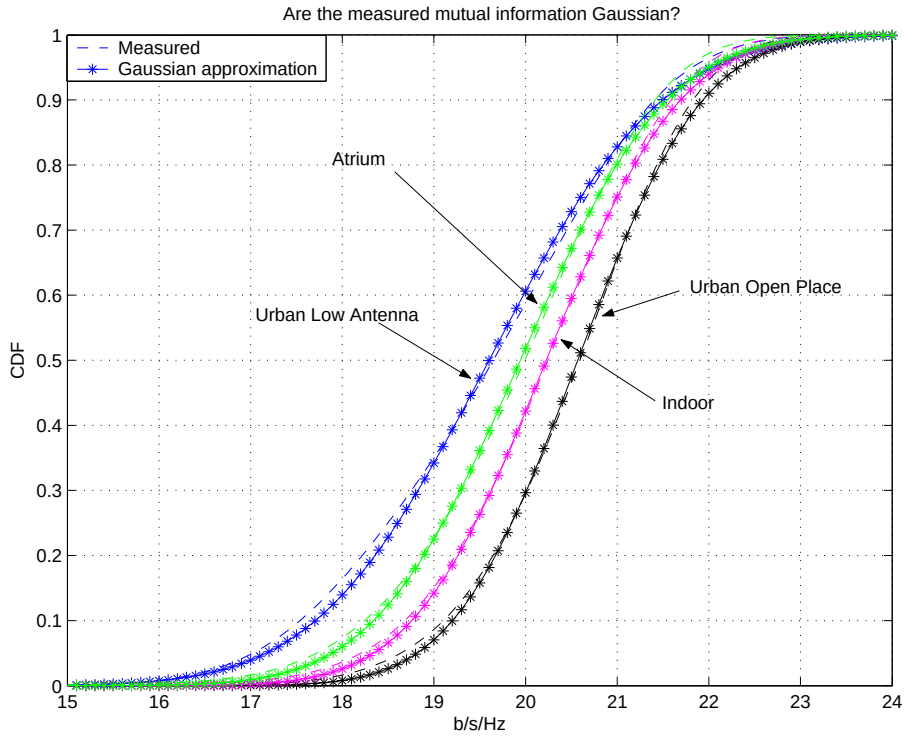

Fig. 17. Are the measured mutual information Gaussian

we will see how close are the maximum entropy models from the measurements.

C. What about frequency selectivity?

In the models derived in [1], we argued that frequency selectivity does not affect the mutual information. In Figure 18, 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, 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%.

D. Results

In Figure 19, 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.

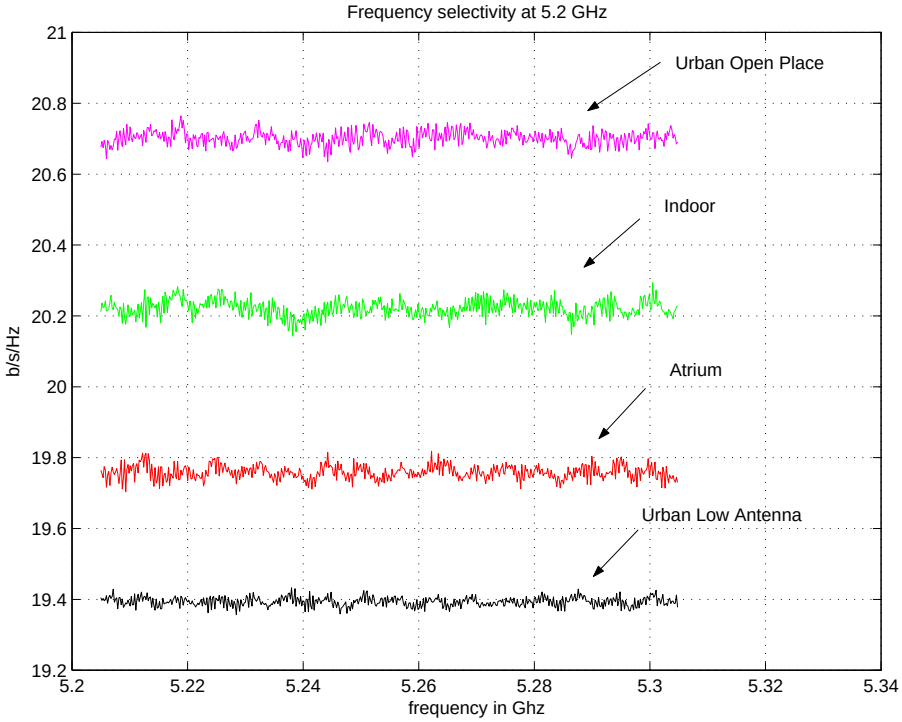

Fig. 18. Frequency selectivity for various scenarios at 5.2 GHz

• 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 $s = 8$ and gives the same performance (this is due to the fact that our models are consistent) as the i.i.d Gaussian case.

• DoD model: In this case also, the curve does not fully fit the measurement and gives the same performance as the i.i.d Gaussian for $s_1 = 8$.

• Double directional model: The double directional model does not fit accurately the data with a number of scatterers equal to $s = 8$ and $s_1 = 8$. 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 20, 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 1 b/s/Hz is revealed.

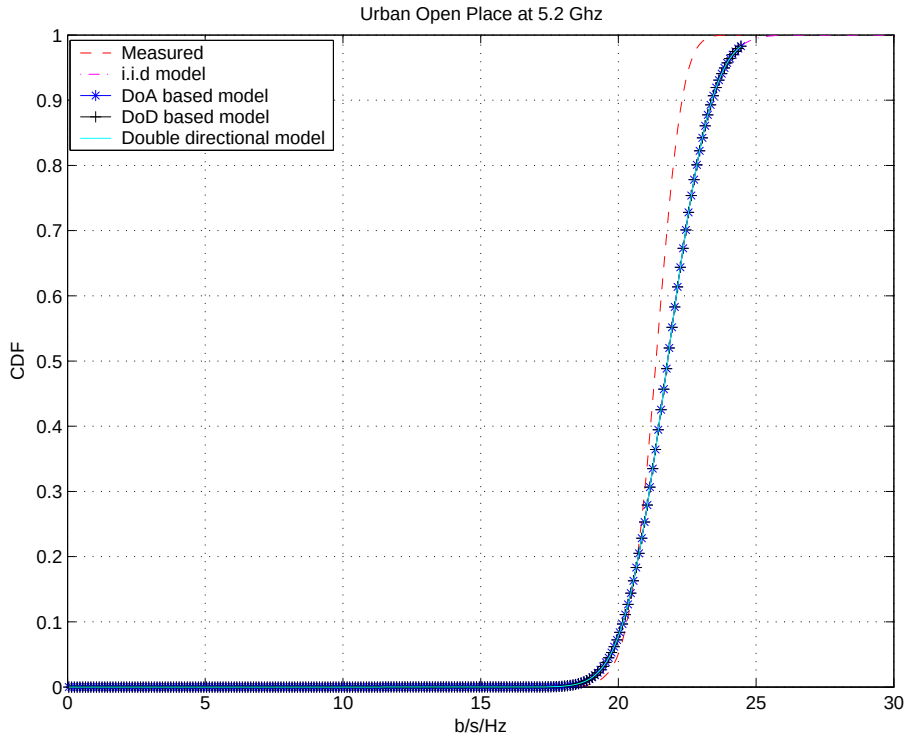

Fig. 19. Urban Open Place at 5.2GHz.

• DoA model: The DoA based model fits the data with a number of scatterers equal to $s = 7$.

• DoD model: The DoD based model gives the same performance as the DoA one and fits the data with $s_1 = 7$.

• Double directional model: The double directional model gives accurate results with a number of scatterers equal to $s = 8$ and $s_1 = 7$. One can also notice that the number of scatterers at 5.2 GHz is higher than at 2.1 GHz.

In Figure 21, 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 $s = 6$

• DoD model: The DoD model gives the same performance as the DoA one with $s_1 = 6$

• Double directional model: The double directional model gives accurate results with $s = 7$ and $s_1 = 7$.

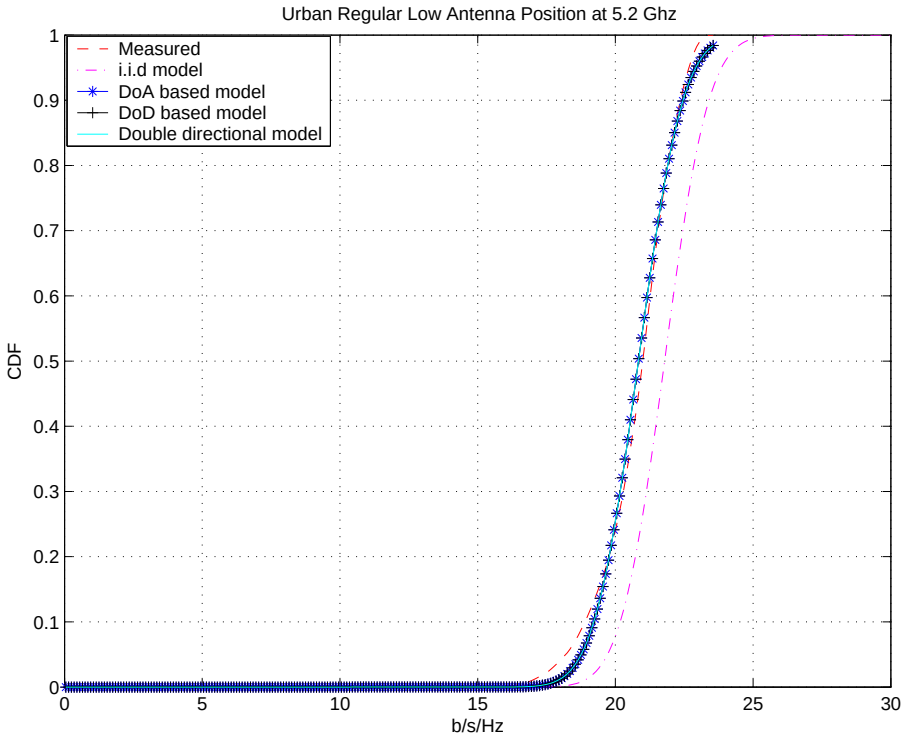

Fig. 20. Urban Regular, Low Antenna Position at 5.2 GHz.

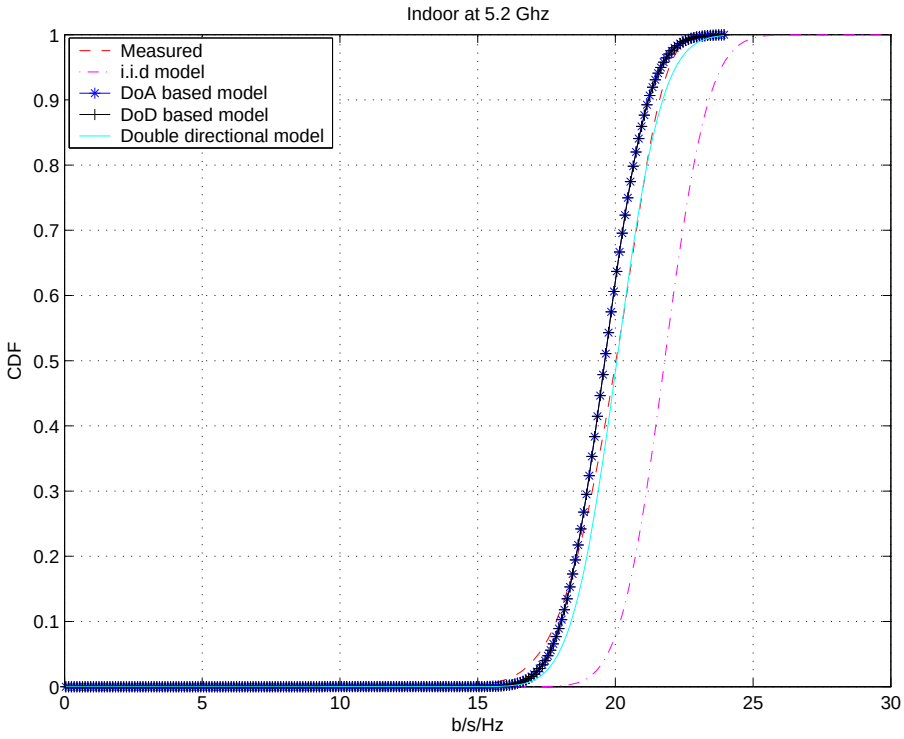

Fig. 21. Indoor scenario at 5.2 GHz.

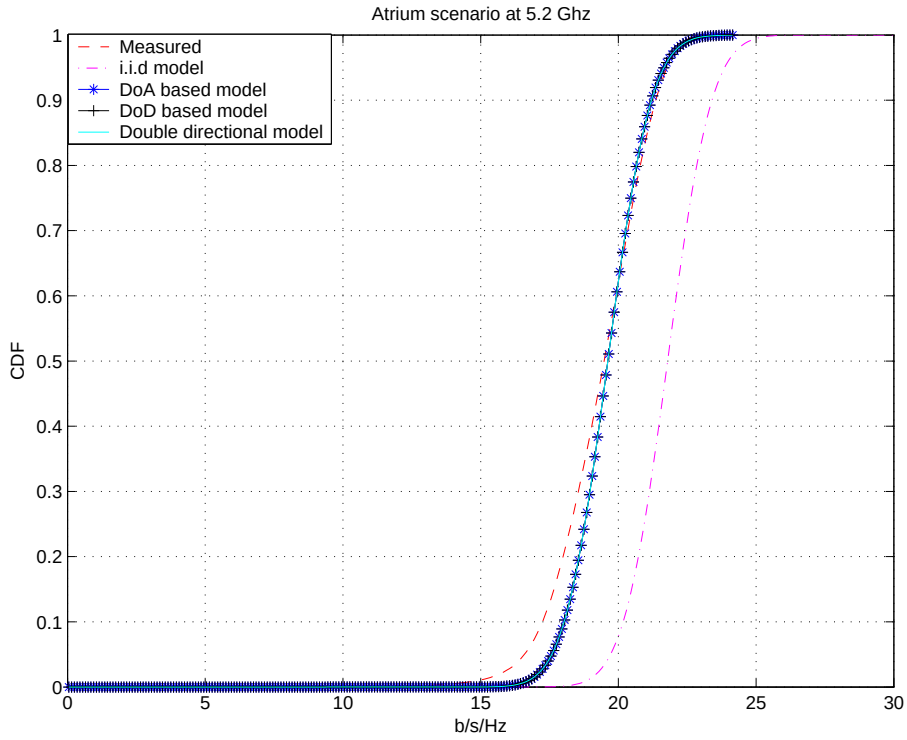

Fig. 22. "Atrium" scenario at 5.2 GHz.

In Figure 22, 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: 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 $s = 6$. However, a little mismatch appears at the lower edge of the curve.
- DoD model: The DoD model also fits the data for $s_1 = 6$ and gives the same performance as the DoA model.
- Double directional model: The double directional fits the data for $s = 6$ and $s_1 = 8$ 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 or the non-zero mean.

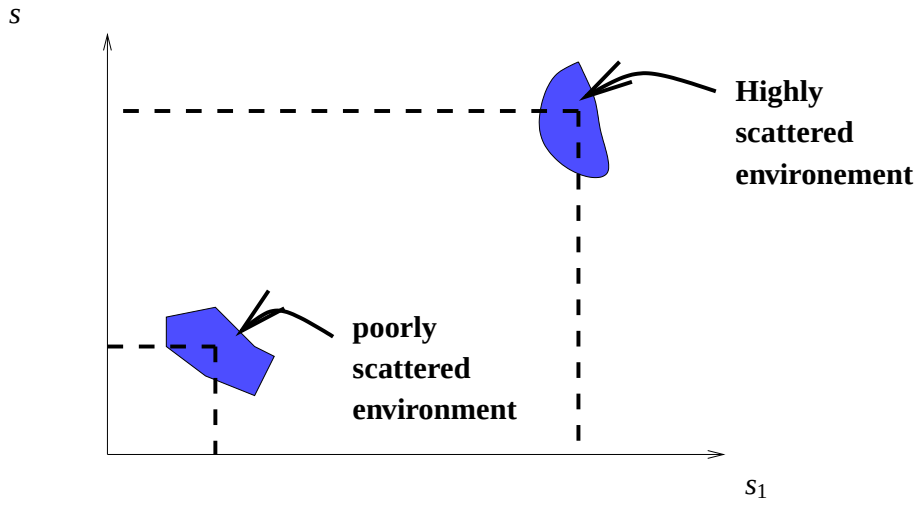

Fig. 23. Characterization of the environment with s and s_1.

V. 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, s_1): Depending on their value, these numbers would characterize entirely the environment as shown in Figure 23[13]. A quick look at the theoretical mutual information equations of section. I-B 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 [1], 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 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

[13]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.

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 [12] 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 25 the optimal number of scatterers for the Urban Regular Low Antenna scenario with 2×2, 3×3, 4×4, 5×5, 6×6, 7×7 and 8×8 antennas (different antennas have been used for the same scenario) at 5.2 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[14] (for which all the scatterers can be captured), we have plotted in figure (24) the cdf of the mutual information for a 2×2, 3×3, 4×4, 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

[14]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

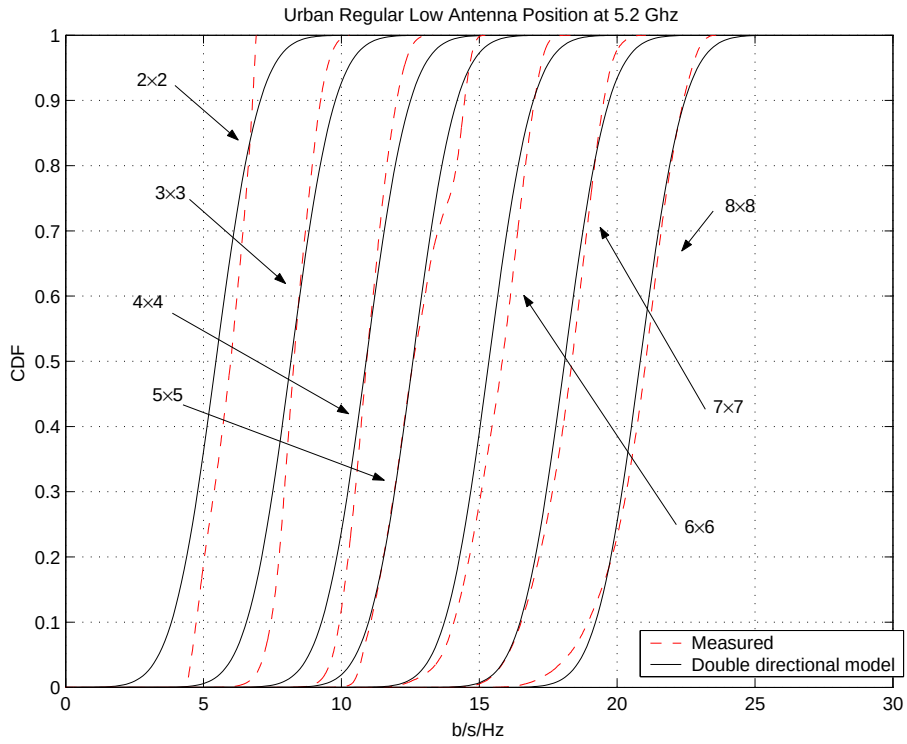

Fig. 24. 6×6 antenna system.

ones).

VI. 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.
- 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.

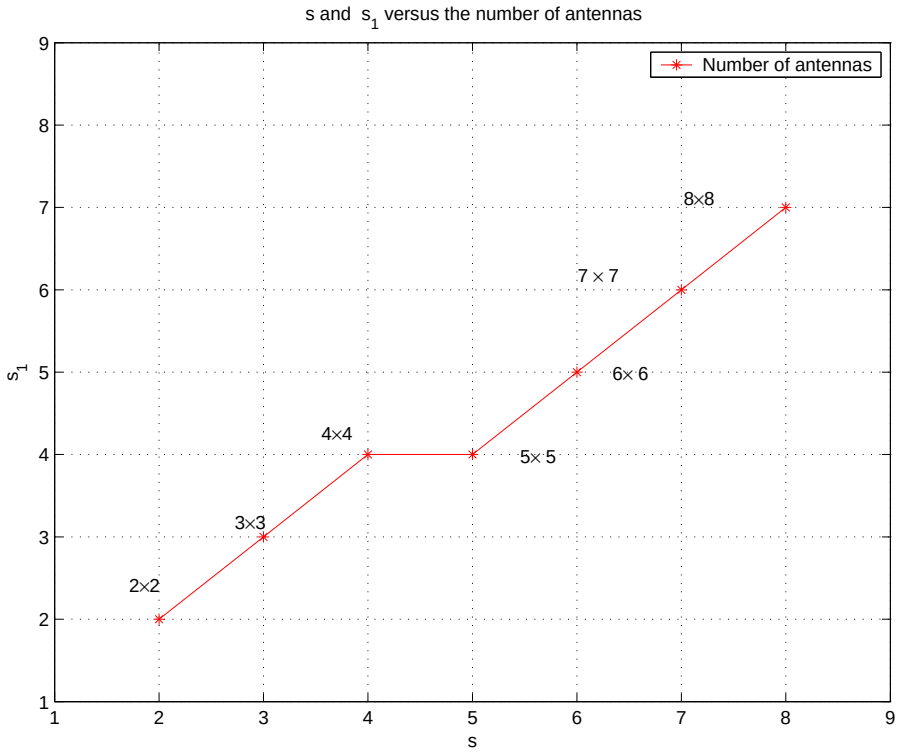

Fig. 25. s and s_1 versus the number of antennas.

- 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.

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.

VII. Appendix

We give hereafter[15] a method to rank several models. When judging the appropriateness of a model, one has to derive the posterior probability of the model. Bayes rule gives the posterior probability for the i^{th} model according to: [16]

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

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: $\frac{P(M_1|I)}{P(M|I)}$. It favors one model or the other before the observation. As an example, suppose that the information $\{I =$ 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 $\frac{P(M_1|I)}{P(M|I)} > 1$.

Let us now evaluate the second term for two models of interest: the DoA based model M_{doa} and the double directional model M_{double}.

Model M_{doa}:

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

with

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

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

[16] We use here the notations and meanings of Jaynes [2] and Jeffrey [13]: $P(M_i \mid Y, I)$ 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.

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 $\mathbf{u}_i^r \mathbf{v}_r = (\cos({\beta^r}_i)\tilde{\imath} + \sin({\beta^r}_i)\tilde{\jmath})(v_r \cos(\alpha_r)\tilde{\imath} + v_r \sin(\alpha_r)\tilde{\jmath}) = v_r \cos({\beta^r}_i - \alpha_r)$

The set of parameters on which the model is defined is

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

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

$$Y(t,f) = \frac{1}{\sqrt{s}} \Phi \left(\Theta \bigodot D(f,t) \right) X(f) + N(f)$$

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

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

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

$$\frac{1}{(2\pi\sigma^2)^{\frac{N_1 N_r}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{N} \sum_{j=1}^{N_1} \left(Y(t_j,f_i) - \frac{1}{\sqrt{s}} \Phi(\Theta \bigodot D(t_j,f_i)) X(f_i) \right)^H \left(Y(t_j,f_i) - \frac{1}{\sqrt{s}} \Phi(\Theta \bigodot D(t_j,f_i)) X(f_i) \right)}$$

and

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

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 [1]) and only the prior on Θ and s remain to be given. We give these two priors now:

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

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

• How can we assign a prior probability $P(s \mid M_{\text{doa}}, I)$ 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, one could fit the data 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 $P(s \mid M_{\text{doa}}, I) = \frac{1}{N\times N_1}$. 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.

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

$$\begin{aligned} P(Y \mid M_{\text{doa}}, I) = \int & \frac{1}{(2\pi\sigma^2)^{\frac{N_1 N r}{2}}} \\ & e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{N_1}\left(Y(t_j,f_i)-\frac{1}{\sqrt{s}}\Phi(\Theta\odot D(t_j,f_i))X(f_i)\right)^H\left(Y(t_j,f_i)-\frac{1}{\sqrt{s}}\Phi(\Theta\odot D(t_j,f_i))X(f_i)\right)} \\ & P(\Phi \mid s, M_{\text{doa}}, I)P(s \mid M_{\text{doa}}, I)P(v_r \mid M_{\text{doa}}, I)P(\alpha_r \mid M_{\text{doa}}, I)P(\beta^r \mid M_{\text{doa}}, I) \\ & P(\tau \mid M_{\text{doa}}, I)P(\Theta \mid M_{\text{doa}}, I)d\Phi d\Theta ds d\tau dv_r d\alpha_r d\beta^{\mathbf{r}} \end{aligned}$$

which gives:

$$
\begin{aligned}
P(Y \mid M_{\text{doa}}, I) &= \frac{1}{N \times N_1} \sum_{s=1}^{N \times N_1} \int_0^{2\pi} \int_0^{\infty} \int_0^{v_{\lim}} \int_0^{\tau_{\max}} \frac{1}{(2\pi\sigma^2)^{\frac{N_1 N r}{2}}} \prod_{i=1}^{N} \prod_{j=1}^{N_1} \\
& e^{-\frac{1}{2\sigma^2}\left(Y(t_j,f_i)-\frac{1}{\sqrt{s}}\Phi(\Theta \bigodot D(t_j,f_i))X(f_i)\right)^H \left(Y(t_j,f_i)-\frac{1}{\sqrt{s}}\Phi(\Theta \bigodot D(t_j,f_i))X(f_i)\right)} \\
& (\frac{1}{\tau_{\max}})^{s \times t} \frac{1}{v_{\lim}} (\frac{1}{2\pi})^{r \times s} \frac{1}{2\pi} (\frac{1}{2\pi})^{s} \\
& d\phi_{11} ... d\phi_{rs} d\theta_{11} ... d\theta_{st} d\tau_{11} ... d\tau_{st} dv_r d\alpha_r d\beta^r{}_1 ... d\beta^r{}_s
\end{aligned} \tag{6}
$$

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_{double}:

Let us now derive model M_{double}:

$$
H(f,t) = \frac{1}{\sqrt{s s_1}} \Phi \left(\Theta \bigodot D(f,t) \right) \Psi
$$

with

$$
D(f,t) = \begin{pmatrix}
e^{-j2\pi f \tau_{1,1}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r \mathbf{v}_r)} & \dots & e^{-j2\pi f \tau_{1,s_1}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r \mathbf{v}_r)} \\
\vdots & \ddots & \vdots \\
e^{-j2\pi f \tau_{s,1}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_s^r \mathbf{v}_r)} & \dots & e^{-j2\pi f \tau_{s,s_1}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_s^r \mathbf{v}_r)}
\end{pmatrix}
$$

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

$$
p_{\text{double}} = \{\Phi, s, \Psi, s_1, \tau, v_r, , \alpha_r, \beta_{\mathbf{r}} \Theta\} = \{p_{\text{doa}}, \Phi, s_1\}
$$

by adding two new parameters Ψ and s_1 and going to the new subspace $S_{p_{\text{double}}}$ in such a way that $\Psi = F_{n_t}$ $(n_t = s_1)$ represents model M_{doa}. Indeed, in this case, we have:

$$
\begin{aligned}
\left(\Theta \bigodot D(f,t) \right) F_{n_t} &= \begin{pmatrix}
\sum_{i=1}^{n_t} \theta_{1i} e^{-j2\pi f \tau_{1,i}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r \mathbf{v}_r)} & \dots & \sum_{i=1}^{n_t} \theta_{1i} e^{-j2\pi f \tau_{1,i}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_1^r \mathbf{v}_r)} e^{j2\pi \frac{(n_t-1)i}{n_t}} \\
\vdots & \ddots & \vdots \\
\sum_{i=1}^{n_t} \theta_{si} e^{-j2\pi f \tau_{s,i}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_s^r \mathbf{v}_r)} & \dots & \sum_{i=1}^{n_t} \theta_{si} e^{-j2\pi f \tau_{s,i}} e^{j2\pi \frac{ft}{c}(\mathbf{u}_s^r \mathbf{v}_r)} e^{j2\pi \frac{(n_t-1)i}{n_t}}
\end{pmatrix} \\
&= \begin{pmatrix}
\sum_{i=1}^{n_t} \theta_{1i} e^{-j2\pi f (\tau_{1,i}-\tau_{1,1})} & \dots & \sum_{i=1}^{n_t} \theta_{1i} e^{-j2\pi f (\tau_{1,i}-\tau_{1,n_t})} e^{j2\pi \frac{(n_t-1)i}{n_t}} \\
\vdots & \ddots & \vdots \\
\sum_{i=1}^{n_t} \theta_{si} e^{-j2\pi f (\tau_{s,i}-\tau_{s,1})} & \dots & \sum_{i=1}^{n_t} \theta_{si} e^{-j2\pi f (\tau_{s,i}-\tau_{s,n_t})} e^{j2\pi \frac{(n_t-1)i}{n_t}}
\end{pmatrix} \bigodot D(f,t) \\
&= \Theta_1 \bigodot D(f,t)
\end{aligned}
$$

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

We recall here the model for a given frequency:

$$Y(f,t) = \frac{1}{\sqrt{ss_1}}\Phi\left(\Theta \bigodot D(f,t)\right)\Psi X(f) + N(f)$$

The same methodology applies and we have:

$$\begin{aligned} P(Y \mid M_{\text{double}}, I) \quad = \quad & \int \frac{1}{(2\pi\sigma^2)^{\frac{N_1 N_r}{2}}} \\ & e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{N_1}\left(Y(t_j,f_i)-\frac{1}{\sqrt{ss_1}}\Phi(\Theta\bigodot D(t_j,f_i))\Psi X(f_i)\right)^H\left(Y(t_j,f_i)-\frac{1}{\sqrt{ss_1}}\Phi(\Theta\bigodot D(t_j,f_i))\Psi X(f_i)\right)} \\ & P(\Phi \mid s, M_{\text{double}}, I)P(s \mid M_{\text{double}}, I)P(\Psi \mid s_1, M_{\text{double}}, I)P(s_1 \mid M_{\text{double}}, I) \\ & P(v_r \mid M_{\text{double}}, I)P(\alpha_r \mid M_{\text{double}}, I)P(\beta_{\mathbf{r}} \mid M_{\text{double}}, I)P(\tau \mid M_{\text{double}}, I) \\ & P(\Theta \mid M_{\text{double}}, I)d\Phi d\Psi d\Theta ds ds_1 d\tau dv_r d\alpha_r d\beta^{\mathbf{r}} \end{aligned}$$

and

$$\begin{aligned} P(Y \mid M_{\text{double}}, I) \quad = \quad & (\frac{2}{N\times N_1})^2 \sum_{s=1}^{\frac{N\times N_1}{2}}\sum_{s_1=1}^{\frac{N\times N_1}{2}}\int_0^{2\pi}\int_0^{\infty}\int_0^{v_{\lim}}\int_0^{\tau_{\max}} \frac{1}{(2\pi\sigma^2)^{\frac{N_1 N_r}{2}}}\prod_{i=1}^{N}\prod_{j=1}^{N_1} \\ & e^{-\frac{1}{2\sigma^2}\left(Y(t_j,f_i)-\frac{1}{\sqrt{ss_1}}\Phi(\Theta\bigodot D(t_j,f_i))\Psi X(f_i)\right)^H\left(Y(t_j,f_i)-\frac{1}{\sqrt{ss_1}}\Phi(\Theta\bigodot D(t_j,f_i))\Psi X(f_i)\right)} \\ & (\frac{1}{\tau_{\max}})^{s\times s_1}\frac{1}{v_{\lim}}(\frac{1}{2\pi})^{n_r\times s}(\frac{1}{2\pi})^{s_1\times n_t}\frac{1}{2\pi}(\frac{1}{2\pi})^{s} \\ & d\phi_{11}...d\phi_{n_r s}d\psi_{11}...d\psi_{s_{1n_t}}d\theta_{11}...d\theta_{sn_t}d\tau_{11}...d\tau_{sn_t}dv_r d\alpha_r d\beta^r{}_1...d\beta^r{}_s \end{aligned} \quad (7)$$

Although tedious, all the framework is therefore provided for ranking the two models and it can be easily shown (see [2] and [14]) that Bayesian probability will always choose the model with the least parameters when both models have the same maximum likelihood.

References

[1] M. Debbah and R. Müller, "MIMO Channel Modelling and the Principle of Maximum Entropy, Part I: Model Construction," *submitted to IEEE Transactions on Information Theory*, can be downloaded at http://www.eurecom.fr/∼debbah, 2003.

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

[3] 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.

[4] J. Capon, "High-Resolution Frequency Wavenumber Spectrum Analysis," *Proceedings of the IEEE*, vol. 8, no. 57, pp. 1408–1418, 1969.

[5] R.O Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *Proceedings of RADC: Spectral Estimation Workshop, Rome*, pp. 243–258, 1979.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

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

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

[14] M. Debbah, "MIMO Channel Modelling and the Principle of Maximum Entropy: An Information Theoretic Point of View ," Flows Technical report 2003, can be downloaded at http://www.flows-ist.org.