Multiple Antenna Technologies

Mérouane Debbah

February 10, 2008

Contents

Preface

In 1948, Shannon determines the achievable rate in a point to point link through the famed capacity formula:

$$C = W \log_2(1 + \mathrm{SNR})$$

where W is the bandwidth and SNR is the Signal to Noise Ratio. Although extremely simple, this formula has put a huge burden on engineers since Shannon did not provide any feasible scheme to achieve the limiting rate in digital transmissions. Indeed, it was only with the invention of Turbo-codes (and now LDPC (Low Density Parity Check codes)) by C. Bérou of ENST Bretagne in the 90's that engineers were able to fully exploit the channel capabilities. Moreover, for 50 years, engineers were limited in terms of transmission by two degrees of freedom: bandwidth and power. Since the capacity scale linearly with the bandwidth and the logarithmically with the SNR, engineers have focused their attention on bandwidth. However, in the wireless arena, bandwidth is a rare resource which can not be used indefinitely and has to be shared by multiple technologies.

Aware of this limiting factor, a young researcher of Bell Labs, E. Telatar (now Professor at EPFL in Lausanne) derives in 1995 the capacity of a multiple link to multiple link transmission (also known as MIMO (Multiple Input Multiple Output)) in another famed technical report of Bell Labs [1] later published in [2]. Surprisingly, Telatar shows that without any increase of transmitted power (or bandwidth), the achievable rate for certain type of multiple antenna channels is in fact

$$C = \min(n_t, n_r) W \log_2(1 + \mathrm{SNR})$$

at high SNR (n_t and n_r are respectively the number of transmitting and receiving antennas). This striking result has led since ten years to many advances in wireless communications towards the space frontier.

At first sight, one could argue that one could pack more and more antennas with an indefinite increase of capacity. Is it the case? How should the antennas be displayed? What are the construction schemes that achieve that rate?

This course is intended to give some partial answers to the previous questions by providing a brief introduction to Multiple Antenna systems. In the first three chapters,we will focus on channel modelling. Two approaches are provided. In chapter 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 chapter 2, we use 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. In chapter 3, measurements performed recently in Oslo at 2.1 GHz and 5.2 GHz give us some insight on the behavior of the wireless link and show the relevance of the models proposed. In chapter 4, we introduce some useful metrics for MIMO systems and finally, in chapter 5, we discuss some practical schemes exploiting the space dimension.

Chapter 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 [3] and from [4].

1.1 Doppler effect and deterministic fading

Consider the situation of Fig. 1.1, 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

$$\begin{aligned} r(t) &= \mathrm{Re}\{x(t-\tau_0(t))\exp(j2\pi f_c(t-\tau_0(t))\} \\ &\approx \mathrm{Re}\{x(t-\tau_0)\exp(j2\pi(f_c+\xi_0)t)\exp(j\phi_0)\} \end{aligned} \tag{1.1}$$

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 (1.1) 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.

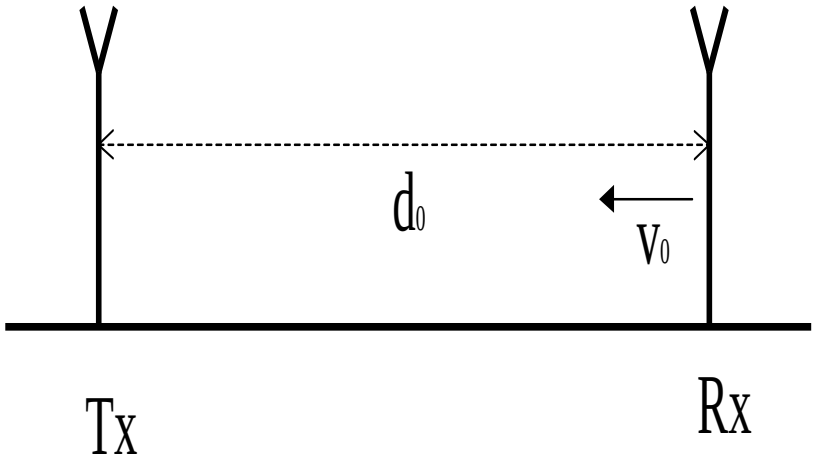

Figure 1.1: Line of sight propagation with a mobile terminal.

Next, consider the situation of Fig. 1.2, 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

$$\begin{aligned} r(t) &\approx \mathrm{Re}\{x(t-\tau_0)\exp(j\phi_0)\exp(j2\pi(f_c+\xi_0)t) + ax(t-\tau_1)\exp(j\phi_1)\exp(j2\pi(f_c+\xi_1)t)\} \\ &= \mathrm{Re}\left\{\left(\sum_{i=0}^{1}\rho_i x(t-\tau_i)\exp(j2\pi\xi_i t)\right)\exp(j2\pi f_c t)\right) \end{aligned} \tag{1.2}$$

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 (1.2) 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

$$\begin{aligned} |y(t)| &= |x\rho_0|\left|1 + \frac{\rho_1}{\rho_0}\exp(j2\pi(\xi_1-\xi_0)t)\right| \\ &= A\sqrt{1+b^2+2b\cos(2\pi\Delta\xi t + \Delta\phi)} \end{aligned} \tag{1.3}$$

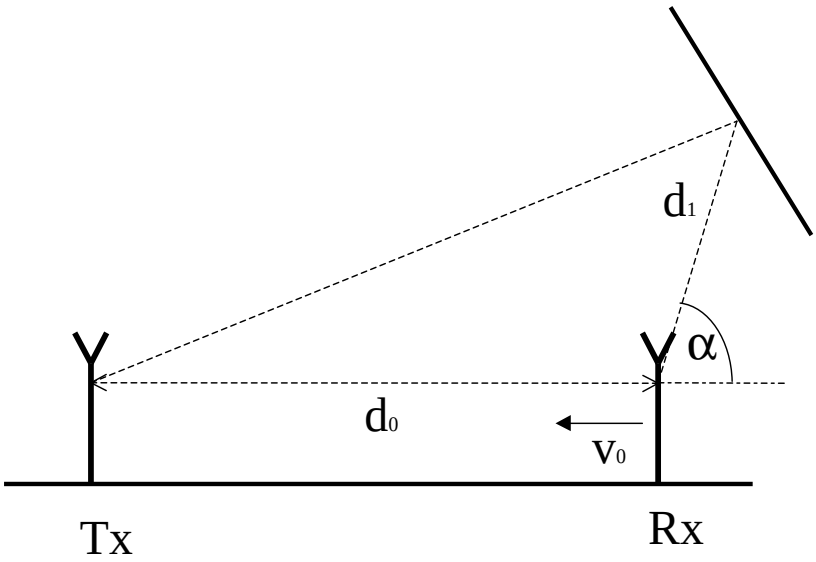

Figure 1.2: Two-ray propagation with a mobile terminal.

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 (1.3) 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*.

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. 1.3, 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

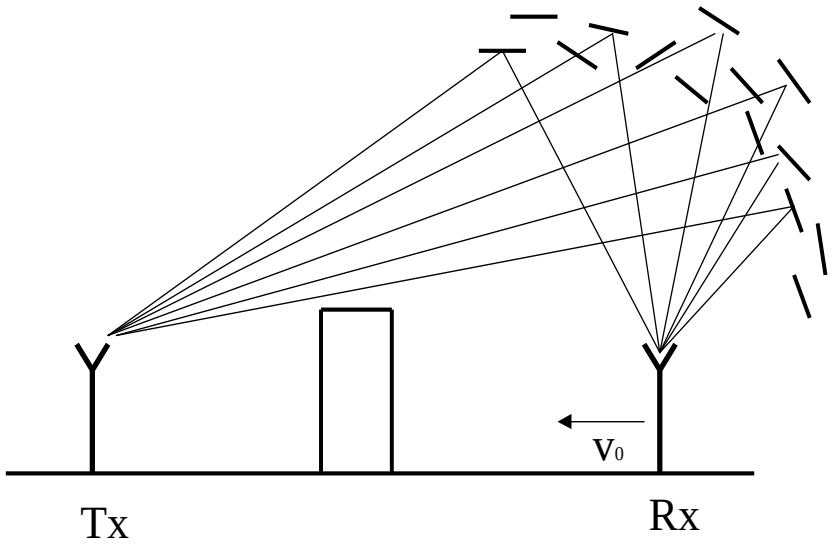

Figure 1.3: Scattering from many elements, without line of sight propagation.

done before, we can write the complex envelope of the received signal as

$$y(t) = \sum_i \rho_i x(t - \tau_i) \exp(j2\pi\xi_i t) \tag{1.4}$$

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 [5, 4], is that the scattering coefficients ρ_i are zero-mean pairwise uncorrelated, i.e., $E[\rho_i \rho_j^*] = 0$ for all $i \neq j$. 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.

[1] All channel models specified by ETSI as standard test models for GSM belong to this class [6].

Then, (1.4) can be written as

$$\begin{aligned} y(t) &= \sum_{p=0}^{P-1} \left(\sum_{i \in \mathcal{P}_p} \rho_i \exp(j2\pi\xi_i t) \right) x(t-\tau_p) \\ &= \sum_{p=0}^{P-1} c_p(t) x(t-\tau_p) \end{aligned} \tag{1.5}$$

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

$$c(t,\tau) = \sum_{p=0}^{P-1} c_p(t)\delta(\tau - \tau_p) \tag{1.6}$$

and

$$C(t,f) = \sum_{p=0}^{P-1} c_p(t) e^{-j2\pi f \tau_p} \tag{1.7}$$

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 [7] 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]. $\Diamond$

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 (1.4) in the following integral form

$$\begin{aligned} y(t) &= \int_{\mathcal{D}} \int_{\mathcal{T}} \rho(\xi,\tau) x(t-\tau) \exp(j2\pi\xi t) d\xi d\tau \\ &= \int_{\mathcal{T}} c(t,\tau) x(t-\tau) d\tau \end{aligned} \tag{1.8}$$

where we define the channel impulse response

$$c(t,\tau) = \int_{\mathcal{D}} \rho(\xi,\tau) \exp(j2\pi\xi t) d\xi \tag{1.9}$$

[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)

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 [5] 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 [8], 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.

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 (1.9), 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

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

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

$$\Phi(\Delta t,\Delta f)=\int\int \sigma(\xi,\tau)\exp(j2\pi\xi\Delta t)\exp(-j2\pi\Delta f\tau)d\xi d\tau \tag{1.10}$$

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

$$P(\tau)=\int\sigma(\xi,\tau)d\xi=\phi(0,\tau) \tag{1.11}$$

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

$$T_d=\left(\frac{\int_{\mathcal{T}}(\tau-\bar{\tau})^2P(\tau)d\tau}{\int_{\mathcal{T}}P(\tau)d\tau}\right)^{1/2} \tag{1.12}$$

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

$$D(\xi)=\int\sigma(\xi,\tau)d\tau=\int\Phi(\Delta t,0)\exp(-j2\pi\xi\Delta t)d\Delta t \tag{1.13}$$

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

$$B_d=\left(\frac{\int_{\mathcal{D}}(\xi-\bar{\xi})^2D(\xi)d\xi}{\int_{\mathcal{D}}D(\xi)d\xi}\right)^{1/2} \tag{1.14}$$

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 [4, 9]. 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 (1.4), 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. 1.4. 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. 1.4). 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

$$\begin{aligned} r_c(\Delta t) &= E\left[\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \rho_\alpha \rho_\beta^* \exp(j2\pi F_d\cos\alpha t)\exp(-j2\pi F_d\cos\beta(t-\Delta t))d\alpha d\beta\right] \\ &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} p(\alpha)g(\alpha)\delta(\alpha-\beta)\exp(j2\pi F_d\cos\alpha t)\exp(-j2\pi F_d\cos\beta(t-\Delta t))d\alpha d\beta \\ &= \int_{-\pi}^{\pi} p(\alpha)g(\alpha)\exp(j2\pi F_d\cos\alpha\Delta t)d\alpha \end{aligned} \tag{1.15}$$

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 $D(\xi) = \int r_c(\Delta t)\exp(-j2\pi\xi\Delta t)d\Delta t$.

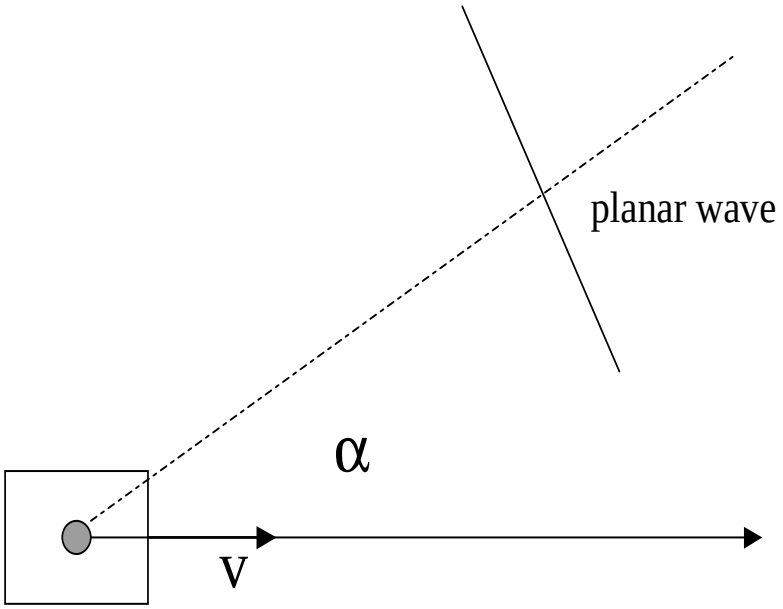

Figure 1.4: Planar geometry for Doppler spectra calculation.

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 (1.15) 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 [10]. 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 [9]. The corresponding Doppler spectrum is given by

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

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 $W \gg 1/T$).

[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 1.1: Classes of fading channels and compatible signal types.

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 1.1 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 $B_d T$.

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

$$p_\alpha(z) = \frac{2z}{\Omega}\exp(-z^2/\Omega) \tag{1.16}$$

and g is exponentially distributed as

$$p_g(z) = \frac{1}{\Omega}\exp(-z/\Omega) \tag{1.17}$$

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, $c(t) \sim \mathcal{N}_{\mathbb{C}}(u, \sigma^2)$. 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

$$p_\alpha(z) = \frac{2(1+K)z}{\Omega} \exp\left(-\frac{1+K}{\Omega}\left(z^2 + \frac{\Omega K}{1+K}\right)\right) I_0\left(2z\sqrt{\frac{K(1+K)}{\Omega}}\right) \tag{1.18}$$

and

$$p_g(z) = \frac{1+K}{\Omega} \exp\left(-\frac{1+K}{\Omega}\left(z + \frac{\Omega K}{1+K}\right)\right) I_0\left(2\sqrt{\frac{zK(1+K)}{\Omega}}\right) \tag{1.19}$$

where $I_0(z)$ is the 0-th order modified Bessel function of the first kind [10]. Pdfs (1.18) and (1.19) 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. 1.5 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 [11] 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 [12] and reference therein).

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

$$\begin{aligned} p_\alpha(z) &= \frac{(1+q^2)z}{q\Omega} \exp\left(-\frac{(1+q^2)^2 z^2}{4q^2\Omega}\right) I_0\left(\frac{(1-q^4)z^2}{4q^2\Omega}\right) \\ p_g(z) &= \frac{(1+q^2)}{2q\Omega} \exp\left(-\frac{(1+q^2)^2 z}{4q^2\Omega}\right) I_0\left(\frac{(1-q^4)z}{4q^2\Omega}\right) \end{aligned} \tag{1.20}$$

  for $0 \le q \le 1$.

- Nakagami-n: same as Rice with $n^2 = K$.
- Nakagami-m (land mobile, indoor channels, ionospheric radio links).

$$\begin{aligned} p_\alpha(z) &= \frac{2m^m z^{2m-1}}{\Omega^m \Gamma(m)} \exp\left(-\frac{mz^2}{\Omega}\right) \\ p_g(z) &= \frac{m^m z^{m-1}}{\Omega^m \Gamma(m)} \exp\left(-\frac{mz}{\Omega}\right) \end{aligned} \tag{1.21}$$

for $1/2 \leq m \leq \infty$ ($\Gamma(x)$ denotes the Gamma function [10]).

Figs. 1.6 and 1.7 show the Nakagami-q and -m pdfs for some values of q and m.

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 [6], 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.

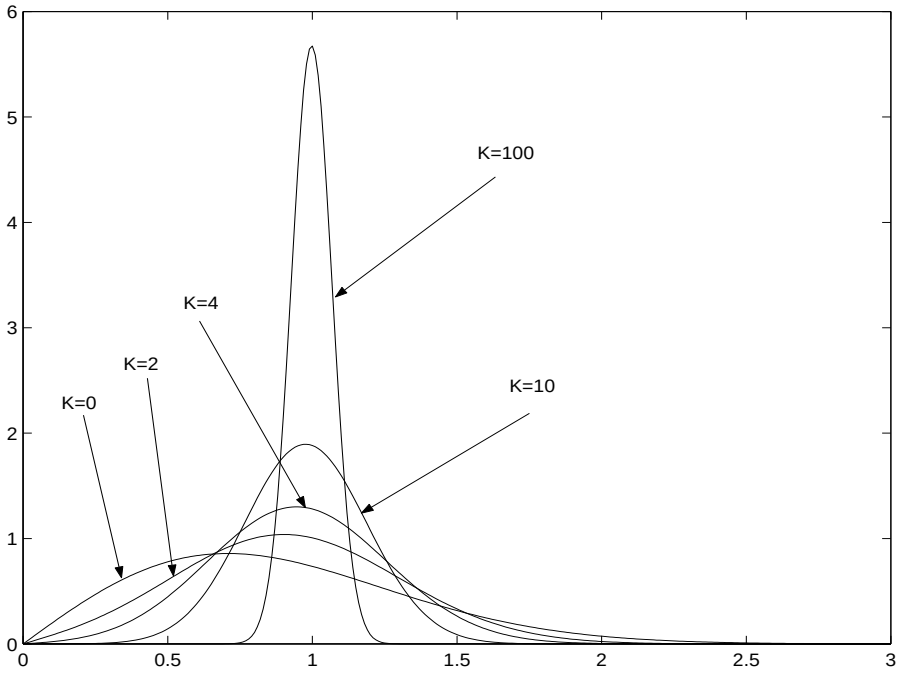

Figure 1.5: Rice pdf for some values of K and $\Omega = 1$.

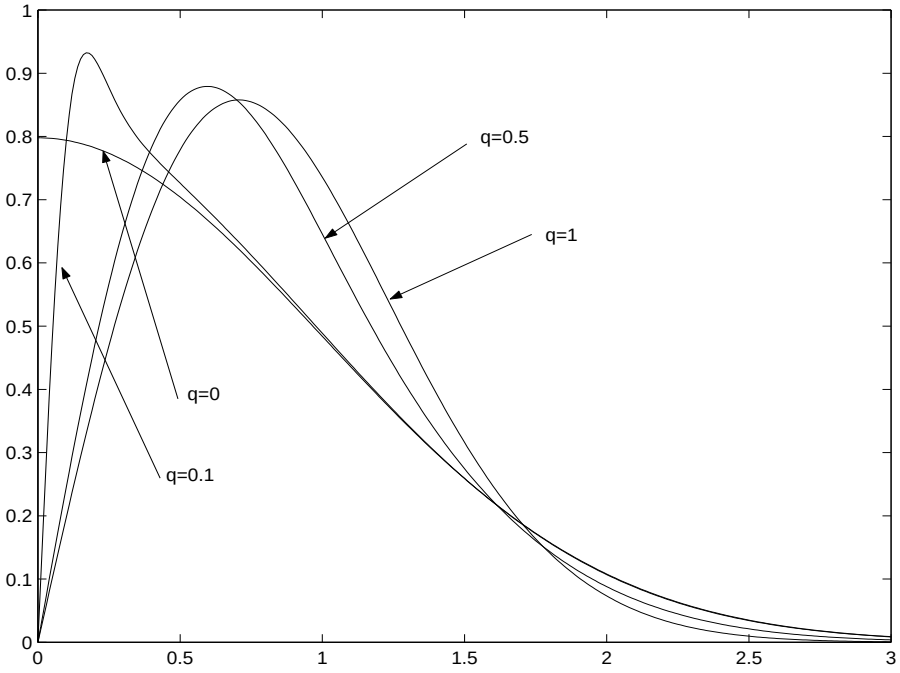

Figure 1.6: Nakagami-q pdf for some values of q and $\Omega = 1$.

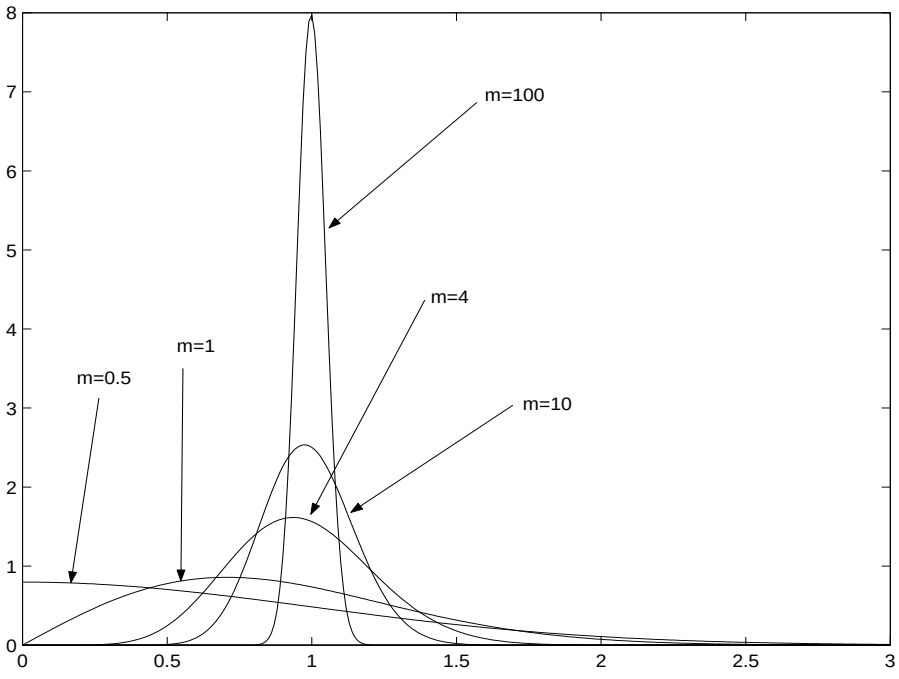

Figure 1.7: Nakagami-m pdf for some values of m and $\Omega = 1$.

Chapter 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 [13, 14] 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 [15, 16]) for inferring on channel models. The goal of the modelling methodology is twofold:

- to define a set of rules, called hereafter *consistency axioms*, where only our state of knowledge needs to be defined.
- to use a measure of uncertainty, called hereafter *entropy*, in order to avoid the arbitrary introduction or assumption of information that is not available.

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

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

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

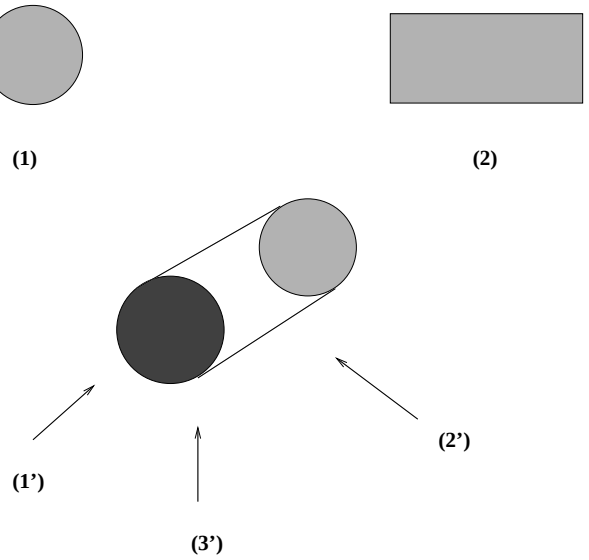

Figure 2.1: Duality wave-corpuscule?

To explain this point of view, the author recalls an experiment made by his teacher during a tutorial explanation on the duality behavior of light: photon or wave. The teacher took two students of the class, called here A and B for simplicity sake. To student A, he showed view (1') (see Figure 2.1) of a cylinder and to student B, he showed view (2') of the same cylinder. For A, the cylinder was a circle and for B, the cylinder was a rectangle. Who was wrong? Well, nobody. Based on the state of knowledge (1'), representing the cylinder as a circle is the best one can do. Any other representation of the cylinder would have been made on unjustified assumptions (the same applies to view (2')). Unless we have another state of knowledge (view (3')), the true nature of the object will not be found.

Our channel modelling will not pretend to seek reality but only to represent view (1') or view (2') in the most accurate way (i.e if view (1') is available then our approach should lead into representing the cylinder as a circle and not as a triangle for example). If the model fails to comply with measurements, we will not put into doubt the model but conclude that the information we had at hand to create the model was insufficient. We will take into account the failure as a new source of information and refine/change our question in order to derive a new model based on the principle of maximum entropy which complies with the measurements. This procedure will be routinely applied until the right question (and therefore the right answer) is found. When performing scientific inference, every question asked, whether right or wrong, is important. Mistakes are eagerly welcomed as they lead the path to better understand the propagation environment. Note that the approach devised here is not new and has already been used by Jaynes [17] and Jeffrey [18]. 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.

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 [16]). 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:

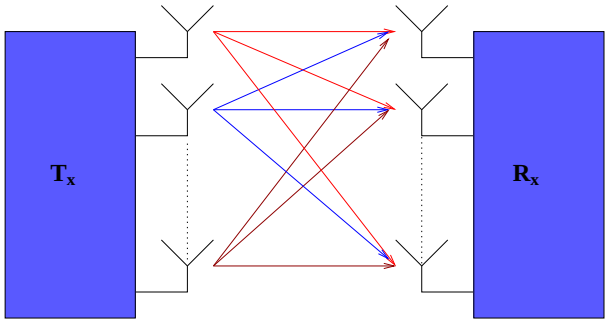

Figure 2.2: MIMO channel representation.

$$y(t) = \sqrt{\frac{\rho}{n_t}} \int \mathbf{H}_{n_r \times n_t}(t,\tau) x(t-\tau) d\tau + n(t) \tag{2.1}$$

and

$$Y(f,t) = \sqrt{\frac{\rho}{n_t}} \mathbf{H}_{n_r \times n_t}(f,t) X(f) + N(f) \tag{2.2}$$

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[2]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:

$$\mathbf{H}(f,t) = \begin{pmatrix} h_{11}(f,t) & \dots & \dots & h_{1n_t}(f,t) \\ \vdots & \dots & \dots & \vdots \\ \vdots & \dots & \dots & \vdots \\ h_{n_r 1}(f,t) & \dots & \dots & h_{n_r t}(f,t) \end{pmatrix} \tag{2.3}$$

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]

[2]It was the work of Telatar [1] (later published as [2]) that triggered research in multi-antenna systems. For contribution [2], 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.

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:

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

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 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.

2.3 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[3]. 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,

$$\int d\mathbf{H} \mid h_{ij} \mid^2 P(\mathbf{H}) = \sigma^2 \qquad \text{(Finite energy assumption)} \tag{2.4}$$

$$\int dP(\mathbf{H}) = 1 \qquad \text{(P(}\mathbf{H}\text{) is a probability distribution)} \tag{2.5}$$

[3]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.

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:

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

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 [17] 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 [17] 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[4].

2.4 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 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:

[4]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.

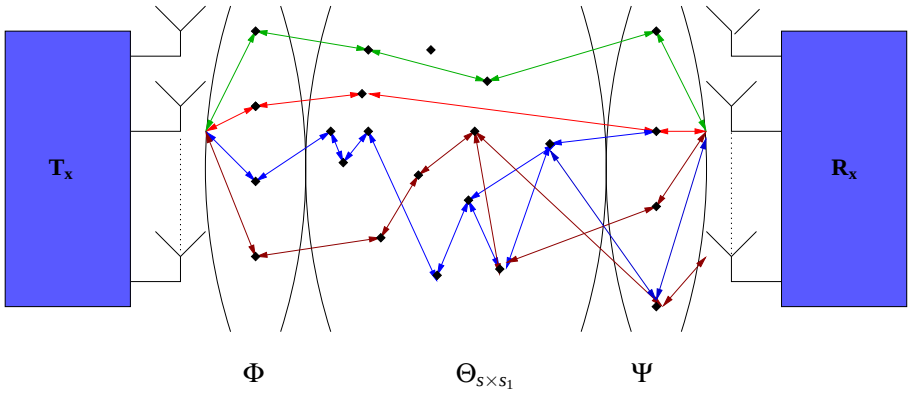

Figure 2.3: Moving antennas.

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

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

$$\begin{aligned} \mathbf{H}(f,t) = & \frac{1}{\sqrt{s_t s_r}} \begin{pmatrix} e^{j(\phi_{1,1}+2\pi \frac{f \mathbf{u}_{11}^r \mathbf{v}_r}{c} t)} & \dots & e^{j(\phi_{1,s_r}+2\pi \frac{f \mathbf{u}_{1s_r}^r \mathbf{v}_r}{c} t)} \\ \vdots & \ddots & \vdots \\ e^{j(\phi_{n_r,1}+2\pi \frac{f \mathbf{u}_{n_r 1}^r \mathbf{v}_r}{c} t)} & \dots & e^{j(\phi_{n_r,s_r}+2\pi \frac{f \mathbf{u}_{n_r s_r}^r \mathbf{v_r}}{c} t)} \end{pmatrix} \begin{pmatrix} P_1^r & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & P_{s_r}^r \end{pmatrix} \\ & \mathbf{\Theta}_{s_r \times s_t} \bigodot \mathbf{D}_{s_r \times s_t}(f) \\ & \begin{pmatrix} P_1^t & 0 & \dots \\ 0 & \ddots & 0 \\ \vdots & 0 & P_{s_t}^t \end{pmatrix} \begin{pmatrix} e^{j(\psi_{1,1}+2\pi \frac{f \mathbf{u}_{11}^t \mathbf{v}_t}{c} t)} & \dots & e^{j(\psi_{1,n_t}+2\pi \frac{f \mathbf{u}_{1n_t}^t \mathbf{v}_t}{c} t)} \\ \vdots & \ddots & \vdots \\ e^{j(\psi_{s_t,1}+2\pi \frac{f \mathbf{u}_{s_1 1}^t \mathbf{v}_t}{c} t)} & \dots & e^{j(\psi_{s_t,n_t}+2\pi \frac{f \mathbf{u}_{s_t n_t}^t \mathbf{v_t}}{c} t)} \end{pmatrix} \end{aligned}$$

$\bigodot$ 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[5]. 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.

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

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

This relation is a generalization of equation 1.7 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:

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

[5]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.

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

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

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

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

  which is independent of f.

2.5 Other Models

2.5.1 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:

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

where $\mathbf{\Phi}$ and $\mathbf{\Theta}$ are two random matrices with zero mean unit variance i.i.d entries and $\mathbf{A}$ is a diagonal matrix (representing the attenuations). This model is intended to represent the fact

that each signal bounces off a scattering object exactly once. $\mathbf{\Phi}$ represents the steering directions from the scatterers to the receiving antennas while $\mathbf{\Theta}$ represents the steering directions from the transmitting antennas to the scatterers. Measurements in [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]

2.5.2 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.

2.5.3 The "Kronecker" model

In a paper "Capacity Scaling in MIMO Wireless Systems Under Correlated fading", Chuah et al. study the following Kronecker [6] model:

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

[6]The model is called a Kronecker model because $\mathbb{E}(\mathrm{vec}(\mathbf{H})^H \mathrm{vec}(\mathbf{H})) = \mathbf{R}_{n_r} \bigotimes \mathbf{R}_{n_t}$ 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].

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

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

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:

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

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

2.5.4 The "Keyhole" Model

In [30], Gesbert et al. show that low correlation[7] is not a guarantee of high capacity: cases where the channel is rank deficient can appear while having uncorrelated entries (for example when a

[7] "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].

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:

$$\mathbf{H} = \mathbf{R}_{n_r}{}^{\frac{1}{2}} \mathbf{g}_r \mathbf{g}_t{}^H \mathbf{R}_{n_t}{}^{\frac{1}{2}} \tag{2.6}$$

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:

$$\mathbf{H} = \frac{1}{\sqrt{s_r s_t}} \mathbf{\Phi}_{n_r \times s_r} \mathbf{\Theta}_{s_r \times s_t} \mathbf{\Psi}_{s_t \times n_t} \tag{2.7}$$

The Keyhole model can be either:

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

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

Chapter 3

Measurement results

3.1 Measurement set-up

In this section[1], we provide some results on a wideband outdoor measurement campaign carried out in Oslo during summer 2002 and summer 2003 [35]. The measurements were performed at a center frequency of 2.1 GHz and 5.2 GHz with a bandwidth of 100 MHz in three different urban scenarios: a regular street grid scenario, an open city place and an indoor cell site. In each scenario, many routes have been measured: at 2.1 GHz, 150 routes have been measured (and in each case, many snapshots have been measured) whereas at 5.2 GHz, 79 routes have been measured. The measurements performed at 2.1 GHz are relevant for UMTS whereas those at 5.2 GHz are valuable for IEEE 802.11a.

- The street grid scenario is in Oslo downtown and corresponds to Concrete/Brick buildings. The buildings are around 20-30 m high (see Figure 3.1). In this scenario, two different receiver positions were tested, one high position on a roof terrace and the other one, a low position on street level. The area is often referred as "Kvadraturen".
- 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 3.2). The surrounding buildings are of variable size. In this scenario, the receiver was placed above some arcades.
- For the indoor scenario, the measurements were performed in a modern office building with open indoor areas. The building (Telenor headquarters building at Fornebu) has a irregular structure and is mostly of glass and steel. Measurements were taken in two different parts of the building. Inside a work zone (see Figure 3.3 and 3.4) and in a common area called the "Atrium" (see Figure 3.5).

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.

[1]The author is grateful to Helmut Hofstetter and P.H. Lehne for useful discussions and providing the measurements. A comprehensive introduction to the measurement set-up can be found in [33] and [34]. For the different scenario routes, notations of [33] and [34] will be used.

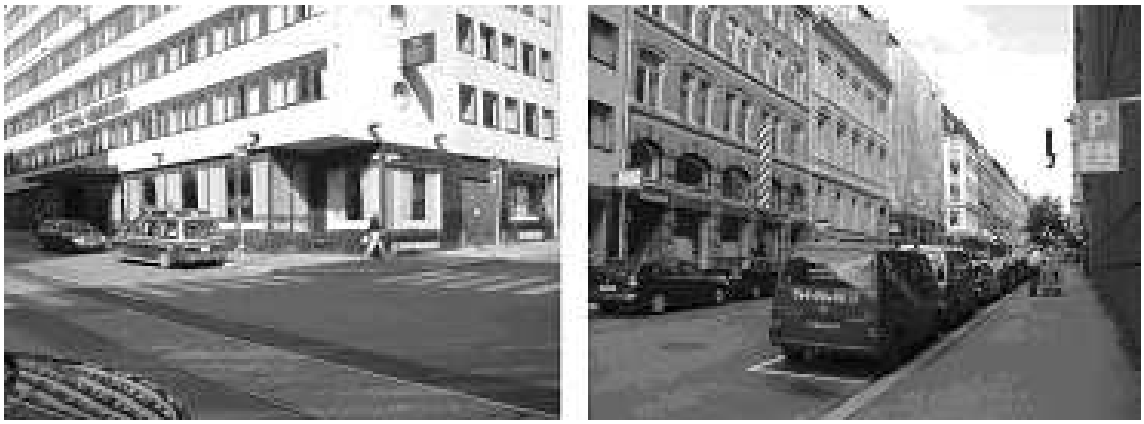

Figure 3.1: Urban Regular site.

Figure 3.2: Urban Open Place.

Figure 3.3: Telenor Headquarters: work zone.

Figure 3.4: Telenor Headquarters: work zone.

Figure 3.5: Telenor Headquarters: "Atrium".

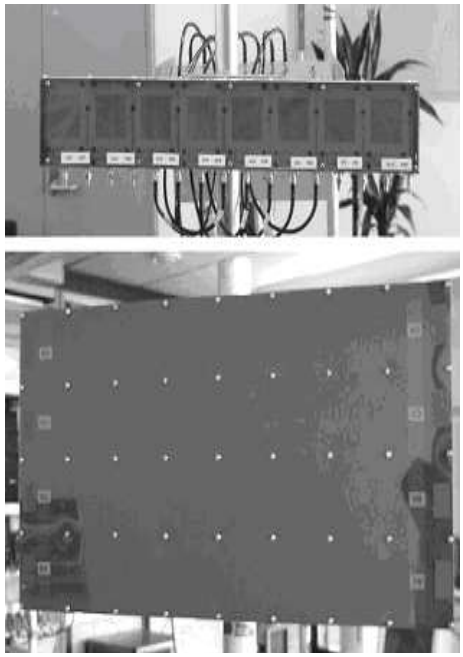

Figure 3.6: Antenna: The receiver is the 8×4 planer array while the transmitter is the 8 element uniform linear array

3.2 2.1 GHz

3.2.1 Channel sounder

The transmitter is an 8 element uniform linear array (ULA) while the receiver antenna is an 8×4 planar array, i.e. two dimensional with 8 elements horizontally and 4 vertically, giving a total of 32 elements (see Figure 3.6). In all the cases, the receiver acted as the base station and only 8 elements were used. The transmitter antenna was connected using the 4 center elements with both polarizations.

The main channel sounder specification are listed on the following Table 3.1. The sounder was manufactured by SINTEF Telecom and Informatics in Trondheim, Norway, on assignment from Telenor.

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 3.1: Channel sounder specification at 2.1 GHz.

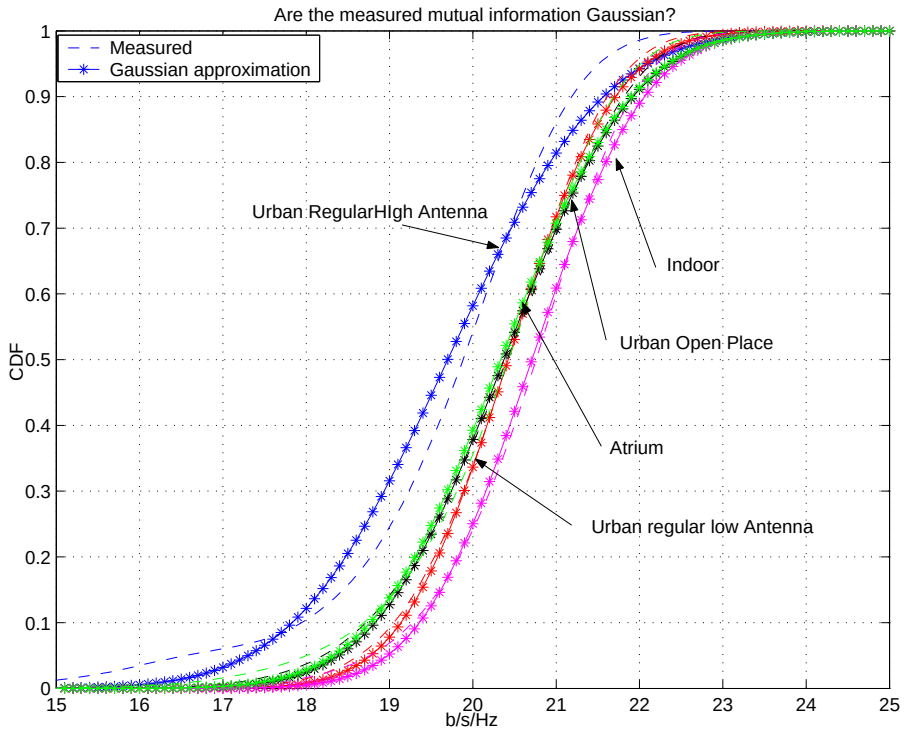

Figure 3.7: Are the measured mutual information Gaussian at 2.1 GHz?

3.2.2 Are the measured mutual information Gaussian?

In Figure 3.7, 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 cdf (cumulative density function) 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:

$$F(C) = 1 - Q\left(\frac{C - t\mu_{\text{empirical}}}{\sigma_{\text{empirical}}}\right)$$

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

As one can see, the mutual information has a Gaussian behavior in every case except for the Urban Regular High Antenna case (we think that this is due to possible error measurements).

3.2.3 What about frequency selectivity?

In the double directional model derived in chapter 2, we argued that frequency selectivity does not affect the mutual information. In Figure 3.8, 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

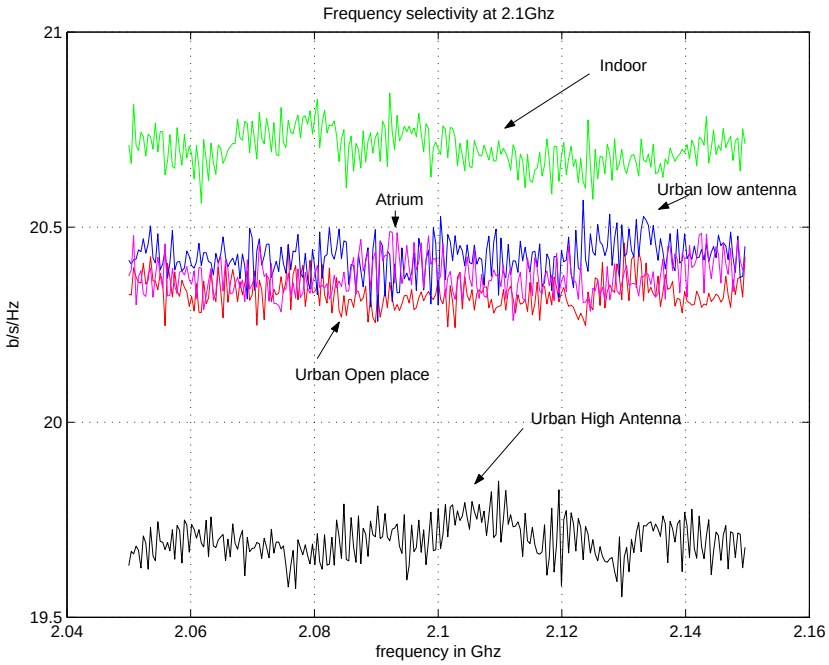

Figure 3.8: Frequency selectivity for many scenarios at 2.1 GHz

scenario[2]. As one can observe, for the different frequencies, the mutual information does not really vary which is adequate with our model structure: the highest variation occurs for the urban regular high antenna position and is about 0.3 b/s/Hz which makes a relative variation of ($\frac{19.85-19.55}{19.55} = 0.015$) around 1.5%.

Note that the mutual information is smaller in the urban regular high antenna position than in the low antenna scenario. This may be due to the fact that there are less scattering objects when the antenna is high. Note also that the mutual information in the indoor scenario is higher than the outdoor case due to a possibly higher number of scattering objects.

3.2.4 Is the model mutual information complying?

In Figure 3.9, we plotted the measured mutual information in the indoor scenario and the mutual information with our model in the simplified case of equal powers of the scatterers. As one can observe, the i.i.d. Gaussian model overestimates the mutual inforamtion whereas the simplified model gives better results. Better compliance could be achieved if one takes the power of the scatterers into account.

3.3 5.25 GHz

3.3.1 Channel sounder

In this case, broadband antennas, vertical patch arrays were used at the transmitter and the receiver. They were designed and manufactured by the Instituto Superior Técnico, Instituto of

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

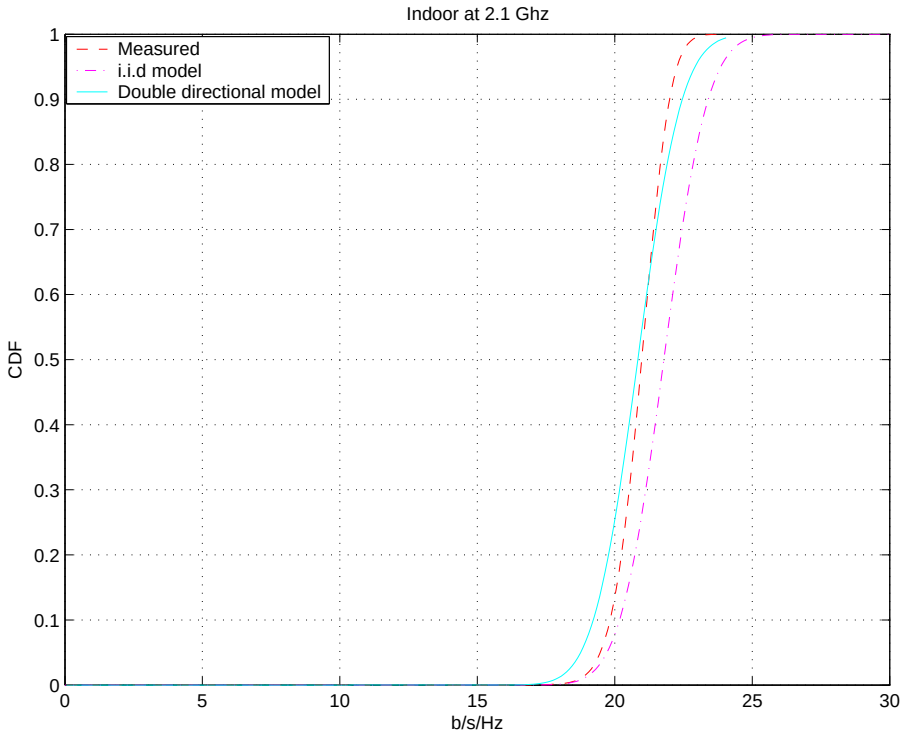

Figure 3.9: Indoor measurements versus model at 2.1 GHz

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 specifications are listed on the following Table (3.2).

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

Table 3.2: Channel sounder specification at 5.255 GHz.

3.3.2 Are the measured mutual information Gaussian?

In Figure 3.11, 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 cdf of each scenario based on the first and second measured moment. As one can see, the mutual information has a Gaussian behavior (more accurate than in the 2.1 GHz campaign).

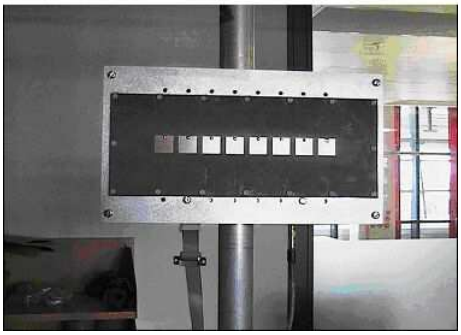

Figure 3.10: Antenna at 5.25 GHz.

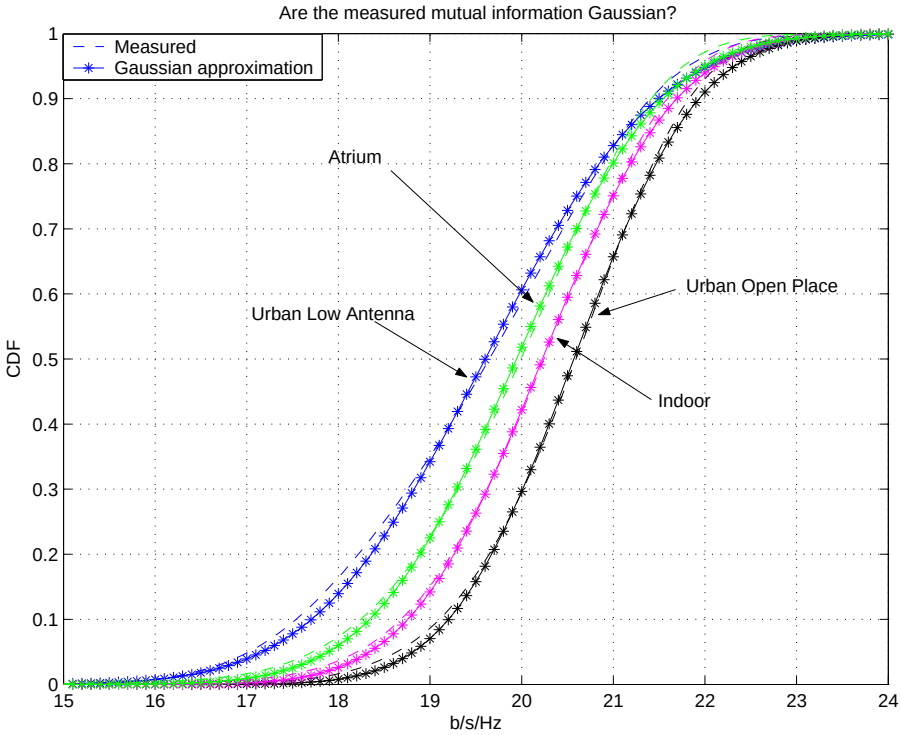

Figure 3.11: Are the measured mutual information Gaussian at 5.2 GHz?

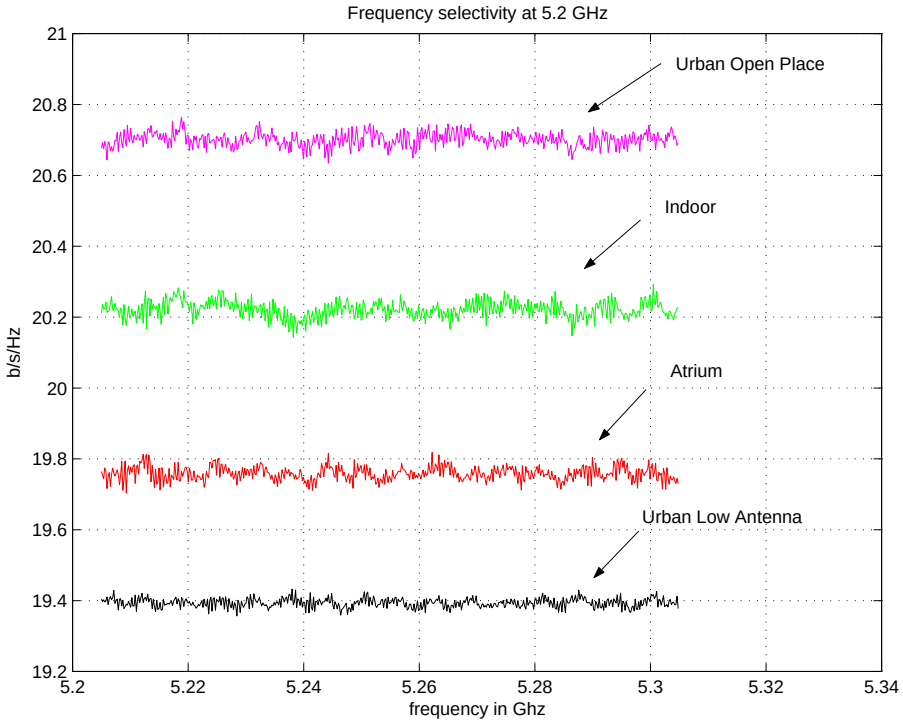

Figure 3.12: Frequency selectivity for various scenarios at 5.2 GHz

3.3.3 What about frequency selectivity?

We previously argued that frequency selectivity does not affect the mutual information. In Figure 3.12, 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%.

3.3.4 Is the model mutual information complying?

In Figure 3.13, we plotted the measured mutual information in the indoor scenario and the mutual information with our model in the simplified case of equal powers of the scatterers. As one can observe, the i.i.d. Gaussian model overestimates the mutual information whereas the simplified model gives better results. Better compliance could be achieved if one takes the power of the scatterers into account.

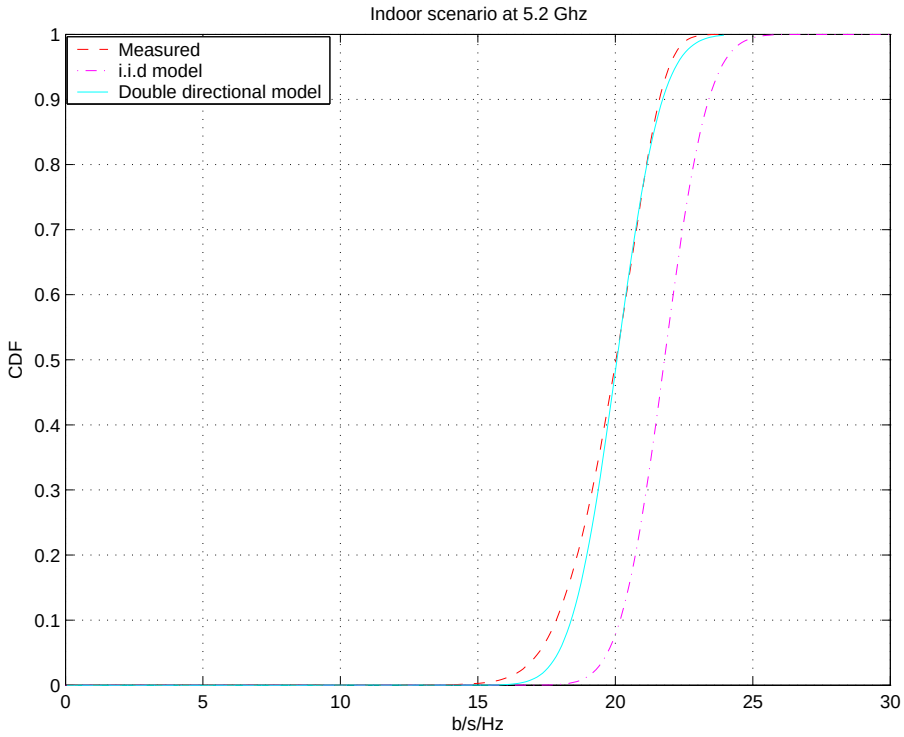

Figure 3.13: Indoor measurements versus model at 5.2 GHz

Chapter 4

Some Metrics

4.1 Mutual Information

Before starting any discussions on MIMO capacity, let us first review the pioneering work of Telatar[1] (later published as [2]) that triggered research in multi-antenna systems[1]. In this paper, Telatar develops the channel capacity of a general MIMO channel. Assuming perfect knowledge of $\mathbf{H}$ at the receiver, the mutual information I^M between input and output is given by[2]:

$$\begin{aligned} I^M(\mathbf{x};(\mathbf{y},\mathbf{H})) &= I^M(\mathbf{x};\mathbf{H}) + I^M(\mathbf{x};\mathbf{y} \mid \mathbf{H})) \\ &= I^M(\mathbf{x};\mathbf{y} \mid \mathbf{H}))^3 \\ &= \text{Entropy}(\mathbf{y} \mid \mathbf{H}) - \text{Entropy}(\mathbf{y} \mid \mathbf{x},\mathbf{H})^4 \\ &= \text{Entropy}(\mathbf{y} \mid \mathbf{H}) - \text{Entropy}(\mathbf{n} \mid \mathbf{H}) \end{aligned}$$

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

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

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

If the $x_i, i = 1...n_t$ are Gaussian[5]:

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

[1] For contribution [2], 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.

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

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

As a consequence[6], the ergodic capacity of an $n_r \times n_t$ MIMO channel with Gaussian entries and covariance matrix $\mathbf{Q}$ ($\mathbf{Q} = \mathbb{E}(\mathbf{x}\mathbf{x}^H)$)[7] is :

$$\overline{C} = \max_{\mathbf{Q}} \mathbb{E}_{\mathbf{H}} \left(C(\mathbf{Q}) \right) \tag{4.1}$$

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

In the original paper [1], Telatar exploits the isotropical property of Gaussian i.i.d $\mathbf{H}$ to show that in this case, ergodic capacity is achieved with $\mathbf{Q} = \mathbf{I}$. However, this result has been proved only for a Gaussian i.i.d channel matrix and was not extended to other types of matrices. In correlated fading, $C(\mathbf{I})$ is called the average mutual information with covariance $\mathbf{Q} = \mathbf{I}$. It has never been proved that capacity was close to this mutual information except for certain particular cases (see [2, 36]). $C(\mathbf{I})$ underestimates the achievable rate[8]: indeed, even though the channel realization is not known, the knowledge of the channel model (is it i.i.d Rayleigh fading? is it i.i.d Rice fading? is it correlated Rayleigh fading with a certain covariance matrix?...) can be taken into account in order to optimize the coding scheme at the transmitter. There is no reason why one should transmit independent substreams on each antenna. It is as if one stated that the space-time codes designed through the rank and determinant criterion [37] were optimal for **all kinds of MIMO channels**.

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

- If the channels are ergodic, $C(\mathbf{Q})$ can be averaged over many channel realizations and the corresponding capacity is defined as : $\overline{C} = \max_{\mathbf{Q}} \mathbb{E}(C(\mathbf{Q}))$
- If the channels are static, there is only one channel realization and an outage probability for each positive rate of transmission can be defined:[9]

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

Once again, when deriving the outage capacity, if the channel distribution is known, then it is possible at the transmitter side to optimize the covariance of the transmitting signal. However,

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

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

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

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

this is not an obvious task[10] and in all the following we will therefore derive the outage mutual information with Gaussian input covariance matrix $\mathbf{Q} = \mathbf{I}$ (**and not the outage capacity!**).

Many results have already been derived on the ergodic capacity of channels based on different channel models taking into account correlation [24, 38, 39, 40] or not [1]. However, very few have been devoted to the outage capacity [41, 42, 43] or deriving the capacity distribution [44, 45].

Recently, using tools from random matrix theory, it was shown in [13, 14] that in many cases the mutual information of MIMO models have asymptotically a Gaussian behavior. Random matrices were first proposed by Wigner in quantum mechanics to explain the measured energy levels of nuclei in terms of the eigenvalues of random matrices. When Telatar [1] (in the context of antenna capacity analysis) and then nearly simultaneously Tsé & Hanly [46] and Verdu & Shamai [47] (for the analysis of uplink unfaded CDMA equipped with certain receivers) introduced random matrices, the random matrix theory entered the field of telecommunications[11]. From that time, random matrix theory has been successively extended to other cases such as uplink CDMA fading channels [49] OFDM [50], downlink CDMA [51], multi-user detection [52]. One of the useful features of random matrix theory is the ability to predict, under certain conditions, the behavior of the empirical eigenvalue distribution of products or sums of matrices. The results are striking in terms of closeness to simulations with reasonable matrix size and enable to derive linear spectral statistics for these matrices with only few meaningful parameters.

Main Results. In the case of the i.i.d. Gaussian channel, the following theorem holds:

Theorem 1 *With the Gaussian i.i.d. model, as* $n_t \to \infty$ *with* $n_r = \gamma n_t$, $\ln\det\left(\mathbf{I}_{n_r} + \frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H\right) - n_t\mu(\gamma,\rho)$ *converges in distribution to a Gaussian* $N(0,\sigma^2(\gamma,\rho))$ *random variable where*[12]*:*

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

and

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

with

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

Proof 1 *We give here only a sketch of the proof for deriving the value of* μ_{iid} *which corresponds*

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

[11] It should be noted that in the field of array processing, Silverstein already used in 1992 random matrix theory [48] for signal detection and estimation.

[12] ln is the natural logarithm such as $\ln(e) = 1$. When this notation is used, the mutual information is given in nats/s. When the notation $\log_2(x) = \frac{\ln(x)}{\ln(2)}$ is used, the results are given in bits/s.

to the asymptotic mutual information per transmitting antenna:

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

The second equality comes from the determinant identity $det(I+AB) = det(I+BA)$. λ_i are the eigenvalues of matrix $\frac{1}{n_t}\mathbf{H}^H\mathbf{H}$ and $F_{n_t}(\lambda)$ is the empirical eigenvalue distribution of matrix $\frac{1}{n_t}\mathbf{H}^H\mathbf{H}$ defined by: $dF_{n_t}(\lambda) = \frac{1}{n_t}\sum_{i=1}^{n_t}\delta(\lambda-\lambda_i)$. It is now well established that the empirical eigenvalue distribution $F_{n_t}(\lambda)$ converges weakly to a non-random distribution (see chapter 7) defined by:

$$f_{iid}(x) = (1-\gamma)\delta(x) + \frac{1}{2\pi x}\sqrt{((1+\sqrt{\gamma})^2 - x)(x-(1-\sqrt{\gamma})^2)}$$

defined in the interval $[(1-\sqrt{\gamma})^2, (1+\sqrt{\gamma})^2]$.

The asymptotic mean value is therefore equal to:

$$\begin{aligned}
\mu_{iid}(\gamma,\rho) &= \int_{(1-\sqrt{\gamma})^2}^{(1+\sqrt{\gamma})^2} \ln(1+\rho\lambda) f(\lambda) d\lambda \\
&= \gamma \ln(1+\rho-\rho\alpha(\gamma,\rho)) + \ln(1+\rho\gamma-\rho\alpha(\gamma,\rho)) - \alpha(\gamma,\rho)
\end{aligned}$$

with

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

Remark 1 *For the double directional model, it can be shown that the mean mutual information scales at high SNR as:*

$$\mathbb{E}(C(\mathbf{I})) = min\left(n_t, n_r, s_r \int_{\lambda>0} dS_{doa}(\lambda), s_t \int_{\lambda>0} dS_{dod}(\lambda)\right) log_2(\rho)$$

$\int_{\lambda>0} dS_{doa}(\lambda))$ and $\int_{\lambda>0} dS_{dod}(\lambda))$ express the correlation factor of the s_r and s_t scatterers respectively. The factor $min\left(n_t, n_r, s_r \int_{\lambda>0} dS_{doa}(\lambda), s_t \int_{\lambda>0} dS_{dod}(\lambda)\right)$ is also known as the multiplexing gain.

Interestingly, the environment is the limiting factor (through the scatterers and their correlation). As a consequence, there is no use in putting more antennas than the number of degrees of freedom the environment is able to provide.

Deriving the outage mutual information: let q denote the outage probability and ${I^M}_q$ the corresponding outage mutual information with covariance $\mathbf{Q} = \mathbf{I}$, then:

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

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

We define[13]

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

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

4.2 MMSE SINR Considerations

The MMSE receiver has several attributes that makes it appealing for use. The MMSE receiver is known to generate a soft decision output which maximizes the output Signal to interference plus Noise ratio (see [53]) (whereas in a AWGN channel with no interference, the match filter maximizes the output SNR). This advantage combined with the low complex implementation of the receiver (due in part to its linearity) has triggered the search for other MMSE based receivers such such as the MMSE DFE (Minimum Mean Square Decision Feedback Equalizer). The MMSE DFE [54, 55] is at the heart of very famous schemes such as BLAST [56]. BLAST which stands for Bell Labs Layered Space Time, was invented by Foschini at Bell Labs in 1998. It is highly related to spatial multiplexing introduced in a 1994 Standford University patent by A. Paulraj [57]: the idea is based on successive interference cancellation where each layer is decoded, re-encoded and subtracted from the transmitted signal. This approach has triggered a lot of implementation research schemes as it was shown by Varanasi and Guess[58][14] to be optimal. The MMSE receiver is therefore at the heart of many schemes and studying the SINR distribution enables to design and understand the performance of many MMSE based receivers (MMSE DFE [54, 55, 59], MMSE Parallel Interference cancellation [60])

As far as the MMSE SINR is concerned and based on the model 2.2, the output of the MMSE

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

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

detector $\hat{\mathbf{x}} = [\hat{x}_1, \ldots, \hat{x}_{n_t}]^T$ is given by

$$\begin{aligned} \hat{\mathbf{x}} &= E\left(\mathbf{x}\mathbf{y}^H\right) \left[E(\mathbf{y}\mathbf{y}^H)\right]^{-1} \mathbf{y} && (4.2) \\ &= \sqrt{\frac{\rho}{n_t}}\mathbf{H}^H \left(\frac{\rho}{n_t}\mathbf{H}\mathbf{H}^H + \mathbf{I}_{n_r}\right)^{-1} \mathbf{y} && (4.3) \\ &= \sqrt{\frac{\rho}{n_t}}\mathbf{H}^H (\mathbf{A})^{-1} \mathbf{y} \ . && (4.4) \end{aligned}$$

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

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

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

where

$$\eta_{\mathbf{h}_k} = \frac{\rho}{n_t}\mathbf{h}_k^H (\mathbf{A})^{-1} \mathbf{h}_k \ . \tag{4.5}$$

and

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

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

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

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

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

Writing $\mathbf{HH}^H = \mathbf{UU}^H + \mathbf{h}_k\mathbf{h}_k^H$ and invoking the matrix inversion lemma[15], we get after some simple algebra another useful expression for this SINR (see e.g. [46]):

$$\mathrm{SINR}^k = \mathbf{h}_k^H \left(\mathbf{UU}^H + \frac{n_t}{\rho}\mathbf{I}_{n_r} \right)^{-1} \mathbf{h}_k \ . \tag{4.6}$$

For a fixed n_r and n_t, it is extremely difficult to get insight on the performance of the MMSE receiver from the expressions (4.6) and (4.5). As a consequence, in order to obtain interpretable expressions, we will focus on an asymptotical analysis of the SINR. Moreover, it has been shown in [61] and [62] that the additive noise τ_k can be considered as Gaussian when n_t and n_r are large enough. In this case, τ_k is an asymptotically zero mean Gaussian noise of variance and one can easily derive performance measures such as BER or spectral efficiency with MMSE equalization. More precisely, the following result can be proven:

Main Results. We recall some important results proved by Tse et al. [63]:

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

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

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

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

and

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

This theorem extends results of Tse et al. [46][16] where it is shown that in a large system limit, the SINR of a user at the output of the MMSE receiver converges to a deterministic limit. The proof uses ideas of Silverstein [64]. In fact, for many linear receivers (MMSE, matched filter, decorrelator), asymptotic normality of the receiver output (in the general the ouput decision statistic) can be proven [65, 62].

Outage BER. Hence, let q denote the outage probability and p_q the corresponding outage probability of error with QPSK uncoded constellations. In this case, we have:

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

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

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

Therefore,

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

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

4.3 Mutual Information and MMSE

The MMSE receiver plays a central role in telecommunications. Recently, it was shown that the MMSE estimator, which has its roots in Signal Processing, is also fundamental in Information Theory. Nice discussions on the Shannon [66, 67] versus Wiener [68, 69] legacy are given by Forney [70] and Guo [71].

It may seem strange that it took more than 50 years to discover quite fundamental relationships between the input output mutual information and the minimum mean square error of an estimate. Astonishingly, it is shown in [71] that the derivative of the mutual information (nats) in the SISO case with respect to the SNR is equal to half the MMSE. In the MIMO case, similar results can be proven [72].

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

$$\mathbf{R}_y = \mathbb{E}(\mathbf{y}\mathbf{y}^H) = \frac{\rho}{n_t}\mathbf{H}\mathbf{Q}\mathbf{H}^H + \mathbf{I} \tag{4.7}$$

$$\mathbf{R}_{xy} = \mathbb{E}(\mathbf{x}\mathbf{y}^H) = \sqrt{\frac{\rho}{n_t}}\mathbf{Q}\mathbf{H}^H \tag{4.8}$$

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

$$\hat{\mathbf{x}} = \mathbf{R}_{xy}{\mathbf{R}_y}^{-1}\mathbf{y} \tag{4.9}$$

and the covariance matrix of the MMSE receiver is:

$$\begin{aligned}
\mathbf{R}_{\text{MMSE}} &= \mathbb{E}\left[(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})^H\right] && (4.10)\\
&= \mathbf{R}_{xy}{\mathbf{R}_y}^{-1}\mathbf{R}_{xy}^H - 2\mathbf{R}_{xy}\mathbf{R}_y^{-1}\mathbf{R}_{xy}^H + \mathbf{Q} && (4.11)\\
&= \mathbf{Q} - \mathbf{R}_{xy}{\mathbf{R}_y}^{-1}\mathbf{R}_{xy}^H && (4.12)
\end{aligned}$$

It follows from the inversion lemma that:

$$\begin{aligned}
\mathbf{R}_{\text{MMSE}}^{-1} &= \mathbf{Q}^{-1} + \mathbf{Q}^{-1}\mathbf{R}_{xy}\left(\mathbf{R}_y - \mathbf{R}_{xy}^H\mathbf{Q}^{-1}\mathbf{R}_{xy}\right)^{-1}\mathbf{R}_{xy}^*\mathbf{Q}^{-1} && (4.13)\\
&= \mathbf{Q}^{-1} + \frac{\rho}{n_t}\mathbf{H}^H\left(\mathbf{R}_y - \mathbf{H}\mathbf{Q}\mathbf{H}^*\right)^{-1}\mathbf{H} && (4.14)\\
&= \mathbf{Q}^{-1} + \mathbf{H}^H\mathbf{H} && (4.15)
\end{aligned}$$

Finally, the capacity is given by:

$$\begin{aligned} C &= \log_2 \det\left(\mathbf{I} + \frac{\rho}{n_t}\mathbf{H}\mathbf{Q}\mathbf{H}^H\right) && (4.16)\\ &= \log_2 \det\left(\mathbf{I} + \frac{\rho}{n_t}\mathbf{H}^H\mathbf{H}\mathbf{Q}\right) && (4.17)\\ &= \log_2 \det\left(\mathbf{Q}\right) + \log_2 \det\left(\mathbf{Q}^{-1} + \mathbf{H}^H\mathbf{H}\right) && (4.18) \end{aligned}$$

Hence, the channel capacity can be rewritten:

$$C = \log_2 \frac{\det\left(\mathbf{Q}\right)}{\det\left(\mathbf{R}_{\mathrm{MMSE}}\right)} \quad (4.19)$$

The expression relates, in a simple manner, the channel capacity to the covariance matrix of the MMSE estimate of $\mathbf{x}$. In this form, the channel capacity formula has an intuitive appeal. In fact, the MMSE estimate $\hat{\mathbf{x}}$ lies (with high probability) in a "small cell" centered around the codeword $\mathbf{x}$. The volume of the cell is proportional to $\det(\mathbf{R}_{\mathrm{MMSE}})$. The volume of the codebook space (in which $\mathbf{x}$ lies with high probability) is proportional to $\det(\mathbf{R}_{\mathbf{x}})$. The ratio $\rho = \frac{\det(\mathbf{R}_{\mathbf{x}})}{\det(\mathbf{R}_{\mathrm{MMSE}})}$ gives the number of cells that can be packed into the codebook space without significant overlapping. The "center" of each such cell, the codeword, can be reliably detected, for instance, using $\hat{\mathbf{x}}$. As a consequence, one can communicate reliably using a codebook of size ρ, which contains $\log_2(\rho)$ information bits. This provides an intuitive motivation to the capacity formula, in the same vein as [73].

Chapter 5

Transmission/Reception with Multiple Antennas

In this chapter we introduce some techniques based on multiple antenna transmission and reception that have gained considerable attention in the last few years [74, 75, 2]. Classical reception diversity techniques are a longstanding well-known method to improve the quality of a wireless link, and have been thoroughly analyzed (see [76, 3, 77, 78, 79, 12] and references therein). More recently, transmission with multiple antennas has been recognized as perhaps the most efficient technique to improve the wireless channel spectral efficiency, at the expense of additional complexity in the transmitter and in the receiver.

Space-time coding, i.e., sending encoded signals through several transmit antennas, can achieve at the same time an increase in the system diversity, in order to combat the effect of fading, and an increase in the transmission spectral efficiency by creating effectively a set of virtual parallel channels between the transmitter and the receiver.

5.1 Antenna diversity reception

Consider two identical receivers at relative distance d. Let $c_1(t,\tau)$ and $c_2(t,\tau)$ be the fading channels characterizing the propagation between the transmitter and receivers 1 and 2, respectively. The correlation between the fading channels $c_1(t,\tau)$ and $c_2(t,\tau)$ depends on the geometry of the scattering elements and on the radiation patterns of transmit and receive antennas. For simplicity, we shall assume an isotropic radiation pattern for the receiving antennas and the planar scattering geometry of Fig. 5.1, where both antennas receive scattered plane waves with uniform probability from all angles $\alpha \in [-\pi, \pi)$. We assume that d is much smaller than the distance between the antennas and the scattering elements. Then, all scattered signals are received with the same delay τ_0 (the channel is frequency non-selective, since there is only one path). Moreover, the fading Doppler spectrum is of Jakes' type, with time autocorrelation function $J_0(2\pi F_d \Delta t)$, with $F_d = f_c v/c$ being the maximum Doppler shift, where f_c and v are the carrier frequency and the relative speed between the scatterers and the receiving antennas.

Let $\lambda = c/f_c$ denote the carrier wavelength, and let $\Delta t = d/v$ be the time necessary to move from antenna 2 to antenna 1, at speed v. Then, the corresponding *space* autocorrelation

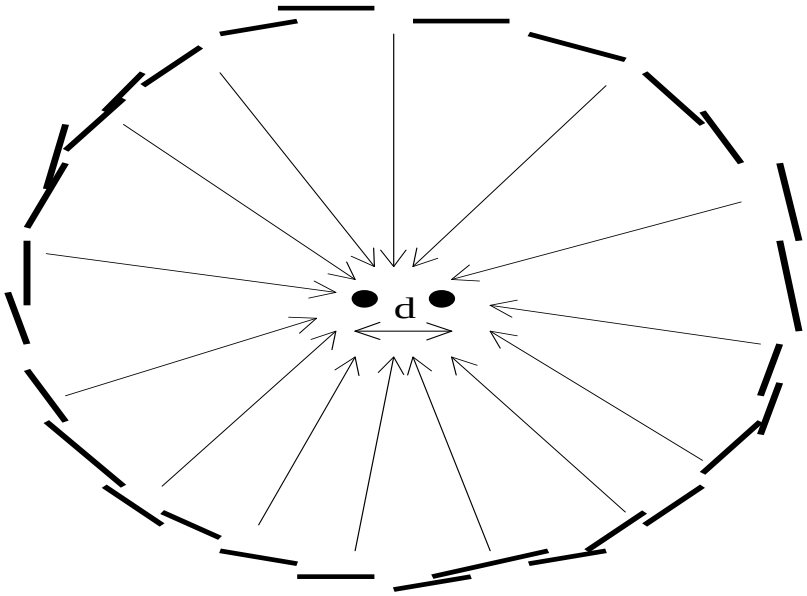

Figure 5.1: Isotropic planar scattering geometry.

function of fading experienced by the two antennas is

$$\begin{aligned} E[c_1(t)c_2(t)^*] &= E[c_1(t)c_1(t-\Delta t)^*] \\ &= J_0(2\pi F_d \Delta t) \\ &= J_0(2\pi d/\lambda) \end{aligned} \tag{5.1}$$

where we have used the fact that $c_2(t) = c_1(t - \Delta t)$. From Fig. 5.2, representing the function $J_0(x)$, we see that for d/λ of the order of some unit, the fading at the two antennas is virtually uncorrelated.

The above consideration motivates *diversity* reception based on multiple antennas: multiple observations of the transmit signal can be obtained from multiple receiving antennas. Each observation is characterized by different statistically independent fading. Hence, the probability that all channels fade at the same time is small and the chances that at least some of the observations have a good instantaneous SNR is increased by increasing the number of receiving antennas. In the following, we make this statement more precise by considering an idealized simplified model, with r receiving antennas with mutually independent frequency non-selective fading.

The discrete-time signal received at the i-th antenna is given by (see (**??**))

$$y_{i,n} = \sqrt{\mathcal{E}} c_{i,n} a_n + \nu_{i,n}$$

where $\nu_{i,n} \sim \mathcal{N}_{\mathbb{C}}(0, N_0)$. Noises at different antennas are mutually independent. At this point, the multiple observations obtained from the r receiving antennas are totally analogous to the

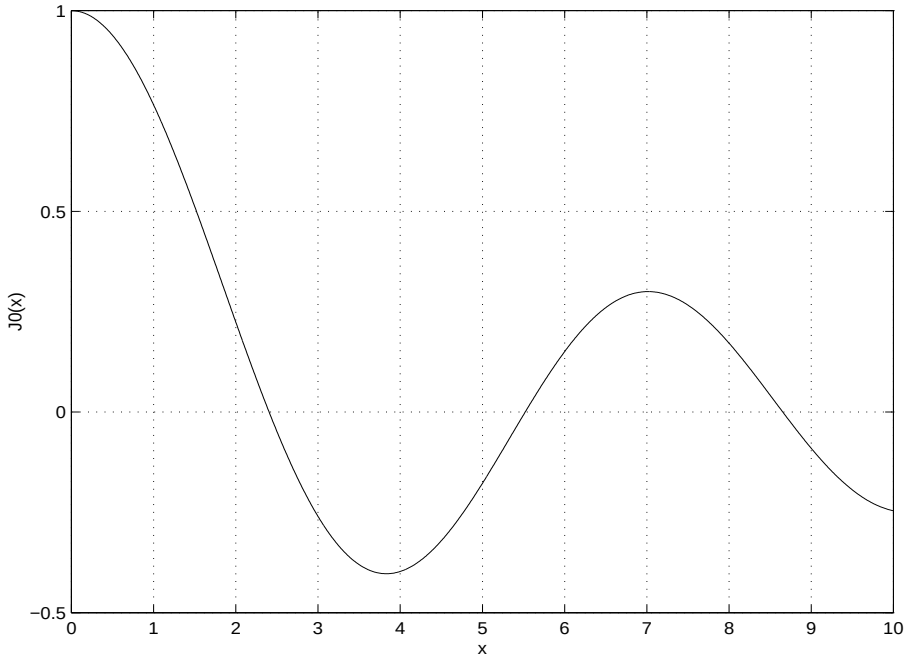

Figure 5.2: Bessel $J_0(x)$ function.

multiple observations obtained from the fingers of a rake receiver (this is seen immediately by replacing the antenna index i by the path index p). The vector observation $\mathbf{y}_n = (y_{1,n}, \ldots, y_{n_r,n})^T$ can be written as

$$\mathbf{y}_n = \sqrt{\mathcal{E}}\mathbf{c}_n a_n + \boldsymbol{\nu}_n \tag{5.2}$$

where we define $\mathbf{c}_n = (c_{1,n}, \ldots, c_{n_r,n})^T$ and $\boldsymbol{\nu}_n = (\nu_{1,n}, \ldots, \nu_{n_r,n})^T$. The optimal receiver consists of a Maximal-Ratio Combiner (MRC) followed by either a symbol-by-symbol detector (uncoded modulation) or a soft decoder.

After normalization, the output of the MRC is given by

$$y_n = \sqrt{\mathcal{E}}\alpha_n a_n + v_n$$

where $\alpha_n = \sqrt{\sum_{i=1}^{n_r} |c_{i,n}|^2}$ and $v_n \sim \mathcal{N}_{\mathbb{C}}(0, N_0)$.

The performance of uncoded modulation with MRC can be calculated as for the rake receiver, either by using the technique based on HQF-GRV, or by using the technique based on the integral expression of the Gaussian Q-function.

The quality of wireless links with very slowly-varying fading (quasi-static fading) is often characterized by the probability that the SNR at the detector input falls below a certain threshold β_0, that depends on the coding and modulation scheme used. The SNR threshold determines the required quality-of-service (QoS) of the link. If SNR $< \beta_0$, the required QoS is not met and we say that the link is in *outage* state. The outage probability $P_{\text{out}} = P(\text{SNR} < \beta_0)$ is directly obtained by the SNR cdf, evaluated in β_0.

5.1.1 SNR cdf at the output of the MRC

For simplicity, we drop the time index n. Suppose independent normalized Ricean fading for all antennas, i.e., $c_i \sim \mathcal{N}_{\mathbb{C}}(\sqrt{K/(1+K)}, 1/(1+K))$. Then, the SNR

$$\mathrm{SNR}_{\mathrm{mrc}} = \frac{\mathcal{E}}{N_0} \sum_{i=1}^{n_r} |c_i|^2$$

is distributed according to the non-central Chi-squared pdf with $2r$ degrees of freedom

$$f_{\mathrm{mrc}}(z) = \frac{1+K}{\gamma} \left(\frac{z(1+K)}{\gamma n_r K} \right)^{(n_r-1)/2} \exp\left(-(1+K)\left(\frac{n_r K}{1+K} + \frac{z}{\gamma} \right) \right) I_{n_r-1}\left(2\sqrt{\frac{z n_r K(1+K)}{\gamma}} \right)$$

where we let $\gamma = \mathcal{E}/N_0$ and where $I_{n_r-1}(x)$ is the modified Bessel function of the first kind of order $n_r - 1$.

For $K = 0$ (Rayleigh fading), the SNR pdf becomes a central Chi-squared with $2n_r$ degrees of freedom

$$f_{\mathrm{mrc}}(z) = \frac{1}{\gamma} \frac{1}{(n_r-1)!} \left(\frac{z}{\gamma} \right)^{n_r-1} e^{-z/\gamma}$$

For $K \to \infty$, the channel becomes AWGN, and the SNR converges in distribution to the constant value $n_r\gamma$.

The SNR cdf, for independent identically distributed Ricean fading, is given by

$$F_{\mathrm{mrc}}(z) = 1 - Q_r\left(\sqrt{2n_r K}, \sqrt{\frac{2z(1+K)}{\gamma}} \right)$$

where $Q_{n_r}(a,b)$ is the generalized Marcum Q-function [80], defined by

$$Q_{n_r}(a,b) = \int_b^\infty x(x/a)^{n_r-1} e^{-(x^2+a^2)/2} I_{n_r-1}(ax)dx$$

The cdf of the output SNR for independent identically distributed Rayleigh fading is given by

$$F_{\mathrm{mrc}}(z) = 1 - e^{-z/\gamma} \sum_{\ell=0}^{n_r-1} \frac{1}{\ell!} \left(\frac{z}{\gamma} \right)^{\ell}$$

5.1.2 Selection combining

A suboptimal but simpler combining scheme consists of selecting the output of the antenna with largest instantaneous SNR. This method is called **selection combining**. With selection combining, the detector input is given by

$$y_n = c_{\max,n} a_n + \nu_n$$

where $c_{\max,n}$ is defined as the fading coefficient with maximum squared magnitude, over all $i = 1, \ldots, n_r$, for given time n, and where $\nu_n \sim \mathcal{N}_{\mathbb{C}}(0, N_0)$. The resulting output SNR is given by (we drop again the time index)

$$\mathrm{SNR}_{\mathrm{sel}} = \frac{\mathcal{E}}{N_0} |c_{\max}|^2$$

This technique does not "combine" the multiple observations, however it can be seen as a linear combiner with suboptimal combining coefficients equal either to 0 or to 1.

If the fading coefficients of all antennas are i.i.d., we can write the SNR cdf as

$$\begin{aligned} F_{\text{sel}}(z) &= P(\text{SNR} \leq z) \\ &= P\left(\gamma \max_{i=1,\ldots,n_r} |c_i|^2 \leq z\right) \\ &= P\left(|c_1|^2 \leq z/\gamma, \ldots, |c_{n_r}|^2 \leq z/\gamma\right) \\ &= \prod_{i=1}^{n_r} P\left(|c_i|^2 \leq z/\gamma\right) \\ &= F_g(z/\gamma)^{n_r} \end{aligned} \tag{5.3}$$

where g denotes the power gain of any of the n_r channels (they are identically distributed) and $F_g(z)$ denotes its cdf.

The SNR pdf is obtained by deriving its cdf with respect to z:

$$f_{\text{sel}}(z) = \frac{n_r}{\gamma} \left(F_g(z/\gamma)\right)^{n_r - 1} f_g(z/\gamma)$$

For example, in the case of normalized Rayleigh fading we have

$$f_{\text{sel}}(z) = \frac{n_r}{\gamma} \sum_{\ell=0}^{n_r-1} \binom{n_r - 1}{\ell} (-1)^\ell e^{-z(\ell+1)/\gamma}$$

for $z \geq 0$, and $f_{\text{sel}}(z) = 0$ for $z < 0$.

Fig. 5.3 shows the SNR pdf's for MRC and Selection Combining in the case of Rayleigh fading with $n_r = 1, 2, 4, 8$ independent antennas.

5.2 Spatial multiplexing: the V-BLAST scheme

The idea of *spatial multiplexing* is very simple: n_t independent data streams are sent in parallel, at the same time, from n_t transmit antennas. If the number of receiving antennas n_r is sufficiently large, the data streams can be separated at the receiver and independently detected successfully. Eventually, the n_t-input n_r-output MIMO channel behaves as a set of n_t parallel channels, thus increasing the effective transmission rate by a factor n_t.

The first proposed and best known scheme for spatial multiplexing is the so called V-BLAST (Vertical Bell-Labs Space-Time) system [56]. Since the scheme applies to uncoded transmission, we drop once again the time index and let $\mathbf{x} = (x_1, \ldots, x_{n_t})^T$ be the vector of modulation symbols (taking values into a signal set $\mathcal{A}$) to be transmitted at time n from the t transmit antennas and $\mathbf{H} \in \mathbb{C}^{n_r \times n_t}$ be the channel matrix such that $[\mathbf{H}]_{i,j} = h_{i,j}$ is the complex fading coefficient between transmit antenna j and receiving antenna i. Then, the received signal vector is given by

$$\mathbf{y} = \sqrt{\mathcal{E}/n_t}\mathbf{H}\mathbf{x} + \boldsymbol{\nu} \tag{5.4}$$

Notice that the energy term is divided by n_t since in order to keep the total transmit energy per symbol interval equal to $\mathcal{E}$, with n_t symbols sent in parallel over the n_t antennas, each symbol must have average energy $\mathcal{E}/n_t$.

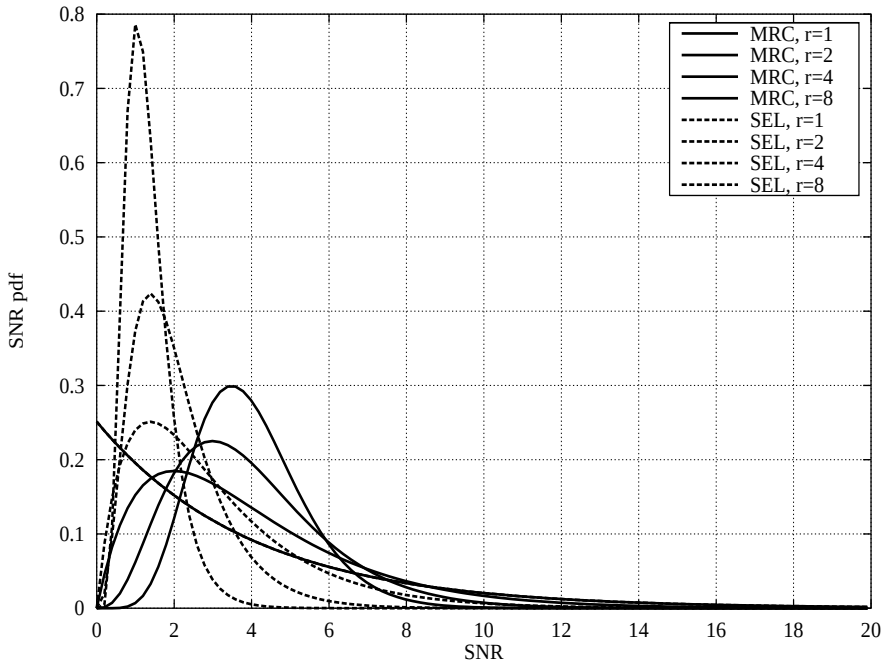

Figure 5.3: Output SNR pdf for MRC and Selection Combining for independent Rayleigh fading.

The ML detection of the symbol vector $\mathbf{x}$ when the receiver has perfect knowledge of $\mathbf{H}$ is immediately given by

$$\widehat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathcal{A}^t} \left| \mathbf{y} - \sqrt{\mathcal{E}/n_t}\mathbf{H}\mathbf{x} \right|^2 \tag{5.5}$$

The above detection rule might be too complex for large n_t and/or large constellation size $|\mathcal{A}|$. In fact, this requires the evaluation of $|\mathcal{A}|^{n_t}$ squared Euclidean distances. By comparing (5.4) with the vectored form of the fading ISI channel (**??**) we notice that the two problems are formally analogous. Therefore, decision-feedback as well as linear equalization can be used as a suboptimal but simpler detection method.

By replicating the derivation of the MMSE-DFE and of the ZF-DFE we obtain the following receiver schemes, commonly known as "nulling and canceling", since for each symbol in $\mathbf{x}$, the already detected symbols are subtracted while the undetected interfering symbols are filtered ("nulled") out by the receiver.

MMSE-DFE V-BLAST receiver. Assume that symbol detection is performed in the "bottom-up" order, i.e., $n_t, n_t - 1, \ldots, 1$. Then, if symbols $n_t, \ldots, i+1$ have been successfully detected, they can be subtracted exactly form the received signal $\mathbf{y}$, so that the signal after subtraction is given by

$$\mathbf{y} - \sqrt{\mathcal{E}/n_t} \sum_{\ell=i+1}^{n_t} \mathbf{h}_\ell x_\ell = \underbrace{\sqrt{\mathcal{E}/n_t}\mathbf{h}_i x_i}_{\text{useful signal}} + \underbrace{\sqrt{\mathcal{E}/n_t} \sum_{\ell=1}^{i-1} \mathbf{h}_\ell x_\ell}_{\text{interference}} + \underbrace{\boldsymbol{\nu}}_{\text{noise}} \tag{5.6}$$

where we let $\mathbf{h}_\ell$ denote the ℓ-th row of $\mathbf{H}$ and where we have put in evidence the useful signal for the detection of x_i. The (unbiased) MMSE filter for the detection of x_i with observation given by (5.6) is given by

$$\mathbf{f}_i = \frac{\sqrt{\mathcal{E}/n_t}}{\mu_i}\mathbf{\Sigma}_i^{-1}\mathbf{h}_i \tag{5.7}$$

where

$$\mathbf{\Sigma}_i = N_0\mathbf{I} + \frac{\mathcal{E}}{n_t}\sum_{\ell=1}^{i-1}\mathbf{h}_\ell\mathbf{h}_\ell^H$$

and where $\mu_i = \frac{\mathcal{E}}{n_t}\mathbf{h}_i^H\mathbf{\Sigma}_i^{-1}\mathbf{h}_i$ is the instantaneous SINR conditioned with respect to the channel matrix at the input of the detector for symbol i.

Hence, the V-BLAST MMSE detection scheme works as follows: for $i = n_t, n_t-1, \ldots, 1$, compute the MMSE estimate of symbol x_i as

$$z_i = \mathbf{f}_i^H\left(\mathbf{y} - \sqrt{\mathcal{E}/n_t}\sum_{\ell=i+1}^{n_t}\mathbf{h}_\ell\widehat{x}_\ell\right)$$

where $\widehat{x}_\ell$ denote the symbol decisions already available. Then, treat z_i as if it was the output of a virtual Gaussian noise channel with SNR equal to μ_i, and detect symbol x_i as

$$\widehat{x}_i = \arg\min_{a\in\mathcal{A}}|z_i - a|^2$$

As far as performance analysis is concerned, unfortunately even for the simple case where $\mathbf{H}$ has i.i.d. elements $\sim \mathcal{N}_{\mathbb{C}}(0,1)$ the statistics of the SINR μ_i is not known analytically, so that it is impossible to compute in closed form the symbol error probability $\Pr(\widehat{x}_i \neq x_i)$ (even assuming no error propagation in the decision feedback).

ZF-DFE V-BLAST receiver. Assume that $n_r \geq n_t$. Then, for any continuous fading distribution the channel matrix $\mathbf{H}$ has rank n_t with probability 1. In this case, we apply the QR decomposition $\mathbf{H} = \mathbf{QR}$ where $\mathbf{Q}$ is $n_r \times n_t$ such that the columns of $\mathbf{Q}$ are orthonormal, and $\mathbf{R}$ is a $n_t \times n_t$ upper triangular matrix.

By multiplying $\mathbf{y}$ by $\mathbf{Q}^H$ we obtain

$$\mathbf{y}^{(w)} = \sqrt{\mathcal{E}/n_t}\mathbf{Rx} + \mathbf{v} \tag{5.8}$$

where $\mathbf{v} = \mathbf{Q}^H\boldsymbol{\nu}$ is still white with entries $\mathcal{N}_{\mathbb{C}}(0, N_0)$, in fact,

$$E[\mathbf{vv}^H] = E[\mathbf{Q}^H\boldsymbol{\nu}\boldsymbol{\nu}^H\mathbf{Q}] = N_0\mathbf{Q}^H\mathbf{Q} = N_0\mathbf{I}$$

It is possible to show that $\mathbf{y}^{(w)}$ is the analogous of the output of the Whitened Matched Filter (WMF) for ISI channels, generalized to the case where the channel matrix $\mathbf{H}$ is not necessarily Toeplitz.

It is also possible to see easily that the matrix $\mathbf{Q}$ implements the zero-forcing forward filter for the ZF-DFE. Indeed, the i-th component of $\mathbf{y}^{(w)}$ contains only the contribution of symbols $i, \ldots, n_t$. We have

$$y_i^{(w)} = \sqrt{\mathcal{E}/n_t}r_{i,i}x_i + \sqrt{\mathcal{E}/n_t}\sum_{\ell=i+1}^{t} r_{i,\ell}x_\ell + v_i$$

By detecting the symbols in the order $n_t, n_t-1, \ldots, 1$, we can subtract from $y_i^{(w)}$ the decisions about the interfering symbols. The, the V-BLAST ZF detection works as follows: for $i = t, t-1, \ldots, 1$, compute the estimate of x_i as

$$z_i = y_i^{(w)} - \sqrt{\mathcal{E}/n_t} \sum_{\ell=i+1}^{n_t} r_{i,\ell} \widehat{x}_\ell$$

Then, treat z_i as if it was the output of a virtual Gaussian noise channel, and detect symbol x_i as

$$\widehat{x}_i = \arg \min_{a \in \mathcal{A}} \left| z_i - \sqrt{\mathcal{E}/n_t} r_{i,i} a \right|^2$$

As far as analysis in independent Rayleigh fading is concerned, we have the following result:

Lemma. Let $\mathbf{h}_1, \ldots, \mathbf{h}_{i-1}, \mathbf{h}_i$ be n_r-dimensional vectors with i.i.d. entries $\sim \mathcal{N}_{\mathbb{C}}(0,1)$ (we assume $n_r \geq i$). Let $\mathbf{P}$ be the orthogonal projector onto the orthogonal complement of the subspace generated by $\mathbf{h}_1, \ldots, \mathbf{h}_{i-1}$. Then, the random variable $\mu_i = |\mathbf{P}\mathbf{h}_i|^2$ is central Chi-squared with $2(n_r - i + 1)$ degrees of freedom, i.e., its pdf is given by

$$p_{\mu_i}(z) = \frac{z^{n_r - i} e^{-z}}{(n_r - i)!}, \quad z \geq 0 \tag{5.9}$$

Proof. Notice that $\mathbf{P}$ is independent of $\mathbf{h}_i$. Then, by definition of orthogonal projector, $\mathbf{P}$ can be written as $\mathbf{U}\mathbf{U}^H$, where the columns of $\mathbf{U}$ are an orthonormal basis for the orthogonal complement of the subspace generated by $\mathbf{h}_1, \ldots, \mathbf{h}_{i-1}$. With probability 1 this subspace has dimension $n_r - i + 1$, therefore, $\mathbf{g} = \mathbf{U}^H \mathbf{h}_i$ is a $n_r - i + 1$ dimensional Gaussian vector with i.i.d. entries $\sim \mathcal{N}_{\mathbb{C}}(0,1)$. Finally, it is well known that $\mu_i = |\mathbf{g}|^2$ is Chi-squared distributed with $2(n_r - i + 1)$ degrees of freedom. □

Then, notice that the i-th detector input z_i, when there are no errors in the decision feedback, is conditionally Gaussian with instantaneous SNR given by

$$\mathrm{SNR}_i = \frac{\mathcal{E}}{n_t N_0} |r_{i,i}|^2$$

where $r_{i,i} = \mathbf{P}\mathbf{h}_i$, therefore, SNR_i is Chi-squared distributed with $2(n_r - i + 1)$ degrees of freedom. The pairwise error probability assuming no decision feedback errors can be upperbounded by using the technique based on (**??**), and yields

$$P(x_i \to x_i') \leq \frac{1}{2} \left(1 + \frac{\bar{\gamma}}{n_t} \frac{d_i^2}{4} \right)^{-(n_r - i + 1)} \tag{5.10}$$

where, as usual, we let $\mathcal{E}/N_0 = \bar{\gamma}$, and $d^2 = |x_i - x_i'|^2$. We see that the i-th symbol enjoys a diversity order of $n_r - i + 1$. The overall error probability is then dominated by the first detection, has has diversity order $n_r - n_t + 1$.

5.3 Space-time coding

We have seen that spatial multiplexing alone involves some loss in diversity. This is essentially due to the fact that the independent data sequences sent by the t antennas mutually interfere at the receiver. Therefore, the receiver spends some of its degrees of freedom in order to separate the individual sequences and create "parallel channels".

Clearly, the optimal way of transmitting over multiple antennas is to use channel coding across antennas and time. This approach is generally referred to as *Space-Time Coding* (STC) [75]. The system block diagram is represented in Fig. 5.4. An encoder with rate b/n_t bit/complex symbol and code book $\mathcal{C}$ produces n_t modulation symbols every b input information bits. In order to send a code word $\mathbf{c} \in \mathcal{C}$ over the n_t-antenna channel, the code word is formatted as a $n_t \times N$ array by a *Space-Time* parser, and then mapped over some signal set alphabet. Let

$$\mathbf{c} = (c_{1,1}, \ldots, c_{n_t,1}, c_{1,2}, \ldots, c_{n_t,2}, \ldots, c_{1,N}, \ldots, c_{n_t,N})$$

be a code word of $\mathcal{C}$, defined over some discrete symbol alphabet $\mathbb{F}$. The parsed version of $\mathbf{c}$ is the $n_t \times N'$ array $\mathbf{C} = \mathcal{P}(\mathbf{c})$ over $\mathbb{F} \cup \{\emptyset\}$, where $\emptyset$ is a special "null" symbol that is mapped by the modulator onto the zero complex symbol, $\mathcal{P}$ denotes the parsing function and $N' \geq N$. Finally, the modulated signal array to be sent in N' symbol intervals is given by $\mathbf{X} = \mathcal{F}(\mathbf{C})$. STC schemes differ by the underlying code $\mathcal{C}$, the space-time parser and the modulation mapping.

We assume quasi-static frequency non-selective fading, so that the ensemble of the channels going from each transmitter to each receiver, during the time span of a code word, are described by a random (but constant with time) complex $n_r \times n_t$ matrix $\mathbf{H}$. As before, we can write the received signal vector of length n_r during the n-th symbol interval as

$$\mathbf{y}_n = \sqrt{\mathcal{E}/n_t}\mathbf{H}\mathbf{x}_n + \boldsymbol{\nu}_n \tag{5.11}$$

By collecting the signal vectors $\mathbf{y}_n$ for $n = 1, \ldots, N'$ into the $n_r \times N'$ array $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_{N'}]$ we obtain the matrix channel model

$$\mathbf{Y} = \sqrt{\mathcal{E}/n_t}\mathbf{H}\mathbf{X} + \mathbf{N} \tag{5.12}$$

In order to illustrate the utility of STC and to gain insight in the code design, we calculate the Chernoff upper bound on the PEP for space-time codes in the case of coherent detection with perfect CSI.

Since $\mathbf{Y}$ is conditionally Gaussian with independent entries (for given $\mathbf{H}$ and $\mathbf{X}$), the ML decoding rule is given by

$$\widehat{\mathbf{c}} = \arg \min_{\mathbf{c} \in \mathcal{C}} \left\| \mathbf{Y} - \sqrt{\mathcal{E}/n_t}\mathbf{H}\mathcal{F}(\mathcal{P}(\mathbf{c})) \right\|_{\mathrm{F}}^2 \tag{5.13}$$

where $\| \cdot \|_{\mathrm{F}}^2$ denotes the squared Frobenious norm defined by

$$\|\mathbf{A}\|_{\mathrm{F}}^2 = \sum_{i=1}^{n_t} \sum_{n=1}^{N'} |a_{i,n}|^2$$

As usual, the conditional PEP of two code sequences $\mathbf{c}, \mathbf{c}' \in \mathcal{C}$ for given channel matrix $\mathbf{H}$ is given by

$$P(\mathbf{c} \to \mathbf{c}'|\mathbf{H}) = P(\Delta \leq 0)$$

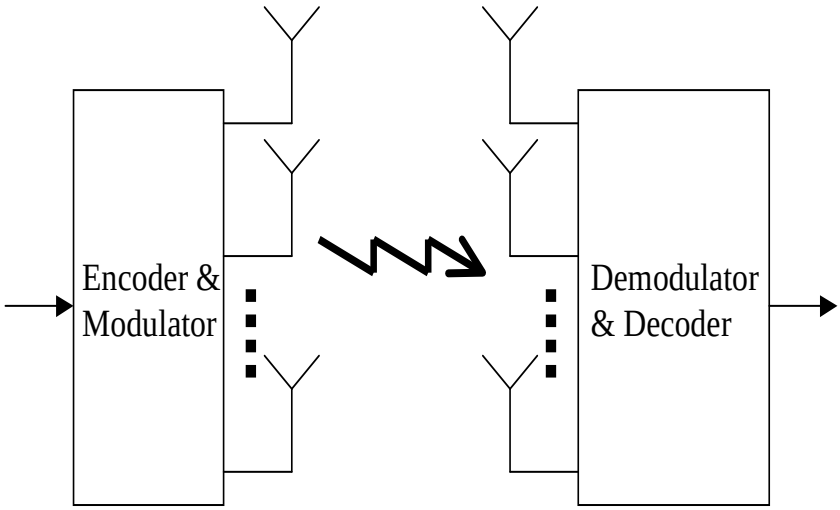

Figure 5.4: Block diagram of a space-time coding system.

where Δ is the decision metric difference, given by

$$\begin{aligned}\Delta &= \sum_{n=1}^{N'} \left|\mathbf{y}_n - \sqrt{\mathcal{E}/n_t}\mathbf{H}\mathbf{x}'_n\right|^2 - \left|\mathbf{y}_n - \sqrt{\mathcal{E}/n_t}\mathbf{H}\mathbf{x}_n\right|^2 \\ &= \frac{\mathcal{E}}{n_t}\sum_{n=1}^{N'} \left|\mathbf{H}(\mathbf{x}'_n - \mathbf{x}_n)\right|^2 - \eta \end{aligned} \tag{5.14}$$

where

$$\eta \sim \mathcal{N}(0, 2N_0\mathcal{E}/n_t \sum_{n=1}^{N'} \left|\mathbf{H}(\mathbf{x}'_n - \mathbf{x}_n)\right|^2)$$

and where we let $\mathbf{X} = \mathcal{F}(\mathcal{P}(\mathbf{c}))$ and $\mathbf{X}' = \mathcal{F}(\mathcal{P}(\mathbf{c}'))$.

Let $D = \sum_{n=1}^{N'} |\mathbf{H}(\mathbf{x}'_n - \mathbf{x}_n)|^2$. The characteristic function of Δ is given by

$$\Phi_\Delta(s) = \exp\left(-(s - s^2N_0)(\mathcal{E}/n_t)D\right) \tag{5.15}$$

whose minimum on the real non-negative axis is obtained for $s = 1/(2N_0)$. By substituting this into (5.15), we obtain the Chernoff upper bound

$$P(\mathbf{x} \to \mathbf{x}'|\mathbf{H}) \leq \exp\left(-\frac{\mathcal{E}}{4n_tN_0}D\right) \tag{5.16}$$

In order to obtain the PEP, we need to average with respect to $\mathbf{H}$. For simplicity, we assume independent Rayleigh fading for all elements of $\mathbf{H}$, i.e., the entries $h_{i,j}$ of $\mathbf{H}$ are i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$.

We rewrite D as

$$\begin{aligned} D &= \sum_{n=1}^{N'} \left|\mathbf{H}(\mathbf{x}'_n - \mathbf{x}_n)\right|^2 \\ &= \sum_{n=1}^{N'} |\mathbf{H}\mathbf{d}_n|^2 \\ &= \sum_{i=1}^{n_r}\sum_{j=1}^{n_t}\sum_{j'=1}^{n_t} h_{i,j}h^*_{i,j'}\sum_{n=1}^{N'} d_{j,n}d^*_{j',n} \\ &= \mathrm{tr}\left(\mathbf{H}\mathbf{D}\mathbf{D}^H\mathbf{H}^H\right) \end{aligned} \tag{5.17}$$

where we have defined $\mathbf{d}_n = \mathbf{x}'_n - \mathbf{x}_n$, and the $n_t \times N'$ matrix difference

$$\mathbf{D} = \mathbf{X}' - \mathbf{X}\begin{bmatrix} d_{1,n} & \cdots & d_{1,N'} \\ \vdots & & \vdots \\ d_{n_t,1} & \cdots & d_{n_t,N'} \end{bmatrix} \tag{5.18}$$

Since $\mathbf{D}\mathbf{D}^H$ is a $n_t \times n_t$ Hermitian symmetric positive semidefinite matrix, we have $\mathbf{D}\mathbf{D}^H = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, where $\mathbf{U}$ is unitary and $\mathbf{\Lambda}$ is diagonal, with diagonal elements λ_j real non-negative, equal to the eigenvalues of $\mathbf{D}\mathbf{D}^H$. Thus,

$$\begin{aligned} D &= \mathrm{tr}\left(\mathbf{H}\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H\mathbf{H}^H\right) \\ &= \mathrm{tr}\left(\widetilde{\mathbf{H}}\mathbf{\Lambda}\widetilde{\mathbf{H}}^H\right) \\ &= \sum_{j=1}^{n_t}\lambda_j\sum_{i=1}^{n_r}|\widetilde{h}_{i,j}|^2 \end{aligned} \tag{5.19}$$

where we have defined $\widetilde{\mathbf{H}} = \mathbf{H}\mathbf{U}$, with elements $\widetilde{h}_{i,j}$. Since $\mathbf{U}$ is unitary, the random variables $\widetilde{h}_{i,j}$ are also i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0,1)$. Then, by using (5.19) in (5.16) and by averaging with respect to the fading, we obtain the desired Chernoff bound on the PEP as

$$\begin{aligned} P(\mathbf{c}\to\mathbf{c}') &\le E\left[\exp\left(-\frac{\mathcal{E}}{4tN_0}\sum_{j=1}^{n_t}\sum_{i=1}^{n_r}\lambda_j|\widetilde{h}_{i,j}|^2\right)\right] \\ &= \prod_{j=1}^{n_t}\prod_{i=1}^{n_r}E\left[\exp\left(-\frac{\mathcal{E}}{4n_tN_0}\lambda_j|\widetilde{h}_{i,j}|^2\right)\right] \\ &= \prod_{j=1}^{n_t}\left[1+\frac{\mathcal{E}}{4n_tN_0}\lambda_j\right]^{-n_r} \end{aligned} \tag{5.20}$$

By using the fact that the λ_j's are non-negative, we obtain the simpler bound

$$P(\mathbf{c}\to\mathbf{c}') \le \left(1 + n_r\mathrm{tr}(\mathbf{D}\mathbf{D}^H)\frac{\mathcal{E}}{4n_tN_0} + \Gamma(\mathbf{c},\mathbf{c}')^{\rho n_r}\left(\frac{\mathcal{E}}{4n_tN_0}\right)^{\rho n_r}\right)^{-1} \tag{5.21}$$

where ρ is the rank of $\mathbf{D}$, i.e., it is the number of non-zero eigenvalues λ_j's, and where we define the "STC gain"

$$\Gamma(\mathbf{c},\mathbf{c}') = \left(\prod_{j:\lambda_j>0} \lambda_j\right)^{1/\rho} \tag{5.22}$$

Notice also that $\mathrm{tr}(\mathbf{D}\mathbf{D}^H)$ is the SED between the two code word arrays $\mathbf{X}$ and $\mathbf{X}'$ (seen as $n_t N'$ vectors).

From (5.21) we observe that for small $\mathcal{E}/N_0$ the most significant performance measure is SED, while for large $\mathcal{E}/N_0$ the performance is dominated by the rank ρ and by the STC gain. The total diversity order of the coded system is given by the product $r\rho_{\min}$, where n_r is the number of receiving antennas and $\rho_{\min}$ is the minimum rank of the matrix $\mathbf{D}$, for all possible pairs of code words $\mathbf{c},\mathbf{c}' \in \mathcal{C}$. We shall refer to $\rho_{\min}$ as the code rank-diversity. The STC gain should be as large as possible for the error events with minimum rank $\rho_{\min}$. In particular, if $\rho_{\min} = n_t$, i.e., $\mathbf{D}$ is full-rank for all code word pairs, we have that $\Gamma(\mathbf{c},\mathbf{c}') = (\prod_{j=1}^{n_t}\lambda_j)^{1/n_t} = (\det(\mathbf{D}\mathbf{D}^H))^{1/n_t}$. For this reason, the above design criteria for large SNR is sometimes referred to as *rank and determinant* criteria.

A necessary condition for full-rank is that $N' \geq n_t$. In other words, in order to exploit the full diversity order of the system, the code block length N' must be at least equal to the number of transmit antennas.

5.3.1 Generalized orthogonal designs and Alamouti's code

In [81], a family of space-time block codes are found from the theory of *Generalized Orthogonal Designs* (GOD). These schemes differ from the general space-time paradigm seen before, as the modulator mapping is applied directly to the information symbols, and then the modulated information symbols are mapped onto code word arrays by a widely linear map.

Widely linear maps are mappings involving linear combinations of the input variables and of their complex conjugates. For every widely linear map $\mu : \mathbb{C}^k \to \mathbb{C}^{n_t \times N}$ there exists a linear map $\widetilde{\mu} : \mathbb{R}^{2k} \to \mathbb{C}^{n_t \times N}$ such that if $\mathbf{x}$ is the input of μ, then $\widetilde{\mathbf{x}} = (\mathrm{Re}\{\mathbf{x}\}, \mathrm{Im}\{\mathbf{x}\})$ is the input of $\widetilde{\mu}$, and such that $\mu(\mathbf{x}) = \widetilde{\mu}\widetilde{\mathbf{x}}$.

Let $\mathbf{x} = (x_1, \ldots, x_q)$ be a vector of q modulation symbols in the signal set $\mathcal{A}$. The corresponding code word array of a space-time block code obtained by widely-linear mapping $\mathbf{x}$ is given by

$$\mathbf{M}(\mathbf{x}) = \sum_{\ell=1}^{q} x_\ell \mathbf{A}_\ell + \sum_{\ell=1}^{q} x_\ell^* \mathbf{B}_\ell \tag{5.23}$$

where $\mathbf{A}_\ell$ and $\mathbf{B}_\ell$ are $n_t \times N$ matrices. The rate of the GOD code is given by $R = qb/N$, where $\log_2 |\mathcal{A}| = b$ is the number of bit/symbol of the signal set $\mathcal{A}$.

The space-time block code defined by (5.23) is a GOD if, for all $\mathbf{x} \in \mathbb{C}^q$ it satisfies the row orthogonality condition

$$\mathbf{M}(\mathbf{x})\mathbf{M}(\mathbf{x})^H = \left(\sum_{\ell=1}^{q} |x_\ell|^2\right)\mathbf{I} \tag{5.24}$$

From the above condition, it follows immediately that a GOD space-time block code has full

rank-diversity. In fact, for any $\mathbf{x} \neq \mathbf{x}'$, the difference of the code word arrays is given by

$$\mathbf{D} = \mathbf{M}(\mathbf{x}) - \mathbf{M}(\mathbf{x}') = \mathbf{M}(\mathbf{x} - \mathbf{x}')$$

and $\det(\mathbf{D}\mathbf{D}^H) = \left(\sum_{\ell=1}^{q} |x_\ell - x'_\ell|^2\right)^t$. It is clear that $\det(\mathbf{D}\mathbf{D}^H) > 0$ for all $\mathbf{x} \neq \mathbf{x}'$. Hence, the rank of $\mathbf{D}$ is equal to n_t.

The other interesting feature of GOD space-time codes is that their ML decoder is very simple. In fact, the general ML decision metric in (5.13) in this case becomes

$$\begin{aligned}
\|\mathbf{Y} - \sqrt{\mathcal{E}/n_t}\mathbf{H}\mathbf{M}(\mathbf{x})\|_F^2 &= \sum_{i=1}^{n_r} |\mathbf{y}^i - \sqrt{\mathcal{E}/n_t}\mathbf{h}^i\mathbf{M}(\mathbf{x})|^2 \\
&= \sum_{i=1}^{n_r} |\mathbf{y}^i|^2 - 2\sqrt{\mathcal{E}/t}\mathrm{Re}\{\mathbf{y}^i\mathbf{M}(\mathbf{x})^H(\mathbf{h}^i)^H\} + (\mathcal{E}/n_t)\mathbf{h}^i\mathbf{M}(\mathbf{x})\mathbf{M}(\mathbf{x})^H(\mathbf{h}^i)^H \\
&= \|\mathbf{Y}\|_F^2 - 2\sqrt{\mathcal{E}/n_t}\mathrm{Re}\left\{\mathrm{tr}\left(\mathbf{H}^H\mathbf{Y}\mathbf{M}(\mathbf{x})^H\right)\right\} + (\mathcal{E}/n_t)\left(\sum_{\ell=1}^{q}|x_\ell|^2\right)\|\mathbf{H}\|_F^2
\end{aligned} \tag{5.25}$$

where $\mathbf{y}^i$ and $\mathbf{h}^i$ denote the i-th row of $\mathbf{Y}$ and of $\mathbf{H}$, respectively.

Since the trace is a linear operation and the elements of $\mathbf{M}(\mathbf{x})$ are linear combinations of symbols x_ℓ and x_ℓ^*, after neglecting the ininfluent term $\|\mathbf{Y}\|_F^2$ which does not depend on the modulation symbols, we can rewrite the ML decision metric in the decoupled form

$$\sum_{\ell=1}^{q}\left(-\mathrm{Re}\left\{\alpha_\ell x_\ell + \beta_\ell x_\ell^*\right\} + \sqrt{\mathcal{E}/n_t}|x_\ell|^2\|\mathbf{H}\|_F^2\right)$$

where each term inside the summation depends only on x_ℓ, and where α_ℓ and β_ℓ are coefficients that can be easily calculated from the code generating arrays $\{\mathbf{A}_\ell, \mathbf{B}_\ell\}$ and from the output of the matrix channel matched filter $\mathbf{H}^H\mathbf{Y}$. Since the symbols x_ℓ are independent and the terms of the above metric are decoupled, the ML metric is minimized by minimizing separately each term with respect to its own variable, i.e., the ML decoding rule is, for all $\ell = 1, \ldots, q$,

$$\widehat{x}_\ell = \arg\min_{a\in\mathcal{A}}\left\{-\mathrm{Re}\left\{\alpha_\ell a + \beta_\ell a^*\right\} + \sqrt{\mathcal{E}/n_t}|a|^2\|\mathbf{H}\|_F^2\right\}$$

The complexity of this rule is $q|\mathcal{A}|$ instead of $|\mathcal{A}|^q$ as in the general "brute force" case.

Example: Alamouti's scheme. Finding GOD space-time block codes with optimal rate (i.e., with the minimum length N for given n_t and ratio q/N) is in general a difficult problem. Some constructions are given in [81]. Perhaps the most remarkable one is the $n_t = 2, N = 2$ GOD known as Alamouti's scheme, which is optimal in terms of rate. Moreover, it is possible to show that Alamouti's scheme is also optimal in an information theoretic sense for the case $n_t = 2$, $n_r = 1$. In fact, in this case the capacity of the original channel (when the transmitter does not know the channel matrix $\mathbf{H}$) is the same as that of the channel including Alamouti's encoder as part of the channel. Hence, Alamouti's encoding is information lossless for the case of two transmit antennas and one receiving antenna.

Alamouti's GOD scheme is defined by the mapping

$$(x_1, x_2) \mapsto \begin{bmatrix} x_1^* & -x_2 \\ x_2^* & x_1 \end{bmatrix} \tag{5.26}$$

It is easy to check that the matrices defined above are proportional to unitary matrices for all $x_1, x_2 \in \mathbb{C}$.

Consider for example the transmission of Alamouti's code over the $n_t = 2$, $n_r = 1$ channel with matrix $\mathbf{H} = [h_1, h_2]$. The channel output is given by

$$\mathbf{Y} = [y_1, y_2] = \sqrt{\mathcal{E}/2}[h_1, h_2] \begin{bmatrix} x_1^* & -x_2 \\ x_2^* & x_1 \end{bmatrix} + [\nu_1, \nu_2]$$

The matrix matched filter output is given by

$$\begin{aligned} \mathbf{H}^H\mathbf{Y} &= \begin{bmatrix} h_1^* y_1 & h_1^* y_2 \\ h_2^* y_1 & h_2^* y_2 \end{bmatrix} \\ &= \sqrt{\mathcal{E}/2} \begin{bmatrix} |h_1|^2 & h_1^* h_2 \\ h_2^* h_1 & |h_2|^2 \end{bmatrix} \begin{bmatrix} x_1^* & -x_2 \\ x_2^* & x_1 \end{bmatrix} + \begin{bmatrix} h_1^* \nu_1 & h_1^* \nu_2 \\ h_2^* \nu_1 & h_2^* \nu_2 \end{bmatrix} \\ &= \sqrt{\mathcal{E}/2} \begin{bmatrix} |h_1|^2 x_1^* + h_1^* h_2 x_2^* & -|h_1|^2 x_2 + h_1^* h_2 x_1 \\ h_2^* h_1 x_1^* + |h_2|^2 x_2^* & -h_2^* h_1 x_2 + |h_2|^2 x_1 \end{bmatrix} + \begin{bmatrix} h_1^* \nu_1 & h_1^* \nu_2 \\ h_2^* \nu_1 & h_2^* \nu_2 \end{bmatrix} \end{aligned} \tag{5.27}$$

The ML decisions on the symbols x_1 and x_2 are given by (see (5.25))

$$\begin{aligned} \widehat{x}_1 &= \arg \min_{a \in \mathcal{A}} \left\{ -2\mathrm{Re}\{h_1^* y_1 a + h_2^* y_2 a^*\} + \sqrt{\mathcal{E}/2}|a|^2(|h_1|^2 + |h_2|^2) \right\} \\ \widehat{x}_2 &= \arg \min_{a \in \mathcal{A}} \left\{ -2\mathrm{Re}\{h_2^* y_1 a - h_1^* y_2 a^*\} + \sqrt{\mathcal{E}/2}|a|^2(|h_1|^2 + |h_2|^2) \right\} \end{aligned} \tag{5.28}$$

We notice that, when x_1, x_2 are the transmitted symbols, then

$$\begin{aligned} h_1 y_1^* + h_2^* y_2 &= \sqrt{\mathcal{E}/2}(|h_1|^2 + |h_2|^2)x_1 + h_1 \nu_1^* + h_2^* \nu_2 \\ h_2 y_1^* - h_1^* y_2 &= \sqrt{\mathcal{E}/2}(|h_1|^2 + |h_2|^2)x_2 + h_2 \nu_1^* - h_1^* \nu_2 \end{aligned} \tag{5.29}$$

Hence, the decision rule (5.28) can be interpreted as the ML decision rule at the output of a Maximal Ratio Combiner with observations

$$\begin{aligned} \widetilde{y}_1 &= \sqrt{\mathcal{E}(|h_1|^2 + |h_2|^2)/2}x_1 + v_1 \\ \widetilde{y}_2 &= \sqrt{\mathcal{E}(|h_1|^2 + |h_2|^2)/2}x_2 + v_2 \end{aligned}$$

where v_1 and v_2 are i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0, N_0)$.

From this we see clearly that Alamouti's scheme creates two parallel non-interfering channels, each having diversity 2. $\Diamond$

Chapter 6

A Tribute to Shannon

Cet article est tiré de "hommage à Shannon" de Gérard Batail.

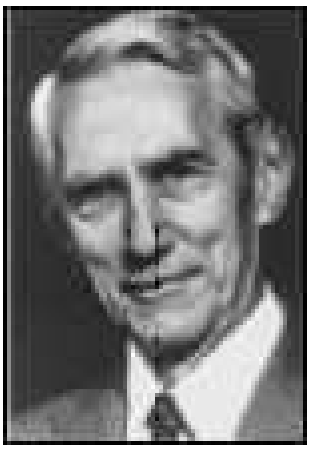

Figure 6.1: Shannon (1916-2001).

Claude Edwood Shannon est né le 30 avril 1916 à Petoskey, Michigan, aux Etats-Unis. Son père, homme d'affaires et magistrat, descendait des premiers colons du New Jersey. Sa mère, fille d'immigrés allemands, professeur de langues, dirigeait le collège (high school) de Gaylord, Michigan, où il a passé toute son enfance. Il admire alors Edison, lointain cousin de la famille, et fait preuve d'ingéniosité dans divers bricolages et inventions en mécanique, électricité et radioélectricité. Il quitte le collège de Gaylord en 1932 pour entrer à l'Université du Michigan à Ann Arbor. Il y obtient en 1936 le grade de bachelor en ingénierie électrique ainsi qu'en mathématiques. Il entre alors comme assistant de recherche au Massachusetts Institute of Technology (MIT) près de Boston, poste à temps partiel qui lui permet de continuer ses études. Sa thèse de master of science est consacrée à l'application de l'algèbre de Boole aux circuits de relais et commutateurs. Elle est publiée en 1938, rencontre un très grand succès et lui vaut en 1940 l' Alfred D. Nobel prize qui récompense annuellement aux Etats-Unis le travail d'un ingénieur de moins de 30 ans (ne pas confondre ...). En 1938, il quitte le département d'ingénierie électrique pour le département de mathématiques, à l'instigation du vice-président

du MIT Vannevar Bush (qui deviendra, pendant et après la guerre, un conseiller écouté du président Franklin Roosevelt). Bush était un ingénieur d'une imagination visionnaire, qui avait inventé des machines anticipant sur l'ordinateur mais trahies par les moyens technologiques du temps. Il venait d'accéder à la présidence de la Carnegie Institution à Washington, dont une branche étudiait la génétique (et l'eugénisme, qui ne sera discrédité qu'après que la guerre aura révélé l'usage monstrueux qui en a été fait). Avec le mémoire de Shannon, la conception de circuits de commutation était passée du statut d'art à celui de science, grâce à la formalisation mathématique du problème, et Bush espérait qu'une démarche similaire du même Shannon féconderait la génétique. De retour au MIT après un stage au laboratoire de génétique de la Carnegie Institution à Cold Spring Harbor, Shannon écrit, sous la direction de l'algébriste Frank L. Hitchcock, sa thèse intitulée 'An algebra for theoretical genetics'. Incidemment, le travail de Shannon a été examiné par Barbara Burks, psychologue spécialiste de la 'génétique des génies', membre de la société américaine d'eugénisme. Son diagnostic est dénué d'ambiguité: le jeune Shannon est un génie qu'elle compare, dans une lettre adressée à Bush en 1939, à Pascal réinventant à 12 ans la géométrie d'Euclide . Le doctorat (Ph.D.) est décerné à Shannon au printemps 1940. Il passe l'été de cette même année aux Bell Telephone Laboratories (Bell Labs) pour appliquer, avec succès, la méthode qui a fait l'objet de son mémoire de 1938 à la simplification des circuits de commutation (un enjeu considérable pour la conception des centraux téléphoniques). Après avoir travaillé pendant l'année scolaire 1940-1941 à l'Institute for Advanced Studies de Princeton, sous la direction de Hermann Weyl, il retourne aux Bell Labs en 1941, appelé à s'intégrer à une équipe de recherche (dont les principaux membres étaient H.S. Black et H.W. Bode) sur les systèmes de défense anti-aérienne: un problème d'une pressante actualité en ce temps de guerre. Les travaux de cette équipe ont abouti à la mise au point et à la fabrication du système de conduite de tir anti-aérien M6 qui a permis à l'Angleterre de limiter les dégâts causés par les fusées V1 et V2 et a contribué à donner aux Alliés la maitrise du ciel, pas décisif vers leur victoire. Le contexte de la guerre a aussi amené Shannon à travailler sur la cryptographie comme consultant du National Defense Research Committee (NDRC), créé dès avant l'entrée en guerre des Etats-Unis et présidé par Vannevar Bush. C'est à ce titre qu'il a rencontré Alan Turing à plusieurs reprises. Il semble que la cryptographie ait été pour Shannon une source d'inspiration, mais aussi et surtout un masque, honorable en ce temps de guerre, pour les études qu'il avait déjà entreprises sur la théorie des communications et de l'information: elles ne contribuaient pas à l'effort de guerre et leur utilité éventuelle ne pouvait être justifiée qu'á posteriori. Les Bell Labs étaient une très féconde pépinière de chercheurs et ingénieurs dans tous les domaines de la physique et des mathématiques. La théorie de l'information n'est pas la moindre production des Bell Labs. L'invention du transistor en est une autre, miraculeusement complémentaire de la théorie de l'information à laquelle elle a fourni, ainsi qu'à la technique des ordinateurs, des moyens qui manquaient cruellement en 1948. Shannon est resté 15 ans aux Bell Labs qu'il n'a quittés que pour aller enseigner au MIT. Des témoins dignes de foi ont rencontré Shannon dans les couloirs des Bell Labs, juché sur un monocycle et jonglant (avec trois balles, précisait l'intéressé). Par delà l'anecdote, cela atteste le goût immodéré du jeu qui était une dominante de sa personnalité, l'intérêt pour les équilibres précaires et, bien sûr, un anticonformisme qu'il ne craignait pas d'afficher. Peut-être était-ce d'ailleurs un moyen paradoxal de se protéger des indiscrétions: Shannon ne se livrait guère et vivait retiré. C'est ainsi qu'il se débarrassait des journalistes qui essayaient de l'interviewer en leur faisant visiter sa collection de 'jouets'. Shannon aimait le jeu, tous les jeux. Les jeux de hasard, les échecs, la musique (il jouait de la clarinette et collectionnait les instruments de toutes sortes) et, plus

encore peut-être, les jouets savants qu'il construisait. Son vif intérêt pour la roulette l'avait rendu indésirable dans les casinos. Faut-il ranger parmi les jeux de hasard les investissements financiers? Shannon y avait réussi au point de faire fortune, ce qui lui avait permis de ne plus dépendre pécuniairement des Bell Labs. Il jouait excellemment aux échecs (lors d'un voyage en URSS, en 1965, il avait brillamment résisté au champion du monde Mikhail Botvinnik, manquant de peu la nullité), ce qui l'a naturellement conduit à s'intéresser à la conception de machines expertes en ce jeu. Son article de 1950 "Programming a computer for playing chess" en a fait un des pionniers en ce domaine. Professeur invité au MIT en 1956, il y devint enseignant permanent en 1959, encadrant notamment les thèses de chercheurs dont beaucoup ont fait une brillante carrière en théorie de l'information et codage. Il y est resté jusqu'en 1972. Il s'est retiré dans une grande maison près d'un lac, à Winchester (Massachusetts), où il pouvait s'adonner à ses passe-temps favoris. Ses derniers articles publiés dans le domaine de la théorie de l'information l'ont été en 1967, co-signés de R.G. Gallager et E.R. Berlekamp. L'un des derniers articles publiés sous son seul nom, en 1959, "Probability of error for optimal codes in a Gaussian channel", d'ailleurs remarquable, contient peut-être l'une des clés du comportement de Shannon vis-à-vis de la théorie de l'information. Un mot que l'on n'a pas l'habitude de lire dans les écrits scientifiques y est répété à plusieurs reprises: tedious. De fait, Shannon a dû effectuer des calculs d'algèbre élémentaire longs et peu passionnants pour obtenir des bornes de probabilité d'erreur (avec l'aide, mentionnée explicitement, de sa femme Betty), en contraste évident avec l'exaltation qui n'avait pu qu'accompagner les découvertes des débuts. Sans nul doute, la peur de l'ennui était une motivation majeure pour cet amateur passionné de jeux. Son célèbre théorème sur le décodage de canal établissait l'existence de codes rendant la probabilité d'erreur arbitrairement petite pourvu que l'entropie de la source soit inférieure à la capacité du canal, résultat dont la démonstration reposait sur un procédé extraordinaire qui ne laissait aucun espoir de réalisation effective: le codage aléatoire, inspiré probablement par la cryptographie et peut-être aussi par Darwin. Pourquoi Shannon n'a-t-il pas contribué à la recherche de procédés de codage efficaces pour les critères de la théorie de l'information, mais explicites et non plus aléatoires? Ses co-auteurs des articles précités, Gallager et Berlekamp, ont été l'un et l'autre des acteurs éminents de cette recherche qui allait s'apparenter à la quête du Graal (et se poursuit aujourd'hui encore, bien qu'elle ait fait un progrès décisif avec l'invention des turbo codes en 1993, il est vrai que le chemin qui y a conduit a été tortueux et peu conforme aux directions initiales de ces recherches). Etait-ce la conscience de l'importance des efforts qu'il fallait consentir, celle que ce travail ne pourrait être que collectif, lent et risquait d'être passablement ennuyeux? Peut-être aussi la réticence de Shannon à toute finalité utilitaire, illustrée par la gratuité des 'jouets' qu'il construisait?

Chapter 7

Useful Results on random matrix theory

7.1 Random matrix theory

Random matrix theory has been a part of advanced multivariate statistical analysis since the end of the 1920's. Random matrices were first proposed by Eugene Wigner in quantum mechanics where the statistics of experimentally measured energy levels of nuclei were explained in terms of the eigenvalues of random matrices. When Tsé [46] and Verdu [47] introduced nearly simultaneously random matrices in 1999 to model uplink unfaded CDMA systems equipped with certain types of receivers, random matrix theory entered the field of telecommunications. From that time, random matrix theory has known considerable interest to study the performance of multiple antenna schemes [24, 1]

This section is intended to give a brief survey on random matrix theory. Our aim is to give a better understanding of matrix theory in order to apply the work of Tsé and Verdu to i.i.d Linear Precoders for OFDM. Let us give first some basic definitions.

Let μ be a probability measure on $\mathbb{R}$. Its Stieltjes transform:

$$G_\mu(z) = \int_{-\infty}^{+\infty} \frac{d\mu(t)}{t-z}$$

is defined on $\mathbb{C}/\mathbb{R}$. When z lies in the upper half-plane $\mathbb{C}^+ = (z \in \mathbb{C} : \mathrm{Im}(z) > 0)$, the transform $G_\mu(z)$ is an analytic function in $\mathbb{C}^+$ possessing the following properties: $G_\mu(\mathbb{C}^+) \subset \mathbb{C}^+$ and $\mid G_\mu(z) \mid \leq \frac{1}{\mathrm{Im}(z)}$

Let μ be compactly supported. Then $G_\mu(z)$ is analytic in a neighborhood of ∞. Since $(z-t)^{-1} = \sum_{k=0}^{\infty} t^k z^{-k-1}$, it is obvious that $G_\mu(z)$ has the following expansion at $z = \infty$:

$$-G_\mu(z) = z^{-1} + \sum_{k=0}^{\infty} m_k(\mu) z^{-k-1}.$$

where $m_k(\mu) = \int t^k d\mu(t) (k \in \mathbb{Z}^+)$

Writing $z = x + iy$, it is possible to recover μ from its Cauchy transform up to a factor. When μ is absolutely continuous with respect to the Lebesgue measure, its density f(x) is given by:

$$f(x) = +\frac{1}{\pi}\lim_{y\to 0}\mathrm{Im}G_\mu(x+iy)$$

The random matrix theory in consideration here (see [82, 83] for more details) deals with the limiting distribution of random Hermitian matrices of the form $A+WDW^H$. Here, $W(N\times K), D(K\times K),$ and $A(N\times N)$ are independent, with W containing i.i.d. entries having finite second moments, D is diagonal with real entries, A is Hermitian and $K/N\to\alpha>0$ as $N\to\infty$. The behavior is expressed in terms of the limiting distribution function F^{A+WDW^H} of the eigenvalues of $A+WDW^H$ (i.e. $F^{A+WDW^H}(x)$ is the proportion of eigenvalues of $A+WDW^H\leq x$). The remarkable result of random matrix theory is the convergence in some sense, of $F^{A+WDW^H}(x)$ to a non-random F.

The papers vary in the assumption on T, W and A. We will only take here the case of interest.

Theorem 3 *Let A be a $N\times N$ hermitian matrix, nonrandom, for which F^A converges weakly as $N\to\infty$ to a distribution function $\mathbb{A}$. Suppose F^D converges weakly to a nonrandom probability distribution function denoted $\mathbb{D}$. Suppose the entries of $\sqrt{N}W$ i.i.d. for fixed N with unit variance (sum of the variances of the real and imaginary parts in the complex case). Then the eigenvalue distribution of $A+WDW^H$ converges weakly to a deterministic F. Its Stieljes transform G(z) satisfies the equation:*

$$G(z) = G_{\mathbb{A}}\left(z-\alpha\int\frac{\tau d\mathbb{D}(\tau)}{1+\tau G(z)}\right) \tag{7.1}$$

7.2 An application example:

We give hereafter the eigenvalue distribution of matrices defined by:

$$\mathbf{W}_{N,K}\mathbf{W}_{N,K}^H+\sigma^2\mathbf{I}$$

where $\sqrt{N}\mathbf{W}_{N,K}$ is a $N\times K$ matrix with i.i.d. entries with zero mean and variance one and $N\to\infty$ such as $\frac{K}{N}\to\alpha$ fixed. This example is intended to show to what extent (in terms of the matrix dimension) theoretical and practical results fit. Moreover, we give the basic machinery that the reader can use to derive any limiting eigenvalue distribution of matrices defined as in Theorem 3.

Denote $A=\sigma^2\mathbf{I}$ and $D=\mathbf{I}_{K,K}$. In this case, $d\mathbb{A}(x)=\delta(x-\sigma^2)$ and $d\mathbb{D}(x)=\delta(x-1)$. Applying Theorem 3, the following result is obtained:

$$\begin{aligned}
G(z) &= G_{\sigma^2\mathbf{I}}\left(z-\alpha\int\frac{\tau\delta(\tau-1)d\tau}{1+\tau G(z)}\right) &(7.2)\\
&= G_{\sigma^2\mathbf{I}}(z-\frac{\alpha}{1+G(z)}) &(7.3)\\
&= \int\frac{\delta(\sigma^2-\lambda)d\lambda}{\lambda-z+\frac{\alpha}{1+G(z)}} &(7.4)\\
&= \frac{1}{\sigma^2-z+\frac{\alpha}{1+G(z)}}. &(7.5)
\end{aligned}$$

$G_{\sigma^2\mathbf{I}}(z)$ is the Cauchy transform of the eigenvalue distribution of matrix $\sigma^2\mathbf{I}$. The solution of the second order equation 7.5 yields:

$$G(z) = \frac{1-\alpha}{2(\sigma^2 - z)} - \frac{1}{2} + \frac{1}{2(\sigma^2 - z)}\sqrt{((\sigma^2 - z + \alpha - 1)^2 + 4(\sigma^2 - z))}.$$

The asymptotic eigenvalue distribution is therefore given by:

$$f(\lambda) = \begin{cases} [1-\alpha]^+\delta(x) + \frac{\alpha}{\pi(\lambda-\sigma^2)}\sqrt{\lambda - \sigma^2 - \frac{1}{4}(\lambda - \sigma^2 + 1 - \alpha)^2} & \text{if } \sigma^2 + (\sqrt{\alpha}-1)^2 \leq \lambda \leq \sigma^2 + (\sqrt{\alpha}+1)^2 \\ 0 & \text{otherwise} \end{cases}$$

where $\delta(x)$ is a unit point mass at 0 and $[z]^+ = \max(0, z)$.

A remarkable result is that the eigenvalues of such matrices have a compact support while the entries may take any value (subject to a zero mean and variance one distribution). We have plotted the theoretical and practical eigenvalue distribution for various size matrices. Only one realization of matrix $\mathbf{W}_{N,K}\mathbf{W}_{N,K}^H + \sigma^2\mathbf{I}$ has been studied. As the dimensions increase, the practical results fit extremely well the theoretical results as shown in fig.A.1 and fig.A.2.

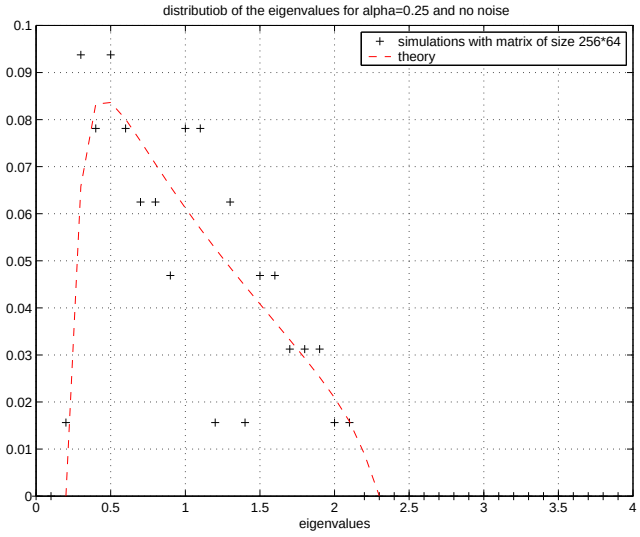

Figure 7.1: matrix size: 256*64, no noise, $\alpha = 0.25$.

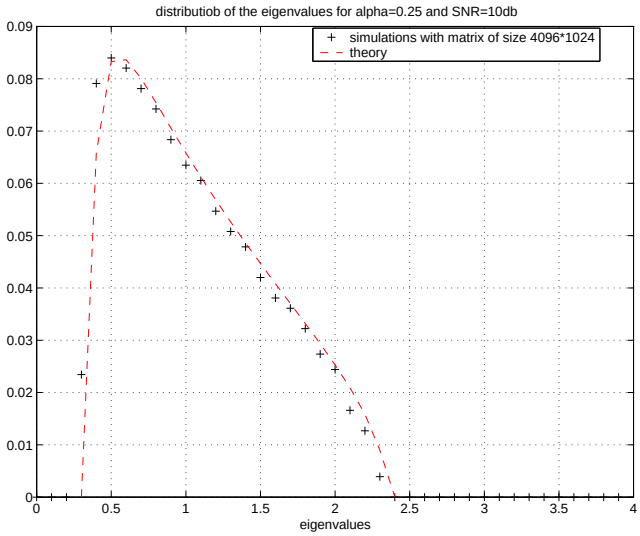

Figure 7.2: matrix size: 4096*1024, 10dB, $\alpha = 0.25$.

Chapter 8

Some useful results

This section provides general useful theorems and definitions. For specific developments, the reader can consult [84].

8.1 Useful Inequalities and Results

- **Jensen's Inequality**
  If $\Phi(x)$ is a convex function of x and $f(\omega)$ and $\Phi(f(\omega))$ are integrable then:

$$\int \Phi(f(\omega))dP \geq \Phi\left(\int f(\omega)dP\right)$$

- **Holder's Inequality**

$$E\left(\mid XY \mid\right) \leq \left(E(\mid X \mid^{\frac{1}{p}}\right)^{p}\left(E(\mid Y \mid^{\frac{1}{q}}\right)^{q}$$

$$\frac{1}{p}+\frac{1}{q}=1$$

- **Schwarz's Inequality (case p=2)**

$$E\left(\mid XY \mid\right) \leq \left(E(\mid X \mid^{\frac{1}{2}}\right))^{2}\left(E(\mid Y \mid^{\frac{1}{2}}\right))^{2}$$

- **Markov's Inequality**

$$P[\mid X \mid \geq a] \leq \frac{E\left(\mid X \mid^{r}\right)}{a^{r}}$$

  $r > 0$.

- **The Poisson distribution**

$$P(k) = e^{-m}\frac{m^{k}}{k!}$$

- **The Laplace Distribution**

$$f(x) = \frac{1}{2a} e^{\frac{-|x|}{a}}$$

- **The Rayleigh distribution**

$$f(x) = \frac{2}{\alpha} x e^{-\frac{x^2}{\alpha}}$$

- **The Cauchy distribution**

$$f(x) = \frac{1}{\pi} \frac{a}{a^2 + (x-m)^2}$$

- **Gaussian Characteristic function**
  For $\mathbf{X} \in N(\mu, \Lambda)$, the Gaussian Characteristic function has the following form:

$$\phi_{\mathbf{X}}(t) = e^{it'\mu - \frac{1}{2}t'\Lambda t}$$

- **Remark**
  If $\det\mathbf{\Lambda} > 0$, the probability distribution of $\mathbf{X}$ is truly n-dimensional in the sense that it cannot be concentrated on a subspace of lower dimension. If $\det\mathbf{\Lambda} = 0$, it can be concentrated on such a subspace; we call it a singular case. In this case, no density exists.

- **The Central Limit Theorem**
  Let $\mathbf{X}_1, \mathbf{X}_2, ...$ be independent, identically distributed random variables with finite expectation μ and finite variance σ^2, and set $\mathbf{S}_n = \mathbf{X}_1 + \mathbf{X}_2 + ... + \mathbf{X}_n, n \geq 1$. Then

$$\frac{\mathbf{S}_n - n\mu}{\sigma\sqrt{N}} \to N(0,1)$$

  in distribution as $n \to \infty$

- **Stirling Formula**

$$\log n! = n\log n - n + \sqrt{2\pi n} + \frac{1}{12n} + O(\frac{1}{n^2})$$

- **Boltzmann Entropy**
  If μ has a multivariate normal distribution $N(m, \Sigma)$ whose density is

$$\frac{1}{(2\pi)^{\frac{n}{2}} \mid \Sigma\mid^{\frac{1}{2}}} e^{-\frac{1}{2}(x-m)^H \Sigma^{-1}(x-m)}$$

  with the mean vector m and the covariance matric Σ, then its entropy is equal to:

$$\frac{n}{2}\log(2\pi e) + \frac{1}{2}\log(\mid \Sigma \mid)$$

8.2 Probability theory results

Let (Ω, A, P) be a probability space.

Definition 1 *A sequence f_n of random variables is said to converge to a function f everywhere or pointwise if*

$$\lim_{n\to\infty} f_n(\omega) = f(\omega)$$

for every $\omega \in \Omega$.

Definition 2 *A sequence f_n of random variables is said to converge to a measurable function f almost everywhere or almost surely (usually abbreviated as a.s) if there exists a measurable set N with P(N)=0 such that:*

$$\lim_{n\to\infty} f_n(\omega) = f(\omega)$$

for every $\omega \in N^c$

Theorem 4 (Borel Cantelli Lemma)
let $\{A_n\}_{n=0}^{\infty}$ be a sequence of events and let A be the event consisting of the occurrence of an infinite number of events A_n, $n = 1, \ldots$ Then if

$$\sum_{n=1}^{\infty} P(A_n) < \infty$$

then

$$P(A) = 0$$

Definition 3 *A sequence f_n of measurable functions is said to converge to a measurable function f in probability if*

$$\lim_{n\to\infty} P[\omega :\mid f_n(\omega) - f(\omega) \mid \geq \epsilon] = 0$$

for every $\epsilon > 0$.

Definition 4 *A sequence α_n of probability measures on the real line $\mathbb{R}$ with distribution functions $F_n(x)$ is said to converge weakly to a limiting probability measure α with distribution function $F(x)$ (in symbols $\alpha_n \Rightarrow \alpha$ or $F_n \Rightarrow F$, F is also called the limit distribution) if*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for every x that is a continuity point of F.

Definition 5 Measure
A measure is formally defined as a non-negative map $m : F \to \mathbb{R}$ (the reals) such that $m(\emptyset) = 0$ and if A_n is a countable sequence in F and the A_n are pairwise disjoint, then

$$m\left(\bigcup A_n\right) = \sum_n m(A_n)$$

If, in adition, m(X)=1, then m is said to be a Probability Measure.

Definition 6 Banach space
A normed linear Space which is complete in the normdetermined Metric. A Hilbert Space is always a Banach space, but the converse need not hold.

Definition 7 Complete Metric Space
A complete metric space is a metric space in which every Cauchy sequence is convergent. Examples include the Real Numbers with the usual metric.

Definition 8 Hilbert space
A Hilbert space is an inner product space which, as a metric space, is complete.

Bibliography

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

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

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

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

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

[6] 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.

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

[8] 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.

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

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

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

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

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

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

[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.

[33] 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.

[34] 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.

[35] 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.

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

[37] D. Gesbert, M. Shafi, D. Shiu, P.J Smith, and A. Naguib, "From Theory to Practice: an Overview of MIMO Space-Time Coded Wireless Systems," *IEEE Journal on Selected Areas in Communications*, pp. 281– 302, vol. 21, no. 3 2003.

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

[39] L. Hanlen and A. Grant, "Capacity analysis of correlated mimo channels," in *IEEE International Symposium on Information Theory*, Yokohama, Japan, July 2003.

[40] A. Grant, "Rayleigh fading multiantenna channels," in *Eurasip Journal on Applied Signal Processing, special issue on Space-Time coding*, 2002, vol. 3, pp. 316–329.

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

[42] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multi-Antenna Channel-Hardening and its Implications for Rate Feedback and Scheduling," *IEEE Trans. on Information Theory*, pp. vol. 50, no. 9, Sept. 2002.

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

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

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

[46] D.N.C Tse and S. Hanly, "Linear Multi-user Receiver: Effective Interference, Effective Bandwidth and User Capacity," *IEEE Trans. on Information Theory*, pp. 641–657, Mar. 1999.

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

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

[49] S. Shamai and S. Verdu, "The Impact of Frequency-Flat Fading on the Spectral Efficiency of CDMA," *IEEE Trans. on Information Theory*, pp. 1302–1326, May 2001.

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

[51] J.M. Chaufray, W. Hachem, and Ph. Loubaton, "Asymptotical Analysis of Optimum and Sub-Optimum CDMA Downlink MMSE Receivers," in *IEEE International Symposium on Information Theory*, Lausanne, Switzerland, July 2002.

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

[53] U. Madhow and M. Honig, "MMSE Interference Supression for Direct-Sequence Spread Spectrum CDMA," *IEEE Trans. on Communications*, vol. 42, pp. 3178–3188, Dec. 1994.

[54] J.M. Cioffi, G.P. Dudevoir, M.V. Eyuboglu, and G.D. Forney, "MMSE Decision Feedback Equalizers and Coding- Part I: Equalization Results," *IEEE Trans. on Communications*, pp. 2582–2594, Oct. 1995.

[55] J.M. Cioffi, G.P. Dudevoir, M.V. Eyuboglu, and G.D. Forney, "MMSE Decision Feedback Equalizers and Coding- Part II: Coding Results," *IEEE Trans. on Communications*, pp. 2595–2604, Oct. 1995.

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

[57] A. J. Paulraj and T. Kailath, "Increasing Capacity in Wireless BroadcastSystems Using Distributed Transmission/Directional Reception," U.S Patent, No. 5345599, 1994.

[58] M. Varanasi and T. Guess, "Achieving vertices of the capacity region of the synchronous correlated-waveform multiple-access channel with decision-feedback receivers," in *IEEE International Symposium on Information Theory*, 1997, p. 270.

[59] N. Al-Dahir and J.M. Cioffi, "MMSE decision-feedback equalizers:finite-length results," *IEEE Trans. on Information Theory*, pp. 961–975, jul 1995.

[60] Albert M. Chan and Gregory W. Wornell, "A Class of Asymptotically Optimum Iterated-decision Multiuser Detectors," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, Utah, May 2001.

[61] H. Poor and S. Verdú, "Probability of Error in MMSE Multiuser Detection," *IEEE Trans. on Information Theory*, vol. 43, no. 3, pp. 858–871, May 1997.

[62] J. Zhang, E. Chong, and D. Tse, "Output MAI Distributions of Linear MMSE Multiuser Receivers in CDMA Systems," *IEEE Trans. on Information Theory*, pp. 1128–1144, Mar. 2001.

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

[64] J.W. Silverstein, "Weak Convergence of Random Functions defined by the Eigenvectors of Sample Covariance Matrices," *Annals of Probability*, vol. 18, pp. 1174–1194, 1990.

[65] D. Guo, S. Verdu, and L.K Rasmussen, "Asymptotic normality of linear multiuser receiver outputs," *IEEE Trans. on Information Theory*, vol. 48, no. 12, pp. 3080 – 3095, Dec. 2002.

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

[67] C. E. Shannon, "Communication Theory of Secrecy Systems," *The Bell Labs Technical Journal*, pp. 656–715, May, vol. 28, No 4 1949.

[68] N. Wiener, *Cybernetics*, J. Wiley and Sons, Inc, New York, 1948.

[69] N. Wiener, *Extrapolation, Interpolation and Smooting of Stationary Time Series*, J. Wiley and Sons, Inc, New York, 1949.

[70] Jr G. D. Forney, "On the Role of the MMSE Estimation in Approaching the Information-Theoretic Limits of Linear Gaussian Channels: Shannon meets Wiener," in *41th Annual Allerton Conf. on Comm. Control and Comput., Monticello, Il*, Oct. 2-4, 2003.

[71] D. Guo, S. Shamai, and S. Verdu, "Mutual information and mmse in gaussian channels," in *IEEE International Symposium on Information Theory*, Chicago, IL, USA, 2004, p. 347.

[72] P. Stoica, Y. Jiang, and J. Li, "On MIMO Channel Capacity: An Intuitive Discussion," pp. 83–84, May 2005.

[73] C. E. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE*, pp. 10–21, Jan, vol. 37, No 1 1949.

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

[75] V. Tarokh, N. Seshadri, and A.R. Calderbank, "Space-Time Codes for High Data Rate Wireless Communication: Performance Criterion and Code Construction," *IEEE Trans. on Information Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[76] M. Schwartz, W. Bennet, and S. Stein, *Communication systems and techniques*, McGraw-Hill, New York, 1966.

[77] J. Ventura-Traveset, G. Caire, E. Biglieri, and G.Taricco, "Impact of Diversity Reception on Fading Channels with Coded Modulation. Part I: Coherent Detection," *IEEE Trans. on Communications*, vol. 45, no. 5, May 1997.

[78] J. Ventura-Traveset, G. Caire, E. Biglieri, and G.Taricco, "Impact of Diversity Reception on Fading Channels with Coded Modulation. Part II: Differential Block Detection," *IEEE Trans. on Communications*, vol. 45, no. 6, June 1997.

[79] J. Ventura-Traveset, G. Caire, E. Biglieri, and G.Taricco, "Impact of Diversity Reception on Fading Channels with Coded Modulation. Part III: Co-Channel Interference," *IEEE Trans. on Communications*, vol. 45, no. 7, July 1997.

[80] M. Simon and M. S. Alouini, "A Unified Approach to the Probability of Error for Noncoherent and Differentially Coherent Modulations over Generalized Fading Channels," *IEEE Trans. on Communications*, vol. 46, no. 12, pp. 1625–1638, December 1998.

[81] V. Tarokh, H. Jafarkhani, and A.R. Calderbank, "Space-Time Block Codes from Orthogonal Designs," *IEEE Trans. on Information Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.

[82] V.A. Marchenko and L.A. Pastur, "Distribution of Eigenvalues for Some Sets of Random Matrices," *Math USSR Sb.*, vol. 1, pp. 457–483, 1967.

[83] J.W. Silverstein and Z.D. Bai, "On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices," *J. Multivariate Anal.*, vol. 54, no. 2, pp. 175–192, 1995.

[84] W. Rudin, *Fourier Analysis on Groups*, Wiley Classics Library. Wiley, 1990.