Robust Waterfilling strategies for the fading channel

Master SICOM (Université de Nice Sophia-Antipolis)

Alberto Suárez Real

Advisors: Merouane Debbah
Eitan Altman

Email : {albsuarez}@gmail.com

Abstract

This work is the result of a 3.5 month intership at INRIA Sophia Antipolis under the supervision of professors Merouane Debbah and Eitan Altman. In spite of the number of papers studying capacity and power allocation issues for wireless communications, not so many consider the increasingly important case of only partial (statistical) knowledge of the channel, without assuming any particular model for it. As a consequence, in this work some bounds on the capacity are obtained for the performance of different communication channels under those assumptions: In the first part, both SISO and MIMO channels (without power allocation) are considered where the actual realizations are not available at the transmitter (or corrupted by an additive noise), while in the second part, strategies for resource allocation in the context of OFDM systems are obtained. Two main performance measures are derived depending on the characteristics of the channel: the worst case and average capacity, but always under partial knowledge at the transmitter.

Resumé

Ce travail est le résultat d'un stage de 3.5 mois à l'INRIA Sophia Antipolis sous la supervision des professeurs Merouane Debbah et Eitan Altman. Malgré la grand quantité de papiers qui étudient thèmes de capacité et allocation de puissance dans les systèmes de communications mobiles, il sont pas si nombreux ceux qui étudient le chaque fois plus important cas ou la connaissance est seulement partiel (statistique), sans supposer aucun model en particulier. Donc, dans ce travail, quelques bornes sont obtenues pour la performance sous ces conditions, dans la première partie, systèmes SISO et MIMO sont considérés, dans lesquels, les réalisations instantanés du canal ne sont pas disponibles dans le émetteur (ou affectés par un bruit additif), et pour la dernière, stratégies pour allocation des ressources dans le contexte des systèmes OFDM sont obtenues, en considérant deux mesures de performance: capacité moyenne, ou dans le pire cas, selon les caractéristiques d'ergodicité du canal, mais toujours sous une connaissance partielle au émetteur.

Contents

List of Figures

Chapter 1

Introduction

The growth in wireless communication development during the last few years and the need of increasingly high transmission rates to be able to accommodate the new and forthcoming services, has resulted in a great amount of papers investigating the limits for the different kinds of related communication techniques. Different assumptions about the available information at either transmitter and receiver [1, 2] have been considered, using as the main measure of performance the channel capacity first introduced by Shannon [3], as the maximal achievable transmission rate without any errors.

Some of the new techniques recently introduced, such as OFDM and MIMO have promised possibilities of remarkable improvements in the resulting transmission rates achievable in a fading environment, by increasing noticeably the spectral efficiencies with respect to established systems, thus receiving a great deal of attention by researchers, reflected in the numbers of papers published on these topics during the last few years.

The first published papers focused on assumptions that can be judged as too idealistic to be used on actual communications systems, since they usually considered complete knowledge of the channel at both the receiver and the transmitter end, as [4] for the MIMO case, but these works allowed to show the possibilities opened by these systems, allowing a distinct gain over the already then known MISO and SIMO systems by adding multiple antennas to both transmitter and receptor. The perfect channel state information assumption at the transmitter was based on the possibility to retransmit this information from receiver to transmitter via a feedback link, or in duplex system using the same frequency and time intervals close enough with respect to the channel coherence time, so as to allow to use the information obtained when acting as receiver for transmitting purposes.

However, the steady increase in transfer rates and subsequent faster variation of the channel parameters with respect to the symbol time, have made it impossible to achieve perfect transmitter channel state information by means of a limited rate feedback link, and even in duplex systems the time or frequency bands used for transmission and reception are different in many systems, so those models cannot be considered valid anymore. This has motivated the later appearance of more realistic models in which an imperfect knowledge of the channel was considered, either as a result of errors committed in the channel estimation [5], because of the need to perform a coarse quantization of the parameters in order to accommodate them to a low-rate feedback link or because of the need to predict the channel state from previous information, due to delays when using a feedback link from receiver to transmitter, and forbids the accesss to the actual realization of the channel. But in these works, the uncertainty appears as a deviation from a certain, a priori known model.

Situations considering the possible lack of any instantaneous channel state information at the transmitter or even also at the receiver have also been considered [6]. However, in most of these results, even then, the channel statistics are assumed to be perfectly known, and to be one of the most often used statistics for this kind of channels (Rayleigh, Rice, Nakagami), which imposes unnecessary constraints, not directly derived from the actual knowledge about the channel. All this may result in a not very good fit to the actual channel under consideration, and thus not very accurate results or predictions for the transmission possibilities, since in different scenarios it has been shown that even slight deviations from the assumed distribution or model might result in an important performance degradation.

So, an interesting open problem for which only partial results exists (mainly in the context of asymptotical analysis, either for the SNR under consideration [7, 8, 9] or for the number of antennas in the MIMO case [10]),is to study the available transmission rates of only some channel statistics, reducing the assumptions about a particular model to the minimum, and making use of only some channel statistics knowledge at the transmitter. In this work, the estimation at the receiver will be assumed to be performed without error, while only some statistical knowledge is available at the transmitter, possibly sent through a limited capacity feedback link from the receiver.

The evaluation of the penalties derived from a certain limited statistical knowledge might be compared to the resources involved to achieve a bigger amount of information about the channel, and thus allow to achieve an optimal level of compromise between these two issues. So as to evaluate or bound the system performance, several measure tools can be used, most often the ergodic entropy, as will be in our case, considering a fast fading channel, which allows to use its averaging properties for the transmission of each symbol, since it will typically extend through many independent realizations of the fading process, whilst in other cases the outage capacity formulation will be more suitable if the fading is slow enough so that this averaging property is not available, and any given rate might not be achievable at certain instants, so the rate achievable during at least a certain percentage of time (except for the outages) is a more meaningful performance measure, or even robust capacity notions, in which the goal is to maximise the minimal rate over all possible channel realizations, to ensure the system performance under possible deviations from the considered channel characteristics, but this requires a limited set of states, all of them to be separated away from 0. As stated above, in these part only the first case will be considered, that is a fast fading channel with i.i.d. realizations.

The other problem under consideration in this work relates to the OFDM channel, in which not only the transmission rates should be determined, but also the power allocated to each of the different subchannels. Again the solution to this problem is well known in the case of complete channel state information at the transmitter, given by the waterfilling algorithm. There has also been abundant research in this field, but mainly relating to the minimization of error probabilities for a fixed modulation, and simpler algorithms to perform an approximate waterfilling, while reducing the complexity but not nearly as much attention has been payed to the consideration of only partial or statistical knowledge about the channel, which will be the main focus in this work, considering both worst-case and average capacities, applicable to different scenarios in terms of channel ergodicity, and different degrees of available knowledge at the transmitter.

Chapter 2

Capacity modelling under channel uncertainty

2.1 Channel modelling versus capacity prediction

The usual trend in communications systems when faced to the problem of assessing the behaviour under certain channel conditions with some degree of uncertainty included in them, has been to use some kind of model for the channel under consideration. For that purpose, it could be obtained either from the fitting of experimental data from measurements or from theoretical model selection techniques as could be mainly the maximum entropy principle, promulgated by Jaynes [11] in the statistical inference field. From that model, it would be possible to calculate system performance, and it could be optimized with respect to some possible parameters.

This scheme is depicted in Fig. 2.1 . However, there is an implicit problem with this procedure: the practical measures obtained, of the constraints imposed to the problem whether because of physical requirements or any other considerations as might be regulatory issues, will allow a number of different possible channels. And using a model implies choosing one of them, which would not necessarily be the actual one. In fact different methods using for the selection of the model, will result in a different result. So the results obtained either directly or in terms of upper and lower bounds for the system performance are just a subensemble of the whole space of possibilities if all the states allowed by the original constraint were considered. This may result in a system performing worst than the lower bound predicted by a particular model analysis, since the worst cases may be excluded from it, and there is no guarantee that the one chosen

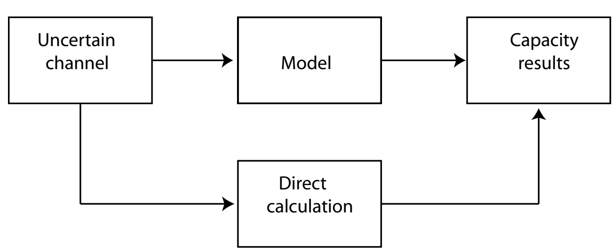

Figure 2.1: Scheme of modelling versus direct calculation

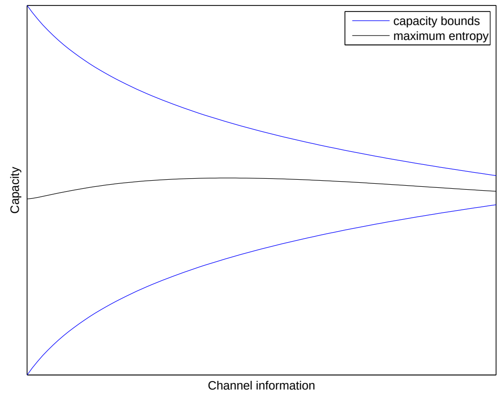

Figure 2.2: Capacity as a function of the information available about the channel

will be the right one. This is reflected in Fig. 2.2 where capacity is shown as a function of the knowledge available about the channel. As a bigger amount of information is known, the gap betwwen upper and lower bounds reduces, since in the limit, when the statistical behaviour is completely characterized, the exact capacity can be calculated. Meanwhile, the result obtained by maximum entropy knowledge for a certain degree of knowledge will lay between those two limits and will also vary with new available information

From the problems just considered results the intention to consider a more general scenario, thus avoiding this inherent loss of information and avoiding arbitrary choices. There is obviously also a price to pay, the increased complexity of the problem when less assumptions are being taken and the impossibility to use particular expressions corresponding to a certain channel. It may also happen that even choosing the appropriate constraints, the actual limit solutions found may not be reasonable for a real channel, and might lead thus to a bound which could be far too pessimistic to be useful when applied to a real communications systems, so in general a compromise will have to be achieved between these two extremes.

2.2 SISO Channels

For the problem under consideration, let us initially consider the study of a SISO channel characterised by

$$y = \rho h x + n \tag{2.1}$$

where y is the observation at the receiver, x the transmitted signal, h the realization of a fading process, ρ the average signal to noise ratio and n that of a white gaussian noise process. The only constraint is a normalised average energy of the channel $E[|h|^2] = 1$, so as to establish a fair comparison between different statistical behaviours, since the multiplying factor ρ will include the actual energy, this way also σ^2 can be assumed to be 1 without loss of generality.

In this case, when channel state information is available error-free at the receiver, and unavailable at the transmitter, assuming fast fading conditions and complete statistical knowledge of the channel also at the trasmitter, the capacity is given by the well known formula

$$C = \int_{-\infty}^{\infty} \log(1 + \rho|h|^2)p(h)\mathrm{d}h \tag{2.2}$$

However, when not all the statistical knowledge is available at the transmitter, it becomes impossible to calculate the capacity to determine the maximum possible transmission rates. Then, for a given state of knowledge, upper and lower bounds on the capacity will have to be derived for the different possible channels by solving the following optimization problem:

$$\begin{aligned} &\max_p \quad (\text{resp.} \min_p) \quad C_{av} = \max_p(\text{resp.} \min_p) \int_{-\infty}^{\infty} \log(1 + \rho|h|^2)p(h)\mathrm{d}h \\ &s.t \qquad E_h[|h|^2] = 1 \end{aligned} \tag{2.3}$$

For the maximisation, an immediate upper bound to the expression of the capacity can be obtained by applying Jensen inequality to $\log(1 + x)$, since it is a concave function of x, getting as a result the capacity of the AWGN channel with the same average energy, $E_{av} = E[|h|^2]$, this will always be an upper bound for the capacity as long as the energy constraint is imposed, no matter which others are, but those will make it achievable or not. On the other hand, for the minimisation there is not such an immediate solution, and as we will see afterwards, under the imposed constraint, the minimum is achieved by a degenerate distribution, which obtains a zero capacity. So with only that information about the channel the range of possible average channel capacities goes from 0 to C_{awgn}, and to obtain meaningful results some other constraints would have to be imposed, even though there is no other immediate one which might be applied to any case under consideration and different solutions might appear: constraints on higher order moments of the probability distribution of the channel transfer function, since the first ones might be quite easily measurable by the transmitter to calculate the achievable rates using that information, limited available peak energy, a given set of possible states with some fixed values, among others, and it would be mostly the nature of the problem which would suggest which one to choose. The problem in general is that the objective function under consideration is not convex, but concave, which complicates the solution of the optimization problem, since it will generally lay on the boundary determined by the constraints.

2.2.1 Finite-level channel

A simplified case which might allow to extract some interesting conclusions consists in considering a discrete set of possible states for the channel, which, when making this number go to infinity might as well be applied to the continuous case. The expression for the capacity would be the same, just replacing the integral in 2.3 by a sum over the different states, leading to

$$C_{av} = \sum_{i=1}^{N} log(1 + \rho|h_i|^2)p(h_i) \tag{2.4}$$

We may start by considering just two states, h_1 and h_2, with respective probabilities p_1 and $p_2 = 1 - p_1$. As mentioned before, if we impose no further constraints that the average energy $E_{av} = p_1|h_1|^2 + (1 - p_1)|h_2|^2 = 1$, the maximum capacity would be given by the non-random distribution of the same energy. And considering the distribution

$$h = \begin{cases} 0 & \text{with} \quad p_1 = 1 - \epsilon \\ h_0 & \text{with} \quad p_2 = \epsilon \end{cases}$$

with $|h_0|^2 = \frac{1}{\epsilon}$, so as to satisfy the energy constraint, which would result in a capacity given by

$$C = p_1 \log(1 + \rho|h_1|^2) + (1 - p_1)\log(1 + \rho|h_2|^2) = \epsilon \log(1 + \frac{\rho}{\epsilon}) \tag{2.5}$$

which, when taking ϵ going to zero, can easily be seen to vanish. So there is no positive lower bound for the capacity under this assumption, and this result can be immediately extended to a discrete model with an arbitrary number of states, by making the probability of the rest of them equal to zero, or to the continuous case, resulting then in delta functions centered at the same points, since those solutions do not contradict the imposed constraints. However it is evident that this bounds, although theoretically achievable with the chosen constraints, will give a very pessimistic prediction for most real channels, since for instance, in the most widely used probabilistic distributions for a continuum of possible channels, there are no upper nor lower bounds for the instantaneous values of the coefficients, which would result in a zero capacity bound, and correspondingly, for the upper bound, there will be states arbitrarily close to the mean energy, or that value will be by itself a possible outcome. With just 2 states it would not be possible to impose more constraints, since there are not enough degrees of freedom, so now an arbitrary number, N, of states will have to be considered, with $N > 2$.

Of course, when new constraints are introduced, the set of possibilities is reduced, so the previously mentioned bounds would still be valid, but they will become possibly too loose to be useful, so some tighter ones must be searched for.

2.2.2 Fixed states

We start in this case with N arbitrarily fixed levels, which we will consider without loss of generality to be ordered by increasing energies $\{|h_1|^2 \leq |h_2|^2 \leq ... \leq |h_N|^2\}$ and with associated probabilities $\{p_1, ..., p_N\}$. At the same time, the average energy constraint is retained. We will see that in this case the minimum capacity will be achieved by distributing the probability between just the 2 extreme values, that is $p_1 = p, p_N = 1 - p, p_{i \neq 1,N} = 0$, with p obtained from the energy constraint as $p = \frac{|h_N|^2 - 1}{|h_N|^2 - |h_1|^2}$, resulting in a capacity

$$C = \frac{|h_N|^2 - 1}{|h_N|^2 - |h_1|^2} \log_2(1 + |h_1|^2) + \frac{1 - |h_1|^2}{|h_N|^2 - |h_1|^2} \log_2(1 + |h_N|^2) \tag{2.6}$$

Where the energy levels are considered to have been normalized, so that $|h_1|^2 \leq 1$ To see it is indeed the minimum, let us suppose now a certain probability p_i would be 'transferred' from h_1 and h_n to any other state h_i with all the other thing remaining equal. In that case we may find λ so that $p_i\lambda|h_1|^2 + p_i(1-\lambda)|h_N|^2 = p_i|h_i|^2$, and applying Jensen's inequality to this subevent, it can be seen that the minimum capacity is effectively obtained by using just the 2 extreme valued states. This result might also be readily extended to the continuous case, the only problem being that the most usually found channel probability distributions are unbounded. For the upper bound, by an analogous procedure it can be shown that the maximum capacity will be obtained by distributing the probability between the 2 states closer to the average energy, one below and the other above.

Meanwhile if we want to construct a model from which to calculate the quantities of interest, the correct approach would be to use the maximum entropy model under the constraints under consideration. If only the energy constraint is imposed, this is well known [11, 12] to induce an exponential distribution for the channel energy, as would be the case for instance in a Rayleigh channel. If in addition the discrete set of energy levels is fixed,the probability distribution would be characterized by the optimisation problem

$$\begin{aligned} \max_{p} \quad & H = \sum_{i=1}^{N} p_i \log \frac{1}{p_i} \\ s.t \quad & \sum_{i=1}^{N} p_i = 1 \\ & \sum_{i=1}^{N} p_i |h_i|^2 = 1 \end{aligned} \tag{2.7}$$

which results also in an exponential distribution, with the decay adjusted to satisfy the constraint [13] $p_i = \frac{e^{\lambda |h_i|^2}}{\sum_{i=1}^{N} e^{\lambda |h_i|^2}}$. Obviously, in that case we will not obtain a range of values for the possible capacity, but just one value corresponding to the choice of a model taken, thus losing the information about the other possibilities that might be achievable.

2.2.3 Other bounds

A different approach consists in getting some bounds directly from the expression of the capacity in a continuous case. As will be seen later in the more general MIMO setting, when the aim is to obtain expressions for the capacity for low SNR, expansions (e.g polynomial) can be used, either directly of the logarithm inside the integral, or using some of those functions to apply Jensen's inequality. However, the rapid growth of the polynomial functions as the SNR increases, renders their use limited to a small range, typically up to just 1 or 2 dB. This will also happen to any function that asymptotically outweighs the logarithm for high SNR, however,the lower the order, the wider will be the range of application of the obtained bound, and so a possibility is to use the approximation $|h| \approx \log(1 + |h|^2)$, which is known to give good result for values of up to SNR=10dB, with a reduced error [14]. However it is not a lower bound, but just an approximation, surpassing the objective function for some values of $|h|$, but some functional concave in $|h|$ can be constructed, as in eq. 2.8, containing this approximation so that again Jensen inequality can be applied to obtain a bound on the expected capacity, which would be a function of just $\sqrt{SNR}$, instead of SNR^2 if we were using a higher order polynomial.

$$f(x) = -k\sqrt{x} + ln(x+1) \tag{2.8}$$

Where the value of k can be optimized so as to get the tighter bound while guaranteeing that it is a lower bound. To get it we calculate the second derivative

$$f''(x) = \frac{k}{4} x^{-3/2} - (x+1)^{-2} \tag{2.9}$$

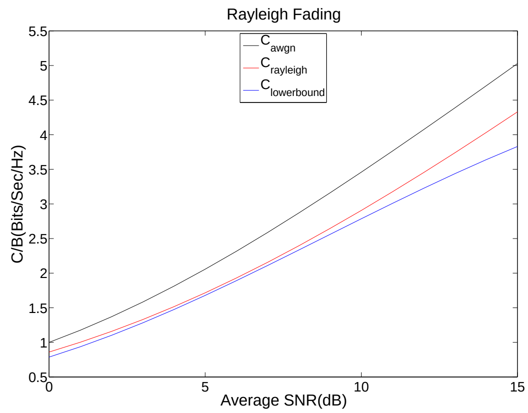

Figure 2.3: Actual capacity and lower bound with the approximation in $\sqrt{x}$

where the positive term is bigger for x going to 0 or ∞, so the value of k will be found for the maximum of the ratio between both terms. The obtained value is $k = \frac{3\sqrt{3}}{4}$, and applying Jensen's inequality as mentioned before the capacity can be lower bounded as

$$-k\sqrt{E[|h|^2]} + ln(1 + E[|h|^2]]) \leq E[ln(1 + |h|^2]) - k\sqrt{|h|^2]] \tag{2.10}$$

$$C(nats) \geq C_{AWGN}(nats) - k(\sqrt{E[|h|^2]} - E[|h|]) \tag{2.11}$$

$$C \geq C_{AWGN} - 1.87(\sqrt{E[|h|^2]} - E[|h|]) \tag{2.12}$$

In fig. 2.3, the results of the lower bound, compared to the actual capacity, when applied to a Rayleigh channel, have been shown. It can be appreciated that for values of SNR of up to 10 dB, the approximation can be considered quite good, but it obviously begins to lose tightness for higher values of SNR, however, for them the approximations available for the high SNR regime might already be used, so the bound under consideration might be useful in filling the mid SNR zone gap between both asymptotic zones, achieving a result for the capacity by just using the information about the first two moments of $|h|^2$

2.3 MIMO channels

The MIMO channel can be considered as a generalization of the SISO one, since by just taking the arbitrary number of antennas to be 1 at both transmitter and receiver, some of the results considered in this section apply directly to the previous one. However, the presence of multiples antennas also allows the use of some other techniques not available in the SISO case, most notably, when considering an asymptotically increasing number of antennas, growing at both transmitter and receiver with a fixed ratio between both quantities, allows to use results from random matrix theory [15] to characterize the system capacity, and what is more, those result can be found very often to approximate extremely well the actual capacity for a quite reduced number of antennas.

A general MIMO fading system with n_t antennas at the transmitter and n_r at the receiver will be characterised by the matrix equation

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{2.13}$$

where $\mathbf{x}$ and $\mathbf{y}$ are respectively the transmitted and received signal vectors, $\mathbf{n}$ is a vector of zero-mean complex Gaussian noise, spatially and temporally white and H is a $n_r x n_t$ matrix, where the coefficients are arbitrarily and identically distributed, with average energy $E[H_{ij}]^2 = \mathcal{E}$

The capacity for the general MIMO channel assuming perfect instantaneous channel knowledge at the receiver is given by

$$\begin{aligned} C &= \max_{Q} E[\log_2 \det(\mathbf{I} + \mathbf{H}\mathbf{Q}\mathbf{H}^{\dagger})] \\ s.t. \quad & Tr(Q) = P \end{aligned} \tag{2.14}$$

where $Q = E[\mathbf{x}\mathbf{x}^{\dagger}]$ is the covariance of the input signal, and in general will have to be calculated so as to maximize the capacity, but for the case where H has iid entries it was shown by Telatar [4], that capacity is achieved by an isotropic input, and that would also be the used one when no information is available at the transmitter, neither instantaneous nor about the possible correlation between the different subchannels. In that case, using the singular value decomposition of the matrix H, the capacity can be expressed in terms of its eigenvalues as

$$C = \sum_{i=1}^{n_{min}} \log_2(1 + \frac{\rho}{n_t}\lambda_i) \tag{2.15}$$

where $\frac{\rho}{n_t}$ is the average power in each receiving antenna, $\{\lambda_i\}$ the set of positive eigenvalues of $\mathbf{H}\mathbf{H}^{\dagger}$ and we define the minimum between the number of transmit and receive antennas, $n_{min} = \min\{n_t, n_r\}$.

So, having decomposed the channel in a set of parallel subchannels, the result obtained previously for SISO systems may be applied, substituting the distribution of the channel for the distribution of an unordered eigenvalue of the MIMO channel. Unfortunately, as far as we know, there are no expressions for the distribution of the mean as there exist for the greater and smaller eigenvalues, but this value might be obtained numerically.

2.3.1 Asymptotical results: Low and high SNR regimes

The low SNR regime is acquiring an increasing importance, since in many modern wireless communication systems, most users are working under these conditions. Moreover, it allows the analysis by means of series expansions, obtaining particularly simple expressions in function of the first statistical moments of the channel distribution to calculate the capacity. Assuming the receiver knows the realization of the channel matrix H, but the transmitter doesn't (nor it has statistical knowledge, in the sense of correlations between the different entries of H), we can follow the steps of Verdu in [9] to obtain the capacity, this time in function of the SNR

$$C = E\Big[\log \det[\boldsymbol{I} + \frac{1}{n_t}\boldsymbol{H}^{\dagger}\boldsymbol{H}\rho]\Big] \tag{2.16}$$

And taking into account that for an nxn matrix,

$$\frac{\mathrm{d}}{\mathrm{d}u}\log\det[\boldsymbol{I}+u\boldsymbol{A}]|_{u=0}=\mathrm{trace}(\boldsymbol{A})\log(e) \tag{2.17}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}u^2}\log\det[\boldsymbol{I}+u\boldsymbol{A}]|_{u=0}=-\mathrm{trace}(\boldsymbol{A}^2)\log(e) \tag{2.18}$$

thus for $\rho \to 0$ the expression for the capacity can be expanded in a Taylor series, with the normalised power per transmitter antenna resulting in an individual SNR $\frac{\rho}{n_t}$ as

$$C=\frac{\rho}{n_t}\ \mathrm{trace}E[\boldsymbol{H}^\dagger\boldsymbol{H}]-\frac{1}{2}\frac{\rho^2}{{n_t}^2}\mathrm{trace}E[(\boldsymbol{H}^\dagger\boldsymbol{H})^2] \tag{2.19}$$

If the elements of $\boldsymbol{H}$ are i.i.d. with zero mean and normalised energy,

$$\mathrm{trace}E[\boldsymbol{H}^\dagger\boldsymbol{H}]=\sum_{i=1}^{n_r}E[(\boldsymbol{H}^\dagger\boldsymbol{H})_{ii}]=n_r\sum_{j=1}n_tE[(\boldsymbol{H}^\dagger)_{ij}(\boldsymbol{H})_{ji}]=n_rn_tE[|H_{ij}|^2] \tag{2.20}$$

$$[(\mathbf{H}^\dagger\mathbf{H})^2]_{ij}=\sum_{l=1}^{n_t}(\mathbf{H}^\dagger\mathbf{H})_{il}(\mathbf{H}^\dagger\mathbf{H})_{lj}=\sum_{l=1}^{n_t}\sum_{p=1}^{n_r}\sum_{q=1}^{n_r}H^*_{pi}H_{pl}H^*_{ql}H_{qj} \tag{2.21}$$

$$E[H^*_{pi}H_{pl}H^*_{ql}H_{qi}]=\begin{cases} l=i,p=q & E[|H_{ij}|^4] \quad n_r terms \\ l=i,p\neq q & E[|H_{ij}|^2]^2 \quad n_r(n_r-1)terms \\ l\neq i,p=q & E[|H_{ij}|^2]^2 \quad n_r(n_t-1)terms \\ l\neq i,p\neq q & E[H_{ij}]^2E[H^*_{ij}]^2=0 \end{cases}$$

$$\mathrm{trace}E[(\boldsymbol{H}^\dagger\boldsymbol{H})^2]=n_tn_r(E[|H_{ij}|^4]+(n_r+n_t-2)E[|H_{ij}|^2]^2) \tag{2.22}$$

so that the expression in Eq. 2.19 can be further simplified as

$$C=\rho\ n_r-\frac{1}{2}\frac{n_r}{n_t}\rho^2\Big(\kappa(|H_{ij}|)+(n_r+n_t-2)\Big) \tag{2.23}$$

It can be noted that the first order term depends just of the number of receive antennas, but is independent of the channel under consideration, whilst the second-order one takes into account also the number of transmit antennas and is influenced by the channel distribution, but only through its kurtosis, defined as $\kappa(|H_{ij}|)=\frac{E[|H_{ij}|^4]}{E[|H_{ij}|^2]^2}$, which represents a measure of the dispersion in the channel distribution and thus decreases capacity due to its concavity.

For the high SNR case, we might obtain a first approximation to the asymptotic behaviour as the slope with respect to the logarithm. For that, we should consider the quotient

$$\lim_{\mathrm{SNR}\to\infty}\frac{C(\mathrm{SNR})}{\log_2\mathrm{SNR}} \tag{2.24}$$

often named as 'pre-log' or multiplexing gain, which, using the above mentioned decomposition in the eigenvalues can be expressed as

$$\lim_{\rho\to\infty} \frac{\sum_{i=1}^{n_{min}} \log_2(1+\frac{\rho}{n_t}\lambda_i)}{\log_2 \mathrm{SNR}} = \tag{2.25}$$
$$\lim_{\rho\to\infty} \frac{\sum_{i=1}^{n_{min}} \left[\log_2(\rho)+\log_2(\frac{1}{\rho}+\frac{\lambda_i}{n_t})\right]}{\log_2 \rho} =$$
$$n_{min}$$

Again just the slope would be the same for any given channel verifying the energy constraint, irrespective of the coefficient probability distributions or even the possible correlations present in the channel matrix, whilst in practice significative discrepancies may be appreciated, resulting in noticeable differences in the amount of power needed to achieve a given capacity. So a meaningful quantity is the so called power offset [7], defined as the zero-order term in the affine expansion of the capacity

$$\mathcal{S}_\infty = \lim_{\rho\to\infty}\left(\log_2(\rho) - \frac{\mathrm{C}(\rho)}{\lim_{\mathrm{SNR}\to\infty}\frac{\mathrm{C}(\rho)}{\log_2\rho}}\right) \tag{2.26}$$

which thus represent the excess energy needed to obtain the same performance as in an unfaded channel with orthogonal dimensions. It was calculated [16] using random matrix results for i.i.d distributed H but otherwise arbitrary distribution, and asymptotically in the number of antennas with fixed ratio $\beta = \frac{n_r}{n_t}$ as

$$\mathcal{S}_\infty = \begin{cases} (\beta-1)\log_2\frac{\beta-1}{\beta} + \log_2 e & \beta \geq 1 \\ \frac{1-\beta}{\beta}\log_2(1-\beta) + \log_2(\beta e) & \beta < 1 \end{cases}$$

2.3.2 Asymptotical results: High number of antennas

As opposed to what occurred in the previous section, where results were only available for low or high SNR regimes, the assumption of a big number of antennas (equal at transmitter and receiver $n_t = n_r = n$) allows to solve the capacity for arbitrary values of SNR [10], since then, for any arbitrary statistical distribution of the entries of $\mathbf{H}$, as long as they are independent and zero mean, the law of the eigenvalues of $\boldsymbol{H}^\dagger\boldsymbol{H}$ scaled by the dimension, $\frac{\lambda_i}{n}$ converges to the fixed quarter-circle law. Thus, by its integration the capacity per dimension (i.e. per antenna) can be obtained as

$$C = 2\log_2(1+\sqrt{4\rho+1)} - \frac{\log_2 e}{4\rho}\left(\sqrt{4\rho+1}-1\right) \tag{2.27}$$

It can be noted that this result, when particularized for the high high SNR case coincides with that of the previous section particularized for the same number of receive and transmit antennas.

Chapter 3

OFDM power allocation strategies

Even though the avantages of multicarrier modulations have been known for quite a long time, it has not been until recently that the interest in OFDM has raised, as a system that could achieve an increase in the transmission rates for wireless systems, being as it is particularly suited for multipath fading environments and combating the ensuing intersymbol interference, which has made it to be included in current and next generation wireless standards as IEEE 802.11a, IEEE 802.16 or IEEE 802.20.

The broadband signal results in a set of independent subchannels by means of an orthogonal transformation that can be efficiently implemented with the use of FFT, another of its main advantages. So the resulting system can be modelled by the set of equations corresponding to the different subchannels, and each of them is narrow enough so that the fading affecting it can be considered as flat.

$$y_i = h_i x_i + n_i \qquad i = 1, ..., N \tag{3.1}$$

As a result the total capacity of the OFDM system will be given be the sum of the capacities from each of the individual subchannels

$$C_{OFDM} = \sum_{i=1}^{N} C_i = \sum_{i=1}^{N} \log_2(1 + \frac{p_i |h_i|^2}{\sigma^2}) \tag{3.2}$$

where P_i is the power assigned to each of the subchannels, and perfect channel knowledge is assumed at the receiver, while the one present at the transmitter will influence in what way it is possible to distribute the power among the different subchannels, together with the constrains imposed on it, and whether they are for the average or instantaneous power.

OFDM is also employed in DSL systems (usually known as DMT in that case), and given the fixed nature of the channel, full channel state information can be assumed to be available at both receiver and transmitter, so the optimal power allocation is given by the well known waterfilling algorithm [13] among the different subchannels, by which better channels are allocated more power according to the formula

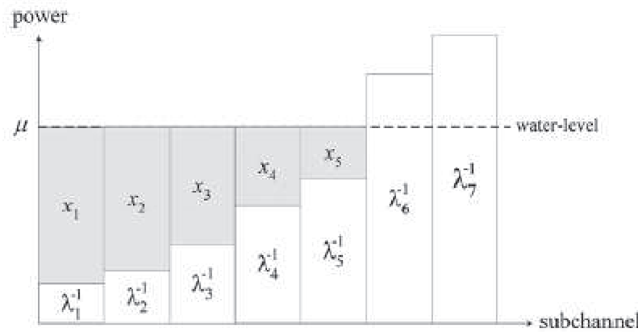

Figure 3.1: The waterfilling algorithm

$$p_i = \left(\frac{1}{\mu} - \frac{\sigma^2}{|h_i|^2}\right)^+ \tag{3.3}$$

$$\text{where } x^+ = \max(0, x) \tag{3.4}$$

$$\text{and } \mu \text{ is fixed to verify the power constraint} \quad \sum_{i=1}^{N} \left(\frac{1}{\mu} - \frac{\sigma^2}{|h_i|^2}\right)^+ = P \tag{3.5}$$

The result is illustrated in Fig. 3.1 and can be obtained in a simple way by means of the following iterative algorithm

Waterfilling Iterative Algorithm

1. Let P_1 be the subchannel with biggest SNR. Initialize $P_i = 0 \quad \forall i$
2. While $\left\|P - \sum_{i=1}^{N} P_i\right\| >$ tolerance
   3. $\Delta = \dfrac{P - \sum_{i=1}^{N} P_i}{N}$
   4. $P_1 = P_1 + \Delta$
   5. New water-level $\quad \lambda = P_1 + \sigma^2 |h_1|^2$
   6. Remaining powers $\quad P_i = \lambda - \dfrac{\sigma^2}{|h_i|^2} \qquad i = 2, ..., N$

end

However, when we are considering a wireless environment, two main differences appear: the channel gains for each subchannel are not fixed all through the transmission anymore so we will be interested in the mean throughput since it will now be a random variable (also the notion of worst case capacity, appearing in slow fading channels will be considered), and in addition, as has already been mentioned, it is often not possible to track the channel perfectly at the transmitter, so the information will only be partial. If it were complete, we would have again a waterfilling solution, but now at the same time bidimensionally in frequency (through the different subchannels) and time, to account for the temporal variations of channel gains. However, it might not be desirable to transmit a variable amount of power, so in that case it would be possible to restrain the application of the waterfilling algorithm to the subchannels, reverting to the precedent scenario of the DSL case, but with the relative allocations being time dependent (instead of a fixed threshold with would be the result of the bidimensional waterfilling when the constraint is imposed just on the average energy and not on the instantaneous one), and that is the scenario we will be considering, which can be formulated as the convex optimisation problem

$$\max_{S} \quad E\Big[\sum_{i=1}^{N} \log_2(1 + \frac{p_i|h_i|^2}{\sigma^2})\Big] \tag{3.6}$$

$$s.t \qquad \sum_{i=1}^{N} p_i \leq P \tag{3.7}$$

$$p_i \geq 0 \tag{3.8}$$

At the other extreme, if no CSI is available at the transmitter the only possibility would be to allocate the power equally to all the different subchannel. Intermediate settings appear when the knowledge is only partial, with the particular case of statistical knowledge, that will be considered here. Furthermore, the fading distribution of the different subchannels will be assumed to be equal, although this might not be necessarily the case, but with possibly different parameters characterizing it, which would suit well a multiuser system with OFDMA access, in which the mean energy of the channel available for each user and other possible parameters will vary, as a function, among other things, of its relative position and distance to the base station, and it is something that is also much easier to track that the whole instantaneous channel state information. Some related work is developed in [17], but in that case the partial knowledge is the result of an error committed in the estimation of the channel coefficients. Depending on the ergodicity or not of the channel, average or worst case capacities will be considered respectively in the following two sections

3.1 Ergodic capacity

When we are considering an ergodic channel, as mentioned before, the most useful notion is that of average capacity to measure its performance.

Here two main situations will be considered: knowledge of the full statistical distribution for each of the subchannels, or just of their respective mean energies. In any of them, since no instantaneous information is available the allocation will have to be fixed a priori once the statistics are known, and be kept like that during all the transmission irrespective of the instantaneous fluctuations of the different subchannels. For the former, it would be possible to numerically calculate the optimal allocation, while for the latter an intuitive approach, although a priori suboptimal, would be to apply the waterfilling algorithm to the mean energy values to obtain the corresponding powers. These three strategies require decreasing amounts of information about the channel, and so would be the resources needed to get them. Particularly appealing would be the mean information case, since in case the incurred loss would not be too significant, it would mean just some reduced amount of information would be needed, while allowing good performance.

By considering the duality gap associated with the convex optimization problem, a general bound for the loss incurred from using a given arbitrary power allocation instead of the optimal waterfilling solution is given by [18]

$$\Gamma = \frac{1}{\ln 2}\Bigg[\sum_{i=1}^{m}\big(\frac{S_i}{\min_j\{S_j + \frac{\sigma^2}{|h|^2}\}} - \frac{S_i}{s_i + \frac{\sigma^2}{|h|^2}\}}\big)\Bigg] \tag{3.9}$$

where m is the number of subchannels which are allocated a positive amount of power. For our problem we will be interested in the expectation of the gap over the channel probability distribution. For the case of Rayleigh fading, the instantaneous energy is exponentially distributed and since the subchannels are independent of each other

$$\text{prob}\Big(\min_j\{P_j + \frac{\sigma^2}{|h_j|^2}\} \geq x\Big) = \tag{3.10}$$

$$\prod_{i=1}^{m}\text{prob}\big(P_j + \frac{\sigma^2}{|h_j|^2} \geq x\big) = \tag{3.11}$$

$$\prod_{i=1}^{m}\text{prob}\big(|h_j|^2 \leq \frac{\sigma^2}{x - P_j}\big) = \tag{3.12}$$

$$\prod_{i=1}^{m}\big(1 - e^{-\frac{\sigma^2}{(x-P_j)E[|h_j|^2]}}\big) \tag{3.13}$$

and the associated pdf can be obtained by differentiating with respect to the variable x as

$$f(x) = \sum_{i=1}^{m} e^{-\frac{\sigma^2}{(x-P_j)E[|h_j|^2]}}\frac{\sigma^2}{E[|h_j|^2]}\frac{1}{(x-P_j)^2}\prod_{k=1,k\neq i}^{m}\big(1 - e^{-\frac{\sigma^2}{(x-P_j)E[|h_j|^2]}}\big) \tag{3.14}$$

We may particularize this expression for the mentioned solution of performing the waterfilling on the average energies of the different subchannels, which gives the power assignment

$$P_i = \left(\frac{1}{\mu} - \frac{\sigma^2}{E[|h_i|^2]}\right)^+ \tag{3.15}$$

with μ fixed to verify the power constraint $$\sum_{i=1}^{N}\big(\frac{1}{\mu} - \frac{\sigma^2}{E[|h_i|^2]}\big)^+ = P \tag{3.16}$$

When all the statistical knowledge about the channel is known there appear some slight differences. The optimization problem is the same as for the knowledge of just the mean in eq. 3.6 , but we may profit from the extra information. The dual problem may be formulated with the corresponding Lagrangian multiplier λ as

$$\max_{P} \quad E\Big[\sum_{i=1}^{N} \log_2(1+\frac{P_i|h_i|^2}{\sigma^2}) - \lambda\Big(\sum_{i=1}^{N} P_i - P\Big)\Big] \tag{3.17}$$

and deriving with respect to the P_i and equating to 0 to obtain the maximum, the following set of equations is obtained

$$\begin{aligned} &E\Big[\frac{\frac{\rho_i|h_i|^2}{\sigma^2}}{(1+\frac{P_i\rho_i|h_i|^2}{\sigma^2})}\Big] - \lambda = \\ &\frac{1}{P_i}\Big(1 - E\big[\frac{\sigma^2/P_i\rho_i}{\sigma^2/P_i\rho_i + |h_i|^2}\big]\Big) - \lambda = 0 \qquad i = 1, ..., N \end{aligned} \tag{3.18}$$

This would be a general expression for any given distribution. In the case of an exponential distribution of the channel energy instantaneous values, as is the case for Rayleigh distributed coefficients, the expectation can be further explicited as

$$\begin{aligned} &E\big[\frac{\sigma^2/P_i\rho_i}{\sigma^2/P_i\rho_i + |h_i|^2}\big] = \\ &\int_0^{\infty} \frac{\sigma^2/P_i\rho_i}{\sigma^2/P_i\rho_i + \mathcal{E}} e^{-\mathcal{E}_i} d\mathcal{E}_i = \\ &\frac{\sigma^2}{P_i\rho_i} e^{\frac{\sigma^2}{P_i\rho_i}} E_i(\frac{\sigma^2}{P_i\rho_i}) \\ &\text{where} \quad \mathcal{E}_i = |h_i|^2 \quad \text{and } E_i \text{ denotes the exponential integral} \end{aligned} \tag{3.19}$$

getting thus an explicit expression of the power assigned to each subchannel that must be constant for all those that are assigned a positive amount of power

$$\frac{1}{P_i}\Big(\frac{\sigma^2}{P_i\rho_i} e^{\frac{\sigma^2}{P_i\rho_i}} E_i(\frac{\sigma^2}{P_i\rho_i})\Big) = \lambda \qquad 1 = 1, ..., m \tag{3.20}$$

A priori, it would not be possible to apply the same iterative algorithm as in the usual waterfilling case, since here we do not have anymore a linear function of the water level to assign the powers to each of the subchannels, so that for a given increment in the power assigned to the stronger one, the increase in some other subchannel may be bigger, thus eventually surpassing the total amount of available power. However the properties of the function under consideration, essentially its monotonicity, allow to use a bisection numerical algorithm to solve for λ and guarantee its convergence, getting a solution that complies with the power constraint.

In fact, for the knowledge of just the mean, a problem would appear, since the transmitter would be unable to calculate the rate to which transmit, since for that the whole channel distribution would be needed. Nevertheless, the idea would still be useful, taking into account the negligible performance difference of both systems, even with full statistical knowledge, this scheme might be applied, given its much lower complexity, obtaining that way the power allocation, and using afterwards the full statistics to calculate the transmission rates.

3.2 Worst case noise

We start by considering the related problem of the worst case noise [19]. In this case, the noise will have any possible power spectral density $q = (q_1, ..., q_N)$, with the constraint $\sum_{i=1}^{N} q_i \leq Q$ The model becomes a compound channel, with the capacity given by

$$\min_q \max_p \sum_{i=1}^{N} \log(1 + \frac{|h_i|^2 p_i}{q_i}) \tag{3.21}$$

By noting that C is concave in p, convex in q and the constraints also, the solution will be unique and can be obtained by forming the Lagrangian

$$\mathcal{L} = \sum_{i=1}^{N} \log_2(1 + \frac{|h_i|^2 p_i}{q_i}) + \mu\Big(\sum_{i=1}^{N} q_i - Q\Big) - \frac{1}{\epsilon}\Big(\sum_{i=1}^{N} p_i - P\Big) \tag{3.22}$$

which differentiating and equating to 0 gives the conditions:

$$\begin{aligned} p_i &\geq \epsilon - \frac{q_i}{|h_i|^2} && \text{with equality if } p_i \geq 0 \\ q_i &\geq \frac{|h_i|^2 p_i}{{q_i}^2 + |h_i|^2 p_i q_i} && \text{with equality if } q_i \geq 0 \end{aligned} \tag{3.23}$$

Assuming $p_i > 0 \; \forall i$, and that for some $\rho > 0, p_i = \rho q_i$ and $\epsilon\mu = \rho$ we can substitute in the conditions above to obtain

$$\begin{aligned} p_i &= \epsilon \frac{\rho |h_i|^2}{1 + \rho |h_i|^2} \\ q_i &= \frac{1}{\mu} \frac{\rho |h_i|^2}{1 + \rho |h_i|^2} \end{aligned} \tag{3.24}$$

which verify the LKT conditions, and verify $p_i = \rho q_i \forall i$, so that $\rho = P/Q$. Then the power allocation is obtained as

$${p_i}^* = P \frac{\frac{\frac{P}{Q}|h_i|^2}{1+\frac{P}{Q}|h_i|^2}}{\sum_{k=1}^{N} \frac{\frac{P}{Q}|h_k|^2}{1+\frac{P}{Q}|h_k|^2}} \tag{3.25}$$

and the worst noise by

$${q_i}^* = \frac{Q}{P} {p_i}^* \tag{3.26}$$

resulting in a capacity

$$C = \sum_{i=1}^{N} \log_2(1 + \frac{P}{Q}|h_i|^2) \tag{3.27}$$

which would be the same as in the case of a frequency flat noise but with no channel knowledge at the transmitter.

3.3 Worst case channel

Another case of interest consists in analyzing the worst case capacity when the noise variances are fixed (possibly unknown at the transmitter) and the carrier gains are allowed to variate while verifying a certain constraint. This is specially true for cases, such as slow fading, in which the averaging capacities of the channel may not be available for a reasonable delay constraint for the transmission. In that case, transmission at the worst rate guarantees error free communication under any possible conditions of the channel, although it might give a pessimistic result, and consideration of outage capacities, allowing for a certain amounts of time in which transmission is not possible might be an alternative.

In this section we assume again that the channel realizations are not known at the transmitter, and the channel is characterized by a constraint on the minimum total energy available, guaranteeing thus a minimum rate

$$\sum_{i=1}^{N} |h_i|^2 \geq \mathcal{E} \tag{3.28}$$

The problem is then given by the optimization with respect to the powers assigned to each channel of

$$\max_{p} \min_{h} \sum_{i=1}^{N} \log_2(1 + \frac{|h_i|^2 p_i)}{{\sigma_i}^2} \tag{3.29}$$

which can be interpreted in the context of game theory, as a game between the user which must choose the powers assigned to each subchannel and a malicious nature that will choose the worst possible channel for those powers, under the imposed constraints.

Several possibilities will be considered depending on the degree of knowledge at the transmitter, and mainly if the noise powers for each of the subchannels are known or not.

3.3.1 Unknown subchannel noise powers

Since no knowledge is available, there is no reason to give preference to one of the subchannels with respect to the others, and the uniform power allocation seems the most natural choice.

In effect, we will show that it is indeed the optimal one. For $p_i = \frac{P}{N}$, the minimum capacity, given the concave nature of the log function, will be achieved by the channel concentrating all its energy in the worst subcarrier. This can be seen by noting the quasiconvex character of the logarithm [20] and the resulting first order condition for a global constrained minimum at x^*

$$\Delta f(x^*)(y - x) > 0 \; \forall y \quad \text{feasible} \tag{3.30}$$

in our case, we can write $C = \sum_{i=1}^{N} \log_2(1+k_i g_i)$ and considering the k_i ordered increasingly, $x* = (\mathcal{E}, 0, ..., 0)$ and $\Delta f(x^*) = (\frac{k_1}{1+k_1\mathcal{E}}, k_2, ..., k_N)$, so its immediate to verify that the condition is satisfied.

And since the corresponding noise variances are unknown at the transmitter, they only introduce a probabilistic behaviour, but which doesn't affect the choice, since the probability to be affected by a certain noise would be independent of our election.

On the other hand, if a different power assignment was chosen, it would mean that there would be a channel, i, with $p_i \leq \frac{P}{N}$, so using the same distribution for the channel, concentrating all energy on the subchannel i, would result in a lower capacity. Thus, it has been shown that the uniform allocation is optimum, resulting in a worst case capacity $C_{wc} = \log_2(1 + \frac{P\mathcal{E}}{N\sigma_{max}^2})$ since the channel chosen will be the noisiest one. However, some knowledge about the noise would have to be needed in order to calculate the rate.

3.3.2 Known subchannel noise powers

Now the average noise powers $\{{\sigma_1}^2, ..., {\sigma_N}^2\}$, possibly different, are assumed to be known at the transmitter, so this information may be used to increase the worst case capacity with respect to a uniform allocation. The problem may be characterized again as

$$\max_p \min_h \sum_{i=1}^{N} \log_2(1 + \frac{|h_i|^2 p_i)}{{\sigma_i}^2} \tag{3.31}$$

Again, the channel will be chosen to minimize the capacity, concentrating all its energy $\mathcal{E}$ on the most unfavorable carrier in terms of $\frac{p_i}{{\sigma_i}^2}$. Then the optimal assignment must guarantee that all the ratios are equal $\frac{p_i}{{\sigma_i}^2} = k$. Inserting this condition into the power constraint, the value of the ratio is obtained as $k = \frac{P}{\sum_{i=1}^{N} {\sigma_i}^2}$, and $p_i = \sigma_i^2 \frac{P}{\sum_{i=1}^{N} {\sigma_i}^2}$, thus the optimal strategy consists in performing the inversion of noise variances, with a worst case capacity given by $C_{wc} = \log_2(1 + \frac{P\mathcal{E}}{\sum_{i=1}^{N} \sigma_i^2})$. Evidently, in the case of equal noise variances, the previous result just reduces to a uniform power allocation.

3.4 Simulation results

In Fig.3.2 and Fig.3.4 the results of the average rate for N=64 subchannels with average subchannel gains obtained randomly, and the different considered strategies are shown, with respect to the average SNR and $\frac{E_b}{N_0}$ respectively, obtained through Monte Carlo simulation. It can be seen that the gain obtained by using the whole distribution instead of just the mean is negligible, and that is the case in other scenarios with different average energies for each of the subchannels or a different number of subchannels, that have also been simulated, without obtainig remarkable differences with respect to the example shown here. Even the difference to the optimal waterfilling performed with the instantaneous channel realizations is not too important, and less so as SNR increases (obviously so, since then all the strategies tend to the equal allocation among subchannels, which is asymptotically optimal). Also the gain obtained by the waterfilling on the mean compared to the equal power allocation is a function of the divergence between subchannel mean energies, since that is what determines the difference in the allocations. The case of

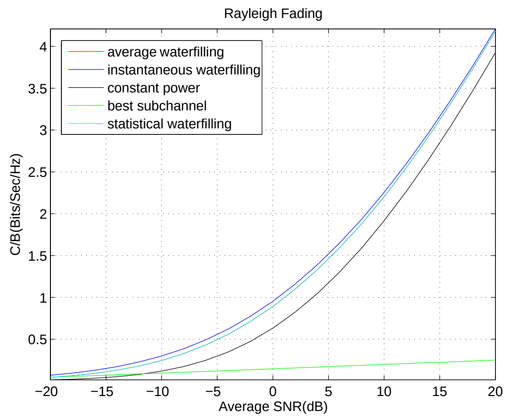

Figure 3.2: Capacity of an OFDM system with 64 carriers as a function of SNR

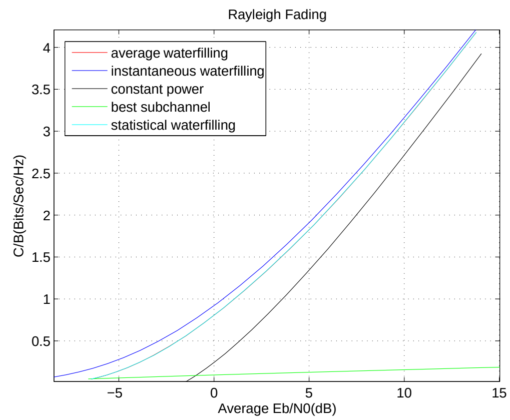

Figure 3.3: Capacity of an OFDM system with 64 carriers as a function of E_b/N_0

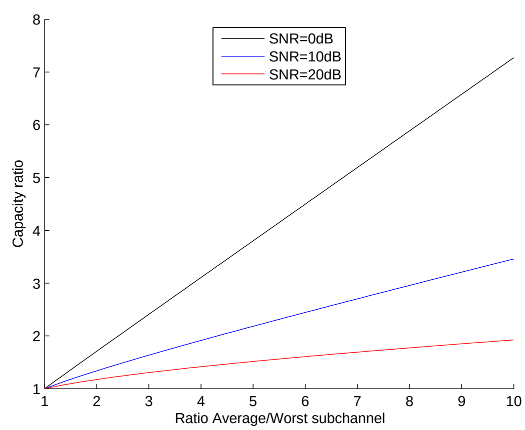

Figure 3.4: Ratio of worst case capacities with or without knowledge of the noise on each carrier

transmission only through the best subchannel has been added for reference since it is optimal for the low SNR scenario.

As for the worst case capacity, it has been shown that the gain obtained by the knowledge of the average noise power in each subchannel will be a function of just the ratio between the average and worst. In Fig.3.4 the ratio between the capacity with and without the information is shown in function of the above mentioned parameter for different values of the average SNR. It can be appreciated, that the influence is bigger for lower SNRs, but in any case, it can be very noticeable if some of the subchannels is rather weak.

Chapter 4

Conclusions and further work

Some bounds have been obtained, allowing to get a characterization of the performance of a fading communications system under different degrees of knowledge at the transmitter. The knowledge is always supposed to be incomplete, as is the case in usual wireless communication systems.

However, the results are not always sufficiently precise to be readily used, and different constraints allowing for tighter characterizations would have to be considered in further work, and other performance measures as outage capacity will help to get a better characterization of the systems.

Moreover, only some of the possible cases have been considered, having many natural extensions in terms of the study of multiuser systems, in which the interferences between them also plays a role, or the consideration of MIMO systems taking into account more complex models including correlation between the different components. Also the combined MIMO-OFDM systems, considered as possible candidates for new wireless standards would be part of that study.

Bibliography

[1] E. Biglieri, J. Proakis, and S. Shamai(Shitz), “Fading channels: Information-Theoretic and Communications Aspects,” *IEEE Transactions on Information Theory*, pp. 2619–2692, Oct. 1998.

[2] A.J. Goldsmith and P.P. Varaiya, “Capacity of fading channels with channel side information,” *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[3] C.E. Shannon, “A mathematical theory of communication,” *Bell Sys. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.

[4] I.E. Telatar, “Capacity of multi-antenna gaussian channels,” *European Transactions on Telecommunications and Related Technologies*, vol. 10, no. 6, pp. 585–596, 1999.

[5] M. Medard, “The effect upon channel capacity in wireless communication of perfect and imperfect knowledge of the channel,” *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 933–946, may 2000.

[6] A. Lapidoth and S.M. Moser, “Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels.,” *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2426–2467, 2003.

[7] A. Lozano, A.M. Tulino, and S. Verdu, “High-SNR power offset in multiantenna communication,” *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4134–4151, Dec. 2005.

[8] A. Lozano, A.M. Tulino, and S. Verdú, “Multiple-antenna capacity in the low-power regime,” *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2527–2544, Oct. 2003.

[9] S. Verdu, “Spectral Efficiency in the Wideband Regime,” *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[10] C-N. Chuah, D. Tse, J. M. Kahn, and R. A. Valenzuela, “Capacity Scaling in MIMO Wireless Systems Under Correlated Fading,” *IEEE Transactions on Information Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

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

[12] J.N. Kapur and H.K. Kesavan, *Entropy Optimization Principles with Applications*, Academic Press, 1997.

[13] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley, 1991.

[14] M. Dohler and H. Aghvami, "On the approximation of mimo capacity," *IEEE Transactions on Wireless Communications*, vol. 4, no. 1, pp. 30–34, Jan. 2005.

[15] A.M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*, Now Publishers, 2004.

[16] A. Lozano and A.M. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Transactions on Information Theory*, vol. 48, no. 12, pp. 3117–3128, Dec. 2002.

[17] Y.Yao and G.B. Giannakis, "Rate-maximizing power allocation in ofdm based on partial channel knowledge," *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 1073–1083, May 2005.

[18] W. Yu and J. Cioffi, "Constant power water-filling: Performance bound and low-complexity implementation," *IEEE Transactions on Communications*, vol. 54, no. 1, pp. 23–28, Jan. 2006.

[19] E.A. Jorswieck and H. Boche, "Performance analysis of capacity of mimo systems under-multiuser interference based on worst-case noise behavior," *EURASIP Journal on Wireless Communications and Networking*, vol. 2, pp. 273–285, Jan. 2005.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.