Notation

a	Scalar.
$\boldsymbol{a}$	Vector.
$\boldsymbol{A}$	Matrix.
$\boldsymbol{A}^{-1}$	Inverse of the matrix $\boldsymbol{A}$.
$\boldsymbol{A}^{\dagger}$	Pseudo-inverse of the matrix $\boldsymbol{A}$.
$\boldsymbol{A}^{T}$	Transpose of a matrix.
$\boldsymbol{A}^{*}$	Complex conjugate of a matrix.
$\boldsymbol{A}^{H}$	Transpose and complex conjugate of a matrix (Hermitian).
$\mathbf{0}$	Vector/matrix of zeros of the appropriate dimensions.
$\boldsymbol{I}$	Identity matrix of the appropriate dimensions.
$\operatorname{rank}(\boldsymbol{A})$	Rank of the matrix $\boldsymbol{A}$.
$\ker(\boldsymbol{A})$	Kernel/null-space of the matrix $\boldsymbol{A}$.
$\operatorname{blkdiag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N)$	Block diagonal matrix formed with the matrices $\{\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N\}$.
$\operatorname{diag}(a_1, \ldots, a_N)$	Diagonal matrix whose main diagonal is $\{a_1, \ldots, a_N\}$.
$\|\boldsymbol{a}\|_2$	Norm-2 (Euclidean norm) of the vector, see (1.45).
$\boldsymbol{\nabla}_{\boldsymbol{x}} f(\boldsymbol{x})$	Gradient of a function $f(\boldsymbol{x})$, (1.49)
$\mathbb{E}\{\cdot\}$	Statistical expectation.
$\log_2(\cdot)$	Base 2 logarithm.
$\ln(\cdot)$	Natural (base e) logarithm.
$[a]^{+}$	$\max(0, a)$.

Acronyms

AWGN Additive White Gaussian Noise.

BD Block Diagonalization.

BS Base Station.

KKT Karush-Kuhn-Tucker.

PAPC Per Antenna Power Constraint.

PBPC Per Base Station Power Constraint.

SVD Singular Value Decomposition.

TPC Total Power Constraint.

Chapter 1

System Model

1.1 System Model

The system that will be considered throughout this work aims to represent the downlink of a canonical cellular network, comprising several identical cells, layed out over a regular hexagonal grid.

When studying a cellular network, the cells located at the edge of the network will not experience the same conditions as the cells in the center of the network. A typical way to deal with this situation is to consider a scenario that wraps around (Figure 1.1) so that cells on one side of the scenario affect cells on the opposite side. Another option is to consider a scenario with more cells than necessary, and then analyze the behavior of the cells located within the center of the network (Figure 1.2) , so that the exterior cells account for the interference, equaling the conditions of all the cells in the network.

Each cell in the system under study will be served with a single Base Station (BS) that is equipped with t transmit antennas. Each of the users considered in the system has r receive antennas.

A system with M BS and N users can then be modelled as

$$\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \tag{1.1}$$

$\boldsymbol{y}$ represents the signal received at all the users and is defined as

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix} \in \mathbb{C}^{Nr \times 1} \tag{1.2}$$

where $\boldsymbol{y}_i \in \mathbb{C}^{r \times 1}$ is the signal received at the i-th user.

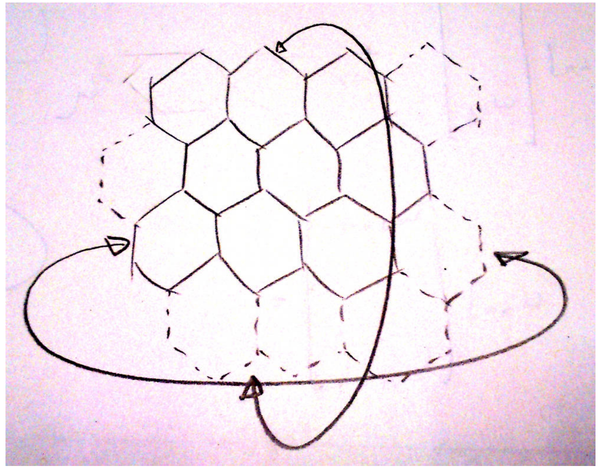

Figure 1.1: Wrap around scenario. Cells on one side of the scenario influence the cells on the other side as if they were next to each other.

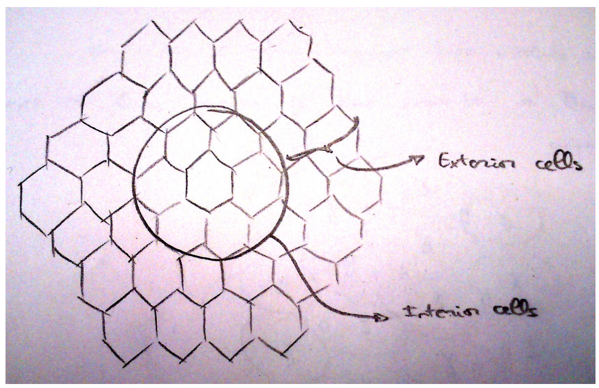

Figure 1.2: Oversized scenario. The behavior of the network is analyzed in the central cells of the network, and the exterior cells compensate the network edge effects.

$\boldsymbol{H}$ is the channel matrix from all the BS to all the users, with the following structure

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_{11} & \cdots & \boldsymbol{H}_{1M} \\ \vdots & \ddots & \vdots \\ \boldsymbol{H}_{N1} & \cdots & \boldsymbol{H}_{NM} \end{bmatrix} \in \mathbb{C}^{Nr \times Mt} \tag{1.3}$$

where $\boldsymbol{H}_{ij} \in \mathbb{C}^{r \times t}$ represents the channel matrix from the j-th BS to the i-th user. It will include the path loss due to propagation, small scale fading, shadowing, and any other characteristic of the radio channel that needs to be taken into consideration.

$\boldsymbol{x}$ is the signal transmitted from all the BS, and it is composed of

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_M \end{bmatrix} \in \mathbb{C}^{Mt \times 1} \tag{1.4}$$

where $\boldsymbol{x}_j \in \mathbb{C}^{t\times 1}$ is the signal transmitted by the j-th BS. Additionally, the power transmitted by the j-th BS can be calculated from the transmitted signal as

$$P_{j,\mathrm{tx}} = \mathrm{Tr}\left(\boldsymbol{x}_j \boldsymbol{x}_j^H\right) = \boldsymbol{x}_j^H \boldsymbol{x}_j \tag{1.5}$$

and each BS will have an independent power constraint

$$P_{j,\mathrm{tx}} \leq P_{j,\max} \tag{1.6}$$

Finally $\boldsymbol{n}$ represents the Additive White Gaussian Noise (AWGN) at all the receivers

$$\boldsymbol{n} = \begin{bmatrix} \boldsymbol{n}_1 \\ \vdots \\ \boldsymbol{n}_N \end{bmatrix} \in \mathbb{C}^{Nr\times 1} \tag{1.7}$$

with $\boldsymbol{n}_i \in \mathbb{C}^{r\times 1}$ accounts for the Gaussian noise at the i-th receiver. Throughout this work $\boldsymbol{n}_i$ is considered to be formed by *iid* entries, drawn from a zero mean, σ_i^2 variance Gaussian distribution, $\boldsymbol{n}_i \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_i^2 \boldsymbol{I}\right)$. The noise variance will be assumed the same for all the receivers.

In a general scenario, there may be cooperation among the BS in the system so that the information intended for a particular user will be transmitted by several or all the BS. Or, equivalently, each BS transmits a combination of the information of several users

$$\boldsymbol{x}_j = \boldsymbol{W}_{j1}^{(\mathrm{tx})} \boldsymbol{s}_1 + \cdots + \boldsymbol{W}_{jN}^{(\mathrm{tx})} \boldsymbol{s}_N \tag{1.8}$$

where $\boldsymbol{s}_i \in \mathbb{C}^{\ell\times 1}$ is the vector of information symbols to be transmitted to user i, with ℓ being the number of simultaneous symbols or streams to be transmitted to that user. $\boldsymbol{W}_{ji}^{(\mathrm{tx})} \in \mathbb{C}^{t\times \ell}$ is the precoding matrix used at the j-th transmitter for the data of the i-th user.

It will be assumed that the information symbols are independent and drawn from a Gaussian distribution such that $\boldsymbol{s}_i \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{R}_{\boldsymbol{s}_i}\right)$, where $\boldsymbol{R}_{\boldsymbol{s}_i} = \mathrm{diag}\left\{p_{i1}, \ldots, p_{i\ell}\right\} \in \mathbb{R}^{\ell\times\ell}$ contains the power allocated to each of the symbols in $\boldsymbol{s}_i$. The transmitted power can be expressed as

$$P_{j,\mathrm{tx}} = \sum_{i=1}^{N} \mathrm{Tr}\left(\boldsymbol{W}_{ji}^{(\mathrm{tx})} \boldsymbol{R}_{\boldsymbol{s}_i} \boldsymbol{W}_{ji}^{(\mathrm{tx}),H}\right) \tag{1.9}$$

The choice of the precoding matrices and of the receiving filter will determine the transmission strategy used. This Thesis will focus mainly on Block Diagonalization (BD) [1] which is described in Section 1.2.

On the receiver side, no cooperation among the users will be considered, so each user may perform, independently, additional processing of the received signal by applying a linear filter or equalizer

$$\hat{\boldsymbol{s}}_i = \boldsymbol{W}_i^{(\mathrm{rx})} \boldsymbol{y}_i \tag{1.10}$$

where $\boldsymbol{W}_i^{(\mathrm{rx})} \in \mathbb{C}^{r \times \ell}$ is the equalizer used at the i-th receiver.

Combining (1.1)–(1.8) it is possible to rewrite (1.1) as

$$\boldsymbol{y} = \boldsymbol{H} \boldsymbol{W}^{(\mathrm{tx})} \boldsymbol{s} + \boldsymbol{n} \tag{1.11}$$

where

$$\boldsymbol{s} = \begin{bmatrix} \boldsymbol{s}_1 \\ \vdots \\ \boldsymbol{s}_N \end{bmatrix} \in \mathbb{C}^{N\ell \times 1} \tag{1.12}$$

and the global precoding matrix is

$$\boldsymbol{W}^{(\mathrm{tx})} = \begin{bmatrix} \boldsymbol{W}_{11}^{(\mathrm{tx})} & \cdots & \boldsymbol{W}_{1N}^{(\mathrm{tx})} \\ \vdots & \ddots & \vdots \\ \boldsymbol{W}_{M1}^{(\mathrm{tx})} & \cdots & \boldsymbol{W}_{MN}^{(\mathrm{tx})} \end{bmatrix} \in \mathbb{C}^{Mt \times N\ell} \tag{1.13}$$

and it can be partitioned as

$$\boldsymbol{W}^{(\mathrm{tx})} = \left[\boldsymbol{W}_1^{(\mathrm{tx})}, \dots, \boldsymbol{W}_N^{(\mathrm{tx})} \right] \tag{1.14}$$

with

$$\boldsymbol{W}_i^{(\mathrm{tx})} = \begin{bmatrix} \boldsymbol{W}_{11}^{(\mathrm{tx})} \\ \vdots \\ \boldsymbol{W}_{M1}^{(\mathrm{tx})} \end{bmatrix} \in \mathbb{C}^{Mt \times \ell} \tag{1.15}$$

The channel matrix can then be partitioned as

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_1 \\ \vdots \\ \boldsymbol{H}_N \end{bmatrix} \tag{1.16}$$

where $\boldsymbol{H}_i \in \mathbb{C}^{r \times Mt}$ is the channel matrix from all the BS to the i-th user.

As it has already been mentioned, receiver cooperation is not going to be considered, so looking at a particular user, *e.g.* the i-th user, the signal that is received will be

$$\boldsymbol{y}_i = \boldsymbol{H}_i \boldsymbol{W}^{(\mathrm{tx})} \boldsymbol{s} + \boldsymbol{n}_i \tag{1.17}$$

which can be rewritten as

$$\boldsymbol{y}_i = \boldsymbol{H}_i \boldsymbol{W}_i^{(\mathrm{tx})} \boldsymbol{s}_i + \underbrace{\sum_{\substack{j=1 \\ j \neq i}}^{N} \boldsymbol{H}_i \boldsymbol{W}_j^{(\mathrm{tx})} \boldsymbol{s}_j}_{\text{Interference}} + \boldsymbol{n}_i \tag{1.18}$$

so that it can be readily seen how other users' data appear as an interference term that degrades the received signal.

Defining the term of interference plus noise as

$$\boldsymbol{z}_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} \boldsymbol{H}_i \boldsymbol{W}_j^{(\mathrm{tx})} \boldsymbol{s}_j + \boldsymbol{n}_i \tag{1.19}$$

and using (1.18), the ergodic (mean) rate for the i-th user is given by [2], [3]

$$R_i = \mathbb{E}\left\{ \log_2 \left| \boldsymbol{I} + \boldsymbol{H}_i \boldsymbol{W}_i^{(\mathrm{tx})} \boldsymbol{R}_{\boldsymbol{s}_i} \boldsymbol{W}_i^{(\mathrm{tx}),H} \boldsymbol{H}_i^H \boldsymbol{R}_{\boldsymbol{z}_i}^{-1} \right| \right\} \tag{1.20}$$

where $\boldsymbol{R}_{\boldsymbol{z}_i} \in \mathbb{C}^{r \times r}$ is the covariance matrix of the noise plus the interference term in (1.18)

$$\boldsymbol{R}_{\boldsymbol{z}_i} = \boldsymbol{z}_i \boldsymbol{z}_i^H = \left(\sum_{\substack{j=1 \\ j \neq i}}^{N} \boldsymbol{H}_i \boldsymbol{W}_j^{(\mathrm{tx})} + \boldsymbol{n}_i \right) \left(\sum_{\substack{j=1 \\ j \neq i}}^{N} \boldsymbol{H}_i \boldsymbol{W}_j^{(\mathrm{tx})} + \boldsymbol{n}_i \right)^H \tag{1.21}$$

In the rest of the work, there are a set of assumptions that will be made, mainly to guarantee the feasibility of some of the results obtained:

- The number of users will be the same as the number of BS, *i.e.* $N = M$.
- The total number of antennas transmitting will be greater or equal than the total number of antennas at the receiver side, this is $Mt \geq Nr$.
- The number of streams transmitted to each user must be $\ell \leq r$.
- There is no correlation, neither at the transmitters nor at the receivers, so that $\boldsymbol{H}$ is full rank or, equivalently, $\operatorname{rank}(\boldsymbol{H}) = \min(Nr, Mt)$. As $Mt \geq Nr$, then $\operatorname{rank}(\boldsymbol{H}) = Nr$.

1.2 Block Diagonalization

One possibility to cancel the inter-user interference is to diagonalize the channel matrix. Perfect diagonalization is only possible if $Mt \geq Nr$ [4], and it is achieved using the following precoding matrix

$$\boldsymbol{W}^{(\mathrm{tx})} = \boldsymbol{H}^{\dagger} \tag{1.22}$$

This solution is optimum only when every user has only one antenna. In the case under study, with multiantenna receivers, complete diagonalization of the channel matrix is suboptimal since each user is able to coordinate the processing of its received signal.

In [1] is stated that the optimum solution under the constraint that all inter-user interference be zero is obtained with $\boldsymbol{H}\boldsymbol{W}^{(\mathrm{tx})}$ being block diagonal. In [1], BD is proposed as an algorithm to obtain a precoding matrix that is able to block diagonalize the channel matrix. This algorithm is described next.

In order to meet the condition of zero inter-user interference, it is necessary to cancel the interference term in (1.18), and this is equivalent to meet the following

$$\boldsymbol{H}_i \boldsymbol{W}_j^{(\mathrm{tx})} = \boldsymbol{0} \quad \forall j \neq i \tag{1.23}$$

Let $\widetilde{\boldsymbol{H}}_i$ be the channel matrix $\boldsymbol{H}$ without $\boldsymbol{H}_i$, *i.e.*

$$\widetilde{\boldsymbol{H}}_i = \begin{bmatrix} \boldsymbol{H}_1 \\ \vdots \\ \boldsymbol{H}_{i-1} \\ \boldsymbol{H}_{i+1} \\ \vdots \\ \boldsymbol{H}_N \end{bmatrix} \in \mathbb{C}^{(N-1)r \times Mt} \tag{1.24}$$

Then the condition (1.23) can be obtained making $\boldsymbol{W}_i^{(\mathrm{tx})}$ lie in the null space or kernel of $\widetilde{\boldsymbol{H}}_i$. This is possible only if the dimension of the null space is greater than zero, *i.e.* $\mathrm{rank}\left(\ker\left(\widetilde{\boldsymbol{H}}_i\right)\right) > 0$.

Now, with the dimensions of $\widetilde{\boldsymbol{H}}$, the rank of its null space is

$$\mathrm{rank}\left(\ker\left(\widetilde{\boldsymbol{H}}_i\right)\right) = Mt - \mathrm{rank}\left(\widetilde{\boldsymbol{H}}_i\right) \tag{1.25}$$

But it is assumed that $\boldsymbol{H}$ is full rank, ergo $\mathrm{rank}\left(\widetilde{\boldsymbol{H}}_i\right) = (N-1)r = \widetilde{L}_i$, and then

$$\mathrm{rank}\left(\ker\left(\widetilde{\boldsymbol{H}}_i\right)\right) = Mt - \widetilde{L}_i > 0 \tag{1.26}$$

so it is guaranteed that a precoding matrix $\boldsymbol{W}_i^{(\mathrm{tx})}$ that lies in the null space of $\widetilde{\boldsymbol{H}}_i$ exists.

The simplest way to obtain such $\boldsymbol{W}_i^{(\mathrm{tx})}$ involves using the Singular Value Decomposition (SVD) of the matrix $\widetilde{\boldsymbol{H}}_i$.

Let $\widetilde{\boldsymbol{H}}_i$ be decomposed as

$$\widetilde{\boldsymbol{H}}_i = \widetilde{\boldsymbol{U}}_i \widetilde{\boldsymbol{\Lambda}}_i \left[\widetilde{\boldsymbol{V}}_i^{(1)}, \widetilde{\boldsymbol{V}}_i^{(0)}\right]^H \tag{1.27}$$

where $\widetilde{\boldsymbol{V}}_i^{(0)} \in \mathbb{C}^{Mt\times(Mt-\widetilde{L}_i)}$ contains the last $Mt - \widetilde{L}_i$ right singular vectors of $\widetilde{\boldsymbol{H}}_i$, corresponding to the singular values equal to zero. $\widetilde{\boldsymbol{V}}_i^{(0)}$ forms an orthonormal basis of the null space of $\widetilde{\boldsymbol{H}}_i$, and thus its columns can be used to cancel the inter-user interference

$$\widetilde{\boldsymbol{H}}_i \widetilde{\boldsymbol{V}}_i^{(0)} = \boldsymbol{0} \tag{1.28}$$

Using these matrices as precoding the result is

$$\widehat{\boldsymbol{H}} = \boldsymbol{H}\left[\widetilde{\boldsymbol{V}}_1^{(0)}, \dots, \widetilde{\boldsymbol{V}}_N^{(0)}\right] = \begin{bmatrix} \boldsymbol{H}_1\widetilde{\boldsymbol{V}}_1^{(0)} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{H}_N\widetilde{\boldsymbol{V}}_N^{(0)} \end{bmatrix} \tag{1.29}$$

that, as it can be seen, has a block diagonal structure, which gives the name to the algorithm proposed in [1].

The next problem that BD solves is the maximization of the sum rate of the system, given the block diagonal structure in (1.29). The precoding matrix $\boldsymbol{W}_i^{(\mathrm{tx})}$ will be considered to be

$$\boldsymbol{W}_i^{(\mathrm{tx})} = \widetilde{\boldsymbol{V}}_i^{(0)} \boldsymbol{W}_i' \tag{1.30}$$

where $\boldsymbol{W}_i' \in \mathbb{C}^{(Mt-\widetilde{L}_i)\times\ell}$ will take care of the rate maximization.

Introducing (1.30) into (1.20) the ergodic capacity simplifies to

$$R_i^{\text{no interf}} = \mathbb{E}\left\{\log_2\left|\boldsymbol{I} + \frac{1}{\sigma_i^2}\widehat{\boldsymbol{H}}_i \boldsymbol{W}_i' \boldsymbol{R}_{\boldsymbol{s}_i} \boldsymbol{W}_i'^{,H} \widehat{\boldsymbol{H}}_i^H\right|\right\} \tag{1.31}$$

where $\widehat{\boldsymbol{H}}_i = \boldsymbol{H}_i \widetilde{\boldsymbol{V}}_i^{(0)} \in \mathbb{C}^{r\times(Mt-\widetilde{L}_i)}$.

In order to maximize the rate, consider the SVD

$$\widehat{\boldsymbol{H}}_i = \widehat{\boldsymbol{U}}_i \begin{bmatrix} \widehat{\boldsymbol{\Lambda}}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \left[\widehat{\boldsymbol{V}}_i^{(1)}, \widehat{\boldsymbol{V}}_i^{(0)}\right]^H \tag{1.32}$$

where $\widehat{\boldsymbol{\Lambda}}_i = \mathrm{diag}\left\{\hat{\lambda}_{i1}^{1/2}, \ldots, \hat{\lambda}_{ir}^{1/2}\right\} \in \mathbb{C}^{r \times r}$ contains the non zero singular values of $\widehat{\boldsymbol{H}}_i$, which has rank $\left(\widehat{\boldsymbol{H}}_i\right) = r$. And $\widehat{\boldsymbol{V}}_i^{(1)} \in \mathbb{C}^{(Mt - \widetilde{L}_i) \times r}$ contains the first r right singular vectors of $\widehat{\boldsymbol{H}}_i$, and it will be used as $\boldsymbol{W}_i'$, yielding the following precoding matrix

$$\boldsymbol{W}_i^{(\mathrm{tx})} = \widetilde{\boldsymbol{V}}_i^{(0)} \widehat{\boldsymbol{V}}_i^{(1)} \tag{1.33}$$

The BD also provides the receiver filter to be used at each user which will be

$$\boldsymbol{W}_i^{(\mathrm{rx})} = \widehat{\boldsymbol{U}}_i^{H} \tag{1.34}$$

Using all of the above, the rate that the i-th user can obtain is given by the expression

$$R_i^{\mathrm{BD}} = \mathbb{E}\left\{\sum_{k=1}^{\ell} \log_2\left(1 + \frac{\hat{\lambda}_{ik} p_{ik}}{\sigma_i^2}\right)\right\} \tag{1.35}$$

where the only parameters left to be computed are the power allocated to each of the ℓ streams of each user, and different options to do it will be discussed in Section 1.3.

1.3 Power Allocation

In Section 1.2, the BD algorithm has been described to get the precoding matrix to be used at the transmitter and the equalization filter to be used at the receiver side. After BD has been used, the power should be allocated to each of the data streams of each user, this is, the p_{ij} in (1.35) should be calculated in order to achieve a given performance, and subject to particular constraints.

The need for a power allocation algorithm comes from the restriction on the maximum power available for transmission, which may be due to physical limitations, or regulatory issues.

There are several different power constraints:

- Per Antenna Power Constraint (PAPC): The maximum power is constrained for each antenna at the transmitter. This option is specially well suited for distributed antenna systems [5], [6].
- Per Base Station Power Constraint (PBPC): In this case the maximum power is limited per base station instead of per antenna. This option is more appropriate for scenarios where all the transmitting antennas are collocated and may share a power budget, so that the transmission power can be arbitrarily allocated to each of the transmitter antennas.

A Total Power Constraint (TPC) can also be considered, which assumes that the maximum power is shared among all the transmitters in the system. Although this system is more easily analyzed, it is very unrealistic so it will not be considered in this work, except in Subsection 1.3.3 where a TPC is used to obtain an intermediate result.

For the sake of simplicity, PBPC will be used for the different analyses. In any case, PAPC can be seen as a particularization of PBPC as the derivations shown in this section can be applied to a PAPC system considering instead of each BS to have t transmit antennas, t single antenna BS.

The problem that needs to be solved is, in general, the maximization of some function of the rate of each of the users subject to a PBPC

$$\begin{aligned} \underset{\{\boldsymbol{R}_{\boldsymbol{s}_i}\}}{\text{maximize}} \quad & f\left(R_i^{\mathrm{BD}}\left(\boldsymbol{R}_{\boldsymbol{s}_1},\right),\dots,R_N^{\mathrm{BD}}\left(\boldsymbol{R}_{\boldsymbol{s}_N}\right)\right) \\ \text{subject to} \quad & P_{j,\mathrm{tx}} \leq P_{j,\max}, \quad j=1,\dots,M \end{aligned} \tag{1.36}$$

One common metric used for the maximization is the weighted sum-rate of the system, so that the function $f(\cdot)$ is equal to

$$f\left(R_i^{\mathrm{BD}},\dots,R_N^{\mathrm{BD}}\right) = \sum_{i=1}^{N} \alpha_i R_i^{\mathrm{BD}} \tag{1.37}$$

where $\alpha_i \in [0,1]$ can be seen as different priorities for different users, and they are assumed to be

$$\sum_{i=1}^{N} \alpha_i = 1 \tag{1.38}$$

and in the particular case where all the $\alpha_i = 1/N$, then the function $f(\cdot)$ represents the sum-rate of the system.

Calling

$$\overline{\boldsymbol{W}}_j^{(\mathrm{tx})} = \left[\boldsymbol{W}_{j1}^{(\mathrm{tx})},\dots,\boldsymbol{W}_{jN}^{(\mathrm{tx})}\right] \in \mathbb{C}^{t\times N\ell} \tag{1.39}$$

the precoding matrix of the j-th BS, and

$$\boldsymbol{R}_{\boldsymbol{s}} = \operatorname{blkdiag}\left(\boldsymbol{R}_{\boldsymbol{s}_1},\dots,\boldsymbol{R}_{\boldsymbol{s}_N}\right) \in \mathbb{C}^{N\ell\times N\ell} \tag{1.40}$$

the matrix containing the power assigned to all the streams of all the users, the power constraint in (1.36) can then be reformulated as

$$\mathrm{Tr}\left(\overline{\boldsymbol{W}}_j^{(\mathrm{tx})}\boldsymbol{R_s}\overline{\boldsymbol{W}}_j^{(\mathrm{tx}),H}\right) \leq P_{j,\max} \tag{1.41}$$

Now the term inside the trace operator can be written explicitly as a function of p_{ik} in order to make it easier to analyze. First define

$$\overline{\boldsymbol{W}}_j^{(\mathrm{tx})} = \left[\bar{\boldsymbol{w}}_{j,11}, \dots, \bar{\boldsymbol{w}}_{j,1\ell}, \dots, \bar{\boldsymbol{w}}_{j,N\ell}\right] \tag{1.42}$$

where $\bar{\boldsymbol{w}}_{j,ik} \in \mathbb{C}^{t\times 1}$ is the ik-th column of $\overline{\boldsymbol{W}}_j^{(\mathrm{tx})}$, *i.e.* the precoding that is used at the j-th BS for the k-th stream of the i-th user. And then:

$$\begin{aligned}\overline{\boldsymbol{W}}_j^{(\mathrm{tx})}\boldsymbol{R_s}\overline{\boldsymbol{W}}_j^{(\mathrm{tx}),H} &= \left[\bar{\boldsymbol{w}}_{j,11}, \dots, \bar{\boldsymbol{w}}_{j,N\ell}\right]\begin{bmatrix} p_{11} & & 0 \\ & \ddots & \\ 0 & & p_{N\ell}\end{bmatrix}\begin{bmatrix}\bar{\boldsymbol{w}}_{j,11}^H \\ \vdots \\ \bar{\boldsymbol{w}}_{j,N\ell}^H\end{bmatrix} = \\ &= p_{11}\bar{\boldsymbol{w}}_{j,11}\bar{\boldsymbol{w}}_{j,11}^H + \dots + p_{N\ell}\bar{\boldsymbol{w}}_{j,N\ell}\bar{\boldsymbol{w}}_{j,N\ell}^H \\ &= \sum_{i=1}^{N}\sum_{k=1}^{\ell} p_{ik}\bar{\boldsymbol{w}}_{j,ik}\bar{\boldsymbol{w}}_{j,ik}^H\end{aligned} \tag{1.43}$$

and the trace is

$$\mathrm{Tr}\left(\overline{\boldsymbol{W}}_j^{(\mathrm{tx})}\boldsymbol{R_s}\overline{\boldsymbol{W}}_j^{(\mathrm{tx}),H}\right) = \sum_{i=1}^{N}\sum_{k=1}^{\ell} p_{ik}\left\|\bar{\boldsymbol{w}}_{j,ik}\right\|_2^2 \tag{1.44}$$

where

$$\left\|\bar{\boldsymbol{w}}_{j,ik}\right\|_2^2 = \mathrm{Tr}\left(\bar{\boldsymbol{w}}_{j,ik}\bar{\boldsymbol{w}}_{j,ik}^H\right) = \bar{\boldsymbol{w}}_{j,ik}^H\bar{\boldsymbol{w}}_{j,ik}. \tag{1.45}$$

The sum-rate maximization problem can then be formulated in *standard form* [7] as

$$\begin{aligned} \underset{p_{ik}}{\text{minimize}} \quad & -\sum_{i=1}^{N}\sum_{k=1}^{\ell}\log_2\left(1+\frac{\hat{\lambda}_{ik}p_{ik}}{\sigma_i^2}\right) \\ \text{subject to} \quad & \sum_{i=1}^{N}\sum_{k=1}^{\ell} p_{ik}\left\|\bar{\boldsymbol{w}}_{j,ik}\right\|_2^2 - P_{j,\max} \leq 0, \qquad j = 1,\dots,M \\ & -p_{ik} \leq 0, \qquad \begin{array}{l} i = 1,\dots,N \\ k = 1,\dots,\ell \end{array}\end{aligned} \tag{1.46}$$

In the next sections, different alternatives for obtaining these powers are presented and described.

1.3.1 Optimal Power Allocation

In (1.46), the function $\log_2(\cdot)$ is convex on p_{ik} and the sum of convex functions is also convex, so the objective function in (1.46) is convex. The constraints are affine and therefore convex too. The optimization problem in (1.46) is a convex optimization problem that can be solved using a myriad of numerical technics [7]. Nonetheless, it would be interesting to analyze a bit further the problem in order to get some insight about it.

The problem in (1.46) satisfies *Slater's condition* [7] since the objective function is convex and all the inequality constraints are affine, hence *strong duality* holds. This means that the optimum value of the primal problem is equal to the optimum value of the *Lagrange dual problem*, so that this can be used to find out the solution to the primal, original, problem.

Under these conditions, and considering that the objective function is differentiable with respect to p_{ik}, Karush-Kuhn-Tucker (KKT) conditions [7] are necessary and sufficient for optimality of a solution, and they can be used to analyze the optimization problem in search for an optimal solution. The KKT conditions for (1.46) are

$$\begin{array}{rcll}
\sum_{i=1}^{N}\sum_{k=1}^{\ell} p_{ik}^{*}\left\|\bar{\boldsymbol{w}}_{j,ik}\right\|_2^2 - P_{j,\max} & \leq & 0, & j=1,\dots,M \\
p_{ik}^{*} & \leq & 0, & \begin{array}{l} i=1,\dots,N \\ k=1,\dots,\ell \end{array} \\
\nu_j^{*} & \geq & 0, & j=1,\dots,M \\
\mu_{ik}^{*} & \geq & 0, & \begin{array}{l} i=1,\dots,N \\ k=1,\dots,\ell \end{array} \\
\nu_j^{*}\left(\sum_{i=1}^{N}\sum_{k=1}^{\ell} p_{ik}\left\|\bar{\boldsymbol{w}}_{j,ik}\right\|_2^2 - P_{j,\max}\right) & = & 0, & j=1,\dots,M \\
-\mu_{ik}^{*}p_{ik}^{*} & = & 0, & \begin{array}{l} i=1,\dots,N \\ k=1,\dots,\ell \end{array} \\
\boldsymbol{\nabla}_{\boldsymbol{p}}\mathcal{L}\left(\boldsymbol{p}^{*},\boldsymbol{\nu}^{*},\boldsymbol{\mu}^{*}\right) & = & \mathbf{0} &
\end{array} \tag{1.47}$$

where the superscript * represents a feasible solution of the optimization problem, $\boldsymbol{\nabla}_{\boldsymbol{p}}$ is the gradient with respect to the powers $\boldsymbol{p} = [p_{11}, \dots, p_{N\ell}]^T$, $\boldsymbol{\nu} = [\nu_1, \dots, \nu_M]^T$ and $\boldsymbol{\mu} = [\mu_{11}, \dots, \mu_{N\ell}]^T$ are the Lagrange multipliers, and $\mathcal{L}$ represents the *Lagrangian* associated with the problem (1.46), and it is defined as

$$\begin{aligned}
\mathcal{L}\left(\boldsymbol{p},\boldsymbol{\nu},\boldsymbol{\mu}\right) = & -\sum_{i=1}^{N}\sum_{k=1}^{\ell}\log_2\left(1+\frac{\hat{\lambda}_{ik}p_{ik}}{\sigma_i^2}\right)+ \\
& \sum_{j=1}^{M}\nu_j\left(\sum_{i=1}^{N}\sum_{k=1}^{\ell}p_{ik}\left\|\bar{\boldsymbol{w}}_{j,ik}\right\|_2^2 - P_{j,\max}\right) - \\
& \sum_{i=1}^{N}\sum_{k=1}^{\ell}\mu_{ik}p_{ik}
\end{aligned} \tag{1.48}$$

The gradient of a function $f(\boldsymbol{x})$ with respect to $\boldsymbol{x} \in \mathbb{C}^{n\times 1}$ is defined as

$$\boldsymbol{\nabla}_{\boldsymbol{x}} f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(\boldsymbol{x}) \\ \vdots \\ \frac{\partial}{\partial x_n} f(\boldsymbol{x}) \end{bmatrix} \tag{1.49}$$

First the gradient of the objective function is calculated, by computing the partial derivatives

$$\frac{\partial}{\partial p_{ik}} \left\{ -\sum_{i=1}^{N} \sum_{k=1}^{\ell} \log_2 \left(1 + \frac{\hat{\lambda}_{ik} p_{ik}}{\sigma_i^2} \right) \right\} = \frac{-\hat{\lambda}_{ik}}{\ln(2) \left(\sigma_i^2 + \hat{\lambda}_{ik} p_{ik} \right)} \tag{1.50}$$

And the same for the inequality constraints

$$\begin{aligned} &\frac{\partial}{\partial p_{ik}} \left\{ \sum_{i=1}^{N} \sum_{k=1}^{\ell} p_{ik} \left\| \bar{\boldsymbol{w}}_{j,ik} \right\|_2^2 - P_{j,\max} \right\} = \left\| \bar{\boldsymbol{w}}_{j,ik} \right\|_2^2 \\ &\frac{\partial}{\partial p_{ik}} \{ p_{ik} \} = 1 \end{aligned} \tag{1.51}$$

So that the condition of the gradient of the Lagrangian vanishing, in (1.46) can be writen as

$$\frac{-\hat{\lambda}_{ik}}{\ln(2) \left(\sigma_i^2 + \hat{\lambda}_{ik} p_{ik}^* \right)} + \sum_{j=1}^{M} \nu_j^* \left\| \bar{\boldsymbol{w}}_{j,ik} \right\|_2^2 - \mu_{ik}^* = 0, \qquad \begin{array}{l} i = 1, \ldots, N \\ k = 1, \ldots, \ell \end{array} \tag{1.52}$$

It can be seen that μ_{ik} is a slack variable that takes into account the non-negativeness of the powers p_{ik}, and it can be ommited to get the equation

$$\frac{-\hat{\lambda}_{ik}}{\ln(2) \left(\sigma_i^2 + \hat{\lambda}_{ik} p_{ik}^* \right)} + \sum_{j=1}^{M} \nu_j^* \left\| \bar{\boldsymbol{w}}_{j,ik} \right\|_2^2 \geq 0, \qquad \begin{array}{l} i = 1, \ldots, N \\ k = 1, \ldots, \ell \end{array} \tag{1.53}$$

Calling

$$L_{ik} = \sum_{j=1}^{M} \nu_j^* \left\| \bar{\boldsymbol{w}}_{j,ik} \right\|_2^2 \tag{1.54}$$

(1.53) can be solved for p_{ik}^*

$$p_{ik}^* \leq \frac{1}{\ln(2)\, L_{ik}} - \frac{\sigma_i^2}{\hat{\lambda}_{ik}}, \qquad \begin{array}{l} i = 1, \ldots, N \\ k = 1, \ldots, \ell \end{array} \tag{1.55}$$

The result in (1.55) resembles the classical *water-filling* solution, except that now the water level is not fixed, and it depends on the precoders. The coupling existing among the power constraints of the different BS makes it impossible to find a closed-form solution for the values of p_{ik}.

Nevertheless, this analysis motivates the development of suboptimal schemes that are described in the following sections.

1.3.2 Modified Water-Filling

[8] proposes a simplification to the original problem, in order to make it more tractable. The coupling of the power constraints in (1.46) makes it impossible to get a simple solution for the optimal power allocation problem. [8] approaches the problem by first considering an equivalent virtual BS so that the problem is cast with a single power constraint.

In order to do so, instead of having a power constraint for each of the BS consider a single power constraint given by the most restrictive BS in the original problem. Define

$$\Omega_{ik}^{\mathrm{BS}} = \max_{j=1,\dots,M} \left\|\bar{\boldsymbol{w}}_{j,ik}\right\|_2^2 \tag{1.56}$$

as the weights of the single virtual BS corresponding to each of the users' streams. The optimization problem becomes then

$$\begin{aligned} \underset{p_{ik}}{\text{minimize}} \quad & -\sum_{i=1}^{N}\sum_{k=1}^{\ell}\log_2\left(1+\frac{\hat{\lambda}_{ik}p_{ik}}{\sigma_i^2}\right) \\ \text{subject to} \quad & \sum_{i=1}^{N}\sum_{k=1}^{\ell} p_{ik}\Omega_{ik}^{\mathrm{BS}} - P_{\mathrm{BS,max}} \leq 0 \\ & -p_{ik} \leq 0, \qquad\qquad \begin{array}{l} i=1,\dots,N \\ k=1,\dots,\ell \end{array} \end{aligned} \tag{1.57}$$

where $P_{\mathrm{BS,max}}$ represents the most restrictive power constraint among all of the BS.

This problem meets the same conditions as the original problem so that a similar analysis can be used. First formulate the Lagrangian of the new problem as

$$\begin{aligned} \mathcal{L}(\boldsymbol{p},\nu,\boldsymbol{\mu}) = & -\sum_{i=1}^{N}\sum_{k=1}^{\ell}\log_2\left(1+\frac{\hat{\lambda}_{ik}p_{ik}}{\sigma_i^2}\right) + \\ & \nu\left(\sum_{i=1}^{N}\sum_{k=1}^{\ell}p_{ik}\Omega_{ik}^{\mathrm{BS}} - P_{\mathrm{BS,max}}\right) - \\ & \sum_{i=1}^{N}\sum_{k=1}^{\ell}\mu_{ik}p_{ik} \end{aligned} \tag{1.58}$$

And its gradient is given by

$$\nabla_{\boldsymbol{p}}\mathcal{L}\left(\boldsymbol{p}^{*},\nu^{*},\boldsymbol{\mu}^{*}\right)=\frac{-\hat{\lambda}_{ik}}{\ln\left(2\right)\left(\sigma_{i}^{2}+\hat{\lambda}_{ik}p_{ik}^{*}\right)}+\nu^{*}\Omega_{ik}^{\mathrm{BS}}-\mu_{ik}^{*}=0,\qquad \begin{array}{l} i=1,\dots,N\\ k=1,\dots,\ell \end{array} \tag{1.59}$$

Using the KKT condition that the gradient of the Lagrangian should vanish, and considering μ_{ik}^{*} a slack variable, and solving for p_{ik}^{*}, the following inequality is obtained

$$p_{ik}^{*}\leq\frac{1}{\ln\left(2\right)\nu^{*}\Omega_{ik}^{\mathrm{BS}}}-\frac{\sigma_{i}^{2}}{\hat{\lambda}_{ik}},\qquad \begin{array}{l} i=1,\dots,N\\ k=1,\dots,\ell \end{array} \tag{1.60}$$

which, together with the constraint of the powers being non-negative, can be written as

$$p_{ik}^{*}=\left[\frac{1}{\ln\left(2\right)\nu^{*}\Omega_{ik}^{\mathrm{BS}}}-\frac{\sigma_{i}^{2}}{\hat{\lambda}_{ik}}\right]^{+},\qquad \begin{array}{l} i=1,\dots,N\\ k=1,\dots,\ell \end{array} \tag{1.61}$$

The solution of this simplified problem is given by the water-filling solution with a variable water level and, in this case, an uncoupled solution for each of the data streams for each user. This allows for the use of standard and efficient methods to find the power allocation [9].

Clearly, the definition of the new problem makes it more restrictive than the original, and its solution will be also a feasible solution for the original problem, albeit not the optimal. The results in [8] show how under some conditions, the solution achieved like this can be rather close to the optimum one.

1.3.3 Scaled Water-Filling

In [10] the same power allocation problem as in (1.46) is dealt with by considering a TPC, so that the optimization problem becomes

$$\begin{array}{lll} \underset{p_{ik}}{\text{minimize}} & -\sum_{i=1}^{N}\sum_{k=1}^{\ell}\log_{2}\left(1+\frac{\hat{\lambda}_{ik}p_{ik}}{\sigma_{i}^{2}}\right) & \\ \text{subject to} & \mathrm{Tr}\left(\boldsymbol{W}^{(\mathrm{tx})}\boldsymbol{R}_{\boldsymbol{s}}\boldsymbol{W}^{(\mathrm{tx}),H}\right)-MP_{\max}\leq 0 & \\ & -p_{ik}\leq 0, & \begin{array}{l} i=1,\dots,N\\ k=1,\dots,\ell \end{array} \end{array} \tag{1.62}$$

where it has been assumed that $P_{j,\max}=P_{\max}\forall j$.

Under this TPC, the solution is readily derived by water-filling [9], but the resulting $\boldsymbol{R}_{\boldsymbol{s}}^{\mathrm{TPC}}$ may violate the individual power contraints of each BS.

In order to meet each PBPC, the matrix $\boldsymbol{R}_{\boldsymbol{s}}^{\mathrm{TPC}}$ must be scaled so that the final power allocation is given by

$$\boldsymbol{R}_{\boldsymbol{s}}^{\mathrm{SWF}} = \beta \boldsymbol{R}_{\boldsymbol{s}}^{\mathrm{TPC}} \tag{1.63}$$

where the scaling factor $\beta \in (0,1)$ is calculated as

$$\beta = \frac{P_{\max}}{\max\limits_{j=1,\dots,M} \mathrm{Tr}\left(\boldsymbol{W}_j^{(\mathrm{tx})} \boldsymbol{R}_{\boldsymbol{s}}^{\mathrm{TPC}} \boldsymbol{W}_j^{(\mathrm{tx}),H}\right)} \tag{1.64}$$

The results in [10] show, as well, that this simplified approach can deliver near-optimum performance.

1.3.4 Uniform Power allocation

The simplest, both conceptually and computationally, alternative that can be considered to solve the power allocation in (1.46) consists on considering a uniform power allocation.

This approach assigns the same power to all the data streams of all the users. Formally this means

$$p_{ik} = p_s, \qquad \begin{array}{l} i = 1, \dots, N \\ k = 1, \dots, \ell \end{array} \tag{1.65}$$

where the power p_s should be computed taking into account the PBPC for each of the BS.

Recall from (1.41) the power transmitted by the j-th BS, where now the matrix $\boldsymbol{R}_{\boldsymbol{s}}$ is given by

$$\boldsymbol{R}_{\boldsymbol{s}} = p_s \boldsymbol{I} \tag{1.66}$$

and (1.41) becomes

$$p_s \,\mathrm{Tr}\left(\overline{\boldsymbol{W}}_j^{(\mathrm{tx})} \overline{\boldsymbol{W}}_j^{(\mathrm{tx}),H}\right) \leq P_{j,\max}, \quad j = 1, \dots, M \tag{1.67}$$

The new power allocation problem can be formulated as

$$\begin{aligned} &\underset{p}{\text{maximize}} && p \\ &\text{subject to} && p \,\mathrm{Tr}\left(\overline{\boldsymbol{W}}_j^{(\mathrm{tx})} \overline{\boldsymbol{W}}_j^{(\mathrm{tx}),H}\right) \leq P_{j,\max}, \qquad j = 1, \dots, M \end{aligned} \tag{1.68}$$

which is a *linear programming* optimization problem, and it can be solved efficiently using classical methods, *e.g.* bisection method [11].

Bibliography

[1] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, pp. 461–471, February 2004.

[2] T. M. Cover and J. A. Thomas, *Elements of information theory.* John Wiley & Sons, 2012.

[3] B. Holter, "On the capacity of the MIMO channel: A tutorial introduction," in *Proc. IEEE Norwegian Symposium on Signal Processing*, pp. 167–172, 2001.

[4] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, pp. 1691–1706, July 2003.

[5] W. Choi and J. G. Andrews, "Downlink performance and capacity of distributed antenna systems in a multicell environment," *Wireless Communications, IEEE Transactions on*, vol. 6, pp. 69–73, Jan 2007.

[6] S.-R. Lee, S.-H. Moon, J.-S. Kim, and I. Lee, "Capacity analysis of distributed antenna systems in a composite fading channel," *Wireless Communications, IEEE Transactions on*, vol. 11, pp. 1076–1086, March 2012.

[7] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge university press, 2009.

[8] A. García Armada, M. Sánchez-Fernández, and R. Corvaja, "Constrained power allocation schemes for coordinated base station transmission using block diagonalization," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, no. 1, pp. 1–14, 2011.

[9] J. M. Cioffi, *EE379C - Advanced Digital Communication - Class Notes - Chapter 4.*

[10] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath Jr., "Networked MIMO with clustered linear precoding," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 1910–1921, April 2009.

[11] R. L. Burden and J. D. Faires, *Numerical Analysis.* Cengage Learning, 2010.