Wireless Communications and Propagation Aspects

Mérouane Debbah
Alcatel-Lucent Chair on Flexible Radio
SUPELEC
Plateau de Moulon
3 rue Joliot-Curie
B.P. 91192
GIF SUR YVETTE CEDEX, France
merouane.debbah@supelec.fr.

ABSTRACT. A recent article in IEEE spectrum dedicated to wireless technologies provided some figures recalled in the following table.

Technology	Data Rate (Mb/s)	Output Power (mW)	Range (meters)	Frequency Band
Bluetooth	1-2	100	100	2.4 GHz
IrDa	4	100 mW/sr	1-2	Infrared
Ultrawideband	100-500	1	10	3.1-10.6 GHz
IEEE802.11a	54	40-800	20	5GHz
IEEE802.11b (Wi-Fi)	11	200	100	2.4 GHz
UMTS	2Mb/s	250	more than 100 meters	2 GHz

Table 1. Comparing Wireless Technologies.

A closer look at the numbers shows how diverse the wireless link is going to be. Why are there so many technologies for the same goal: the transmission of information through a wireless medium? Is there any relationship between data rate, output power, range and frequency band? Which technology to choose? This course is intended to give some partial answers to the previous questions by presenting several technologies, their advantages as well as their drawbacks.
In these notes, we will focus on channel modeling. Two approaches are provided. In part 1, the SISO (Single Input Single Output) time varying channel response is modeled as the sample path of a random process, whose statistics depend on the underlying physical model. In part 2, we use an information theoretic approach. The information provided by the physical environment to model the MIMO (Multiple Input Multiple Ouput) link. In this case, a general procedure to translate information into probability assignement is given through the use of the principle of maximum entropy.

Debbah, M. (2007) Wireless Communications and Propagation Aspects. In *Emerging Wireless Technologies* (pp. 4-1 – 4-34). Educational Notes RTO-EN-IST-070, Paper 4. Neuilly-sur-Seine, France: RTO. Available from: http://www.rto.nato.int.

1 SISO channel modeling

This chapter presents an introduction to the mathematical modeling of time-varying linear **S**ingle **I**nput **S**ingle **O**utput (SISO) channels typical of mobile wireless communications, known as *fading channels*. The material is mainly taken from [1] and from [2].

1.1 Doppler effect and deterministic fading

Consider the situation of Fig. 1, where there exists a single line-of-sight (LOS) propagation path between the transmitter and the receiver, but where the position of the receiver relative to the transmitter changes in time. Let the distance between transmitter and receiver at time t be given by

$$d_0(t) = -v_0 t + d_0$$

(this is equivalent to approximate a general time-varying distance $d_0(t)$ with its first-order Taylor expansion). Let the transmitted (bandpass) signal be

$$s(t) = \mathrm{Re}\{x(t)\exp(j2\pi f_c t)\}$$

where $x(t)$ is the complex envelope and f_c is the carrier frequency. In the absence of other impairments, the received signal is just a delayed version of the transmitted signal, where the propagation delay is given by

$$\tau_0(t) = d_0(t)/c = -(v_0/c)t + d_0/c$$

and where c denotes the speed of light. In terrestrial wireless applications, v_0/c is very small. However, the term $\xi_0 = f_c v_0/c$ can be non-negligible since the carrier frequency f_c is normally large. For example, in GSM we have $f_c \approx 900$ MHz. Then, a mobile travelling at $v_0 = 100$ km/h yields $\xi_0 = 83.33$ Hz. The received signal can be written as

$$\begin{aligned} r(t) &= \mathrm{Re}\{x(t-\tau_0(t))\exp(j2\pi f_c(t-\tau_0(t))\} \\ &\approx \mathrm{Re}\{x(t-\tau_0)\exp(j2\pi(f_c+\xi_0)t)\exp(j\phi_0)\} \end{aligned} \quad (1)$$

where we let $\phi_0 = -2\pi f_c d_0/c$ and $\tau_0 = d_0/c$, and we have used the fact that, since v_0/c is small, $x(t-\tau_0(t)) \approx x(t-\tau_0)$ over the (relatively short) observation interval. From (1) we observe that the time-varying propagation delay causes a shift of the carrier frequency by ξ_0 (called *Doppler effect*) and a carrier phase shift by ϕ_0.

Next, consider the situation of Fig. 2, where in addition to the LOS path, there is a second propagation path due to a reflecting object (e.g., a building or a hill). Let $d_i(t) = -v_i t + d_i$, for $i = 0, 1$, be the propagation distances, where v_0 and v_1 denote the relative speed of the receiver with respect to the transmitter and to the reflector, respectively (in this case, $v_1 = -v_0 \cos\alpha$). By defining $\xi_i = f_c v_i/c$, $\phi_i = -2\pi f_c d_i/c$ and $\tau_i = d_i/c$, we obtain the received signal as

$$\begin{aligned} r(t) &\approx \mathrm{Re}\{x(t-\tau_0)\exp(j\phi_0)\exp(j2\pi(f_c+\xi_0)t) + ax(t-\tau_1)\exp(j\phi_1)\exp(j2\pi(f_c+\xi_1)t)\} \\ &= \mathrm{Re}\left\{\left(\sum_{i=0}^{1}\rho_i x(t-\tau_i)\exp(j2\pi\xi_i t)\right)\exp(j2\pi f_c t)\right) \end{aligned} \quad (2)$$

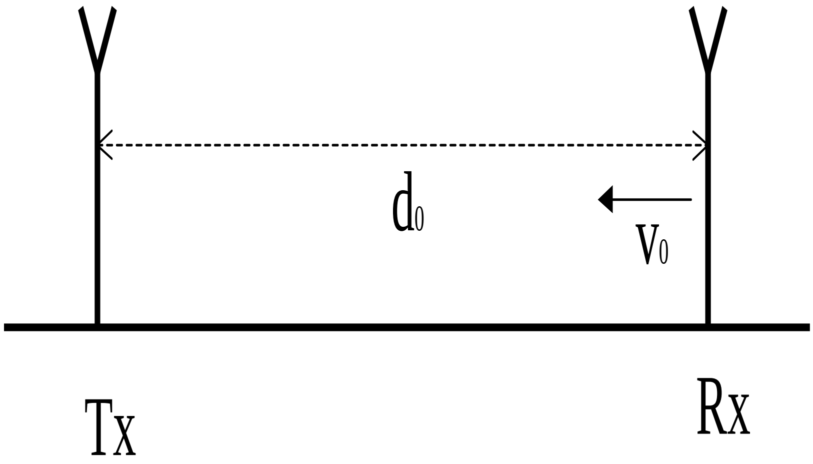

Figure 1: Line of sight propagation with a mobile terminal.

where a is a complex attenuation due to reflection, and we have defined the complex coefficients $\rho_0 = \exp(j\phi_0)$ and $\rho_1 = a\exp(j\phi_1)$. The complex envelope of the received signal (2) is

$$y(t) = \sum_{i=0}^{1} \rho_i x(t-\tau_i)\exp(j2\pi\xi_i t)$$

Suppose that $x(t)$ varies slowly, so that $x(t) \approx x$, constant over the observation interval. Nevertheless, the envelope of the received signal may be subject to time-variations. In fact, we have

$$\begin{aligned} |y(t)| &= |x\rho_0| \left| 1 + \frac{\rho_1}{\rho_0}\exp(j2\pi(\xi_1-\xi_0)t) \right| \\ &= A\sqrt{1+b^2+2b\cos(2\pi\Delta\xi t + \Delta\phi)} \end{aligned} \tag{3}$$

where we let $A = |x\rho_0|$, $\Delta\xi = \xi_1 - \xi_0$ and $\rho_1/\rho_0 = b\exp(j\Delta\phi)$, with $b \in \mathbb{R}_+$. From (3) we see that if the observation interval is larger than $1/\Delta\xi$, the received signal envelope changes between a maximum value $A|1+b|$ to a minimum value $A|1-b|$. The rate of variation is given by the Doppler frequency *spread* $\Delta\xi$, given by the difference between the maximum and the minimum Doppler frequency shifts. The Doppler frequency *spread* characterizes the time selectivity of the channel.

In conclusion, we have seen that with more than one propagation path the receiver signal envelope changes in time even if the transmitted signal envelope is kept constant. This effect is called *fading*, and channels characterized by this type of time-variations are called *fading channels*.

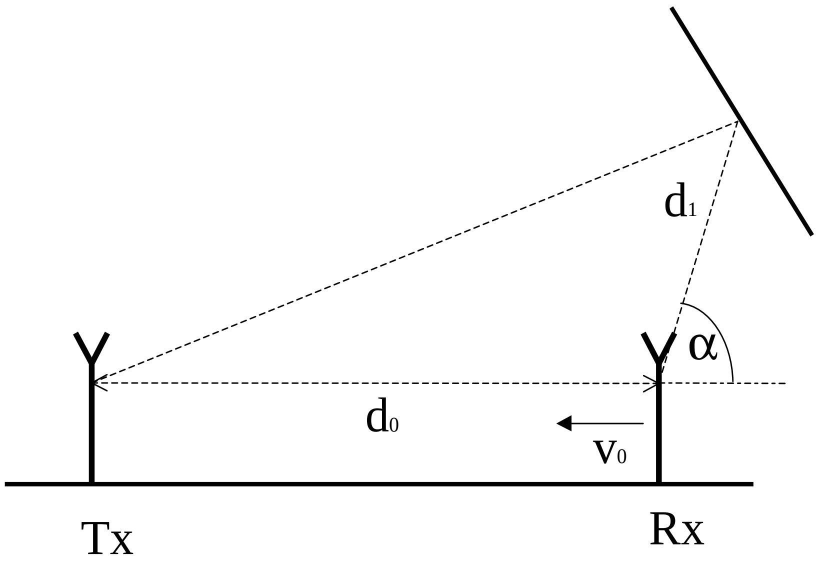

Figure 2: Two-ray propagation with a mobile terminal.

1.2 Wide-sense stationary uncorrelated scattering model

In the previous section we considered the simple case of two paths, whose characteristic is perfectly known. In this case, the time-varying channel is deterministic. However, in a real wireless communication situation, propagation may go through an unknown number of paths, each of which has unknown characteristics. Moreover, the number of paths is normally very large. Therefore, it is convenient to describe the resulting fading channel by a *statistical* model. In other words, the time-varying channel response is modeled as the sample path of a random process, whose statistics depend on the underlying physical model.

Consider the situation of Fig. 3, where propagation goes through a large number of paths, each due to a *scattering element* located in some scattering region. In analogy with what was done before, we can write the complex envelope of the received signal as

$$y(t) = \sum_i \rho_i x(t - \tau_i) \exp(j2\pi\xi_i t) \tag{4}$$

where the index i runs over the scattering elements and where the scattering element i is characterized by the complex coefficient ρ_i and by the Doppler shift ξ_i. When the number of scattering elements is large, determining each ρ_i is either impossible or impractical. Moreover, from an engineering point of view it is more interesting to consider classes of channels with some common features, and describe them in terms of their statistics. Thus, the ρ_i's are modelled as random variables, with some joint distribution. A common simplifying assumption, followed in most literature on the field and supported by physical arguments and experimental evidence [3, 2], is that the scattering coefficients ρ_i are zero-mean pairwise uncorrelated, i.e., $E[\rho_i \rho_j^*] = 0$ for all

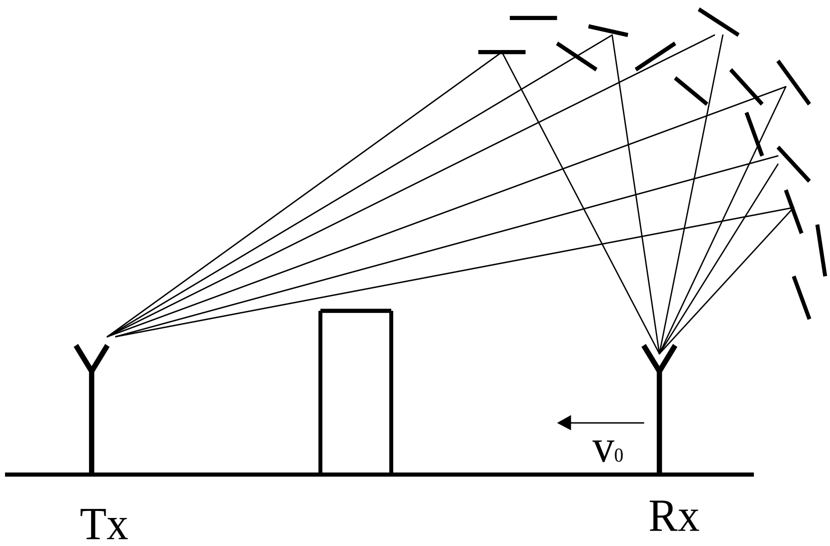

Figure 3: Scattering from many elements, without line of sight propagation.

$i \neq j$. This is referred to as the *Uncorrelated Scattering* (US) assumption, and from now on we restrict to this case.

We start with considering the following particular case, very important in applications. [1]

Discrete multipath channel. Assume that the scattering elements are clustered, so that there exist P clusters. The paths in the same cluster have similar propagation delay. Let τ_p be the delay of the p-th cluster and $\mathcal{P}_p$ the set of indices of scattering elements of the p-th cluster. Then, (4) can be written as

$$\begin{aligned} y(t) &= \sum_{p=0}^{P-1} \left(\sum_{i \in \mathcal{P}_p} \rho_i \exp(j 2\pi \xi_i t) \right) x(t - \tau_p) \\ &= \sum_{p=0}^{P-1} c_p(t) x(t - \tau_p) \end{aligned} \tag{5}$$

This can be interpreted as the output of the complex baseband equivalent time-varying linear channel with impulse response

$$c(t, \tau) = \sum_{p=0}^{P-1} c_p(t) \delta(\tau - \tau_p) \tag{6}$$

[1] All channel models specified by ETSI as standard test models for GSM belong to this class [4].

and

$$C(t,f) = \sum_{p=0}^{P-1} c_p(t) e^{-j2\pi f \tau_p} \tag{7}$$

where $c_p(t) = \sum_{i \in \mathcal{P}_p} \rho_i \exp(j2\pi\xi_i t)$. The above channel can be represented as a *tapped delay-line* with time-varying coefficients $c_p(t)$ and non-uniformly spaced delays τ_p. Each process $c_p(t)$ is given by the superposition of several complex sinusoids at different frequencies $\{\xi_i : i \in \mathcal{P}_p\}$, with different amplitudes and (uniformly distributed) phases, due to the random coefficients $\{\rho_i : i \in \mathcal{P}_p\}$. If each cluster is made of a very large number of scattering elements, each giving a very small contribution to the received signal, we can safely use the Central Limit Theorem [5] and conclude that the $c_p(t)$'s are jointly circularly symmetric complex Gaussian processes with mean zero. Since the clusters $\mathcal{P}_p$ are disjoint, because of the US assumption we have $E[c_p(t)c_q(t')^*] = 0$ for all $p \neq q$ and t, t'. Therefore, the path gains $c_p(t)$ are statistically independent[2]. $\diamondsuit$

Diffuse scattering. Next, we consider the more general case of *diffuse* scattering, where the scattering elements may be arbitrarily distributed over a certain delay interval $\mathcal{T}$, and over a certain Doppler frequency interval $\mathcal{D}$. Let $\rho(\xi,\tau)$ for $\xi \in \mathcal{D}$ and $\tau \in \mathcal{T}$ be the complex coefficient due to an elementary scattering element at Doppler shift ξ and delay τ. We write (4) in the following integral form

$$\begin{aligned} y(t) &= \int_{\mathcal{D}} \int_{\mathcal{T}} \rho(\xi,\tau) x(t-\tau) \exp(j2\pi\xi t) d\xi d\tau \\ &= \int_{\mathcal{T}} c(t,\tau) x(t-\tau) d\tau \end{aligned} \tag{8}$$

where we define the channel impulse response

$$c(t,\tau) = \int_{\mathcal{D}} \rho(\xi,\tau) \exp(j2\pi\xi t) d\xi \tag{9}$$

In words, $c(t,\tau)$ is the response of the channel at time t to an impulse at time $t-\tau$. Notice that $c(t,\tau)$ is obtained from $\rho(\xi,\tau)$ by inverse Fourier transform with respect to the Doppler frequency variable ξ. The fact that the fading channel is a linear time-varying system is an obvious consequence of the fact that we expressed the channel output as the superposition of weighted and delayed replicas of the input, where the weighting coefficients $\rho(\xi,\tau)\exp(j2\pi\xi t)$ depend on t.

Usually, $\rho(\xi,\tau)$ is assumed to be a zero-mean complex circularly symmetric independent Gaussian random field, even though no particular clustering argument is advocated. This assumption is motivated by some experimental evidence (see [3] and references therein) and will be followed here. Since $c(t,\tau)$ is obtained from $\rho(\xi,\tau)$ by a linear transformation, it is also Gaussian with circular symmetry [6], although correlated. It follows that the channel is fully characterized by the second order statistics of $\rho(\xi,\tau)$ or, equivalently, of $c(t,\tau)$. This is given next, in terms of some functions that, because of their importance, have special names.

[2]Note that the result was shown for the single bounce model. It can be shown that the i.i.d. Gaussian model for the path gains $c_p(t)$ persists in other cases if one uses maximum entropy arguments (see chapter 2)

Scattering function. From the US assumption, the autocorrelation between scattering coefficients $\rho(\xi,\tau)$ and $\rho(\xi',\tau')$ is given by

$$E[\rho(\xi,\tau)\rho(\xi',\tau')^*] = \sigma(\xi,\tau)\delta(\xi-\xi')\delta(\tau-\tau')$$

where we have defined the function $\sigma(\xi,\tau) = E[|\rho(\xi,\tau)|^2]$. This is called *scattering function* and describes the average received power from scattering at delay τ and Doppler shift ξ.

Time-delay autocorrelation function. Consider the autocorrelation function $E[c(t,\tau)c(t',\tau')^*]$ of the channel impulse response. By using (9), this can be written as

$$\begin{aligned} E[c(t,\tau)c(t',\tau')^*] &= E\left[\int\int \rho(\xi,\tau)\rho(\xi',\tau')^* \exp(j2\pi\xi t)\exp(-j2\pi\xi' t')d\xi d\xi'\right] \\ &= \int\int \sigma(\xi,\tau)\delta(\xi-\xi')\delta(\tau-\tau')\exp(j2\pi\xi t)\exp(-j2\pi\xi' t')d\xi d\xi' \\ &= \int \sigma(\xi,\tau)\exp(j2\pi\xi(t-t'))d\xi\delta(\tau-\tau') \\ &= \phi(t-t',\tau)\delta(\tau-\tau') \end{aligned}$$

where the function $\phi(\Delta t,\tau) = \int \sigma(\xi,\tau)\exp(j2\pi\xi\Delta t)d\xi$ is the autocorrelation function of the channel impulse response, seen as a random process with respect to the variable t, at delay τ. It is important to notice that $c(t,\tau)$ is wide-sense stationary (WSS) with respect to t, for every τ, since $E[c(t,\tau)c(t',\tau)^*]$ depends only on the difference $t-t'$ and not individually on t and t'. For this reason, this channel model belongs to the more general class of wide-sense stationary US (WSSUS) channels, i.e. time-varying channels whose impulse response $c(t,\tau)$ is WSS with respect to t and uncorrelated with respect to τ.

Time-frequency autocorrelation function. Let $C(t,f) = \int c(t,\tau)\exp(-j2\pi f\tau)d\tau$[3] be the time-varying transfer function of the channel with impulse response $c(t,\tau)$. The autocorrelation function $E[C(t,f)C(t',f')^*]$ is given by

$$\begin{aligned} E[C(t,f)C(t',f')^*] &= E\left[\int\int c(t,\tau)c(t',\tau')^* \exp(-j2\pi f\tau)\exp(j2\pi f'\tau')d\tau d\tau'\right] \\ &= \int\int \phi(t-t',\tau)\delta(\tau-\tau')\exp(-j2\pi f\tau)\exp(j2\pi f'\tau')d\tau d\tau' \\ &= \int \phi(t-t',\tau)\exp(-j2\pi(f-f')\tau)d\tau \\ &= \Phi(t-t',f-f') \end{aligned}$$

where the function $\Phi(\Delta t,\Delta f) = \int \phi(\Delta t,\tau)\exp(-j2\pi\Delta f\tau)d\tau$ is the autocorrelation of the channel transfer function at frequencies separated by Δf and time separated by Δt. From the definition of $\phi(\Delta t,\tau)$ we get the following two-dimensional Fourier transform relation between the time-frequency autocorrelation and the scattering function

$$\Phi(\Delta t,\Delta f) = \int\int \sigma(\xi,\tau)\exp(j2\pi\xi\Delta t)\exp(-j2\pi\Delta f\tau)d\xi d\tau \tag{10}$$

[3]Note that $C(t,f) = \int c(t,\tau)\exp(-j2\pi f\tau)d\tau$ and $\rho(\xi,\tau) = \int c(t,\tau)\exp(-j2\pi\xi t)dt$.

Delay-intensity profile and Doppler spectrum. It is customary to describe fading channels in terms of two one-dimensional functions derived from the scattering function: the delay-intensity profile and the Doppler spectrum. The delay intensity profile is defined by

$$P(\tau) = \int \sigma(\xi, \tau) d\xi = \phi(0, \tau) \tag{11}$$

and describes the total average scattering power at delay τ. The support of $P(\tau)$ is the range $\mathcal{T}$ of the channel delays. Roughly speaking, the "size" of $\mathcal{T}$ is a measure of the *delay spread* T_d of the channel. Several definitions of delay spread are possible. A widely accepted definition is the root-mean-square (RMS) delay spread

$$T_d = \left(\frac{\int_{\mathcal{T}} (\tau - \bar{\tau})^2 P(\tau) d\tau}{\int_{\mathcal{T}} P(\tau) d\tau} \right)^{1/2} \tag{12}$$

where $\bar{\tau} = \int_{\mathcal{T}} \tau P(\tau) d\tau / \int_{\mathcal{T}} P(\tau) d\tau$ is the mean channel delay. The Doppler spectrum is defined by

$$D(\xi) = \int \sigma(\xi, \tau) d\tau = \int \Phi(\Delta t, 0) \exp(-j 2\pi \xi \Delta t) d\Delta t \tag{13}$$

and describes the total average scattering power at Doppler shift ξ. The support of $D(\xi)$ is the range $\mathcal{D}$ of possible Doppler frequency shifts. Roughly speaking, the "size" of $\mathcal{D}$ is a measure of the *Doppler spread* B_d of the channel. Several definitions of Doppler spread are possible (at least as many as of bandwidth). A widely accepted definition is the RMS Doppler spread

$$B_d = \left(\frac{\int_{\mathcal{D}} (\xi - \bar{\xi})^2 D(\xi) d\xi}{\int_{\mathcal{D}} D(\xi) d\xi} \right)^{1/2} \tag{14}$$

where $\bar{\xi} = \int_{\mathcal{D}} \xi D(\xi) d\xi / \int_{\mathcal{D}} D(\xi) d\xi$ is the mean Doppler spread, that is normally equal to zero, since a non-zero average Doppler spread would correspond just to a deterministic offset of the carrier frequency, that can be compensated at the receiver.

Separable fading channels. A simplifying assumption, often used in practice, is that the scattering function is separable, i.e. that $\sigma(\xi, \tau) = D(\xi) P(\tau)$ (without loss of generality, we consider a normalized channel where $\int D(\xi) d\xi = \int P(\tau) d\tau = \int \int \sigma(\xi, \tau) d\xi d\tau = 1$). This is motivated by the fact that $P(\tau)$ depends mainly on the spatial distribution of the scattering elements, while $D(\xi)$ depends mainly on the relative motion between the receiver and the scattering elements. Several reference channel models specified in international telecommunications standards, given in terms of $P(\tau)$ and $D(\xi)$ assume implicitly the separability property.

1.3 Doppler spectra calculation

In this section we focus on the derivation of the Doppler spectrum $D(\xi)$ from the physical characteristics of the environment and of the receiver antenna [2, 7]. We focus on a single delay τ. If the channel is separable, the Doppler spectrum calculated for τ is the actual channel Doppler spectrum. Otherwise, we should repeat the calculation for every τ, obtaining different sections of the scattering function, and integrate over all τ.

In (4), consider only the delays $\tau_i = \tau$. We have

$$y(t) = \sum_i \rho_i x(t-\tau)\exp(j2\pi\xi_i t) = c(t)x(t-\tau)$$

where $c(t) = \sum_i \rho_i \exp(j2\pi\xi_i t)$. By definition, the Doppler spectrum (for delay τ) is the power spectral density of the process $c(t)$. Now, consider the planar situation of Fig. 4. Let α denote the azimuth angle, $g(\alpha)$ be the receiver antenna azimuthal gain and $p(\alpha)$ be the average received power from angle α. With the normalization $\int_{-\pi}^{\pi} p(\alpha)d\alpha = 1$, $p(\alpha)$ can be seen as the probability of receiving a scattered signal component from direction α. From what seen in the previous section about Doppler effect, we can express the Doppler frequency shift ξ caused by a scattering element at angle α as

$$\xi = F_d \cos\alpha$$

where $F_d = f_c v/c$ is the maximum Doppler shift and v is the terminal speed (see Fig. 4). Then, instead of summing over the scattering index i, we can integrate over the angle α and obtain

$$c(t) = \int_{-\pi}^{\pi} \rho_\alpha \exp(j2\pi F_d \cos\alpha t)d\alpha$$

where with some abuse of notation we let $\rho_\alpha d\alpha$ be the received signal component at angle α. The autocorrelation function of $c(t)$ is given by

$$\begin{aligned} r_c(\Delta t) &= E\left[\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \rho_\alpha \rho_\beta^* \exp(j2\pi F_d \cos\alpha t)\exp(-j2\pi F_d \cos\beta(t-\Delta t))d\alpha d\beta\right] \\ &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} p(\alpha)g(\alpha)\delta(\alpha-\beta)\exp(j2\pi F_d\cos\alpha t)\exp(-j2\pi F_d\cos\beta(t-\Delta t))d\alpha d\beta \\ &= \int_{-\pi}^{\pi} p(\alpha)g(\alpha)\exp(j2\pi F_d \cos\alpha\Delta t)d\alpha \end{aligned} \quad (15)$$

where we have used the US assumption, so that $E[\rho_\alpha\rho_\beta^* d\alpha d\beta] = p(\alpha)g(\alpha)\delta(\alpha-\beta)d\alpha d\beta$. Finally, the desired Doppler spectrum is obtained, by definition, via the Fourier transform $D(\xi) = \int r_c(\Delta t)\exp(-j2\pi\xi\Delta t)d\Delta t$.

Example: Jakes' Doppler spectrum. Assuming scattered signals arriving from any direction with equal probability and an omnidirectional receiver antenna, we get $p(\alpha) = 1/(2\pi)$ and $g(\alpha) = 1$. Using these in (15) we obtain

$$r_c(\Delta t) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\exp(j2\pi F_d\cos\alpha\Delta t)d\alpha = J_0(2\pi F_d\Delta t)$$

where $J_0(x)$ is the 0-th order Bessel function of the first kind [8]. The above autocorrelation function is by far the most used and widely accepted autocorrelation model for the mobile channel, and it is known as Jakes' model [7]. The corresponding Doppler spectrum is given by

$$D(\xi) = \begin{cases} \frac{1}{\pi F_d\sqrt{1-\xi^2/F_d^2}} & |\xi| < F_d \\ 0 & \text{elsewhere} \end{cases}$$

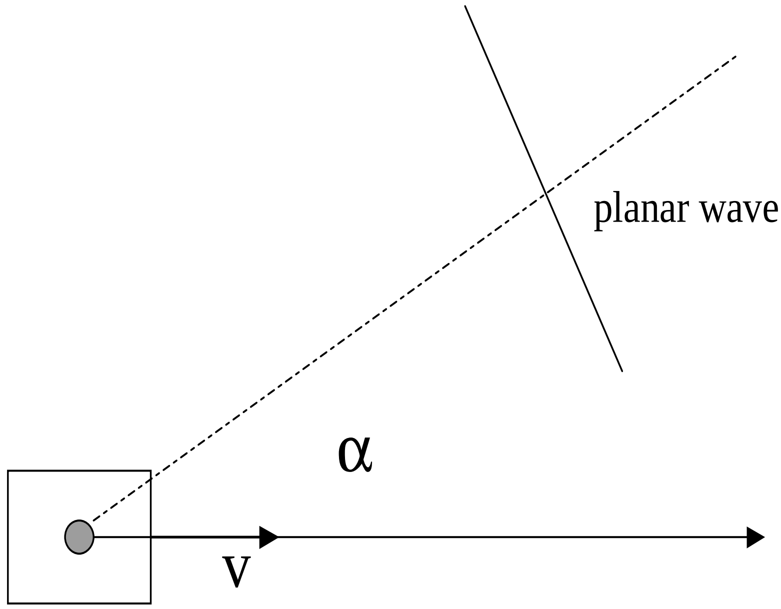

Figure 4: Planar geometry for Doppler spectra calculation.

1.4 Qualitative behavior and simplified models

In this section we discuss some qualitative characteristics of fading channels. These describe the effect of the channel on the transmitted signal and are given in terms of delay spread and Doppler spread relative to the signal duration and bandwidth. Strictly speaking, finite duration signals cannot be strictly bandlimited and vice versa. However, since the discussion carried out in this section is qualitative, we assume that the transmitted passband signal $s(t)$ is mostly concentrated over an interval of duration T and over a bandwidth W around the carrier frequency f_c. Then, its complex envelope $x(t)$ has a spectrum located around the origin, mostly concentrated in the interval $[-W/2, W/2]$. To be consistent with the definitions of RMS delay spread T_d and RMS Doppler spread B_d given previously, we define the RMS signal duration T as

$$T = \left(\frac{\int (t-\bar{t})^2 |x(t)|^2 dt}{\int |x(t)|^2 dt} \right)^{1/2}$$

where $\bar{t} = \int t|x(t)|^2 dt / \int |x(t)|^2 dt$, and the RMS signal bandwidth W as

$$W = \left(\frac{\int (f-\bar{f})^2 |X(f)|^2 df}{\int |X(f)|^2 df} \right)^{1/2}$$

where $X(f)$ is the Fourier transform of $x(t)$ and where $\bar{f} = \int f|X(f)|^2 df / \int |X(f)|^2 df$.

Next, we classify fading channels according to their time-frequency spreading effect and multiplicative distortion on the transmitted signal. Each class yields a simplified channel model, valid within given assumptions. These assumptions involve also the signal bandwidth and duration, so that they are relative with respect to the transmitted signal. The same physical channel may fall in different classes and be modeled in different ways according to the type of signals employed.

If $T_d \ll T$, the duration of the output signal $y(t)$ is about $T + T_d \approx T$, i.e. the channel does not "spread" the signal in time. On the contrary, if this condition is not satisfied, the channel is said to be *dispersive* in time. If $B_d \ll W$, the bandwidth of $y(t)$ is about $W + B_d \approx W$, i.e. the channel does not "spread" the signal in frequency. On the contrary, if this condition is not satisfied, the channel is said to be dispersive in frequency. In almost all terrestrial wireless applications, B_d ranges between 0 and some hundreds of Hz, while W ranges from some tens of kHz to some MHz. For this reason, we shall restrict our discussion to frequency non-dispersive channels. [4]

The time interval Δt_c beyond which the channel decorrelates in time (i.e. for which $|\phi(\Delta t, \tau)| \simeq 0$ for $\Delta t \geq \Delta t_c$) is called *coherence time* and it is roughly given by $\Delta t_c \approx 1/B_d$. For $T \ll 1/B_d$, the channel appears as random but time-invariant over the signal duration. Otherwise, the channel time variations cause multiplicative distortion and the channel is said to be *time-selective*. The frequency interval Δf_c beyond which the channel decorrelates in frequency (i.e., for which $|\Phi(0, \Delta f)| \simeq 0$ for $\Delta f \geq \Delta f_c$) is called *coherence bandwidth* and it is roughly given by $\Delta f_c \approx 1/T_d$. For $W \ll 1/T_d$, the channel appears as random but frequency-flat over the signal bandwidth. Otherwise, the channel frequency variations cause linear distortion and the channel is said to be *frequency-selective*.

Time non-dispersive fading: memoryless channel. If $T_d \ll T$, a train of pulses of duration approximately T will not suffer from intersymbol interference (ISI), since after convolution with $c(t, \tau)$ the received pulses overlap on intervals of duration T_d that is negligible with respect to the pulse duration T. This yields a simplified memoryless channel model, where the output over an interval of duration T (conditioned with respect to the channel impulse response) depends only on the input over the same interval, irrespectively of the past and without affecting the future.

If $W \approx 1/T_d$, the channel is frequency-selective. This condition, together with $T_d \ll T$, implies that $W \gg 1/T$. Therefore, frequency-selectivity with time non-dispersive channels may occur only for *spread signals* (i.e., those signals for which $W \gg 1/T$).

Time dispersive fading: ISI channel. If T_d is not negligible with respect to T, a train of pulses of duration T will suffer from ISI, since after convolution with $c(t, \tau)$ the received pulses overlap. In this case, the channel has memory even if conditioned with respect to the channel impulse response. Time dispersion implies frequency selectivity, since conditions $T \approx T_d$ and $W \ll 1/T_d$ are not compatible, as WT is always greater than 1.

Table 1 summarizes the classes of fading channels examined above and the corresponding compatible signals. For each case, the channel can be either time-selective or not, depending on the product $B_d T$.

[4] This is not the case for other applications, like for example the underwater acoustic channel.

	$T_d \ll T$ (memoryless)	$T_d \approx T$ (ISI)
$W \ll 1/T_d$ (freq. flat)	any signal	impossible
$W \approx 1/T_d$ (freq. selective)	spread signal	any signal

Table 1: Classes of fading channels and compatible signal types.

1.5 Fading gain statistics

In this section we focus on the first-order statistics of the amplitude and power gains of a single-path channel with impulse response

$$c(t,\tau) = c(t)\delta(\tau - \tau_0)$$

This is an example of time non-dispersive, frequency non-selective channel. In fact, the received signal $y(t) = c(t)x(t-\tau_0)$ is just a delayed version of the transmit signal with some multiplicative distortion. For a given t, we define the instantaneous amplitude and power channel gains by $\alpha(t) = |c(t)|$ and $g(t) = |c(t)|^2$, respectively. We fix the time instant t and focus on the random variables $\alpha = \alpha(t)$ and $g = g(t)$. Because of wise-sense stationarity, the probability density functions (pdf) of α and g are independent of t.

For the channel examined in the previous sections, $c(t) \sim \mathcal{N}_{\mathcal{C}}(0,\Omega)$ where $\Omega = E[|c(t)|^2]$ is the average path power gain. In this case, α is Rayleigh distributed as

$$p_\alpha(z) = \frac{2z}{\Omega}\exp(-z^2/\Omega) \tag{16}$$

and g is exponentially distributed as

$$p_g(z) = \frac{1}{\Omega}\exp(-z/\Omega) \tag{17}$$

Rayleigh fading is typically originated by a large number of scattering elements, each contributing for a small fraction of the total received signal power, as in the case examined until now. However, there are other important situations, motivated by different physical propagation conditions, where the channel gain statistics are different. In the following, we review briefly the most important cases.

Rician fading. If there is a direct LOS propagation path between transmitter and receiver in addition to a large number of scattering elements, $c(t) \sim \mathcal{N}_{\mathcal{C}}(u,\sigma^2)$. The non-zero mean u is due to the direct LOS path, while the variance σ^2 is due to scattering, as before. By letting again $\Omega = E[|c(t)|^2] = \sigma^2 + |u|^2$ and by defining the ratio $K = |u|^2/\sigma^2$ between the average LOS to scattered power, we can write the pdf of α and g as

$$p_\alpha(z) = \frac{2(1+K)z}{\Omega}\exp\left(-\frac{1+K}{\Omega}\left(z^2 + \frac{\Omega K}{1+K}\right)\right) I_0\left(2z\sqrt{\frac{K(1+K)}{\Omega}}\right) \tag{18}$$

and

$$p_g(z) = \frac{1+K}{\Omega}\exp\left(-\frac{1+K}{\Omega}\left(z + \frac{\Omega K}{1+K}\right)\right) I_0\left(2\sqrt{\frac{zK(1+K)}{\Omega}}\right) \tag{19}$$

where $I_0(z)$ is the 0-th order modified Bessel function of the first kind [8]. Pdfs (18) and (19) are called *Rice* and *non-central chi-squared with two degrees of freedom*, and K is called *Rician factor*. The two extreme of Rayleigh fading (no LOS component) and no fading (no scattered component) are obtained as special cases of Rician fading for $K = 0$ and $K \to \infty$, respectively. Fig. 5 shows the Rice pdf for some values of K.

Log-normal shadowing. Another effect observed in propagation measurements of mobile wireless channels is the so-called *shadowing*, caused by obstacles like trees, hills or buildings, that do not scatter the signal but attenuate the received signal power. Rayleigh (or Rician) fading varies quite rapidly with respect to shadowing. For example, the coherence time of a mobile with Doppler spread $B_d = 100$ Hz is about $1/B_d = 10$ ms. On the contrary, the receiver may stay in the shadow of some obstacle for some seconds. Therefore, "short-term" fading due to scattering and "long-term" shadowing act on different time scales.

Experiments and propagation analysis shows that in the far-field of the transmit antenna, the long-term attenuation for ground propagation is proportional to $\zeta d^{-\alpha}$, where d is the distance between transmitted and receiver, α is an exponent that depends on the ground physical characteristics and ranges typically from 2 to 4 and ζ is a log-normal random variable [9] such that $E[\zeta] = 1$. In particular, it is customary to express ζ in dB, so that

$$10 \log_{10} \zeta \sim \mathcal{N}\left(-\frac{s^2 \log 10}{20}, s^2\right)$$

The standard deviation s, expressed in dB, is called *shadowing factor* and normally ranges from 2 to 8 dB.

Nakagami fading. Nakagami pdfs often offer the best fit of experimental data (see [10] and reference therein).

- Nakagami-q (satellite links subject to strong ionospheric scintillation).

$$\begin{aligned} p_\alpha(z) &= \frac{(1+q^2)z}{q\Omega} \exp\left(-\frac{(1+q^2)^2 z^2}{4q^2\Omega}\right) I_0\left(\frac{(1-q^4)z^2}{4q^2\Omega}\right) \\ p_g(z) &= \frac{(1+q^2)}{2q\Omega} \exp\left(-\frac{(1+q^2)^2 z}{4q^2\Omega}\right) I_0\left(\frac{(1-q^4)z}{4q^2\Omega}\right) \end{aligned} \tag{20}$$

  for $0 \le q \le 1$.

- Nakagami-n: same as Rice with $n^2 = K$.

- Nakagami-m (land mobile, indoor channels, ionospheric radio links).

$$\begin{aligned} p_\alpha(z) &= \frac{2m^m z^{2m-1}}{\Omega^m \Gamma(m)} \exp\left(-\frac{mz^2}{\Omega}\right) \\ p_g(z) &= \frac{m^m z^{m-1}}{\Omega^m \Gamma(m)} \exp\left(-\frac{mz}{\Omega}\right) \end{aligned} \tag{21}$$

  for $1/2 \le m \le \infty$ ($\Gamma(x)$ denotes the Gamma function [8]).

Figs. 6 and 7 show the Nakagami-q and -m pdfs for some values of q and m.

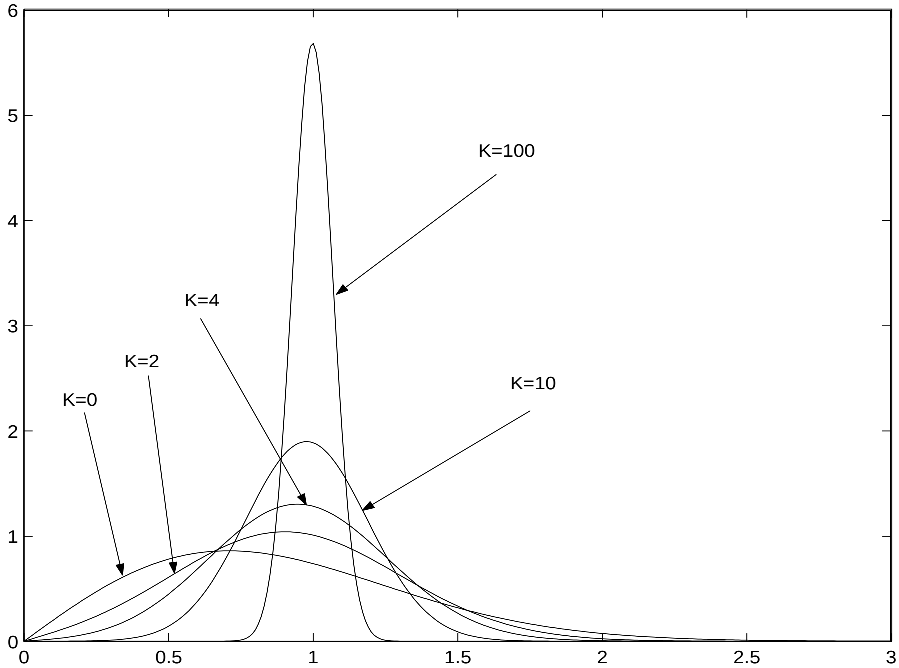

Figure 5: Rice pdf for some values of K and $\Omega = 1$.

Composite channels. In some applications, propagation conditions change at a rate much slower than the coherence time of fading. These effects are modeled by a channel state variable S that ranges over a number of possible channel states, corresponding to different fading statistics. The channel state variable remains in a given state for some time (dwell time) and changes according to some random or deterministic rule.

For example, in low-Earth orbit (LEO) satellite systems, propagation may have a LOS path or be blocked by objects. In the first case, the channel is Rice with average power gain Ω and fairly large parameter K, since the scattered component is normally very weak. In the second case, the channel is Rayleigh with average power $\zeta\Omega$, where ζ is log-normal. In [4], results from field measurements are collected and a two-state Markov model fitting the measurements is proposed. The channel state variable S takes on two possible values: a "good" state (Ricean channel) and a "bad" state (Rayleigh channel with log-normal average power). The transitions between good and bad states are governed by a two-state Markov chain, whose transition probabilities are calculated in order to fit the measurements. The dwell time in both states is much larger than the fading coherence time, so that in each state several realizations of the instantaneous channel gain occur.

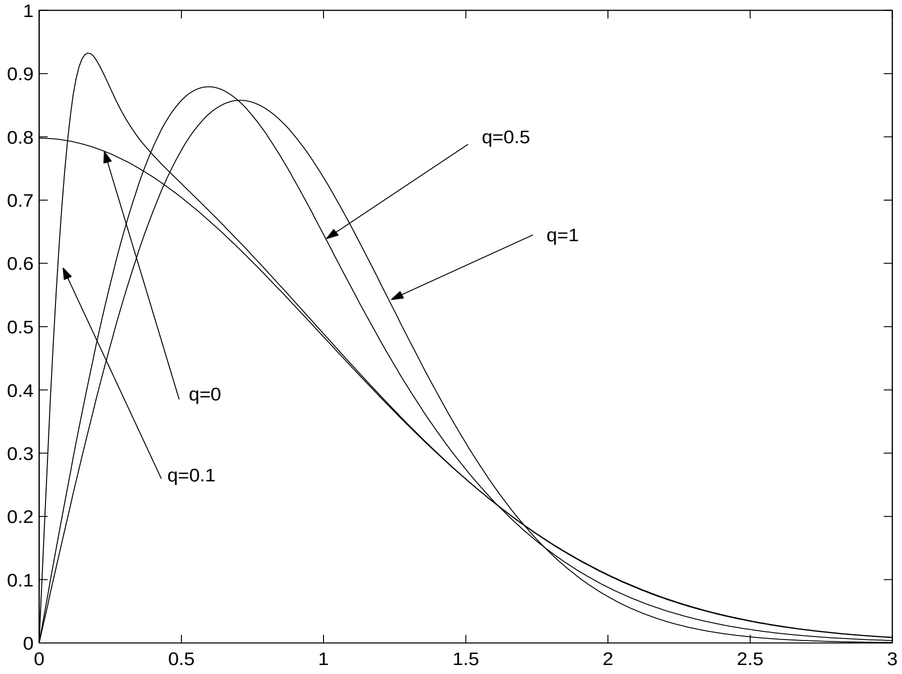

Figure 6: Nakagami-q pdf for some values of q and $\Omega = 1$.

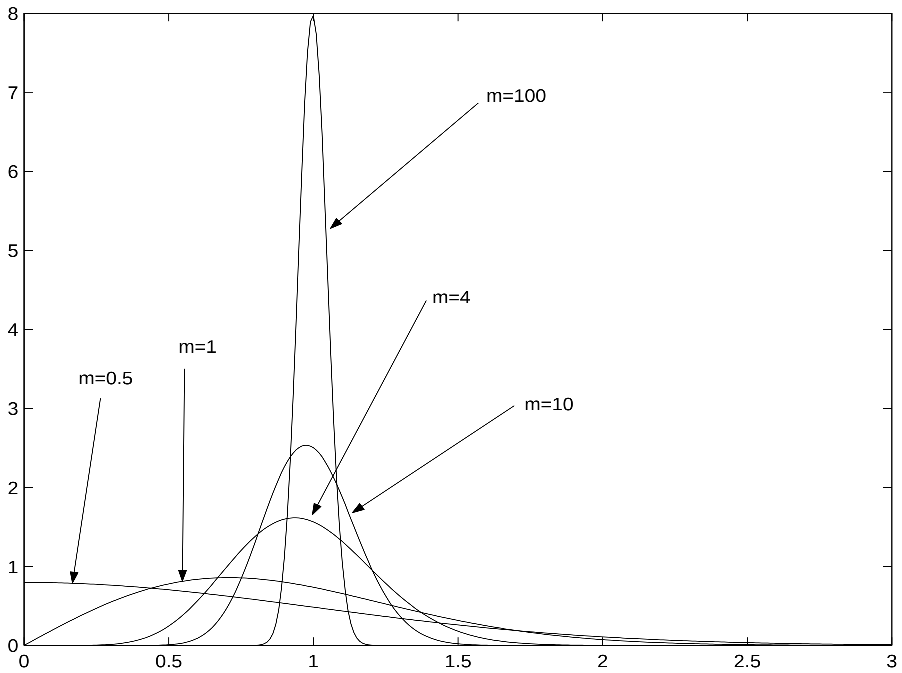

Figure 7: Nakagami-m pdf for some values of m and $\Omega = 1$.

2 MIMO channel modeling

This chapter presents an introduction to the mathematical modeling of time varying linear **M**ultiple **I**nput **M**ultiple **O**utput (MIMO) wireless channels. The material is based on the new approach devised in [11, 12] based on the principle of maximum entropy.

2.1 Channel Modelling Methodology

In this chapter, we provide a methodology (already successfully used in Bayesian spectrum analysis [13, 14]) 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.[5].
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 8) 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. 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

[5]Note that in Bayesian inference, all probabilities are conditional on some hypothesis space (which is assumed to be true).

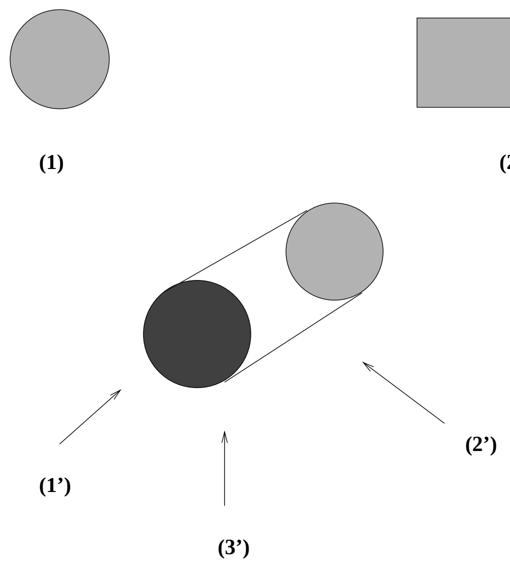

Figure 8: Duality wave-corpuscule?

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 [15] and Jeffrey [16]. 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.**??**)
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.

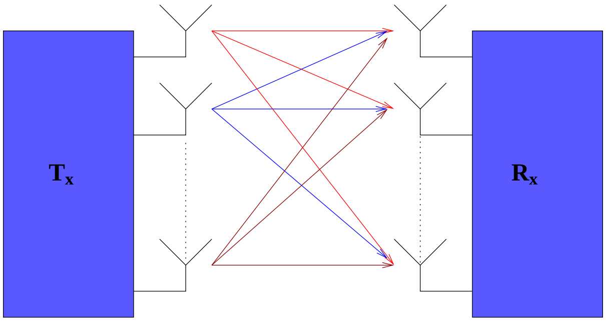

Figure 9: MIMO channel representation.

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 [14]). 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 Constraints

The transmission is assumed to take place between a mobile transmitter and receiver. The transmitter has n_t antennas and the receiver has n_r antennas. Moreover, the input transmitted signal is assumed to go through a time variant linear filter channel. Finally, the interfering noise is supposed to be additive white Gaussian.

The transmitted signal and received signal are related as:

$$y(t) = \sqrt{\frac{\rho}{n_t}} \int \mathbf{H}_{n_r \times n_t}(t,\tau) x(t-\tau) d\tau + n(t) \tag{22}$$

and

$$Y(f,t) = \sqrt{\frac{\rho}{n_t}} \mathbf{H}_{n_r \times n_t}(f,t) X(f) + N(f) \tag{23}$$

ρ is the received SNR, $Y(f)$ is the $n_r \times 1$ received vector (Fourier transform of the time signal $y(t)$), $X(f)$ is the $n_t \times 1$ transmit vector (Fourier transform of the time signal $x(t)$), $N(f)$ is an $n_r \times 1$ additive standardized white Gaussian noise vector (Fourier transform of $n(t)$).
In this section, we will only be interested in the frequency domain modeling (knowing that the impulse response matrix can be accessed through an inverse Fourier transform). We[6]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:

$$\mathbf{H}(f,t) = \begin{pmatrix} h_{11}(f,t) & \dots & \dots & h_{1n_t}(f,t) \\ \vdots & \dots & \dots & \vdots \\ \vdots & \dots & \dots & \vdots \\ h_{n_r1}(f,t) & \dots & \dots & h_{n_rt}(f,t) \end{pmatrix} \tag{24}$$

Here, we describe the MIMO link by a statistical model which takes into account our information of the environment. In this contribution, the goal is to derive a model which is adequate with our state of knowledge. A measure of uncertainty is needed which expresses the constraints of our knowledge and the desire to leave the unknown parameters to lie in an unconstraint space. To this end, many possibilities are offered to express our uncertainty. However, we need an information measure which is consistent (complying to certain common sense desiderata, see [19] to express these desiderata and for the derivation of entropy) and easy to manipulate: we need a simple general principle for translating information into probability assignment. Entropy is the measure of information that fulfills this criteria. The principle of maximum entropy states that, if P is the distribution of a random variable x, one should maximize the following functional under the state of information constraints:

$$D(P) = -\int P(x) \log (P(x))\, dx$$

Let us give an example in the context of spectral estimation of the powerful feature of the maximum entropy approach which has inspired this work. 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

[6]It was the work of Telatar [17] (later published as [18]) that triggered research in multi-antenna systems. For contribution [18], 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.

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 p^{th} auto-regressive (AR) order model process of the form [20]:

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

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 section, we will provide guidelines for creating models from an information theoretic point of view and therefore make extensive use of the principle of maximum entropy.

2.2.1 Gaussian i.i.d. channel model

In this section, we give a precise justification on why and when the Gaussian i.i.d. model should be used[7]. We recall the general model:

$$Y = \sqrt{\frac{\rho}{n_t}}\mathbf{H}X + N$$

Imagine now that the modeler is in a situation where he has no measurements and no knowledge where the transmission took place. The only thing the modeler assumes is that the channel carries some energy, in other words, each complex frequency path has a certain variance σ^2. Knowing only this information, the modeler is faced with the following question: what is the consistent model one can make assuming only the variance (but not the correlation even though it may exist) of the path gains? In other words, based on the fact that: For all i, j,

$$\int d\mathbf{H} \mid h_{ij} \mid^2 P(\mathbf{H}) = \sigma^2 \qquad \text{(Finite energy assumption)} \tag{25}$$

$$\int dP(\mathbf{H}) = 1 \qquad \text{(P}(\mathbf{H})\text{ is a probability distribution)} \tag{26}$$

what distribution $P(\mathbf{H})$ 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 mean and the variance. This statement can simply be expressed if one tries to maximize the following expression using Lagrange multipliers with respect to P:

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

[7] The title insists on purpose on the fact that the Gaussian i.i.d. channel is a model and not an assumption. There is a fundamental difference between the two as the author will try to explain hereafter.

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

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

then this yields:

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

with

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

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: if correlations exist, then the model will try to cope as best it can with this case because it is the one which makes the least assumption on the channel distribution. Independence is not at all an assumption but only the result of the maximum entropy principle. Instead of saying that the i.i.d. model does not contain correlation, it should be more correct to say as in [15] that this probability density function makes allowance for every possible correlation that could be present to exist and so is less informative than correlated distributions. 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 $\beta, \gamma_{ij}, \alpha_{ij}$. The coefficients are solutions of the following constraint equations: For all i, j,

$$\int d\mathbf{H} \mid h_{ij} \mid^2 P(\mathbf{H}) = \sigma^2$$

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

Solving the previous equations yields the following probability distribution:

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

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 each frequency path has different variances such as $\mathbb{E}(\mid h_{ij} \mid^2) = {\sigma_{ij}}^2$. 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 $P(h_{ij}) = \frac{1}{\pi \sigma_{ij}^2} e^{-\frac{|h_{ij}|^2}{\sigma_{ij}^2}}$. 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 $\mathbb{E}(h_{pk}) = \mu_{pk}$ and variance $\mathbb{E}(\mid h_{pk} - \mu_{pk} \mid^2) = \sigma_{pk}^2$, 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 $\{i, j, (i,j) \neq (p,k)\}$ $P(h_{ij}) = \frac{1}{\pi \sigma_{ij}^2} e^{-\frac{|h_{ij}|^2}{\sigma_{ij}^2}}$ and $P(h_{pk}) = \frac{1}{\pi \sigma_{pk}^2} e^{-\frac{|h_{pk} - \mu_{pk}|^2}{\sigma_{pk}^2}}$. 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. 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:

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

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

This model is called a pre-data model [15] in the Bayesian lexicography, in the sense that without knowing anything about the measured data, the best model one can make (best in the sense of maximizing our uncertainty with respect to certain conditions which we know are fulfilled) is the Gaussian i.i.d. model[8].

2.3 Double directional model

The modeler wants to derive a consistent model taking into account the directions of arrival and respective power profile, directions of departure and respective power profile, delay, Doppler effect. As a starting point, the modeler assumes that the position of the transmitter and receiver

[8]Using the maximum entropy principle to describe wave propagation has also been advocated recently. In "The Role of Entropy in Wave Propagation" [21], Franceshetti 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 travelled, the relevant metric is non-Euclidean whereas in sparse lattices, the relevant metric becomes Euclidean as propagation is not constrained along the axis directions.

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

$$\mathbf{s}_{ij}^r(t) = \mathbf{s}_0 e^{j(2\pi(\frac{f_c \mathbf{v}_r \mathbf{u}_{ij}^r}{c}t+f_c t))} e^{j\phi_{ij}}$$

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

- speed $\mathbf{v}_r$.
- 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:

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

The modeler also supposes as a given state of knowledge the fact that 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:

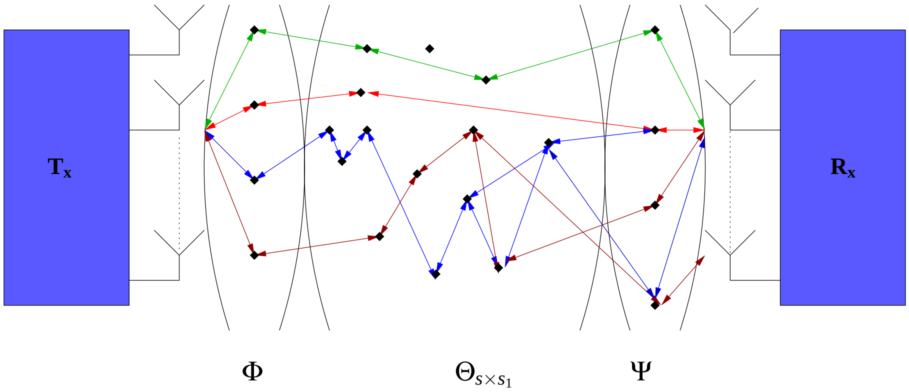

Figure 10: Moving antennas.

$$\mathbf{H}(f,t) = \frac{1}{\sqrt{s_t s_r}} \begin{pmatrix} e^{j(\phi_{1,1}+2\pi\frac{f\mathbf{u}^r_{11}\mathbf{v}_r}{c}t)} & \dots & e^{j(\phi_{1,s_r}+2\pi\frac{f\mathbf{u}^r_{1s_r}\mathbf{v}_r}{c}t)} \\ \vdots & \ddots & \vdots \\ e^{j(\phi_{n_r,1}+2\pi\frac{f\mathbf{u}^r_{n_r1}\mathbf{v}_r}{c}t)} & \dots & e^{j(\phi_{n_r,s_r}+2\pi\frac{f\mathbf{u}^r_{n_rs_r}\mathbf{v}_r}{c}t)} \end{pmatrix} \begin{pmatrix} P^r_1 & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & P^r_{s_r} \end{pmatrix}$$
$$\mathbf{\Theta}_{s_r\times s_t} \bigodot \mathbf{D}_{s_r\times s_t}(f)$$
$$\begin{pmatrix} P^t_1 & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & P^t_{s_t} \end{pmatrix} \begin{pmatrix} e^{j(\psi_{1,1}+2\pi\frac{f\mathbf{u}^t_{11}\mathbf{v}_t}{c}t)} & \dots & e^{j(\psi_{1,n_t}+2\pi\frac{f\mathbf{u}^t_{1n_t}\mathbf{v}_t}{c}t)} \\ \vdots & \ddots & \vdots \\ e^{j(\psi_{s_t,1}+2\pi\frac{f\mathbf{u}^t_{s_11}\mathbf{v}_t}{c}t)} & \dots & e^{j(\psi_{s_t,n_t}+2\pi\frac{f\mathbf{u}^t_{s_tn_t}\mathbf{v}_t}{c}t)} \end{pmatrix}$$

$\bigodot$ represents the Hadamard product defined as $c_{ij} = a_{ij}b_{ij}$ for a product matrix $\mathbf{C} = \mathbf{A} \bigodot \mathbf{B}$. It is straightforward to see that $\mathbf{\Theta}_{s_r\times s_t}$ i.i.d. zero mean Gaussian with variance 1 maximizes entropy under the previous constraints.

Remark A question the reader could ask is whether we should take into account all the information provided, in other words, why have we limited ourselves since the beginning to the variance of each path? We should of course take into account 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 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[9]. As a typical example, when conducting iterative decoding analysis [22], 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.

[9]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.

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

$$\begin{aligned}
\mathbf{H}(f,t) &= \frac{1}{\sqrt{s_r s_t}} \left[\begin{array}{ccc} e^{j(\phi_1+2\pi\frac{f\mathbf{u}_1^r\mathbf{v}_r}{c}t)} & \dots & e^{j(\phi_s+2\pi\frac{f\mathbf{u}_{s_r}^r\mathbf{v}_r}{c}t)} \end{array} \right] \begin{pmatrix} P_1^r & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & P_{s_r}^r \end{pmatrix} \\
& \quad \mathbf{\Theta}_{s_r\times s_t} \bigodot \mathbf{D}_{s_r\times s_t}(f) \begin{pmatrix} P_1^t & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & P_{s_t}^t \end{pmatrix} \left[\begin{array}{c} e^{j(\psi_1+2\pi\frac{f\mathbf{u}_1^t\mathbf{v}_t}{c}t)} \\ \vdots \\ e^{j(\psi_{s_t}+2\pi\frac{f\mathbf{u}_{s_t}^t\mathbf{v}_t}{c}t)} \end{array} \right] \\
&= \frac{1}{\sqrt{s_r s_t}} \left[\begin{array}{ccc} \sum_{k=1}^{s_r}\theta_{k,1}P_k^r e^{j(\phi_k+2\pi\frac{f\mathbf{u}_k^r\mathbf{v}_r}{c}t)}e^{-j2\pi f\tau_{k,1}} & \dots & \sum_{k=1}^{s_r}\theta_{k,s_t}P_k^r e^{j(\phi_k+2\pi\frac{f\mathbf{u}_k^r\mathbf{v}_r}{c}t)}e^{-j2\pi f\tau_{k,s_t}} \end{array} \right] \\
& \quad \begin{pmatrix} P_1^t & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & P_{s_t}^t \end{pmatrix} \begin{pmatrix} e^{j(\psi_1+2\pi\frac{f\mathbf{u}_1^t\mathbf{v}_t}{c}t)} \\ \vdots \\ e^{j(\psi_{s_t}+2\pi\frac{f\mathbf{u}_{s_t}^t\mathbf{v}_t}{c}t)} \end{pmatrix} \\
&= \frac{1}{\sqrt{s_r s_t}} \sum_{l=1}^{s_t}\sum_{k=1}^{s_r} \theta_{k,l}P_k^r P_l^t e^{j(\phi_k+2\pi\frac{f\mathbf{u}_k^r\mathbf{v}_r}{c}t)} e^{j(\psi_l+2\pi\frac{f\mathbf{u}_l^t\mathbf{v}_t}{c}t)} e^{-j2\pi f\tau_{k,l}}
\end{aligned}$$

This relation is a generalization of equation 7 in the case of multifold scattering with the power profile taken into account.

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 [23]. Note that in the case of a ULA geometry (Uniform Linear Array) and in the Fourier directions, we have $u_{ij}^r = u_j^r$ (any column of matrix Φ has a given direction) and $u_{ij}^t = u_i^t$ (any line of matrix Ψ has a given direction). Therefore, the channel model simplifies to:

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

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

$$\mathbf{D}_{s_r\times s_t}(f,t) = \begin{pmatrix} P_1^r P_1^t e^{-j2\pi f\tau_{1,1}} e^{j2\pi\frac{ft}{c}(\mathbf{u}_1^r\mathbf{v}_r+\mathbf{u}_1^t\mathbf{v}_t)} & \dots & P_1^r P_{s_t}^t e^{-j2\pi f\tau_{1,s_t}} e^{j2\pi\frac{ft}{c}(\mathbf{u}_1^r\mathbf{v}_r+\mathbf{u}_{s_t}^t\mathbf{v}_t)} \\ \vdots & \ddots & \vdots \\ P_{s_r}^r P_1^t e^{-j2\pi f\tau_{s_r,1}} e^{j2\pi\frac{ft}{c}(\mathbf{u}_{s_r}^r\mathbf{v}_r+\mathbf{u}_1^t\mathbf{v}_t)} & \dots & P_{s_r}^r P_{s_1}^t e^{-j2\pi f\tau_{s_r,s_t}} e^{j2\pi\frac{ft}{c}(\mathbf{u}_{s_r}^r\mathbf{v}_r+\mathbf{u}_{s_t}^t\mathbf{v}_t)} \end{pmatrix}$$

Although we take into account many parameters, the final model is quite simple. It is the product of three matrices: Matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$ taking into account the directions of arrival and departure; matrix $\mathbf{\Theta}_{s_r \times s_t} \bigodot \mathbf{D}_{s_r \times s_t}$ 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 $\mathbf{\Theta}_{s_r \times s_t} \bigodot \mathbf{D}_{s_r \times s_t}(f,t)$.

Remark 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,

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

  for every i,j and the mean of $\mathbf{H}(f)$ is therefore independent of f.

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

$$\begin{aligned}\text{Cov}(i,j,m,n,f) &= \frac{1}{s_r s_t} \sum_{k=1}^{s_t} \sum_{p=1}^{s_r} \sum_{q=1}^{s_t} \sum_{l=1}^{s_r} \mathbb{E}(\theta_{pk}\theta_{ql}^*) e^{j2\pi f(\tau_{pk}-\tau_{ql})} \\ & P_k^t P^{*t}_q P_p^r P^{*r}_l e^{j2\pi(\psi_{kj}-\psi_{qn})} e^{j2\pi(\phi_{ip}-\phi_{ml})}\end{aligned}$$

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

$$\text{Cov}(i,j,m,n,f) = \frac{1}{s_r s_t} \sum_{k=1}^{s_t} \sum_{p=1}^{s_r} \mid P_k^t \mid^2 \mid P_p^r \mid^2 e^{j2\pi(\psi_{kj}-\psi_{kn})} e^{j2\pi(\phi_{ip}-\phi_{mlp})}$$

  which is independent of f.

2.4 Other Models

2.4.1 Müller's Model

In a paper "A Random Matrix Model of Communication via Antenna Arrays" [24], Müller develops a channel model based on the product of two random matrices:

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

where $\mathbf{\Phi}$ and $\mathbf{\Theta}$ are two random matrices with zero mean unit variance i.i.d entries and $\mathbf{A}$ is a diagonal matrix (representing the attenuations). This model is intended to represent the fact that each signal bounces off a scattering object exactly once. $\mathbf{\Phi}$ 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 [24] 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 [25]

2.4.2 Sayeed's Model

In a paper "Deconstructing Multi-antenna Fading Channels" [26], 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 [23], 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.

2.4.3 The "Kronecker" model

In a paper "Capacity Scaling in MIMO Wireless Systems Under Correlated fading", Chuah et al. study the following Kronecker [10] model:

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

Here, $\mathbf{\Theta}$ is an $n_r \times n_t$ i.i.d zero mean Gaussian matrix, $\mathbf{R}_{n_r}{}^{\frac{1}{2}}$ is an $n_r \times n_r$ receiving correlation matrix while $\mathbf{R}_{n_t}{}^{\frac{1}{2}}$ is a $n_t \times n_t$ transmitting correlation matrix. The correlation is supposed to decrease sufficiently fast so that $\mathbf{R}_{n_r}$ and $\mathbf{R}_{n_t}$ have a Toeplitz band structure. Using a software tool (Wireless System Engineering [29]), 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

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

[10] 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 [27, 28].

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:

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

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

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

Note that this connection with the double directional model has already been reported in [23]. 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.

2.4.4 The "Keyhole" Model

In [30], Gesbert et al. show that low correlation[11] 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 [32], they propose the following model for a rank one channel:

$$\mathbf{H} = \mathbf{R}_{n_r}{}^{\frac{1}{2}} \mathbf{g}_r \mathbf{g}_t{}^H \mathbf{R}_{n_t}{}^{\frac{1}{2}} \tag{27}$$

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

$$\mathbf{H} = \frac{1}{\sqrt{s_r s_t}} \mathbf{\Phi}_{n_r \times s_r} \mathbf{\Theta}_{s_r \times s_t} \mathbf{\Psi}_{s_t \times n_t} \tag{28}$$

The Keyhole model can be either:

[11]"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 [31].

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

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

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

References

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

[2] R. H. Clarke, "A Statistical Theory of Mobile-Radio Reception," pages 957-1000, Bell Syst. Tech J., July-August 1968.

[3] R. S. Kennedy, *Fading Dispersive Communication Channels*, Wiley, New York, USA, 1969.

[4] E. Lutz, D. Cygan, M. Dippold, F. Dolainsky, and W. Papke, "The Land Mobile Satellite Communication Channel: Recording, Statistics and Channel Model," in *IEEE Trans. on Vehic.*, May, 1991.

[5] P. Billingsley, *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, 3nd edition, 1995.

[6] F. Neeser and J.L. Massey, "Proper Complex Random Processes with Application to Information Theory," *IEEE Trans. on Information Theory*, pp. 1293–1302, July 1993.

[7] J. William C. Jakes, Ed., *Microwave Mobile Communications*, John Wiley and sons, New York, 1974.

[8] M. Abramovitz and I. Stegun, *Handbook of Mathematical Functions*, Dover Publications Inc., New York, 1970.

[9] T.S. Rappaport, *Wireless Communications, Principles and Practice*, Prentice-Hall, New Jersey, 1996.

[10] M. K. Simon and M.S. Alouini, "A Unified Approach to the Performance Analysis of Digital Communications over Generalized Fading Channels," in *Proc. of the IEEE*, 1860-1877, Sept. 1998.

[11] 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.

[12] M. Debbah, R. Müller, H. Hofstetter, and P. Lehne, "Validation of Mutual Information Complying MIMO Models," *submitted to IEEE Transactions on Wireless Communications*, can be downloaded at http://www.eurecom.fr/∼debbah, 2004.

[13] 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.

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

[15] E. T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge, 2003.

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

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

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

[19] 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.

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

[21] 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.

[22] J. Boutros and G. Caire, "Iterative Multiuser Joint Decoding: Unified Framework and Asymptotic Analysis ," *IEEE Trans. on Information Theory*, pp. 1772–1793, July 2002.

[23] 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.

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

[25] 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.

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

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.