A COMPARISON OF DYNAMIC COOPERATION CLUSTERING ALGORITHMS AND PILOT ASSIGNMENT METHODS ON PERFORMANCE, COMPLEXITY AND FAIRNESS OF USER CENTRIC MIMO UPLINK

ANTONIO BITTAR BIGONHA

MASTER'S THESIS IN FACULDADE DE TECNOLOGIA

DEPARTMENT OF ELECTRICAL ENGINEERING

FT

UNIVERISTY OF BRASÍLIA

UNIVERSITY OF BRASÍLIA
FT
DEPARTMENT OF ELECTRICAL ENGINEERING

UMA COMPARAÇÃO DE ALGORITMOS DINÂMICOS DE CLUSTERIZAÇÃO E MÉTODOS DE ASSINALAMENTO DE PILOTOS NO DESEMPENHO, COMPLEXIDADE E DISTRIBUIÇÃO DE TAXAS NO UPLINK DE SISTEMAS USUÁRIOS CÊNTRICOS.

ANTONIO BITTAR BIGONHA

SUPERVISOR: DANIEL ARAÚJO

MASTER'S THESIS IN FACULDADE DE TECNOLOGIA

PUBLICATION: PPGEE.TD-817/24

BRASÍLIA/DF: MAY - 2024

UNIVERSITY OF BRASÍLIA
FT
DEPARTMENT OF ELECTRICAL ENGINEERING

A COMPARISON OF DYNAMIC COOPERATION CLUSTERING ALGORITHMS AND PILOT ASSIGNMENT METHODS ON PERFORMANCE, COMPLEXITY AND FAIRNESS OF USER CENTRIC MIMO UPLINK

ANTONIO BITTAR BIGONHA

MASTER'S THESIS SUBMITTED TO DEPARTMENT OF ELECTRICAL ENGINEERING OF FT FROM UNIVERSITY OF BRASÍLIA AS PART OF THE REQUIREMENTS FOR OBTAINING THE MASTER'S DEGREE DEGREE.

APPROVED BY:

_____

Prof. Dr. Daniel Costa Araújo – FGA/University of Brasília
Supervisor

_____

Prof. Dr. Ugo Silva Dias – ENE/University of Brasília
Internal Member

_____

Prof. Dr. Igor Moáco Guerreiro – DEE/Federal University of Ceará
External Member

_____

Prof. Dr. Hugerles Sales Silva – ENE/University of Brasília
Substitute Member

BRASÍLIA, 01 DE MAY DE 2024.

CATALOG FILE

BIGONHA, ANTONIO; DANIEL COSTA ARAÚJO
A Comparison of Dynamic Cooperation Clustering Algorithms and Pilot Assignment Methods on Performance, Complexity and Fairness of User Centric MIMO Uplink [Distrito Federal] 2024.
xxiv, 180p., 210 x 297 mm (ENE/
/UnB, Master's degree, Faculdade de Tecnologia, 2024).
Master's thesis – University of Brasília, FT.
Department of Electrical Engineering
1. User Centric MIMO
2. DCC
3. MMSE Channel Estimation
4. Scalability
I. ENE//UnB
II. Título (série)

REFERENCES

BIGONHA, A., ARAÚJO, D. (2024). A Comparison of Dynamic Cooperation Clustering Algorithms and Pilot Assignment Methods on Performance, Complexity and Fairness of User Centric MIMO Uplink . Master's thesis em Faculdade de Tecnologia, Publication PPGEE.TD-817/24, Department of Electrical Engineering, University of Brasília, Brasília, DF, 180p.

SHARING RIGHTS

Antonio Bittar Bigonha

Department of Electrical Engineering (ENE) -
University of Brasília (UnB)
Campus Darcy Ribeiro
CEP 70919-970 - Brasília - DF - Brasil

This thesis is dedicated to my parents, Antônio Carlos and Márcia, and siblings, Márcio and Roberto, to my family, in special to my late grandfather Antônio da Silva Bigonha, and to my dear Natália

ACKNOWLEDGMENTS

On the emotional aspects, I would like to thank my parents ans my siblings for the ongoing support I have always gotten from them, specially on the hardest times, when I needed the most. Without them I am sure that I would not have been able to finish this thesis, and for that I thank them profusely. I would like to also mention my friends from the Electrical Engineering major, that have been a huge help whenever I needed either emotional support or theoretical knowledge. I would also like to extend my thanks to my friends in Ultra, that have helped me immensely on dealing with the emotional difficulties of writing this work, always counseling me to stay focused. A special thanks for my coaches Tiago and Matheus, who have improved my mental health in order to be able to finish this thesis. I also thank my family and my other friends very much for the continued support each one has given me, always believing in me. Finally, I would like to thank immensely for the support of my dear Natália, who has been my partner and supporter even before I started writing. Throughout the months, I could always count on her comprehension and kindness, as she has stayed by my side, on the bright moments, such as in our trip to present in the Simpósio Brasileiro de Telecomunicações (SBRT), but also on the worst moments, where the uncertainty was present.

On the academic aspect, I would like to thank immensely my supervisor Daniel Araújo for accepting me as his student, in such short time, and with great enthusiasm. He has helped me regain my confidence on researching telecommunications topics and projects that I have thoroughly studied as an undergraduate student, but were nonetheless dormant, always with much patience. I would like to thank professor João Paulo Leite, who has helped me with some telecommunications concepts, as well as simulation aspects even though he is not my supervisor. Finally, I would like to thank professor Ugo Silva Dias. who has taught the subject of Wireless Communications, thereby improving my understanding of key concepts that were crucial for this thesis.

ABSTRACT

User Centric Massive MIMO or Cell Free MIMO propose an alternative to conventional cellular MIMO systems, where the latter aims for a more uniform distribution of data rates among the user equipments. Cell free achieves that by expoiting the macro diversity of the communication media, by connecting the user equipment to various acess points scattered around the grid. Early reasearch on user centric systems have suggested that the user equipment should be served by all of its neighboring access points for near optimal performance. Recent studies, however, have shown that such configuration may be unfeasible when there are many such access points. An algorithm that assigns the user equipment to access points based on interference cancellation metrics has been shown to require less computational power, with minimum performance loss. Such method does not perform well, however, on environments where greater channel estimation quality is required. This monograph proposes alternative methods, that show themselves to be more feasible on such situations. Also, two of those methods of access point assignment do not depend on the channel estimate technique, being suitable even for environments with non-linear channel estimation, opposed to traditional methods. One key point of this work is to establish a comparison between the access points clustering methods on feasibility of conventional networks, such as the increasing number of user equipments, quality of channel estimation, and addition of nerwer acess points. Another key objective is the comparison of such methods to more conventional network configurations such as cellular networks. Another key aspect is the performance over known and unknown channel statistics, and the limitaions of such methods on imperfect channel statistics environments. The final objective, is to compare performance of such methods for different interference cancellation metrics (the pilot assignemnt algorithms).

Keywords: User Centric MIMO, DCC, MMSE Channel Estimation, Scalability.

RESUMO

Título: Uma Comparação de Algoritmos Dinâmicos de Clusterização e Métodos de Assinalamento de Pilotos no Desempenho, Complexidade e Distribuição de Taxas no Uplink de Sistemas Usuários Cêntricos.

User Centric Massive MIMO ou Cell Free MIMO propõem uma alternativa aos convencionais sistemas celulares MIMO, visto que possuem como objetivo uma distribuição mais uniforme das taxas de transmissão entre os usuários. Cell free atinge esse objetivo ao explorar a diversidade geográfica do meio de comunicação, ao conectar o usuário a vários pontos de acesso espalhados ao longo de uma grade. Pesquisas iniciais a respeito de sistemas usuàrio centrado sugeriram que o usuário deveria ser servido por todos os pontos de acesso em sua proximidade, o que resultaria em desempenho quase ótimo. Estudos recentes, no entanto, mostraram que tal configuração pode ser irrealizável na presença de um número elevado de pontos de acesso. Um algoritmo que conecta um usuário aos pontos de acesso por meio de métricas de cancelamento de interferência provou requerer menos potência computacional, com uma perda mínima de desempenho. Tal método possui desempenho insuficiente em ambientes em que uma estimativa de canal de maior qualidade se faz necessária. Este trabalho propõe métodos alternativos, que se mostram mais realizáveis em tais ambientes. Adicionalmente, dois desses algoritmos para assinalamento dos pontos de acesso não dependem da técnica utlizada para estimar o canal, de tal maneira que podem ser usados até em ambientes em que se faça uso de estimativa não linear do canal, contrário a técnicas tradicionais. Um ponto chave deste trabalho é estabelecer uma comparação entre os métodos de assinalamento de pontos de acesso no que se refere à viabilidade de seus usos em redes convencionais, por exemplo em relação ao aumento do número de usuários, qualidade de estimativa do canal e adição de novos pontos de acesso. Outro objetivo é a comparação desses métodos aos convencionais, como sistemas celulares. Outro objetivo é a comparação do desempenho de sistemas com estatísticas perfeitas e imperfeitas de canal, e, finalmente, o desempenho destes métodos para diferenres métricas de cancelamento de interferência (os algoritmos de assinalamento de pilotos).

Palavras-Chave: User Centric MIMO, DCC, Estimaitva de Canal MMSE, Escalabilidade

SUMMARY

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

Greek symbols

α	Pathloss exponent	[m^{-1}]
γ	Pathloss at reference distance	[dB]
τ_c	Total tranmsitted symbols per coherence block	
τ_p	Pilot sequence length	
τ_u	Information transmitted symbols per coherence block	
σ_φ	Azimuth angular standard deviation	
σ_θ	Elevation angular standard deviation	
$\bar{\beta}_l$	Mean channel gain of access point l	
$\beta_{\mathbf{th}}$	Normalized channel gain threshold	[dB]
β	Average channel gain	[dB]
δ_{ki}	Distance between user equipment i and user equipment j	[m]
η_i	Uplink pilot transmit power	[W]
γ_{kl}	Variance of the minimum mean square error channel estimate between user equipment k and access point l	
μ	Vector mean of a multi dimensional random variable	
$\tilde{\beta}_{kl}$	Normalized channel gain between user equipment k and access point l	
$\bar{\varphi}$	Nominal azimuth angle	
π	Ratio of circumference of circle to its diameter	
σ^2	Noise variance	[dBm]
σ_{sf}	Shadowing standard deviation	
$\bar{\theta}$	Nominal elevation angle	

Latin symbols

$\mathbf{S}$	Noise correlation matrix	[V^2]
q	Known estimation signal	[W]
$s(t)$	continuous transmitted signal	
B	Signal bandwidth	
C	Communication channel capacity	[bps/Hz]
$\mathcal{I}$	Mutual information	[bps/Hz]
K	Total number of user equipments	
L	Total number of access points	
M	Total number of antennas	
N	Total number of antennas in access point	

d	Distance between access point and the user equipment	[m]
e	Euller's number	[bps/Hz]
$n(t)$	Continuous noise	[V]
$\mathbf{n}$	Noise vector of the access point antennas for massive MIMO	[V]
p	Uplink transmit power	[mW]
$r(t)$	Continuous received signal	[V]
t	Time	[s]
v	Random interference	
$\mathbf{E}$	Total energy consumed	[J]
$\mathbf{P}$	Average power consumption	[W]

Dimensionless symbols

Dummy	A dummy counter

Subscript symbols

B_c	Channel coherence bandwidth	[Hz]
C_m	Number of complex multiplications required to compute the combining vector of user equipment k, and its channel estimates	
F_{kl}	Shadow fading coefficients between access point l and user equipment k	
$Idle$	The idle state of an energy	
N_c	Number of coherence blocks	
N_t	Number of transmissions	
T_c	Channel coherence time	[s]
$\hat{\mathbf{h}}_{kl}$	Channel estimate between access point l and user equipment k	
$\hat{s}_k$	Data estimate of the transmission from user equipment k	
$\mathbf{C}_{\mathbf{MMSE}}$	Estimation error correlation matrix	
$\mathbf{C}_{kl}$	Channel estimate error correlation matrix between user equipment k and access point l	
$\mathbf{D}_{kl}$	Selection matrix whose terms are the identity if the user equipment and access point are connected and zero otherwise	
$\mathbf{F}_k$	Collective noise vector of user equipment k	[dBm]
$\mathbf{N}_l$	Receiver noise matrix at access point l	
$\mathbf{R}_{kl}$	Spatial correlation matrix between user equipment k and the antennas of access point l	[V^2]
$\mathbf{U}$	Unitary matrix	[s]
$\mathbf{Z}_k$	Total uncertainty in channel estimate of collecitve user equipment k channel	
$\mathbf{\Psi}_{t_k l}$	Spatial correlation plot matrices	

$\mathbf{a}_k$	Large scale fading decoding coefficients for user equipment k	
$\mathbf{g}_i$	Channel between the i-th user equipment and the acces point for massive MIMO	[V]
$\mathbf{g}_{ki}$	Collective noiseless symbol estimate of user equipment k	
$\mathbf{h}_i$	Collective channel impulse reponse vector between user equipment i and all access points	[V]
$\mathbf{v}_k$	Receiver combining vector for user equipment k	
$\mathcal{D}_l$	Set of users equipment served by access point l	
$\mathcal{M}_k$	Selection of access points that are connected to user equipment k	
$\mathcal{P}_k$	Set of user equipments sharing the same pilot sequence as user equipment k	[s]
$\mathcal{S}_k$	Selection of user equipments that are connected to partially the same set of access points of user equipment k	
ϕ_{t_k}	Pilot sequence used by user equipment k	[s]
ρ_{ki}	Cross correlation coefficient between user equipment k and user equipment i	
$\tilde{\mathbf{D}}_k$	Collective binary connection matrix of user equipment k	[dBm]
a_k	Gain of the k-th path	[V]
n_l	Noise coefficient at access point l	
s_i	Transmitted data of the i-th user equipment to the access point for massive MIMO	
t_k	Delay of the k-th path	[s]

Superscript symbols

β_k^{δ}	Channel gains power threshold for user equipment k aross all access points	
$\dot{y}$	The first derivative for a function $y = f(x)$	

AP	access point.
ASD	angular standard deviation.
AWGN	addictive white gaussian noise.
CDF	cumulative distribution function.
CDMA	code division multiple access.
CF	cell free.
CPU	central processing unit.
CSI	channel state information.
DCC	dynamic cooperation clustering.
DL	downlink.
FTDMA	frequency time division multiple access.
GPA	greedy pilot assignment algorithm.
GSM	global system for mobile communication.
HSPA	high speed packet access.
iid	independent and identically distributed.
L-MMSE	local MMSE.
LOS	line of sight.
LP-MMSE	local P-MMSE.
LP-RZF	local P-RZF.
LS	least squares.
LSFD	large scale fading decoder.
LTE	long term evolution.
MIMO	multiple input multiple output.
MMSE	minimum mean squared error estimator.
MR	maximal ratio.
MSE	mean squared error.

n-LSFD	no LSFD.
NLoS	non line of sight.
NMSE	normalized mean squared error.
nopt-LSFD	near optimal LSFD.
NT	normalized threshold.
NTA	normalized threshold algorithm.
OFDM	orthogonal frequency division multiplexing.
OFDMA	orthogonal frequency division multiple access.
ONTU	orthogonal users normalized threshold.
ONTUA	orthogonal normalized threshold users algorithm.
OU	orthogonal users.
OUA	orthogonal users algorithm.
P-MMSE	partial MMSE.
P-RZF	partial regularized zero forcing.
PDF	probability density function.
PT	power threshold.
PTA	power threshold algorithm.
RP	repulsive pilot algorithm.
RPA	random pilot assignment.
RRH	remote radio head.
SE	spectral efficiency.
SINR	signal to interference to noise ratio.
SNR	signal to noise ratio.
TDD	time division duplex.
UE	user equipment.
UL	uplink.
ULA	uniform linear array.
WCDMA	wideband code division multiple access.

LIST OF OPERATIONS

$(\mathbf{A})^H$	Transpose conjugate of matrix $\mathbf{A}$
$(\mathbf{A})^T$	Transpose of matrix $\mathbf{A}$
$\bar{a}$	Mean of random variable a
$\mathrm{diag}\left(\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n\right)$	Operation that forms a resulting matrix with diagonal form, composed by $\mathbf{A}_i$ with $i \in \{1, 2, \ldots, n\}$
$\mathbb{E}\left\{a\right\}$	Expected value of complex valued number a
$\forall$	For all
$\mathbb{V}\{\mathbf{a}\}$	Variance of complex vector $\mathbf{a}$
$\mathbf{A}^{-1}$	Inverse of matrix $\mathbf{A}$
$\mathbf{a} \sim \mathcal{N}_{\mathbb{C}}(\mu, \mathbf{A})$	Random vector $\mathbf{a}$ follows a complex normal probability distribution with mean μ and covariance matrix $\mathbf{A}$
$\mathcal{A} \cup \mathcal{B}$	Union of sets $\mathcal{A}$ and $\mathcal{B}$
$\mathbf{det}(\mathbf{A})$	Determinant of matrix $\mathbf{A}$
$\mathbf{Var}(a)$	Variance of random variable a
$\mathrm{tr}\left(\mathbf{A}\right)$	Trace of matrix $\mathbf{A}$
$\arg\max_a f(a)$	Value of a that maximizes $f(a)$
$\arg\min_a f(a)$	Value of a that minimizes $f(a)$
$\max_a f(a)$	Maximum value of function $f(a)$
$\min_a f(a)$	Minimum value of function $f(a)$
$a \in \mathcal{A}_k \backslash \{b\}$	Element a belongs to set $\mathcal{A}_k$, and $a \neq b$
$a \in \mathcal{A}$	Element a belongs to set $\mathcal{A}$
$a \notin \mathcal{A}$	Element a does not belong to set $\mathcal{A}$
$a \rightarrow b$	a tends to b
$a(t) * b(t)$	Linear convolution between $a(t)$ and $b(t)$
a^*	Complex conjugate of variable a
$f(\mathbf{a})$	Probability density function of random vector $\mathbf{a}$
$f(a\|b)$	Probability density function of random variable a given random variable b
$\|\mathcal{A}_k\|$	Number of elements of set $\mathcal{A}_k$
$\|a\|$	Absolute value of complex valued number a
$\|\|\mathbf{a}\|\|$	Norm of complex valued vector $\mathbf{a}$

1 INTRODUCTION

This chapter aims to introduce the concepts that will be used on this monograph. Section 1.1 has focus on showing the motivation behind studying cell free systems, while section 1.2 establishes a comparison between traditional cellular systems and cell free with respects to limitations and benefits, over topics as signal to noise ratio and signal to noise to interference ratio variations for implementation in 5G systems. Section 1.3 elaborates on the need of finding an access point (AP) clustering algorithm that provides appropriate performance while mantaining system complexity on manageable terms.

1.1 INTRODUCTION AND MOTIVATION

The main objective of any network is to provide users with reliable access to data among various geographical locations. Initially, the focus was on telephone and voice calls, however nowadays with the advent of internet connection, the applications vary ranging from simple text messages, to the uploads of files and videos that might require a large amount of data and processing in order to be reliably transmitted. On primitive communication networks, analog signals were transmitted and decoded, as focus was primarily on telephone communication, which did not require a high quality to be performed. Analog communication is still used nowadays, but digital communication is much more used, since it can lead better with the effects of thermal noise, as the focus is decoding data packets, opposed to the entire analog waveform, as explained in Chapter 7 of [1], that better exemplifies this transition.

When dealing with wireless systems, the effects of the communication channel suffer much more variation than wired systems. For instance, the mobility between the user equipment (UE) and the transceiver responsible for decoding information, the AP is a factor that makes data transmission challenging, since the communication channel changes rapidly in time, making it very hard to estimate. A more detailed explanation can be found in the works of [2]. Over the years various mobile technologies have been proposed: the authors in [3] have studied code division multiple access (CDMA) techniques, that are basically the transmission of information over the same bandwidth by use of spread spectrum techniques. High speed packet access (HSPA) systems have been used extensively on the beginning of this century, along with wideband code division multiple access (WCDMA), and were able to provide download rates of up to 168 Mbps and uploads of up to 22 Mbps. Before that, frequency time division multiple access (FTDMA) techniques were used to transmit information either at the same time over different frequency ranges or at the same frequency over different time slots. This has been used by 2G systems global system for mobile communi-

cation (GSM) that were able to provide a download rate up to 97 kbps. Such systems were not able, however, to connect many UE in the same grid, since available bandwidth was very limited. After HSPA came the long term evolution (LTE), which composed the 4G of wireless systems, the first of which employed multiple input multiple output (MIMO) systems, such as seen in [4], in an effort to explore the spatial diversity of the channels, so that information could be transmitted at the same time over the same bandwidth over different spatial paths. It was this technology that also used the notions of orthogonal frequency division multiple access (OFDMA), which consisted on the modulation of subcarriers in order to turn a frequency selective channel into one frequency flat channel, in order to diminish distortion. A more complete description of cellular systems up to orthogonal frequency division multiplexing (OFDM) can be found in [5].

At this time, the demand for ever higher transmission rates started to grow, and thus alternative technologies to guarantee higher rates with small delays were suggested. One such method was the Massive MIMO that was explored in [6], that consists on employing a very large number of antennas at each AP, so that there are a very large number of multipath components of the transmitted signals that, when combined, provide a robust detection. This, in theory, could lead to an infinite value for the channel capacity, as long as the number of antennas at each AP kept increasing, and some favorable propagation conditions hold, as shown in [7]. On practical networks, however, this was proved to act differently, as spatial correlation between the AP antennas prevented the statistical independence of the multipath components from happening, which was a crucial property for such systems. Thereby, authors such as [8] have proposed an alternative type of MIMO in the form of cell free (CF).

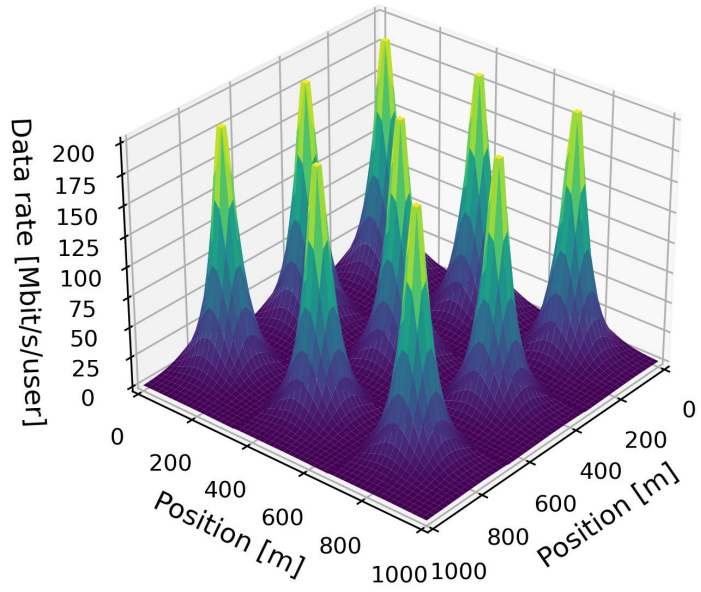
One such thing that exists in conventional networks is the concept of *cell*. A cell is a geographical area (which could take any form, from hexagonal, to square, to circular) where the total signal to noise ratio (SNR) of a transmitted signal is greater than a certain threshold. Conventional networks operate in a way that each UE connects to an AP based on a cell, that provides the largest SNR to it. The cellular network is divided into a *core* and an *edge*. The edge network is the region where the hardware units and the APs are involved with the physical layer communication with the UEs. The core network is responsible for facilitating the services requested by the UE, such as routing of data packages and connection to the internet. The connection between the edge and the core is defined a the *backhaul* links, that can be wired, by use of for example optical fibers, or partially wired using fixed microwave links.

A fact about the electromagnetic signal strength, is that it decays exponentially with the distance of the transmitter to the receiver. The minimum decaying value is quadratic, meaning that it still decays rapidly. This means the closer the UE is to the AP, the better the SNR values it will experience. UEs that are on the cell edge, which is defined as the frontier between cells, will experience low values of SNR, and additionally may experience higher

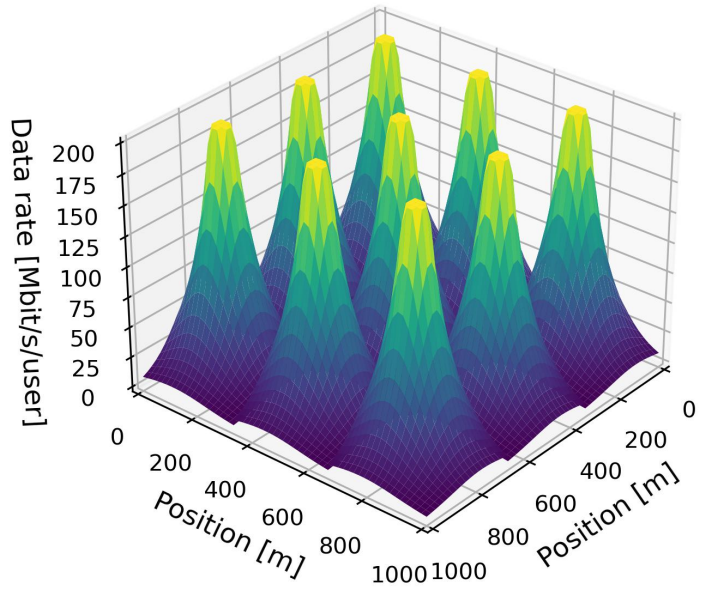
interference from other cells. In fact, it is very common that a cell-edge UE has an SNR gap of up to 40 dB in relation to an UE in a privileged location. With the data rate being an increasing function of the signal to interference to noise ratio (SINR), we see that data rates variations are very common in cellular networks. Fig. 1.1 shows the data rate of the downlink (DL) operation achieved by the UEs at different locations when each AP deploys fixed gain antennas and transmits with the maximum power. We can see that the maximum achievable rate, taking in consideration a system bandwidth of 20 MHz, is 200 Mbps, which happens when the UE is close to one of the AP. On the other hand, UEs who are far from the APs achieve a rather low rate (many of them actually achieve rates below 0.5 Mbps). The lower rates are clearly not enough for the transmission of videos and streaming. They are, nonetheless, much suitable for voice calls, which was the primary reason that cellular systems were implemented. We see that by deploying a large number of antennas at each AP, the rates have a more uniform distribution when compared to the first case, however there are still very large variations, specially in the cell edges. In summary, the peak rates are very substantial for the UEs that are nearby the AP, but they get progressively smaller for the far away UEs. For the actual demand of high rates and specially, connectivity everywhere, we must find a way to obtain rates that are more uniformly distributed. For instance, one of the aspects of 5G is the possibility of connectivity at every location, with minimum information delay and minimum outage (by outage we mean the amount of time the network will not be able to provide coverage). Conventional cellular networks have inherently rates that are distance variant, so we must break free of the cell definition by studying an architecture that does not rely on it: the CF architecture.

1.1.1 Cell Free Systems

First, we shall describe the architecture of CF systems. The channel models and mathematical properties shall be discussed later on in Chapter 2. A cell free network is composed of L APs that jointly serve the UEs that compose the grid. Each one of those APs is connected by the *fronthaul* to a central processing unit (CPU), that in turn is responsible for AP cooperation. Multiple CPUs connected by multiple fronthaul links can exist. Fig. 1.2 represents the architecture of a CF network composed of single antenna APs. Just like cellular networks, CF can also be decomposed into a core and an edge. The APs and the CPUs are all part of the edge and are connected by the fronyhaul links, while the connections between the edge and the core are called backhaul links. Thus, the backhaul links can be used to send/receive data from the internet, and other sources, to facilitate data services. The fronthaul links can be used to share physical layer signals that will be transmitted in the DL, forwarding received uplink (UL) signals that will be decoded, and sharing channel state information (CSI). The fronthaul can also facilitate phase-synchronization between the APs by providing a phase reference. In this work, we shall also assume that they have infinite capacity, are



(a) Each AP is equipped with a single antenna with a 12 dBi gain



(b) Each AP is equipped with 128 omni-directional antennas

Figure 1.1 – Example of the downlink data rate achieved by a UE at different locations in the cellular network, assuming each AP transmits with full power. The cell-edge SNR is 0 dB in 1.1a and the power is assumed to decay as the distance to the power of four. The bandwidth is 20 MHz, and the maximum spectral efficiency (SE) is 10 bit/s/Hz. We can see that the rates vary largely along the grid.

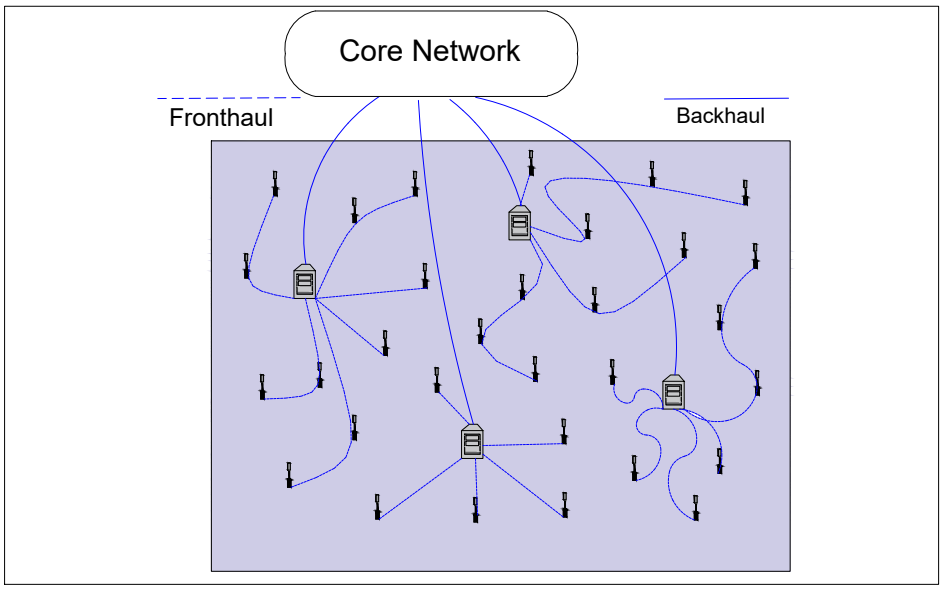
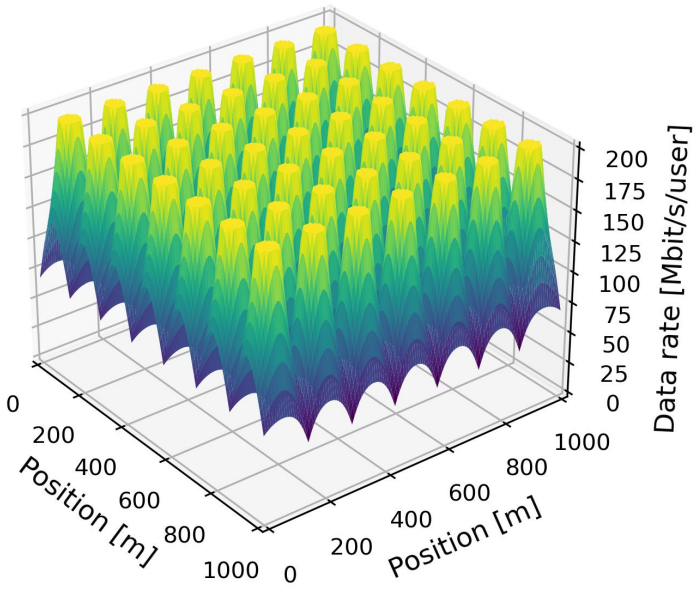


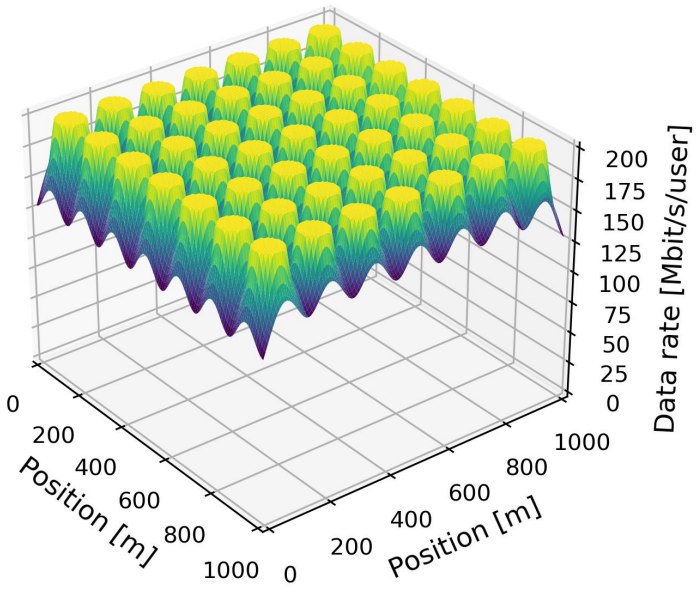
Figure 1.2 – Cell free architecture where many APs are geographically distributed and connected to the CPUs via fronthaul links. The CPUs are all connected to the core network via backhaul links. The APs are jointly serving all of the UEs in the grid.

error free and possess negligible latency, which are common limitations of real life fronthaul links.

In CF networks, no cell boundaries exist for connecting UEs, meaning that all of the APs affecting the UE will be part of the communication protocol. The method to select the APs that will take part in decoding the information data can be of various forms: for example, all of the APs in the area could serve the UE, or only the APs with a SNR above a certain threshold, or other methods of AP assignment that shall be discussed later on Chapter 4. Therefore, not all of the APs in the coverage area must be serving the UE, in fact there could be the possibility of an AP not serving any UE at all. Cellular architecture and CF are very similar in regards to signal processing, as we shall see later in Chapter 2. For that reason, the UE should be able to change from cellular to CF or vice versa without any upgrade to its software. For the simulation setup of Fig. 1.3 we have deployed a total of $L = 49$ APs along a square grid, serving a total of $K = 400$ UEs. Each AP was equipped with $N = 16$ antennas with unitary gain. We can see that the cellular case has data rates varying from 47 to 200 Mbps, again taking in consideration a bandwidth of 20 MHz, while the cell free case has rates that range from 128 to 200 Mbps. This proves once again that increasing the number of antennas at the AP does indeed help, however this act alone is not enough to provide high rates for the UEs that are at the cell edges. Instead, if we look at the cell-free system, we see that even the UEs that are far away from the APs still have fairly high rates, due to cooperation between the APs to decode the intended signals. One important aspect that was used to generate the CF system was that every AP serves every UE. Also, another key point was assuming perfect CSI, thus those rate values represent an upper bound performance and should be only seen as comparison measures between the cellular system and the CF.



(a) Each AP is equipped with a single antenna with a 12 dBi gain



(b) Each AP is equipped with 128 omni-directional antennas

Figure 1.3 – Example of the downlink data rate achieved by a UE at different locations in the cellular network, assuming each AP transmits with full power. The cell-edge SNR is 0 dB in 1.3a and the power is assumed to decay as the distance to the power of four. The bandwidth is 20 MHz, and the maximum SE is 10 bit/s/Hz. We can see that the rates vary largely along the grid.

1.2 BENEFITS AND LIMITATION OF CELL FREE AND CELLULAR MASSIVE MIMO IN 5G

The proposed method of interest in 5G systems is the Cellular Massive MIMO. This concept, that dates back to 2010, says that each AP is equipped with a large number of antennas, usually possessing low gains, as opposed to conventional cellular networks, that employ antennas of high gains and highly directive. The key concept that makes Massive MIMO an option is *spatial diversity*, that is, the property of the signal to propagate along spatially diverse paths. Another key factor is the *spatial multiple access*, that consists on the ability of multiple UEs to be served by the same AP at the same time-frequency interval. This way multiple UEs can be served by the same AP simultaneously in the same cell with negligible interference.

1.2.1 Benefits of Cellular Massive MIMO

The most important aspect of the network for Massive MIMO to be used efficiently is each AP being equipped with a number of antennas such that it far surpasses the number of served UEs. When this is guaranteed, two such phenomena are observed: *favorable propagation* and *channel hardening*. The first refers to the fact that the channels of spatially diverse UEs are nearly orthogonal due to the high directivity of the transmitted and received signal beams, while the latter refers to the possibility of neutralizing the small scale fading variation by combining the replicas of the same signal transmitted and received over many antennas. Initially, it was believed that simple linear receivers such as maximal ratio (MR), that do not possess any interference cancellation whatsoever, were implementable in such systems, as seen in [9]. However, it has been later shown that linear methods that actively suppress interference are more adequate, such as minimum mean squared error estimator (MMSE), as seen in [10] and [11], where the effects of pilot contamination and spatial correlation were discussed as fundamental limitations to the use of simple linear receivers. Nonetheless, the use of linear methods to suppress interference seem to lead to desirable performance outcomes, which makes the use of non-linear and more complex signal detection techniques optional.

1.2.2 Limitations of Cellular Massive MIMO

As seen in Fig. 1.1 and 1.3, the data rates on Massive MIMO technology are greatly increased compared to traditional systems, however they do still vary a considerable amount. Besides that, the antennas at the APs, in commercial real-life systems, are not deployed in an uniform linear array (ULA), as conventionally they are in literature, but rather on planar arrays. Planar arrays exhibit low resolution regarding the azimuth domain, which is not the

case of ULAs, therefore the service quality of such systems is diminished. Therefore, a new architecture that aims to provide more rate uniformity, and also, that allows for the use of ULAs is proposed in the form of CF MIMO.

1.2.3 Benefits of Cell Free MIMO

Among the benefits of CF systems, we can enumerate two that summarize them as viable solutions for network architecture: the ability to manage interference and the presence of small SNR variations.

1.2.3.1 Benefit 1: SNR with smaller variations

Lets assume first a Massive MIMO setup, where there is only one cell and one AP with multiple antennas. We initially do not consider the small scale fading variations, and focus solely on the large scale fading. Moreover, we shall notice that the SNR expression between the l-th AP and the k-th UE, where $\beta(d)$ is the pathloss, p is the transmitted power and σ^2 is the noise power, is given by

$$\text{SNR}_{kl} = \frac{p}{\sigma^2}\beta. \tag{1.1}$$

Since pathloss is dependent on the distance between the UE and the AP, it can be inferred that the pathloss from the UE to all of the antennas of the AP is the same. Therefore, the SNR of Massive MIMO systems can be summarized as

$$\text{SNR}^{\text{MIMO}} = \frac{p}{\sigma^2}M\beta(d), \tag{1.2}$$

where$M$ is the number of AP antennas.

Lets now assume traditional cellular, where the UE can connect between a range of L single antenna APs, and will choose the strongest between them. Therefore, the SNR of this case, which will be called the *small cell*, is given by

$$\text{SNR}^{\text{small-cell}} = \frac{p}{\sigma^2}\underset{l\in\{1,\dots,L\}}{\arg\max}\beta(d_l). \tag{1.3}$$

Finally, when comparing the CF setup, which has the same AP distribution as the small cell, we consider that each one of the APs serve the UE, and therefore, a superposition of the signals of each AP will compose the collective SNR in the form

$$\text{SNR}^{\text{cell-free}} = \frac{p}{\sigma^2}\sum_{l=1}^{L}\beta(d_l). \tag{1.4}$$

We shall compare numerically those setups in Chapter 3, but for now we can observe that

the Massive MIMO SNR is highly dependent on this particular UE location, and can vary greatly with its position. For instance UEs that are far away from the AP are very likely to have low SNR due to high pathloss. When comparing the case of small cell, we see that the chosen AP is the one that has the lowest pathloss to that particular UE, which surely makes fluctuations smaller than Massive MIMO. On the case of CF MIMO. however, the received SNR is a superposition of the ones for each AP to that particular UE, therefore it has always larger values than the single cell, as the latter is only a special case of the former. It also possesses much lower SNR variations, due to the fact that it connected to all APs, so the likelihood of an stronger AP serving the UE increases.

1.2.3.2 Benefit 2: Ability to Manage Interference

We will see that CF also has the ability to suppress interference that surpasses that of conventional cellular networks. Lets consider the Massive MIMO architecture, with the channel $\mathbf{g}_k \in \mathbb{C}^M$ between the UE k and the AP. The signal received at the AP is given by

$$\mathbf{y}^{\text{MIMO}} = \sum_{i=1}^{K} \mathbf{g}_i s_i + \mathbf{n}, \tag{1.5}$$

with s_i being the information data transmitted by UE i, such that $\mathbb{E}\{\{|s_i|^2\}\} = p$ and $\mathbf{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_M, \sigma^2\mathbf{I}_M)$, the receiver noise. In order to decode the information, the AP uses the receiver combining vector $\mathbf{v}_k \in \mathbb{C}^M$ to the received UL signal, in order to obtain the data estimate

$$\hat{s}_k^{\text{MIMO}} = \mathbf{v}_k^H \mathbf{y}^{\text{MIMO}} = \sum_{i=1}^{K} \mathbf{v}_k^H \mathbf{g}_i s_i + \mathbf{v}_k^H \mathbf{n} \tag{1.6}$$

of the UE k. Therefore, the SINR is given by

$$\begin{aligned} \text{SINR}_k^{\text{MIMO}} = \frac{\mathbb{E}\left\{\left|\mathbf{v}_k^H \mathbf{g}_k s_k\right|\right\}^2}{\mathbb{E}\left\{\left|\sum\limits_{\substack{i=1 \\ i\neq k}}^{K} \mathbf{v}_k^H \mathbf{g}_i s_i + \mathbf{v}_k^H \mathbf{n}\right|^2\right\}} &= \frac{\left|\mathbf{v}_k^H \mathbf{g}_k\right|^2 p}{\mathbf{v}_k^H \left(p \sum\limits_{\substack{i=1 \\ i\neq k}}^{K} \mathbf{g}_i \mathbf{g}_i^H + \sigma^2 \mathbf{I}_M\right)^{-1} \mathbf{v}_k} \\ &\leq p \mathbf{g}_k^H \left(p \sum_{\substack{i=1 \\ i\neq k}}^{K} \mathbf{g}_i \mathbf{g}_i^H + \sigma^2 \mathbf{I}_M\right)^{-1} \mathbf{g}_k, \end{aligned} \tag{1.7}$$

where the combiner that was used to obtain the upper bound for the SINR is given by

$$\mathbf{v}_k = \left(p \sum_{\substack{i=1 \\ i\neq k}}^{K} \mathbf{g}_i \mathbf{g}_i^H + \sigma^2 \mathbf{I}_M\right)^{-1} \mathbf{g}_k, \tag{1.8}$$

and can be found by applying the principle of the Rayleigh quotient maximization described in Corollary 5.1.

When dealing with small cells, the only connected AP will be the AP l, that provides the best SINR to the UE, therefore the received UL signal in AP l is:

$$y_l^{\text{small-cell}} = \sum_{i=1}^{K} h_{il} s_i + n_l \tag{1.9}$$

where $n_l \sim \mathcal{N}(0, \sigma^2)$ denotes the noise at the receiver, and $h_{kl} \in \mathbb{C}$ is the channel response between the AP l and UE k. The SINR expression is

$$\text{SINR}_{kl}^{\text{small-cell}} = \frac{\mathbb{E}\left\{|h_{kl} s_k|^2\right\}}{\mathbb{E}\left\{\left|\sum\limits_{\substack{i=1 \\ i\neq k}}^{K} h_{il} s_i + n_l\right|^2\right\}} = \frac{p|h_{kl}|^2}{p\sum\limits_{\substack{i=1 \\ i\neq k}}^{K} |h_{il}|^2 + \sigma^2}. \tag{1.10}$$

Therefore, the AP that has the maximum SINR with respect to UE k is the AP that UE is going to connect to, and the SINR is

$$\text{SINR}_{k}^{\text{small-cell}} = \max_{l \in \{1,\ldots,L\}} \text{SINR}_{kl}^{\text{small-cell}}. \tag{1.11}$$

Notice that the small cell setup chooses the AP that provides the best SINR value, not necessarily the best SNR, since the AP with the best SNR value could potentially introduce a high degree of interference, thereby leading to a smaller performance when compared to the best SINR case.

Finally, we shall consider the CF setup. For that, we suppose the same L APs that were deployed in the small cell setup are used here. Also, we consider that each AP is single antenna. The signal received by the APs in the UL are then given by

$$\mathbf{y}^{\text{cell-free}} = \sum_{i=1}^{K} \mathbf{h}_i s_I + \mathbf{n}, \tag{1.12}$$

where $\mathbf{h}_i = [h_{i1}, \ldots, h_{iL}]^T$ and $\mathbf{n} = [n_1, \ldots, n_L]^T$. In a similar way to the one of Massive MIMO, a combiner vector is used to detect the signal from UE k. Therefore, the estimate of s_k is

$$\hat{s}_k^{\text{cell-free}} = \mathbf{v}_k^H \mathbf{y}^{\text{cell-free}} = \sum_{i=1}^{K} \mathbf{v}_k^H \mathbf{h}_i s_i + \mathbf{v}_k^H \mathbf{n}, \tag{1.13}$$

and the SINR expression is given by

$$\text{SINR}_k^{\text{cell-free}} = \frac{\mathbb{E}\left\{\left|\mathbf{v}_k^H \mathbf{h}_k s_k\right|\right\}^2}{\mathbb{E}\left\{\left|\sum\limits_{\substack{i=1\\ i\neq k}}^{K} \mathbf{v}_k^H \mathbf{h}_i s_i + \mathbf{v}_k^H \mathbf{n}\right|^2\right\}} = \frac{\left|\mathbf{v}_k^H \mathbf{h}_k\right|^2 p}{\mathbf{v}_k^H \left(p \sum\limits_{\substack{i=1\\ i\neq k}}^{K} \mathbf{h}_i \mathbf{h}_i^H + \sigma^2 \mathbf{I}_M\right)^{-1} \mathbf{v}_k}$$

$$\leq p \mathbf{h}_k^H \left(p \sum_{\substack{i=1\\ i\neq k}}^{K} \mathbf{h}_i \mathbf{h}_i^H + \sigma^2 \mathbf{I}_M\right)^{-1} \mathbf{h}_k. \quad (1.14)$$

We can easily see that the SINR expression in Eq, (1.11) is nothing more than a particular case of Eq. (1.14). For evaluating the expressions of the CF and Massive MIMO cases, however, we need to resort to simulations, since the expressions are very similar. For the simulation setup, we shall consider $K = 10$ UEs randomly placed along a square grid of 500 m $\times$ 500 m, where the channel between UE and AP is generated by $\sqrt{\beta(d)}e^{j\phi}$, with the complex exponential denoting the random phase due to multipath propagation in the complex envelope domain, β is evaluated by

$$\beta_{kl}(d_{kl})[\text{dB}] = -30.5 - 36.7 \log_{10}\left(\frac{d_{kl}}{1\text{ m}}\right), \quad (1.15)$$

and the phase ϕ is an uniform random variable ranging from 0 to 2π. In the Massive MIMO setup, we have considered the AP centered at the middle of the square grid, whereas the setup of small-cell and CF have the APs evenly distributed along the square grid. There are once again $L = 49$ APs, and we assume the transmit power to be $p = 20$ mW, and the noise power is $\sigma^2 = -96$ dBm. Also, we assume an height $h = 10$m between the UEs and the APs. The resulting SINR cumulative distribution function (CDF) can be seen in Fig. 1.4. We can immediately notice that the benefits of geographically distributed antennas are much larger than the ones obtained from deploying a large array in one location: for instance there is a variation of almost 40 dB between the lowest and the highest SINR in the massive MIMO case, whereas there is a 30 dB gap between them in the CF case. This indicates that on those conditions, the SINR of CF outperforms that of massive MIMO by at least 10 times. We can also see that the values for the small-cell have an even larger variance than the other two methods, being even more outperformed by the CF architecture.

1.2.4 Limitation of Cell Free MIMO

The benefits of CF makes it a very promising solution for the issue of uniformity in throughput. However, there are some limitations that should be taken in account before one is to implement this technology: the most immediate of them is the synchronization needed between the APs to jointly decode the data, which can be seen in works such as [12], that

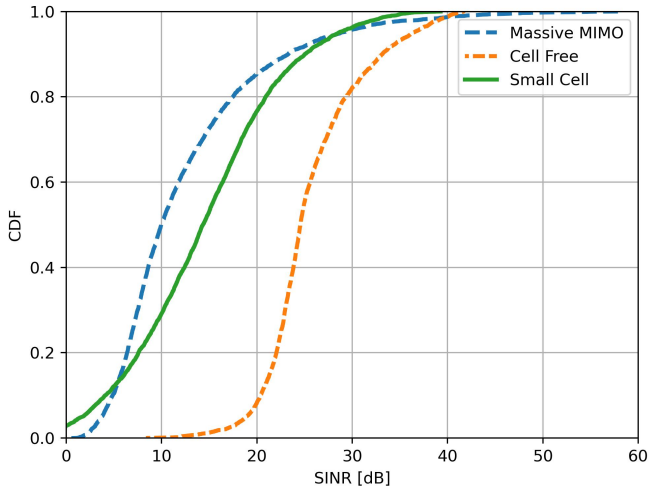


Figure 1.4 – CDF for the SINR of small cell, massive MIMO and CF architectures. The single-antennas APs, for CF and small-cells, are deployed on a square grid of 500m $\times$ 500m and serve $K = 10$ randomly deployed UEs. In the case of massive MIMO the AP is equipped with L antennas, and is located at the center of the grid.

focus on synchronizing the clocks of the APs to jointly decode data. Another key point is the ever growing system complexity, since more APs being connected to the same UE mean more resources needed to compute the combining vector for that particular UE. We will see in Chapter 2 that this is a fundamental issue, preventing the implementation of the traditional CF (when all APs serve all UEs), except on very reduced networks.

1.3 THE IMPORTANCE OF SELECTING THE DYNAMIC COOPERATION CLUSTERING METHOD

We shall see in Chapter 2 that the ideal CF setup, with all of the UEs being served by all APs in the grid, is not viable in conventional networks, since system complexity grows large, and an ever growing amount of resources is needed for signal detection. Addressing that, a discussion about the metric to measure the feasibility of systems has been established in [13], and the authors of [8] have proposed a method such that only a subset of APs is connected to each UE. Since this AP assignment changes as channel conditions change, they called it dynamic cooperation clustering (DCC). They have shown that, by choosing the DCC method appropriately, the throughput loss from the ideal case is negligible, for much more manageable complexity.

Although this DCC method addresses the problem of the ideal CF architecture, it also possesses some limitations: one of them is the growing complexity with channel estimation quality. We will see in Chapter 3 that multiple antennas architectures, and by extension, the

CF architecture, need some measure of knowledge of the communication channel in order to decode the transmitted data. The chosen channel estimation method is linear, in that known signals are transmitted from the UEs to the AP in the UL, and the response is processed into an estimate of the channel. The DCC method is dependent on the length of the transmitted training data (which we should call *pilot*). Therefore, as this length increases, so does the number of APs that are connected to a given UE, thereby increasing system complexity. Although this change in complexity is perfectly manageable in the simulation setups of this monograph, it clearly presents itself as a challenge in conventional networks, with thousands of APs, hundreds of UE and hundreds of pilots.

1.4 CONTRIBUTIONS AND WORK ORGANIZATION

The objective of this work is to introduce a DCC method that benefits from acceptable performance as the one of [8], but also possesses a much lower complexity, that makes signal detection less computationally expensive. To illustrate the motivation to such methods, we shall repeat the simulation that generated Fig. 1.3b, but with $L = 100$ APs, where each one is equipped with $N = 4$ antennas. Additionally, the decaying path loss exponent is 3.67, which sits well with urban environments, according to [14]. The communication channel is known by the receiver and there is no shadowing effect. Also, the APs are deployed following a wrap around topology.

1.4.1 Unscalable and Orthogonal Users Cell Free Architectures

We shall compare two cell free architectures: the first one will be called *unscalable*, because it does not meet the criteria that shall be presented in Chapter 2 in any way. This architecture consists on every AP serving every UE, and thus, the APs can all jointly cooperate to generate a estimate of the transmitted data. The second cell free architecture is called orthogonal users (OU). It consists on allowing the AP to serve only the UEs that have been assigned orthogonal pilot sequences. Since channel estimation is considered perfect on this simulation setup, we shall allow each AP to serve the 100 strongest UEs, in order to better approximate the results to those that will be shown in Chapter 6. The rates in Mbps are shown in Fig. 1.5 and Fig. 1.6

Clearly there is a larger variation of the rates in OU than on the unscalable configuration. Such variation is, however, very small: In the OU architecture the minimum rate is 151 Mbps while the unscalable configuration has a minimum rate of 158 Mbps. We shall see in Chapter 6 that the reduction in the minimum rate value is greatly compensated by the reduction in system complexity.

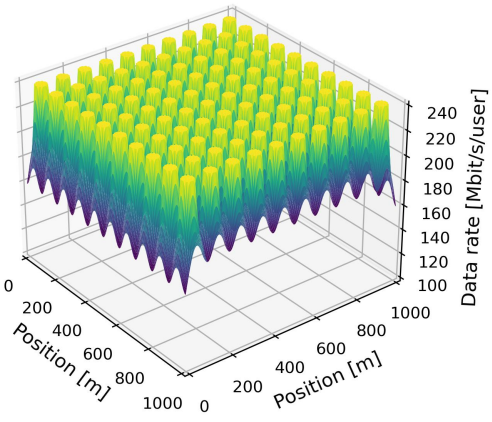


Figure 1.5 – UE throughput in Mbps for the unscalable cell free architecture of an $K = 400$ UEs over a square grid of 1 km^2. The number of APs is $L = 100$ and each one is equipped with $N = 4$ antennas. The system bandwidth is 20 MHz.

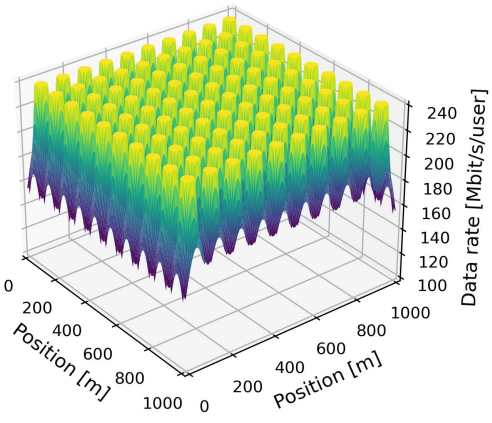


Figure 1.6 – UE throughput in Mbps for the OU cell free architecture of an $K = 400$ UEs over a square grid of 1 km^2. The number of APs is $L = 100$ and each one is equipped with $N = 4$ antennas. The system bandwidth is 20 MHz.

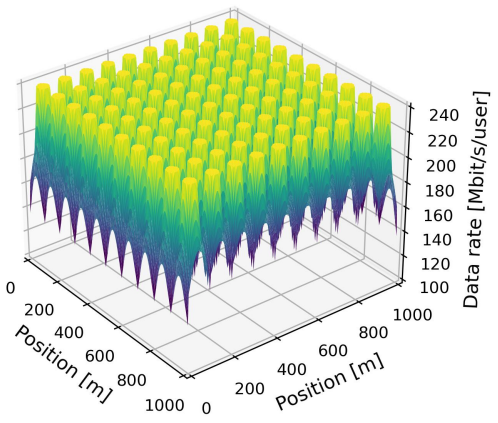


Figure 1.7 – UE throughput in Mbps for the NT cell free architecture of an $K = 400$ UEs over a square grid of 1 km^2. The number of APs is $L = 100$ and each one is equipped with $N = 4$ antennas. The system bandwidth is 20 MHz.

1.4.2 Normalized Threshold and Orthogonal Users Normalized Threshold Architectures

Now let us compare the architectures of normalized threshold (NT) and orthogonal users normalized threshold (ONTU). Both are contributions of this author. The architecture of NT is more detailed in [15], while ONTU is proposed in this monograph in Chapter 4. Both of them rely on a parameter that should be called the *normalized threshold*, γ_{th}. This value could depend on various metrics, such as the channel gains, the correlation between the channels, the shadowing coefficients, etc. We shall define this parameter based on the channel gains, such that

$$\tilde{\beta}_{kl} = \frac{\beta_{kl} - \bar{\beta}_l}{\sqrt{\text{Var}(\beta_l)}} \tag{1.16}$$

is the normalized channel gain, since it consists on the subtraction of the actual channel gain and the mean channel gain vector wit respect to AP l divided by the standard deviation of the same vector, hence the normalized name. The normalized threshold value may be obtained from the percentile of the normalized channel gains. Having defined $\tilde{\beta}_{kl}$, we might discuss the differences of the NT and ONTU architectures. The first one applies the threshold condition to every AP-UE pair while the latter consists on selecting the APs by OU and then choosing the UEs above the threshold condition. It is to be expected that ONTU yields more variation in the rates than OU, since less UEs are served by each AP. Fig. 1.7 and Fig. 1.8 show the rates for those architectures

The threshold value has been chosen as $\gamma_{\text{th}} = \{0.906, 0.127\}$ for the NT and ONTU respectively. Those values correspond to the 80-th and 60-th percentiles of the normalized

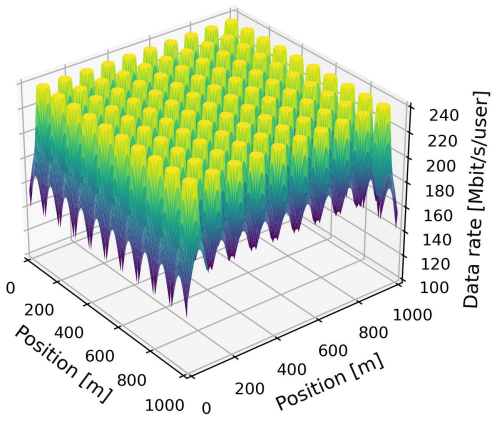


Figure 1.8 – UE throughput in Mbps for the ONTU cell free architecture of an $K = 400$ UEs over a square grid of 1 km^2. The number of APs is $L = 100$ and each one is equipped with $N = 4$ antennas. The system bandwidth is 20 MHz.

channel gains, which means that on average 20% of the AP-UE pairs will be connected for the NT while on average 40% of the UEs connected to a given AP by OU will be served by that AP on the ONTU architecture. Fig 1.7 shows that the lowest rate in the network is of 135 Mbps, which is shorter by 23 Mbps than the uscalable architecture. Still, it provides satisfactory performance. The minimum rate of ONTU is 151 Mbps, which is virtually the same as OU. We will see in Chapter 4 and Chapter 6 that the complexity for ONTU is reduced when compared to OU, which justifies its use.

Additionally, one should notice that pilot contamination is not a factor when dealing with known communication channels, which is the assumption of this section simulations. Thus, the rates of NT under unknown channel assumptions are actually different from the ones shown in this section, since one AP might serve UEs that share the same pilot sequence, which in turn results in pilot contamination, that is discussed in Chapter 3. The architectures of OU and ONTU assume the AP may only serve the UEs that have been allocated orthogonal pilot sequences, thereby the rate values under perfect channel knowledge should be similar to the ones of imperfect channel knowledge.

1.4.3 Power Threshold Architecture

Finally, two alternative architectures are discussed in this section. One such is the power threshold (PT), which can be found in [16], while the other is the Orthogonal PT, which is again a contribution of this monograph. Both methods allow the UEs to be served only by the APs that contribute to a given power threshold percent of the total power of the UEs, that shall be written as β_k^{δ}, and the total power is given by $\sum_{l=1}^{L} \beta_{kl}$. The main difference

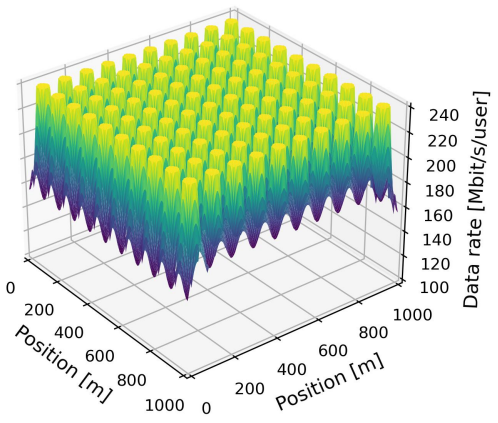


Figure 1.9 – UE throughput in Mbps for the PT cell free architecture of an $K = 400$ UEs over a square grid of 1 km^2. The number of APs is $L = 100$ and each one is equipped with $N = 4$ antennas. The system bandwidth is 20 MHz.

of both methods is that PT allows for UEs with the same pilot sequence to be served by the same AP, while Orthogonal PT has the condition that each AP should serve only the UEs that have orthogonal pilots allocated to them. For the purpose of this section both methods have the same performance, such that the simulation will be performed only for PT. Also, the power percent threshold will be $\beta_k^{\delta} = 0.9$, which corresponds to 90 % of the total UE power. Additionally, for comparison sake, the same simulation should be done for a small cell setup, with the same parameters. Fig. 1.9 and Fig. 1.10 show those results

The minimum rate value for the PT is of 155 Mbps, while it is merely 21 Mbps for the small cell. The rate values are very similar to the ones of the unscalable case, and to the OU and ONTU cases. When comparing to the cellular setup one notices that the rate variations are much more severe than on cell free architectures. Therefore, even the less effective of the cell free setups is still able to provide more rate uniformity than the cellular setup.

1.4.4 Work Organization

We shall establish comparison between the DCC method of [15], that operates with satisfaction on networks where the number of UEs do not change abruptly, and is dependent on a determined threshold value, the proposed methods, and the ones introduced by [8]. We shall also analyze all of the DCC methods in the case of imperfect channel statistics knowledge, since we will see in Chapter 4 that the DCC methods rely on knowledge of those parameters.

This work is organized as follows: Chapter 1 gives motivation for the use of CF networks, as well as the need for the alternative proposed DCC methods. Chapter 2 introduces the

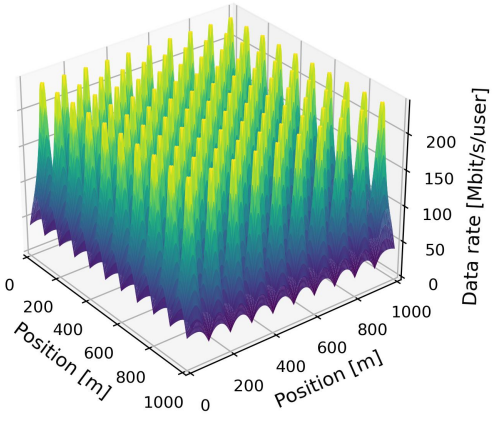


Figure 1.10 – UE throughput in Mbps for the small cell architecture of an $K = 400$ UEs over a square grid of 1 km^2. The number of APs is $L = 100$ and each one is equipped with $N = 4$ antennas. The system bandwidth is 20 MHz.

system model, while Chapter 3 discusses key parameters of importance for CF architecture, such as capacity bounds and channel estimation quality. Chapter 4 introduces the DCC methods, the ones proposed in this work, and the ones proposed by other authors. Chapter 5 discusses UL operation and evaluates performance and complexity metrics for the network. Chapter 6 compares the DCC and pilot methods numerically, in order to further motivate the use of the proposed methods. Finally, Chapter 7 presents the conclusions and suggestions on future research.

Chapter Summary

- Conventional Cellular systems aim to provide throughput to the UEs with manageable interference, by connecting them to the AP that provides the best performance. Systems such as cellular massive MIMO also manages interference by use of concepts such as spatial diversity and multiplexing.
- The throughput variations of cellular systems are often very significant, with the UEs near one of the AP having far greater performance than the ones far from it, or at cell edges. The ones at the edges are also more prone to suffer from interference. CF acts as a viable solution to those cases, since it does not require that each UE be connected to only one AP, allowing for the APs to jointly decode the information. This in turn makes the throughput more distance invariant.
- Among the benefits of CF is the higher value of SNR when compared to traditional cellular setup. Also, it has the ability to manage interference that far

surpasses conventional networks systems. Among the limitations are the increasing system complexity and the need for acceptable AP synchronization.

- The choice of the DCC method is of vital importance to CF system. Along this monograph some DCC methods ill be suggested and compared to find the ones that offer better system performance and simultaneously lower complexity, for large scale network deployment.

2 CELL FREE MIMO SYSTEM MODEL

This chapter aims to introduce the CF model. First, it introduces the concept of coherence block. Then, the CF UL model is explained as well as the notion of dynamic clustering. Next, scalability concepts are introduced and discussed. Finally, it introduces the channel model using the correlated Rayleigh fading model and then favorable propagation and channel hardening are introduced and discussed.

2.1 COHERENCE BLOCK MODEL

Mobile communication systems experience the phenomena of multipath scattering. This refers to the multiple paths the communication signal can propagate, thus being detected at different times at the receiver. Works such as [20] show that the amount of multipath scattering is inversely proportional to the essential bandwidth at which the transmitted signal suffers negligible distortion. This bandwidth is defined as B_c, the channel *coherence bandwidth*. In a similar fashion, the communication channel suffers variations over time, due to the relative displacement between transmitter and receiver. Thus, the rate at which the channel suffers negligible variations in time is defined as T_c, the channel *coherence time*. To account for both multipath propagation and the varying nature of the communication channel, the coherence block model is derived. The *coherence block*, C, is the time-frequency interval at which the channel is constant. It is obtained by

$$C = B_c T_c. \tag{2.1}$$

One can see that the coherence block is greater for systems with less mobility, since the channel varies slowly with time, and greater for systems with large coherence bandwidths, since more information can be transmitted over a larger range of frequencies. We will generally work with systems with high mobility to generate worst case scenarios, which might prove useful to estimate lower bound system behavior.

2.2 CELL FREE MODEL

Consider a communication network consisting of L APs, each deployed with N antennas, that are randomly located in a geographical area. The total number of antennas in the network is given by $M = NL$, and they serve a number of K single-antennas UEs. Such system can be seen in Fig. 2.1

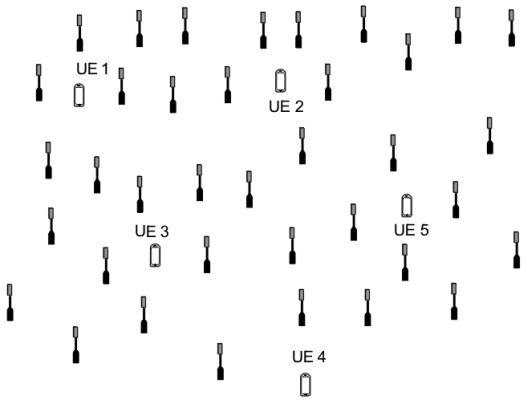


Figure 2.1 – Cell Free architecture.

Traditional cellular systems have the UE connected to the strongest AP in the grid. However, the CF system aims to connect one UE with multiple APs. The objective, as already discussed, is to jointly combine the transmitted signal in order to more reliably decode it. The model was first proposed by [18], where all of the APs serve all of the UEs. From Fig. 2.1, we can see that, taking UE 1 in consideration, there are few APs that are close to it, and a lot more that are far away from it. Since the greatest contribution to reliable estimation should come from the nearest APs, it is reasonable to say that the farthest ones have a much lower influence. We will see in the subsequent sessions that the complexity of the system should not increase much as the number of APs and UEs grow large. Therefore, the proposed method will be proven as unscalable, giving rise to alternative methods of AP clustering. Thus, we define one such method below:

DEFINITION 2.1 DCC is defined as the subset of APs, $\mathcal{M}_k \subset \{1, \ldots, L\}$, that serve UE k.

It is called *dynamic* because the clusters are regularly updated as the channel time-variations occur. Those changes may be due to different UE locations, the inclusion of new APs on the grid, interference, etc. Fig. 2.2 shows the example of a DCC for a set of 5 UEs. We can see from the figure that the colors represent each cluster $\mathcal{M}_k$ of the APs that serve each UE. Also, there are intersections between the clusters, so that into a DCC configuration, it is impossible to separate the APs into disjoint sets that serve disjoint subsets of the UEs. So, there are APs that are serving multiple UEs, which may give rise to interference. Therefore, methods to mitigate this interference will be discussed accordingly in the later sections.

Another important aspect of DCC, is that it should not change for various coherence blocks, since it follows the changes caused by macro propagation effects, such as the *large*

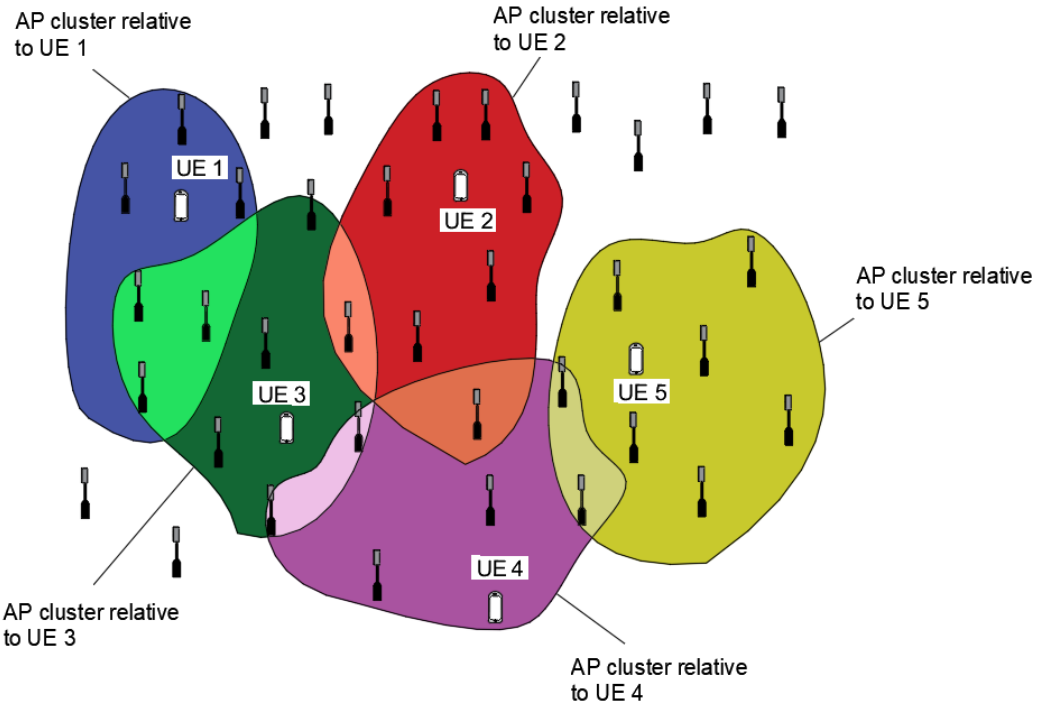


Figure 2.2 – DCC for a cell free MIMO with a large number of APs and a smaler number of UEs.

scale fading. Thus the channel fading variations should occur a fair amount of times before DCC is updated.

Since each UE can be served by any AP, it is convenient to define a matrix $\mathbf{D}_{kl} \in \mathbb{C}^{N \times N}$, for the UE index $k = 1, \ldots, K$ and the AP index $l = 1, \ldots, L$, so that

$$\mathbf{D}_{kl} = \begin{cases} \mathbf{I}_N & l \in \mathcal{M}_k \\ \mathbf{0}_{N \times N} & l \notin \mathcal{M}_k. \end{cases} \tag{2.2}$$

One can notice that the consideration of $\mathbf{D}_{kl} = \mathbf{I}_N$ for all l, k leads to the definition of CF system described by [18]. Also, this notation holds the special case of today's cellular networks, when $\mathcal{M}_k$ has only one element in it, which means that each UE is connected to only one AP.

2.3 TRANSMISSION

The channel between UE k and AP l into a coherence block is given by the complex-valued vector $\mathbf{h}_{kl} \in \mathbb{C}^N$. Each of its elements consist in the channel coefficient, recalling that in the block fading model, the channel is said to be flat, thus there is only one coefficient

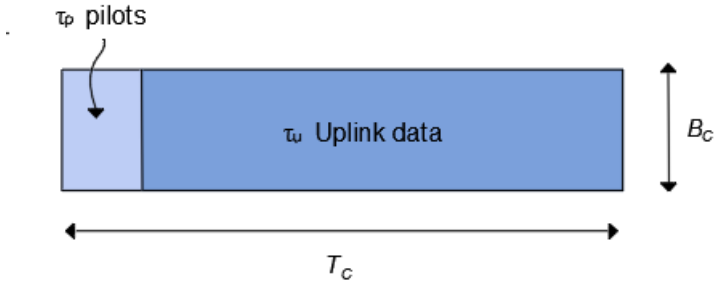


Figure 2.3 – The used transmission protocol of pilots and data over a coherence block of area $T_c B_c$.

to define the channel characteristics, between the UE and the n-th antenna of the AP. It is also interesting to define the collective channel, that consists of all of the APs to UE k, $\mathbf{h}_k \in \mathcal{C}^M$, as

$$\mathbf{h}_k = \begin{bmatrix} \mathbf{h}_{k1} \\ \vdots \\ \mathbf{h}_{kL}. \end{bmatrix} \tag{2.3}$$

Also, recalling the block fading model, the channel takes one random realization throughout the block duration, from a random distribution, to be defined in later sections. One aspect of the MIMO technology is the necessity of the APs to know the channel realizations, or at least an estimate of it, so that the interference between antennas and UEs be correctly mitigated. One efficient way to estimate the channel is to assign pilots, that is, a sequence of known symbols, to the UEs. This has the advantage of being a linear method of channel estimation, however it has the drawback of requiring actual symbols to be transmitted, which takes space from the coherence block. Therefore, the UL transmission requires τ_c symbols, of which τ_u are used to actually transmit useful data and τ_p of which are used to transmit the pilots, according to Fig. 2.3 In real-life implementations, there is also the DL, which can use the time division duplex method taking advantage of channel reciprocity. In this case, the coherence block would be divided between pilots, UL symbols and DL symbols. This monograph, however, covers only UL transmission, so the DL will not be analyzed.

2.4 UPLINK TRANSMISSION MODEL

During the UL transmission, the APs receive from all UEs a superposition of signals. The received signal, $\mathbf{y}_l^{ul} \in \mathbb{C}^N$ at the l-th AP is given by

$$\mathbf{y}_l^{ul} = \sum_{i=1}^{K} \mathbf{h}_{il} s_i + \mathbf{n}_l, \tag{2.4}$$

where $s_i \in \mathbb{C}$ is the UL data signal transmitted by UE i, with power $p_i = \mathbb{E}\{|s_i|^2\}$. The term $\mathbf{n}_l$ is the addictive white gaussian noise (AWGN) at the receiver, $\mathbf{n}_l \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_{ul}^2 \mathbf{I}_N)$. The UL signals can be of useful data or pilots, and the channel realizations are constant over the coherence block length. The noise, however, takes an independent realization at every transmitted symbol.

By taking a look on Eq. (2.4), one can notice that AP l is able to compute an estimate $\hat{s}_{kl}$ of the transmitted symbol s_k from UE k. Based on DCC however, the AP should only do that if $l \in \mathcal{M}_k$. For ease of notation, we should set $\hat{s}_{kl}$ to 0 if $l \notin \mathcal{M}_k$. This can be done by means of a receive combining vector $\mathbf{v}_{kl} \in \mathbb{C}^N$, that is used by AP l to decode information sent from UE k. Then, the effective receive combining vector is represented as

$$\mathbf{D}_{kl}\mathbf{v}_{kl} = \begin{cases} \mathbf{v}_{kl} & l \in \mathcal{M}_k \\ \mathbf{0}_N & l \notin \mathcal{M}_k \end{cases} \tag{2.5}$$

The received symbol is then estimated by taking the inner product between the effective combiner and the transmitted UL signal so that

$$\hat{s}_{kl} = \mathbf{v}_{kl}\mathbf{D}_{kl}\mathbf{y}_l^{\text{ul}}. \tag{2.6}$$

By use of Eq. (2.5):

$$\hat{s}_{kl} = \begin{cases} \mathbf{v}_{kl}^H \mathbf{h}_{kl} s_k + \sum\limits_{\substack{i=1 \\ i \neq k}}^{K} \mathbf{v}_{kl}^H \mathbf{h}_{il} s_i + \mathbf{v}_{kl}^H \mathbf{n}_l & l \in \mathcal{M}_k \\ 0 & l \notin \mathcal{M}_k. \end{cases} \tag{2.7}$$

We can see that the estimated symbol is composed of the desired estimated symbol $\mathbf{v}_{kl}^H \mathbf{h}_{kl} s_k$, the interference from the other UEs, $\sum\limits_{\substack{i=1 \\ i \neq k}}^{K} \mathbf{v}_{kl}^H \mathbf{h}_{il} s_i$, and the noise, $\mathbf{v}_{kl}^H \mathbf{n}_l$. The interference between UEs can be greatly reduced by allocating the pilots wisely. The receive combining vector is highly dependent on the CSI available, so the greater the quality of CSI available, the better the capacity of the combiner to suppress interference will be.

2.5 SCALABILITY CONCEPT AND CRITERIA

The coverage area of the network could be huge, which sometimes make the addition of APs necessary. More UEs can also be added to the network at any give moment. Such acts should not increase the overall capabilities of the existing APs and UEs, On traditional cellular networks, each cell has a maximum number of UEs that can be served by use of spatial multiplexing techniques, so the addition of new APs and UEs will not affect the operation of the cell, which makes cellular networks scalabale.

On the case of CF, however, we can't affirm that so easily, since the addition of those elements would impact on new effective combining vectors, since the overall combining vector is dependent on all of the APs that serve a given UE. Therefore, we should evaluate if the resources of given network remain sufficient even if the number of UEs grows indefinitely, or $K \to \infty$. So we must check if the following tasks remain implementable over this condition:

1. Signal processing for channel estimation
2. Signal processing for data reception and transmission
3. Fronthaul signaling for data and CSI sharing
4. Power allocation optimization

The authors in [8] have shown and discussed those conditions, and we should do the same to prove scalability or unscalability.

DEFINITION 2.2 Scalability - A CF MIMO is said to be *scalable* if all of the four tasks listed above have finite computation complexity as $K \to \infty$ for each AP.

It is important to say that this definition does not imply that an infinite amount of UEs will be added to the network, which would be impossible on real network deployments, but rather to show scalability issues that arise in the asymptotic regime. There are a handful of algorithms proposed that have reasonable complexity if generated at simulations related to academic papers, but unfeasible complexity on real-life applications, where the number of APs and UEs is large.

2.5.1 Example of an Unscalable Network

To better understand the concept of scalability it is reasonable to start with the opposite: unscalability. Let's suppose that the l-th AP is serving all the K UEs in the grid. The signal

detection and channel estimation are all performed locally at the AP. Then we shall analyze if the proposed configuration obeys the above scalability list.

Firstly, the channel estimation method, although not officially defined, will surely scale at lest linearly with K, since it was already mentioned that the method is linear. Besides that, the stored memory for the channel estimates is finite, and the number of UEs is infinite, therefore there is not such memory to store the estimates.

Then AP l should compute the symbol estimate, which, as already seen in dependent on the implementation of the combining vector. If said implementation is linear, then its complexity should again vary at least linearly with K, which makes the operation unscalable.

Next, the AP should forward the processed receives signals over the fronthaul links. Since the number of scalars grow with K, then again if $K \to \infty$ the number of scalars will grow indefinitely, making the system unscalable once more.

Finally, for adequate power allocation, power allocations methods should be deployed. Such methods also scale at least linearly with K, so such operation would be unscalable for an infinite number of UEs.

All of that has been done to show that the main issue is the fact that the AP is serving all of the UEs. Therefore, the method proposed on [18] is unscalable and thus the importance of selecting a cluster of serving APs by means of DCC.

2.5.2 Scalability Criteria

By noticing that $\mathbf{D}_{il}$ is non-zero if AP l serves UE i, we can define the set of UEs served by AP l

$$\mathcal{D}_l = \{i : \mathbf{tr}(\mathbf{D}_{il}) \geq 1, i \in \{1, \ldots, K\}\}. \tag{2.8}$$

The following lemma shows a condition on $\mathcal{D}_l$ that partially guarantees scalability.

LEMMA 2.1 For a network of unlimited UEs, $K \to \infty$, if the cardinality of $\mathcal{D}_l$ remains finite, for $l = 1, \ldots, L$, then the network satisfies the first three scalability conditions.

Proof 2.1: The AP l needs only to compute the channel estimates and combining vectors for the UEs on the set $\mathcal{D}_l$. Therefore, the complexity remains finite as long as the set $\mathcal{D}_l$ remains finite. Besides, the AP needs only send/receive data related to the UEs in the set over the fronthaul links, which is also finite. □

Lemma 2.1 implies that the number of active UEs that each AP serves should be lim-

ited. It is apparent that the AP should serve the UEs that are closer to it, since they would contribute more on the network performance. However, this task in not as simple, given the set $\mathcal{D}_l$ should be small and finite to ensure scalability, which can become challenging as the number of APs and UEs increase. That said, only the fourth condition of scalability remains to be proven, which will be discussed as follows:

LEMMA 2.2 Suppose every UE is assigned a transmit power by only one of its serving APs, and that this AP selects that power based only on information about the $\mathcal{D}_l$ UEs that it serves. Then, the power control scalability constraint is satisfied.

Lemma 2.2 implies that the transmit power should be selected in a distributed manner in order to achieve scalability. Thus, algorithms that rely on convex optimization or exhaustive searches would provide a polynomial complexity with respect to the UEs, and would be unscalable. In order to find a scalable solution, each algorithm should be implemented separately for each AP with no or limited interactions between the APs.

2.6 CHANNEL MODELLING

To better compare the performance of CF systems, reliable channel models should be used. In the case of wireless communications, due to multipath propagation and mobility, the channel variations get very complicated to predict. Thus, those variations appear to be random, come from a random distribution, making it possible to use stochastic models to describe it. The authors in [19] have suggested deterministic and stochastic models. The deterministic model has the clear advantage of more accurately reproduce the reality, however it has the drawback of only be applicable to very specific propagation scenarios, whereas he random model can be applied to a variety of environments. Nevertheless, the stochastic models limit the scope of each to the propagation conditions, such as urban, rural, suburban, etc.

A classic model for non line of sight (NLoS), used by many authors such as [20] is the *uncorrelated Rayleigh fading*, where the channel between AP l and UE k is given by $\mathbf{h}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \beta_{kl}\mathbf{I}_N)$. This happens because there is not a favorable direction for propagation, and moreover, there is mobility of the transmitter, which makes the superposition of the signals of each path become the sum of independent random variables, which by use of the Central Limit Theorem makes the channel Gaussian distributed.

On the coherence block model, the channel is generated as an independent random vector following a Gaussian distribution if Rayleigh fading is considered. Moreover, the variance β_{kl} denotes the *large scale fading*, and describes the average channel quality as the UE moves

around a relatively small area. Those coefficients may be determined by use of deterministic models, such as the geometric pathloss, and stochastic models, such as shadowing, which will be discussed later. The uncorrelated Rayleigh fading is the chosen model of the majority of the academic works because it is easy to implement and is also analytically tractable. However, measurements in real-life applications have shown that the channel measurements between antennas are not independent and makes the channel vectors elements actually correlated, since there are directions to the transmitted beam that are more likely to reach the UE, and also, there is the antenna array geometry, which greatly contributes to correlate the channel coefficients. We will not discuss that in detail, but there are very rare occasions in which uncorrelated Rayleigh fading, and they are generally very specific, such as the one shown in [21].

2.6.1 Correlated Rayleigh Fading and Large Scale Fading Model

This model is chosen because of its relative ease of tractability. It is used considering a NLoS channel with a fair amount of multipath propagation, that makes it possible to model the channel elements as Gaussian random variables. However, the channel variance cannot be considered a multiple of the identity matrix, since the channel is spatially correlated. Thus, the channel between AP l and UE k can be written as

$$\mathbf{h}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{R}_{kl}), \tag{2.9}$$

where $\mathbf{R}_{kl}$ is the spatial correlation matrix. The effects of small scale fading are accounted by the Gaussian distribution, and the effects of geometric pathloss, shadowing and array geometry are all inside the spatial correlation matrix, as explained in [19] and [20]. We will define that the large scale coefficients are the average value of the channel vector, or

$$\beta_{kl} = \frac{1}{N}\text{tr}(\mathbf{R}_{kl}). \tag{2.10}$$

Besides that, we will also assume, quite reasonably, that the channel vectors of different APs are uncorrelated, or $\mathbb{E}\{\mathbf{h}_{kl}\mathbf{h}_{kj}^H\} = \mathbf{0}_{N\times N}$ for $l \neq j$. This holds, because usually the APs are much more spaced than the UE to its connecting AP. Thus, the collective channel follows the distribution:

$$\mathbf{h}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_M, \mathbf{R}_k), \tag{2.11}$$

where $\mathbf{R}_k = \text{diag}(\mathbf{R}_{k1}, \ldots, \mathbf{R}_{kL}) \in \mathbb{C}^{M\times M}$ is the collective spatial correlation matrix.

The model used in this work assumes that the APs are deployed in urban environments with a large amount of UEs. Each AP is also deployed ten meters above the UE, according to the 3GPP Urban Microcell model, which can be found in [14]. The model was designed for 2GHz band, but is also usable for other sub- 6GHz frequency bands. According to this

model, the large scale fading coefficients (channel gains) are obtained, in dB, as

$$\beta_{kl}[\text{dB}] = -30.5 - 36.7 \log_{10}\left(\frac{d_{kl}}{1\text{m}}\right) + F_{kl} \tag{2.12}$$

in which d_{kl} is the physical distance between the AP l and the UE k. The chosen attenuation coefficient of 3.67 reflects a dense urban environment, as already discussed, and the attenuation of -30.5 dB is dependent on the carrier frequency, as well as the geometric pathloss. Also, the term F_{kl} represents the shadowing, which are fluctuations on the signal power that occur because of phenomena such as multiple reflections and scattering. Since each of those phenomena is considered to be independent from another, and since the total power fluctuation is multiplicative, then this phenomena may be interpreted as a sum of random variables, when the power losses are converted to dB. Thus, by use of the Central Limit Theorem, we can say that $F_{kl} \sim \mathcal{N}(0, \sigma_{sf}^2)$, where σ_{sf} represents the shadowing standard deviation. Moreover, we consider the correlated shadow fading model, described in [14]

$$\mathbb{E}\{F_{kl}F_{ij}\} = \begin{cases} \sigma_{sf}^2 2^{-\delta_{kl}/9\text{m}} & l = j \\ 0 & l \neq j \end{cases} \tag{2.13}$$

in which δ_{ki} is the distance between UE k and UE j. Again, the APs are assumed to be far away from each other, and thus, their shadowing coefficients are uncorrelated.

2.6.2 Spatial Correlation Matrix Using Local Scattering Model

The spatial correlation matrix is dependent on two factors: the angular distribution of the multipath components and the array geometry. In the simulations, we will use smaller sized APs. Because of that, it is reasonable to use the ULA, in which the N antennas are equally spaced along an horizontal line. We will consider that the spacing between antennas is the half-wavelength, and by doing that, the (i, j)-th element of the spatial correlation matrix $\mathbf{R}$ can be computed as

$$[\mathbf{R}]_{ij} = \beta \int\int e^{j\pi(i-j)\sin(\bar{\varphi})\cos(\bar{\theta})} f(\bar{\varphi}, \bar{\theta}) d\bar{\varphi} d\bar{\theta}, \tag{2.14}$$

where β is the large scale coefficient, $\bar{\varphi}$ is the azimuth angle and $\bar{\theta}$ is the elevation angle of a multipath component, all of them computed with respect to the broadside of the array, as shown in [22]. The $f(.)$ function is the joint probability density function (PDF) of the angles. This integral could be computed numerically for any given distribution of the angles. however, for simulations purposes, we will use the local scattering model, which assumes that the multipath components are symmetrically distributed across a straight line between the AP and the UE. This is made so that there be a scattering cluster around the UE and

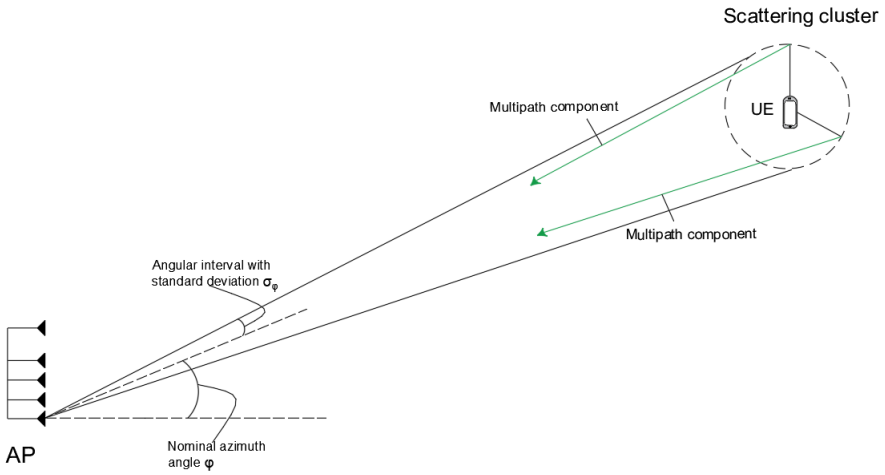


Figure 2.4 – NLoS propagation for the local scattering model. The scatterers are located around the UE. The figure shows two multipath components, and the azimuthal angle between the AP and the UE is φ, with σ_φ being the ASD

the multipath components arrive from nearby angles. Since those variations in the azimuth and elevation angles are said to be independent, then we can assume the jointly Gaussian distribution

$$f(\bar{\varphi}, \bar{\theta}) = \frac{1}{2\pi\sigma_\varphi\sigma_\theta} e^{-\frac{(\bar{\varphi}-\varphi)^2}{2\sigma_\varphi^2}} e^{-\frac{(\bar{\theta}-\theta)^2}{2\sigma_\theta^2}}, \tag{2.15}$$

where $\sigma_\varphi, \sigma_\theta$ are the angular standard deviations (ASDs). Fig. 2.4 shows the model,

This model works with sufficient precision when the angles are small, and thus the ASDs are small. Being a linear model, that relies on a Gaussian distribution, this model also has a closed form expression for the spatial covariance matrix elements and is also computationally inexpensive to compute, if one chooses a computationally inexpensive method to compute the integrals.

2.6.2.1 On the Effects of Spatial Covariance

We will now quickly discuss the effects of spatial covariance on the system. One efficient way to quantify this effect is to study the eigenvalues of $\mathbf{R}_{kl}$. For this, we analyze a matrix for an UE-AP system with $N = 10$ antennas and azimuth and elevation angles of $\varphi = 45^o$ and $\theta = -20^o$. The ASDs are considered the same and are $\sigma = 5^o, 15^o$ and 25^o. Fig. 2.5 shows the eigenvalues for the ASDs and the uncorrelated case. The uncorrelated case of $\mathbf{R}_{kl} = \mathbf{I}_N$ is also considered. As we can see, this case has the eigenvalues to have equal values. Since they are normalized such that $\text{tr}(\mathbf{R})/N = 1$, this value is 0 dB. As the ASD decreases, the eigenvalues are vastly different in values, and also, the gap between them is much larger. For instance , when the ASD is of 5^o, the gap between the first and second eigenvalues is near 10 dB, which means that the first eigenvalue is ten times larger than the

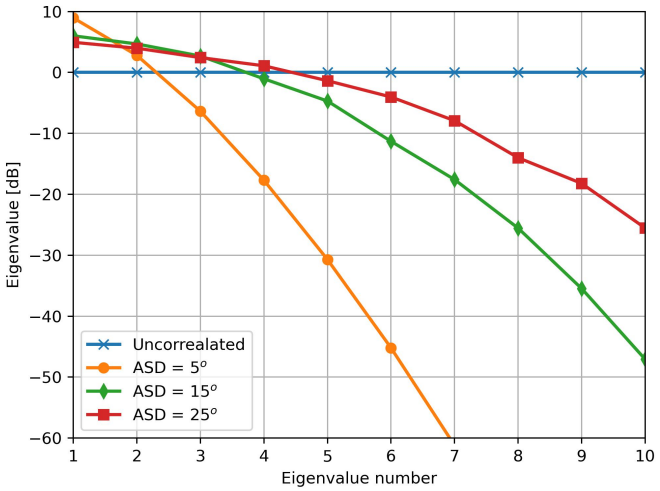


Figure 2.5 – Eigenvalues of the spatial correlation matrix when using the Gaussian local scattering model with $N = 10$ antennas at the AP. The azimuth and elevation angles are $\varphi = 45^o$ and $\theta = -20^o$. The uncorrelated case is also considered.

second. When considering the ASD of 25^o, the gap is much smaller, being near zero. This means that smaller ASDs will have a preferred propagation direction, since the eigenvectors corresponding to the eigenvalues of said direction will be dominant.

Spatial correlation studies show that the large gaps between eigenvalues, when the number of antennas is large, play a key role in mitigating the effects of the interference between UEs. This can be seen in works such as [18] and [22]. On the case of CF however, the spatial correlation matrix cannot do that, since the number of antennas is small at each AP. However, we will see in subsequent sessions that the effect of pilot contamination is much less severe for channel estimation when highly correlated systems are considered.

2.7 FAVORABLE PROPAGATION AND CHANNEL HARDENING DEFINITIONS AND APPLICATIONS

When analyzing MIMO systems, two concepts must always be taken in consideration: *channel hardening* and *favorable propagation*. Those concepts have been extensively used and discussed by authors such as [18], [22], [23] between others. They rely strongly on the number of antennas in the grid and also, on the spatial channel correlation properties. The aim of this section is to provide an insight on when those concepts may be used in theoretical or practical applications, as well as define them for CF systems.

2.7.1 Channel Hardening

The communication channel considers a stochastic behavior, in which every channel vector has one random realization along a given coherence block. Considering that the variance of the channel vectors realizations (which are related to the large scale coefficients) are small, the value of the realizations would be closer to the mean. Thus, although the channel realizations vary between APs and coherence blocks, those variations are small. Thus, the channel coefficients will have the same scalar value for every coherence block. When this happens, the achieved phenomena is called *channel hardening*, since the channel is getting more deterministic (being then hardened).

DEFINITION 2.3 For a set of $\mathcal{M}_k$ serving APs, and power coefficients $p_{k1}, p_{k2}, ..., p_{kL}$, the effective channel to UE k is said to achieve channel hardening if

$$\frac{\sum_{l \in \mathcal{M}_k} \sqrt{\frac{p_{kl}}{\mathbb{E}\{||\mathbf{h}_{kl}||^2\}}} ||\mathbf{h}_{kl}||^2}{\mathbb{E}\left\{\sum_{l \in \mathcal{M}_k} \sqrt{\frac{p_{kl}}{\mathbb{E}\{||\mathbf{h}_{kl}||^2\}}} ||\mathbf{h}_{kl}||^2\right\}} \rightarrow 1 \tag{2.16}$$

in the mean square sense as $N \rightarrow \infty$.

The definition 2.3 implies that the effective channel converges to its mean value as N grows large. If the channel model is the correlated Rayleigh fading, then the following ensures channel hardening:

$$\frac{\sum_{l \in \mathcal{M}_k} p_{kl} \frac{\text{tr}(\mathbf{R}_{kl}^2)}{N \beta_{kl}}}{N \left(\sum_{l \in \mathcal{M}_k} \sqrt{p_{kl} \beta_{kl}}\right)^2} \rightarrow 0 \quad \text{as} \quad N \rightarrow \infty \tag{2.17}$$

Proof 2.2: The convergence required for the channel hardening property is the mean squared sense convergence. Thus, we must apply the variance to the quotient of Eq.

(2.16) so that it goes asymptotically to 0

$$\mathbb{V}\left\{\frac{\sum_{l\in\mathcal{M}_k}\sqrt{\frac{p_{kl}}{\mathbb{E}\{||\mathbf{h}_{kl}||^2\}}}||\mathbf{h}_{kl}||^2}{\mathbb{E}\left\{\sum_{l\in\mathcal{M}_k}\sqrt{\frac{p_{kl}}{\mathbb{E}\{||\mathbf{h}_{kl}||^2\}}}||\mathbf{h}_{kl}||^2\right\}}\right\} = \frac{\sum_{l\in\mathcal{M}_k} p_{kl}\mathbb{V}\left\{\frac{||\mathbf{h}_{kl}||^2}{\sqrt{\mathbb{E}\{||\mathbf{h}_{kl}||^2\}}}\right\}}{\left(\sum_{l\in\mathcal{M}_k} p_{kl}\mathbb{E}\left\{\sqrt{\frac{||\mathbf{h}_{kl}||^2}{\mathbb{E}\{||\mathbf{h}_{kl}||^2\}}}\right\}\right)^2}$$
$$= \frac{\sum_{l\in\mathcal{M}_k} p_{kl}\frac{\text{tr}(\mathbf{R}_{kl}^2)}{\text{tr}(\mathbf{R}_{kl})}}{\left(\sum_{l\in\mathcal{M}_k}\sqrt{p_{kl}\text{tr}(\mathbf{R}_{kl})}\right)^2}$$
$$= \frac{\sum_{l\in\mathcal{M}_k} p_{kl}\frac{\text{tr}(\mathbf{R}_{kl}^2)}{N\beta_{kl}}}{N\left(\sum_{l\in\mathcal{M}_k}\sqrt{p_{kl}\beta_{kl}}\right)^2} \tag{2.18}$$

□

where we have used the independence between the channels of different APs and the fact that $\text{tr}(\mathbf{R}_{kl}) = N\beta_{kl}$. It is important to notice that channel hardening only applies to the asymptotic regime, with an unlimited number of antennas. However, one can define the degree of channel hardening for setups with limited antennas. Also, it has been shown in [22] that spatial correlation greatly reduces the degree of channel hardening. Indeed, if the uncorrelated case has all of the eigenvalues of the matrix $\mathbf{R}_{kl}$ having the same number, and $\text{tr}(\mathbf{R}_{kl}^2)$ is the sum of the squared eigenvalues, then $N\beta_{kl}^2 \leq \text{tr}(\mathbf{R}_{kl}^2) \leq N^2\beta_{kl}^2$, where the lower case occurs when all of the eigenvalues are equal, and the second when only the first eigenvalue is nonzero. By taking a look at Eq. (2.17) one can see that the lower case (uncorrelated) ensures the quotient stays proportional to $1/N$, which ensures that it converges to 0 as $N \to \infty$. On the upper case (highly correlated), the quotient value doesn't depend on N, and thus achieves higher values than the lower case.

2.7.1.1 Impact of the Geographical Distribution and Number of APs

For CF systems, the number of antennas at each AP is smaller than traditional Massive MIMO systems. The number of serving APs is however much larger. One interesting aspect would be if the effect of having a large number of antennas could be replaced by the effect of having a large number of APs for channel hardening. This, we shall analyze the degrees of channel hardening for CF systems. We shall assume, for simplicity, that the channel is uncorrelated,

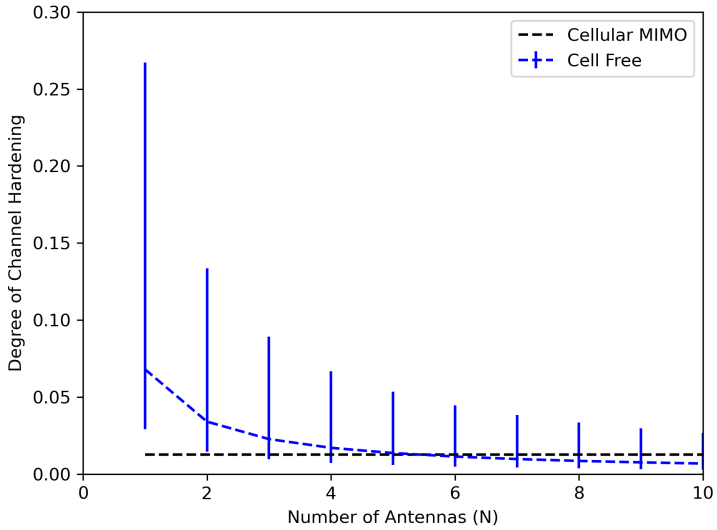


Figure 2.6 – The variance of the ratio between the effective channel and its mean value is plotted for different numbers of AP antennas, according to Eq. (2.19). The dotted black line corresponds to the Cellular MIMO case with 80 antennas, the dotted blue line corresponds to the median CF with 80 uniformly distributed APs, and the bars correspond to the region where 90 % of channel realizations occur.

so $\mathbf{R}_{kl} = \beta_{kl}\mathbf{I}_N$. The expression of Eq. (2.17) can then be written as

$$\frac{\sum_{l\in\mathcal{M}_k} p_{kl}\beta_{kl}}{N\left(\sum_{l\in\mathcal{M}_k}\sqrt{p_{kl}\beta_{kl}}\right)^2}. \tag{2.19}$$

In the case where $|\mathcal{M}_k| = 1$, or one AP has the term $p_{kl}\beta_{kl}$ much larger than the others, Eq. (2.19) becomes approximately $1/N$. In the case that all of the APs have the same β_{kl} term, it becomes $1/(N|\mathcal{M}_k|)$. Thus, we can see that channel hardening can also be obtained by having a large number of serving APs. However, since the number of antennas is still present on the equation, the degree of channel hardening should be smaller than the degree of traditional Massive MIMO systems. This degree should be even smaller if the terms $p_{kl}\beta_{kl}$ vary greatly between APs, which is the likely situation for practical systems, where the APs are spaced by several wavelengths. To check the degree of channel hardening, we consider a 500 m $\times$ 500 m square grid with one UE at its center. A number of $L = 80$ APs are uniformly distributed and transmit with equal power to the UE. The large scale fading model of Eq. (2.12) is used and the channel is said to be independent and identically distributed (iid) Rayleigh. Different values of N are considered. The results are compared to the traditional Massive MIMO system with only one AP equipped with $N = 80$ antennas. From Fig. 2.6, we can see that the degree of channel hardening (lower values corresponds to higher degrees of channel hardening) are lower for the Cellular MIMO case. In fact, for $N = 1$, the variance is fairly big for the CF case, which correspond to a fairly low degree of channel

hardening. It is not up to $N = 5$ antennas that the median value achieves the same value as the Massive MIMO. Still, the majority of the realizations are greater than this case, and some are substantially higher. It is only for $N = 8$ and onward that we can say that the degree of channel hardening is comparable to the traditional system. We can then notice that the total number of antennas on the grid becomes at minimum 8 times greater than the traditional case for comparable;e degree of channel hardening.

The degree of channel hardening is fairly smaller for the analyzed CF case (and even smaller if we consider DCC). This might appear as a nuisance, however, in practice this is not detrimental to the system, it only makes some tasks such as estimating lower bounds for the channel capacity easier, however, it is not mandatory for the operation of CF systems, as discussed in [8].

2.7.2 Favorable Propagation

When there is spatial multiplexing between multiple UEs, there is generally interference between them. The precoding or combiner can then be deigned to meet a balance between getting a strong desired signal power and mitigating interference. However, if the UE channels are orthogonal, then we can safely say that there is no interference and the inter-user interference is mitigated simply by use of MR. When this happens, we say that *favorable propagation* has been achieved.

If we take in consideration the received UL signal, then the interference between the desired signal of UE k and UE i is given by $\sum_{l=1}^{L} \mathbf{v}_{il}^H \mathbf{D}_{il} \mathbf{h}_{kl}$. If we use the MR combiner, that is, $\mathbf{v}_{kl} = \mathbf{h}_{kl}$, the *effective interfering channel* is

$$\sum_{l=1}^{L} \sqrt{p_{il}} \mathbf{h}_{il}^H \mathbf{D}_{il} \mathbf{h}_{kl} = \sum_{l \in \mathcal{M}_i} \sqrt{p_{kl}} \mathbf{h}_{il}^H \mathbf{h}_{kl}. \tag{2.20}$$

The magnitude of the effective interfering channel compared to the effective channel of the desired signal determine how strong is the interference. If the ratio is small, then the interference will be small and thus the system will benefit from favorable propagation. Then, we can define favorable propagation:

DEFINITION 2.4 UE k experiences favorable propagation with respect to UE i if the following is satisfied

$$\frac{\sum\limits_{l \in \mathcal{M}_i} \sqrt{p_{il}} \mathbf{h}_{il}^H \mathbf{h}_{kl}}{\mathbb{E}\left\{ \sum\limits_{l \in \mathcal{M}_k} \sqrt{p_{kl}} ||\mathbf{h}_{kl}||^2 \right\}} \to 0 \tag{2.21}$$

in the mean squared sense as $N \to \infty$.

Definition 2.4 says that the effective interfering channel should go asymptotically to 0 when normalized by the average value of the effective channel. Notice that the numerator depends on the set of APs that serve the interfering UE, whereas the denominator depends on the set of APs that serve the desired UE. If we consider correlated Rayleigh fading, then the following can be said

$$\frac{\sum\limits_{l \in \mathcal{M}_i} p_{il} \text{tr}(\mathbf{R}_{il}\mathbf{R}_{kl})}{\left(N \sum\limits_{l \in \mathcal{M}_k} \sqrt{p_{kl}} \beta_{kl} \right)^2} \to 0 \quad \text{as} \quad N \to \infty. \tag{2.22}$$

Proof 2.3: To prove Eq. (2.22) we need only remember that convergence on the mean squared sense means that the variance should go asymptotically to 0, then

$$\begin{aligned} \mathbb{V}\left\{ \frac{\sum\limits_{l \in \mathcal{M}_i} \sqrt{p_{il}} \mathbf{h}_{il}^H \mathbf{h}_{kl}}{\mathbb{E}\left\{ \sum\limits_{l \in \mathcal{M}_k} \sqrt{p_{kl}} ||\mathbf{h}_{kl}||^2 \right\}} \right\} &= \frac{\sum\limits_{l \in \mathcal{M}_i} p_{il} \mathbb{V}\left\{ \mathbf{h}_{il}^H \mathbf{h}_{kl} \right\}}{\left(\mathbb{E}\left\{ \sum\limits_{l \in \mathcal{M}_k} \sqrt{p_{kl}} ||\mathbf{h}_{kl}||^2 \right\} \right)^2} \\ &= \frac{\sum\limits_{l \in \mathcal{M}_i} p_{il} \text{tr}(\mathbf{R}_{il}\mathbf{R}_{kl})}{\left(\sum\limits_{l \in \mathcal{M}_k} \sqrt{p_{kl}} \text{tr}(\mathbf{R}_{kl}) \right)^2} \\ &= \frac{\sum\limits_{l \in \mathcal{M}_i} p_{il} \text{tr}(\mathbf{R}_{il}\mathbf{R}_{kl})}{\left(N \sum\limits_{l \in \mathcal{M}_k} \sqrt{p_{kl}} \beta_{kl} \right)^2} \end{aligned} \tag{2.23}$$

□

Like channel hardening, we can use the expression of Eq. (2.22) to evaluate a certain degree of favorable propagation. If this equation is near zero, then favorable propagation is approximately achieved. Differently from the channel hardening, which depends on the desired spatial correlation matrix $\mathbf{R}_{kl}$, this time the dependence of Eq. (2.23) is on the desired and interfering spatial correlation matrix product. This analysis is not as trivial, but we can affirm that the numerator gets its maximum when the spatial correlation matrices $\mathbf{R}_{kl}$ and $\mathbf{R}_{il}$ are all identical up to a scaling factor (meaning the UEs have the same spatial directivity, or

are highly correlated). On the other hand, the numerator gets near to zero when the matrices have identical eigenvectors, but the eigenvalues are such that they are matched together on the opposite order, that is, large eigenvalues of a matrix are multiplied by small eigenvalues of the other matrix. This means that the UEs have very different spatial directivity, or their channels are highly uncorrelated. This last case clearly represents the situation that has the largest degree of favorable propagation.

2.7.2.1 Impact of the Geographical Distribution and Number of APs

Let us assume that the desired UE as well as the interfering UE have spatially uncorrelated channels. (that is, $\mathbf{R}_{kl} = \beta_{kl}\mathbf{I}_N$ and $\mathbf{R}_{il} = \beta_{il}\mathbf{I}_N$). The expression of Eq. (2.22) is then simplified to

$$\frac{\sum\limits_{l\in\mathcal{M}_i} p_{il}\beta_{kl}\beta_{il}}{N\left(\sum\limits_{l\in\mathcal{M}_k}\sqrt{p_{kl}}\beta_{kl}\right)^2} \tag{2.24}$$

This expression again decays with the number of antennas N, which means that more antennas guarantee a higher degree of favorable propagation. We can notice that the sum of the numerator depends on the subset of APs that serve the interfering UE, and the denominator on the subset of APs that serve the desired UE. Considering that the UEs are far away from each other, then its serving APs may be also different, and we might say that the coefficients β_{kl} of the numerator will be very small. This means that favorable propagation could be achieved even if $N = 1$. On the other hand, if the subsets $\mathcal{M}_i$ and $\mathcal{M}_k$ are similar, then it might be necessary for the APs to have more antennas to reach the same degree of favorable propagation. Also, the power control plays a key role in this effect. If the power coefficients p_{il} are small compared to p_{kl}, we can also guarantee a higher degree of favorable propagation with few antennas. We shall demonstrate the effect of the number of antennas at each AP as well as the geographical location of the APs. For that, we shall use the same setup as the one that was used to generate Fig. 2.6, with the addition of an interfering UE that is placed either 150 m or 30 m to the east of the desired UE. However, instead of having all of the APs save the UEs, the UE is served by the 6 APs that have the largest large scale fading coefficients, and all of them transmit with equal power. Fig. 2.7 shows those results. We can immediately see from Fig. 2.7 that the situation of nearby interfering-desired UE pair, the variance can assume large values, which mean lower degree of favorable propagation. In fat, although the median is comparable to the reference case (Cellular Massive MIMO), at $N = 10$, the fluctuations of the random realizations is still very large more than half of the time. This shows that we cannot achieve a satisfactory degree of favorable propagation for nearby UEs even if we increase the number of antennas. In fact, we would need to increase the number of antennas to an extent that it would be comparable to the reference case number of antennas.

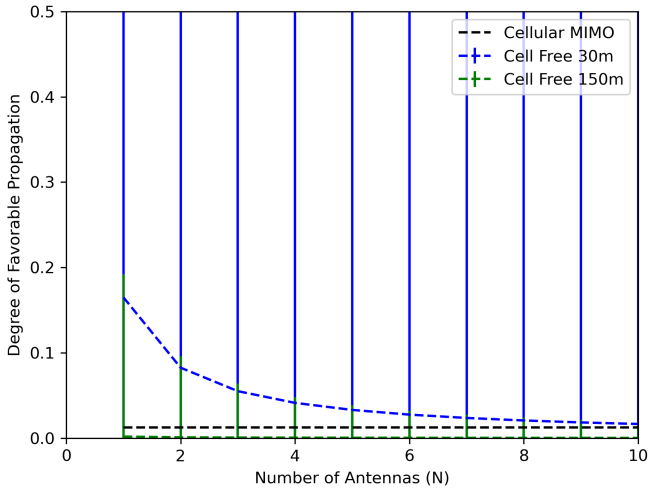


Figure 2.7 – The variance of the ratio between the effective interfering channel and the mean of the effective channel described in Eq. (2.24) is plotted for various N. When the variance is small there is a high degree of favorable propagation between the UE pair. The Cellular Massive MIMO corresponds to a setup of a single AP with $n = 80$ antennas. The CF setup consists of an interfering UE placed at either 30 m or 150 m to the left of the desired UE, both served by the 6 APs providing the best channel gains among a pool of $L = 80$ APs. The line shows the median value and the bars show the 90 % interval of random realizations.

For the case of far away interfering-desired UE pair, however, the degree of favorable propagation is fairly high for a small number of antennas at each AP. In fact, it becomes near identical to the reference case at $N = 4$, although the results at $N = 2$ still show that the variance fluctuations are not as significant as the nearby case. Favorable propagation is a desirable property that makes the use of simple combiner vectors possible. It is not however mandatory for the operation of CF systems. For instance, the case of UEs that are nearby each other, and thus are served by partially the same APs, favorable propagation cannot be achieved. Thus, since such situations are fairly common, it is necessary for the combiner vector to have an interference suppressing method, as it will be discussed later.

Finally, although the mathematical models of favorable propagation and channel hardening may look very similar, the phenomena are not the same: one UE might experience favorable propagation with respect to other, but have a low degree of channel hardening. The opposite could also happen, as well as the UE having none of those properties or both of them. It is important not to confuse the definitions or think they are the same.

Chapter Summary

- The coherence block model is an efficient way to analyze the effects of more complex techniques as for example OFDM with less computational complexity. One must only ensure that the transmitted signal has lower bandwidth than the

channel coherence bandwidth and lower duration than the channel coherence time.

- A CF MIMO system consists of L APs. equipped with N antennas, that jointly serve the UEs in the grid, to jointly process the information transmitted by or received from the UEs over the duration of one coherence block.
- By having a large number of APs antennas, such that $M = LN >> K$, the system has a high number of spatial degrees of freedom, to suppress interference by use of linear methods. Such can also be said to traditional Massive MIMO systems, where each UE connects to only one AP equipped with M antennas. However, on the latter case, the fact that the antennas are near each other makes so that the channel stochastic means converge to the real mean and the interference between UEs be suppressed.
- The latter case also implies that the degrees of channel hardening and favorable propagation are high, whereas the CF case has the APs scattered geographically, thus increasing channel variability and decreasing the degrees of channel hardening and favorable propagation. On the matter of channel hardening, increasing the number of antennas at each AP seems to solve this matter. On the matter of favorable propagation, ensuring the UEs are far from each other seems to increase its degree.
- The situation of large coverage areas makes impracticable to connect every AP to a UE in the grid. Scalable methods must be implemented, such that at every moment new APs amd UEs can be added to the grid with minimum increasing computational capability of each AP and fronthaul capacity.
- The fact of the APs in CF systems having a low number of antennas, the spatial correlation between their channel coefficients must be accounted for, given the fact that the channels in real life applications are spatially correlated. Also, the correlation between the shadowing coefficients of each UE must be considered, since in real life applications the shadowing between different UEs might be correlated.

3 CHANNEL ESTIMATION AND CAPACITY BOUNDS

This chapter aims to introduce channel estimation theory and channel capacity concepts. First, MMSE theory is discussed. Then channel capacity theory is introduced, and lower bounds for it are derived. Finally channel estimation theory is discussed, as well as the factors that could improve or worsen its quality, such as spatial correlation, geographical locations, system architecture and pilot contamination interference.

3.1 MINIMUM MEAN SQUARED ERROR ESTIMATION THEORY

To properly process the signal information from the APs to the UE, MIMO systems must have some degree of knowledge of the channel coefficients. To begin our analysis, we consider a realization of a Gaussian random variable from an observation that is corrupted by independent additive Gaussian noise. There is a large variety of methods to estimate such variables, however we shall use the MMSE method, as defined below:

DEFINITION 3.1 Consider a random variable $\mathbf{x} \in \mathbb{C}^N$ defined over an interval Ω and let the variable $\hat{\mathbf{x}}(\mathbf{y})$ denote an estimator of $\mathbf{x}$ obtained from the observation of $\mathbf{y} \in \mathbb{C}^M$. The choice of estimator $\hat{\mathbf{x}}(\mathbf{y}) : \mathbb{C}^M \to \mathbb{C}^N$ that minimizes the mean squared error (MSE)

$$\mathbb{E}\left\{||\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y})||^2\right\} \tag{3.1}$$

is defined the MMSE estimator of $\mathbf{x}$. It can be obtained as

$$\hat{\mathbf{x}}_{\text{MMSE}}(\mathbf{y}) = \int_{\Omega} \mathbf{x} f(\mathbf{x}|\mathbf{y}) d\mathbf{x}, \tag{3.2}$$

in which $f(\mathbf{x}|\mathbf{y})$ is the conditional PDF of the variable $\mathbf{x}$ given the variable $\mathbf{y}$.

The MMSE estimator of the Gaussian random variable that is corrupted by additive independent noise and independent interference can be obtained in closed from as follows

LEMMA 3.1 Consider we wish to estimate the N-dimensional vector $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{R})$, with a positive semi-definite correlation matrix $\mathbf{R}$, from the observation $\mathbf{y} = \mathbf{x}q + \mathbf{n} \in \mathbb{C}^N$. The transmitted signal $q \in \mathbb{C}$ is known and vector $\mathbf{n} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{S})$ is an independent

noise/interference vector with a positive definite correlation matrix.
The MMSE estimator of $\mathbf{x}$ is

$$\hat{\mathbf{x}}_{\text{MMSE}}(\mathbf{y}) = q * \mathbf{R}\left(|q|^2\mathbf{R} + \mathbf{S}\right)^{-1}\mathbf{y}, \tag{3.3}$$

with the estimation error correlation matrix as

$$\mathbf{C}_{\text{MMSE}} = \mathbf{R} - |q|^2\mathbf{R}\left(|q|^2\mathbf{R} + \mathbf{S}\right)^{-1}\mathbf{R} \tag{3.4}$$

and the MSE as

$$\text{MSE} = \text{tr}\left(\mathbf{R} - |q|^2\mathbf{R}\left(|q|^2\mathbf{R} + \mathbf{S}\right)^{-1}\mathbf{R}\right). \tag{3.5}$$

Proof 3.1: We first begin this demonstration by defining the following

DEFINITION 3.2 An N-dimensional circularly symmetric complex Gaussian random vector $\mathbf{x}$ with mean $\mu \in \mathbb{C}^N$ and positive definite covariance matrix $\mathbf{R} \in \mathbb{C}^{N\times N}$ has the PDF

$$f(\mathbf{x}) = \frac{e^{-(\mathbf{x}-\mu)^H\mathbf{R}^{-1}(\mathbf{x}-\mu)}}{\pi^N \det(\mathbf{R})} \tag{3.6}$$

then , we notice that

$$f(\mathbf{x}) = \frac{e^{-(\mathbf{x})^H\mathbf{R}^{-1}(\mathbf{x})}}{\pi^N \det(\mathbf{R})} \tag{3.7}$$

and

$$f(\mathbf{y}) = \frac{e^{-(\mathbf{y})^H(|q|^2\mathbf{R}+\mathbf{S})^{-1}(\mathbf{y})}}{\pi^M \det(|q|^2\mathbf{R} + \mathbf{S})}. \tag{3.8}$$

We then compute the conditional probability $f(\mathbf{y}|\mathbf{x})$,

$$f(\mathbf{y}|\mathbf{x}) = \frac{e^{-(\mathbf{y}-\mathbf{x}q)^H\mathbf{S}^{-1}(\mathbf{y}-\mathbf{x}q)}}{\pi^M \det(\mathbf{S})} \tag{3.9}$$

and we use the Bayes theorem to obtain the conditional probability $f(\mathbf{x}|\mathbf{y})$,

$$f(\mathbf{x}|\mathbf{y}) = \frac{f(\mathbf{y}|\mathbf{x})f(\mathbf{x})}{f(\mathbf{y})}. \tag{3.10}$$

Thus, the desired conditional probability is given by

$$f(\mathbf{x}|\mathbf{y}) = \frac{\frac{e^{-(\mathbf{y}-\mathbf{x}q)^H\mathbf{S}^{-1}(\mathbf{y}-\mathbf{x}q)}}{\pi^M\det(\mathbf{S})}\frac{e^{-(\mathbf{x})^H\mathbf{R}^{-1}(\mathbf{x})}}{\pi^N\det(\mathbf{R})}}{\frac{e^{-(\mathbf{y})^H(|q|^2\mathbf{R}+\mathbf{S})^{-1}(\mathbf{y})}}{\pi^M\det(|q|^2\mathbf{R}+\mathbf{S})}}$$
$$= \frac{e^{-(\mathbf{x}-(\mathbf{R}q(|q|^2\mathbf{R}+\mathbf{S})^{-1}\mathbf{y})^H(\mathbf{R}^{-1}+|q|^2\mathbf{S})^{-1})(\mathbf{x}-(\mathbf{R}q(|q|^2\mathbf{R}+\mathbf{S})^{-1}\mathbf{y})}}{\pi^N\det((\mathbf{R}^{-1}+|q|^2\mathbf{S}^{-1})^{-1})}, \tag{3.11}$$

where we have used the matrix relationships

$$(\mathbf{A}+\mathbf{x}\mathbf{x}^H)^{-1} = \mathbf{A}^{-1} - \frac{1}{1+\mathbf{x}^H\mathbf{A}^{-1}\mathbf{x}}\mathbf{A}^{-1}\mathbf{x}\mathbf{x}^H\mathbf{A}^{-1} \tag{3.12}$$

$$(\mathbf{A}+\mathbf{x}\mathbf{x}^H)^{-1}\mathbf{x} = \frac{1}{1+\mathbf{x}^H\mathbf{A}^{-1}\mathbf{x}}\mathbf{A}^{-1}\mathbf{x}. \tag{3.13}$$

By taking a look at Eqs. (3.2) and (3.11), we can see that the mean of $f(\mathbf{x}|\mathbf{y})$ is the estimator. Thus,

$$\hat{\mathbf{x}} = q\mathbf{R}(|q|^2\mathbf{R}+\mathbf{S})^{-1}\mathbf{y}. \tag{3.14}$$

Following the same logic, the correlation error matrix is

$$\mathbf{C} = (\mathbf{R}^{-1}+|q|^2\mathbf{S}^{-1})^{-1} = \mathbf{R} - |q|^2\mathbf{R}(|q|^2\mathbf{R}+\mathbf{S})^{-1}\mathbf{R} \tag{3.15}$$

and the MSE is obtained simply by taking the trace of $\mathbf{C}$, by the trace definition, completing the proof. Also, the estimate is distributed as

$$\hat{\mathbf{x}}_{\text{MMSE}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{R}-\mathbf{C}_{\text{MMSE}}). \tag{3.16}$$

□

3.2 CAPACITY BOUNDS

The SE is the mean amount of information that can be transferred per complex-valued sample per bandwidth with an arbitrarily small error. This is a very natural metric in broadband applications, where the required data volume is large. The SE is the chosen metric for analysis of this work, and will be extensively used throughout all of its duration.

When communication occurs over a bandwidth B, the Nyquist-Shannon sampling theorem dictates that the communication signal can be fully recovered by B complex valued samples per second. What the SE describes is how much data can be transmitted per such complex sample. Thus, given that there are B samples per second, the measure of the SE

must be either of bit per complex sample or bit per second per Hertz, bps/Hz. Another related metric is the *information rate*, which is the product of the SE by the bandwidth, resulting in a measure of bit/s.

It is of great importance to define which would be the upper bound on the SE of some architectures. For that, there is the *channel capacity* metric. Its definitions can be found on the work of its original author, [24], and [17], as well as various other information theory papers. In wireless communications, it is usually of interest the situation where the communication channel is composed of a receives signal that is a scaled version of the desired signal plus interference/noise. Those channels are the already defined AWGN channels, since each input signal results into an output signal that is independent of previous and future inputs, thus defining a *memoryless channel*. We shall now begin our analysis for the capacity bounds for random channels, one of the metrics that act as a comparison measure in this monograph.

3.2.1 Capacity Bounds for Random Channels

For this section, we consider fading channels, that, as already seen, are random complex variables that take a new realization every coherence block. From this setup rises the concept of *ergodic capacity*. The transmission generates many realizations of the random variable describing the channel and the word "ergodic" means that all of the statistical properties of the channel are obtainable from a single sequence of channel realizations. We begin with the case of which the only disturbance is Gaussian noise.

LEMMA 3.2 Consider a discrete memoryless channel that has the input $x \in \mathbb{C}$ and output $y \in \mathbb{C}$, given by

$$y = hx + n, \tag{3.17}$$

where $n \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ is the independent noise and the input distribution has power limited by $\mathbb{E}\{|x|^2\} \leq p$. The channel h is a realization of a random variable $\mathbb{H}$ that is independent of the signal and the noise. and known at the receiver. The ergodic channel capacity is then given by

$$C = \mathbb{E}\left\{\log_2\left(1 + \frac{p|h|^2}{\sigma^2}\right)\right\}, \tag{3.18}$$

where the expectation operator is with respect to h. This expression is achieved when selecting the input distribution $x \sim \mathcal{N}_{\mathbb{C}}(0, p)$.

Proof 3.2: We shall fist define the differential entropy

DEFINITION 3.3 Given a random variable y with PDF $f(y)$ defined over the domain Ω_y. The differential entropy of y is given by

$$\mathcal{H}(y) = -\int_{\dagger} \log_2(f(y))f(y)dy. \tag{3.19}$$

If the random variable x with PDF $f(x)$ defined over the domain Ω_x is given, then the conditional differential entropy is

$$\mathcal{H}(y|x) = -\int_{\Omega_y}\int_{\Omega_x} \log_2(f(y|x))f(y|x)f(x)dxdy. \tag{3.20}$$

The differential entropy $\mathcal{H}(y)$ measures the amount of information that the observation of y conveys. If we want to quantify the mutual information between the variables x and y, it is the difference between the differential entropy of x and the differential entropy of x given y:

$$\mathcal{I}(x;y) = \mathcal{H}(x) - \mathcal{H}(x|y). \tag{3.21}$$

Next, given $x \sim \mathcal{N}_{\mathbb{C}}(0,q)$, we evaluate the differential entropy of x, defined in the domain $\mathbb{D}$, as

$$\begin{aligned}\mathcal{H}(x) &= -\int_{\mathbb{D}} \log_2\left(\frac{e^{-|x|^2/q}}{\pi q}\right)\frac{e^{-|x|^2/q}}{\pi q}dx \\ &= \frac{\log_2(e)}{q}\underbrace{\int_{\mathbb{D}}|x|^2\frac{e^{-|x|^2/q}}{\pi q}dx}_{=q} + \log_2(\pi q)\underbrace{\int_{\mathbb{D}}\frac{e^{-|x|^2/q}}{\pi q}dx}_{=1} \\ &= \log_2(e\pi q),\end{aligned} \tag{3.22}$$

where the integrals are solved by applying the properties of the zero-mean Gaussian distribution (the variance and the total probability).
Next, we have $y = hx + n$. We have that the conditional entropy of y given x is only dependent on the noise n, since only the noise is unknown at the output. Thus, we have that

$$\mathcal{H}(n) = \log_2(e\pi\sigma^2). \tag{3.23}$$

Next, we shall compute the differential entropy of y. Since the power of the output signal y is given by $\mathbb{E}\{|y|^2\} = |h|^2\mathbb{E}\{|x|^2\} + \mathbb{E}\{|n|^2\} \leq |h|^2p + \sigma^2$. We can find in [24] and [17] that the entropy is maximized when the random variable has a circularly symmetric

complex Gaussian distribution. Thus, we shall select $x \sim \mathcal{N}_{\mathbb{C}}(0, q)$. By doing that, the maximum of the mutual information between x and y is the channel capacity:

$$C = \sup_{f(x)}(\mathcal{H}(y, \mathbb{H}) - \mathcal{H}(y, \mathbb{H}|x)) = \sup_{f(x)}(\mathbb{E}\left\{\mathcal{H}(y, \mathbb{H} = h) - \mathcal{H}(y, \mathbb{H} = h|x)\right\}) \quad (3.24)$$

Since the expected value considers the value of h as deterministic, we can write

$$C = \log_2\left(e\pi(|h|^2 p + \sigma^2)\right) \log_2(e\pi\sigma^2) = \log_2\left(1 + \frac{p|h|^2}{\sigma^2}\right) \quad (3.25)$$

completing the proof. □

The value of (3.25) is the instantaneous capacity value. If we instead wish to obtain the ergodic capacity, we should write

$$C = \log_2\left(1 + \frac{p\mathbb{E}\{|h|^2\}}{\sigma^2}\right). \quad (3.26)$$

Next, we extend our analysis to the case of an interfering signal on the AWGN channel. The ergodic capacity in this case does not have a closed form expression, since the distribution of such interference is generally not known and also, it is not always additive. However, a lower bound can be obtained if we treat the interference as noise, as seen in [25] and [26]. Lemma 3.3 provides a lower bound for the capacity.

LEMMA 3.3 Consider a memoryless interference channel with input $x \in \mathbb{C}$ and $y \in \mathbb{C}$ given by

$$y = hx + v + n, \quad (3.27)$$

with $n \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ as the independent noise, $h \in \mathbb{C}$ as the channel response, known by the receiver, and $v \in \mathbb{C}$ a random interference with an arbitrary distribution. The input power is bounded as $\mathbb{E}\{|x|^2\} \leq p$.
Suppose now $h \in \mathbb{C}$ is a random realization of the random variable $\mathbb{H}$ and $\mathbb{U}$ is the random variable with realization u that describes the interference variance. The realizations of those random variables are all known by the receiver. If the noise n is conditionally independent of v given h and u, the interference v has conditional zero mean, $\mathbb{E}\{v|h, u\} = 0$, and conditional variance $p_v(h, u) = \mathbb{E}\{|v|^2 h, u\}$, and also the interference is conditionally uncorrelated with the input, that is, $\mathbb{E}\{x^* v|h, u\} = 0$, then the

lower-bound ergodic channel capacity C is

$$C \geq \mathbb{E}\left\{\log_2\left(1+\frac{p|h|^2}{p_v(h,u)+\sigma^2}\right)\right\} \tag{3.28}$$

.

Proof 3.3: First, we express the channel capacity,

$$C = \sup_{f(x)}(\mathcal{H}(x)-\mathcal{H}(x|y)). \tag{3.29}$$

if $\mathbb{H}$ and $\mathbb{U}$ are known at the receiver, the capacity can be lower bounded as

$$\begin{aligned} C &\geq \mathcal{H}(x)-\mathcal{H}(x|y,\mathbb{H},\mathbb{U}) \\ &= \log_2(e\pi p)-\mathbb{E}\left\{\mathcal{H}(x|y,\mathbb{H}=h,\mathbb{U}=u)\right\}, \end{aligned} \tag{3.30}$$

where the result of Eq. (3.30) follows from the fact that $x \sim \mathcal{N}_{\mathbb{C}}(0,p)$, using the results of Eq. (3.22), and then conditioning on particular realizations of $\mathbb{H}$ and $\mathbb{U}$. The expectation of Eq. (3.30) is with respect to h and u. Now, we assume that the variable x is estimated from y by use of MMSE. From Eq. (3.14), its scalar form is given by

$$\hat{x} = \frac{\mathbb{E}\{xy^*\}}{\mathbb{E}\{|y|^2\}}y, \tag{3.31}$$

where

$$\mathbb{E}\{xy^*\} = \mathbb{E}\{xx^*h^*\}+\mathbb{E}\{xv^*\}+\mathbb{E}\{xn^*\} = ph^*, \tag{3.32}$$

since the noise and the input are independent and also the interference has conditional zero mean. Also,

$$\mathbb{E}\{|y|^2\} = p|h|^2+p_v+\sigma^2, \tag{3.33}$$

utilizing the independence between the noise and the interference and the conditional zero mean for the interference. From that, the MSE of the MMSE estimator is as follows

$$\text{MSE} = \mathbb{E}\{|x|^2\}-\frac{\mathbb{E}\{xy^*\}}{\mathbb{E}\{|y|^2\}} = p-\frac{p^2|h|^2}{p|h|^2+p_v+\sigma^2}. \tag{3.34}$$

If we take the upper bound of the conditional differential entropy

$$\mathcal{H}(x|y) \overset{(a)}{=} \mathcal{H}(x-\hat{x}|y) \overset{(b)}{\leq} \mathcal{H}(x-\hat{x}) \overset{(c)}{\leq} \log_2(e\pi\text{MSE}), \tag{3.35}$$

where (a) follows from the fact that the estimate $\hat{x}$ is a known value, and the entropy is invariant to constant displacement, (b) follows from removing the remaining information in y (which also does not change the entropy), and (c) follows from the fact that the largest entropy is obtained from a random variable that is circularly symmetric and Gaussian distributed. Since the random variable $x - \hat{x}$ has zero mean and variance MSE, the expression of Eq. (3.35) is obtained. Finally, by using the expression of Eq. (3.30), we obtain

$$\begin{aligned} C &\geq \log_2(e\pi p) - \log_2(e\pi \text{MSE}) = -\log_2\left(1 - \frac{p|h|^2}{p|h|^2 + p_v + \sigma^2}\right) \\ &= \log_2\left(\frac{p|h|^2 + p_v + \sigma^2}{p_v + \sigma^2}\right) = \log_2\left(1 + \frac{p|h|^2}{p_v + \sigma^2}\right), \end{aligned} \tag{3.36}$$

completing the proof. □

Although the expression of Eq. (3.28) may look simple at the first look, it is actually of significant importance for this work. In later sections, the interference term, p_v, will contain not only the interference of one source, but also the interference of various sources, and also the "noise", or uncertainty of the channel estimation. This expression is the base expression of all of the SE derivations that will be used on this work. In the next section, we discuss channel estimation techniques

3.3 CHANNEL ESTIMATION

This section aims to describe how channel estimation is performed into CF MIMO systems. There are various ways to estimate the channel: one could estimate it by use of non-linear methods such as the ones used in [27], [28] and [29]. For instance, [28] uses an estimation method that relies on the singular value decomposition of the channel matrix in order to separate the subspaces of the desired signal and the interference/noise, in order to estimate the desired channel with minimum interference from other UEs. There are also linear methods of channel estimation. [9] and [30] use a method of transmitting known pilot signals in the UL and observing the received signal. Such a method of estimation will be the chosen method through this work, and we begin to introduce it by describing the system model as follows

3.3.1 Uplink Pilot Transmission

The model is described by Fig. 2.2: K single antennas UEs are served by L APs, each of them equipped with N antennas, and are randomly deployed along the coverage area. The APs are connected to the CPUs via fronthaul links, that can be used for sharing the CSI and channel statistics among the APs, that are needed to decode the UL data. UE k is served by the APs belonging to the subset $\mathcal{M}_k$, which is assumed to be fixed and known everywhere. in order to correctly decode the desired signal from the UE, knowledge of the channel response between the AP and UE is required. It is of importance that AP l has the estimates of the channel vector $\mathbf{h}_{kl}$ from the k-th UE if $l \in \mathcal{M}_k$. The channels are assumed to be constant during the duration of one coherence block, and take independent realizations from one block to another. Therefore, it is necessary to estimate the channel only once per coherence block. From the transmission protocol in Section 2.3, τ_p symbols are transmitted as pilots during each coherence block. Thus, each UE can transmit a pilot sequence of τ_p symbols, and each AP can use the receives signals in order to estimate the channel.

Ideally, every UE would have a different orthogonal pilot sequence in order to cancel the interference between the channel estimate of UEs. However, since the pilot sequence is τ_p-dimensional, we can only have a total of τ_p mutually orthogonal pilot sequence. Also, the length of the coherence block imposes the constraint $\tau_p \leq \tau_c$, that makes impossible to assign mutually orthogonal sequences in real life networks, where the number of UEs, K, is larger than τ_c. Thus, the network will use the set of τ_p pilot sequences, $\phi_1, ..., \phi_{\tau_p} \in \mathbb{C}^{\tau_p}$, that satisfy $||\phi_t||^2 = \tau_p$ for $t = 1, ..., \tau_p$. Also, they must be mutually orthogonal:

$$\phi_{t_1}^H \phi_{t_2} = \begin{cases} \tau_p & t_1 = t_2 \\ 0 & t_1 \neq t_2. \end{cases} \tag{3.37}$$

More than one UE might be assigned the same pilot sequence. We denote the index of the pilot assigned to UE k as $t_k \in \{1, ..., \tau_p\}$ and we define

$$\mathcal{P}_k = \{i : t_i = t_k, i = 1, ..., K\} \subseteq \{1, ..., K\} \tag{3.38}$$

as the set of UEs that use the pilot t_k. Thus, the transmitted pilot signal, scaled by the UL transmitted power, η_k is received by AP l as the received UL matrix $\mathbf{Y}_l^{\text{pilot}} \in \mathbb{C}^{N \times \tau_p}$,

$$\mathbf{Y}_l^{\text{pilot}} = \sum_{i=1}^{K} \sqrt{\eta_i} \mathbf{h}_{il} \phi_{t_i}^T + \mathbf{N}_l \tag{3.39}$$

where $\mathbf{N}_l \in \mathbb{C}^{N \times \tau_p}$ is the receiver noise, whose elements are iid and distributed as $\mathcal{N}_{\mathbb{C}}(0, \sigma^2)$. The received UL signal $\mathbf{Y}_l^{\text{pilot}}$is the observation that AP l can use to estimate the channels to all of its served UEs. The estimation can either be performed locally at AP l, or be

delegated to the CPU via fronthaul link. Since the channels are independent from each other, the estimation occurring locally does not have any loss of quality when compared to the centralized estimation, in the CPU.

3.3.2 MMSE Channel Estimation

We want to estimate the channel $\mathbf{h}_{kl}$ based on the observation $\mathbf{Y}_l^{\text{pilot}}$, in Eq. (3.39). First, we should remove the interference from the UEs that use orthogonal pilots, by multiplying the received signal with the normalized conjugate of the pilot sequence ϕ_{t_k}, yielding $\mathbf{y}_{t_k l}^{\text{pilot}} = \mathbf{Y}_l^{\text{pilot}} \phi_{t_k}^* / \sqrt{\tau_p} \in \mathbb{C}^N$, which is given by

$$\begin{aligned} \mathbf{y}_{t_k l}^{\text{pilot}} &= \sum_{i=1}^{K} \frac{\sqrt{\eta_i}}{\sqrt{\tau_p}} \mathbf{h}_{il} \phi_{t_i}^T \phi_{t_k}^* + \frac{1}{\sqrt{\tau_p}} \mathbf{N}_l \phi_{t_k}^* \\ &= \underbrace{\sqrt{\eta_k \tau_p} \mathbf{h}_{kl}}_{\text{Desired channel}} + \underbrace{\sum_{i \in \mathcal{P}_k \ k} \sqrt{\eta_i \tau_p} \mathbf{h}_{il}}_{\text{Interference}} + \underbrace{\mathbf{n}_{t_k l}}_{\text{Noise}} \end{aligned} \tag{3.40}$$

The first term in Eq. (3.40) is the desired estimate, scaled by the square root of the total pilot power. The second term is the interference generated by the UEs sharing the same pilot sequence and the third term is the noise $\mathbf{n}_{t_k l} = \frac{1}{\sqrt{\tau_p}} \mathbf{N}_l \phi_{t_k}^* \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \sigma^2 \mathbf{I}_N)$. Notice that the received signal $\mathbf{y}_{t_k l}^{\text{pilot}}$ is a sufficient statistic for the estimation of $\mathbf{h}_{kl}$, as affirmed by [10], since the pilot signals used are mutually orthogonal. Thus, we are able to use the following result

COROLLARY 3.1 The MMSE estimate of $\mathbf{h}_{kl}$ based on $\mathbf{y}_{t_k l}^{\text{pilot}}$ is

$$\hat{\mathbf{h}}_{kl} = \sqrt{\eta_k \tau_p} \mathbf{R}_{kl} \boldsymbol{\Psi}_{t_k l}^{-1} \mathbf{y}_{t_k l}^{\text{pilot}}, \tag{3.41}$$

with

$$\boldsymbol{\Psi}_{t_k l} = \mathbb{E}\left\{ \mathbf{y}_{t_k l}^{\text{pilot}} (\mathbf{y}_{t_k l}^{\text{pilot}})^H \right\} = \sum_{i \in \mathcal{P}_k} \eta_i \tau_p \mathbf{R}_{il} + \sigma^2 \mathbf{I}_N, \tag{3.42}$$

the correlation matrix of the received signal. The estimated channel $\hat{\mathbf{h}}_{kl}$ and the estimation error $\tilde{\mathbf{h}}_{kl} = \mathbf{h}_{kl} - \hat{\mathbf{h}}_{kl}$ are independent and distributed as

$$\hat{\mathbf{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \eta_k \tau_p \mathbf{R}_{kl} \boldsymbol{\Psi}_{t_k l}^{-1} \mathbf{R}_{kl}) \tag{3.43}$$

$$\hat{\mathbf{h}}_{kl} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{C}_{kl}) \tag{3.44}$$

with the error correlation matrix

$$\mathbf{C}_{kl} = \mathbb{E}\{\tilde{\mathbf{h}}_{kl}\tilde{\mathbf{h}}_{kl}^{H}\} = \mathbf{R}_{kl} - \eta_k \tau_p \mathbf{R}_{kl} \boldsymbol{\Psi}_{t_k l}^{-1} \mathbf{R}_{kl}. \tag{3.45}$$

Proof 3.4: By taking a look at the received signal of Eq. (3.40), and defining $q = \sqrt{\eta_k \tau_p}$, $\mathbf{n} = \sum_{i \in \mathcal{P}_k \ k} \sqrt{\eta_i \tau_p}\mathbf{h}_{il} + \mathbf{n}_{t_k l}$, $\mathbf{R} = \mathbf{R}_{kl}$ and $\mathbf{S} = \sum_{i \in \mathcal{P}_k \ k} \eta_i \tau_p \mathbf{R}_{il} + \sigma^2 \mathbf{I}_N$. By invoking Lemma 3.1, the MMSE estimate and properties follow directly from it. □

The MMSE estimator has its name because it minimizes the MSE $\mathbb{E}\{||\mathbf{h}_{kl} - \hat{\mathbf{h}}_{kl}||^2\} = \mathbb{E}\{||\tilde{\mathbf{h}}_{kl}||^2\}$ among all possible estimators. This means that the use of non-linear estimators will not reduce the MSE even more. The computation of the channel estimate is then dependent on the knowledge of two matrices:

1. The spatial correlation matrix $\mathbf{R}_{kl}$
2. The matrix sum $\boldsymbol{\Psi}_{t_k l} = \sum_{i \in \mathcal{P}_k} \eta_i \tau_p \mathbf{R}_{il} + \sigma^2 \mathbf{I}_N$ of the correlation of the pilot sharing UE and the noise.

Both matrices depend on the channel statistics and are therefore constant. Thus, the matrix $\sqrt{\eta_k \tau_p}\mathbf{R}_{kl}\boldsymbol{\Psi}_{t_k l}^{-1}$ can be precomputed. This matrix is described by $N^2/2$ complex scalars, which can be exchanged via the fronthaul links in order to make this matrix available everywhere in the grid. The fronthaul signaling is negligible, since the channel statistics are deterministic through the transmission. If the aforementioned matrix is precomputed , the MMSE estimate requires computing $\mathbf{y}_{t_k l}^{\text{pilot}}$ and then multiplying it with $\sqrt{\eta_k \tau_p}\mathbf{R}_{kl}\boldsymbol{\Psi}_{t_k l}^{-1}$ of each UE served by the l-th AP. The inverse matrix operation requires $N\tau_p$ complex multiplications per pilot sequence while the matrix multiplication needs N^2 complex multiplications per UE, as shown in [22]. Thus, the computation complexity for channel estimation at the l-th AP is

$$|\mathcal{D}_l|(N\tau_p + N^2) \quad \text{complex multiplications} \tag{3.46}$$

per coherence block, considering that every UE that it serves use a different pilot sequence (which is desirable, since the interference from neighboring UEs that are served by AP l is canceled. Since this value stays finite for $K \to \infty$, considering $|\mathcal{D}|_l$ finite, then the scalability condition of Lemma 2.1 is satisfied.

3.3.3 Spatial Correlation Matrix Estimation

All of the aforementioned methods assume the fact that channel statistics are known everywhere. However, in practical systems, the knowledge of those parameters are often obtained from estimations. There are various ways of estimating the channel spatial correlation matrix, such as the ones seen in [31], [32], [33] and [34]. We will explore the most straightforward of those methods, consisting on taking samples of the estimated matrix over coherence blocks and averaging it. We begin by allocating pilots to the UEs, so that each pilot is used exactly K/τ_p times, i.e., this means that the number of pilots must be a multiple of the number of UEs. There is also a generalization for any τ_p, but this will not be covered here. Thus, each group of τ_P UEs, using orthogonal pilots, transmit to every AP in the grid. The received signal $\mathbf{Y}_l^{\text{estimate}} \in \mathbb{C}^{N \times \tau_p}$ is given by

$$\mathbf{Y}_l^{\text{estimate}} = \sum_{i \in \mathcal{T}_p} \sqrt{\eta_i} \mathbf{h}_{il} \phi_{t_k} + \mathbf{N}_l, \tag{3.47}$$

where $\mathcal{T}_p$ is the set of all of the transmitting UEs. Then the interference is canceled by performing the same operation of Eq. (3.40) as

$$\mathbf{h}_{kl}^{\text{LS}} = \sum_{i \in \mathcal{T}_p} \frac{\sqrt{\eta_i}}{\sqrt{\tau_p}} \mathbf{h}_{il} \phi_{t_i}^T \phi_{t_k}^* + \frac{1}{\sqrt{\tau_p}} \mathbf{N}_l \phi_{t_k}^* \tag{3.48}$$

where $\mathbf{h}_{kl}^{\text{LS}} \in \mathbb{C}^N$ is the least squares (LS) estimate of the channel between UE k and AP l. Since the pilots of the transmitting UEs are orthogonal. Eq. (3.48) can be simplified as

$$\mathbf{h}_{kl}^{\text{LS}} = \sqrt{\eta_k \tau_p} \mathbf{h}_{kl} + \mathbf{n}_{t_k l} \tag{3.49}$$

that is a scaled up version of the desired channel plus noise. Since the channel spatial correlation matrix is defined as $\mathbb{E}\{\mathbf{h}_{kl}\mathbf{h}_{kl}^H\}$, the estimated channel spatial correlation matrix, $\mathbf{R}_{kl}^{\text{sample}}$ can be expressed by

$$\mathbf{R}_{kl}^{\text{sample}} = \frac{1}{N_c N_t} \sum_{n=1}^{N_c} \sum_{n_t=1}^{N_t} \frac{\mathbf{h}_{kln_tn_c}^{\text{LS}}}{\sqrt{\eta_k \tau_p}} \frac{(\mathbf{h}_{kln_tn_c}^{\text{LS}})^H}{\sqrt{\eta_k \tau_p}} \tag{3.50}$$

where N_t is the total number of transmissions for each estimation of the matrix and N_c is the number of coherence blocks that are used for estimation.

The greater the number of transmissions, the better the spatial correlation matrix estimate. However, the greater the number of coherence blocks that will be used to estimate it, which in turn shrinks the available space for data transmissions.

3.3.4 Normalized MSE

The values of the MSE depends on the channel gains β_{kl}, so a strong channel might have larger errors in absolute values than a weak channel. To better measure the error in a given estimate of the channel, is best to use a normalized version of the MSE, the normalized mean squared error (NMSE). In respect to the channel between AP l and UE k, the NMSE is given by

$$\text{NMSE}_{kl} = \frac{\mathbb{E}\{||\mathbf{h}_{kl} - \hat{\mathbf{h}}_{kl}||^2\}}{\mathbb{E}\{||\mathbf{h}_{kl}||^2\}} = \frac{\text{tr}(\mathbf{C}_{kl})}{\text{tr}(\mathbf{R}_{kl})} = 1 - \frac{\eta_k \tau_p \text{tr}(\mathbf{R}_{kl} \boldsymbol{\Psi}_{t_k l}^{-1} \mathbf{R}_{kl})}{\text{tr}(\mathbf{R}_{kl})} \tag{3.51}$$

and it is a measure of the estimation error variance per antenna. When the estimate is perfect, the NMSE goes to 0. However, when the UL transmit power is $\eta_k = 0$, meaning that the received signal is composed of only interference and noise, the NMSE has its maximum value 1.

In the case of single antennas, the NMSE is obtained as

$$\begin{aligned}
\text{NMSE}_{kl} &= 1 - \frac{\eta_k \tau_p \left(\beta_{kl} \left(\sum\limits_{i \in \mathcal{P}_k} \eta_i \tau_p \beta_{il} + \sigma^2 \right)^{-1} \beta_{kl} \right)}{\beta_{kl}} \\
&= 1 - \frac{\eta_k \tau_p \beta_{kl}}{\eta_K \tau_p \beta_{kl} + \sum\limits_{i \in \mathcal{P}_k \ \{k\}} \eta_i \tau_p \beta_{il} + \sigma^2} \\
&= 1 - \frac{\text{SNR}_{kl}^{\text{pilot}}}{\text{SNR}_{kl}^{\text{pilot}} + \sum\limits_{i \in \mathcal{P}_k \ \{k\}} \text{SNR}_{il}^{\text{pilot}} + 1},
\end{aligned} \tag{3.52}$$

where

$$\text{SNR}_{kl}^{\text{pilot}} = \frac{\eta_K \tau_p \beta_{kl}}{\sigma^2} \tag{3.53}$$

is the *effective* SNR of the pilot transmission from UE k to AP l. Notice that the pilot sequence length τ_p increases this value, given that during the transmission, the receiver captures the transmitted pilot energy without necessarily increasing the noise. Since the total pilot energy, $\tau_p \eta_k$, is dependent on the pilot sequence length and the UL transmit power, it is possible to choose between concentrating all the energy in one of the τ_p samples, or spreading it between all of the samples. The second solution seems more plausible, since it keeps the peak to average power into a realizable range. However, this solution is highly sensitive to hardware imperfections, that might lead to approximately orthogonal pilot sequences being generated from the UE, when it is only able to generate its own pilot sequence. Thus, in real life applications, the system can improve its effective SNR of the pilot transmission by increasing the pilot sequence length and by boosting the pilot power into specific samples.

Also, we see the effect of the interference of the pilot-sharing UEs into the value of the NMSE. This interference will be discussed in the next section.

3.3.5 Pilot Contamination

From Eq. (3.52), we can see that the interference from pilot sharing UEs increases the value of the NMSE, which in turn reduces the channel estimation quality. Assuming that uncorrelated fading, $\mathbf{R}_{kl} = \beta_{kl}\mathbf{I}_N$, we can see that the NMSE has the same expression as Eq. (3.52). This means that the number of antennas, in the case of uncorrelated fading, does not interfere in the quality of channel estimation. On the case of correlated fading, however, it is not as straightforward. For instance, if the interference from the pilot-sharing UEs is increased, it leads to an increase of the NMSE. Also, it depends clearly on the relationship between the spatial correlation matrices $\mathbf{R}_{il}$ and $\mathbf{R}_{kl}$. When the correlation between their vectors is weak, meaning that their UEs channels have widely different multipath components directions, then their corresponding dominating eigenspaces will be widely different, which will somehow increase the quality of channel estimation, as will be confirmed later.

The interference of pilot sharing UEs is called *pilot contamination*. It is different from the interference generated from receiver noise, since besides reducing the channel estimate quality, it also correlates the desired transmitted information with the interference of the pilot-sharing UEs. To see that, let the estimate of $\mathbf{h}_k l$, $\hat{\mathbf{h}}_{kl}$, be the MMSE estimate as

$$\hat{\mathbf{h}}_{kl} = \sqrt{\eta_k \tau_k}\mathbf{R}_{kl}\boldsymbol{\Psi}_{t_k l}^{-1}\mathbf{y}_{t_k l}^{\text{pilot}} \tag{3.54}$$

and the estimate of another UE $i \in \mathcal{P}_k \setminus \{k\}$ using the same pilot as UE k. In that situation, we have that $\boldsymbol{\Psi}_{t_k l} = \boldsymbol{\Psi}_{t_i l}$ and $\mathbf{y}_{t_k l}^{\text{pilot}} = \mathbf{y}_{t_i l}^{\text{pilot}}$, such that Eq. (3.54) can be rewritten as

$$\hat{\mathbf{h}}_{il} = \sqrt{\eta_i \tau_i}\mathbf{R}_{il}\boldsymbol{\Psi}_{t_k l}^{-1}\mathbf{y}_{t_k l}^{\text{pilot}} \tag{3.55}$$

and, by comparing Eq. (3.54) and Eq. (3.55), we have that

$$\hat{\mathbf{h}}_{il} = \sqrt{\frac{\eta_i}{\eta_k}}\mathbf{R}_{il}\mathbf{R}_{kl}^{-1}\hat{\mathbf{h}}_{kl} \quad i \in \mathcal{P}_k. \tag{3.56}$$

Thus, we can see that the two estimates are correlated, even though the two channels are considered statistically independent ($\mathbb{E}\{\mathbf{h}_{kl}\mathbf{h}_{il}^H\} = \mathbf{0}_{N\times N}$, for $i \neq k$). In fact, at the case of uncorrelated fading, the relationship between the estimates will be given by $\hat{\mathbf{h}}_{il} = \sqrt{\frac{\eta_i \beta_i^2}{\eta_k \beta_k^2}}\hat{\mathbf{h}}_{kl}$ which makes $\hat{\mathbf{h}}_{kl}$ a scaled version of $\hat{\mathbf{h}}_{il}$, and thus the estimates are fully correlated. This shows that the reuse of pilots makes the channel estimates correlated, reducing the quality of the estimates. However, pilot contamination is not only detrimental to channel estimation: it also reduces the capability of the combiner to mitigate interference, as will be seen at subsequent chapters.

3.3.6 MMSE Estimation of the Collective Channel

To this moment, we have only estimate the channels between AP l and UE k. However, in order to coherently process the signals, the system must have knowledge of the collective channel $\mathbf{h}_k$ between the UE k to all of the APs in the grid. In the case of the estimated channels, it is not mandatory for the system to have the estimates of UE k to all APs. Instead, it is only necessary for it to have the estimates between UE k and the subset of APs that serve it, that is, the APs in $\mathcal{M}_k \in \{1, ..., L\}$. This means that, following the collective channel properties of Section 2.3, the following partial channel is known

$$\mathbf{D}_k\hat{\mathbf{h}}_k \triangleq \begin{bmatrix} \mathbf{D}_{k1}\hat{\mathbf{h}}_{k1} \\ \vdots \\ \mathbf{D}_{kL}\hat{\mathbf{h}}_{kL} \end{bmatrix} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_M, \eta_k\tau_p\mathbf{D}_k\mathbf{R}_k\boldsymbol{\Psi}_{t_k}^{-1}\mathbf{R}_k\mathbf{D}_k), \tag{3.57}$$

where $\boldsymbol{\Psi}_{t_k}^{-1} = \text{diag}\left(\boldsymbol{\Psi}_{t_k1}^{-1}, \ldots, \boldsymbol{\Psi}_{t_kL}^{-1}\right)$ and the estimation error is $\mathbf{D}_k\tilde{\mathbf{h}}_k = \mathbf{D}_k\mathbf{h}_k - \mathbf{D}_k\hat{\mathbf{h}}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_M, \mathbf{D}_k\mathbf{C}_k)$ in which $\mathbf{C}_k = \text{diag}\left(\mathbf{C}_{k1}, \ldots, \mathbf{C}_{kL}\right)$. The collective NMSE of UE k is then computed as

$$\begin{aligned} \text{NMSE}_k &= \frac{\mathbb{E}\left\{||\mathbf{D}_k\mathbf{h}_k - \mathbf{D}_k\hat{\mathbf{h}}_k||^2\right\}}{\mathbb{E}\left\{||\mathbf{D}_k\mathbf{h}_k||^2\right\}} = \frac{\text{tr}(\mathbf{D}_k\mathbf{C}_k)}{\text{tr}(\mathbf{D}_k\mathbf{R}_k)} = \frac{\sum_{l=1}^{L}\text{tr}(\mathbf{D}_{kl}\mathbf{C}_{kl})}{\sum_{l=1}^{L}\text{tr}(\mathbf{D}_{kl}\mathbf{R}_{kl})} \\ &= 1 - \frac{\eta_k\tau_p\sum_{l\in\mathcal{M}_k}\text{tr}(\mathbf{R}_{kl}\boldsymbol{\Psi}_{t_kl}^{-1}\mathbf{R}_{kl})}{\sum_{l\in\mathcal{M}_k}\text{tr}(\mathbf{R}_{kl})}. \end{aligned} \tag{3.58}$$

If the APs are equipped with only one antenna, then the expression of Eq. (3.58) is written as

$$\begin{aligned} \text{NMSE}_k &= 1 - \frac{\eta_K\tau_p\sum_{l\in\mathcal{M}_k}\frac{\beta_{kl}^2}{\eta_k\tau_p\beta_{kl} + \sum_{i\in\mathcal{P}_k\setminus\{k\}}\eta_i\tau_p\beta_{il} + \sigma^2}}{\sum_{l\in\mathcal{M}_k}\beta_{kl}} \\ &= 1 - \frac{\sum_{l\in\mathcal{M}_k}\frac{(\text{SNR}_{kl}^{\text{pilot}})^2}{\text{SNR}_{kl}^{\text{pilot}} + \sum_{i\in\mathcal{P}_k\setminus\{k\}}\text{SNR}_{il}^{\text{pilot}} + 1}}{\sum_{l\in\mathcal{M}_k}\text{SNR}_{kl}^{\text{pilot}}}, \end{aligned} \tag{3.59}$$

where it is easily seen that the presence of pilot sharing UEs will establish a correlation between the channel estimates, which in turn, will increase the NMSE.

3.4 IMPACT OF SPATIAL CORRELATION, CONTAMINATION AND ARCHITECTURE

This section aims to numerically exemplify the affects of the architecture, pilot contamination and spatial correlation on channel estimation accuracy.

3.4.1 Impact of Cell Free Architecture

To understand the basics of the impact of the CF architecture into channel estimation, we begin by analyzing single antenna APs and the estimation of the collective channel of an arbitrary UE k that is assigned one unique pilot sequence $\mathcal{P}_k = \{k\}$. In this situation, the expression for the NMSE on Eq. (3.59) is reduced to

$$\text{NMSE}_k = 1 - \frac{\sum\limits_{l \in \mathcal{M}_k} \frac{\text{SNR}_{kl}^{\text{pilot}}}{1+\frac{1}{\text{SNR}_{kl}^{\text{pilot}}}}}{\sum\limits_{l \in \mathcal{M}_k} \text{SNR}_{kl}^{\text{pilot}}}. \tag{3.60}$$

If one AP has a much larger SNR than the other APs in $\mathcal{M}_k$, then Eq. (3.60) is reduced to the approximation

$$\text{NMSE}_k \approx 1 - \frac{1}{1+\frac{1}{\text{SNR}_{kl_{\text{max}}}^{\text{pilot}}}} = \frac{1}{\text{SNR}_{kl_{\text{max}}}^{\text{pilot}}+1}, \tag{3.61}$$

with l_{max} being the AP that has the highest effective SNR to the UE k among all of the elements composing $\mathcal{M}_k$. This approximation is exact if $|\mathcal{M}_k| = 1$, or if all the serving APs have the same effective SNR. However, in the case the APs have different values of effective SNR, the NMSE values will be higher than the ones in Eq. (3.60).

To show this, we propose a simulation setup with one UE k, that may be served by $\mathcal{M}_k = 2$ APs or just one of them. We let the value of $\text{SNR}_{k1}^{\text{pilot}}$ be fixed to 20 dB, and the value of $\text{SNR}_{k2}^{\text{pilot}}$ vary from -20 dB to 40 dB. The results are shown in Fig. 3.1 It is possible to see that the NMSE with the two APs serving UE k is higher than the case of only AP 1 serving the UE if $\text{SNR}_{k2}^{\text{pilot}} < \text{SNR}_{k1}^{\text{pilot}}$. However, a lower NMSE is obtained when $\text{SNR}_{k2}^{\text{pilot}} > \text{SNR}_{k1}^{\text{pilot}}$. This shows that having two serving APs does not necessarily improve the quality of the collective channel estimation. To a satisfactory estimate one must ensure also that the effective SNR of both APs is high. This can be seen in the case that for low values of $\text{SNR}_{k2}^{\text{pilot}}$, the NMSE increases to more than one serving AP to the point it starts decreasing, because the effective SNR of AP 2 is high enough.

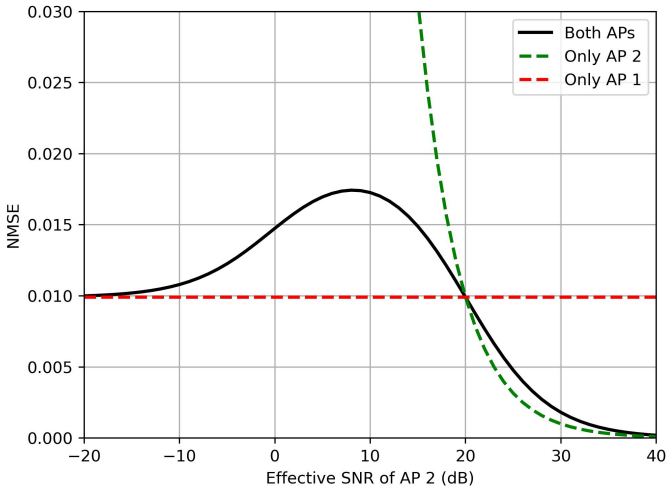


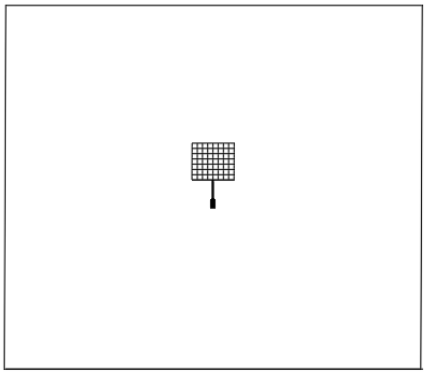
Figure 3.1 – NMSE when at most two APs serve UE k. The value of $\text{SNR}_{k1}^{\text{pilot}}$ if fixed to 20 dB, while the value of $\text{SNR}_{k2}^{\text{pilot}}$ ranges from -20 to 40 dB.

To further explore the influence of the architecture on the NMSE, we propose another simulation scheme, with a square grid of 600 m $\times 600$ m, with $L = 81$ single antenna APs across the grid. The UE is placed randomly in the grid. The bandwidth is 20 MHz, and the noise power is $\sigma^2 = -93$ dBm. Fig. 3.2 shows the simulated setup. The channels are Rayleigh fading, and the channel gains are given by

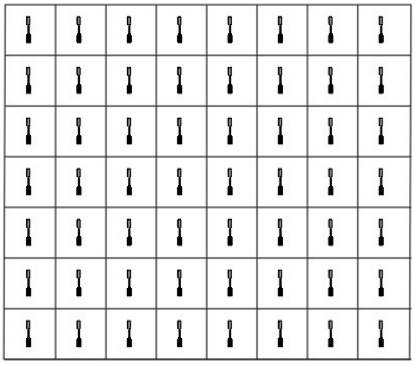
$$\beta_{kl}(d_{kl})[\text{dB}] = -30.5 - 36.7 \log_{10}\left(\frac{d_{kl}}{1 \text{ m}}\right), \tag{3.62}$$

where the propagation distances d_{kl} are computed assuming the APs are 10 m above the UEs. In the case of the Massive MIMO, the channel is considered uncorrelated. Fig. 3.3 shows the NMSE curves for the three described setups. The pilot transmit power η_k varies from 1 mW to 1 W, the pilot sequence length is $\tau_p = 10$. When we compare all of the setups, we can see that the average NMSE is always greater in the Massive MIMO case. This shows that, for channel estimation quality, its always more beneficial to place more APs in the grid than to increase the number of antennas of a single AP. In the case of the small cell setup, the NMSE is lower than the CF, given the chosen AP from the UE is always the one that has the best channel quality, which in turn will have the lowest NMSE. However, one should not confuse the fact that lower channel estimation quality implies lower communication performance in this case. Since the CF uses the combined information of the APs to jointly decode the desired signal, its communication performance is generally higher than the small cell setup, even though channel estimation quality is lower.

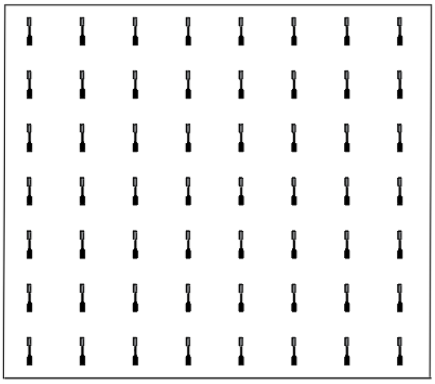
The SE gain of CF systems will be discussed later in this work. However, we will give an indication of the performance gain by evaluating the SNR that is achieved during data



(a) Massive MIMO setup



(b) Small cell setup



(c) Cell free setup

Figure 3.2 – Simulation setups. (a) shows the Massive MIMO setup, (b) the small cell setup, and (c) the CF setup.

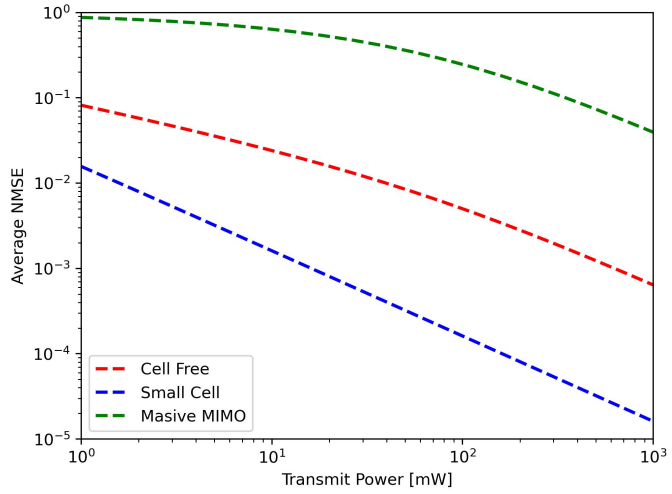


Figure 3.3 – Average NMSE for a single UE in the three setups described in Fig. 3.2

detection. Supposing that only UE k is active in the network, then the local estimate of the transmitted symbol s_k, as defined in Eq. (2.7) is given by

$$\mathbf{s}_{kl} = \underbrace{\mathbf{v}_{kl}^H \hat{\mathbf{h}}_{kl} s_k}_{\text{Signal over the estimated channel}} + \underbrace{\mathbf{v}_{kl}^H \tilde{\mathbf{h}}_{kl} s_k}_{\text{Signal over unknown channel}} + \underbrace{\mathbf{v}_{kl}^H \mathbf{n}_l}_{\text{Noise}}. \tag{3.63}$$

This is the estimated signal at AP l. All of the estimates of the serving APs are sent to the CPU, where they are combined and a final estimate of s_k is obtained:

$$\hat{s}_k = \sum_{l \in \mathcal{M}_k} \mathbf{v}_{kl}^H \hat{\mathbf{h}}_{kl} s_k + \sum_{l \in \mathcal{M}_k} \mathbf{v}_{kl}^H \tilde{\mathbf{h}}_{kl} s_k + \sum_{l \in \mathcal{M}_k} \mathbf{v}_{kl}^H \mathbf{n}_l. \tag{3.64}$$

We will assume, for the sake of simplicity, that MR combining is applied, with $\mathbf{v}_{kl} = \hat{\mathbf{h}}_{kl}$, and compute the resulting SNR as

$$\begin{aligned} \text{SNR}_k^{\text{data}} &= \frac{\mathbb{E}\left\{ \left| \sum_{l \in \mathcal{M}_k} \mathbf{v}_{kl}^H \hat{\mathbf{h}}_{kl} s_k \right|^2 \right\}}{\mathbb{E}\left\{ \left| \sum_{l \in \mathcal{M}_k} \mathbf{v}_{kl}^H \mathbf{n}_l \right|^2 \right\}} = \frac{p_k \mathbb{E}\left\{ \left| \sum_{l \in \mathcal{M}_k} \left\| \hat{\mathbf{h}}_{kl} \right\|^2 \right|^2 \right\}}{\sigma^2 \sum_{l \in \mathcal{M}_k} \mathbb{E}\left\{ \left\| \hat{\mathbf{h}}_{kl} \right\|^2 \right\}} \\ &= \frac{\sum_{l \in \mathcal{M}_k} p_k \operatorname{tr}\left((\mathbf{R}_{kl} - \mathbf{C}_{kl})^2 \right)}{\sum_{l \in \mathcal{M}_k} \sigma^2 \operatorname{tr}\left(\mathbf{R}_{kl} - \mathbf{C}_{kl} \right)} + \sum_{l \in \mathcal{M}_k} \frac{p_k \operatorname{tr}\left(\mathbf{R}_{kl} - \mathbf{C}_{kl} \right)}{\sigma^2}, \end{aligned} \tag{3.65}$$

where the expectations are with respect to the channel estimate. By using the results of Eq. (3.65), we propose the same simulation setup as the one that was used to generate Fig. 3.3. We let $p_k = \eta_k = 20$ mW. Fig. 3.4 shows the SNR distribution for random UE locations. We

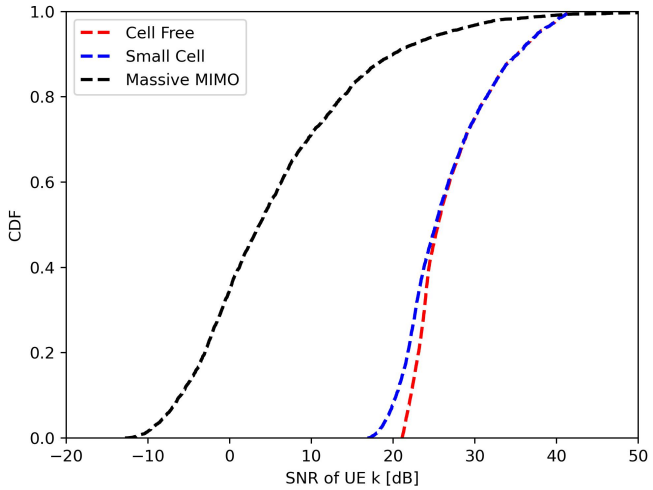


Figure 3.4 – CDF of the SNR achieved in the UL with MR combining and MMSE channel estimation in the same setups as Fig. 3.3

.

can see from it that the CF system achieves the larger values of SNR most of the time. The Massive MIMO setup has the lowest values of the SNR, since it can't exploit the advantage of having APs at different geographical locations. The setup of Small Cell has a performance similar to the CF setup for large values of SNR. This happens due to the fact that for those values, the jointly processing of the APs with weaker channels is overcompensated by the processing of stronger channels in CF. Thus, the CF has the largest gains since it is able to coherently process the information at multiple APs. Also, we will see in later sections that in the presence of multiple UEs it achieves an even better performance, by jointly mitigating the interference from the UEs.

3.4.2 Impact of Pilot Contamination

The next point of interest is to show the effect of pilot contamination for channel estimation. For this, we assume the scenery of a single-antenna APs l, and that it wants to estimate the channel of UE 1, with UE 2 using the same pilot sequence, and acting as interference. From Eq. (3.59), we obtain

$$\text{NMSE}_{k1} = 1 - \frac{\text{SNR}_{1l}^{\text{pilot}}}{\text{SNR}_{1l}^{\text{pilot}} + \text{SNR}_{2l}^{\text{pilot}} + 1} \tag{3.66}$$

The increasing value of pilot contamination on Eq. (3.66) depends on the term $\text{SNR}_{2l}^{\text{pilot}} + 1$, which are the interference and noise. When this term is close to 1, the pilot contamination is small, whereas values much greater than 1 imply a high level of pilot contamination. To show that, we consider 5 values of $\text{SNR}_{1l}^{\text{pilot}} \in \{-20, -10, 0, 10, 20\}$ dB, and the values of $\text{SNR}_{2l}^{\text{pilot}}$

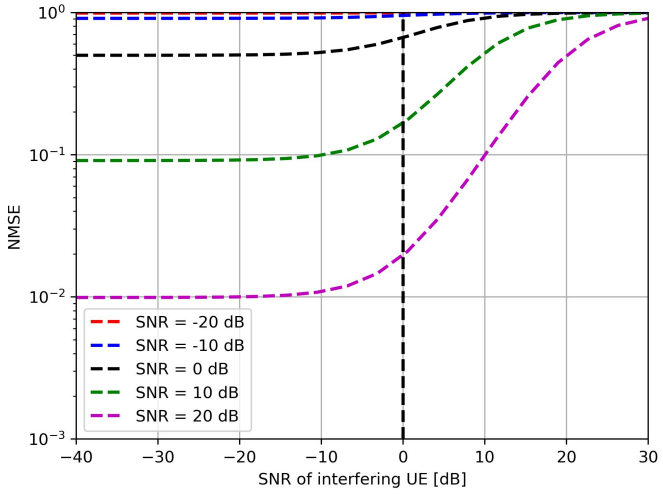


Figure 3.5 – NMSE as function of the varying SNR, $\text{SNR}_{2l}^{\text{pilot}}$ of the interfering pilot signal

ranging from -40 to 30 dB. Fig. 3.5 shows that the NMSE increases when the effective SNR of the interfering UE increases, specially when the effective SNR of the desired UE is high. For instance, when $\text{SNR}_{1l}^{\text{pilot}} = 20$ dB, the NMSE increases drastically when the SNR of the interfering signal reaches -10 dB. From that, we can see that, if the interfering signal is 10 dB weaker than noise, then pilot contamination is negligible. Thus, we can affirm that pilot contamination is an aggravating factor when the interfering UEs are near the desired UE. Since those UEs are probably being served by the same AP, it would be logical to assign different pilot sequences to the UEs being served by the same AP, as proposed by [8]. This is very similar to cellular systems, where UEs that are in the same cell are assigned preferably orthogonal pilot sequences, to reduce pilot contamination. Since the main factor on determining the SNR from the interfering signal is the distance between the AP and the interfering UE, we propose a simulation scheme, similar to the one that was used to generate Fig. 3.4, where the effective SNR, from Eq. (3.53), is evaluated for different distances between the interfering UE and its serving AP, and the pilot sequence length takes values of $\tau_p \in \{1, 5, 10\}$. From Fig. 3.6 we can see that for shorter distances, the pilot contamination is significantly large, given that the values of the effective SNR of the interfering UE are large. Also, for longer pilot sequences, the effect of pilot contamination is ever present for longer distances. We have seen that above the -10 dB line, the pilot contamination is neglectful. This means that at distance of 215 m, for $\tau_p = 1$, 330 m, for $\tau_p = 5$, and 400 m for $\tau_p = 10$, the pilot contamination is neglectful, which means that in this setup, if we have a frequency reuse factor of 1 ($\tau_p = 1$), the interfering UEs should be at least 400 m for the AP in order to have neglectful pilot contamination. Thus, the increasing values of the pilot sequence makes the pilot contamination effect much more manageable. Therefore, one could argue that increasing τ_p would always be beneficial, since the interference from pilot contamination is mitigated. However, we will see in later sections that this is not always the

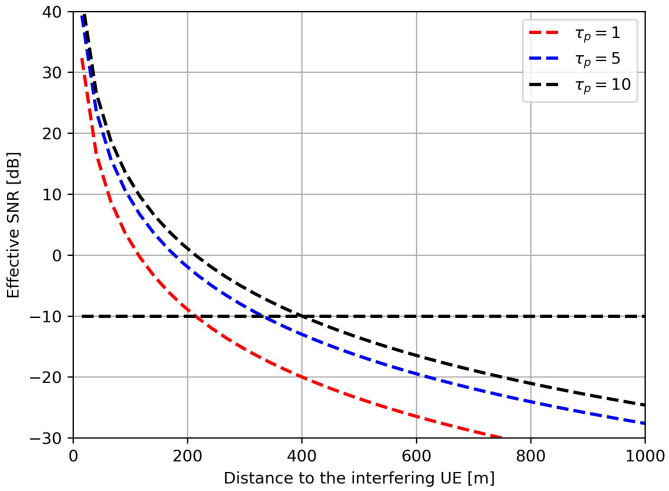


Figure 3.6 – The effective SNR of the interfering UE, from Eq. (3.53) as function of the distance between the interfering UE and the AP. The pilot sequence length is such that $\tau_p \in \{1, 5, 10\}$.

ideal case, since the complexity of signal processing at the AP grows significantly, and the SE gains are negligible.

3.4.3 Impact of Spatial Correlation

This section aims to quantify the effect of spatial correlation on channel estimation of the AP antennas, and on pilot contamination.

3.4.3.1 Impact of Spatial Correlation in Channel Estimation

Let us consider the spatially correlated channel $\mathbf{h} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_N, \mathbf{R})$ between an AP and an UE. We assume the UE uses a unique pilot sequence, so that Eq. (3.58) can be written as

$$\text{NMSE} = 1 - \frac{\eta\tau_p \,\text{tr}\left(\mathbf{R}\left(\eta\tau_p\mathbf{R} + \sigma^2\mathbf{I}_N\right)^{-1}\mathbf{R}\right)}{\text{tr}\left(\mathbf{R}\right)} \tag{3.67}$$

Now, we let $\mathbf{R} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$ be the eigenvalue decomposition of the spatial correlation matrix $\mathbf{R}$, in which the unitary columns of the unitary matrix $\mathbf{V} \in \mathbb{C}^{N\times N}$ are the eigenvectors, and the diagonal matrix $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, ..., \lambda_N)$ has the correspondent non-negative eigenvalues, with $\sum_{n=1}^{N} \lambda_n = \text{tr}\left(\mathbf{R}\right) = N\beta$. By using the eigenvalue decomposition, Eq. (3.67) can be

rewritten as

$$\begin{aligned}\text{NMSE} &= 1 - \frac{\eta\tau_p \,\text{tr}\left(\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H \left(\eta\tau_p \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H + \sigma^2 \mathbf{I}_N\right)^{-1} \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H\right)}{N\beta} \\ &= 1 - \frac{\eta\tau_p}{N\beta}\,\text{tr}\left(\boldsymbol{\Lambda}\left(\eta\tau_p\boldsymbol{\Lambda} + \sigma^2\mathbf{I}_N\right)^{-1}\boldsymbol{\Lambda}\right) \\ &= 1 - \frac{\eta\tau_p}{N\beta}\sum_{n=1}^{N}\frac{\lambda_n^2}{\eta\tau_p\lambda_n + \sigma^2} \\ &= 1 - \frac{1}{N\beta}\sum_{n=1}^{N}\frac{\text{SNR}^{\text{pilot}}\lambda_n^2}{\text{SNR}^{\text{pilot}}\lambda_n + \beta}\end{aligned} \tag{3.68}$$

Now, by defining Lemma 3.4

LEMMA 3.4 For matrices $\mathbf{A} \in \mathbb{C}^{N_1 \times N_2}$ and $\mathbf{B} \in \mathbb{C}^{N_2 \times N_1}$, it is true that

$$(\mathbf{I}_{N_1} + \mathbf{AB})^{-1}\mathbf{A} = \mathbf{A}\,(\mathbf{I}_{N_2} + \mathbf{BA})^{-1} \tag{3.69}$$

$$\text{tr}\,(\mathbf{AB}) = \text{tr}\,(\mathbf{BA})\,. \tag{3.70}$$

We see that the second equality of Eq. (3.68) is obtained from moving the matrices $\mathbf{V}$ and $\mathbf{V}^H$ into the inverse and applying Lemma 3.4. We can see from this equation that the NMSE depends on the eigenvalues, but not on the eigenvectors. Although we cannot change the spatial correlation properties of a practical channel, we can understand how it impacts the NMSE, and by that, try to find a eigenvalue distribution that maximizes or minimizes the NMSE under the constraint $\sum_{n=1}^{N}\lambda_n = N\beta$. For that, we announce Lemma 3.5:

LEMMA 3.5 The function NMSE : $\mathbb{R}_{\geq 0}^{N} \rightarrow \mathbb{R}_{\geq 0}$ of the eigenvalue vector $\lambda = [\lambda_1, ..., \lambda_N]^T$ given in Eq. (3.68) is strictly concave on its domain

Proof 3.5: The proof can be found in Appendix A.1. □

A consequence of Lemma 3.5 is that the maximum value of the NMSE expression of Eq. (3.68), under the constraint $\sum_{n=1}^{N}\lambda_n = N\beta$, is achieved when all the eigenvalues are equal and is an unique solution, $\lambda_n = \beta$, for $n = 1, ..., N$. This shows that the uncorrelated model of the spatial correlation matrix is the one that has the highest NMSE. Now, we focus on finding the spatial correlation matrix that would instead minimize the NMSE expression. For that, we make use of Lemma 3.6:

LEMMA 3.6 Consider the eigenvalues of the spatial correlation matrix $\mathbf{R}$ sorted in descending order

$$\lambda_1 \geq \lambda_2 \geq ...\lambda_r > \lambda_{r+1} = ... = \lambda_N = 0, \tag{3.71}$$

where $r \leq N$ is the rank of $\mathbf{R}$. Let $\lambda = [\lambda_1 ... \lambda_{r-2} \lambda_{r-1} \lambda_r 0 ... 0]^T$ be the eigenvalue vector. Then NMSE$(\lambda) >$ NMSE(λ'), where $\lambda' = [\lambda_1 ... \lambda_{r-2} \lambda_{r-1} + \lambda_r 0 ... 0]^T$.

Proof 3.6: The proof can be found in Appendix A.2. □

Lemma 3.6 shows that is possible to reduce the NMSE by replacing the two smallest eigenvalues of $\mathbf{R}$, λ_{r-1} and λ_r with the sum $\lambda_{r-1} + \lambda_r$ and 0. If we repeat this procedure a sufficient number of times, we will reach the case where $\lambda_1 = N\beta$, and $\lambda_n = 0$, for $n \geq 2$, which has the lowest NMSE. This case corresponds to a rank-one spatial correlation matrix, which is the case of strongest correlation between the channels of the antennas at the AP. Thus, we can conclude that, the bigger the spatial correlation between the antennas, the easier is to estimate the channel, since there is more known data to be provided to the estimator.

Those properties are seen in Fig. 3.7, where the local scattering model spatial correlation matrix, of Eq. (2.14), is used to estimate the NMSE as function of the azimuth ASD σ_φ. The results are averaged over different nominal azimuth angles, and the elevation angle is fixed at $\theta = -10^o$ and the elevation ASD is such that $\sigma_\theta \in \{0^o, 10^o, 20^o\}$. The effective SNR is of 20 dB, and there are $N = 10$ antennas at the AP. We can see from Fig. 3.7 that the NMSE is smaller as the ASD reduces (that is, when the channel is highly correlated). As the ASD increases, the NMSE also increases to the point where it converges to the uncorrelated case, when all eigenvalues from the spatial correlation matrix are equal, β, as opposed to the fully correlated case of $\sigma_\varphi = \sigma_\theta = 0^o$, when there will be only one nonzero eigenvalue of value $N\beta$. Also, we can see that ahead of the ASD $\sigma_\varphi = 40^o$ the spatial characteristics of the channel cannot be exppolited to generate a better NMSE. Thus, as we can see, fully correlated channels leads to better channel estimation quality, which would contradicy conventional MIMO literature, like [19] and [20]. Next, we analyze the effect of the number of AP antennas N on the NMSE. For that, we consider the same setup as the one that generated Fig. 3.7, but we fix the ASDs to be $\sigma_\varphi = \sigma_\theta = 15^o$, and the effective SNR to be such that SNR $\in \{-10, 0, 10, 20\}$ dB. We can see from Fig. 3.8 that the NMSE decreases as the number of AP antennas increases, since the channel between adjacent antennas become strongly correlated. However, there is a point around $N = 12$, where the NMSE stays the same, irrespective of the number of antennas. This happens due to the fact that the ULA

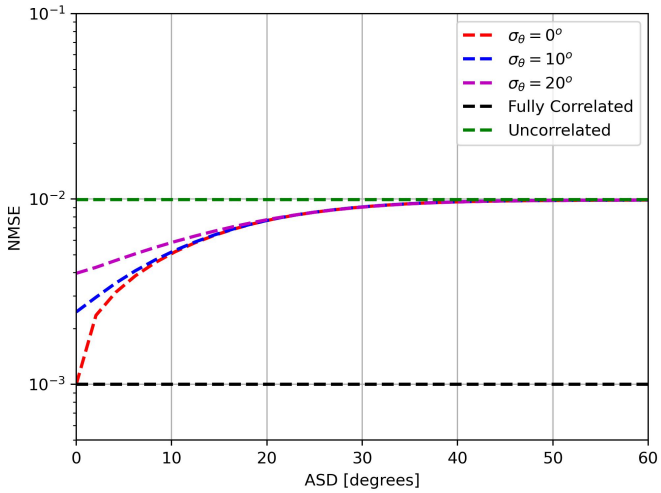


Figure 3.7 – NMSE in the estimation of an arbitrary channel as function of the azimuth ASD. The local scattering model of Eq. (2.14) is used, with Gaussian angular distribution. The results are averaged over all azimuth nominal angles, and the elevation angle is $\theta = -10^o$. The effective SNR is 20 dB and each AP is equipped with $N = 10$ antennas.

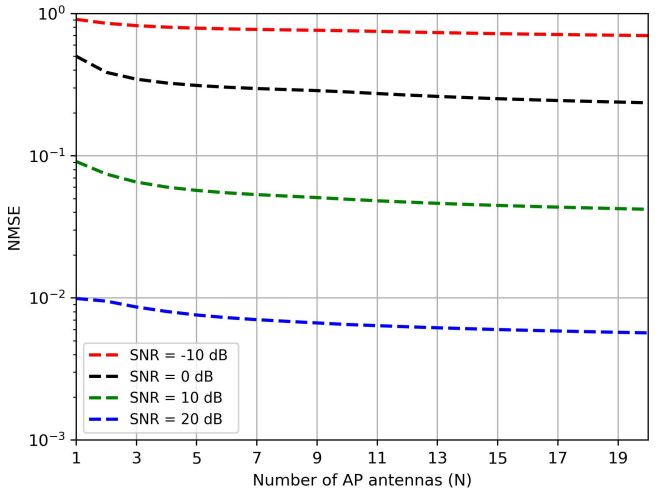


Figure 3.8 – NMSE in the estimation of an arbitrary channel as function of the number of antennas at the AP. The effective SNR is -10, 0, 10 and 20 dB, and the spatial correlation is like Fig. 3.8 with $\sigma_\varphi = \sigma_\theta = 15^o$.

gets so big that the channel of the antennas at its beginning and its end become uncorrelated, irrespective of the effective SNR value.

3.4.3.2 Impact of Spatial Correlation in Pilot Contamination

Next, we shall analyze the effect of spatial correlation on pilot contamination. For that, we consider UE 1 and UE 2, who share the same pilot sequence, and one arbitrary AP. The NMSE of Eq. (3.51) is given by

$$\text{NMSE}_1 = 1 - \frac{\eta\tau_p \,\text{tr}\left(\mathbf{R}_1 \left(\eta_1\tau_p\mathbf{R}_1 + \eta_2\tau_p\mathbf{R}_2 + \sigma^2\mathbf{I}_N\right)^{-1}\mathbf{R}_1\right)}{\text{tr}\left(\mathbf{R}_1\right)} \tag{3.72}$$

Differently from the single-antenna case, the NMSE depends not only on the effective SNR of Eq. (3.53), but also on the full spatial correlation matrices $\mathbf{R}_1$ and $\mathbf{R}_2$. We can also state Lemma 3.7 as

LEMMA 3.7 For the positive semi-definite matrices $\mathbf{A} \in \mathbb{C}^{N\times N}$, $\mathbf{C} \in \mathbb{C}^{N\times N}$, and positive-definite matrix $\mathbf{B} \in \mathbb{C}^{N\times N}$, the following inequality is true:

$$\text{tr}\left(\mathbf{A}\left(\mathbf{B}+\mathbf{C}\right)^{-1}\right) \leq \text{tr}\left(\mathbf{A}\mathbf{B}^{-1}\right), \tag{3.73}$$

where the equality holds only when $\mathbf{C}\mathbf{B}^{-1}\mathbf{A} = \mathbf{0}_{N\times N}$

Proof 3.7: Performing the eigenvalue decomposition of $\mathbf{C} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H = \mathbf{V}_1\boldsymbol{\Lambda}_1\mathbf{V}_1^H$, where $\mathbf{U} \in \mathbb{C}^{N\times N}$ is the unitary matrix of eigenvectors and $\boldsymbol{\Lambda} = \text{diag}\left(\lambda_1, ..., \lambda_N\right)$ with the eigenvalues $\lambda_N \geq 0$, for $n = 1, ..., N$. $\mathbf{V_1} \in \mathbb{C}^{N\times r}$ and $\boldsymbol{\Lambda_1} \in \mathbb{C}^{r\times r}$ are the partitions of $\mathbf{V}$ and $\boldsymbol{\Lambda}$ corresponding to the positive eigenvalues. Then we can state, by applying the matrix inversion properties to the inverse of $\mathbf{B} + \mathbf{V_1}\boldsymbol{\Lambda_1}\mathbf{V_1}^H$, that

$$\begin{aligned}&\text{tr}\left(\mathbf{A}\left(\mathbf{B}+\mathbf{C}^{-1}\right)\right)\\ &= \text{tr}\left(\mathbf{A}\mathbf{B}^{-1}\right) - \text{tr}\left(\mathbf{A}\mathbf{B}^{-1}\mathbf{V}_1\left(\mathbf{V}_1^H\mathbf{B}^{-1}\mathbf{V}_1 + \boldsymbol{\Lambda}_1^{-1}\right)^{-1}\mathbf{V}_1^H\mathbf{B}^{-1}\right)\end{aligned} \tag{3.74}$$

that is necessarily less than $\text{tr}\left(\mathbf{A}\mathbf{B}^{-1}\right)$ if $\mathbf{V}^H\mathbf{B}^{-1}\mathbf{A}\mathbf{B}^{-1}\mathbf{V}_1$ is nonzero, noting that the matrix $\left(\mathbf{V}_1^H\mathbf{B}^{-1}\mathbf{V}_1 + \boldsymbol{\Lambda}_1^{-1}\right)^{-1}$ is positive definite. If $\mathbf{V}^H\mathbf{B}^{-1}\mathbf{A}\mathbf{B}^{-1}\mathbf{V}_1 = \mathbf{0}_{r\times r}$, or in other words, $\mathbf{C}\mathbf{B}^{-1}\mathbf{A} = \mathbf{0}_{N\times N}$, then both sides on Eq. (3.73) are equal. □

Then, from Lemma 3.7, we can say that the NMSE expression of Eq. (3.72) is lower

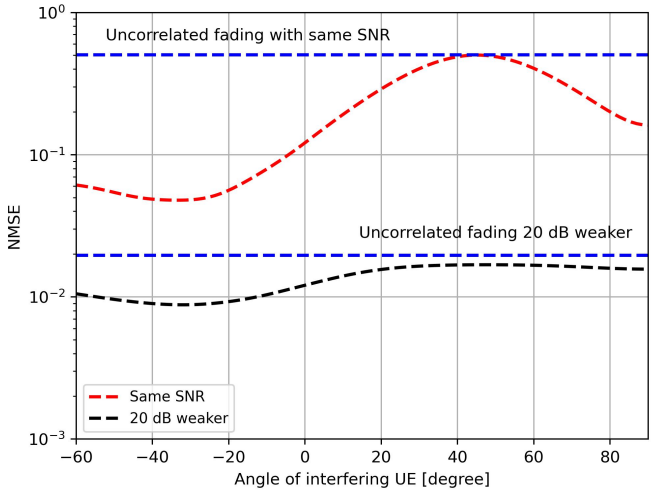


Figure 3.9 – NMSE of the desired channel estimate in the presence of pilot contamination from an interfering UE. There are $N = 8$ antennas, and the local scattering model is used with $\theta_1 = \theta_2 = -10^0$, $\sigma_\varphi = 15^o$ and $\sigma_\theta = 5^o$. The desired UE has an azimuth of $\varphi = 45^o$, while the interfering UE has an azimuth of -60^o and 90^o. The NMSE for uncorrelated fading is shown as a reference case.

bounded by

$$\text{NMSE}_1 \geq 1 - \frac{\eta\tau_p \,\text{tr}\left(\mathbf{R}_1 \left(\eta_1\tau_p\mathbf{R}_1 + \sigma^2\mathbf{I}_N\right)^{-1}\mathbf{R}_1\right)}{\text{tr}\left(\mathbf{R}_1\right)}, \tag{3.75}$$

where the NMSE is equal to the expression in Eq. (3.75) if and only if, the correlation matrices are orthogonal, that is $\mathbf{R_1}\mathbf{R_2} = \mathbf{0}_{N\times N}$. In this case, the NMSE is completely unaffected by the interfering UE. That means that pilot contamination would not occur, in a similar fashion to the favorable propagation phenomena of Chapter 2. Since the orthogonality condition is unlikely to hold in real life applications, it would be sufficient to ensure that $\text{tr}\left(\mathbf{R}_1\mathbf{R}_2\right)$ is as small as possible. The NMSE of the desired channel estimate is shown in Fig. 3.9. The AP is equipped with $N = 8$ antennas, and there are two UEs, the desired UE 1 and the interfering UE 2. The spatial correlation matrix is modeled by the local scattering model, with elevation angles $\theta_1 = \theta_2 = -10^0$, and the ASDs $\sigma_\varphi = 15^o$ and $\sigma_\theta = 5^o$. The azimuth angle of the desired UE is $\varphi = 45^o$, while the azimuth of the interfering UE lies in the range -60^o and 60^o. The effective SNR of the desired signal is 20 dB, and the interfering signal has the same SNR or it is 20 dB weaker. The NMSE is low, and far away from the reference case when the angle of the interfering UE is significantly different from the angle of the desired UE. This gap is much larger when the two signals have the same SNR, since pilot contamination is more aggressive the higher the SNR. This shows that, in the case the interfering UE has a large transmit power, it should be located such that its azimuth angle is different from the desired UE's angle, given the elevation angle is the same. When the interfering signal is much weaker than the desired one there are not such large variations in

the NMSE, therefore the channels might be strongly correlated and still the channel will be estimated with reasonable quality. Thus, there are two ways one could mitigate pilot contamination effect: either by assuring the desired UE has a much larger effective SNR, or if this is not possible, assigning the same pilot to UEs that have much different spatial channel characteristics, or in other words, have orthogonal spatial correlation matrices.

Chapter Summary

- MMSE estimation theory is an important step to understand linear channel estimation techniques, as well as received symbols estimates as will be seen in later sections
- Channel capacity is a measure of the amount of information that can be reliably transmitted through a communication channel. Although the bounds introduced in this chapter may look simple, they are actually quite helpful to understanding the SE metrics that will be introduced and discussed in later chapters.
- Channel estimation is very important to reliably decode the transmitted signal. On the adopted model, since the channel is constant along one coherence block, it needs to be estimated only once in this interval
- MMSE channel estimation uses channel statistics, like the spatial correlation matrix, to obtain an estimate of the channel. The interference generated by reuse of the same pilot sequence by neighboring UEs gives rise to statistical dependence between the estimates, which in turn decreases estimation quality. This phenomena is called pilot contamination.
- Spatial correlation might help to improve channel quality. Given one knows the spatial statistical properties of the UEs in the grid, it is possible to assign pilots in such a way that UEs with similar spatial correlation matrices don't share the same pilot.
- The estimation quality when using many single-antennas APs is much higher than that of many antennas at a single AP. The highest quality of estimation is achieved when the UE is served by only the APs near it. However, CF systems are robust to lower-quality channel estimates.

4 DYNAMIC COOPERATION CLUSTERING ALGORITHMS AND PILOT ASSIGNMENT WITH PERFECT AND IMPERFECT SPATIAL CHANNEL STATISTICS

This chapter aims to introduce pilot assignment algorithms and DCC. Section 4.1 introduces the pilot assignments algorithms and metrics, using the NMSE of channel estimation as a performance metric. Section 4.2 introduces DCC algorithms and discusses its scalability related to various parameters. It also introduces some key parameters that will be fundamental to evaluate system processing complexity at later chapters. Both sections cover the behavior of the algorithms in the situation of imperfect channel statistics also. Finally, Section 4.3 summarizes the algorithms into a table ans discusses the advantages and disadvantages of using each one.

Having studied methods to estimate the communication channels, as well as the pilot contamination interference in Chapter 3, we now shift our focus to the pilot assignment methods, in order to improve the NMSE of the desired channel estimate, and to the formation of DCC. We also analyze the NMSE over the situations of perfect and imperfect knowledge of channel statistics.

4.1 PILOT ASSIGNMENT ALGORITHMS

As seen in Chapter 3, Section 3.3.5, if we consider linear MMSE channel estimation, the reuse of the same pilot sequence by more than one UE gives rise to pilot contamination, that in turn makes the channel estimates correlated, impacting on loss of channel estimation quality.

Therefore, it is of importance to propose pilot assignment methods such that the pilot contamination effect is reduced. Ideally, the pilots would be assigned such that each UE in the grid transmits a different orthogonal pilot sequence. However, since the maximum number of pilot sequences is limited to K, the total number of UEs, there will be the necessity to reuse orthogonal sequences, thus leading to pilot contamination.

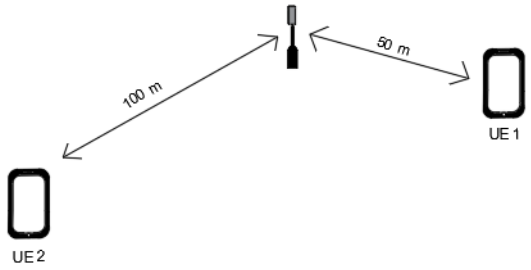


Figure 4.1 – Schematic of two UEs connected to the same AP.

4.1.1 Metrics for Pilot Assignment

Since the interference of the UEs using the same pilot sequence is dependent on the distance between the desired UE and the interfering UEs to the AP that performs channel estimation, it would be adequate to assign orthogonal pilots to the UEs that are nearby the desired UE, and the same pilot sequence to the ones that are far away from it. However, geographical distance is not the only measure of how weak the interference is: as seen in Chapter 3 Section 3.4.3, the spatial correlation properties of the channels play a key role in determining the strength of pilot contamination . For instance, two UEs can be near each other, but have vastly different azimuth and elevation angles to the AP, impacting in weak cross correlation between its channels. Therefore, the same pilot sequence could be assigned to both, having a neglectful impact on channel estimation accuracy.

We can then think on two metrics that would help in assigning the pilots to the UEs. One of them is the average channel gain, or simply channel gain, β, as defined by Eq. (2.12)

$$\beta_{kl}[\text{dB}] = -30.5 - 36.7 \log_{10}\left(\frac{d_{kl}}{1\text{m}}\right) + F_{kl}. \tag{4.1}$$

We can see from this equation that there is a clearly dependency on the distance between the UE and the AP, again suggesting that far away UEs to the AP lead to smaller interference. However, the presence of the term F_{kl}, that corresponds to shadow fading, makes that analysis more complicated. To see that, lets assume that the distance between an arbitrary UE 1 to an AP is $d_1 = 50$m, and the distance between UE 2 and this AP is $d_2 = 100$m, and that their shadowing coefficients are $F_1 = 1$ dB and $F_2 = 12$ dB. Fig. 4.1 shows this scenery The values of β for both of those UE is $\beta_1 = \beta_2 = -92$ dB. This means that the interference, not considering the effect of spatial correlation, will have the same effective SNR as the desired signal, which will introduce a great amount of pilot contamination, even though the interfering UE is distant two times more than the desired UE from the AP. This happens due to the shadowing, that could be more aggressive for one UE and less so for the other.

Notice, however, that if the spatial correlation matrices of both UEs to the AP are orthogonal, then pilot contamination will not occur, even with aggressive shadowing. Therefore, another metric to be considered in the pilot assignment is the amount of cross correlation between the channels. This can be measured by the cross-correlation coefficient given by

$$\rho_{ki} = \frac{\mathrm{tr}\left(\mathbf{R}_k^H \mathbf{R}_i\right)}{\sqrt{\mathrm{tr}\left(\mathbf{R}_k^H \mathbf{R}_k\right) \mathrm{tr}\left(\mathbf{R}_i^H \mathbf{R}_i\right)}}, \tag{4.2}$$

where $\rho_{ki} \in \mathbb{C}$ measures the cross correlation between the spatial channel correlation matrices of UE k and UE i to the AP. Since this is a complex quantity, only its magnitude matters for the analysis: if $|\rho_{ki}| = 0$, then there is no correlation between the channels, and they are orthogonal. However, if $|\rho_{ki}| = 1$, the correlation is at its fullest, and the UEs angles are very similar, making their channel estimates correlated. One interesting aspect of this metric is that it has the potential to cancel pilot contamination interference irrespective of the distance to the AP. This means that it is a quite reliable metric in the situation of different shadowing attenuation of the neighboring UEs using the same pilot sequence. Notice that, if the spatial correlation matrices are the identity (the uncorrelated case), or in the case of single antenna AP, the expression in Eq. (4.2) can be rewritten as

$$\rho_{ki} = \frac{\beta_k \beta_i}{\sqrt{\beta_k^2 \beta_i^2}} = 1 \tag{4.3}$$

showing that in uncorrelated or in single antennas scenarios, the cross-correlation metric is not a great measure, since the channels will posses similar spatial propagation characteristics.

If now we consider that one UE is served by more than one AP, we consider the collective spatial correlation matrices as defined in Chapter 2 Section 2.6.1, then the cross correlation coefficient can be calculated as

$$\rho_{ki} = \frac{\mathrm{tr}\left(\mathbf{R}_k^H \mathbf{R}_i\right)}{\sqrt{\mathrm{tr}\left(\mathbf{R}_k^H \mathbf{R}_k\right) \mathrm{tr}\left(\mathbf{R}_i^H \mathbf{R}_i\right)}} = \frac{\sum_{l=1}^{L} \mathrm{tr}(\mathbf{R}_{kl}^H \mathbf{R}_{il})}{\sqrt{\sum_{l=1}^{L} \mathrm{tr}(\mathbf{R}_{kl}^H \mathbf{R}_{kl}) \sum_{l=1}^{L} \mathrm{tr}(\mathbf{R}_{il}^H \mathbf{R}_{il})}} \tag{4.4}$$

which is the mean cross correlation coefficients of an UE pair with respect to the APs. It measures how correlated the two UE channels are on average per AP.

It is important to notice the cross correlation metric is quite useful to assign the same pilot sequence to the UEs that are near each other and have weak cross correlation. Also, one should notice that, unlike the case of a single AP, the cross correlation coefficient provides

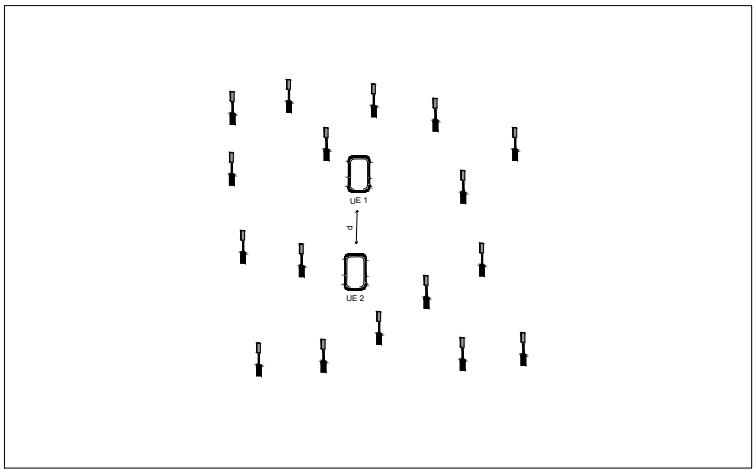


Figure 4.2 – Proposed simulation setup. The APs are randomly placed, and the UEs are places at a varying distance d above and below the grid center.

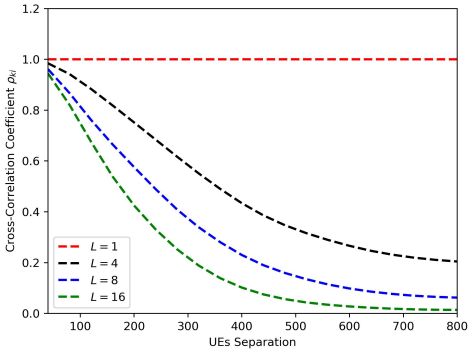
insights into the channel gains of the UEs. We can notice that by doing

$$\rho_{ki} = \frac{\sum_{l=1}^{L} \text{tr}(\mathbf{R}_{kl}^H \mathbf{R}_{il})}{\sqrt{\sum_{l=1}^{L} \text{tr}(\mathbf{R}_{kl}^H \mathbf{R}_{kl}) \sum_{l=1}^{L} \text{tr}(\mathbf{R}_{il}^H \mathbf{R}_{il})}} = \frac{\sum_{l=1}^{L} \beta_{kl}\beta_{il} \text{tr}(\mathbf{U}_{kl}^H \mathbf{U}_{il})}{\sqrt{\sum_{l=1}^{L} \beta_{kl}^2 \text{tr}(\mathbf{U}_{kl}^H \mathbf{U}_{kl}) \sum_{l=1}^{L} \beta_{il}^2 \text{tr}(\mathbf{U}_{il}^H \mathbf{U}_{il})}}$$
$$= \frac{\sum_{l=1}^{L} \beta_{kl}\beta_{il} \text{tr}(\mathbf{U}_{kl}^H \mathbf{U}_{il})}{\sqrt{\sum_{l=1}^{L} \sum_{\substack{m=1 \\ m\neq l}}^{L} \beta_{kl}^2 \beta_{im}^2 \text{tr}(\mathbf{U}_{kl}^H \mathbf{U}_{kl}) \text{tr}(\mathbf{U}_{im}^H \mathbf{U}_{im}) + \sum_{l=1}^{L} \beta_{kl}^2 \beta_{il}^2 \text{tr}(\mathbf{U}_{kl}^H \mathbf{U}_{kl}) \text{tr}(\mathbf{U}_{il}^H \mathbf{U}_{il})}}, \quad (4.5)$$

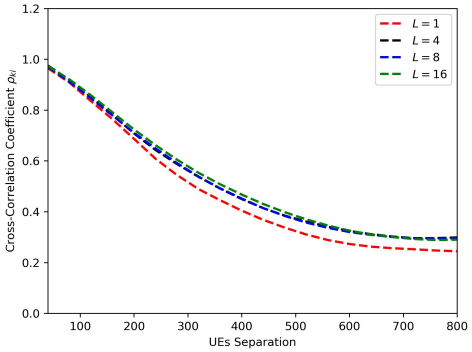
where $\mathbf{U} = \mathbf{R}/\beta$, thus it is the spatial correlation matrix without the channel gain. Assuming the channels of both UEs to the AP are strongly correlated, such that $\mathbf{U}_i = \mathbf{U}_k$ for every l. The channel gains, however are different. Then Eq. (4.5) is written as

$$\rho_{ki} = \frac{\sum_{l=1}^{L} \beta_{kl}\beta_{il} \text{tr}(\mathbf{U}_{kl}^H \mathbf{U}_{kl})}{\sqrt{\sum_{l=1}^{L} \sum_{\substack{m=1 \\ m\neq l}}^{L} \beta_{kl}^2 \beta_{im}^2 \text{tr}^2(\mathbf{U}_{kl}^H \mathbf{U}_{kl}) + \sum_{l=1}^{L} \beta_{kl}^2 \beta_{il}^2 \text{tr}^2(\mathbf{U}_{kl}^H \mathbf{U}_{kl})}} = \frac{\sum_{l=1}^{L} \beta_{kl}\beta_{il}}{\sqrt{\sum_{l=1}^{L} \sum_{\substack{m=1 \\ m\neq l}}^{L} \beta_{kl}^2 \beta_{im}^2 + \sum_{l=1}^{L} \beta_{kl}^2 \beta_{il}^2}}. \quad (4.6)$$

This leads to very interesting results: although the UE channels are greatly correlated, the cross correlation coefficient does not necessarily go to 1, as would in the case of Cellular systems, where each UE only connects to one AP. To better check the influence of the channel gains in ρ_{ki}, we propose the simulation scheme, with the same parameters of the one that generated Fig. 4.4. This time, however, the grid is 1 km $\times$ 1 km, and the UEs are placed at a distance d from the middle of the grid, according to Fig. 4.2 The cross correlation coefficient



(a) Cross correlation coefficient ρ_{ki}^{β} versus distance.



(b) Cross correlation coefficient $\rho_{ki}^{\mathbf{U}}$ versus distance.

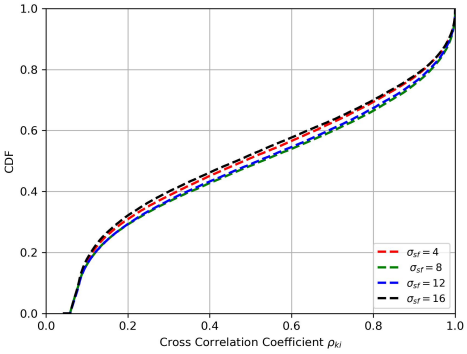
Figure 4.3 – Cross correlation coefficients for varying distance d and number of APs, $L \in \{1, 4, 8, 16\}$. The coefficients are generated assuming the cross correlation definition of Eqs. (4.7) and (4.8), where 0 imply no correlation between the channels and 1 imply maximum correlation between them.

is divided into two distinct groups, defined by

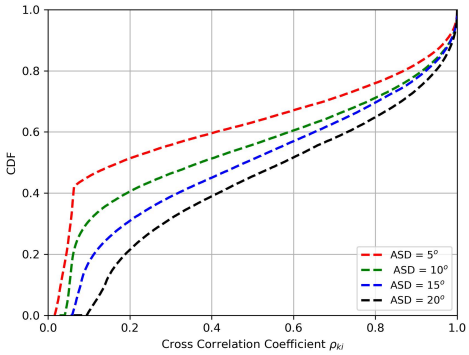
$$\rho_{ki}^{\beta} = \frac{\sum_{l=1}^{L} \beta_k \beta_i}{\sqrt{\sum_{l=1}^{L} \beta_k^2 \sum_{l=1}^{L} \beta_i^2}}, \tag{4.7}$$

$$\rho_{ki}^{\mathbf{U}} = \frac{\sum_{l=1}^{L} \mathrm{tr}(\mathbf{U}_{kl}^H \mathbf{U}_{il})}{\sqrt{\sum_{l=1}^{L} \mathrm{tr}(\mathbf{U}_{kl}^H \mathbf{U}_{kl}) \sum_{l=1}^{L} \mathrm{tr}(\mathbf{U}_{il}^H \mathbf{U}_{il})}}. \tag{4.8}$$

This separation is made assuming that all of the spatial correlation matrices are the identity (the uncorrelated case), for ρ_{ki}^{β}, and that the channel gains are equal for every AP, for $\rho_{ki}^{\mathbf{U}}$. Also, the number of APs is in the range $L \in \{1, 4, 8, 16\}$. The mean cross correlation coefficients are given in Fig. 4.3 We can see from Fig. 4.3a that ρ_{ki}^{β} is maximum for the case of single AP. As L increases, the cross correlation value decreases. This happens because the channel gains become more diverse as the number of APs increases, since the UEs have different channel gains for each AP, given the cross correlation coefficient is given by the mean cross correlation coefficient ρ_{kil}^{β} across the APs. Also, the value of ρ_{ki}^{β} decreases with the distance, since the UEs besides having diverse channel gains from each AP, are also physically far away from each other, which makes their interference even smaller, thus reducing the correlation. In the case of the cross correlation between the normalized spatial correlation matrices, $\rho_{ki}^{\mathbf{U}}$, its behavior is very similar irrespective of the value of L. One could also argue why it decreases with the UEs separation, since it is independent of the channel gains. The answer is, although it is independent of the channel gains, it depends on the azimuth and elevation angles between the APs and the UEs. Therefore, it stands to reason that far away UEs will have different angles, which in turn decreases cross correlation, as would the single AP case.



(a) Cross Correlation versus shadowing standard deviation



(b) Cross Correlation versus ASD

Figure 4.4 – Cross Correlation Coefficient CDF for varying ASDs and shadowing standard deviations. The coefficients are generated assuming the cross correlation definition of Eq. (4.2), where 0 imply no correlation between the channels and 1 imply maximum correlation between them

4.1.2 Random Pilot Assignment

The simplest of pilot assignment algorithms is the random pilot assignment (RPA). This algorithm consists on randomly allocating pilots to the UEs, such that each pilot is used at least once. Clearly, this method has the advantage of not requiring coordination between the AP and the UEs, but also has the drawback of suffering for pilot contamination, since the probability of a sequence being used by a neighboring UE is of $1/\tau_p$. To better show how this method of pilot assignment can increase the influence, we shall simulate a system with $K = 40$ UEs that are served by an arbitrary AP. The simulation grid is of 1 km $\times$ 1 km, and the shadowing standard deviation $\sigma_{sf} \in \{4, 8, 12, 16\}$ dB. The number of AP antennas is $N = 4$, and the spatial correlation matrices are evaluated according to the local scattering model, with both the ASDs $\sigma_\varphi = \sigma_\theta \in \{5^o, 10^o, 15^o, 20^o\}$. The channel gains β_k are calculated using Eq. (2.12). The cross correlation is evaluated for each pair of UEs, according to Eq. (4.2). The CDF of the cross correlation coefficients are shown in Fig. 4.4

We can see from Fig. 4.4a that shadowing variance does not interfere in the distribution of the cross-correlation coefficients. This is a reflex of the fact those coefficients depend only on the correlation of the spatial properties of the UE channels. Since σ_{sf} only influences the value of β, the coefficients distribution is not heavily altered by them. On the case of ASD, is different, as seen in Fig. 4.4b: the higher the ASD, the more dispersed will be the coefficients. Thus, the probability of having a higher correlation between the UEs is greater, the greater the value of the ASD. This figure also shows why it is not advisable to assign the pilots randomly. Since even in the case of low ASD, $\sigma = 5^o$, the probability of having a correlation coefficient above 0.8, $P(\rho_{kl} \geq 0.8) = 0.25$, meaning that in 25% of the cases, the interfering UE will have high correlation with the desired UE. If they are placed near each

other the effect of pilot contamination interference would greatly degrade channel estimation quality. Therefore, more efficient methods will be discussed.

4.1.3 Repulsive Pilot Assignment Algorithm

Since the RPA method is unable to lower the interference caused by pilot contamination, we propose another method to mitigate this effect. The main objective of any pilot assignment algorithm is to maximize or minimize a given metric f given some constraints,

$$\begin{aligned} &\text{Maximize} \quad f(t_1, \ldots, t_K) \\ &\text{subject to} \quad t_k \in \{1, \ldots, \tau p\}, \quad k = 1, \ldots, K. \end{aligned} \tag{4.9}$$

Ideally, the maximization of such metric would mean the exhaustive search, between all of the possible pilot sequences assignments, for the assignment that maximizes f. The exhaustive search is virtually impossible computationally to be performed when the number of UEs grow large. For instance, when the number of UEs is K and the number of possible pilot sequences is τ_p, the number of possible pilot assignments is τ_p^K, which for $K = 20$ and $\tau_p = 5$ (small values) is already a whooping 3.2×10^6 possible sequences. Therefore, instead of finding a global maximum, we should be able to design an algorithm that finds a sub-optimal value. One such algorithm is the repulsive pilot algorithm (RP), defined in [35]. It consists on clustering the UEs based on a metric, such that each cluster carries elements that share a dissimilarity between them (hence the name repulsive). For that, let us consider a binary matrix $\mathbf{X}$, where its elements are $x_{k,t} = 1$ if the UE k belongs to the cluster (or pilot) t, and $x_{k,t} = 0$ otherwise. So the repulsive clustering is obtained by solving the following problem

$$\begin{aligned} \text{Maximize} \quad & \sum_{t=1}^{\tau p} \sum_{k=1}^{K-1} \sum_{k'=k+1}^{K} x_{k,t} x_{k',t} f_r(k, k') \\ \text{subject to} \quad & \sum_{t=1}^{\tau p} x_{k,t} = 1, \quad k \in \{1, \ldots, K\} \\ & \left\lceil \frac{K}{\tau_p} \right\rceil \leq \sum_{k=1}^{K} x_{k,t} \leq \left\lceil \frac{K}{\tau_p} \right\rceil + 1, \quad t \in \{1, \ldots, \tau p\} \\ & x_{k,p} \in \{0, 1\}, \quad k \in \{1, \ldots, K\}, \quad p \in \{1, \ldots, \tau p\}; \end{aligned} \tag{4.10}$$

where $f_r(k, k')$ is a metric, or score, that depends on UE k and UE k'. It can be predefined and static, such as an Euclidean function, or it can be learned, for example, by use of neural networks. The main objective of this function is to assign a repulsion score such that the overall repulsion score in Eq. (4.10) is increased. A simple, yet efficient heuristic algorithm to evaluate this score is given by Algorithm 1. This algorithm works by assigning pilots randomly to the UEs, by use of the pilot sequence vector $\mathbf{t}$. This vector is such that each

Algorithm 1 Repulsive Pilot Assignment

Input: Number of clusters τ_p, set of UEs $\mathcal{K}$
Output: Pilot assignment vector $\mathbf{t}$
Randomly divide $\mathcal{K}$ UEs into τ_p equal sized clusters $\mathcal{C}$
while Overall repulsion function is increasing **do**
  for $C_1, C_2 \in \mathcal{C}$ **do**
    for $u \in C_1$ and $w \in C_2$ **do**
      if exchanging clusters of u and w increases the overall repulsion function **then**
        Swap the clusters of u and w.
      end if
    end for
  end for
end while
for $t = 1 : \tau_p$ **do**
  Assign pilot ϕ_{t_k} to UEs in cluster $\mathbf{t}$
end for

sequence is used at the same amount among all the UEs. Then, the overall score of Eq. (4.10) is evaluated. Defining the group of UEs that use the same pilot sequence t_1 as C_1, one of the elements, u is swapped, from cluster C_1, with an element w, of cluster C_2. By doing that, the overall repulsion score of Eq. (4.10) is again evaluated. In case this score increases, then the elements u and w are permanently swapped. Otherwise, they swap back to their original clusters. This is repeated for every u and w of every cluster. After passing through every cluster, the new overall repulsion score value is compared to the old value. If the new value is bigger, then repeat the process. Otherwise, end the algorithm.

This solution is sub-optimal, in the sense that it does not achieve the global maximum of the overall repulsion score. It also selects randomly the initial pilot assignment sequence. Since the assignment is randomly generated, one random sequence could generate an outcome that is widely diverse if compared to the outcome of another pilot assignment sequence. For instance, one sequence could lead to an outcome that has lesser pilot contamination than the other sequence. However, experimental results show that the difference in pilot contamination is not as huge, so the random sequence is not a limiting factor after all, as we will see in later sections.

For now, we shall define the simulation setup that will be thoroughly used in this monograph, and is proposed by [22] and [8], according to Table 4.1,

Having defined the simulation setup, we now simulate the average NMSE of UE k. We

Parameter	Value
Number of APs	$L = 100$
AP deployment	Uniformly deployed (wrap-around topology)
Number of AP antennas	$N = 4$
Number of UEs in the grid	$K = 40$
Grid dimensions	1 km $\times$ 1 km
Noise variance	$\sigma_n^2 = -94$ dBm
Uplink transmit power	$p = 20$ dBm
Shadowing standard deviation	$\sigma_{sf} = 4$ dB
Bandwidth	$B = 20$ MHz
Pilot sequence length	$\tau_p = 10$
Coherence blok length	$\tau_c = 200$
Channel gain at 1 km	$\gamma = -140.6$ dB
Pathloss exponent	$\alpha = 3.67$
Height difference between the UEs and the APs	10 m
Angular standard deviation	$\sigma_\varphi = \sigma_\theta = 15^o$

Table 4.1 – Parameters for the simulation setup of this monograph. Those will always be considered, unless explicitly said otherwise.

compare the following metrics for the repulsive function $f_r(k, k')$:

$$f_r(k, k') = ||\mathbf{d}_k - \mathbf{d}_{k'}||^2 \tag{4.11}$$

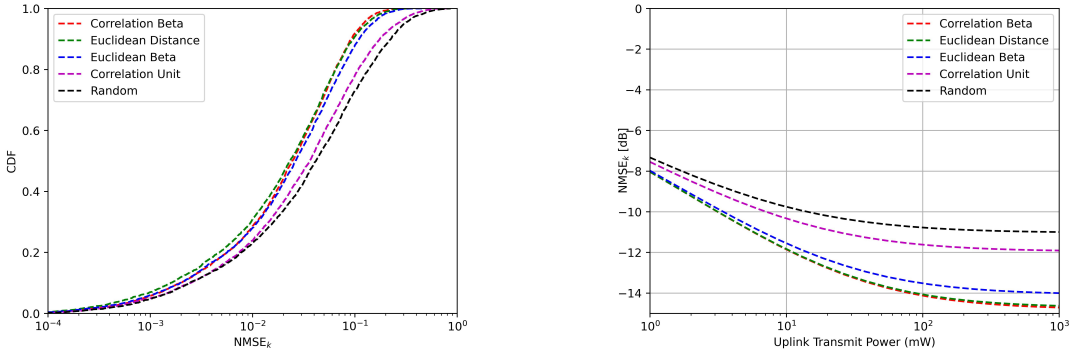
$$f_r(k, k') = ||\beta_k - \beta_{k'}||^2 \tag{4.12}$$

$$f_r(k, k') = 1 - \rho^{\beta}_{kk'} \tag{4.13}$$

$$f_r(k, k') = \rho^{\mathbf{U}}_{kk'}, \tag{4.14}$$

where $||\mathbf{a}-\mathbf{b}||^2$ is the squared Euclidean distance between $\mathbf{a}$ and $\mathbf{b}$, that could also be written as $\sqrt{\sum_{l=1}^{L}(a_l - b_l)}^2$. The average NMSE of an arbitrary UE k was calculated according to Eq. (3.58). In Fig. 4.5a, the CDF of the NMSE for the RP method metrics and for the RPA are shown, while in Fig. 4.5b the mean NMSE is shown as function of the uplink transmit power p varying from 1 mW to 1 W for the same pilot assignments algorithms as Fig. 4.5a. We can see from Fig. 4.5 that the NMSE values are the lowest when using the distance metric of the RP. This is to be expected, since assigning the same pilot to UEs that are far apart from each other is the main objective of any pilot assignment algorithm. Therefore, if the distance between the UEs and the APs is known, it is much easier to infer whether the UEs are far away or not. Next comes the channel gains metric. As already discussed, this metric is similar to the distance metric if the random variations of shadow fading are small. Since those variations have a standard deviation of $\sigma_{sf} = 4$ dB, the resulting NMSE is very similar to the NMSE of distances knowledge.

The RPA has the highest values of NMSE, as expected. The correlation metrics for RP



(a) NMSE CDF

(b) NMSE versus uplink transmit power p

Figure 4.5 – NMSE CDF and average NMSE versus uplink transmit power p for RP method using the metrics of Eq. (4.11) to Eq. (4.14). RPA is also used for a reference case.

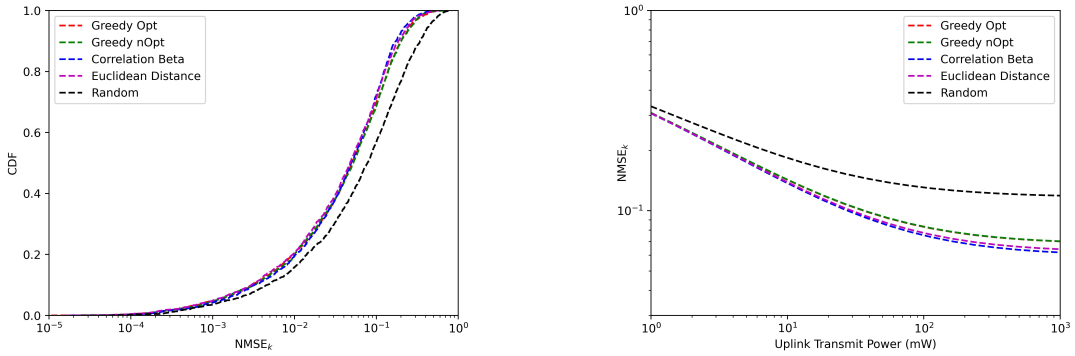
lead to a NMSE, that in the case of Eq, (4.13) becomes similar to the metric of Eq. (4.11), and leads to even better results than Eq. (4.12) This happens because, as seen in Fig. 4.3a, as the number of APs gets large, the cross-correlation coefficients tend to zero. Therefore, the RP algorithm interprets those small coefficients as having a higher repulsion score, and thus the overall repulsion score of Eq. (4.10) gets progressively bigger. The same thing happens for the metric of Eq. (4.14), however, since this metric measures the correlation between the spatial properties of each UE-AP link, it does not have any knowledge of the UEs distance. Therefore. it cannot precisely assign the same pilot to UEs that are far apart, instead relying on correlation of the $\mathbf{U}$ matrices. For instance, picture a UE k that is far away from another UE i. Suppose that those UE are weakly correlated. According to the metric of Eq. (4.14), the repulsion score in that case would be small, which would prevent the two UEs from sharing the same pilot, even though they are far apart. Thus, the NMSE of this metric is slightly lower than the RPA, given that spatial correlation at least assigns some metric, as opposed to just randomly assign pilots with no metrics whatsoever. Finally, the NMSE values decrease with the uplink transmit power, which is to be expected, since the effective SNR increases. However, the interference of the neighboring UEs also increases, given that all UEs transmit with equal power p. This leads to an stabilization of the NMSE value irrespective of the pilot assignment method.

4.1.4 Greedy Pilot Assignment Algorithm

Another method of pilot assignment is the greedy pilot assignment algorithm (GPA) proposed by [8]. It assigns the pilots to τ_p random UEs and evaluates at which pilot does the interference of neighboring UE is minimized. Then it assigns the pilot to this UE. Algorithm 2 summarizes this. This method can be performed by selecting the AP that has the best channel condition to the UE, or in other words. highest channel gain. Then, for each pilot

Algorithm 2 Greedy Pilot Assignment

Output: Pilot assignment vector $\mathbf{t}$
for $k = 1, ..., \tau_p$ **do**
    $t_k \leftarrow k$ ▷ Assign orthogonal pilots tot he first τ_p UEs
end for
for $k = \tau_p + 1, ..., K$ **do**
    $l \leftarrow \underset{l \in \{1,...,L\}}{\arg\max} \beta_{kl}$ ▷ Find the best AP for UE k
    $\tau \leftarrow \underset{t \in \{1,...,\tau p\}}{\arg\min} \sum_{\substack{i=1 \\ t_i = t}}^{k-1} \beta_{il}$ ▷ Find the pilot with least interference at AP l
    $t_k \leftarrow \tau$ ▷ Assign pilot τ to UE k
end for



(a) NMSE CDF

(b) NMSE versus uplink transmit power p

Figure 4.6 – NMSE CDF and average NMSE versus uplink transmit power p for RP and GPA method using the metrics of Eq. (4.11) and Eq. (4.12), as well as GPA with UEs separations known (optimal), and unknown (non-optimal). RPA is also used for a reference case.

t, this AP computes the sum of the interference generated by UEs sharing the same pilot, and select the pilot τ that has the minimum interference. Again, we can identify the random property making this option of pilot assignment sub-optimal: In case a different set of UEs that are nearby each other were assigned orthogonal pilots, thus initializing the assignment algorithm, the interference minimization would surely lead to a situation of assigning the same pilot to UEs that are far apart. To check this, we propose a simulation scheme similar of that from Table. 4.1, but with $K = 20$, $L = 50$ and $\tau_p = 5$. The reason for the decrease in the value of the variables, specially τ_p is that, since the GPA with knowledge of the distance between the UEs requires that an exhaustive search be made to assign orthogonal pilots to the UEs that are nearby, thus enhancing the complexity of the pilot assignment algorithm, which would become too large if we used all of the parameters of Table 4.1. The corresponding NMSE distribution, as well as the NMSE variations for different uplink transmit power are shown in Fig. 4.6a and Fig. 4.6b As we can see from Fig. 4.6, the RPA has the highest values of NMSE, for reasons already discussed. In respect to the GPA, it can be seen that the optimal case ,using the distance between the UEs, and the non-optimal have similar

performances. This means that the non-optimal solution, that does not require exhaustive search, nor distance estimations, can be deployed with satisfactory results. In the case of RP, the results are even better, with the channel gains correlation metric having the better performance, even higher than the distance metric. This might indicate that RP should always be used, but we must heed the warning: it is only usable when the UEs can be grouped into equal sized clusters, which requires that K be a multiple of τ_p. Therefore, it has limited uses, whereas the GPA can be used without this restriction.

4.1.5 NMSE for Pilot Assignment Algorithms with Imperfect Channel Statistics

So far, we have always assumed that the statistics, namely the spatial correlation matrices $\mathbf{R}$ are known everywhere and every time needed. However, it is interesting to investigate the behavior of the pilot assignment algorithm under imperfect channel spatial correlation matrices knowledge. Since the assignment methods are all dependent on the statistics from those matrices, it is likely that imperfect knowledge impacts the value of the NMSE, and thus system performance.

In the case of known $\mathbf{R}$, increasing the UL transmit power would affect the NMSE value because it would make the noise interference more negligible. However, if we take a look at Eq. (3.49) and Eq. (3.50) from Chapter 3,

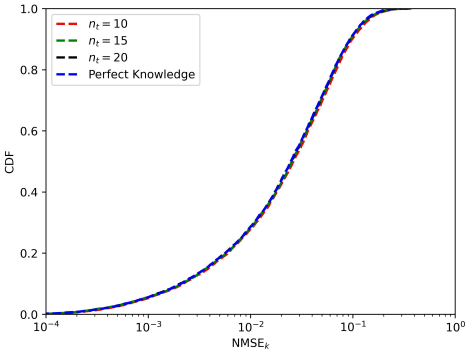
$$\mathbf{h}_{kl}^{\text{LS}} = \sqrt{\eta_k \tau_p}\mathbf{h}_{kl} + \mathbf{n}_{t_k l} \tag{4.15}$$

and

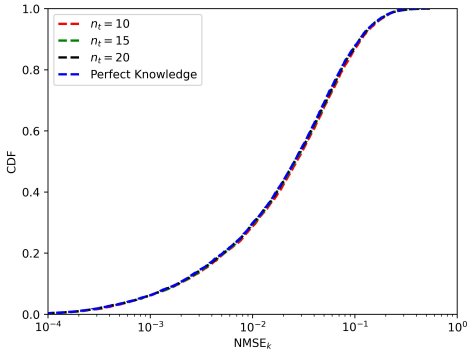
$$\mathbf{R}_{kl}^{\text{sample}} = \frac{1}{N_c N_t} \sum_{n=1}^{N_c} \sum_{n_t=1}^{N_t} \frac{\mathbf{h}_{kl}^{\text{LS}}}{\sqrt{\eta_k \tau_p}} \frac{(\mathbf{h}_{kl}^{\text{LS}})^H}{\sqrt{\eta_k \tau_p}}, \tag{4.16}$$

the quality of the estimation depends on the noise power, therefore the increasing value of the UL transmit power would improve system performance in two ways instead of one. To prove this, we propose the simulation scheme with the same parameters as Table 4.1. Also, each UE transmits a total of $N_t \in \{10, 15, 20\}$ times over a total of $N_c = 20$ coherence blocks to obtain an estimate $\mathbf{R}_{kl}^{\text{sample}}$. The pilot assignments used are RPA, the RP with the channel gains correlation metric and the GPA. The CDF of the NMSE is obtained for the pilots assignments, and also, we simulate the mean NMSE by varying the UL transmit power between 1 mW and 1 W. The NMSE is obtained for all pilot assignments, under the assumption of perfect and imperfect knowledge of the spatial correlation matrix, in the case of imperfect correlation, the parameters for estimation are $N_t = 15$ and $N_c = 20$. The results are shown in Fig. 4.7

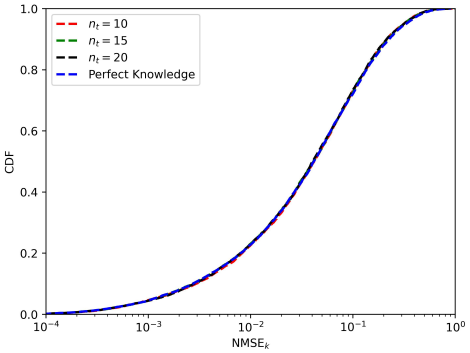
We can see from Figs. 4.7a, 4.7b and 4.7c, that the CDF for the imperfect channel



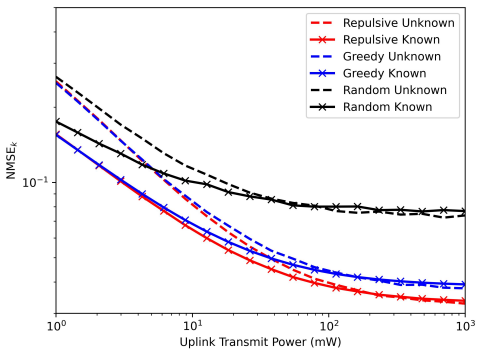
(a) CDF for the NMSE of RP.



(b) CDF for the NMSE of GPA.



(c) CDF for the NMSE of RPA.



(d) NMSE versus UL transmit power for RP, RPA and GPA with perfect and imperfect channel statistics.

Figure 4.7 – NMSE CDF for imperfect and imperfect channel statistics for GPA,RPA and RP. The chosen metrics for the RP is the channel gains correlation, and non-optimal GPA. The mean NMSE versus UL transmit power is also shown, for perfect and imperfect channel statistics.

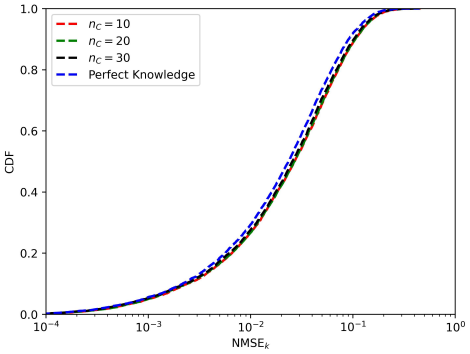
statistics knowledge is very similar to the one of perfect knowledge. The case of $N_t = 10$ still has a little gap between the curves, but for $N_t = 15$ and forward, the CDF's curves are virtually the same. Also, from Fig. 4.7d, the gap between imperfect and perfect spatial correlation matrix estimation is dependent on the UL transmit power. When it is smaller, for example, 1 mW, the difference between the NMSEs is clearly seen, since the spatial correlation matrix becomes less reliable. However, as the power increases, the estimate gets progressively better, thus making the performances of imperfect and perfect spatial channel knowledge similar. Also, as already seen, the RPA has the worst performance, followed by GPA, and finally the RP, with perfect and imperfect channel statistics knowledge.

It is important to notice that the value of N_t has a limit dictated by the coherence block length. For instance, if we assume that τ_p UEs transmit simultaneously, then in a total of N_t transmissions, the number of symbols that will be used for those transmissions is

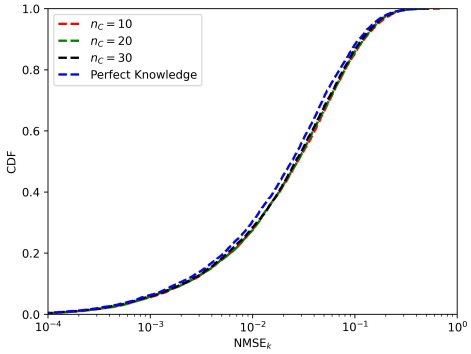
$$\tau_l = K \times N_t. \tag{4.17}$$

Therefore, if N_t is such that $\tau_l > \tau_c$, then there cannot be this amount of transmissions, since some of them would be made into another coherence block, which would include a different realization of the channel. Therefore, we must ensure that $N_t \leq \tau_c/K$. For example, in our simulation schemes the value of $\tau_c = 200$, so $N_t = 200/40 = 5$, therefore there can be a maximum of $N_t = 5$ transmissions for each UE per coherence block. The larger the coherence block length, the more robust will be the spatial covariance matrix estimation, since more transmissions will be made by the UEs.

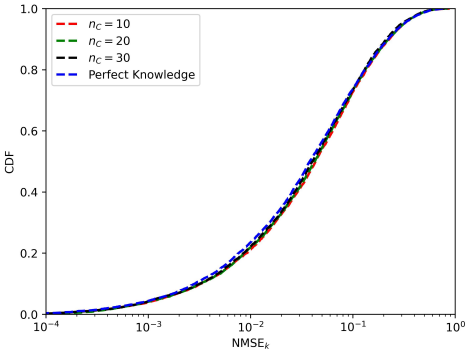
We now explore the effect of the number of coherence blocks into channel statistics estimation. It is obvious from Eq. (3.50) that the use of more coherence blocks for channel statistics estimation increase the quality of estimation. Unlike the parameter N_t, that is bounded by the coherence block length, the value of N_c is bounded only by the number of coherence blocks until the channel statistics suffer significant variation. This value usually takes values on the magnitude range of hundreds. Therefore, if we use a value of N_c that is in the range of dozens, then the number of coherence blocks used for transmission would still be considerable. Therefore, we now propose the same simulation scheme as the one that was used to simulate Fig. 4.7, but we assume that each UE transmits a total of $N_t = 5$ times per coherence block, and also that the number of used coherence blocks for channel statistics estimation lies in the range $N_c \in \{10, 20, 30\}$. Fig 4.8 contains those results. We can see from Figs. 4.8a, 4.8b and 4.8c, that the CDF for the imperfect channel statistics knowledge is very similar to the one of perfect knowledge. Even for a small number of used coherence blocks, $N_c = 10$, and for an even smaller number of transmissions $N_t = 5$, the quality of estimation is such that the estimated NMSE CDF approaches the real NMSE CDF. Also, we can see from Fig. 4.8d, that the UL transmit power plays an important role into the quality of estimation. Since for smaller power, the noise is predominant, the quality of estimation



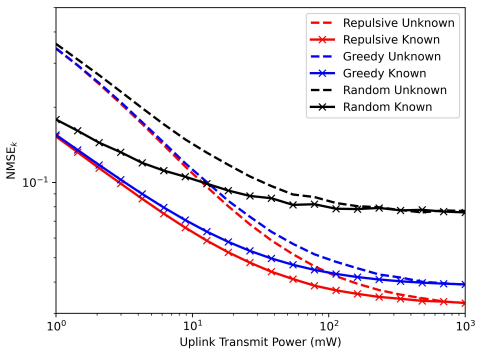
(a) CDF for the NMSE of RP.



(b) CDF for the NMSE of GPA.



(c) CDF for the NMSE of RPA.



(d) NMSE versus UL transmit power for RP, RPA and GPA with perfect and imperfect channel statistics.

Figure 4.8 – NMSE CDF for imperfect and imperfect channel statistics for GPA,RPA and RP. The chosen metrics for the RP is the channel gains correlation, and non-optimal GPA. The mean NMSE versus UL transmit power is also shown, for perfect and imperfect channel statistics.

is decreased, and even with $N_c = 20$ can the estimated NMSE be comparable to the real NMSE. As the power increases, the quality of estimation also increases, to the point where $p = 100$ mW, the values of the NMSEs are almost equivalent. As a result, increasing the transmit power is an option of reliably estimating channel statistics. However, if we increase the value of p, we enter the curve part where the NMSE is constant irrespective of power, so we should increase it wisely. Also, this method of estimating the spatial correlation matrices is only one of the many possible methods. If one does not have as much coherence blocks available for estimation, or the length τ_c does not allow for many transmissions, one could use alternative methods to estimate $\mathbf{R}_{kl}$.

4.2 DYNAMIC COOPERATION CLUSTERING METHODS

Having defined the pilot assignment algorithms, we now focus on developing clustering methods for the APs. Ideally, the UE would connect to every available AP in the grid, thus, by the definition of $\mathcal{M}_k$, $\mathcal{M}_k \in \{1, 2, ..., L\} \, \forall k \in \{1, 2, ..., K\}$. However, this clustering method is not scalable on the senses defined in Chapter 2, therefore alternative clustering methods will be analyzed.

When proposing any clustering method, it is important to select which metric would be more suitable for the problem at hand. This is the focus of the next section.

4.2.1 Metrics for AP Clustering

When choosing the metric for clustering, it is important to choose a metric that would take new realizations over a longer period. For example, if one is to choose any coherence-block related metric, then the interval at which the metric would be applicable would be limited to τ_c, the length of the coherence block. This interval is too short for any application of AP clustering, as it would require that a new AP clustering formation be made for each channel realization. Thus, a metric that changes slowly with time will have to be chosen. With that in mind, two possible variables come naturally: the spatial correlation matrices, and by those, the channel gains. They are considered constant for hundred of coherence blocks, and thus the AP clustering should be remade after a couple of hundreds of coherence blocks, which seems plausible. Having defined that, we can introduce the SNR metric, defined by

$$\text{SNR}_{kl} = \frac{\eta_k \beta_{kl} \tau_p}{\sum\limits_{i \in \mathcal{P}_k \setminus \{k\}} \eta_i \beta_{il} \tau_p + \sigma^2}, \tag{4.18}$$

where η_k is the UL transmitted power of UE k. This metric measures the SNR of the desired UE k to an AP l, and by doing that measures if the UE would have a reasonable signal

strength to connect to the AP if compared to the noise. Another suggested metric would be the normalized channel gain coefficients, that is

$$\tilde{\beta}_{kl} = \frac{\beta_{kl} - \bar{\beta}_l}{\sqrt{\text{Var}(\beta_l)}} \tag{4.19}$$

where

$$\bar{\beta}_l = \frac{1}{K} \sum_{k=1}^{K} \beta_{kl} \tag{4.20}$$

$$\text{Var}(\beta_l) = \frac{1}{K-1} \sum_{k=1}^{K} (\beta_{kl} - \bar{\beta}_l)^2. \tag{4.21}$$

Thus, we define a channel gain that is normalized across the UEs, that is, the values of those channel gains will have equal weights independent of the analyzed AP. Fundamentally, this means that the distribution of $\tilde{\beta}_{kl}$ is independent on the number of APs, therefore more can be added on the grid without changing its behavior. Also, if all UE have the same weight irrespective of the AP, then we would prevent a situation where the AP is nearby a lot of UEs and gets connected to them all, therefore leading to an overload on the number of UE of which it should process the signals, which might cause the complexity to significantly increase, as will be discussed in Chapter 5.

Another useful metric is a variation of Eq. (4.19), proposed by [16], where instead of the normalized channel gains, we define

$$\beta_k^{\delta} = \frac{\sum\limits_{l \in \mathcal{M}_k} \beta_{kl}}{\sum\limits_{l=1}^{L} \beta_{kl}}, \tag{4.22}$$

where the formation of $\mathcal{M}_k$ is dependent on a predetermined number of APs that each UE should be connected to, and by that select this amount of the largest values of either normalized channel gains or actual channel gains for the APs to the desired UE. This metric, as well as the normalized channel gains metric does not take in consideration pilot contamination. That is, since the same AP could be connected to pilot-sharing UEs, the interference from pilot contamination could severely weaken the processing of an AP that is farthest from its desired UE than the interfering pilot-sharing UE.

To prevent this, one approach is to define a secondary metric, the correlation between the spatial correlation matrices of the UEs connected to a certain AP, or

$$\rho_{kil} = \frac{\text{tr}\left(\mathbf{R}_{kl}^H \mathbf{R}_{il}\right)}{\sqrt{\text{tr}\left(\mathbf{R}_{kl}^H \mathbf{R}_{kl}\right) \text{tr}\left(\mathbf{R}_{il}^H \mathbf{R}_{il}\right)}} \tag{4.23}$$

and by that, apply the fact that weak correlation means very different spatial propagation properties. If the pilot sharing UEs have a weak correlation coefficient between them, then they can be served by the same AP with negligible pilot contamination. Therefore, even if the UEs that share the same pilot are geographically close, the pilot contamination effect will be negligible if their correlation is low.

4.2.2 Normalized Threshold AP clustering

One such way of clustering the APs in a DCC manner, is by defining a given threshold β_{th} to the normalized channel gains, as defined in Eq. (4.19). The UEs will connect to those APs with whom they have a channel gain above the threshold value β_{th}. Thus, this DCC algorithm, initially proposed by [15], which we will call normalized threshold algorithm (NTA), is given by Algorithm 3 From this algorithm, we should notice that there is no dependence

Algorithm 3 Normalized Threshold Algorithm

Input: Channel gains β_{kl}
Output: DCC sets $\mathcal{M}_1, \mathcal{M}_2 ... \mathcal{M}_K$
for $k = 1 : K$ **do**
  for $l = 1 : L$ **do**
    $\tilde{\beta}_{kl} = \frac{\beta_{kl} - \bar{\beta}_l}{\sqrt{\text{Var}(\bar{\beta}_l)}}$ ▷ Normalize the channel gain coefficients along the UEs axis
    if $\tilde{\beta}_{kl} \geq \beta_{\text{th}}$ **then**
      $l \in \mathcal{M}_k$ ▷ Add AP l to cluster $\mathcal{M}_k$
    else
      $l \notin \mathcal{M}_k$
    end if
  end for
end for

on the pilot sequence length. In fact, there is no dependence on the pilot assignment vector $\mathbf{t}$ at all. This has the advantage of being a scalable DCC method with respect to the number of pilots, but it also carries the disadvantage of introducing an additional source of pilot contamination: since the APs are clustered based solely on the channel gains, there is a high probability that pilot sharing UEs will be connected to the same AP.

The better the pilot assignment method, the less interference caused by pilot contamination, however, since it allocates the same pilot that are far away from each other, and will most probably be connected to different APs. There will be cases, however, in which the distance of the AP from the pilot sharing UE will be comparable, so the pilot contamination will be more aggressive. Also, there will be situations where the UE will have very weak channel gains with respect to all APs in the grid, and will connect to a handful of them in order to improve signal detection quality. If this UE shares a pilot with a stronger UE, then certainly the stronger one will interfere in the channel estimation of the weaker one, thereby

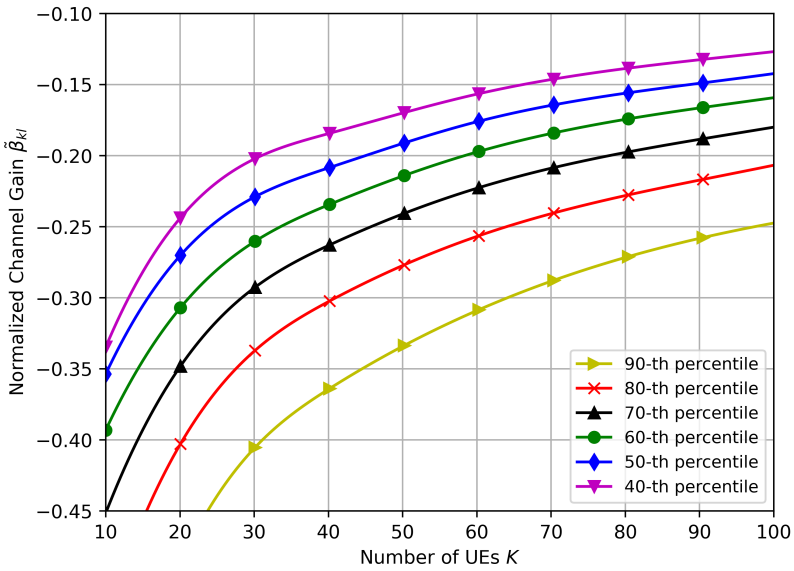


Figure 4.9 – Percentiles of the normalized channel gain coefficients versus number of UEs K. Notice that the percentile number represents the percentage of normalized channel gain coefficients that are above it.

reducing its performance.

It is also interesting to analyze how the normalized coefficient behave by varying the grid parameters, such as the number of APs, the number of UEs, and also the number of AP antennas. With respect to the antennas, it is possible to say that it will not interfere in the values of the normalized coefficients, since they rely on the channel gains, which are independent of N. With respect to the number of APs L, we can say that it will also not interfere in the values of $\tilde{\beta}$ since the normalization occurs in the UEs axis, that is, irrespective of the analyzed AP, the normalized channel gains will be a random variable with zero mean and unit variance. With respect to the number of UEs, however, the dependence will be observed: a low number of UEs will have a sample size that is so small that random variations occur much more aggressively. In resume, the number of UEs is small, thus the channel gains are more diverse, since one UE could be placed more closely to the analyzed AP while the other is placed near it, causing a much more varied channel gains pattern. To show how these variations occur, we propose the reference scheme simulation of Table. 4.1, and we vary the number of UEs. Fig. 4.9 shows the percentiles of the normalized channel gains versus the number of UEs in the grid, That is, it shows the percentage of normalized channel gains that are above a certain value. This value is highly dependent on the number of UEs: when K is small, the value is higher, and when K is large the value is higher, and also tends to be constant, as already discussed. This values can also be a measure of the amount of UEs that are connected to an AP, or the amount of APs that are connected to a given UE. From that, we can see that the NTA DCC is unscalable with respect to the number of UEs: If we fix

a value β_{th}, and vary the number of UEs, then as K gets progressively bigger, so does the percentiles. If we take the 80-th percentile for $K = 40$, we have that its value is -0.3. If now we consider that $K = 60$, this value is closer to -0.25. Supposing we fixed a threshold of $\beta_{\text{th}} = -0.3$, then in the first case an average of 80% of the AP would be connected to the UE. In the second case, this percentage would be near 90%, thus making the system more unscalable, since the more APs connected to the UE, the more operations needed to decode its signal.

This family of curves can be used to estimate a threshold coefficient that connects a desired percentage of the APs to the UE while maintaining scalability, so the system is unscalable only if the threshold value does not change.

4.2.3 Power Threshold Algorithm

Similar to NTA, the power threshold algorithm (PTA), which has been proposed by [16], selects the APs that satisfy a given threshold condition. However, differently from simply selecting them based on the channel gains alone, this time the APs that have the higher channel gains are selected, by means of the power metric β_k^{δ}, defined in Eq. (4.22). This metric creates the AP clustering based on the APs whose channel gains contribute to δ of the total UE power, that it would have if all of the APs were connected to it. This metric does not depend also in the pilot assignment vector $\mathbf{t}$, thereby it also may be an additional source of pilot contamination. In order to reduce the effect of pilot contamination another DCC method will be proposed.

Algorithm 4 Power Threshold Algorithm

Input: Channel gains β_{kl}, Pilot Assignment vector $\mathbf{t}$
Output: DCC sets $\mathcal{M}_1, \mathcal{M}_2 ... \mathcal{M}_K$
for $k = 1 : K$ **do**
$\mathcal{M}_k = \left\{ l \in 1, 2, ..., L : \frac{\sum_{l \in \mathcal{M}_k} \beta_{kl}}{\sum_{l=1}^{L} \beta_{kl}} \leq \delta \right\}$ ▷ Select the APs that contribute to at least $\delta\%$ of the total power from all of the APs in the grid to the UE.
end for

4.2.4 Orthogonal Power Threshold Algorithm

In order to reduce the pilot contamination effect, we propose an alternative to PTA that allows only the UEs that have been allocated orthogonal pilots to be served by the same AP, which we shall call Orthogonal PTA. We first define $\mathcal{K}_l$ as the set of UEs that are served by AP l, and then we propose an algorithm Algorithm 5, which is given as follows This algorithm consists on many parts. First, we select the AP cluster that is responsible for a δ percentage of the total UE power, for each UE. Then, since the probability of selecting pilot-sharing UEs for a given AP is high, we search the UEs that share the same pilot sequence

Algorithm 5 Orthogonal Power Threshold Algorithm

Input: Channel gains β_{kl}, Pilot Assignment vector $\mathbf{t}$
Output: DCC sets $\mathcal{M}_1, \mathcal{M}_2 ... \mathcal{M}_K$
for $k = 1 : K$ **do**
$\mathcal{M}_k = \left\{ l \in 1, 2, ..., L : \frac{\sum_{l \in \mathcal{M}_k} \beta_{kl}}{\sum_{l=1}^{L} \beta_{kl}} \leq \delta \right\}$ ▷ Select the APs that contribute to at least $\delta\%$ of the total power from all of the APs in the grid to the UE.
end for ▷ If we wish to mitigate pilot contamination interference, we proceed
for $l = 1 : L$ **do**
for $t = 1 : \tau_p$ **do**
$\mathcal{U}_{tl} = \{k \in \mathcal{K}_l : k \in \mathcal{P}_k\}$ ▷ Find the set of UEs that share the same pilot sequence t_k.

if $|\mathcal{U}_{tl}| > 1$ **then** ▷ If there are UEs sharing the same pilot t
for $k \in \mathcal{U}_{tl}$ **do** ▷ Evaluate the percentage of the power that the UEs sharing pilot t carry over the set $\mathcal{M}_k$

$$p_k = \frac{\sum_{l' \in \mathcal{M}_k} \beta_{kl'}}{\sum_{l'=1}^{L} \beta_{kl'}}$$

▷ Evaluate the percentage of the power of UE k if AP l is disconnected

$$p'_k = \frac{\sum_{\substack{l' \in \mathcal{M}_k \\ l' \neq l}} \beta_{kl'}}{\sum_{l'=1}^{L} \beta_{kl'}}$$

$p\% = 1 - p'_k/p_k$ ▷ Calculate the percentage variation of the total power of each UE.
end for
$k_{\text{min}} \leftarrow \underset{k \in \mathcal{U}_{tl}}{\arg\max} \quad p\%$ ▷ Select the UE that has the highest percent change if AP l was disconnected, and disconnect all other UEs.
$\mathcal{M}_{k_{\text{min}}} \leftarrow \mathcal{M}_{k_{\text{min}}} \cup \{l\}$
end if
end for
end for

and are connected to AP l. Then, we recalculate Eq. (4.22) with the l-th AP disconnected. The UE that has the highest variation of the overall percent power is reconnected to AP l. This way we are ensuring that each AP is connected to UEs with orthogonal pilot sequences, and also the total power of each UE is not gravely reduced, by removing APs that have low channel gains with respect to the UE.

This method is much more scalable than the NTA, since the AP has a tendency of selecting the UEs that have orthogonal pilots, which would account for a maximum of τ_p UEs. Also, the pilot contamination is greatly reduced with this method.

4.2.5 Orthogonal Users Algorithm

Another DCC algorithm has been proposed by the authors in [8]. There, it is simply known as DCC, but in this work we shall call it orthogonal users algorithm (OUA). It consists on connecting each AP with the strongest τ_p UE that have orthogonal pilots, as seen in Algorithm 6 This algorithm has the advantage of reducing pilot contamination, since each

Algorithm 6 Orthogonal Users Algorithm

Input: Channel gains β_{kl}, Pilot Assignment vector $\mathbf{t}$
Output: DCC sets $\mathcal{M}_1, \mathcal{M}_2 ... \mathcal{M}_K$
for $l = 1 : L$ **do**
  for $t = 1 : \tau_p$ **do**
    $i \leftarrow \underset{k \in \{1,...,K\}: t_k = t}{\arg\max} \beta_{kl}$ ▷ Find UE that AP l serves on pilot t
    $\mathcal{M}_i \leftarrow \mathcal{M}_i \cup \{l\}$
  end for
end for

AP is connected to the UE that have orthogonal pilots allocated to them ,hence the name of the algorithm, and also is capable of guaranteeing a reasonable performance for each UE, since each of them are very likely to be connected to a handful of strong APs. It is however unscalabe when we take the number of orthogonal sequences τ_p into consideration, since its complexity grows with τ_p: If for instance $\tau_p = 40$ in the reference simulation scenario, then we have the case of all serving APs serving the UE, hence unscalable. We shall then propose another DCC algorithm that aims to reduce pilot contamination, but also make the system scalable with respect to τ_p.

4.2.6 Orthogonal Normalized Threshold Users Algorithm

Since the main limitation of the OUA is the unscalability with respect to the length of the pilot sequence, we propose yet a new algorithm that attempts to make it scalable: the orthogonal normalized threshold users algorithm (ONTUA). It acts as a middle ground for

the NTA and OUA, since it selects the UE based on a threshold coefficient and it also selects the UEs that are assigned orthogonal pilot sequences. The method is described in Algorithm 7 Algorithm 7 shows that, by discarding the UEs that have channel gains below a given

Algorithm 7 Orthogonal Normalized Threshold Users Algorithm

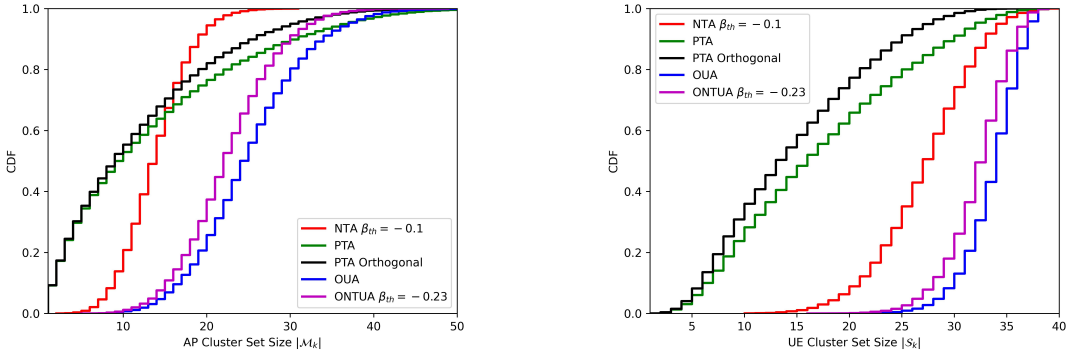
Input: Channel gains β_{kl}, Pilot Assignment vector $\mathbf{t}$
Output: DCC sets $\mathcal{M}_1, \mathcal{M}_2 ... \mathcal{M}_K$
for $l = 1 : L$ **do**
  for $t = 1 : \tau_p$ **do**
    $i \leftarrow \underset{k \in \{1,...,K\}: t_k = t}{\arg\max} \beta_{kl}$ ▷ Find UE that AP l serves on pilot t
    $\tilde{\beta}_{il} = \frac{\beta_{il} - \bar{\beta}_l}{\sqrt{\text{Var}(\beta_l)}}$ ▷ Calculate the normalized channel gain coefficient for UE i
    if $\tilde{\beta}_{il} \geq \beta_{\text{th}}$ **then**
      $\mathcal{M}_i \leftarrow \mathcal{M}_i \cup \{l\}$ ▷ If the normalized channel gain between UE i and AP l is above the threshold, connect i to l
    else
      $\mathcal{M}_i \leftarrow \mathcal{M}_i$
    end if
  end for
end for

threshold , that is, the UEs that have weak channel gains, one could reduce signal detection complexity, while keeping performance similar to OUA, since the discarded APs would have a low contribution to the UE performance if compared to the contribution of the connected APs.

Also, one should notice that the choice of the threshold value is rather different in this case than in the NTA case. On the latter, the threshold value would select the AP-UE pairs on the entire grid. On the ONTUA this value only selects the UEs that are connected to a given AP. Therefore, the ONTUA threshold value would be generally lower than the one of NTA, since we would expect a high percentage of τ_p to be connected to every AP, which would ensure we wouldn't discard a strong AP-UE pair. Still, Fig, 4.9 can still be used to select the threshold β_{th}. For example, if we check the reference case, in which $K = 40$, in order to have an average of 60% of the τ_p UEs connected to a given AP, we would need to select a value of $\beta_{\text{th}} = -0.23$. If the same value was used to the NTA, we would select on average 60% of the UEs for each AP, resulting in a much more unscalable system.

4.2.7 Complexity Metrics for DCC Algorithms over Known and Unknown Channel Statistics

In this section, we shift our focus to establish a comparison between the scalability of the proposed DCC methods. To do this, we use as metric the length of the AP cluster, $|\mathcal{M}_k|$, and



(a) DCC AP length $|\mathcal{M}_k|$ CDF

(b) Partially served UEs clusters set length $|\mathcal{S}_k|$

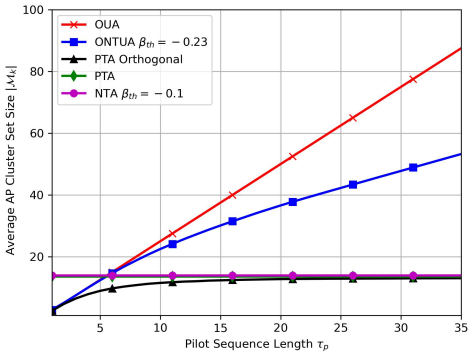
Figure 4.10 – CDF for the length of the DCC cluster sets of APs, $|\mathcal{M}_k|$ and the partially served UEs cluster, $|\mathcal{S}_k|$

we define

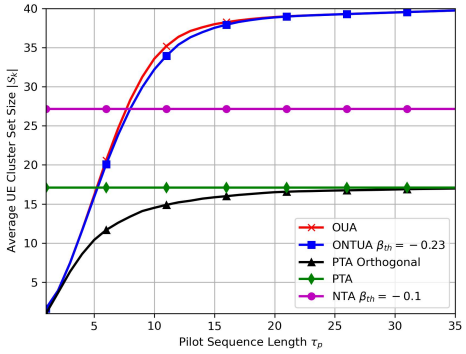
$$\mathcal{S}_k = \{i : \mathbf{D}_k \mathbf{D}_i \neq \mathbf{0}_{LN \times LN}\} \tag{4.24}$$

as the set of UEs that are partially served by the same set of APs that serve UE k. We now compare the DCC cardinality $\mathcal{M}_k$, defined in Definition 2.1 for each of the DCC algorithms analyzed. For the NTA, we consider $\beta_{\text{th}} = -0.1$, and for the ONTUA we shall consider $\beta_{\text{th}} = -0.23$. For the PTA, we make $\delta = 0.95$. We consider both the cases of PTA and Orthogonal PTA. The results are shown in Fig. 4.10. From Fig. 4.10, we can see that the PTA curve has the largest variance in the length of the clusters. This happens due to the large variability of the channel gains in the grid. Also, it has a high probability of having $|\mathcal{M}_k| = 1$, since this method selects the APs based on the contribution of their channel gains to the overall power of the desired UE, an AP that has a large channel gain to an UE will be very likely to carry almost all of the power of the UE, specially in the case the other APs have weak channel gains to that UE. At the same time, there will be cases where all of the APs are placed far away from the UE, configuring a situation of a large value of $|\mathcal{M}_k|$. The same can be applied to $|\mathcal{S}_k|$. In the case of the PTA Orthogonal, the results are a little bit different, since it disconnects from each AP the UEs that have the same pilot sequence assigned to them. However, in order to maintain the overall UE power high, not as much APs are disconnected, leading to results similar to those of regular PTA.

When we analyze the curve for NTA, we see that it has the smallest values of $|\mathcal{M}_k|$ among all curves. Since this threshold value according to Fig. 4.9 corresponds to a percentile lower to the ones shown in the figure. This means that less APs will be connected to the UEs, therefore reducing the values of both $|\mathcal{M}_k|$ and $|\mathcal{S}_k|$. In the case of $|\mathcal{S}_k|$, its values are bigger than the PTA because for the latter, a vast number of UEs will be served by only one AP, such that it will have fewer common serving APs with the other UEs. As expected, the results for the OUA and ONTUA curves are very similar, however the latter has smaller values due to the fact that some of the connected UEs of OUA are disconnected with ONTUA. Since the



(a) Average DCC AP length $|\mathcal{M}_k|$

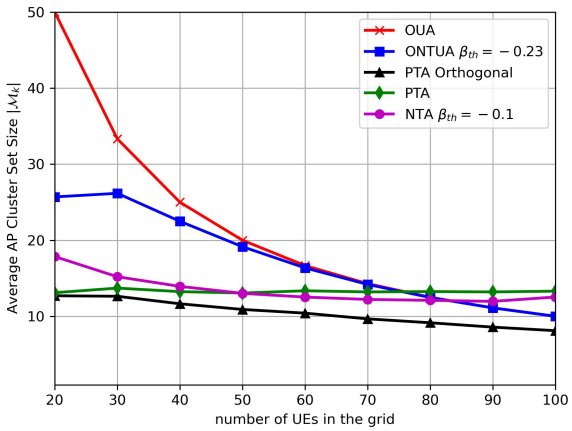


(b) Average Partially served UEs clusters set length $|\mathcal{S}_k|$

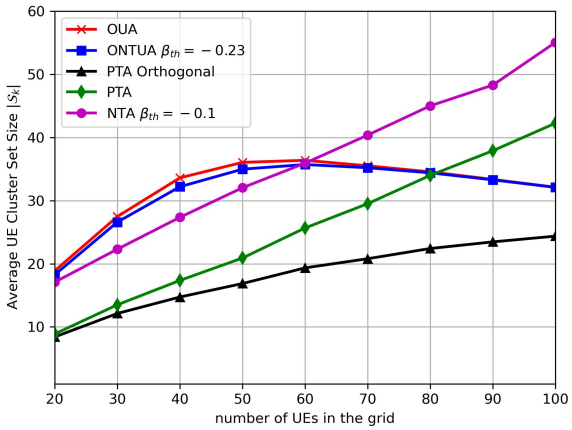
Figure 4.11 – Average length of the DCC cluster sets of APs, $|\mathcal{M}_k|$ and the partially served UEs cluster, $|\mathcal{S}_k|$ versus the pilot sequence length τ_p for the DCC methods of Section 4.2.

value of $\tau_p = 10$ is not as big, this difference is small, however we will see that for larger values of τ_p this gap gets significantly larger. For that, we repeat the simulation scheme that was used to generate Fig. 4.9, but instead we consider the effect of the pilot sequence length, to obtain the average value of the lengths of $\mathcal{M}_k$ and $\mathcal{S}_k$. The results can be seen in Fig. 4.11. Analyzing the lengths of $\mathcal{M}_k$, we notice that the NTA and the orthogonal-regular PTA do not have a strong influence of the pilot sequence length τ_p. That is, they remain constant irrespective of the value of τ_p, in the case of PTA Orthogonal, it has a small variation for small values of τ_p, therefore they are scalable with respect to τ_p. The OUA and ONTUA however, have a strong dependence on τ_p. The relationship between OUA and τ_p is such that it is linear, thus clearly unscalable. On the case of ONTUA, however, the length of $\mathcal{M}_k$ tends to stabilize around $L = 60$ for higher pilot sequence lengths, since the AP only selects the UEs that are orthogonal and have channel quality above the threshold. With respect to the lengths of $\mathcal{S}_k$, we can say again that OUA and ONTUA have a high dependence on τ_p while NTA and orthogonal-regular PTA do not. For the case of the PTA Orthogonal, it has a lower value of $|\mathcal{S}_k|$, because of the discarded APs. The NTA and ONTUA have a growing value of $|\mathcal{S}_k|$ until it gets stabilized at the value $K = 40$. This happens because, since the AP is connected to at maximum τ_p UEs, then as τ_p gets larger, the probability that every UE will have a common serving AP with other UEs gets bigger.

Despite the fact that the set of partially served UEs is maximized for larger τ_p for the ONTUA, which one might think could lead to unscalabilty, we will see in later chapters that it is the value of $|\mathcal{M}_k|$ that impacts the system complexity value the most. Therefore, opposed to OUA, which has larger values of $|\mathcal{M}_k|$, ONTUA remains scalable. Finally, we shall analyze the behavior of the algorithms for UE variations. We shall analyze a system of $K \in \{20, 30, ..., 100\}$ UEs. On real life applications, this is the scenery of greater relevance, since we shall analyze if the given DCC algorithms allow for the addition of more UEs to



(a) Average DCC AP length $|\mathcal{M}_k|$

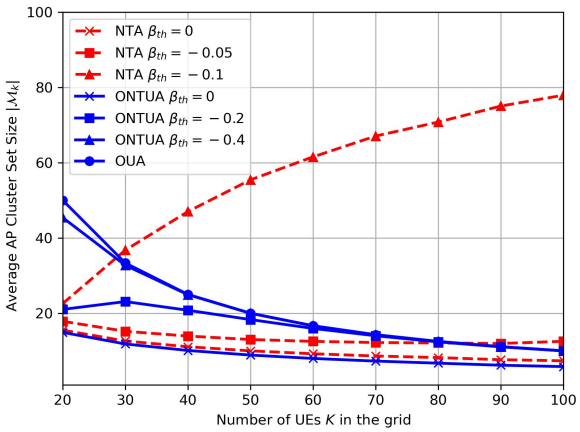


(b) Average Partially served UEs clusters set length $|\mathcal{S}_k|$

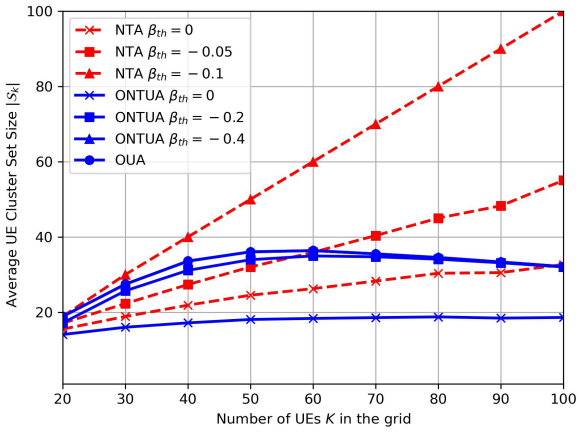
Figure 4.12 – Average length of the DCC cluster sets of APs, $|\mathcal{M}_k|$ and the partially served UEs cluster, $|\mathcal{S}_k|$ versus the total number of UEs in the grid, K, for the DCC methods of Section 4.2.

the grid without the need of much greater processing power by the APs. Fig 4.12 shows those results We can see from Fig. 4.12a, that the AP clusters size decrease in all of the cases, except the NTA and regular PTA. In the case of the latter, it actually increases with the addition of more UEs. This is explained by the fact that, with more UEs, the probability of them being placed at an "unfavorable" location, that is, a location at which all channel gains from the APs are weak, is increased, therefore more AP will be needed in order to obtain β_k^{δ} percent of the overall power, thus increasing the AP cluster size. As for the NTA, it remains constant because the addition of more UEs makes that some possess high channel gains to the AP while others posses lower channel gains, creating a balance such that the number of APs each UE is connected to stays virtually the same.

However, if we shift the focus for the partially served UE clusters, from Fig. 4.12b, we see that the ONTUA and OUA have an increasing mean value of $|\mathcal{S}_k|$ until $K = 60$, after



(a) Average DCC AP length $|\mathcal{M}_k|$



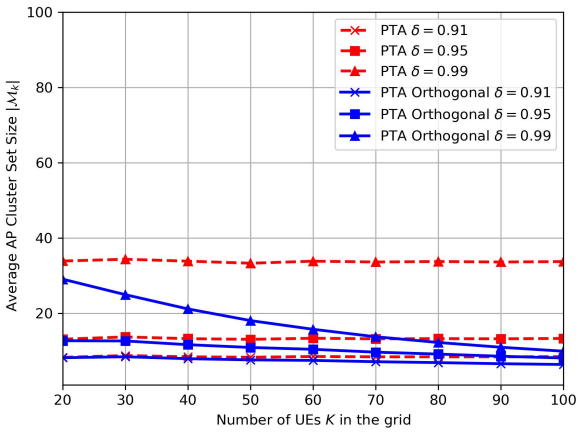
(b) Average Partially served UEs clusters set length $|\mathcal{S}_k|$

Figure 4.13 – Average length of the DCC cluster sets of APs, $|\mathcal{M}_k|$ and the partially served UEs cluster, $|\mathcal{S}_k|$ versus the total number of UEs in the grid, K, for the DCC NTA and ONTUA for different values of β_{th}. OUA is shown as a reference case.

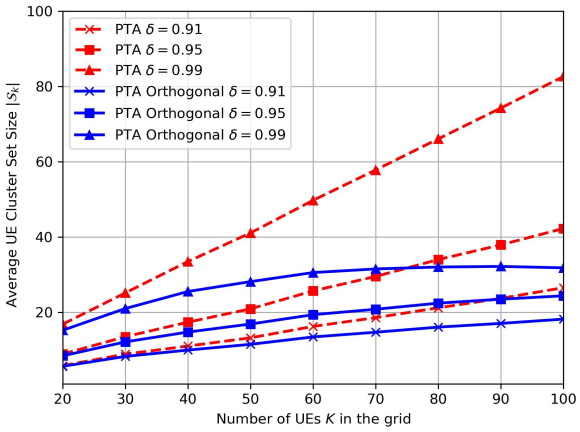
which it decreases. This happens because when the number of UEs is high, the number of partially served UEs is decreased, due to the fact each AP can serve only up to τ_p UEs. The same can be said for PTA Orthogonal, albeit it increases and decreases at a slower pace. For the case of PTA and NTA, since there are no such constraints of a maximum of UE per AP, the number of partially served UEs increase with the increasing value of K. This in turn, contributed to unscalability, as we will see later.

In order to further analyze the DCC algorithms clusters behavior, we now explore the variations of the threshold coefficient β_{th} for the NTA and ONTUA, and δ for the PTA and Orthogonal PTA. For that, we set $\beta_{\text{th}} \in \{0, -0.1, -0.2\}$ for the NTA and $\beta_{\text{th}} \in \{0, -0.2, -0.4\}$ for the ONTUA, as well as $\delta \in \{0.91, 0.95, 0.99\}$ for the PTA and Orthogonal PTA. The results can be seen in Figs. 4.13 and 4.14

We can see from Fig 4.13 that NTA is clearly unscalable for a high value of β_{th}, since



(a) Average DCC AP length $|\mathcal{M}_k|$



(b) Average Partially served UEs clusters set length $|\mathcal{S}_k|$

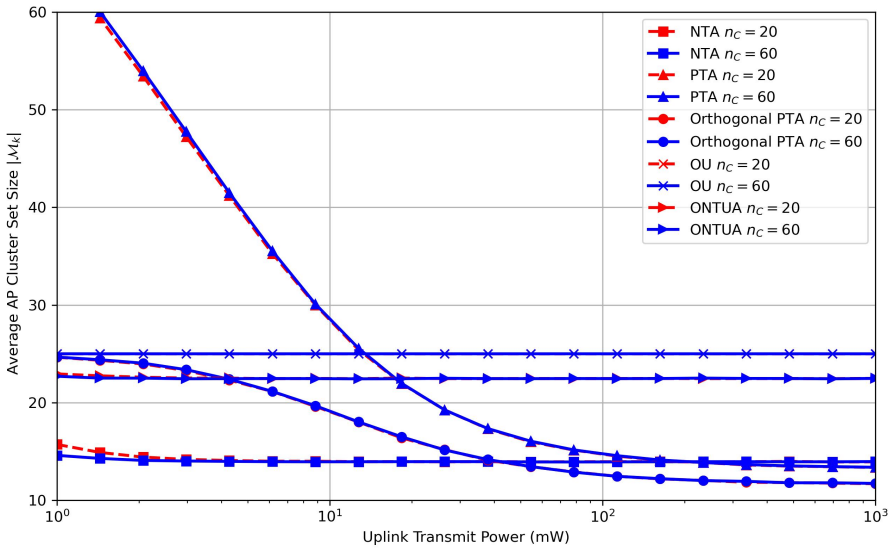
Figure 4.14 – Average length of the DCC cluster sets of APs, $|\mathcal{M}_k|$ and the partially served UEs cluster, $|\mathcal{S}_k|$ versus the total number of UEs in the grid, K, for the DCC PTA and Orthogonal PTA, by varying the parameter δ.

$|\mathcal{M}_k|$ increases with the number of UEs for $\beta_{\text{th}} = -0.2$. If we take a look at Fig. 4.9, we notice that this value of β_{th} is equivalent to say that the percentile of AP-UE pairs connected is 40 for $K = 30$, 50 for $K = 50$ and 70 for $K = 90$. That means that with the increasing value of K, the percentage of UEs that will be connected to a certain AP will grow. which will demand significantly more processing power of each AP, thus making the system unscalable. However, for values of β_{th} above -0.1, we have the percentiles curves shown in Fig. 4.9 such that the percentile of connected AP-UE pairs is below 40%, meaning that even though the number of UEs may rise, this value will keep the percent of connected APs under control. As for the case of ONTUA, the values of β_{th} are interpreted differently from the NTA. On the first case, $\beta_{\text{th}} = 0$ implies that a very small percentile of the UEs that are connected to each AP will be effectively connected to that AP. That is, the cluster sizes are smaller, since each UE is connected to few APs. However, as we lower the values of the threshold, the number of connected UEs for each AP is increases, up to a maximum of τ_p, making the clusters size bigger. However, since there is a limit to the number of connected UEs per AP, the system remains scalable as K increases. For instance, when $\beta_{\text{th}} = -0.4$, for $K \geq 30$, more than 90% of the UEs will be connected to each AP, which is very similar to OUA. As K increases this percent approaches its maximum value, and ONTUA and OUA achieve the same result. Therefore, it is of utmost importance to select a value of threshold such that it connects on average the strongest UEs while removing the weaker ones. Intuitively, we would say that the 60-th percentile coefficient would be adequate, hence the chosen value of $\beta_{\text{th}} = -0.23$ for $K = 40$, as seen in Fig. 4.9.

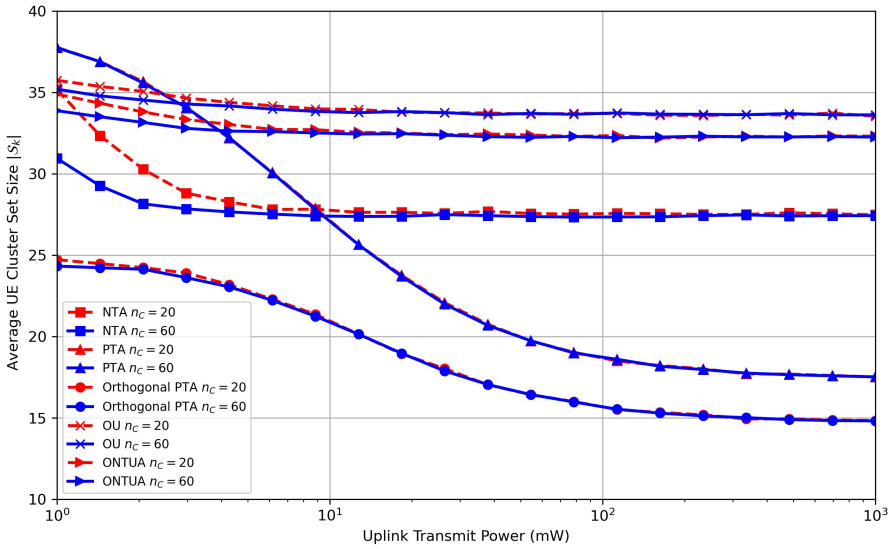
By now taking a look at Fig. 4.14, we can see that the mean PTA AP cluster size is independent of K, which would assume unscalability, as the cluster size should reduce with K. When we shift our focus to the UE cluster size, this becomes even more apparent, since the size scale linearly with K, with different inclinations, but linearly nonetheless. For Orthogonal PTA, however, since each AP can only be connected to UE that have orthogonal pilot sequences, at most τ_p UEs could be connected to a given AP. Therefore, the system is always scalable with the increasing values of K, as expected.

Finally, we shall explore the behavior of the DCC methods in the presence of imperfect channel statistics. Since the methods are highly dependent on the channel gains β_{kl}, not only for DCC formation, but as well as for pilot assignment, it is important to analyze if the DCC algorithms can still be applied with those constraints. For that, we let $\beta_{\text{th}} = -0.1$ for the NTA case and -0.23 for the ONTUA, $\delta = 0.95$ for both PTA and Orthogonal PTA, and we use $n_t = 5$ and $n_C \in \{20, 60\}$. The UL transmit power is varied from 1 mW to 1 W

Fig 4.15 shows that the cluster size variations are rather severe for the PTA and Orthogonal PTA. This happens because, since noise influence is stronger for lower UL transmit power, the spatial correlation matrix cannot be accurately estimated, and therefore the estimated channel gains are rather different from the actual channel gains. Since those algo-



(a) Average DCC AP length $|\mathcal{M}_k|$



(b) Average Partially served UEs clusters set length $|\mathcal{S}_k|$

Figure 4.15 – Average length of the DCC cluster sets of APs, $|\mathcal{M}_k|$ and the partially served UEs cluster, $|\mathcal{S}_k|$ versus the UL transmit power p in mW, for the DCC methods of Section 4.2.

rithms rely on gathering the APs that contribute to a given percentage of the UE total power, and the estimated channel gains are likely to have small values, due to the LS channel estimation being mainly composed of noise, it is likely that more APs will be needed in order for each UE to reach the desired power constraint. In the case of Orthogonal PTA, some APs are disconnected, and therefore the variations are smaller. However, like PTA, those variations are only stabilized once the UL power p is close to 100 mW. For the NTA case, since the main factor for the clustering is the normalization of the channel gains, the relative difference between their magnitudes is more relevant than the actual overall magnitude. Therefore, it remains constant irrespective of the value of p. The same can be said for ONTUA and OUA, since they do not rely on the actual order of magnitude of the channel gains, but on the relative difference between their magnitudes, they are constant with respect to p. We can also see that the number of coherence blocks used for estimation, n_C, offers insights only when p is low. In that case, since the noise is dominant, more samples would be needed in order to mitigate its effects, leading to the offset that can be seen between the curves of $n_C = 20$ and $n_C = 60$ in Fig. 4.15b.

4.3 A SUMMARY OF THE ALGORITHMS STUDIED ON THIS CHAPTER

Throughout this chapter, various pilot assignments and DCC algorithms were introduced. Table 4.2 shows the reference of which every algorithm was taken, as well as their advantages and disadvantages with respect to the NMSE and DCC cluster sizes

With respect to the variance of the cluster sizes in PTA algorithms, we will see in later chapters that it implies that some UEs are not able to achieve higher values of SE as other lower variance algorithms such as NTA and ONTUA. In the case of the RP, the distance metric of (4.11) was initially proposed by [35], the euclidean distance of channel gains metric of Eq (4.11) was proposed by [15], and the correlation metrics of Eq. (4.13) and Eq. (4.14) were proposed by this work.

Algorithm	Reference	Advantage	Disadvantage
RPA		Does not require coordination between the APs	Leads to a high degree of pilot contamination
RP	[35]	Besides reducing pilot contamination, assigns the same number of UEs to one pilot	Only usable if the number of UEs is a multiple of τ_p
GPA	[8]	Usable for any combination of UEs and pilot sequences	Has a lower NMSE if compared to RP
NTA	[15] Proposed	Scalable with respect to τ_p	Unscalable with respect to K
PTA	[16]	Scalable with respect to τ_p	Unscalable with respect to K and large variance of the cluster sizes
Orthogonal PTA	Proposed	Scalable with respect to τ_p and K	Large variance of the cluster sizes
OUA	[8]	Scalable with respect to K	Unscalable with respect to τ_p
ONTUA	Proposed	Scalable with respect to K and τ_p	Small performance loss compares to OUA. This loss will be covered in Chapter 6

Table 4.2 – Proposed algorithms of Chapter 4

Chapter Summary

- Pilot assignment algorithms are essential when the objective is to increase the NMSE by reducing pilot contamination
- We proposed three such algorithms: RPA, RP and GPA. RPA does not require AP coordination, but has the highest values of channel NMSE due to pilot contamination. RP has the lowest values of NMSE at the cost of being limited to situations where K is a multiple of τ_p. GPA has a slightly higher NMSE than RP, but it can be used with any τ_p-K relationship.
- With respect to the AP clustering algorithms we have proposed five algorithms: NTA, PTA, Orthogonal PTA, OUA and ONTUA
- NTA has the advantage of having cluster sizes invariant to the pilot sequence

length τ_p. It has, however, the disadvantage of having increasing cluster sizes with the variation of the UEs in the grid, K.

- PTA and Orthogonal PTA have the advantage of guaranteeing a minimum power percent value for each UE, which in turn guarantees a minimum performance. They are both scalable with τ_p, however PTA is unscalable with K.
- OUA and ONTUA eliminate pilot contamination between the UEs connected to the same AP, since they have orthogonal pilots assigned to them. They are both scalable with K. OUA, however is unscalable with τ_p. ONTUA is scalable with τ_p at the expense of performance, as we will see in later chapters.
- In a situation of imperfect channel statistics, NTA, ONTUA and OUA can be still used with satisfactory performance. PTA and Orthogonal PTA cannot be used under weak spatial correlation matrices estimations, since they are highly dependent on the actual value of the channel gains, and become increasingly unscalable the worse the estimation quality.

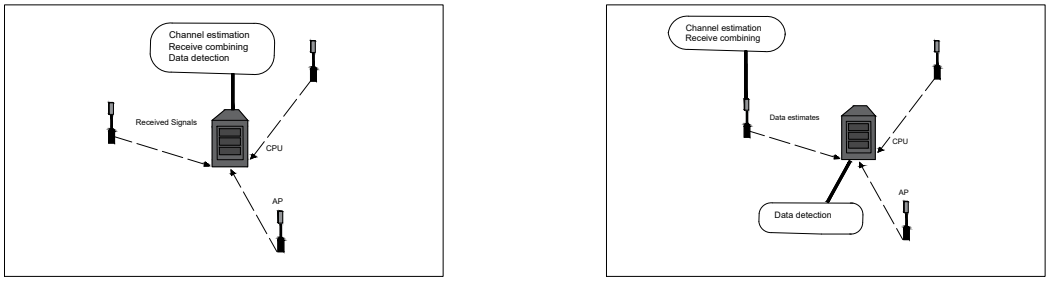
5 CELL FREE MASSIVE MIMO UPLINK SPECTRAL EFFICIENCY AND COMPLEXITY OVER CENTRALIZED AND DISTRIBUTED SYSTEMS

The goal of this chapter is to introduce the centralized and distributed operation for UL. Section 5.1 introduces the centralized operation and the SE for various combining vectors methods. Section 5.2 introduces the distributed operation and the SE for various combining vectors methods and introduces the concept of Large Scale Fading Decoding, used to enhance system performance. Section 5.3 summarizes the complexity comparison for various UL data detection methods.

Having introduced the DCC and the pilot assignment algorithms, we now shift our focus to the performance metrics. We shall evaluate the SE of the systems as a performance metric. since as already discussed, the SE measures the amount of information that can be transmitted over a fixed bandwidth.

When we are analyzing the UL, two implementations, all of which are characterized by the degree of cooperation among the APs, are suggested: the *centralized operation* and the *distributed operation*. The centralized operation consists on gathering all of the information received by the APs, by means of the fornthaul links, at a central unit, the CPU. The CPU then performs channel estimation and data detection. The distributed operation consists on gathering the information received by an arbitrary AP, and at this AP, perform channel estimation and data detection. Fig. 5.1 shows the distributed and centralized operations

Regardless of the operation, channel estimation will always be performed by use of MMSE, and data detection will make use of receive combining techniques, such as the optimal receiver (which we will see, is unscalable), and the scalable receivers.



(a) Centralized Operation

(b) Distributed Operation

Figure 5.1 – Centralized and distributed UL operations. The centralized system performs channel estimation, receive combining and data detection at the CPU, while distributed operation performs everything but data detection at the AP.

5.1 CENTRALIZED OPERATION

The centralized UL operation of CF MIMO is the most advanced implementation of studied systems, where the AP serves merely as remote radio heads (RRHs) that sends their received baseband signals to the CPU, where it is actively processed. An arbitrary AP l sends the τ_p received pilot signals, $\mathbf{y}_{tl}^{\text{pilot}}$, as defined in Chapter 3, and the received UL data signal $\mathbf{y}_l^{\text{ul}}$, defined in Chapter 2 to the CPU, that in turn performs channel estimation, receive combining and data detection.

At the CPU, for each UE k, the pilot signals are used to compute the partial MMSE estimates of the collective channels from all the UEs to the APs that serve UE k, that is $l \in \mathcal{M}_k$. The received signals at the serving APs are then jointly used to compute an estimate $\hat{s}_k$ of the signal s_k that is transmitted by UE k, or

$$\begin{aligned} \hat{s}_k &= \sum_{l=1}^{L} \hat{s}_{kl} \\ &= \sum_{l=1}^{L} \mathbf{v}_{kl}^H \mathbf{D}_{kl} \mathbf{y}_L^{\text{ul}} = \mathbf{v}_k^H \mathbf{D}_k \mathbf{y}^{\text{ul}}, \end{aligned} \tag{5.1}$$

with $\mathbf{v}_k = \left[\mathbf{v}_{k1}^T \dots \mathbf{v}_{kL}^T\right]^T \in \mathbb{C}^{LN}$ as the centralized combining vector and $\mathbf{y}^{\text{ul}} \in \mathbb{C}^{LN}$ as the collective data UL signal, given as

$$\mathbf{y}^{\text{ul}} = \begin{bmatrix} \mathbf{y}_1^{\text{ul}} \\ \vdots \\ \mathbf{y}_L^{\text{ul}} \end{bmatrix} = \sum_{i=1}^{K} \mathbf{h}_i s_i + \mathbf{n} \tag{5.2}$$

with $\mathbf{n} = \left[\mathbf{n}_1^T \dots \mathbf{n}_L^T\right]^T \in \mathbb{C}^{LN}$ being the noise collective vector. The received UL signal in

Eq. (5.2) is composed of all of the transmitted UEs signals, that is, its dependent on all of the UEs channels $\mathbf{h}_i$, Despite that, when computing the combiner vectors, the CPU will only take into consideration the channels of the APs such that $l \in \mathcal{M}_k$, by means of the collective $\mathbf{D}_k$ matrix. Therefore, the CPU does not necessarily need to estimate all of the channels from the UEs to the APs, but rather only those APs that are on the set $\mathcal{M}_k$.

Also from Eq. (5.2), we see that the expression is equivalent to the single cell Massive MIMO with correlated fading, if the CPU is treated as a receiver equipped with $M = LN$ antennas. However, there are some key differences between those setups:

1. The pilot sequences are frequently reused by the UEs managed by the CPU, which is generally not the case in single cell Massive MIMO

2. The APs are distributed over distinct geographical locations. Thus, the collective channel is distributed as $\mathbf{h}_k \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_{LN}, \mathbf{R}_k)$, and the spatial correlation matrix $\mathbf{R}_k = \text{diag}\left(\mathbf{R}_{k1}, \ldots, \mathbf{R}_{kL}\right) \in \mathbb{C}^{LN \times LN}$ has a diagonal structure , which is not the case for the single cell Massive MIMO, where the spatial correlation matrix is usually not diagonally structured.

3. Not all of the antennas of the system are being used for signal detection, only those of the serving APs of UE a particular UE k.

Regardless of the differences between the architectures, we can still compute an achievable SE using the methods described in [18] and [23]. All of the methods will assume the CPU has perfect or imperfect knowledge of the spatial correlation matrices of every UE $\{\mathbf{R}_k : k = 1, \ldots, K\}$.

5.1.1 Spectral Efficiency for Centralized Operation

By using the capacity lower bounds that were defined in Chapter 3, Section 3.2, we can derive a SE expression under imperfect CSI. To do that, we consider the estimated symbol expression of Eq. (5.1) and use Eq. (5.2) to obtain

$$\hat{s}_k = \underbrace{\mathbf{v}_k^H \mathbf{D}_k \hat{\mathbf{h}}_k s_k}_{\text{Desired signal over estimated channel}} + \underbrace{\mathbf{v}_k^H \mathbf{D}_k \tilde{\mathbf{h}}_k s_k}_{\text{Desired signal over unknown channel}} + \underbrace{\sum_{\substack{i=1 \\ i \neq k}}^{K} \mathbf{v}_k^H \mathbf{D}_k \mathbf{h}_i s_i}_{\text{Interference}} + \underbrace{\mathbf{v}_k^H \mathbf{D}_k \mathbf{n}}_{\text{Noisel}}. \tag{5.3}$$

The desired parcel of the estimated symbol was divided into one parcel of the estimated channel and another of the unknown channel. Only the former can effectively be used for

signal detection, while the latter will be treated as noise. Additionally, according to Section 3.2, the interference should also be treated as noise. With that in mind, we can announce Theorem 5.1

Theorem 5.1

An achievable SE of an arbitrary UE k, in the UL centralized operation is given by

$$\text{SE}_k^{\text{c}} = \frac{\tau_u}{\tau_c}\mathbb{E}\left\{\log_2\left(1+\text{SINR}_k^{\text{c}}\right)\right\} \quad \text{bit/s/Hz} \tag{5.4}$$

with the instantaneous effective SINR given by

$$\text{SINR}_k^{\text{c}} = \frac{p_k\left|\mathbf{v}_k^H\mathbf{D}_k\hat{\mathbf{h}}_k\right|^2}{\sum_{\substack{i=1\\i\neq k}}^{K} p_i\left|\mathbf{v}_k^H\mathbf{D}_k\hat{\mathbf{h}}_i\right|^2 + \mathbf{v}_k^H\mathbf{Z}_k\mathbf{v}_k + \sigma^2\left\|\mathbf{D}_k\mathbf{v}_k\right\|^2} \tag{5.5}$$

where

$$\mathbf{Z}_k = \sum_{i=1}^{K} p_i\mathbf{D}_k\mathbf{C}_k\mathbf{D}_k \tag{5.6}$$

with the expectation with respect to the channel estimates. The matrix $\mathbf{C}_i$ is the error correlation matrix of channel $\mathbf{h}_i$.

Proof 5.1: The proof can be found on Appendix A.3. □

The pre-log factor τ_c/τ_c is the fraction of the symbols that are being used to transmit data. The SINR is composed of a numerator that corresponds to the desired signal over the estimated channels, and a denominator that consists on interference from the other UEs, from the unknown channel in the form of $\mathbf{Z}_k$, and from noise. This SE expression bears a similarity to the single antenna fading expression derived in Lemma 3.3, showing the data signal can be encoded and the receiver signal decoded as if communication occurred over a traditional AWGN channel.

The SE expression is not on closed form, due to the expected value operator in it. However, it can be easily computed by means of Monte Carlo simulations, meaning it can be approximated by an average over a large number of realizations. In other words, we could generate a handful of channel realizations over various sets of coherence blocks, and then average the result to obtain an approximation for the SE.

The derived Theorem 5.1 is general in the sense that it serves the purpose of systems with multiple antennas APs with correlated or uncorrelated fading channels of an arbitrary

distribution, as well as any arbitrary DCC setup. Multiple authors have considered the case of $N = 1$, such as [36], [37] and [38]. The case of multiple antennas with uncorrelated fading has been analyzed by authors such as [39] and [40] and with correlated fading having being analyzed among others by [41] and [13]. We will analyze the system by use of this expression for various DCC methods, showing how versatile it is. In the unscalable case, where all of the APs serve all the UEs, the SINR expression of Eq. (5.5) can be rewritten as

$$\text{SINR}_k^{\text{c}} = \frac{p_k \left| \mathbf{v}_K^H \hat{\mathbf{h}}_k \right|^2}{\sum_{\substack{i=1 \\ i \neq k}}^{K} p_i \left| \mathbf{v}_K^H \hat{\mathbf{h}}_i \right|^2 + \mathbf{v}_k^H \left(\sum_{i=1}^{K} p_i \mathbf{C}_i + \sigma^2 \mathbf{I}_{LN} \right) \mathbf{v}_k}, \tag{5.7}$$

which is a simplified version.

5.1.2 Optimal Receive Combining

We want to find the combining vector $\mathbf{v}_K$ that maximizes the SE in Eq. (5.4). For that, we notice that Eq. (5.5) can be written in such a manner that it has the form of a generalized Rayleigh quotient:

$$\text{SINR}_k^{\text{c}} = \frac{p_K \left| \mathbf{v}_k^H \mathbf{D}_k \hat{\mathbf{h}}_k \right|^2}{\mathbf{v}_k^H \left(\sum_{\substack{i=1 \\ i \neq k}}^{K} p_i \mathbf{D}_k \hat{\mathbf{h}}_i \hat{\mathbf{h}}_i^H \mathbf{D}_k + \mathbf{Z}_k + \sigma^2 \mathbf{D}_k \right) \mathbf{v}_k}. \tag{5.8}$$

With that in mind, the optimal combining vector is as it follows:

COROLLARY 5.1 The instantaneous SINR of Eq. (5.4) is maximized by the MMSE combining vector

$$\mathbf{v}_k^{\text{MMSE}} = p_k \left(\sum_{i=1}^{K} p_i \mathbf{D}_k \left(\hat{\mathbf{h}}_i \hat{\mathbf{h}}_i^H + \mathbf{C}_i \right) \mathbf{D}_k + \sigma^2 \mathbf{I}_{LN} \right)^{-1} \mathbf{D}_k \hat{\mathbf{h}}_k \tag{5.9}$$

that yields the maximum SINR value

$$\text{SINR}_k^{\text{c}} = p_k \hat{\mathbf{h}}_k^H \mathbf{D}_k \left(\sum_{\substack{i=1 \\ i \neq k}}^{K} p_i \mathbf{D}_k \hat{\mathbf{h}}_i \hat{\mathbf{h}}_i^H \mathbf{D}_k + \mathbf{Z}_k + \sigma^2 \mathbf{I}_{LN} \right)^{-1} \mathbf{D}_k \hat{\mathbf{h}}_k \tag{5.10}$$

Proof 5.2: The proof begins by defining the generalized Rayleigh quotient

$$\frac{\left|\mathbf{v}^H\mathbf{h}\right|^2}{\mathbf{v}^H\mathbf{B}\mathbf{v}} \tag{5.11}$$

and by adopting $\mathbf{v} = \mathbf{v}_k$, $\mathbf{h} = \sqrt{p_k}\mathbf{D}_k\hat{\mathbf{h}}_k$ and $\mathbf{B} = \sum_{i=1,i\neq k}^{K} p_i\mathbf{D}_k\hat{\mathbf{h}}_i\hat{\mathbf{h}}_i^H\mathbf{D}_k + \mathbf{Z}_k + \sigma^2\mathbf{I}_{LN}$. Then by the Rayleigh quotient maximization principle, the proof is complete. □

The optimal combining vector is composed of 1) the estimates $\mathbf{D}_k\hat{\mathbf{h}}_k$ of the k-th UE channels, and 2) the inverse of a matrix that is correspondent to $\mathbb{E}\left\{\mathbf{D}_k\mathbf{y}^{\text{ul}}(\mathbf{y}_{\text{ul}})^H\mathbf{D}_k|\{\hat{\mathbf{h}}_i\}\right\}$, the conditional correlation matrix of the received signal, given the channel estimates. The first part is responsible to maximize the desired signal power, while the second is intended to strike a balance between maximizing the desired signal power and suppressing interference, when the combining vector is multiplied by the received signal. [8] shows that this is equivalent to a spatial whitening filter. Moreover, we notice that $\mathbf{D}_k\hat{\mathbf{h}}_i$ is always nonzero, except when $\mathbf{D}_k = \mathbf{0}_{LN\times LN}$. This shows that the CPU needs to compute all of the channel estimates $\left\{\hat{\mathbf{h}}_i : i = 1,\ldots,K\right\}$ corresponding to any AP l that is serving UE k. The total number of complex multiplications for channel estimations and for computing the combining vector over one coherence block are taken from [22] and are given by

$$C_m^{\text{MMSE}} = \underbrace{(N\tau_p + N^2)K|\mathcal{M}_k|}_{\text{channel estimation}} + \underbrace{\frac{(N|\mathcal{M}_k|)^2 + N|\mathcal{M}_k|}{2}K + (N|\mathcal{M}_k)^2 + \frac{(N|\mathcal{M}_k|)^3 - N|\mathcal{M}_k|}{3}}_{\text{combining vector}}. \tag{5.12}$$

From Eq. (5.12) we can see that the complexity grows linearly with K, which in turn makes the system unscalable, according to the scalability remarks of Chapter 2 . We shall then look for alternative combining vectors.

5.1.3 Scalable Receiver Combiners for Centralized Operation

We shall now introduce alternative combining methods that aim to guarantee scalability to centralized systems.

5.1.3.1 Maximal Ratio Combiner

We have already used the unscalable MR combiner, which consists only on the MMSE channel estimate of the collective UE channel. To keep it scalable, we only need to estimate the channels that are effectively used by UE k under DCC. Therefore, the scalable MR

combiner is given by

$$\mathbf{v}_k^{\text{mr}} = \mathbf{D}_k \hat{\mathbf{h}}_k \tag{5.13}$$

which in turn maximizes the numerator of the instantaneous SINR expression of Eq. (5.4), thus maximizing the desired signal power. The existence of interference and noise, however, is neglected, which makes the performance poor in the presence of strong interference, since not all of the APs will have a high degree of favorable propagation to UE k. In regard to the number of complex multiplications, since the combining vector is composed of the channel estimates, it is necessarily only to estimate the channel in order to compute it. Therefore,

$$C_m^{\text{MR}} = \underbrace{(N\tau_p + N^2)|\mathcal{M}_k|}_{\text{channel estimation}}. \tag{5.14}$$

This is the least complex combining method, however also the one that has the lowest performance. Thus, we shall introduce two other combining methods that are scalable and also achieve a better performance in the presence of interference.

5.1.3.2 Partial MMSE Combining

In large networks, it is plausible to assume that almost all of the interference that affects UE k is generated by a subset of UEs that are nearby k. We can use this information to reduce the complexity of the optimal MMSE combiner and provide a scalable solution. Therefore, we make only the UEs that are served partially by the same APs as UE k be included in the inverse matrix of Eq. (5.9), that is, the UEs that belong to the set $\mathcal{S}_k$, as defined in Chapter 4. By use of this set, an alternative partial MMSE (P-MMSE) scheme can be defined:

$$\mathbf{v}_k^{\text{P-MMSE}} = p_k \left(\sum_{i \in \mathcal{S}_k} p_i \mathbf{D}_k \hat{\mathbf{h}}_i \hat{\mathbf{h}}_i^H \mathbf{D}_k + \mathbf{Z}_{\mathcal{S}_k} + \sigma^2 \mathbf{I}_{LN} \right)^{-1} \mathbf{D}_k \hat{\mathbf{h}}_k \tag{5.15}$$

with

$$\mathbf{Z}_{\mathcal{S}_k} = \sum_{i \in \mathcal{S}_k} p_i \mathbf{D}_k \mathbf{C}_i \mathbf{D}_k. \tag{5.16}$$

Contrary to the optimal combiner, this scheme does not aim to achieve an optimal SE value, but rather to obtain a more scalable solution while maintaining a reasonable system performance. Also, in the situation where only the UEs in the set $\mathcal{S}_k$ are active, the optimal MMSE and the P-MMSE combiners are the same. In regard to the number of complex multiplications, we can also resort to [22] to obtain

$$C_m^{\text{P-MMSE}} = \underbrace{(N\tau_p + N^2)|\mathcal{S}_k||\mathcal{M}_k|}_{\text{channel estimation}} + \underbrace{\frac{(N|\mathcal{M}_k|)^2 + N|\mathcal{M}_k|}{2}|\mathcal{S}_k| + (N|\mathcal{M}_k|)^2 + \frac{(N|\mathcal{M}_k|)^3 - N|\mathcal{M}_k|}{3}}_{\text{combining vector}}. \tag{5.17}$$

Eq. (5.17) shows the importance of what was discussed in Chapter 4 related to the cluster sets $\mathcal{M}_k$ and $\mathcal{S}_k$, since those are the key values that define system complexity. Since the expression is independent of the number of UEs K, it might be scalable, depending on the value of $|\mathcal{S}_k|$. Regardless, it is more scalable than its optimal counterpart, and we will see that it also exhibits similar performance. We shall introduce another combining method that aims to reduce complexity even more without too much lowering of system performance.

5.1.3.3 Partial Regularized Zero-Forcing Combining

The complexity of computing the MMSE and P-MMSE combiners scale with the term $N|\mathcal{M}_k|$. Since in CF networks, the number of AP antennas is usually smaller than single cell Massive MIMO, the value of N is generally not an issue. On the other hand, $|\mathcal{M}_k|$ might be large when there are a large number of APs nearby UE k, or for example, when a large number of APs is needed to satisfy a given power threshold, such as the one discussed in Chapter 4. Therefore, it is interesting to explore if further complexity reduction is achievable without compromising performance as much. For that, we notice that if the channel conditions of all of the interfering UEs in $\mathcal{S}_k$ are good, then the estimation error correlation matrices are negligible, thus we can neglect the term $\mathbf{Z}_{\mathcal{S}_k}$, to obtain the partial regularized zero forcing (P-RZF) combiner

$$\mathbf{v}_k^{\text{P-RZF}} = p_k \left(\sum_{i \in \mathcal{S}_k} p_i \mathbf{D}_k \hat{\mathbf{h}}_i \hat{\mathbf{h}}_i^H \mathbf{D}_k + \sigma^2 \mathbf{I}_{LN} \right)^{-1} \mathbf{D}_k \hat{\mathbf{h}}_k. \tag{5.18}$$

This change enables a reformulation of the expression. For instance, let $\hat{\mathbf{H}}_{\mathcal{S}_k} \in \mathbb{C}^{LN \times |\mathcal{S}_k|}$ be the stacked collective channel vectors $\hat{\mathbf{h}}_i$ with indices $i \in \mathcal{S}_k$. in which the first column is $\hat{\mathbf{h}}_k$. Let us also define $\mathbf{P}_{\mathcal{S}_k} \in \mathbb{R}^{|\mathcal{S}_k| \times |\mathcal{S}_k|}$ a diagonal matrix containing the transmit powers p_i for $i \in \mathcal{S}_k$, following the order of the columns of $\hat{\mathbf{H}}_{\mathcal{S}_k}$. Then we can rewrite Eq. (5.18) as

$$\begin{aligned}
\mathbf{v}_k^{\text{P-RZF}} &= \left[\left(\mathbf{D}_k \hat{\mathbf{H}}_{\mathcal{S}_k} \mathbf{P}_{\mathcal{S}_k} \hat{\mathbf{H}}_{\mathcal{S}_k}^H \mathbf{D}_k + \sigma^2 \mathbf{I}_{LN} \right)^{-1} \mathbf{D}_k \hat{\mathbf{H}}_{\mathcal{S}_k} \mathbf{P}_{\mathcal{S}_k} \right]_{[:,1]} \\
&= \left[\mathbf{D}_k \hat{\mathbf{H}}_{\mathcal{S}_k} \mathbf{P}_{\mathcal{S}_k} \left(\hat{\mathbf{H}}_{\mathcal{S}_k}^H \mathbf{D}_k \mathbf{D}_k \hat{\mathbf{H}}_{\mathcal{S}_k} \mathbf{P}_{\mathcal{S}_k} + \sigma^2 \mathbf{I}_{|\mathcal{S}_k|} \right)^{-1} \right]_{[:,1]} \\
&= \left[\mathbf{D}_k \hat{\mathbf{H}}_{\mathcal{S}_k} \left(\hat{\mathbf{H}}_{\mathcal{S}_k}^H \mathbf{D}_k \hat{\mathbf{H}}_{\mathcal{S}_k} + \sigma^2 \mathbf{P}_{\mathcal{S}_k}^{-1} \right)^{-1} \right]_{[:,1]}
\end{aligned} \tag{5.19}$$

where the $[.]_{[:,1]}$ denotes the operation of taking only the first column of the matrix. The name P-RZF comes from the fact the expression in Eq. (5.19) has similar format as the pseudo-inverse $\mathbf{D}_k \hat{\mathbf{H}}_{\mathcal{S}_k} \left(\hat{\mathbf{H}}_{\mathcal{S}_k}^H \mathbf{D}_k \hat{\mathbf{H}}_{\mathcal{S}_k} \right)^{-1}$ of the partial channel matrix $\hat{\mathbf{H}}_{\mathcal{S}_k}^H \mathbf{D}_k$. The name regularized comes from the fact the pseudo-inverse has been regularized by the matrix that contains the noise variance and the transmit powers. Thus, it does not suffer the same losses

that traditional zero-forcing algorithms, such as the one described in [17] does. Thus, it forces the interference between the UEs to zero, which might seem an advantage. However, in situations where the UEs have similar channels this cancellation might lead to large losses of the desired signal power, as it will be interpreted as interference by the combiner.

With respect to the number of complex multiplications, [22] shows that it is the following

$$C_m^{\text{P-RZF}} = \underbrace{(N\tau_p + N^2)|\mathcal{S}_k||\mathcal{M}_k|}_{\text{channel estimation}} + \underbrace{\frac{|\mathcal{S}_k|^2 + |\mathcal{S}_k|}{2} N|\mathcal{M}_k| + |\mathcal{S}_k|^2 + |\mathcal{S}_k|N|\mathcal{M}_k + \frac{|\mathcal{S}_k|^3 - |\mathcal{S}_k|}{3}}_{\text{combining vector}}. \tag{5.20}$$

This expression is also independent of K, which might indicate scalability. Besides that, we can see that Eq. (5.20) grows more rapidly with $|\mathcal{S}_k|$ instead of $N|\mathcal{M}_k|$ as would the P-MMSE. Since $|\mathcal{S}_k|$ is usually much lower than $N|\mathcal{M}_k|$ in CF systems, the complexity is greatly reduced. This comes with a cost of SE as already discussed, due to the cancellation nature of the pseudo-inverse matrix.

5.1.4 Combining Vector Performance Comparison

Having introduced the centralized operation combining vectors, we now numerically compare the performance of all of the combiners. In order to observe a cleaner influence of the combiners without the bias of DCC, we shall consider the unscalable case where all APs serve all UEs. Also, for pilot assignment we resort to RP, from Chapter 4, with the correlation metric. Finally, for simulation purposes we shall downsize the parameters of Table 4.1, because of the unscalability, which would require too much resources and time to simulate the system with those parameters. Instead, we consider $K = 20$, $L = 50$, $\tau_p = 5$ and $\tau_c = 100$. The CDF of the combining vectors can be seen in Fig. 5.2. We notice that the performances of MMSE and P-MMSE are the same. This happens because under the assumption of every AP serving every UE, the subset $\mathcal{M}_k$ is such that $\mathcal{M}_k = l \in \{1, 2, \ldots, L\}\, \forall k$, and also $\mathcal{S}_k = k \in \{1, 2, \ldots, K\}\, \forall k$, so that all of the APs are counted as partially serving all UEs, since all of them are serving the same set of UEs. Therefore, performances are the same.

In regard to signal detection quality, MR combining has the worst performance since al already discussed, it requires a high degree of favorable propagation to work properly, as it lacks the capacity of handling interference. When we analyze P-MMSE and P-RZF, we can see that there is a gap in performance, since the P-RZF combining vector neglects the estimation error correlation matrices, which in turn makes it unable to distinguish what is the desired signal and what is interference when their channel conditions are similar. When we talk about system complexity, we are using the unscalable solution, $|\mathcal{M}_k| = L$ and $|\mathcal{S}_k| = K$ $\forall k$. We shall make comparisons with the various pilot assignments and DCC methods later

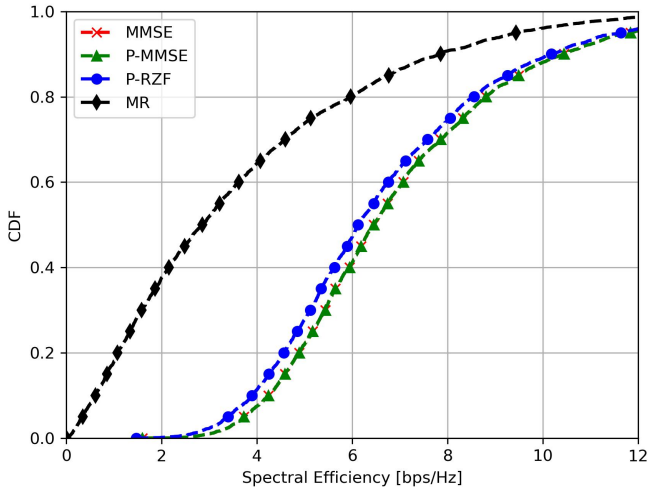


Figure 5.2 – SE for the CF UL system where all of the APs serve all UEs. The combining vectors P-MMSE,MMSE, MR and P-RZF are used.

in this chapter.

5.2 DISTRIBUTED OPERATION

The centralized operation delegates channel estimation, computation of the combining vectors and the data detection to the CPU. Distributed operation aims to delegate only data detection to the CPU. Channel estimation and combining vector computation are then delegated to each one of the APs of the grid. This is supported by the fact each AP could easily be equipped with a baseband processor, thus we could design a system where CPU capabilities are defined by the number of UEs, but invariant to the number of APs. That is, we could add more APs to the grid without having to upgrade the CPU, as we would in centralized operation. To achieve that, each AP l receives pilot signals $\left\{\mathbf{y}_{tl}^{\text{pilot}} : t = 1, \ldots, \tau_p\right\}$ in order to locally estimate the channels $\left\{\hat{\mathbf{h}}_i : i \in \mathcal{D}_l\right\}$. For each UE k, the AP can use the local estimates and compute and select a combining vector $\mathbf{v}_{kl}$. Then the estimated symbol $\hat{s}_{kl}$ at the AP is then obtained as

$$\hat{s}_{kl} = \mathbf{v}_{kl}^H \mathbf{D}_{kl} \mathbf{y}_l^{\text{ul}}. \tag{5.21}$$

Next, the local estimates of the APs are gathered at the CPU, and there they are combined into a final UE data estimate. The CPU performs a linear combination of the local data estimates to obtain

$$\hat{s}_k = \sum_{l=1}^{L} a_{kl}^* \hat{s}_{kl} = \sum_{l=1}^{L} a_{kl}^* \mathbf{v}_{kl}^H \mathbf{D}_{kl} \mathbf{y}_l^{\text{ul}} \tag{5.22}$$

where $a_{kl} \in \mathbb{C}$ is the weight that the CPU assigns to the local estimate of the symbol $\hat{s}_k$ at AP l. To shrink fronthaul signaling, the APs only send the sata estimates, not the channel estimates to the CPU. Thus, the CPU selects the weights $\{a_{kl} : k = 1, \ldots, K, l \in \mathcal{M}_k\}$ as a function of the channel statistics. For instance, the UE that has a high SNR with respect to an AP should be prioritized, that is, should be granted a larger weight, than another UE that has a smaller SNR with respect to it. Also, interference management and combining scheme should be accounted for when selecting the weights. Those weights are denoted as large scale fading decoders (LSFDs), and receive that name due to being selected basing on the channel statistics, which are dependent on the large scale fading coefficients. We shall derive the SE expression for distributed operation by considering the use of the LSFD weights next

5.2.1 Spectral Efficiency for Distributed Operation

We shall now compute the SE expression for UE k in distributed operation. For that, we remember that

$$\mathbf{y}_l^{\text{ul}} = \sum_{i=1}^{K} \mathbf{h}_{il} s_i + \mathbf{n}_l \tag{5.23}$$

and we plug Eq. (5.23) into 5.22, so that we have

$$\hat{s}_k = \left(\sum_{l=1}^{L} a_{kl}^* \mathbf{v}_{kl}^H \mathbf{D}_{kl} \mathbf{h}_{kl} \right) s_k + \sum_{\substack{i=1 \\ k \neq i}}^{K} \left(\sum_{l=1}^{L} a_{kl}^* \mathbf{v}_{kl}^H \mathbf{D}_{kl} \mathbf{h}_{il} \right) s_i + n'_k \tag{5.24}$$

where $n'_k = \sum_{l=1}^{L} a_{kl}^* \mathbf{v}_{kl}^H \mathbf{D}_{kl} \mathbf{n}_l$ is the noise. We shall use the substitution propose in [8] for brevity and define the vector $\mathbf{g}_{ki} \in \mathbb{C}^L$ as

$$\mathbf{g}_{ki} = \begin{bmatrix} \mathbf{v}_{k1}^H \mathbf{D}_{k1} \mathbf{h}_{i1} \\ \vdots \\ \mathbf{v}_{kL}^H \mathbf{D}_{kL} \mathbf{h}_{iL} \end{bmatrix} \tag{5.25}$$

which is the vector whose elements are the receive combined channels between UE i to UE k for each AP. By using this notation, we can express Eq. (5.24) as

$$\hat{s}_K = \mathbf{a}_k^H \mathbf{g}_{kk} + \sum_{i=1//i \neq k}^{K} \mathbf{a}_k^H \mathbf{g}_{ki} s_i + n'_k \tag{5.26}$$

where $\mathbf{a}_k = [a_{k1}, \ldots, a_{kL}]^T \in \mathbb{C}^L$ is the LSFD weight vector of UE k. The realizations of the $\mathbf{g}$ vectors change at each coherence block, while the realizations of $\mathbf{a}_k$ are constant for a large number of coherence blocks. The effective channel $\mathbf{a}_k^H \mathbf{g}_{kk}$ is unknown at the CPU, however the average of it $\mathbb{E}\left\{\mathbf{a}_k^H \mathbf{g}_{kk}\right\} = \mathbf{a}_k^H \mathbb{E}\left\{\mathbf{g}_{kk}\right\}$ is known if the receive combining vector has been correctly selected. Thus, we can calculate the achievable SE.

Theorem 5.2

An achievable SE of an arbitrary UE k, in the UL distributed operation is given by

$$\text{SE}_k^{\text{d}} = \frac{\tau_u}{\tau_c} \mathbb{E}\left\{ \log_2 \left(1 + \text{SINR}_k^{\text{d}} \right) \right\} \quad \text{bit/s/Hz} \tag{5.27}$$

with the instantaneous effective SINR given by

$$\text{SINR}_k^{\text{d}} = \frac{p_k \left| \mathbf{a}_k^H \mathbb{E}\left\{\mathbf{g}_{kk}\right\} \right|^2}{\mathbf{a}_k^H \left(\sum_{i=1}^{K} p_i \mathbb{E}\left\{\mathbf{g}_{ki} \mathbf{g}_{ki}^H\right\} - p_k \mathbb{E}\left\{\mathbf{g}_{kk}\right\} \mathbb{E}\left\{\mathbf{g}_{kk}^H\right\} + \mathbf{F}_k \right) \mathbf{a}_k} \tag{5.28}$$

where

$$\mathbf{F}_k = \sigma^2 \, \text{diag}\left(\mathbb{E}\left\{ \|\mathbf{D}_{k1} \mathbf{v}_{k1}\| \right\}, \ldots, \mathbb{E}\left\{ \|\mathbf{D}_{kL} \mathbf{v}_{kL}\| \right\} \right) \in \mathbb{R}^{L \times L} \tag{5.29}$$

Proof 5.3: The proof can be found on Appendix A.4. □

The expression of Eq. (5.28) is a ratio between the desired signal power $p_k \left|\mathbf{a}_k^H \mathbb{E}\left\{\mathbf{g}_{kk}\right\}\right|^2$ and the interfering signals and noise, as well as a parcel of the desired signal power that comes from imperfect channel knowledge. This SE expression bears a similarity to the single antenna fading expression derived in Lemma 3.3, showing the data signal can be encoded and the receiver signal decoded as if communication occurred over a traditional AWGN channel.

The SE expression is not on closed form, due to the expected value operator in it. However, it can be easily computed by means of Monte Carlo simulations, meaning it can be approximated by an average over a large number of realizations. In other words, we could generate a handful of channel realizations over various sets of coherence blocks, and then average the result to obtain an approximation for the SE.

Theorem 5.2 serves the purpose of systems with multiple antennas APs with correlated or uncorrelated fading channels of an arbitrary distribution, as well as any arbitrary DCC setup. It should hold for any combining vector as well. We would like, however, to derive a combination of combiner and LSFD weights that makes the SE as high as possible.

5.2.2 Optimal Combining Vector and LSFD Weights

We start by taking a look at the combining vector. Differently from the centralized operation, the AP cannot compute a combiner that is optimal network-wide, since it does not hold the information of the other AP channel estimates. However. it can optimize a local performance metric, for instance the local MSE metric

$$\mathbb{E}\left\{\left|s_k - \mathbf{v}_{kl}^H \mathbf{D}_{kl} \mathbf{y}_l^{\text{ul}}\right|^2 \mid \left\{\mathbf{h}_{il} : i = \hat{1}, \ldots, K\right\}\right\}. \tag{5.30}$$

The combining vector that minimizes the MSE expression of Eq. (5.30) is

$$\mathbf{v}_{kl}^{\text{L-MMSE}} = p_k \left(\sum_{i=1}^{K} p_i \mathbf{D}_{kl} \left(\hat{\mathbf{h}}_{il}\hat{\mathbf{h}}_{il}^H + \mathbf{C}_{il}\right) \mathbf{D}_{kl} + \sigma^2 \mathbf{I}_N\right)^{-1} \mathbf{D}_{kl}\hat{\mathbf{h}}_{kl} \tag{5.31}$$

This can be proven by evaluating the conditional expectation of Eq. (5.30) and equating its first derivative with respect to $\mathbf{v}_{kl}$ to zero, as it is a minimization problem. We call this combiner local MMSE (L-MMSE) because it is like the centralized MMSE, except that it minimizes the MSE locally. Thus, it would equate the MMSE combining vector in a situation where only AP l serves UE k. Since on this situation $|\mathcal{M}_k| = 1$, we have that the computational complexity of this combiner also grows linearly with K, being also unscalable. By applying the adequate modifications in Eq. (5.12), we have that the number of

complex multiplications for L-MMSE combining is

$$C_m^{\text{L-MMSE}} = \underbrace{(N\tau_p + N^2)K|\mathcal{M}_k|}_{\text{channel estimation}} + \underbrace{\frac{N^2+N}{2}K|\mathcal{M}_k| + N^2|\mathcal{M}_k| + \frac{N^3-N}{3}|\mathcal{M}_k|}_{\text{combining vector}} \quad (5.32)$$

Notice that the square and cubic terms now are independent of $|\mathcal{M}_k|$, due to combining vector computation being delegated to the AP.

Now we shall discuss the optimal LSFD weights. This vector is optimized with knowledge of the channel statistics. We notice that Eq. (5.28) has the form of a generalized Rayleigh quotient with respect to $\mathbf{a}_k$. Thus, the SINR can be maximized by the optimal LSFD weights vector $\mathbf{a}_k^{\text{opt}}$ according to Corollary 5.2:

COROLLARY 5.2 The effective SINR in Eq. (5.28) is maximized for UE k by

$$\mathbf{a}_k^{\text{opt}} = p_k \left(\sum_{i=1}^{K} p_i \mathbb{E}\left\{\mathbf{g}_{ki}\mathbf{g}_{ki}^H\right\} + \mathbf{F}_k + \mathbf{D}_k \right)^{-1} \mathbb{E}\left\{\mathbf{g}_{kk}\right\} \quad (5.33)$$

with $\mathbf{D}_k \in \mathbb{R}^{L\times L}$ being the diagonal matrix with entry (l,l) being one if $l \notin \mathcal{M}_k$ and zero otherwise. This in turn leads to the maximum SINR value of

$$\text{SINR}_k^{\text{d}} = p_k \mathbb{E}\left\{\mathbf{g}_{kk}^H\right\} \left(\sum_{i=1}^{K} p_i \mathbb{E}\left\{\mathbf{g}_{ki}\mathbf{g}_{ki}^H\right\} - p_k \mathbb{E}\left\{\mathbf{g}_{kk}\right\}\mathbb{E}\left\{\mathbf{g}_{kk}^H\right\} + \mathbf{F}_k + \mathbf{D}_k \right)^{-1} \mathbb{E}\left\{\mathbf{g}_{kk}\right\} \quad (5.34)$$

Proof 5.4: Taking a look at the proof of Corollary 5.1, we notice that expression in Eq. (5.28) is a generalized Rayleigh quotient with $\mathbf{v} = \mathbf{a}_k$, $\mathbf{h} = \sqrt{p_k}\mathbb{E}\left\{\mathbf{g}_{kk}\right\}$ and $\mathbf{B} = \sum_{i=1}^{K} p_i \mathbb{E}\left\{\mathbf{g}_{ki}\mathbf{g}_{ki}^H\right\} - p_k \mathbb{E}\left\{\mathbf{g}_{kk}\right\}\mathbb{E}\left\{\mathbf{g}_{kk}^H\right\} + \mathbf{F}_k + \mathbf{D}_k$. The term $\mathbf{D}_k$ has been added so that matrix $\mathbf{B}$ becomes positive definite and should not impact the result as $\mathbf{D}_k\mathbb{E}\left\{\mathbf{g}_{kk}\right\} = \mathbf{0}_L$. □

The computing of $\mathbf{a}_k^{\text{opt}}$ requires the knowledge of the L-dimensional vector $\mathbb{E}\left\{\mathbf{g}_{kk}\right\}$, the $L \times L$ matrix $\mathbb{E}\left\{\mathbf{g}_{ki}\mathbf{g}_{ki}^H\right\}$, for $i = 1, \ldots, K$, and the L diagonal elements of matrix $\mathbf{F}_k$. Those vectors can all be computed on the APs based on the receives UL signals and the local combining vectors. However, the CPU is unable to acquire this information by its own and needs to receive it from the APs by the fronthaul links. From that, we can see that the number of parameters that is sent to the CPU grows with K, which also makes this LSFD weights vector unscalable. We also see that the computation of this vector can be carried out while

ignoring the $L - |\mathcal{M}_k|$ APs that are not serving UE k. Therefore, computing this vector is only a matter of inverting a $|\mathcal{M}_k| \times |\mathcal{M}_k|$ matrix and multiplying it by an $|\mathcal{M}_k|$ length vector. Thus, the number of complex multiplications for the optimal LSFD weights is

$$C_m^{\text{LSFD-opt}} = |\mathcal{M}_k|^2 + \frac{|\mathcal{M}_k|^3 - |\mathcal{M}_k|}{3}. \tag{5.35}$$

5.2.3 Scalable Combining Vectors and LSFD Weights

The previous section dervied the optimal combining vector and LSFD weights that maximizes the SE. However, such methods are not scalable, the combiner complexity growing linearly with K, and the statistical information required to compute the LSFD weights growing also with K, thus requiring an ever growing number of parameters be sent to the CPU via fronthaul link. Therefore, we shall derive more scalable combining vectors and LSFD weights.

5.2.3.1 Maximal Ratio Combining

The simplest scalable solution is to use MR combining which is given by

$$\mathbf{v}_{kl}^{\text{MR}} = \mathbf{D}_{kl}\hat{\mathbf{h}}_{kl}. \tag{5.36}$$

The main benefit of using this combining scheme is the low complexity, which again will be given by

$$C_m^{\text{MR}} = (N\tau_P + N^2)|\mathcal{M}_K| \tag{5.37}$$

Another benefit of this combiner is the possibility of computing all of the expected values of Theorem 5.2 analytically, so there is a closed form for the SE.

COROLLARY 5.3 If MR combining is used, the expectations of Eq, 5.28 become

$$\left[\mathbb{E}\left\{\mathbf{g}_{ki}\right\}\right]_l = \begin{cases} \sqrt{\eta_k\eta_i}\tau_p \operatorname{tr}\left(\mathbf{D}_{kl}\mathbf{R}_{il}\Psi_{t_kl}^{-1}\mathbf{R}_{kl}\right) & i \in \mathcal{P}_k \\ 0 & i \notin \mathcal{P}_k \end{cases} \tag{5.38}$$

with $\left[\mathbb{E}\left\{\mathbf{g}_{ki}\mathbf{g}_{ki}^H\right\}\right]_{lr} = \left[\mathbb{E}\left\{\mathbf{g}_{ki}\right\}\right]_l \left[\mathbb{E}\left\{\mathbf{g}_{ki}^*\right\}\right]_r$ for $r \neq l$ and

$$\begin{aligned}\left[\mathbb{E}\left\{\mathbf{g}_{ki}\mathbf{g}_{ki}^H\right\}\right]_{ll} &= \eta_k\tau_p \operatorname{tr}\left(\mathbf{D}_{kl}\mathbf{R}_{il}\mathbf{R}_{kl}\Psi_{t_kl}^{-1}\mathbf{R}_{kl}\right) \\ &\quad + \begin{cases} \eta_k\eta_I\tau_p^2\left|\operatorname{tr}\left(\mathbf{D}_{kl}\mathbf{R}_{il}\Psi_{t_kl}^{-1}\mathbf{R}_{kl}\right)\right|^2 & i \in \mathcal{P}_k \\ 0 & i \notin \mathcal{P}_k \end{cases}\end{aligned} \tag{5.39}$$

and

$$[\mathbf{F}_k]_{ll} = \sigma^2 \eta_k \tau_p \, \mathrm{tr}\left(\mathbf{D}_{kl} \mathbf{R}_{il} \Psi_{t_k l}^{-1} \mathbf{R}_{kl}\right) \tag{5.40}$$

Proof 5.5: The proof can be found on Appendix A.5. □

With Corollary 5.3, we can obtain a closed form expression for the SE of MR combining. However, such expression is lengthy, so we opt to present the simpler case of $N = 1$.

COROLLARY 5.4 The case where each AP has $N = 1$ antenna, the MMSE channel estimate of the channel $h_{kl} \in \mathbb{C}$ has variance

$$\gamma_{kl} = \frac{\eta_k \tau_p \beta_{kl}^2}{\sum_{i \in \mathcal{P}_k} \eta_I \tau_p \beta_{il} + \sigma^2}. \tag{5.41}$$

The instantaneous effective SINR is given by

$$\frac{p_k \left| \sum\limits_{l \in \mathcal{M}_k} a_{kl}^* \gamma_{kl} \right|^2}{\sum\limits_{i=1}^{K} p_i \sum\limits_{l \in \mathcal{M}_k} |a_{kl}|^2 \gamma_{kl} \beta_{il} \gamma_{kl} + \sum\limits_{i \in \mathcal{P}_k \setminus \{k\}} p_i \left| \sum\limits_{l \in \mathcal{M}_k} a_{kl}^* \gamma_{kl} \sqrt{\frac{\eta_i}{\eta_k}} \frac{\beta_{il}}{\beta_{kl}} \right|^2 + \sigma^2 \sum\limits_{l \in \mathcal{M}_k} |a_{kl}|^2 \gamma_{kl}} \tag{5.42}$$

Proof 5.6: By using the γ_{kl} definition, we can rewrite the expected value in Corollary 5.3 as $[\mathbb{E}\{\mathbf{g}_{ki}\}]_l = \sqrt{\frac{\eta_i}{\eta_k}} \frac{\beta_{il}}{\beta_{kl}} \gamma_{kl}$ for $i \in \mathcal{P}_k$, and 0 otherwise. Also, $[\mathbf{F}_k]_{ll} = \sigma^2 \gamma_{kl}$ and

$$[\mathbb{E}\{\mathbf{g}_{ki} \mathbf{g}_{ki}^H\}]_{ll} = \beta_{il} \gamma_{kl} + \begin{cases} \frac{\eta_i \beta_{il}^2}{\eta_k \beta_{kl}^2} \gamma_{kl}^2 & i \in \mathcal{P}_k \\ 0 & i \notin \mathcal{P}_k \end{cases} \tag{5.43}$$

for $l \in \mathcal{M}_k$ and zero otherwise. By inserting those values in Eq. (5.28) the proof is complete. □

The expression of Eq. (5.42) shows the behavior of the CF operation. The desired signal term $p_k \left| \sum\limits_{l \in \mathcal{M}_k} a_{kl}^* \gamma_{kl} \right|^2$ contains the desired signal power weighted by the sum of the contributions of the serving APs. The contribution of each AP is proportional to the variance of its channel estimate γ_{kl}, that was used to compute the combining vector. The combination of

the contributions from different APs are done coherently and are represented as the square of the sum. The denominator contains the interference, which is non-coherently combined, and is represented as the sum of the interfering powers at each of the $l \in \mathcal{M}_k$ serving APs. Each of those terms is multiplied by a scaling factor $|a_{kl}|^2\gamma_{kl}$ that contains the LSFD weights and the contribution of MR combining. The second term of the denominator is the interference caused by pilot contamination. Finally, the third term refers to the noise power. If we consider an unitary LSFD vector $\mathbf{a}_k = [1, \ldots, 1]^T$, then the SE expression of the distributed operation is equivalent tp that of centralized operation (by using MR combining). Thus, centralized MR combining can be used in a distributed fashion. Despite that, distributed MR combining still suffers from the same problem as its centralized counterpart, which is the inability to manage interference. Therefore, we shall explore combining vector schemes that manage interference better.

5.2.3.2 Local P-MMSE Combining

Taking inspiration from the P-MMSE combiner, we can derive a local P-MMSE alternative, that aims to reduce L-MMSE's complexity, by considering only the statistics of the UEs that are served by AP l, that is, $\left\{\hat{\mathbf{h}}_{il} : i \in \mathcal{D}_l\right\}$. By considering only the UEs in $\mathcal{D}_l$, we can rewrite Eq. (5.31) as

$$\mathbf{v}_{kl}^{\text{LP-MMSE}} = p_k \left(\sum_{i \in \mathcal{D}_l} p_i \mathbf{D}_{kl} \left(\hat{\mathbf{h}}_{il} \hat{\mathbf{h}}_{il}^H + \mathbf{C}_{il} \right) \mathbf{D}_{kl} + \sigma^2 \mathbf{I}_N \right)^{-1} \mathbf{D}_{kl} \hat{\mathbf{h}}_{kl}. \tag{5.44}$$

The number of complex multiplications of this combining method is given by

$$C_m^{\text{LP-MMSE}} = \underbrace{(N\tau_p + N^2) \sum_{l \in \mathcal{M}_k} |\mathcal{D}_l|}_{\text{channel estimation}} + \underbrace{\frac{N^2 + N}{2}) \sum_{l \in \mathcal{M}_k} |\mathcal{D}_l| + N^2 |\mathcal{M}_k| + \frac{N^3 - N}{3} |\mathcal{M}_k|}_{\text{combining vector}}, \tag{5.45}$$

which is independent of K, and thus scalable. If we compare it to P-MMSE, we can see that it has a much lower complexity per UE, since it requires the inversion of a $N \times N$ matrix as opposed to a $N|\mathcal{M}_k| \times N|\mathcal{M}_k|$ matrix. Also, one could notice that local P-MMSE (LP-MMSE) and L-MMSE coincide in the case where the l-th AP serves all UEs. As opposed to MR the expected values of Eq. (5.28) cannot be obtained in a closed form, however they can easily be simulated by Monte Carlo techniques.

5.2.3.3 Local P-RZF

As opposed to the centralized case, where the complexity of P-MMSE could be as large that it would require much more computation capacity of the CPU, its local counterpart

LP-MMSE already has a rather low complexity per UE. Therefore, the existence of a local P-RZF (LP-RZF) is not needed, as it would reduce an already low complexity at the cost of reducing system performance, which differently from centralized operation, is a handicap distributed systems do not need to maintain.

5.2.3.4 Scalable LSFD

The optimal LSFD vector is not scalable for implementation in large networks. The main reason for that is the fact that all UEs in the network affect the interference levels at all APs, determining how accurate the local estimates are. To find a scalable way, we need to limit the amount of interfering UEs that will be considered in computation. The simplest solution is to assign equal weight to all of the APs by having $\mathbf{a}_K = [1, \ldots, 1]^T$, as proposed in [42]. In that case, the CPU takes the local data estimates of UE k, and combines them to obtain a final data estimate

$$\hat{s}_k = \sum_{l=1}^{L} \hat{s}_{kl}. \tag{5.46}$$

This is called no LSFD (n-LSFD), since the LSFD weights are non existent. This was largely used in previous CF and Massive MIMO application, before the concept of LSFD were introduced. The main benefit of this approach, is that no statistical knowledge is needed at the CPU to compute Eq. (5.46). The drawback is that, since each AP is given the same importance, APs with high SNR with respect to UE k are given the same importance as APs with low SNR with respect to it. Therefore, one AP could introduce a high amount of interference at an AP that has a high SNR, and by that actually decrease the SE.

The CPU should then use at least some statistical information about the channels in order to assing the weights, much like the optimal LSFD weights vector. By observing the peculiarity of the P-MMSE combining method when compared to the MMSE, we see that the former uses information of only the UEs that are partially served by the same set of APs, that is $i \in \mathcal{S}_k$, since those should be the UEs causing the majority of interference at the estimates of UE k. Therefore, like P-MMSE combining, we can select an near optimal LSFD weight vector, such that it uses only the estimates of the UEs that are present in $|\mathcal{S}_k|$. Thus, the optimal LSFD vector can be approximated as

$$\mathbf{a}_k^{\text{n-opt}} = p_k \left(\sum_{i \in \mathcal{S}_k} p_i \mathbb{E}\left\{ \mathbf{g}_{ki} \mathbf{g}_{ki}^H \right\} + \mathbf{F}_k + \tilde{\mathbf{D}}_k \right)^{-1} \mathbb{E}\left\{ \mathbf{g}_{kk} \right\}, \tag{5.47}$$

which is called near optimal LSFD (nopt-LSFD) vector proposed by [43]. It is called nearly optimal because it has performance similar to the optimal LSFD weights, which will be proven numerically. In regards to complexity, it has the same number of complex multiplications as its optimal counterpart, since the dimensions of its matrices are equal. However,

unlike the optimal solution, the nopt-LSFD does send fewer statistical parameters over the fronthaul links to the CPU, and this number is independent of K, which makes fronthaul signaling scalable

5.2.4 Fronthaul Signaling for Distributed Operations

The fronthaul signaling that is required by the distributed operation is quantified as follows. The l-th AP sends the estimates $\hat{s}_{kl}$ for $k \in \mathcal{D}_l$ to the CPU, corresponding to a total of $\tau_u|\mathcal{D}_k|$ complex scalars over one coherence block. When the DCC is scalable with respect to K, so is fronthaul signaling. Besides that, the statistical parameters needed to compute the LSFD vectors must be sent to the CPU, and this number is different if we use the optimal and near optimal solutions as we will quantify in this section.

The following parameters must be sent from each AP l to the CPU for computing the optimal LSFD vector for an UE $k \in \mathcal{D}_l$:

- $\mathbb{E}\{[\mathbf{g}_{ki}]_l\}$, for $i = 1, \ldots, K$;
- $\mathbb{E}\left\{|[\mathbf{g}_{ki}]_l|^2\right\}$, for $i = 1, \ldots, K$;
- $[\mathbf{F}_k]_{ll}$.

We notice that $\mathbb{E}\{[\mathbf{g}_{ki}]_l[\mathbf{g}_{ki}^*]_r\} = \mathbb{E}\{[\mathbf{g}_{ki}]_l\}\,\mathbb{E}\{[\mathbf{g}_{ki}^*]_r\}$, for $l \neq r$. This means each AP must send complex scalars that correspond to the expected value of $\mathbf{g}_{ki}$ and its complex conjugate with respect to all other APs. Therefore, each AP must send $3K + 1$ complex scalars. However, by discounting the repeated scalars, we see that the total number of complex scalars is $(3K + 1)/2$. Each AP sends those values to each of those served $i \in \mathcal{D}_l$ UEs. Therefore $(3K + 1)/2 \sum_{l=1}^{L} |\mathcal{D}_l|$ values are needed in total. This number relies on K, so it is unscalable, as already commented. If we instead use the nopt-LSFD vector, then AP l must send $(3|\mathcal{S}_k| + 1)/2$ complex scalars, and the total number of complex scalars the CPU should receive is given by $\sum_{l=1}^{L} \sum_{k \in \mathcal{D}_l} (3|\mathcal{S}_k| + 1)/2$. This value does not depend on K, so this alternative provides an scalable solution to the network if allied to any scalable combining vector solution. Finally, in the case of n-LSFD we have that no channel statistics must be sent to the CPU, so the number of complex scalar that must be sent is zero. A summary of complex scalars that must be sent over the fronthaul links for each of the methods is given in Table 5.1.

We simulate the same system that was used to generate Fig. 5.2, only that time in a distributed manner. The results can be seen in Fig. 5.3. We notice that under the unscalable case of every AP serving every UE, the performances of L-MMSE and LP-MMSE coincide. That happensa because $\mathcal{S}_k = \{1, \ldots, K\}$ for every k. The same thing can be said about

Scheme	Number per coherence block	Statistics
Centralized	$(\tau_p + \tau_u)NL$	-
Distributed: opt LSFD	$\tau_u \sum_{l=1}^{L} \lvert\mathcal{D}_l\rvert$	$\frac{3K+1}{2} \sum_{l=1}^{L} \lvert\mathcal{D}_l\rvert$
Distributed: nopt-LSFD	$\tau_u \sum_{l=1}^{L} \lvert\mathcal{D}_l\rvert$	$\sum_{l=1}^{L} \sum_{k \in \mathcal{D}_l} \frac{3\lvert\mathcal{S}_k\rvert+1}{2}$
Distributed: n-LSFD	$\tau_u \sum_{l=1}^{L} \lvert\mathcal{D}_l\rvert$	-

Table 5.1 – The number of complex scalars that must be sent to the CPU from the APs over the fronthaul per coherence block and per channel statistics (that is over a new spatial correlation matrix realization). Centralized and distributed operations are considered

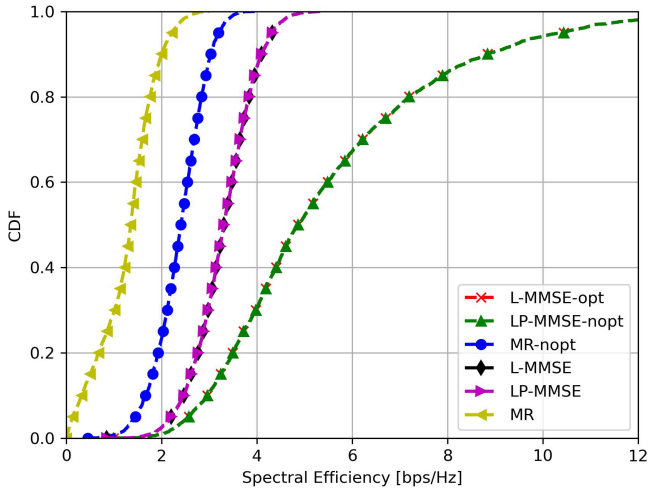


Figure 5.3 – SE for the CF UL system where all of the APs serve all UEs. The combining vectors L-MMSE,LP-MMSE and MR with and without LSFD are used.

the optimal and near optimal LSFD weights. We can also notice the importance of properly computing the LSFD weights: LP-MMSE and L-MMSE have a large performance drop when they do not operate with LSFD compared to their LSFD counterparts. Their performances are however still better than any MR curve, because they still manage to mitigate interference. When we take a look at the MR curves, we can see that MR with LSFD has greater performance than MR without LSFD, for already discussed reasons. Therefore, we stress another time the importance of efficiently selecting the LSFD weights in order to better make use of the distributed operation benefits.

5.3 COMPLEX MULTIPLICATIONS FOR DISTRIBUTED AND CENTRALIZED OPERATIONS

Also, the number of complex multiplications needed to estimate the channel and compute the combining vector, as well as compute the LSFD weights is shown in Table 5.2 for the distributed schemes and Table 5.3 for centralized schemes

Scheme	Number of complex multiplications
opt and n-opt LSFD	$\|\mathcal{M}_k\|^2 + \frac{\|\mathcal{M}_k\|^3 - \|\mathcal{M}_k\|}{3}$
L-MMSE	$(N\tau_p + N^2)K\|\mathcal{M}_k\| + \frac{N^2+N}{2}K\|\mathcal{M}_k\| + N^2\|\mathcal{M}_k\| + \frac{N^3-N}{3}\|\mathcal{M}_k\|$
LP-MMSE	$(N\tau_p + N^2)\sum_{l\in\mathcal{M}_k}\|\mathcal{D}_l\| + \frac{N^2+N}{2})\sum_{l\in\mathcal{M}_k}\|\mathcal{D}_l\| + N^2\|\mathcal{M}_k\| + \frac{N^3-N}{3}\|\mathcal{M}_k\|$
MR	$(N\tau_p + N^2)\|\mathcal{M}_K\|$

Table 5.2 – Number of complex multiplications per coherence block to compute the local channel estimates and combining vectors of UE k. Different local combining methods are listed.

Scheme	Number of complex multiplications
MMSE	$(N\tau_p + N^2)K\|\mathcal{M}_k\| + \frac{(N\|\mathcal{M}_k\|)^2+N\|\mathcal{M}_k\|}{2}K + (N\|\mathcal{M}_k)^2 + \frac{(N\|\mathcal{M}_k\|)^3-N\|\mathcal{M}_k\|}{3}$
P-MMSE	$(N\tau_p + N^2)\|\mathcal{S}_k\|\|\mathcal{M}_k\| + \frac{(N\|\mathcal{M}_k\|)^2+N\|\mathcal{M}_k\|}{2}\|\mathcal{S}_k\| + (N\|\mathcal{M}_k)^2 + \frac{(N\|\mathcal{M}_k\|)^3-N\|\mathcal{M}_k\|}{3}$
P-RZF	$(N\tau_p + N^2)\|\mathcal{S}_k\|\|\mathcal{M}_k\| + \frac{\|\mathcal{S}_k\|^2+\|\mathcal{S}_k\|}{2}N\|\mathcal{M}_k\| + \|\mathcal{S}_k\|^2 + \|\mathcal{S}_k\|N\|\mathcal{M}_k + \frac{\|\mathcal{S}_k\|^3-\|\mathcal{S}_k\|}{3}$
MR	$(N\tau_p + N^2)\|\mathcal{M}_K\|$

Table 5.3 – Number of complex multiplications per coherence block to compute the centralized channel estimates and combining vectors of UE k. Different local combining methods are listed.

We shall use such tables in the next chapter, in order to compute the complexity of the described DCC methods.

Chapter Summary

- There are two ways in which the UL in CF systems could be implemented: distributed and centralized
- In the centralized operation, the channel estimation, combining vector computation and data detection are all delegated to the CPU. In the distributed operation channel estimation and combining vector computation are locally made at each AP, with the CPU being responsible for data detection.
- The MMSE combiner for centralized operation is optimal, in which it maximizes the SE for MMSE channel estimation. It is unscalable however. A more

scalabe approach is to use P-MMSE, which has a similar performance at the expense of complexity. If complexity is still an issue, then P-RZF can be used at the expense of performance. Finally, if there is a sufficient degree of favorable propagation, or if complexity reduction is the main goal, then MR combining is an alternative.

- In distributed operation, the CPU has the power to attribute weights to the data estimates of each AP. This weights are based on the channel conditions of each AP and are called LSFD. The appropriate computation of LSFD weights is of vital importance to the operation of such systems.
- The combining method that maximizes SE for distributed operation with MMSE channel estimation is the L-MMSE. It is, however, unscalable, so we have the LP-MMSE, that ha a similar performance at the expense of complexity. If there is a high degree of favorable propagation between the UEs, then MR combining can be used with satisfactory results. However, unlike the centralized operation, complexity is usually not such an issue in distributed operation.
- The optimal LSFD weights maximize the SE for distributed operation, but require an ever growing number of scalars be sent to the CPU over the fronthaul links as the number of UEs increases. To remedy that, the near optimal solution is proposed, as it does not demand an ever growing number of complex scalars be sent to the CPU, since this number is independent of K.

6 EXPERIMENTAL RESULTS

This chapter aims to establish a comparison betwen the DCC methods as well as the pilot aasignment algorithms introduced in Chapter 4. section 6.1 compares the performance and complexity of the DCC methods for various network parameters. section 6.2 compares performance for the various pilot assignment methods. Finally section 6.3 compares performance for unknown channel statistics.

Having introduced the distributed and centralized UL operations in Chapter 5, as well as the complexity expressions for computing the combining vectors, channel estimates and sending statistical information to the CPU through the fronthaul links, we would like now to establish comparisons between the DCC and pilot assignments methods described in 4. For that, we resort to the simulation setup described in Table 4.1.

6.1 DCC METHODS COMPARISON

For this section we simulate a CF MIMO system with the described DCC methods introduced in Chapter 4. For the pilot assignment method, we use RP with the correlation metric of the large scale fading coefficients of Eq. (4.13). We compare the DCC methods for the various combiners discussed in Chapter 5 over centralized and distributed UL operation.

6.1.1 Spectral efficiency for DCC methods

We aim to compare the SE values for the DCC methods and different combining vector techniques. Fig. 6.1 shows the SE if MMSE combining method is used We can see that,

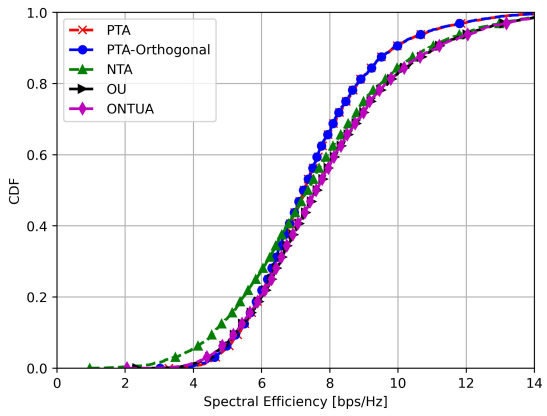


Figure 6.1 – SE CDF for MMSE combining vector for various DCC methods

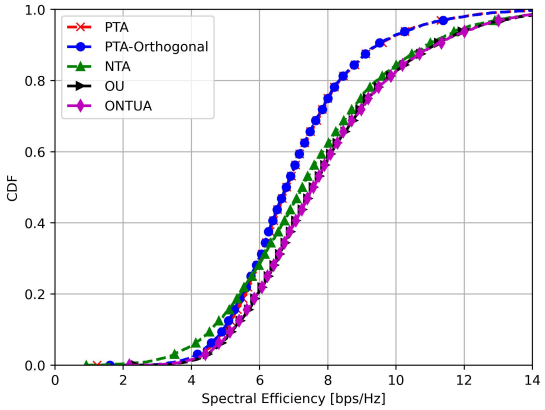


Figure 6.2 – SE CDF for P-MMSE combining vector for various DCC methods

for the lower percentiles of the CDF curve, corresponding to the percentage of time the SE is higher than the corresponding value, the NTA has the weakest performance. That is to be expected, since this method does not take in account the interference between the UEs that are connected to the same AP, which in turn results in smaller values of SE for the UEs that have weaker channel conditions. When we consider PTA, OUA, Orthogonal-PTA and ONTUA, their performance for lower percentiles are very similar. On the case of PTA and ONTUA this happens because the UEs that are connected to the AP have mutually orthogonal pilot sequences, and therefore do not interfere between each other. For the case of PTA such similarity is explained by the fact that the threshold power percentage is high, meaning that the UE with the weakest channel condition is likely to be connected to many APs and in turn achieve a satisfactory SE value. Now taking a look at the higher percentile values, it is clear that the PTA and Orthogonal-PTA methods have a much lower performance than the other methods. This is due to the UE having excellent channel condition being connected to few APs, since they already obey the power threshold criteria. Since the other DCC methods rely on connecting more APs to the UE, even if it has the strongest channel conditions, the performance is greater. The same conclusions can be extended to the P-MMSE and P-RZF combining vector of Fig. 6.2 and Fig. 6.3

When we analyze MR combining of Fig. 6.4 we reach a more interesting conclusion for the lower SE percentiles: performance is actually greater for the NTA method. This is not very intuitive, since NTA is prone to having performance loss due to the lack of interference cancellation between the neighboring UEs as already discussed. However, when we refer to MR we remember that it also lacks the capacity to mitigate the interference between pilot-sharing UEs. Therefore, since NTA will connect a large number of UEs and APs pairs. In fact, from Fig. 4.9, we see that this corresponds to more than 60% for $K = 40$. Therefore, even though the UEs connected to the same AP may share one or more pilot sequences, the fact that more APs will serve each UE makes that NTA has a greater performance than the

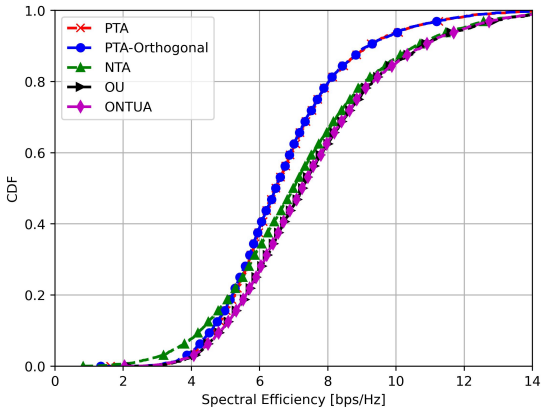


Figure 6.3 – SE CDF for P-RZF combining vector for various DCC methods

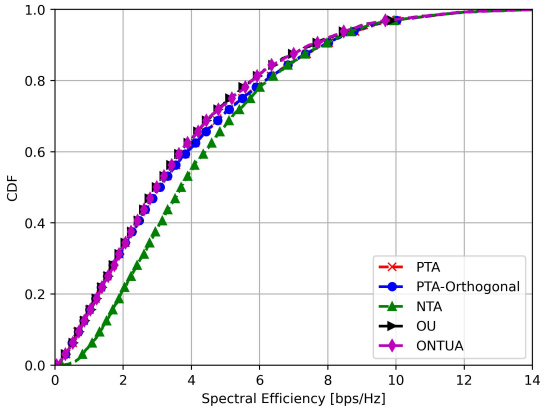


Figure 6.4 – SE CDF for MR combining vector for various DCC methods

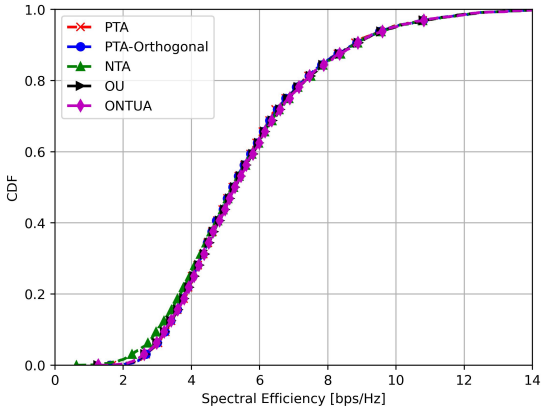


Figure 6.5 – SE CDF for optimal L-MMSE combining vector for various DCC methods

other methods for MR.

From centralized operation methods, we can see that NTA and ONTUA act in order to enhance the overall SE of the system, that is, they act in a manner such that the UE with good channel conditions also benefit from connecting to a large number of APs and by doing that, enhance their already high SE. On the other hand, PTA and Orthogonal-PTA act in a manner such that they prioritize a fairer distribution of the SEs in the grid, that is, they try to enhance the performance of the worst channel condition UE to an acceptable level. One could argue that ONTUA and OUA also have this behavior, but we have to keep in mind that they allow for a large number of APs to be connected to a given UEs irrespective of individual channel condition. However, we have to consider that OUA and ONTUA also aims to cancel interference between the neighboring UEs being served by the same AP, allowing for a fairer distribution of grid SE in a level.

We now analyze the results of optimal L-MMSE combining in Fig. 6.5 From it we can see that the performance of every DCC method is very similar. This derives from the fact the CPU performs optimal fusing of the data estimates of each AP, which in turn makes the SE values proportional to the channel conditions of the UEs. That is, if an UE has an average channel condition, then the optimal fusing by means of the LSFD weights allied by the interference mitigation of L-MMSE will be such that the UE will have a performance that is on pair with the quality of its collective channel. Also, the L-MMSE combining vector does not depend on the set $\mathcal{S}_k$ of neighboring interfering UEs that, as already seen at Chapter 4, have a rather diverse behavior for the various DCC methods. The near optimal LP-MMSE combining vector of Fig. 6.6 on the other hand depends on the neighboring interfering UEs set. Since OUA and ONTUA have an interference mitigation method, their performance is higher, specially for the lower SE percentiles. NTA is slightly worse than them performance wise, since it lacks the ability to mitigate said interference. PTA and Orthogonal-PTA have the worst performance due to the fact the near optimal LSFD weights cancel the benefit

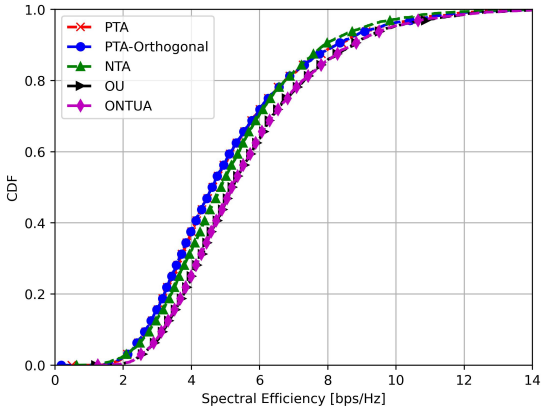


Figure 6.6 – SE CDF for near optimal LP-MMSE combining vector for various DCC methods

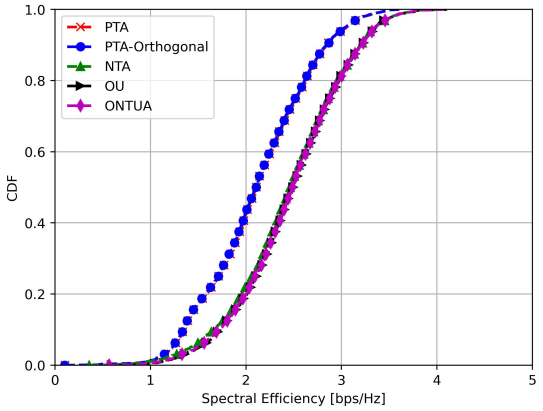


Figure 6.7 – SE CDF for near optimal MR combining vector for various DCC methods

of the weaker UEs of connecting to more APs, because they are dependent on the set of neighboring interfering UEs, which in PTA case might share the same pilot sequence. Near optimal MR of Fig. 6.7 has the same behavior of Fig. 6.6, since the near optimal LSFD weights are actually detrimental to the PTA and Orthogonal-PTA performance. When we analyze the effects of n-LSFD LP-MMSE on the SE, we see that, unlike the optimal or nearly optimal cases, the performance for PTA and Orthogonal-PTA greatly increases. This happens because the UEs with weak channel conditions are benefited from connecting to a handful of APs in order to reach the power threshold percentage, and are not "punished" by LSFD the way they would be under optimal or nearly optimal conditions. The results can be seed in Fig. 6.8 and Fig. 6.9 We notice that for LP-MMSE, the lack of LSFD actually increases the SE for higher percentiles, however for lower percentiles performance for PTA and Orthogonal-PTA is still low. This is explained by the fact the LP-MMSE combining vector relies on the set $\mathcal{S}_k$, which is composed of the interference that weaken their performance. Regardless, it is clear those DCC methods work better in the case of n-LSFD, which might explain why

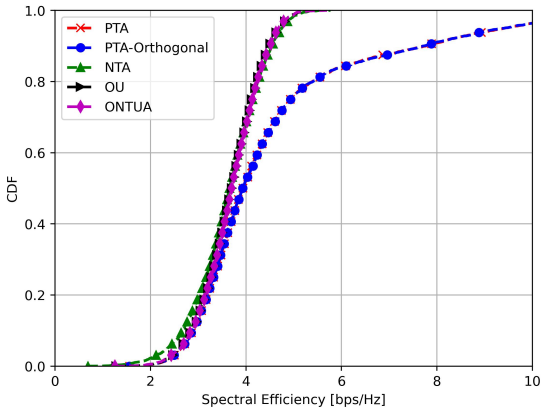


Figure 6.8 – SE CDF for n-LSFD L-MMSE combining vector for various DCC methods

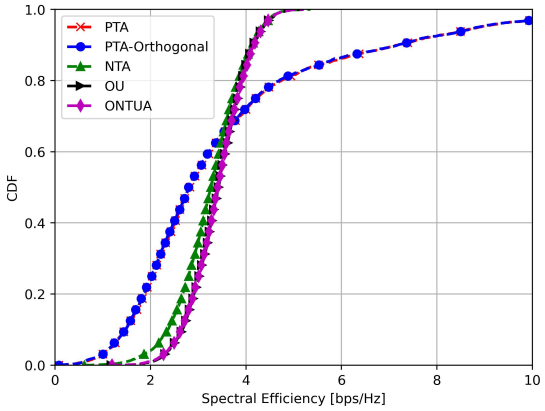


Figure 6.9 – SE CDF for n-LSFD LP-MMSE combining vector for various DCC methods

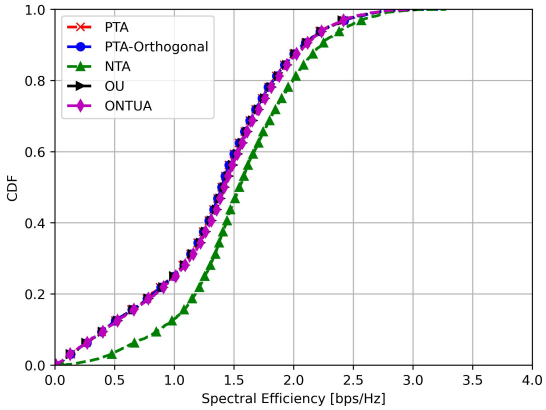


Figure 6.10 – SE CDF for n-LSFD MR combining vector for various DCC methods

PTA was widely used in earlier literature, when the LSFD weighting technique had not yet been proposed.

Finally, the SE for distributed MR combining is analyzed. Since the distributed MR has a similar behavior to its centralized counterpart, the same assumptions that were made for centralized MR can be used. The NTA method has the best performance overall due to the MR combining method lacking the ability to mitigate interference, which in turn makes the UE with the highest amount of connected APs more likely to have a higher value of SE.

On a general basis for distributed operation, the DCC scheme that exhibits the larger performance gains is the OUA and the ONTUA. Their ability to mitigate interference between the UEs being served by the same AP is greatly beneficial for performance, specially when allied to nearly-optimal or optimal LSFD (except in the case of MR combining without LSFD). NTA stands on the middle, since, despite lacking the ability to mitigate interference it still allows for a large number of UE-AP pairs to be connected. Finally, the PTA and Orthogonal-PTA exhibit the worst performance in an optimal or nearly-optimal LSFD scenario, although they exhibit the best performance when there is no LSFD. In situations where the frotnhaul link capacity is limited. thus making use of near optimal LSFD impossible, then those DCC methods might be useful.

6.1.2 Complexity for the DCC methods

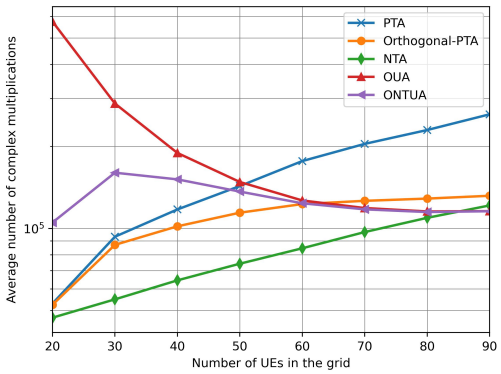
Now we aim to establish a comparison between the DCC methods' complexities for the various combining vectors over centralized and distributed operations. Table 6.1 shows the average number of complex multiplications per UE per coherence block to compute the combining vectors or perform channel estimation.

		DCC Method				
Operation	Combiner	PTA	Orthogonal PTA	NTA	OUA	ONTUA
Centralized	MMSE	117398	101797	64471	189259	151377
	P-MMSE	110095	93678	59225	189257	151348
	P-RZF	6502	5838	8693	12684	11892
	MR-cent	298	281	310	560	509
Distributed	opt L-MMSE	3148	3066	2899	3594	3390
	n-opt LP-MMSE	873	771	625	1614	1351
	L-MMSE	2732	2727	2748	2820	2805
	LP-MMSE	457	432	475	840	766
	n-opt MR	715	620	461	1334	1095
	MR-dist	298	281	310	560	509

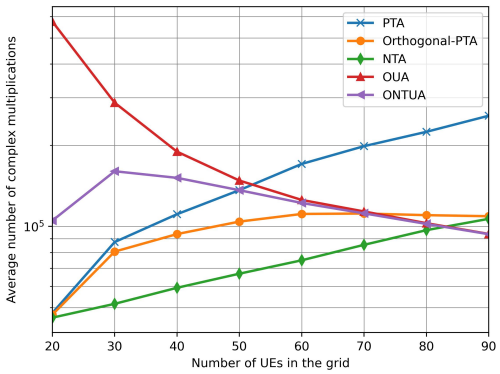
Table 6.1 – Average number of complex multiplications per coherence block to compute the combining vector, the channel estimates and, when needed, the LSFD weights vector

From Table 6.1, we see that, for centralized operation, MMSE has the largest complexity while MR has the lowest. As already discussed, OUA has the largest number of required complex multiplications, since each AP is always connected to τ_p UEs, which makes the minimum value of $|\mathcal{S}_k|$ be τ_p. Moreover, ONTUA has lower complexity than OUA in all cases, deriving from the fact ONTUA is a subset of OUA. In regard to SE, however, they have very similar performance, which makes ONTUA a viable solution for AP since it has high performance with lower complexity. NTA has less than half of the required complex multiplications of OUA and ONTUA for P-MMSE and MMSE, at the expense of some SE loss, while PTA and Orthogonal-PTA have half that number while keeping the SE for lower percentiles on a satisfactory level. For MR combining, the complexity for NTA is lower than OUA and ONTUA, and still the SE is higher. Thus centralized MR has the peculiarity of being able to lower complexity while simultaneously increasing performance.

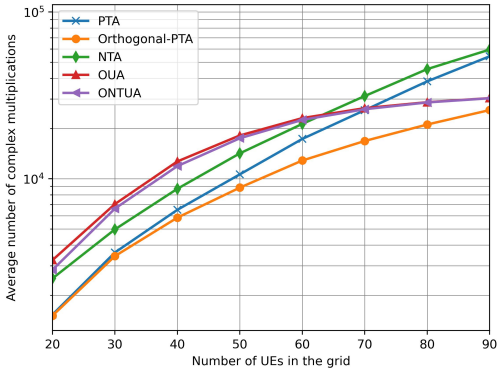
For distributed operation, it is possible to see that L-MMSE has the largest complexity, irrespective of the DCC method and the existence or not of LSFD. Complexity-wise, the performance of all DCC methods for those combining vectors is very similar, so there should be no aggressive SE trade off of using either of those DCC algorithms. For LP-MMSE, however, the complexity of ONTUA and OUA are the largest. We should recall, however, that the two lowest performance schemes, PTA and Orthogonal-PTA, also suffer from large performance losses under near optimal LSFD weights data fusing. Therefore, for a better SE-complexity trade off it would be adequate to choose NTA. In the case of no LSFD, one could then reach the ideal situation of the least complex method being also the one with the best performance. The conclusions for distributed MR are the same as those from centralized



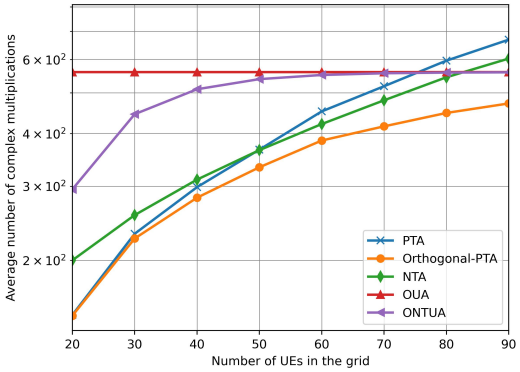
(a) MMSE



(b) P-MMSE



(c) P-RZF



(d) MR

Figure 6.11 – Average number of complex multiplications for the centralized operation combining vectors for the PTA, Orthogonal-PTA ,NTA, OUA and ONTUA versus total number of UEs in the grid

MR. For the nearly optimal LSFD MR, the complexity for NTA is the lowest among the methods that could be used without significant performance losses. Therefore, it is again the chosen DCC method.

6.1.3 Complexity and performance comparison for DCC methods versus number of UEs

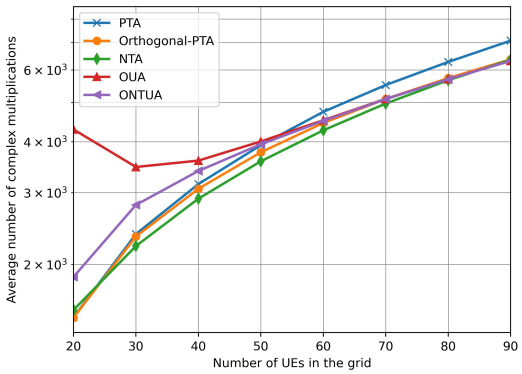
Having compared the complexities of the DCC methods under an specific scenario, $K = 40$, we now explore the scalability of such methods with respect to the number of UEs in the grid. We compare once again the performance for the combining vectors introduced in Chapter 5. Fig 6.11 shows the results for centralized operation It is possible to immediately identify that the PTA and NTA methods are unscalable with respect to K, since the average complexity increases with the total number of UEs in the grid. For NTA this is explained in Fig. 4.9, from Chapter 4: as the number of UEs increases, so does the percentile of connected UE-AP pairs. As for PTA, when K increases, so does the number of connected AP-UE pair,

and consequently, the lengths of the sets $\mathcal{M}_k$ and $\mathcal{S}_k$. With respect to OUA, ONTUA and Orthogonal-PTA, all of them have some form of interference mitigation, meaning UEs that share the same pilot sequences will never connect to the same AP. Therefore, the maximum number of UEs that might be connected to the same AP is τ_p. As K increases, the number of connected UEs partially served by the same AP tends to decrease, therefore lowering the length of the set $\mathcal{S}_k$, and in turn, average complexity. Another insight is that the complexities of ONTUA and OUA converge as K increases. This is explained by again looking at Fig. 4.9 and noticing that the simulated threshold value of -0.23 leads to more UEs being served by the same AP up to a limit of τ_p, that is reached as K increases.

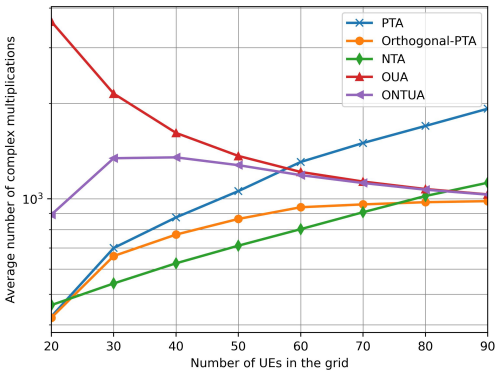
Shifting the focus to the combining vectors, we see that, as anticipated, MMSE is unscalable, since its complexity does not decrease with K. P-MMSE was introduced as an scalable alternative, and we can see form the scalable DCC schemes, that complexity indeed reduces as K increases. One interesting fact is that P-RZF, which we should recall, was an alternative to complexity reduction of the P-MMSE combiner, has its complexity increased with K. This is explained by the fact the length of the partially served UEs set $\mathcal{S}_k$ increases from $K = 20$ to $K = 60$. After that, it starts decreasing, but the complexity stabilizes at a much higher value than $K = 20$. Therefore, besides the SE loss, P-RZF also experiences an increased complexity for large K, meaning that P-MMSE would be more suitable for those scenarios. With respect to MR, we see that the complexity is almost unaffected by K for OUA, since this combiner has an overall low complexity.

We now shift the focus to the centralized operation, which is seen in Fig. 6.12. Once again, the DCC methods that do not cancel interference, such as PTA and NTA), are the ones unscalable with respect to the number of UEs. The remaining methods are scalable, given the used combining vector is also scalable. Contrary to this is the case of L-MMSE and opt-L-MMSE, which, as previously discussed, are unscalable. As for the combiner, LP-MMSE with no LSFD has Orthogonal-PTA as the one that has the lowest complexity with respect to K. It is also the one that has the best performance. For nearly optimal LP-MMSE, the performance for any PTA is weak, so the viable and scalable solutions are OUA and ONTUA. Finally, with respect to nearly optimal MR and MR without LSFD, we have that on the former either OUA or ONTUA could be used in order to reduce complexity and attain good performance. On the latter case, the performance is very similarfor the three interference canceling DCC methods, therefore we should use Orthogonal-PTA since it is less complex.

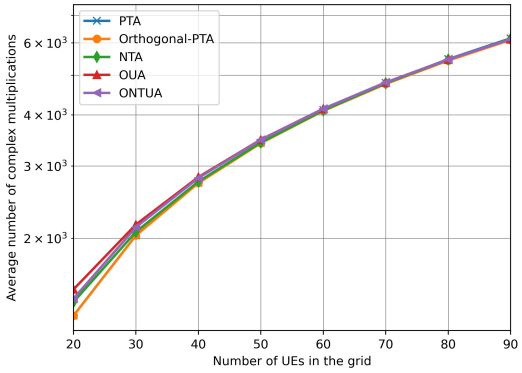
On the performance, when adding more UEs but keeping other grid parameters constant, the newly added UEs will have a high probability of being assigned an already used pilot sequence to estimate its channels. Therefore, the SE values are expected to decay with K. For simulation purposes, we opt not to use any of the unscalable combining vectors, as they would require a large amount of computation to obtain the SE values. Thus, we shall use the



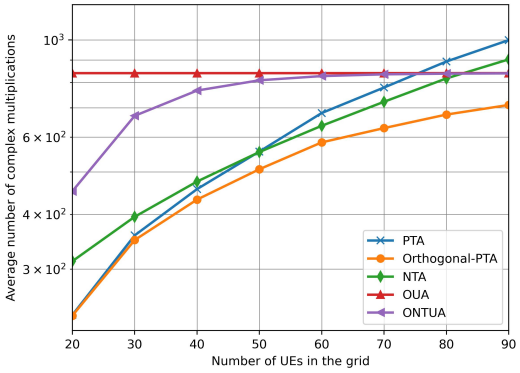
(a) optimal L-MMSE



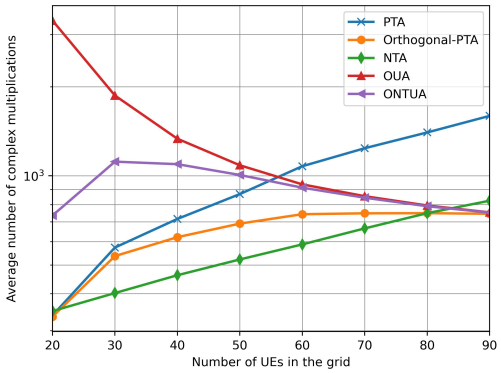
(b) near optimal LP-MMSE



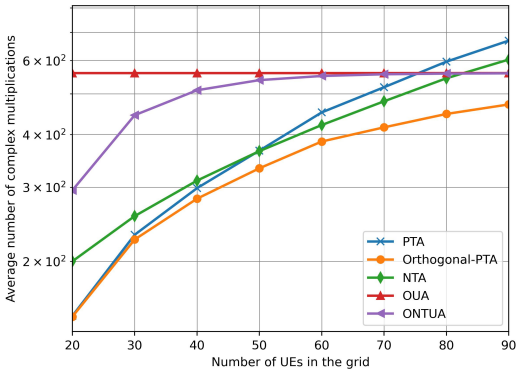
(c) L-MMSE



(d) LP-MMSE



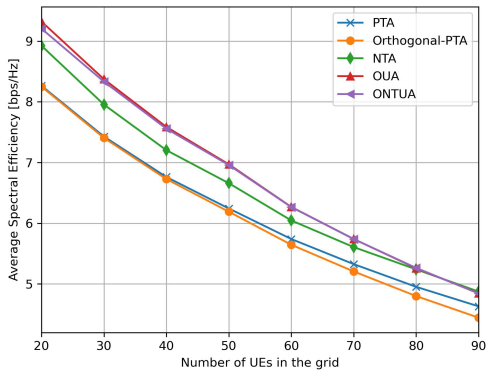
(e) near optimal MR



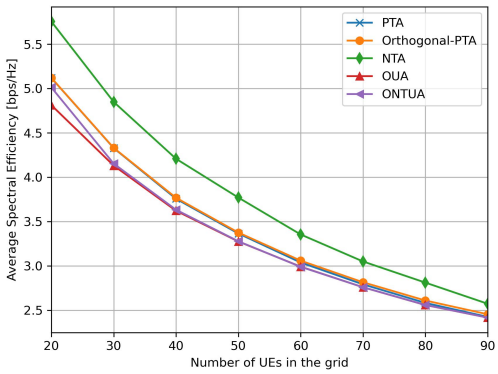
(f) MR

Figure 6.12 – Average number of complex multiplications for the distributed operation combining vectors for the PTA, Orthogonal-PTA ,NTA, OUA and ONTUA versus total number of UEs in the grid

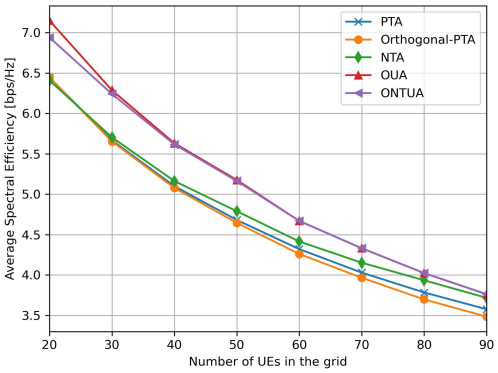
P-RZF and MR combiners for centralized operation and nearly optimal LP-MMSE and MR for distributed operation. Fig. 6.13 shows those results It cab be seen that the performance is



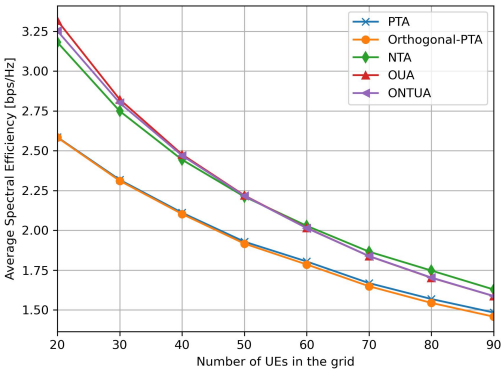
(a) P-RZF



(b) MR



(c) nopt-LP-MMSE

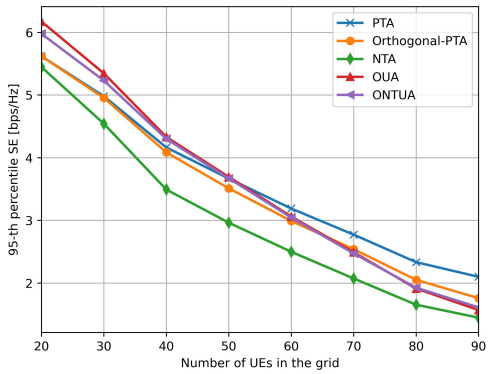


(d) nopt-MR

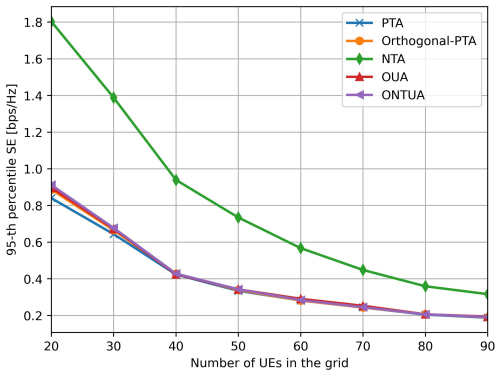
Figure 6.13 – Average SE for the centralized and distributed operations combining vectors for the PTA, Orthogonal-PTA ,NTA, OUA and ONTUA versus total number of UEs in the grid

generally higher when using the OUA and ONTUA methods. They are also the most scalable methods with respect to the number of UEs in the grid, alongside Orthogonal-PTA. With respect to Orthogonal-PTA, we notice that its complexity is much lower than the two other methods when K is not as large. Therefore, despite its SE loss when compared to OUA and ONTUA, the preferred DCC method one should apply when striving for complexity reduction is Orthogonal-PTA since, despite performance loss, complexity is much lower. If one instead wishes for performance, then ONTUA for P-RZF should be used and NTA for centralized MR.

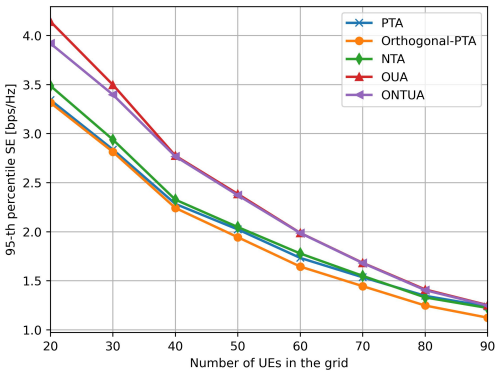
When comparing distributed operation, we notice the considerable performance loss of PTA and Orthogonal-PTA methods when LSFD weights are used. Therefore, ONTUA or OUA should be used for distributed case. Moreover, if complexity reduction is desired, ONTUA would be the ideal choice. For lower values of K, NTA could be also used, since it has lower complexity than OUA and ONTUA with similar SE results.



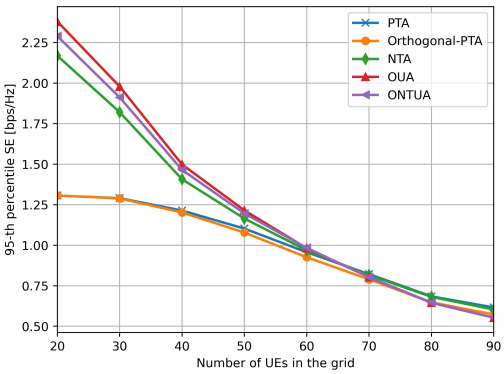
(a) P-RZF



(b) MR



(c) nopt-LP-MMSE



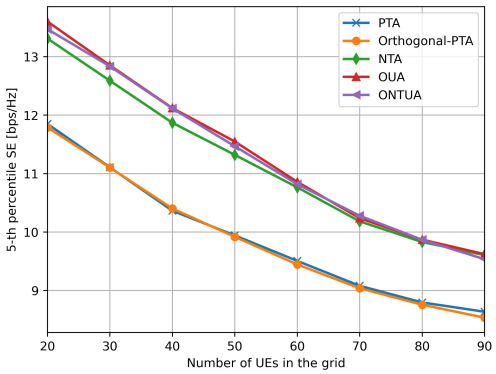
(d) nopt-MR

Figure 6.14 – 95-th percentile SE for the centralized and distributed operations combining vectors for the PTA, Orthogonal-PTA ,NTA, OUA and ONTUA versus total number of UEs in the grid

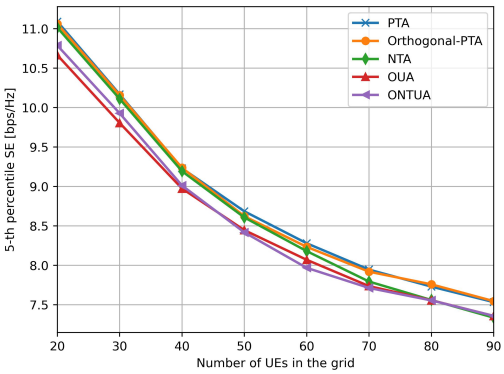
We can also conclude with more certainty that PTA and Orthogonal-PTA do not work as efficiently when the subject is increasing the network SE, since their performance is always the lowest, as evidenced in Fig. 4.12 of Chapter 4, regardless of the combining vector. We see that NTA and ONTUA do a better job performance-wise, and their SE values are very close. Therefore ONTUA is a great alternative for OUA, since it generally requires much less complex multiplications to compute the combining vector and the channel estimates.

We now compare the performance by considering the 95-th and 5-th percentiles of the SE. We use the same combining methods as Fig. 6.13 and the results are shown in Fig. 6.14 and Fig. 6.15

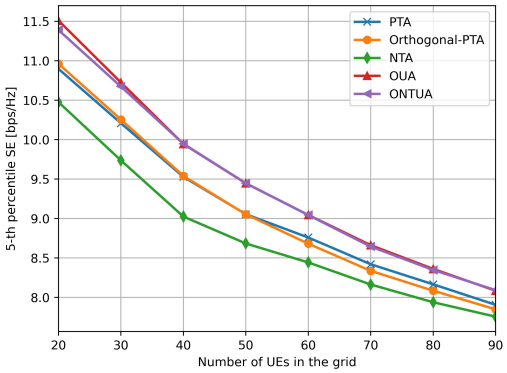
From Fig. 6.14 we can see that the NTA generally is incapable of providing a satisfactory SE for the UEs with worse channel condition. The main reason for that is the lack of interference mitigation techniques that OUA has, to exemplify. The exception for that is centralized MR for reasons already discussed. However, as the number of UEs grow large, the performance gets increasingly similar to that of OUA, since the percentile of connected APs to



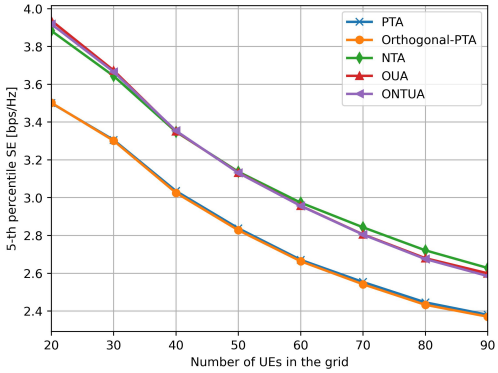
(a) P-RZF



(b) MR



(c) nopt-LP-MMSE



(d) nopt-MR

Figure 6.15 – 5-th percentile SE for the centralized and distributed operations combining vectors for the PTA, Orthogonal-PTA ,NTA, OUA and ONTUA versus total number of UEs in the grid

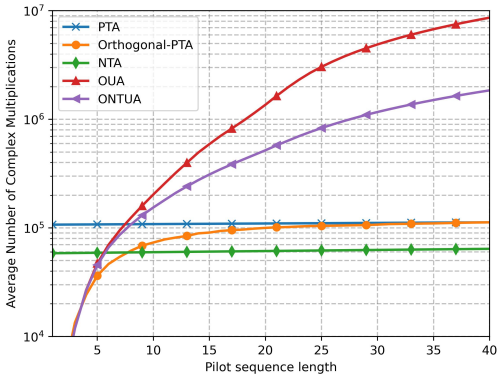
each UE gets larger. Considering PTA and Orthogonal-PTA, the performance for distributed operation with LSFD weights is the weakest, for reasons already exposed. For centralized operation. however, we can see that they possess a rather satisfactory performance while keeping complexity manageable, such as the (Orthogonal-PTA case. Therefore, for centralized operation, PTA and Orthogonal-PTA are methods with great fairness values, since the weakest channel UE is likely to have a reasonable value of SE.

When we take a look at Fig. 6.15, the obvious result is that OUA and ONTUA are the methods that guarantee the highest performance for the UEs with the greatest channel conditions. As opposed to the 5-th percentile case, the PTA and Orthogonal-PTA methods do not have such a great performance when considering the best UEs. That happens because those methods aim to establish an power percentage threshold, which imposes all of the UEs having that percent of the total grid power they would have when connecting to all APs. Since ONTUA and OUA do not have such imposition, the high quality channel UEs can increase their throughput. For this situation, NTA also works well, since the lack of interference management is likely to affect the weaker UEs, but not the strongest. It is preferred, however, to use ONTUA due to its low complexity when compared to NTA, an unscalable method.

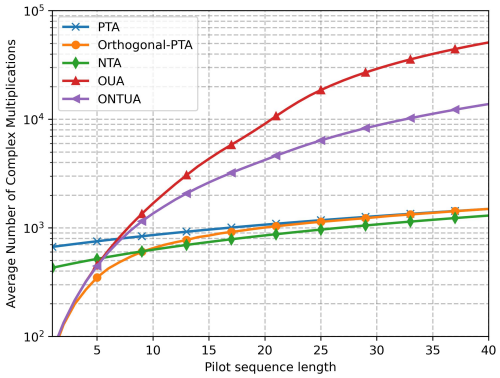
From the simulations, we can then conclude, for centralized operation, that the ONTUA method is the one that has the highest SE values while simultaneously being scalable with respect to K. OUA has also those characteristics, but has almost the same performance as ONTUA, with greater complexity. Also, when the goal is to prioritize fairness in the SE distribution, then Orthogonal PTA might be used, with reduced complexity. NTA and PTA, while having acceptable performance, are unscalable, thus should not be used. When the focus is on maximizing the net SE while increasing K, then undoubtedly ONTUA should be chosen, since it has the best performance and lowest complexity among the scalable methods. If we consider distribution operation, however, one should choose ONTUAor OUA, since the PTA and Orthogonal-PTA are inefficient in scenarios where there is LSFD, which modern distributed operation most commonly systems use. NTA once again is an alternative, but its scalability makes it a not so viable solution. Our final conclusion, therefore, is that ONTUA is the DCC method that offers satisfactory performance in terms of fairness, and in terms of the overall grid SE, while being scalable with respect to K, and, moreover, having the lowest complexity among the scalable methods.

6.1.4 Complexity and performance versus pilot sequence length

We would like now to check scalability for the DCC methods with respect to the pilot sequence length τ_p. Intuitively, we can see that OUA is unscalable on that regard. That can be proven when considering the case of $\tau_p = K$. Since each AP is allowed to be connected



(a) P-MMSE



(b) nopt LP-MMSE

Figure 6.16 – Average number of complex multiplications for the centralized and distributed operations combining vectors for the PTA, Orthogonal-PTA ,NTA, OUA and ONTUA versus pilot sequence length τ_p

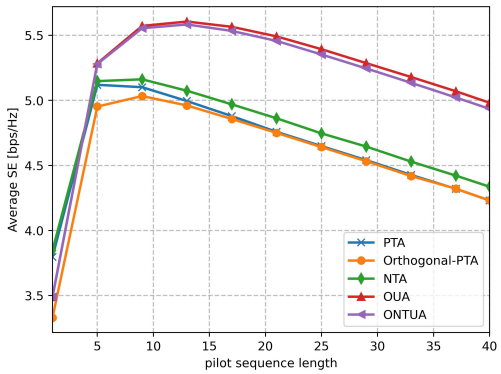
to τ_p UEs, then on that case, K UEs will be connected to each AP. Therefore, all APs will serve all the UEs in te grid, and we have seen that this leads to unscalability. When we consider instead ONTUA, we see that a maximum of τ_p UEs can be connected to each AP. This number, however, is determined by the threshold limit, which we have set as -0.23. By Fig. 4.9 we have that this value allows on average 40 % of the total τ_p UEs to be connected to each UE, which grows with τ_p, but at a much slower rate than OUA. NTA and PTA do not depend on τ_p, and therefore are scalable. Finally, Orthogonal-PTA follows the same logic as ONTUA, so it is unscalable, but its complexity grows at a much lower rate than OUA.

We now present the complexity for centralized and distributed operations. For simulation purposes, we have selected the P-MMSE combiner for the centralized operation and nopt-LP-MMSE for the distributed operation, Also, we have used the GPA for pilot allocation, since it does not depend on the condition of K being a multiple of τ_p. The results are shown in Figure 6.16: As one would expect, the average number of complex multiplications is a crescent function with respect to τ_p, even in the cases when it was stated that the DCC method was invariant, since channel estimation complexity still depends on τ_p. It is seen that PTA, Orthogonal-PTA and NTA are the methods that have the lower complexity values, while OUA and ONTUA have the highest. For both PTA methods this is explained by the fact many UEs will connect to few APs in order to satisfy the power percent threshold. In Orthogonal-PTA even lower AP-UE pairs will be connected, since each AP is only allowed to serve orthogonal UEs. Although ONTUA complexity still seems pretty high, when compared to PTA and Orthogonal-PTA, one has to keep in mind that, for $\tau_p = 20$, the OUA method has almost three times the complexity of ONTUA, which, when we are leading with complexity orders of 10^5 is a very significant gap. Considering the case of no frequency reuse $\tau_p = 40$, we have an even bigger gap, of 8 times the complexity of ONTUA. Therefore, although the number of complex multiplication is still high for ONTUA, it is still scalable and lead

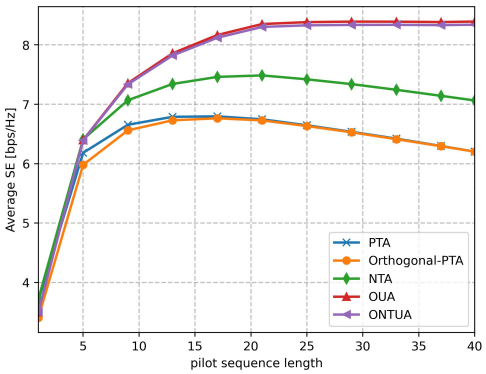
to satisfactory performance, as we see in Fig. 6.17, that shows the 5-th, average and 95-th percentile of the SE. For simulation purposed, we have selected the P-RZF instead of the P-MMSE combiner, since it has a similar performance, as we have seen, only a small SE loss between them. We can immediately notice that the increased value of τ_p is more detrimental to performance in the distributed operation than on centralized. That happens due to inferior quality of channel estimation, since it is done locally, and without information of the other APs channel, in the distributed case. Thus, the overhead caused by limiting the amount of useful data has a much large impact on this system than in the centralized. Also, we can see that the performance of ONTUA and OUA are very similar in every situation, further showing why ONTUA can be chosen as an efficient DCC method.

Now analyzing the average SE, we notice that PTA and Orthogonal-PTA have the lowest performance, because they aim to guarantee a power threshold that in fact does not allow the UEs with good channel condition to achieve the higher SE they could achieve, since usually those will only be connected to one or two APs. The same being extended to the average channel UEs, which will be usually connected to few APs. The NTA, achieves higher average performance, because more AP are connected to each UE. Also, as the length of the pilot sequence grows, the amount of interference between UEs that are served by the same AP tends to decrease, up to a point where useful data is limited and the overhead becomes too big. Finally, for ONTUA and OUA, performance is the highest, because they allow for, on average, the same number of AP to be connected to each UE.

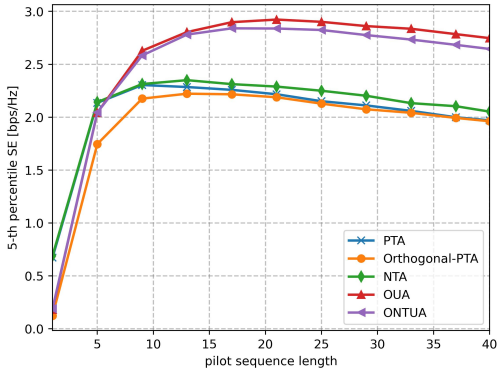
Focusing on the UEs with the worst channel conditions, we can see that, for distributed operation, both PTA and the NTA have the same performance, much lower than the other methods, due to the lack of interference mitigation techniques allied with lower channel estimation quality inherent to distributed operation. For centralized operation, the performance of both PTA methods is greater than the NTA, because they aim to provide satisfactory performance for the weakest UEs, due to the already discussed fairness metric. Finally, when analyzing the best channel condition UEs, we can see that the performance for distributed operation has a curious behavior: it decreases significantly after $\tau_p = 10$. In fact, for some DCC methods, once the pilot sequence length is close to its maximum value, the performance is worse than the case of full frequency reuse, when all the UEs share the same pilot. Although counter intuitive at first, this is explained by the fact the stronger UEs do not benefit from increasing quality channel estimation the same amount weaker UEs do. Additionally, they do benefit from lower pilot sequence lengths, since more useful data can be transmitted by coherence block length. Therefore, with less useful data and no additional benefit from channel estimation quality, the 95-th percentile SE decreases with the length τ_p. On centralized operation a similar behavior is noticed, however not as strong, since channel estimation is better (thus, the effects of reduced performance are apparent at $\tau_p = 20$). In regards to the DCC method, the only ones that doe not provide a satisfactory performance are the



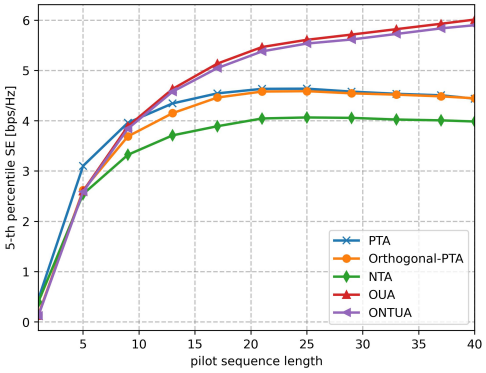
(a) near optimal L-MMSE



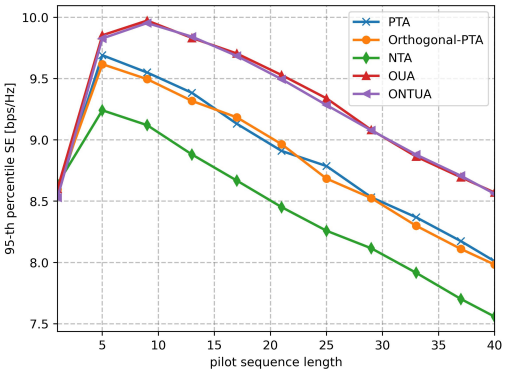
(b) P-RZF



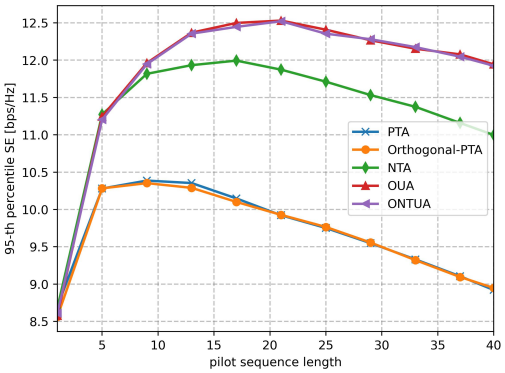
(c) near optimal LP-MMSE



(d) P-RZF

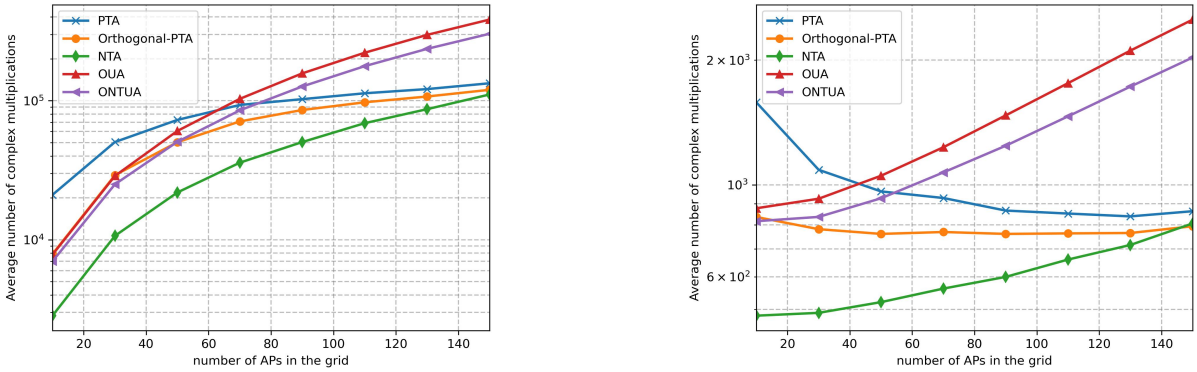


(e) near optimal LP-MMSE



(f) P-RZF

Figure 6.17 – Average, 5-th percentile and 95-th percentile SE for the distributed and centralized operations combining vectors for the PTA, Orthogonal-PTA ,NTA, OUA and ONTUA versus pilot sequence length



(a) P-MMSE

(b) nopt LP-MMSE

Figure 6.18 – Average number of complex multiplications for the centralized and distributed operations combining vectors for the PTA, Orthogonal-PTA ,NTA, OUA and ONTUA versus number of APs in the grid

PTA. NTA does have a performance gap in comparison to OUA and ONTUA, however the complexity for NTAis far smaller for higher pilot sequence lengths. In general, OUA and ONTUA have the better performances, with ONTUA having the benefit of scalability.

One could notice that complexity reduction is not such a benefit when leading with high mobility systems, in which the overhead significantly impact the SE for large pilot sequences length. However, when dealing with high mobility systems, complexity reduction becomes a factor due to the advantage of using large τ_p values without large overhead. Therefore, in such cases ONTUA becomes a solid alternative to OUA.

6.1.5 Complexity and performance versus number of APs

Analyzing the scalability and performance of the DCC methods, when we add APs to the grid. From intuition, one can say that a large number of APs will lead to an increasing value of the cluster $\mathcal{M}_k$ size, irrespective of the method. However, we must notice that OUA will always assign τ_p UEs to each added AP, and therefore lead to an ever increasing value of the AP cluster length for UE k. Since the addition of new APs will not affect the normalized channel gains distribution, since they are normalized with respect to the UEs number, it is expected that the same amount of UE-AP pairs be formed, which in turn should make the complexity of NTA increasing with the number of APs, but still lower than OUA and ONTUA. Such assumptions can be observed at Fig. 6.18. We have assumed that RP has once again be used for pilot assignment. For the PTA and Orthogonal-PTA, we see that the complexity curve grows slower in the centralized case, while in distributed operation it actually decreases. The reason is that, as new APs are added to the grid, the likelihood of an UE being connected to one or few APs increases. For instance, if there is an UE such that it has weak channels to every AP in the grid, and one new AP is placed nearby it, at

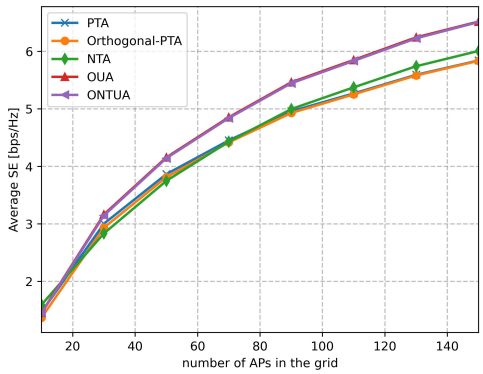
this moment, instead of connecting to a several number of APs in order to satisfy the power threshold, the UE now has to connect to one or few, which in turn impacts on complexity reduction. Therefore, the most scalable methods are PTA and Orthogonal-PTA in distributed operation, and NTA on centralized.

Analyzing the performance for the 5-th percentile, 95-th percentile and average SE, which are shown in Fig. 6.19. As it would be expected, the performance increases with the number of APs, because of more degrees of freedom for propagation, provided by the geographical location of the newly added APs. With respect to the average SE value, we can notice that OUA and ONTUA have virtually the same performance, once again proving that ONTUA is an effective alternative in complexity to OUA. Also, for distributed operation, the performance of NTA and both PTA methods are similar, and lower than OUA and ONTUA. For centralized operation, the performance of NTA is higher, but still lower than OUA and ONTUA. The reason for that difference between operations is that nopt LP-MMSE uses the LSFD for fusing the data estimates, which in turn makes interference from the UEs being served by the same AP even more aggressive. When looking at performance for the weakest UEs, we see that the NTA and both PTA methods have the lowest SE values, in distributed operation, for reasons already discussed. When taking a look at centralized operation, however, PTA becomes a very solid alternative, due to its objective in fairness. Allied with the significant complexity reduction, Orthogonal-PTA is therefore a strong candidate for an scalable system with respect to the number of APs when the objective is a fairer SE distribution. Finally, when we analyze the performance of the strong UEs, the distributed operation can be explained in the same fashion as the 5-th SE percentile curves. For centralized operation, however, we see that an increased number of APs lead to an increase in the number of connected UE-AP pairs, which in turn increases the performance of the strong UEs, since they are connected to more APs and do not suffer the effects of interference as strongly as weaker UEs, assuming an efficient pilot allocation method has been used, such as RP.

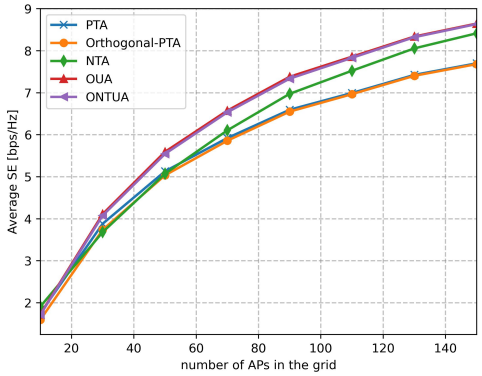
Finally, we can conclude that, for distributed operation, the method that offers better scalability with satisfactory SE in all of the cases is ONTUA. However, when leading with centralized operation, the NTA method acts as efficiently and with reduced complexity when increasing performance for the UEs with the best channel is the main goal, and the Orthogonal-PTA method is the most efficient when the goal is a fair SE distribution among the grid.

6.1.6 Effect of threshold on ONTUA and Orthogonal-PTA on performance and complexity

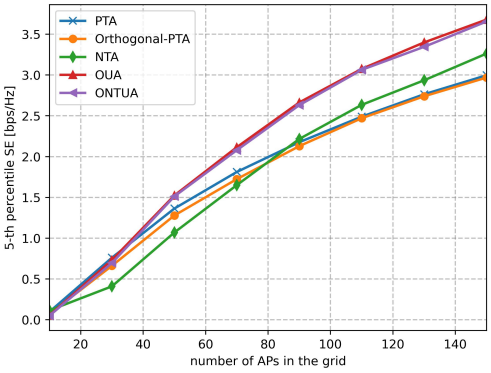
The NTA and the PTA methods have been shown to be unscalable with respect to K. We would like to verify the scalability for ONTUA and Orthogonal-PTA by varying their



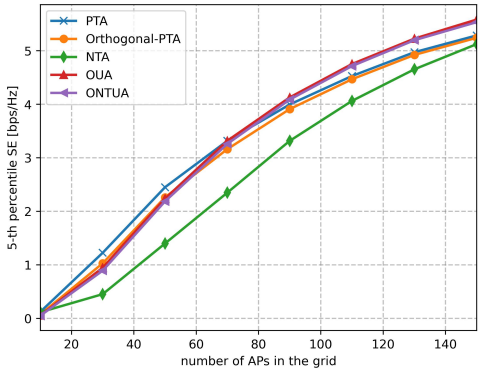
(a) near optimal L-MMSE



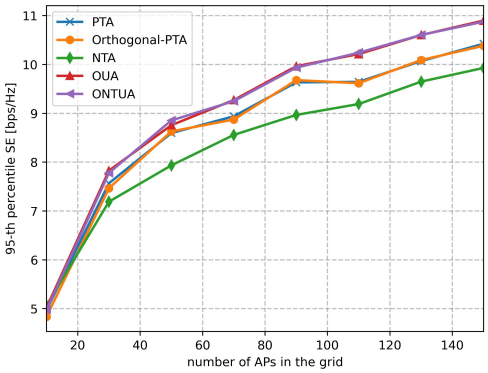
(b) P-RZF



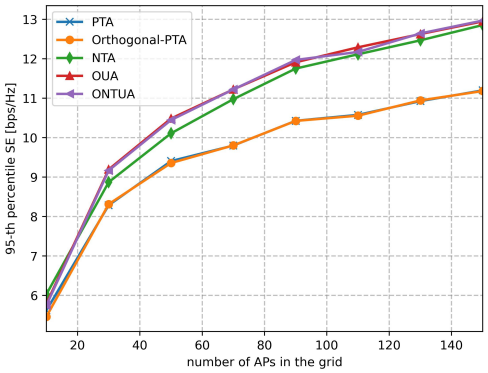
(c) near optimal LP-MMSE



(d) P-RZF

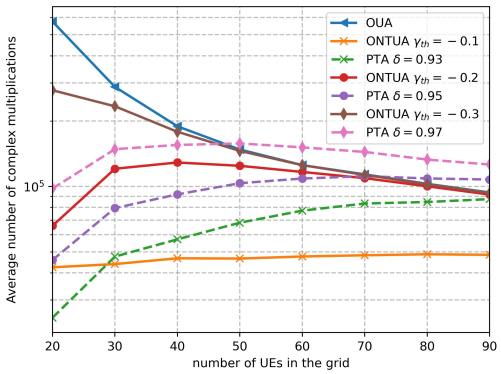


(e) near optimal LP-MMSE

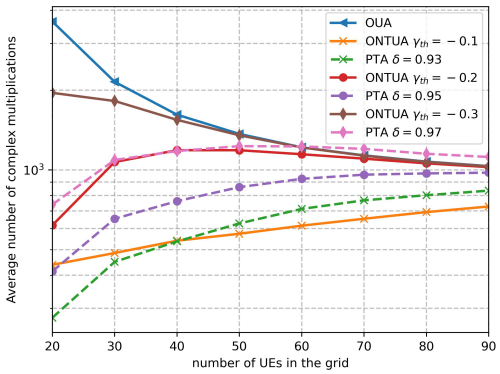


(f) P-RZF

Figure 6.19 – Average, 5-th percentile and 95-th percentile SE for the distributed and centralized operations combining vectors for the PTA, Orthogonal-PTA ,NTA, OUA and ONTUA versus number of APs in the grid



(a) P-MMSE



(b) nopt LP-MMSE

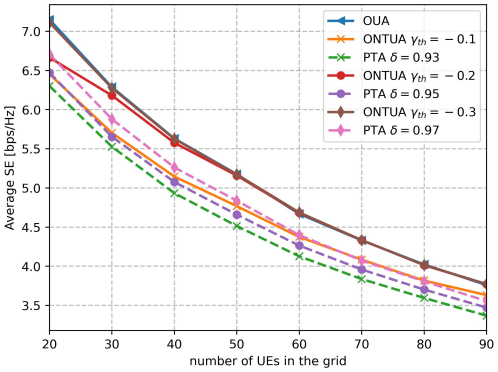
Figure 6.20 – Average number of complex multiplications for the centralized and distributed operations combining vectors for the , Orthogonal-PTA , OUA and ONTUA versus number of UEs in the grid, for various system parameters.

threshold parameters, We would like also to investigate the performance under such variations. For that, we refer to the same system we have been using to simulate the figures of this chapter, but we consider $\lambda_{\text{th}} \in \{-0.1, -0.2, -0.3\}$, and also $\delta \in \{93, 95, 97\}\,\%$. We once again use the combining vector P-MMSE for investigating complexity and P-RZF for the SE, in centralized operation. In distributed operation, we use nopt LP-MMSE for all of the cases. The results are shown in Fig. 6.20 and Fig. 6.21.

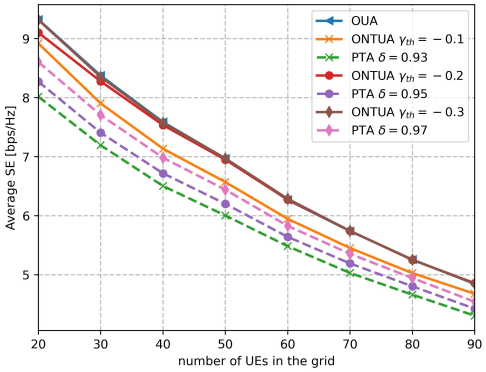
From Fig. 6.20, we can see that system complexity increases as the threshold value γ_{th} gets lower. That is to be expected, since a lower value means that more UEs will be allowed to connect to the AP. We can observe the dependence of the threshold on the number of UEs: for $K = 40$, the ONTUA with $\gamma_{\text{th}} = -0.3$ is such that its complexity is the same as OUA, which strongly suggests that every AP is connected to τ_p UEs. That happens because at this value, the percentile of connected UEs is equal to 100 %, which, in the case of ONTUA means the percent of the strongest τ_p UEs that are connected to a given AP. For $\gamma_{\text{th}} = -0.2$, the complexities of OUA and ONTUA become similar when $K = 60$, for the same reasons.

Comparing the Orthogonal-PTA method, it is clear that the higher the power threshold, the higher the complexity. In fact, we can argue against the increasing value of this parameter, since for $\delta = 97\%$, the complexity becomes similar to that of ONTUA and OUA, for worse performance, as we will see next.

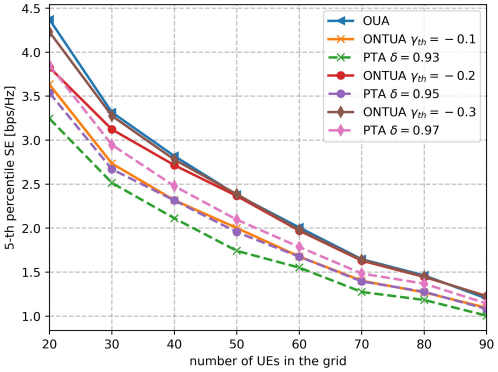
Regarding the average SE, we can see that the Orthogonal-PTA method leads to poor performance, however the complexity drop makes it a viable choice. When the power threshold δ is high, however, there are no significant performance gains and the complexity value gets near to those of ONTUA. Therefore, Orthogonal-PTA is only a viable choice if we set the power threshold to values below $\delta = 95$. On that regard, we notice by looking at the 5-th percentile of the SE, that results for Orthogonal-PTA are rather satisfying for centralized



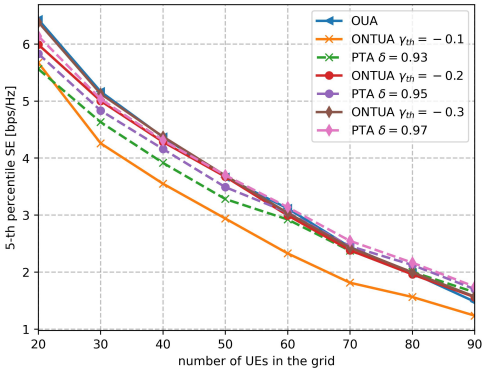
(a) near optimal L-MMSE



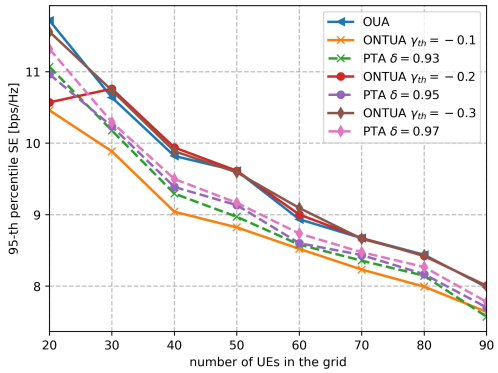
(b) P-RZF



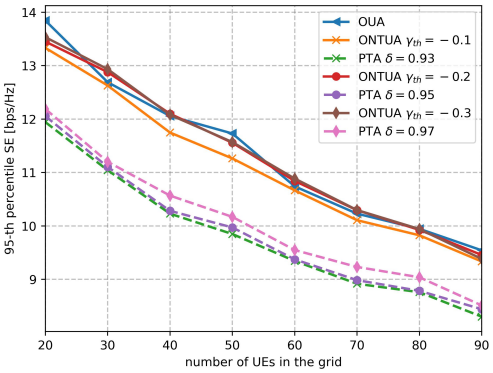
(c) near optimal LP-MMSE



(d) P-RZF



(e) near optimal LP-MMSE



(f) P-RZF

Figure 6.21 – Average, 5-th percentile and 95-th percentile SE for the centralized and distributed operations combining vectors for the , Orthogonal-PTA , OUA and ONTUA versus number of UEs in the grid, for various system parameters.

operation, even with lower threshold values. For distributed operation, they can still be acceptable, due to the reduced complexity, however, given that complexity is not such an issue in distributed operation, ONTUA still is more viable. Finally, when analyzing the 95-th percentile, we see that Orthogonal-PTA is definitely not a viable choice, since the SE values are pretty small. One curious fact is that ONTUA with the highest threshold value $\gamma_{\text{th}} = -0.1$ has a similar performance to OUA in centralized operation. That can be explained by the fact ONTUA still lets the strongest UEs connect to the AP. On distributed case, however, the presence of the LSFD weights makes that the AP is assigned a unfavorable weight for the UE, but reasonable overall, and with the lack of other connected APs to the UE, it experiences the SE loss seen in Fig. 6.21f.

To summarize, when the main goal is to provide satisfactory performance for the UEs with best channel quality, then ONTUA might be used with a low threshold value, which has the lowest complexity values, if centralized operation is considered. For the distributed, we can again make use of ONTUA, but with a higher threshold value. For the case of the lower percentiles, we can use Orthogonal-PTA to guarantee satisfactory performance and lower complexity, in centralized operation. For distributed, we can use ONTUA, with lower threshold values. Finally, when the main goal is to improve the average SE, then Orthogonal-PTA is not a viable solution, and neither is ONTUA with lower threshold values. To better results, we see that we must use a threshold value of γ_{th} of around -0.2, the reason why this value was chosen as -0.23 for almost every scenario.

6.1.7 Summary of the DCC methods scalability and performance

Having compared the DCC methods for various system parameters, we now summarize them, with regard to scalability, in Table 6.2.

	Centralized Operation			Distributed Operation		
Parameter / DCC Method	UEs	Pilots	APs	UEs	Pilots	APs
PTA	No	Yes	Yes	No	Yes	Yes
Orthogonal PTA	Yes	Yes	Yes	Yes	Yes	Yes
NTA	No	Yes	Yes	No	Yes	Yes
OUA	Yes	No	No	Yes	No	No
ONTUA	Yes	Yes	No	Yes	Yes	No

Table 6.2 – Scalability for the DCC methods. Centralized operation assumes use of P-MMSE and distributed use of nopt LP-MMSE.

Just by looking at scalability alone we can select the Orthogonal-PTA as the most scalable method. PTA, NTA and ONTUA also offer scalability for at least one of the three parameters,

with the most unscalable method being OUA, being only scalable with respect to the number of UEs. With that in mind, we analyze Table 6.3 with the SE performance.

	Average SE			5-th Percentile SE			95-th Percentile SE		
Parameter / DCC Method	UEs	Pilots	APs	UEs	Pilots	APs	UEs	Pilots	APs
PTA	1.3	2.3	0.9	0.8	1.5	0.2	1.7	3.0	1.8
Orthogonal PTA	1.3	2.3	0.9	0.8	1.5	0.3	1.7	3.0	1.8
NTA	0.4	1.5	0.2	0.95	2.0	0.8	0.4	1.0	0.2
ONTUA	0.05	0.02	0.01	0.3	0.01	0.02	0.1	0.03	0.01

Table 6.3 – Largest SE gap for the DCC methods. Centralized operation is assumed and the P-RZF combiner is used. The "Best" legend correspond to the DCCmethod with the highest values, and the other values correspond to the highest SE gap between the DCC method and the highest DCC method.

	Average SE			5-th Percentile SE			95-th Percentile SE		
Parameter / DCC Method	UEs	Pilots	APs	UEs	Pilots	APs	UEs	Pilots	APs
PTA	0.9	0.8	0.6	1.0	0.6	0.6	0.5	0.5	0.5
Orthogonal PTA	0.9	0.8	0.6	1.0	0.6	0.6	0.5	0.4	0.4
NTA	0.9	0.7	0.5	0.8	0.5	0.3	1.0	0.9	0.9
ONTUA	0.4	0.05	0.01	0.3	0.1	0.01	0.1	0.01	0.01

Table 6.4 – Largest SE gap for the DCC methods. Distributed operation is assumed and the nopt LP-MMSE combiner is used. The "Best" legend correspond to the DCCmethod with the highest values, and the other values correspond to the highest SE gap between the DCC method and the highest DCC method.

From Tables 6.2, 6.3 and 6.4, we can see that, when analyzing the total number of UEs in the grid, OUA has the best SE performance. Since it is scalable, it would be adequate to chose this as the DCC method. However, since the gap between the SE values of ONTUA and OUA is small, and ONTUA has lower complexity, then ONTUA will be the chosen method, regardless if distributed or centralized operation is considered.

When analyzing the pilot sequence length parameter, then ONTUA is the method with scalability and satisfactory SE values. However, at the expense of some SE loss, we can choose to use NTA when the objective is to increase the SE of the strongest UEs, and Orthogonal-PTA if we wish for more fairness. For distributed operation an alternative to ONTUA, if some SE loss is allowed, is NTA, whatever the SE maximization goal is, due to reduced complexity. Finally when analyzing the effect of adding APs to the grid, we notice that ONTUA is not scalable, so we should choose between Orthogonal-PTA and NTA. For centralized case, we clearly see that NTA is the chosen method when the goal is to provide

a higher average SE or to enhance the SE of the strongest UEs. If the goal is to enhance the SE for the weakest UEs, then Orthogonal-PTAshould be used instead. For the distributed case NTA should be used in every case, since it has low complexity and the SE performance is very similar to Orthogonal-PTA, which is more complex. After those analysis, we can safely say that, if one DCC method should be chosen for an arbitrary network, then ONTUA should be chosen, since it overall demands lesser computation power than OUA, while having very similar performance. When complexity is high, it is not usually as unscalable, as for example, optimal MMSE: as we see from Fig. 6.18a, the complexity value tends to stabilize on a given value as the number of APs increases, and given Table 6.2, the only unscalable ONTUA parameter is the number of APs in the grid, L. For distributed operation, the complexity grows more linearly with L, but still it must not be an issue, as complexity on those systems is generally much manageable than centralized operation.

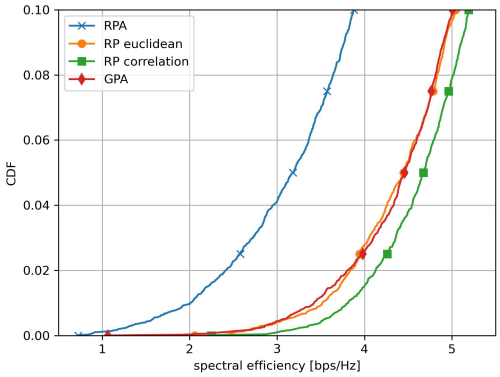
6.2 PILOT ASSIGNMENT METHODS PERFORMANCE COMPARISON

Comparing the pilot assignment methods that were introduced in Chapter 4. Unlike the DCC methods, we shall not simulate scenarios such as varying K and L, so the simulations should not take a large amount of computer power, and therefore we could use a more refined combining vector in centralized operation as P-MMSE instead of the P-RZF combiner.

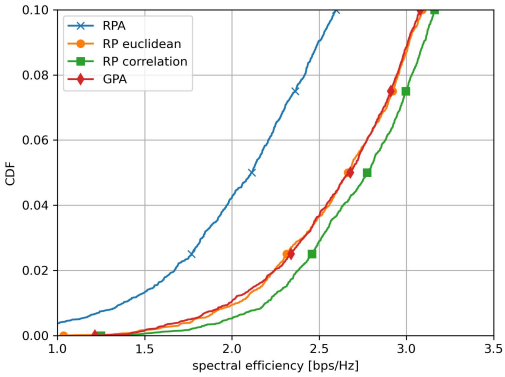
6.2.1 Comparison between pilot assignments CDF performance

As ONTUA is the DCC method that has the best trade off between complexity and performance, according to Section 6.1, we shall use it for AP clustering. Moreover, we will simulate with RPA, GPA and RP, with the correlation metric of Eq. 4.13, and the euclidean distance metric of Eq. 4.13. The CDF of the SE is shown in Fig. 6.22a

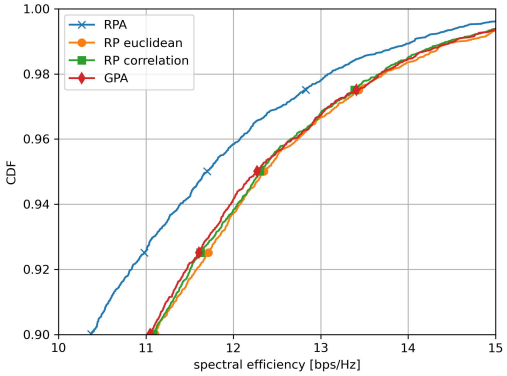
From Fig. 6.22a and Fig. 6.22b, that represent the CDF distribution of the SE up to 90% of the time, we see that RPA has the lowest values, due to the weaker UEs suffering the most from pilot contamination interference. The other methods that attempt to effectively mitigate pilot contamination have a much higher performance, with RP with the correlation metric being the most effective. GPA and RP with the euclidean distance metric have almost similar performances. The reason for performance gap is that the GPA metric assigns orthogonal pilots to the τ_p UEs randomly, which might lead to UEs that are geographically far apart, or experience very different spatial propagation characteristics, to be assigned distinct pilots. Therefore, the chance of a nearby UE sharing the same pilot sequence increase. In RP, the pilots are randomly allocated, but then they are rearranged into a new configuration,



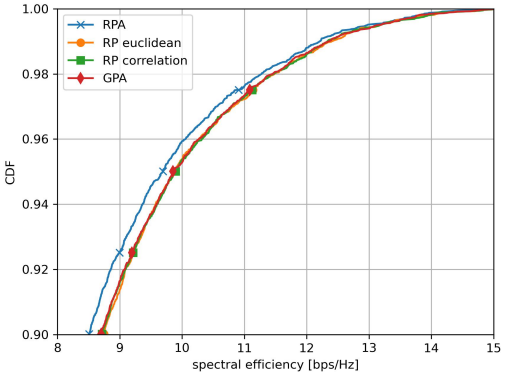
(a) P-MMSE



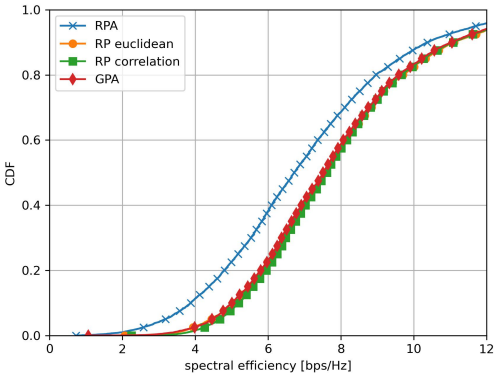
(b) near optimal LP-MMSE



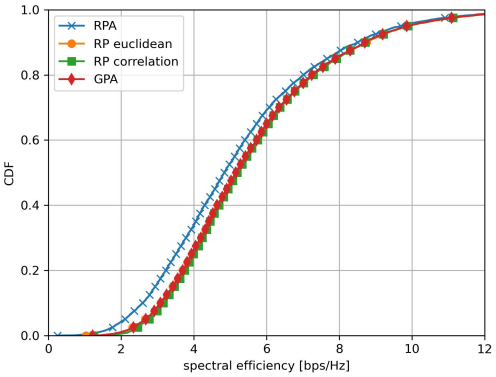
(c) P-MMSE



(d) near optimal LP-MMSE



(e) P-MMSE



(f) near optimal LP-MMSE

Figure 6.22 – CDF SE for the centralized and distributed operations combining vectors for the RPA, RP and GPA

whereas the τ_p UEs in GPA will always have mutually orthogonal pilots. As for the difference between RP metrics performance, it is explained because the euclidean distance metric does not take in account the correlation between large scale fading coefficients, but rather the distance between the collective vectors.

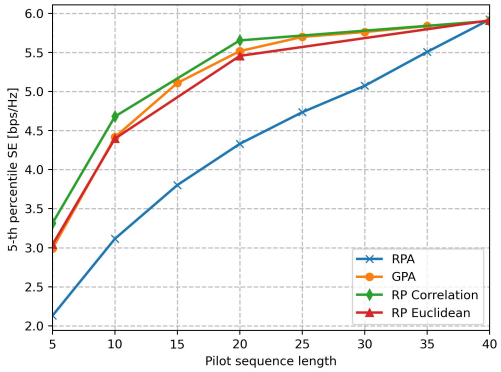
When comparing the overall SE CDF , we see that, apart from RPA, every other pilot assignment method has pretty much the same performance. When analyzing the SE up to 10% of the time, we see that RP and GPA have once again similar performance, except for RPA. This is explained by the fact the UEs with the best channel conditions do not suffer as much from the ineffectiveness of a bad pilot assignment, although they can still suffer from random allocation, as there is a chance two strong UEs that are geographically closer be assigned the same pilot. Therefore, RPA should be considered just a worst-case scenario, while RP with the correlation metric should be used if one desires a fair SE distribution. GPA and RP with the euclidean distance metric could be used in every other situation alongside RP with the correlation metric.

Regarding the operation, we see that centralized operation tends to achieve higher values of SE, since all of the collective channels are known for combining vector computation, but also that RPA has a much aggressive influence on this operation when compared to the distributed. This happens because distributed operation has the advantage of LSFD weights, that help mitigating the interfering UEs signals powers by assigning the weights of each AP. For instance, if an AP has many interfering UEs but is assigned a lower weight, then the signal of a desired UE, that connects to this AP and would be subjected to a strong interference, is fused with a small interference contribution at the CPU.

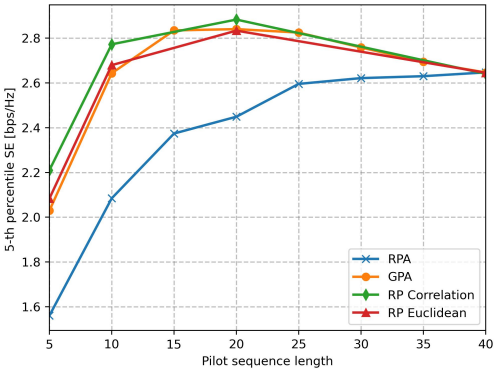
6.2.2 Comparison between pilot assignment performance versus pilot sequence length

Now we compare the behavior of the pilot assignment methods with the variations of the pilot sequence length τ_p. It is expected that system performance is the same irrespective of the method, in the special case of $\tau_p = K$, since there is no frequency reuse. Also, it would be expected that the gap between performance gets smaller the largest the value of τ_p, since channel estimation is improved and the interference between UEs diminishes, since there are less of them using the same pilot. Also, since we are using ONTUA, it would be expected the SE values saturate for a given pilot sequence length, and after to this point, its values gets progressively smaller, again for every pilot assignment method. The values of SE can be seen of Fig. 6.23

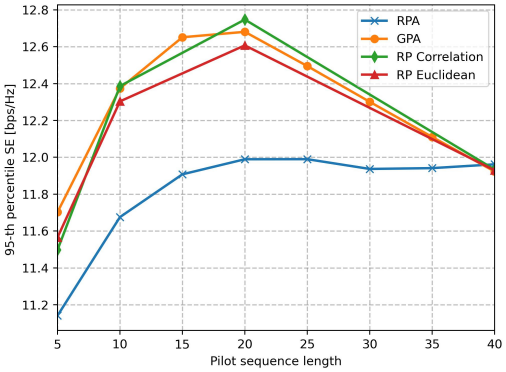
By Figs. 6.23a and 6.23b, we see that the performance for RPA is the smallest. GPA and RP with the Euclidean metric possess similar performance for smaller τ_p, but as it grows larger, GPA has the better performance. For the case of RP with the correlation metric, it has



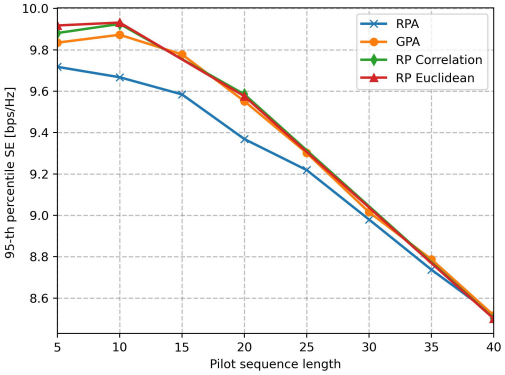
(a) near optimal P-MMSE



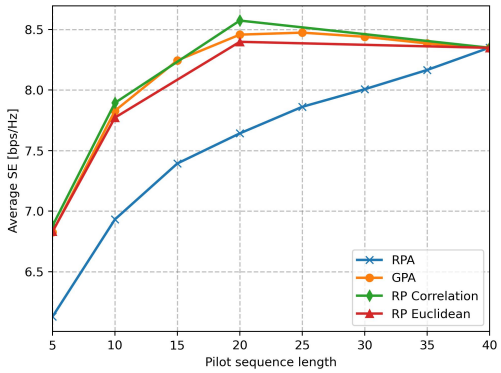
(b) nopt LP-MMSE



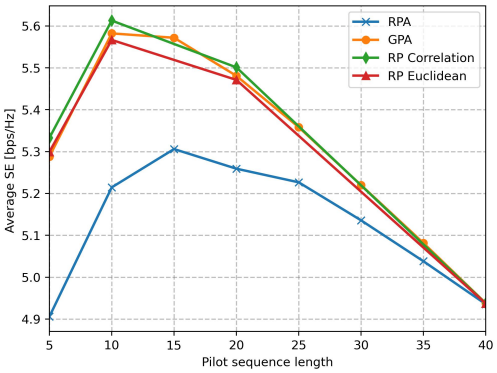
(c) P-MMSE



(d) nopt LP-MMSE



(e) P-MMSE



(f) nopt LP-MMSE

Figure 6.23 – Average, 5-th percentile and 95-th percentile SE for the centralized and distributed operations combining vectors for RPA, RP and GPA versus pilot sequence length, for various system parameters.

the best performance for smaller and larger τ_p, once again proving the point that this pilot assignment method has the objective of fair SE distribution. When we consider Figs. 6.23c and 6.23d, tha gap between GPA and RP with the Euclidean metric gets larger for centralized operation. This can be explained by the fact the Euclidean metric only considers the norms of the differences between the collective large scale fading coefficients. Therefore, it does not take in account the correlation of such vectors, which makes it possible to assign a distinct pilot sequence to UEs that are far apart, thus increasing the likelihood of UE interference. Since GPA minimizes the interference on a greedy manner, it takes to account the interference on the strongest AP only. On the distributed operation the gap is not as intense because of the LSFD weights presence, that tends to influence the behavior of all metrics, which in turn diminishes the likelihood of a strong UE interfering on an equally strong UE, that would in turn result in a SE drop. Finally, when analyzing the average network SE, we see that once again RP with the Euclidean metric has lesser SE values than the other methods, except from RPA, for reasons already stated. The performance of GPA and RP with the correlation metric are very similar, however. From this scenery, we can conclude once again that RPA consists on a reference case, and thus has limited applicability in real world networks, due to its huge performance loss, at least on the considered correlated Rayleigh channel model. Also, RP is a valid alternative for pilot assignment, but one should notice that it requires the collective large scale fading coefficients computation, while GPA requires only the strongest AP large scale coefficient for computation. In distributed operations, it could pose as a more scalable method, since it only requires the computation of one statistical parameter, whereas the other method would require the computation of the entire collective channel.

6.3 DCC AND PILOT ASSIGNMENT METHODS OVER UNKNOWN CHANNEL STATISTICS

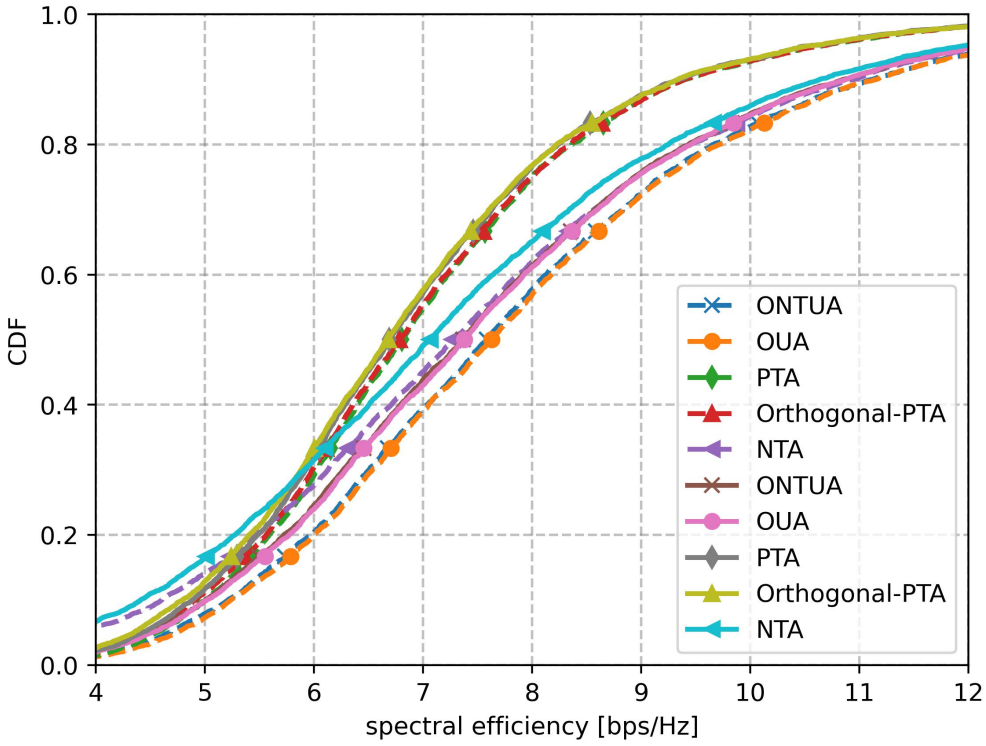
So far we have analyzed performance for various DCC methods and pilot assignment algorithms by assuming full knowledge of the spatial correlation matrices. Moreover, the described methods all require knowledge of the large scale coefficients, which are dependent on the channel statistics. Therefore, for satisfactory performance, some degree of knowledge of the spatial correlation matrices is required. This section aims to explore system performance on the case of imperfect channel statistics knowledge, and by that, evaluate if such methods could still be employed under those limiting situations with satisfactory reliability.

6.3.1 DCC methods SE comparison with imperfect statistics

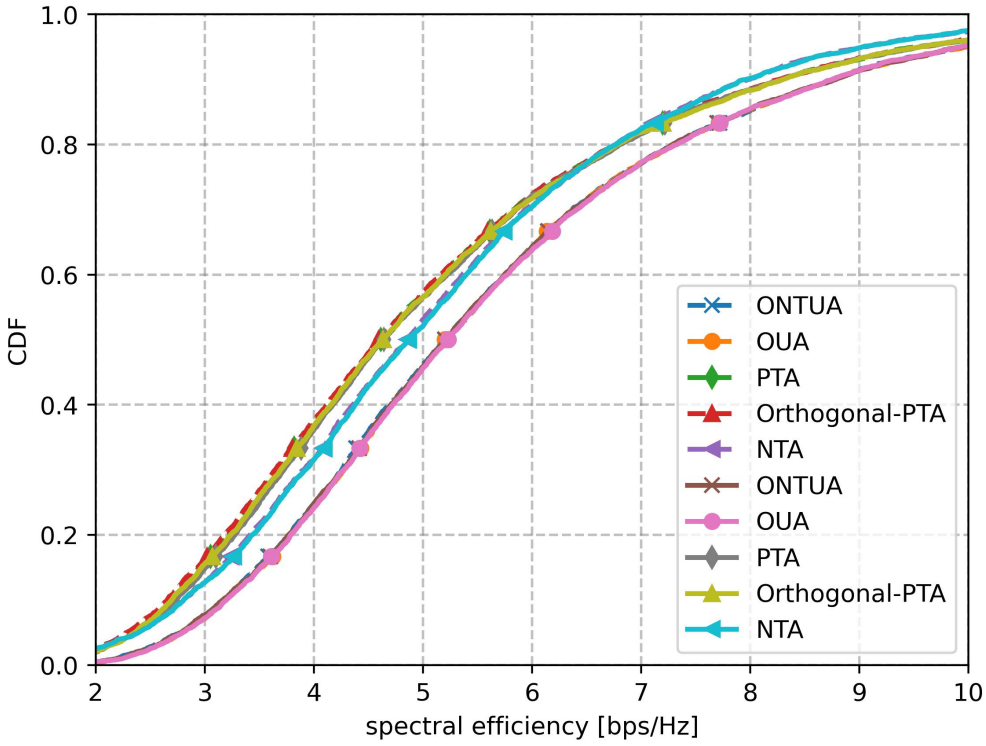
For comparison purposes, we shall select the RP method with the correlation metric. Also, since we consider $\tau_c = 200$, we have that a maximum of 200 transmissions may

be performed over each coherence block. Since for spatial correlation matrix estimation, we assume that τ_p UEs are simultaneously transmitting orthogonal pilot sequences while the others are not transmitting at all, if we consider $\tau_p = 10$, we have that 10 UEs will transmit at the same time. If we also consider $K = 40$, then the maximum number of transmissions that each UE is allowed per coherence block is equal to 5. Therefore, this is the chosen value for the number of transmissions $n_t = 5$. For the number of coherence blocks used for estimation, we can consider a large range of values, given that the channel statistics are invariant usually for hundreds of coherence blocks. Initially, we shall select $n_C = 100$, but we shall analyze the system for other n_C values. The results of DCC methods for known statistics and for unknown with $n_t = 5$ and $n_C = 100$ are given in Fig. 6.24a and Fig. 6.24b. Besides the considerations that were made in Section 6.1, we can see that the imperfect statistics, which are the dotted lines curves, knowledge does not seem to have a big impact on the SE values. The method that seems to have the smallest performance gap is Orthogonal-PTA, given that this method relies on the percentage power threshold to assign the UEs to the AP. Therefore, an imperfect knowledge of the large scale coefficients is not likely to impact which APs should be connected to the AP in order to satisfy the threshold. Despite that, on the OUA and ONTUA cases, the gap is larger, because the imperfect knowledge of the large scale fading coefficients will make very likely that one AP that has a weaker channel conditions to all UEs selects the UE with worst channel gain, therefore reducing performance. The same can be applied to NTA. We can also observe that imperfect statistics are more likely to impact negatively on the performance of weaker UEs. Stronger UEs usually have a larger channel gain with respect to an AP than the others. Therefore, an imperfect knowledge of the channel gains will not interfere with the ability of the DCC method of selecting the better UE for the AP or the better AP to the UE. To better explore the effects of spatial correlation matrices estimation, we should conduct an analysis on the behavior of the system for transmitted UL power variations. The spatial correlation matrices are still estimated over a range of coherence blocks $n_C = 100$, but now each UE transmits with a power p ranging from 1mW to 1W. The results are shown in Fig. 6.25. Some interesting conclusions may be made when we take a look on the distributed operation SE of Fig. 6.25b: we notice that the gap between the SE without and with perfect channel statistics is almost non existent, irrespective of the UL transmit power. This behavior is explained by the presence of the LSFD weights: the main limiting factor of imperfect large scale fading coefficients knowledge is the possibility of inefficient DCC AP assignment. However, since distributed operation actually lets each AP be defined a weight relative to an UE, and also, that the weaker UEs are the most impacted from imperfect statistics, the LSFD weights help diminishing the weights of highly impacted UEs, therefore reducing the impact of wrong DCC choices.

Another aspect of discussion arises when we analyze the behavior of imperfect PTA and Orthogonal-PTA systems for lower transmit powers: since in this case the SNR for the spatial correlation matrices estimation is low, then each of the large scale fading coefficients is

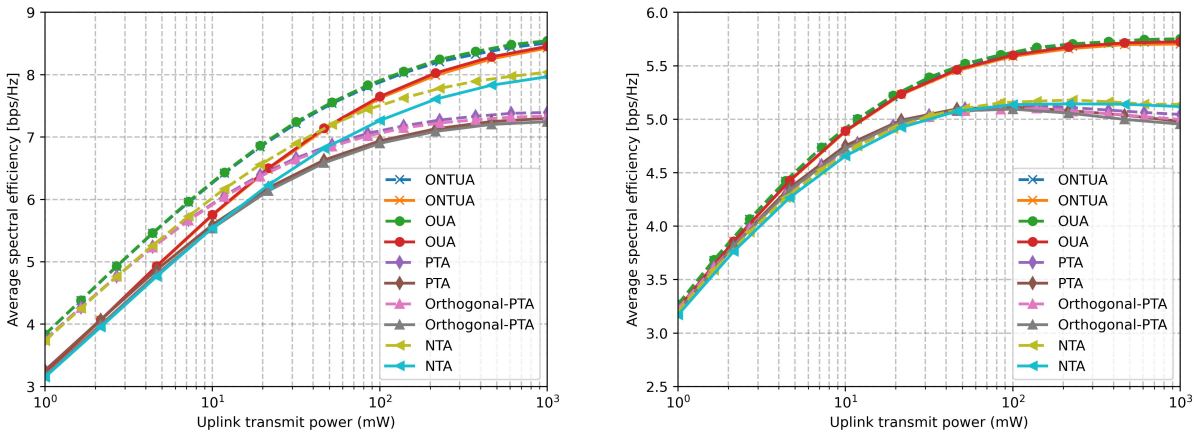


(a) P-MMSE



(b) nopt LP-MMSE

Figure 6.24 – CDF SE for the centralized and distributed operations combining vectors for PTA, ONTUA, OUA, NTA and Orthogonal-PTA, for known and unknown channel statistics.



(a) P-MMSE

(b) nopt LP-MMSE

Figure 6.25 – Average SE for the centralized and distributed operations combining vectors for PTA, ONTUA, OUA, NTA and Orthogonal-PTA, for known and unknown channel statistics versus UL transmit power p in mW.

likely to assume a value that has the noise distribution. Therefore, since those DCC methods rely on a given power threshold, more APs will be connected to each UE to satisfy that threshold, which will then make the complexity increase. Therefore those DCC methods should be avoided in the case of imperfect channel statistics allied with low SNR. Now taking a look at the centralized operation of Fig. 6.25a, we see that the gap between perfect and imperfect statistics is more clearly shown. As expected, the gap is larger for low transmitted power, since the coefficients are poorly estimated, and tends to zero as the transmit power increases. One logical solution would be to transmit with increased power in order to obtain better estimates, however, besides the obvious complications increasing the transmit power brings to the system, such as increased cost on amplifiers, there is also the fact of pilot contamination, which limits performance no matter how much p in increased. Furthermore, we can see the performance gap is small at $p = 100$ mW, which corresponds to a part of the SE curve that has not saturated, which shows that we do not need to increase p as much in order to achieve satisfactory system performance.

6.4 COMPARISON BETWEEN DCC CELL FREE AND CELLULAR SYSTEMS

After establishing the comparison between the DCC methods in the case of imperfect channel statistics knowledge, we shift our focus to comparing the overall performance of CF architecture with some of the analyzed DCC methods and traditional cellular systems such as the Massive MIMO and Small Cell. The simulation setup is the same as the one that

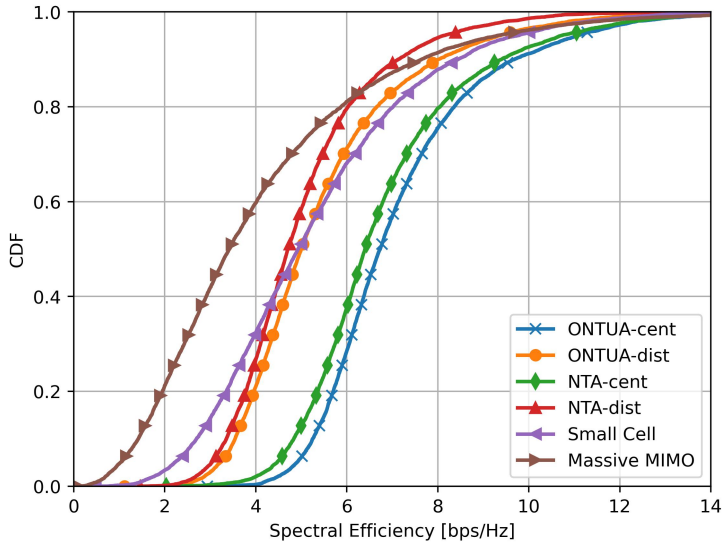


Figure 6.26 – Throughput for centralized and distributed CF operations for ONTUA and NTA, small cell and Massive MIMO architectures. P-MMSE and n-opt LP-MMSE are used in centralized and distributed operations respectively, and L-MMSE is used in both cellular setups.

was used to simulate all of this works figures, according to Table 4.1. However, instead of having the APs randomly deployed along the square grid, we shall have the configuration of Fig. 3.2, with the APs being deployed following in a square manner. For CF and small cell setups, there are $L = 100$ APs with $N = 4$ antennas, whereas the Massive MIMO has $L = 4$ APs equipped with $N = 100$ antennas, where each AP stands at the center of a 500 m $\times$ 500 m square. For the case of small cell, the connected AP will not be the one that possess the highest SINR to a particular UE, instead obeying $l \in \mathcal{M}_k$, where $\mathcal{M}_k$ i obtained from OUA, or

$$\ell = \underset{l \in \mathcal{M}_k}{\arg\max} \beta_{kl}. \tag{6.1}$$

In the case of Massive MIMO, each UE is connected to the AP that possess the highest average channel gain β_{kl} relative to it. Additionally, each one of the cells contain $\tau_p = 10$ UEs, that have been assigned each one orthogonal pilot sequences. The pilot assignment method as well as the correct UE deployment are explained in [8]

For the CF architecture, we have chosen ONTUA and NTA as the DCC methods, since they offer less complexity and similar throughput than their OUA and PTA counterparts. Also, we wish to establish a comparison for distributed and centralized operation. Thus, we shall use P-MMSE combining for centralized operation and n-opt LP-MMSE combining for distributed, all with scalability in mind. For the cellular setups, we shall use L-MMSE combining. The Fig. 6.26 shows that centralized operation, not surprisingly, has the best results, since, besides having the APs jointly cooperating to decode the data, there is also

knowledge of all of the channel estimates for combining vector computing, which in turn results in better interference cancellation. ONTUA has, as expected, higher performance than NTA, but only up to a 0.5 bps/Hz margin, which still makes it a viable solution. When analyzing distributed operation, we can see that it still has overall better performance than cellular Massive MIMO, while it has, for higher percentiles, lower performance than the small cell setup. Still, for lower percentiles, which are the main focus fro a point of view f fairness, it greatly surpasses the small cell, having a gap of approximately 1 bps/Hz at the 1-th percentile. Therefore, as evidenced in Chapter 1, the CF architecture has a larger uniformity of SE values when compared to the cellular setup, be it Massive MIMO or small cell.

Chapter Summary

- There were two main factors that were measured in this chapter: the scalability of the proposed methods and the network performance.
- When talking about scalability and performance, we have established four such factors for it: the possibility of adding UEs and APs to the grid, as well as the possibility of sending statistical information over the fronthaul links and increasing channel estimation quality with minimum additional computing resources and with satisfactory network performance
- When analyzing the combining vectors we observed that the optimal solutions, in the form of MMSE and L-MMSE, albeit optimal are not scalable. Also the optimal LSFD weights are also not scalable with respect to the fronthaul links. We have then selected the more scalable combiners P-MMSE, P-RZF and LP-MMSE.
- From a general case, we observed that the DCC method that has the lowest complexity while maintaining satisfactory performance is ONTUA, when talking about the entire network. When we focus on fairness, then Orthogonal-PTA presents itself as an alternative solution, and when focusing on increasing the performance of the UEs with the best channel condition, NTA is generally a good solution.
- In regards to the pilot assignment, we see that RPA is only a reference scenario, as it has a rather low performance. RP and GPA are almost equal performance-wise with RP having a tiny better performance in terms of fairness
- Finally, when dealing with imperfect statistics, we see that performance is reduced whatever the DCC method is. Also, we notice that increasing the transmit

power is not always beneficial, and that we should instead use more coherence blocks for the training phase of estimating the spatial correlation matrices.

7 CONCLUSIONS

This chapter summarizes the discussions that were made in earlier chapters, specially chapter 6. Section 7.1 presents the conclusions and insights on the DCC methods and pilot assignments that were discussed in this monograph, regarding performance and complexity. It also compare the different network architectures to provide an insight on the pros of using CF architecture. Section 7.2 suggests the path of future researches on topics and systems that were not contemplated on this monograph.

7.1 CONCLUSIONS AND INSIGHTS ON DCC METHODS AND PILOTS ASSIGNMENTS

CF Massive MIMO systems aim to guarantee a more uniform throughput for each UE in a network grid, when compared to cellular systems. Contrary to traditional Massive MIMO systems, it does not deploy a large array at each AP, instead relying on the geographical distribution of the antennas to obtain scattering richness and increase throughput.

The main objective of connecting a cluster of APs to an UE is to coherently and jointly decode the signal intended for this UE in order to obtain a more robust estimation of the desired transmitted data. We have seen in chapters 1 and 3, that it outperforms the small cell setup, which consists on the connection between UE and strongest AP relative to it: small cell cannot use the information of other APs to coherently and jointly decode the intended information, hence the inferior performance.

7.1.1 DCC methods insights and conclusions

Ideally, each UE would be connected to every AP in the grid. However, we have seen in Chapter 2, that such arrangement possesses a complexity that grows linearly with the number of UEs in the grid, therefore adding UEs to the grid would cause an increasing use of computation resources. To counter that, we have introduced algorithms to connect each UE to a subset of APs that keeps performance similar to the case of every AP connected. We call those DCC algorithms, since they vary dynamically as the macro channel conditions, that is, the channel statistics that vary slowly with time, change. Such methods can take a range of parameters to create an AP cluster for each UE, ranging from correlation from maximization, learning techniques among others. Various network measures can also be accounted for clustering, such as transmit power, small scale fading and spatial correlation matrices. Among those, we have chosen the large scale coefficients as the network metric, as they

do not change rapidly with time (usually they remains constant for hundreds of coherence blocks), and even after this interval, their variation are not so drastic, so they can be obtained by use of adaptive techniques.

Additionally, they possess information on the physical distances between AP and UE, as well as the correlation of the collective AP coefficients between two UEs. The clustering can also be done with some goal for the network performance, such as increasing the overall network performance or fairness SE distribution. However, one must keep in mind that such methods should work under variable network conditions, as the number of connected UEs might change with time, as well as the AP number in the grid. Also, in some cases, more robust channel estimation must be done in order to guarantee an agreeable level of service for the UEs, thus the number of pilots must be increases. The methods should be able to accommodate those changes without requiring a large amount of additional computation resources.

One sure way to guarantee that the use of computer resources stay finite is to assign a limited number of APs to each UE, or vice versa. One such method was connecting exactly a maximum of τ_p UEs to each AP, so, no matter the increasing in the values of K and L, each APis still able to process the information efficiently. From those conditions, we have studied the OUA, ONTUA and Orthogonal PTA. Both ONTUA and Orthogonal PTA are viable solution when increasing the value of τ_p, whereas OUA requires too much computer power for large τ_p.

Another way to guarantee that the AP cluster is limited is to assign a threshold below which there is no connection. We have proposed ONTUA, NTA, PTA an Orthogonal PTA. PTA and NTA act as pure threshold algorithms, where NTA is responsible to assign UEs that have reasonable channel quality to the APs, while PTA aims to assign the APs to the UE in order to satisfy a power threshold for each UE. From those, we can clearly see that they do not rely on τ_p, so they can be used on networks that require better channel estimation. However, they also cannot accommodate a large variation of K and L, since they would assign more APs to each UE, thus increasing the AP cluster size indefinitely. When analyzing interference mitigation between the UEs being served by the same AP, we consider OUA, ONTUA and Orthogonal PTA. OUA connects the strongest UEs relative to n AP and that are assigned orthogonal pilot sequences. ONTUA does the sane, however it discards the UEs that have a small channel gain (below the given threshold). On doing so, it reduces system complexity, at the expense of a small, in many times negligible performance loss, since the removed AP are very likely to influence little on the UE performance. This is specially helpful when dealing with situations that require great channel estimation quality, as the increasing number of pilots lead to a complexity increase that is much lower than the one presented by OUA. Regarding Orthogonal PTA. it also aims to connect orthogonal pilots to the UEs being served by the same AP, however, unlike the other two DCC methods, it relies

on the power threshold, which forces the UEs with greater channel conditions to connect to fewer APs, while the weaker UEs are served by more APs. The result is some performance loss for the strongest UEs, since they do not connect to as much APs as OUA and ONTUA.

The combining vector also influences on the choice of the DCC method: one strong example is the MR, where there is no form of interference mitigation between neighboring UEs. Therefore, methods as OUA, ONTUA and Orthogonal PTA are not ideal, whereas ones that connect the APs based on thresholds alone are more efficient, as NTA and PTA. The UL operation also influences that choice: distributed operation for instance has the presence of the LSFD weights, that change the weight of the APs in the grid, thus "confusing" algorithms based on power thresholds. Centralized operation does not possess such peculiarity, thus those algorithms can be used freely on this kind of operation.

When analyzing the feasibility of every method, we notice that ONTUA and OUA can have a local knowledge of the channel gains (that is, each AP can know the information of channel gain only from itself), so that DCC be also performed locally, thus reducing the fronthaul load even more in distributed operation. The other methods are all dependent on the UEs choosing the APs, therefore clustering for them should be made in a centralized manner. Knowledge of the channel gains is highly desirable in every method, and we have seen that imperfect knowledge influences the DCC methods efficiency differently. For example, methods that rely on power thresholds suffer greatly from imperfect channel statistics knowledge, increasing signal detection complexity. Methods such as OUA do ot suffer as much, since imperfect knowledge is not likely to impact the distribution of the larger channel gains the same way it would impact the smaller. NTA works in a similar fashion, and ONTUA acts the same as OUA.

One aspect that has not been covered on this monograph is the effect of the pilot assignment into DCC performance. Intuitively, we can say that, if two UEs with a strong channel to an arbitrary AP share the same pilot sequence, then methods such as OUA and ONTUA would not allow those two UEs to be connected to that AP, which would in turn prevent one of them to connect to a strong AP, thereby reducing performance. For methods that do not have interference mitigation techniques such as NTA and PTA, the performance drop would be even ore aggressive: with a bad pilot assignment, the interfering UEs could be placed nearby each other, which would result in two stronger interfering UEs being connected to the same AP, which would lower the performance of both, thus the performance gap between the non and interference mitigation methods would be even larger. Also, for distributed operation we would have another aggravating situation in the form of the LSFD weights: since their scalable counterpart rely on the subset of interfering UEs on the desired one, a bad pilot assignment would also weaken the estimation of those weights, which would in turn promote an even more aggressive performance loss on distributed operation if compared to centralized.

Finally, we have evaluated performance and complexity for variations of the threshold parameters of each method. We have perceived that the threshold value must be carefully chosen on the case of NTA, as an arbitrary choice could lead to a huge increase in system complexity. In fact, the larger the number of UEs in the grid, the lesser the variance of this threshold value, as we can see from Fig, 4.9 of chapter 4.. Therefore, for larger K, it would be even more important to choose appropriately, as the slight variation of the threshold could lead to a large number of APs being connected to each UE. Thus, it is important to construct the normalized channel gains curves and from them pick the appropriate threshold. When considering ONTUA, we do not have the same problem as NTA, however a bad choice could lead to very few UEs connected to each AP, in some cases none at all, or the asymptotic case of τ_p of them connected to each AP, which is exactly the OUA configuration. For PTA we have seen that the increasing value of the power threshold increases the performance very little at the expense of a huge complexity increase. Therefore, it is recommended to stick with the 95% power value, as it guarantees acceptable performance while maintaining complexity on a feasible level.

7.1.2 Pilot assignment algorithms insights and conclusions

When dealing with MMSE channel estimation, the CSI is obtained from the transmitted UL pilots. If those pilot signals are all transmitted over a distinct frequency range, then there should be no interference from neighboring UEs into channel estimation, and therefore thermal noise would be the only limiting factor. On practical networks it is usually not possible to allocate orthogonal pilot sequences to every UE, since bandwidth is limited, so some frequencies ranges must be used by more than one UE. This makes the channel estimates correlated, and lowers quality of estimation. This effect is shrunk the farther apart the UEs are between each other, since signal strength decays rapidly with distance. Therefore, a method to mitigate this interference would be to reuse pilots for the UEs that are distant between each other, since the farther away UE would most probably not connect to the AP the nearer would., and consequently, assign orthogonal pilot sequences to the UEs connected to sets of APs that share many intersections.

Usually the situation to be avoided is the one where a UE with average channel condition shares a pilot with an UE that has a great channel condition. In this situation, the stronger UE would greatly affect the estimation of the weaker UE channel, which would greatly affect the SE of the weaker one. Also, thee could be two UEs with high channel gains sharing the same sequence, which would result on both channel estimated being impacted, resulting in a huge performance loss of two UEs that should have had high SE values. Thus, removing neighboring pilots interference from the grid is an act that aims to improve overall performance, be it for weaker or stronger UEs. There are, however, different degrees for this improvement: stronger UEs do not need such quality in the channel estimates, merely

requiring that they do not share a pilot with another strong UE. For the case of the weaker ones, however, channel estimation quality is of utmost importance, since they cannot rely on high channel gains to obtain satisfactory performance. Thus, we see that efficient pilot assignment methods are used in order to guarantee a given degree of fairness in the network.

One method of pilot assignment that has been suggested and observed in chapters 4 and 6 is the RPA. This method consists on assigning the pilots to each UE in a random manner, with the condition that each pilot is used at least once. The advantage of such method is that it requires no form of coordination between the CPU and UEs for the assignment, as well as no additional statistics. The biggest disadvantage is that each UE is likely to be sharing a pilot sequence with a neighboring UE. For instance, if only twp pilot sequences are used, the probability of two UEs that are nearby sharing the same pilot is as high as 50%, which incurs in high performance loss. With that in mind, we have also proposed alternative methods, that rely on the channel statistics to be performed. The reason for that is once again the fact channel statistics remains invariant for thousands of coherence blocks, so this kind of assignment must be repeated after a large amount of time, opposed to repeating it once every coherence block.

One method that was proposed is GPA, which can also be found in [8], where each UE selects the AP that has the highest channel gain as its "master" AP. Then, τ_p UEs are randomly assigned mutually orthogonal pilots, and the master AP begins inviting UEs to form a connection, if they are such that the interference between the pilot sharing UEs is minimum. This method obviously requires coordination between the APs and the CPU, as each AP must have channel statistics knowledge of the entire grid in order to assign the pilots. One discussed limitation of this method was the initial random allocation, as it could lead to a situation where UEs that re far away are allocated orthogonal pilots, thereby reducing efficiency. Another limitation is the use of statistics of only the master AP, as opposed to those of every AP in the grid. With that in mind, we have analyzed the RP, originally proposed by [35]. This algorithm aims to first assign the same number of pilots, or if not the same, a similar number, to the UEs in the network grid. By doing that interference is likely to be uniformly distributed among the UEs, as all of them will have an equal number of pilot sharing interfering UE. This, allied with an efficient pilot assignment can lead to satisfying interference mitigation.

The RP is dependent on the metric for the repulsion score: the distance metric has been proposed by [35], while the euclidean distance has been implemented initially by [15]. The correlation metrics have been proposed by this monograph. When we have analyzed the distance metric, we have seen that it succeeds in placing the same pilot in UEs with vastly different collective channel gains. This happens because it has knowledge of the real distances between the APs and UEs, so it can efficiently design a metric that "repels" the pilot sharing UEs to far away locations into the grid. Since this is an ideal case, as distance must

be estimated in conventional networks, another studied method was the euclidean distance between the collective channel gains. The performance of this metric is inferior to that of distances, since we account for the channel gains. It is, however, superior than RPA, since UEs that have a large collective euclidean distance are assigned the same pilot. This metric takes in account the channel gain difference, which can lead to wrong assumptions. For instance, suppose the collective euclidean distance is small for tow UEs. That would only mean that those UEs have similar channel gains, however they could still be far away from each other, which would make assigning distinct pilot sequences fruitless, since they would not interfere strongly on the channel estimates at the first place. To establish a stronger indication of whether the UEs are nearby, we have proposed the collective channel gains correlation metric: this is a much more reliable method, since UEs with lower correlations are very likely to have vastly different channel gains, and therefore are more likely to be spaced far away from each other, while UEs with a high correlation do have very d=similar channel gains, which in fact increases the likelihood of them being nearby each other. In fact, this metric is so efficient that performance is almost similar as the case where the distances are known. Finally, we tried proposing the spatial correlation metric, which measures the correlation between the collective spatial correlation matrices of UEs. However, opposed to the correlation between the channel gains, this metric is not very effective when dealing with pilot assignment. For instance, two UEs can be highly correlated spatial-wise and still be far away from each other. The algorithm would interpret the high degree of correlation as a signal to assign orthogonal pilots to both UEs, when in reality the same pilot sequence could be reused without compromising the performance of both UEs. Thus, spatial correlation is not such an interesting metric when we wish to discuss pilot assignment.

Another point of interest is that GPA has a slight lower performance than RP with the correlation between collective channel gains metric when we consider fairness. This happens due to the randomness of GPA in allocating orthogonal pilots in the beginning of GPA, whereas RP do not have any randomness associated to it, besides the initial configuration of the assignment, which is then completely changed due to the repulsion score behavior. Thus RP has the ability to mitigate interference better which in turn reflects on better performance for the UEs with bad channel conditions. For the ones with good to excellent channel conditions, RP and GPA eliminate the possibility of two strong UEs sharing the same pilot sequence, and therefore, performance is very similar in both cases. When dealing with imperfect channel statistics, we notice that the pilot assignment methods do not suffer variations as high as the ones experienced by the DCC methods. This is explained by the fact pilot assignment does not need such a refined estimate of the channel gains to assign the pilots, as correlation values that are high will still be high, as well as the master AP will still generally have a much larger channel gain. The exception is in the case of very poor statistics, for low UL transmit power, where the estimate is dominated by noise, and thus the metrics become corrupted.

Finally, we have noticed that systems with high mobility, that is, systems where the coherence block length is small suffer from increasing channel estimation quality, since the amount of useful data is limited in favor of the training process. This overhead decreases performance, specially for the stronger UEs, who do not benefit as much for increasing quality channel estimates. Thus, it is crucial for high mobility networks to have efficient pilot assignment methods, since the value of τ_p must be as small as not to cause the discussed overhead, while channel estimation quality should not suffer as much in order to provide satisfactory performance.

7.1.3 Conclusions on different architectures

Finally, having discussed DCC methods and pilot assignments, we now provide reasons for the implementation of CF architecture. The main objective of any MIMO technology is to exploit the scattering richness of the media in order to coherently combine the copies of the intended signal, that passes through distinct paths, to obtain a robust approximation of the transmitted data. Massive MIMO aims to use this information allied to the proposition that, when the number of antennas at a given AP grows large, the mean channel value becomes deterministic, in what is known as channel hardening. Such proposition has thoroughly been used by authors as [9], [30] and [31], that both consider a scenario of spatially uncorrelated APs, that is, where the signals propagate without any preferred direction. Unfortunately, such scenarios are very rare in real network environments, where there is spatial correlation between the antennas of a given AP, and where the signals propagate more in certain angular directions than others. In such cases, the desired degree of channel hardening to obtain reliable performance can require the AP to be equipped with a large array, sometimes containing hundreds of antennas.

We notice that a configuration of hundreds of antennas at each AP could lead to challenging implementation, so CF MIMO arises as an alternative: instead of deploying a very large array in only one AP, we distribute a range of sub-arrays at each AP, so that each one is equipped with a much smaller number of antennas. We immediately notice that the geographical variability of the system increases drastically, and also, that spatial correlation acts as an ally rather than a limitation, since the UEs with different spatial characteristics can be allocated the same pilot without considerable interference in their channel estimates. Also, the possibility of the UE being connected to many APs makes it more likely that it will be connected to an AP that has a high channel gain, thereby avoiding the cell edge problem that arises in Massive MIMO and on more traditional networks architectures. Therefore, CF provides a more uniform distribution of the network throughput among the UEs, while it does not considerably rises the network throughput, as it is the objective of Massive MIMO. Also, it requires a much higher degree of coordination among the APs and the CPUin order to jointly decode the signals, as well as correctly applying the DCC methods and pilot

assignments.

When comparing to traditional network systems, in the form of small cells, where each AP connects only to the strongest AP, we notice the per-user throughput still has a very large variance when compared to CF. This again happens because the APs cannot cooperate in order to provide a more robust estimate of the transmitted data. Although pilot assignment specially becomes much more simple on this case, since we are not leading with the collective channels, but rather with the single AP-UE link. Nevertheless, the throughput is much more uniform in CF rather than single cell, additionally having no cell-edge whatsoever, which helps cement CF MIMO as a strong alternative for modern networks architecture.

7.2 FUTURE RESEARCH

Throughout this work, we have covered a lot of crucial aspects in the implementation of CF MIMO networks. We have considered some models to compose the simulation scenery in order to compare performance and scalability. One assumption was that of MMSE channel estimation. This method of channel estimation is a linear technique, however, there are many authors who have proposed non-linear methods of channel estimation, such as blind methods proposed by [28] or semi-blind methods such as the one in [29]. Since those techniques do not rely on pilot transmissions, we cannot assign the APs using most of the DCC methods discussed in Chapter 4, that rely on assigning orthogonal UEs for each AP. Thus, one could design new DCC methods that attempt to mitigate interference between channels that are estimated by non-linear methods. On that regard, many possibilities come into place, since one could make use of reinforcement learning techniques, in order to obtain a precise knowledge of the network variations, and by that design a DCC method that accompanies those changes.

Another point of interest is that we have used the correlated log-normal for generating the shadowing variations of the large scale fading coefficients, as well as considering an urban environment where the attenuation coefficient α is 3.65. It would be interesting to apply the same algorithms over environments where the distribution of APs is sparse, and where the attenuation is not as aggressive. For example, if there is less attenuation then perhaps the behavior of threshold algorithms change accordingly. Either way, the distribution of the channel gains would surely change, which makes the more important the correct design on the curves of Fig. 4.9. Also, we have assumed the correlated Rayleigh fading model for the communication channel. This model holds for scenarios where there is a large amount of scattering, with the multipath signal components having on average the same power. There are, however, situations where there is the presence of line of sight (LOS) signal component, which has a much greater power than the other multipath components. As DCC and

pilot assignment formation are independent on the small scale variations, they should not be affected by this difference, however the spatial correlation model will surely be altered, and with it, system performance. Authors such as [44] have already proposed models for spatially correlated Rician fading for Massive MIMO and [45] have done the same for CF. It would be interesting to apply the same to the CF systems with the DCC methods that were proposed, as well as the pilot assignments. Additionally, there are alternative channel models that could also be explored, such as the Nakagami model, or eventually the κ-μ model, which is a very general model for conventional networks.

On conventional networks, the rate at which the large scale fading suffer variations is very slow compared to small fading, usually taking hundreds or even thousands of coherence blocks to observe a change at all. Still, even when it changes, it has a tendency not to do so as abruptly. For instance, unless the UE is experiencing high mobility, it is likely that over the next channel gains variation it still stays at the same geographical location. That makes the channel gain variations for that particular UE very small. On our proposed simulation scheme, we have assumed total large scale fading variations for each setup, and have performed DCC entirely at each one of them. Still, since the large scale variations do not change as harshly, there could be some form of adaptive clustering, where DCC would only need to be redone in situations where the channel gain suffers considerable variations. This in turn would surely save a huge amount of computer processing, which would make the described DCC methods all the more implementable. Authors such as [46] have already proposed such methods for some DCC formations, so it would be a possibility to do the same with the proposed DCC methods of this monograph and check whether they can work on adaptive environments.

On this monograph, P-MMSE for the centralized operation and n-opt LP-MMSE have extensively been used to compare system performance, because they are sub optimal with respect to mitigating interference while at the same time provide scalability to the system. Although they do improve performance, they are still computationally costly, and there are situations where they are not entirely needed. For instance, lets suppose a situation where the spatial correlation matrices of two UEs are such that they are orthogonal, and that the other UEs that share pilots with those UE are in similar situation. Than, in this case, where there is no form of interference between neighboring UE, the combining method of P-MMSE would yield values of performance very similar to those of MR, with the latter providing a much more scalable solution. Let us now suppose that those UEs do not possess correlation among the channel estimates, which would also make the P-RZF combining method effective, as well as less costly than P-MMSE. Then a viable solution to improving scalability would be the design of some form of artificial intelligence that selects the combining method based on channel conditions (that is, based on correlation, spatial correlation and measure of interference) in order to select one combiner that improves spectral efficiency while at the same time

reduces computation work. The same could also be made for the number of antennas used for signal detection at each AP: UEs with high channel gains would perhaps not need a high degree of diversity in order to decode the information, so the AP could employ less antennas in this case, and even use only one of them, which would surely save a lot of computational expenses.

Still another point is that no form of spatial multiplexing between the UEs was considered, since they were all single antennas. In the case where each UE is equipped with more than one antenna, the spatial correlation matrices as well as the channel coefficients matrix would suffer changes, as would the performance values. Although the model still holds for this case, it would be nonetheless helpful to investigate the effects of the proposed DCC and pilot assignment methods for those systems. Also, no analysis was conducted whatsoever on the number of antennas at each AP. From chapter 2 we have seen that increasing the number of antennas improves the SNR only until a certain point, specially in scenarios with frequency reuse. However, a deeper analysis should be conducted to find the optimal number of AP antennas that provides increased performance.

Finally, no power control algorithms were used in this monograph. Those methods aim to allocate different transmit powers for each UE in the grid, in order to guarantee various objectives, such as more fairness or higher network throughput. Also, power control techniques could change behavior of some of the introduced DCC methods. For instance PTA relies heavily on the total power of each UE to the grid APs. If power control techniques are used, then probably the weakest UEs would be compensated with higher transmit power (if fairness is the goal), which in turn would make them more likely to connect to less APs in order to satisfy the power threshold conditions. Since exploring power control was not the main objective of this monograph, we have opted not to cover it and left it for future research. Those techniques were used to estimate the channel via UL pilots in [47], although they could additionally be used to simulate the systems described in this monograph. Regarding the transmission protocol, we have only covered the UL transmission. Other researches could also explore the effects of the described methods in the DL. For instance, the DL operation makes use of the so called precoders, that aim to decode the information transmitted by the APs, by using the CSI acquired from UL pilot training, if we consider time division duplex (TDD) and channel reciprocity. Power control techniques also play a crucial role here. Future works could simulate the same scenarios that were analyzed in this monograph, but now with focus on DL operation.

Chapter Summary

- Among the introduced DCC methods, the one with the best trade off between performance and complexity seems to be ONTUA. In centralized operation, if we wish for a method with great fairness and low complexity, then we should

use Orthogonal PTA. If we wish to improve the overall network with less complexity, then we should use NTA. For distributed operation complexity is not such an issue, therefore we can use ONTUA.

- Regarding pilot assignment, we can either use GPA or RP. For a fairness objective RP has a slight better performance than GPA, when the collective channel gains correlations metric is used. In a theoretical basis, the best metric for RP would be the distance. Since it is usually not known, we use the correlation instead. Euclidean distance of the collective channel gains and correlation of the collective spatial correlation matrices are not as efficient.

- Conventional small cell network systems provide the advantage of connecting the UE with the AP providing the best channel gain, which greater reduces complexity. On the other hand, it cannot use the information of other APs to jointly decode information, and also it suffers from cell edge effect. Massive MIMO has the advantage of exploiting spatial diversity to reliably decode the signals. However, it has the disadvantage of requiring a sometimes unreal number of antennas to obtain the so important degree of spatial diversity. CF systems have the main goal of providing uniform throughput to the network, by jointly decoding the signals. It has an overall high complexity and requires perfect cooperation and synchronization between the APs and the CPU

- Non-linear channel estimation has not been covered in this monograph, as well as DL operation, spatial multiplexing of the UEs information and distinct channel models as Rician, Nakagami and $\kappa - \mu$. Future researches could investigate the effect of such scenarios on the DCC and pilot assignment methods that were analyzed in this monograph. Additionally, some kind of artificial intelligence could be developed to better select the combining/precoding vector based on channel statistics and interference patterns. Also, this same intelligence could be used to select the number of AP antennas to be used for detection of each UE signal. This would efficiently reduce system complexity along with offloading the APs.

BIBLIOGRAPHY

1 MOHER, M. *Communication Systems*. [S.l.]: Wiley, 2009.

2 RAPPAPORT, T. S. *Wireless Communications: Principles and Practice*. [S.l.]: Cambridge University Press, 2024.

3 CHEVALLIER, C. et al. WCDMA (UMTS) deployment handbook. *Planning and Optimization/ Christophe Chevallier–John Wiley & Sons LTD, England*, Wiley Online Library, v. 390, n. 6, 2006.

4 AGEE, B. G. Exploitation of internode MIMO channel diversity in spatially distributed multipoint communication networks. In: *Wireless Personal Communications: Bluetooth and Other Technologies*. [S.l.]: Springer, 2000. p. 109–120.

5 DAHLMAN, E.; PARKVALL, S.; SKOLD, J. *4G: LTE/LTE-advanced for mobile broadband*. [S.l.]: Academic Press, 2013.

6 MARZETTA, T. L. How much training is required for multiuser MIMO? In: IEEE. *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*. [S.l.], 2006. p. 359–363.

7 BJÖRNSON, E.; HOYDIS, J.; SANGUINETTI, L. Massive MIMO has unlimited capacity. *IEEE Transactions on Wireless Communications*, IEEE, v. 17, n. 1, p. 574–590, 2017.

8 DEMIR, Ö. T. et al. Foundations of user-centric cell-free massive MIMO. *Foundations and Trends® in Signal Processing*, Now Publishers, Inc., v. 14, n. 3-4, p. 162–472, 2021.

9 RUSEK, F. et al. Scaling up MIMO: Opportunities and challenges with very large arrays. *IEEE Signal Processing Magazine*, IEEE, v. 30, n. 1, p. 40–60, 2012.

10 HOYDIS, J.; BRINK, S. T.; DEBBAH, M. Massive MIMO in the UL/DL of cellular networks: How many antennas do we need? *IEEE Journal on Selected Areas in Communications*, IEEE, v. 31, n. 2, p. 160–171, 2013.

11 NEUMANN, D.; JOHAM, M.; UTSCHICK, W. On MSE based receiver design for massive MIMO. In: VDE. *SCC 2017; 11th International ITG Conference on Systems, Communications and Coding*. [S.l.], 2017. p. 1–6.

12 LI, B. et al. Cooperative time synchronization and robust clock parameters estimation for time-sensitive cell-free massive MIMO systems. *IEEE Transactions on Wireless Communications*, p. 1–1, 2024.

13 BJÖRNSON, E.; SANGUINETTI, L. Scalable cell-free massive MIMO systems. *IEEE Transactions on Communications*, IEEE, v. 68, n. 7, p. 4247–4261, 2020.

14 3GPP. Further advancements for E-UTRA physical layer aspects. *IEEE Journal on Selected Areas in Communications*, 3GPP, v. 36, n. 814, 2017.

15 BIGONHA, A.; ARAÚJO, D.; SILVA, J. Dynamic clustering and pilot assignment for optimizing cell-free massive MIMO uplink. *XLI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT2023)*, SBRT, v. 41, n. 1, 2023.

16 NGO, H. Q. et al. On the total energy efficiency of cell-free massive MIMO. *IEEE Transactions on Green Communications and Networking*, IEEE, v. 2, n. 1, p. 25–39, 2017.

17 LATHI, B. P. *Modern Digital and Analog Communication Systems*. [S.l.]: Oxford University Press, Inc., 1990.

18 YANG, H.; MARZETTA, T. L. Capacity performance of multicell large-scale antenna systems. In: *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. [S.l.: s.n.], 2013. p. 668–675.

19 MISRA, S. Wireless communications"(molisch, a.; 2011)[book review]. *IEEE Wireless Communications*, IEEE, v. 19, n. 1, p. 5–5, 2012.

20 HAMPTON, J. R. *Introduction to MIMO Communications*. [S.l.]: Cambridge University Press, 2013.

21 PIZZO, A.; MARZETTA, T. L.; SANGUINETTI, L. Spatially-stationary model for holographic MIMO small-scale fading. *IEEE Journal on Selected Areas in Communications*, IEEE, v. 38, n. 9, p. 1964–1979, 2020.

22 BJÖRNSON, E. et al. Massive MIMO networks: Spectral, energy, and hardware efficiency. *Foundations and Trends® in Signal Processing*, Now Publishers, Inc., v. 11, n. 3-4, p. 154–655, 2017.

23 BJÖRNSON, E.; LARSSON, E. G.; MARZETTA, T. L. Massive MIMO: Ten myths and one critical question. *IEEE Communications Magazine*, v. 54, n. 2, p. 114–123, 2016.

24 SHANNON, C. E. A mathematical theory of communication. *The Bell System Technical Journal*, Nokia Bell Labs, v. 27, n. 3, p. 379–423, 1948.

25 BIGLIERI, E.; PROAKIS, J.; SHAMAI, S. Fading channels: Information-theoretic and communications aspects. *IEEE Transactions on Information Theory*, IEEE, v. 44, n. 6, p. 2619–2692, 1998.

26 MEDARD, M. The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. *IEEE Transactions on Information Theory*, IEEE, v. 46, n. 3, p. 933–946, 2000.

27 MORALES, C. M.; ESLAVA, G. S. Linear and non-linear channel prediction performance for a MIMO-OFDM system. In: IEEE. *2014 IEEE 5th Latin American Symposium on Circuits and Systems*. [S.l.], 2014. p. 1–4.

28 MÜLLER, R. R.; COTTATELLUCCI, L.; VEHKAPERÄ, M. Blind pilot decontamination. *IEEE Journal of Selected Topics in Signal Processing*, IEEE, v. 8, n. 5, p. 773–786, 2014.

29 WAN, F.; ZHU, W.-P.; SWAMY, M. Semiblind sparse channel estimation for MIMO-OFDM systems. *IEEE Transactions on Vehicular Technology*, IEEE, v. 60, n. 6, p. 2569–2582, 2011.

30 JOSE, J. et al. Pilot contamination and precoding in multi-cell TDD systems. *IEEE Transactions on Wireless Communications*, IEEE, v. 10, n. 8, p. 2640–2651, 2011.

31 BJÖRNSON, E.; SANGUINETTI, L.; DEBBAH, M. Massive MIMO with imperfect channel covariance information. In: IEEE. *2016 50th Asilomar Conference on Signals, Systems and Computers*. [S.l.], 2016. p. 974–978.

32 HAGHIGHATSHOAR, S.; CAIRE, G. Massive MIMO pilot decontamination and channel interpolation via wideband sparse channel estimation. *IEEE Transactions on Wireless Communications*, IEEE, v. 16, n. 12, p. 8316–8332, 2017.

33 NEUMANN, D.; JOHAM, M.; UTSCHICK, W. Covariance matrix estimation in massive MIMO. *IEEE Signal Processing Letters*, IEEE, v. 25, n. 6, p. 863–867, 2018.

34 UPADHYA, K.; VOROBYOV, S. A. Covariance matrix estimation for massive MIMO. *IEEE Signal Processing Letters*, IEEE, v. 25, n. 4, p. 546–550, 2018.

35 MOHEBI, S.; ZANELLA, A.; ZORZI, M. Repulsive clustering based pilot assignment for cell-free massive MIMO systems. In: *2022 30th European Signal Processing Conference (EUSIPCO)*. [S.l.: s.n.], 2022. p. 717–721.

36 NAYEBI, E. et al. Performance of cell-free massive MIMO systems with MMSE and LSFD receivers. In: IEEE. *2016 50th Asilomar Conference on Signals, Systems and Computers*. [S.l.], 2016. p. 203–207.

37 RIERA-PALOU, F. et al. Clustered cell-free massive MIMO. In: IEEE. *2018 IEEE Globecom Workshops (GC Wkshps)*. [S.l.], 2018. p. 1–6.

38 YANG, H.; LARSSON, E. G. Can massive MIMO support uplink intensive applications? In: IEEE. *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. [S.l.], 2019. p. 1–6.

39 BASHAR, M. et al. On the performance of backhaul constrained cell-free massive MIMO with linear receivers. In: IEEE. *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. [S.l.], 2018. p. 624–628.

40 CHEN, Z.; BJÖRNSON, E. Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry. *IEEE Transactions on Communications*, IEEE, v. 66, n. 11, p. 5205–5219, 2018.

41 BJÖRNSON, E.; SANGUINETTI, L. Making cell-free massive MIMO competitive with MMSE processing and centralized implementation. *IEEE Transactions on Wireless Communications*, IEEE, v. 19, n. 1, p. 77–90, 2019.

42 NGO, H. Q. et al. Cell-free massive MIMO versus small cells. *IEEE Transactions on Wireless Communications*, IEEE, v. 16, n. 3, p. 1834–1850, 2017.

43 DEMIR, Ö. T.; BJÖRNSON, E.; SANGUINETTI, L. Cell-free massive MIMO with large-scale fading decoding and dynamic cooperation clustering. In: VDE. *WSA 2021; 25th International ITG Workshop on Smart Antennas*. [S.l.], 2021. p. 1–6.

44 ÖZDOGAN, Ö.; BJÖRNSON, E.; LARSSON, E. G. Massive MIMO with spatially correlated rician fading channels. *IEEE Transactions on Communications*, IEEE, v. 67, n. 5, p. 3234–3250, 2019.

45 ÖZDOGAN, Ö.; BJÖRNSON, E.; ZHANG, J. Performance of cell-free massive MIMO with rician fading and phase shifts. *IEEE Transactions on Wireless Communications*, IEEE, v. 18, n. 11, p. 5299–5315, 2019.

46 RENNA, R. B. D.; LAMARE, R. C. de. Adaptive LLR-based APs selection for grant-free random access in cell-free massive MIMO. In: IEEE. *2022 IEEE Globecom Workshops (GC Wkshps)*. [S.l.], 2022. p. 196–201.

47 MAI, T. C. et al. Pilot power control for cell-free massive MIMO. *IEEE Transactions on Vehicular Technology*, IEEE, v. 67, n. 11, p. 11264–11268, 2018.

APPENDIX

A APPENDIX

A.1 Proof from Lemma 3.5:

The result is proven by simply showing the Hessian matrix $D^2\text{NMSE}(\lambda)$ is negative definite for $\gamma_n \geq 0$. The Hessian matrix $D^2\text{NMSE}(\lambda)$ is given by

$$D^2\text{NMSE}(\lambda) = \frac{1}{N\beta}\text{diag}\left(\frac{-2\eta\tau_p\sigma^2}{(\eta\tau_p\lambda_1+\sigma^2)^3}, ..., \frac{-2\eta\tau_p\sigma^2}{(\eta\tau_p\lambda_N+\sigma^2)^3}\right). \tag{A.1}$$

and, since its elements are all negative, this matrix is negative definite, thus completing the proof.

A.2 Proof from Lemma 3.6:

First, we notice that the elements of λ and λ' are similar except the r-th and the $(r-1)$-th ones. Thus, the difference between NMSE(λ) and NMSE(λ') is the summation of the terms of Eq. 3.68 for $n \in \{r-1, r\}$. Thus, the difference is

$$\begin{aligned}
&\frac{\eta\tau_p}{N\beta}\left(\frac{(\lambda_{r-1}+\lambda_r)^2}{\eta_t a u_p(\lambda_{r-1}+\lambda_r)+\sigma^2} - \frac{\lambda_{r-1}^2}{\eta\tau_p\lambda_{r-1}+\sigma^2} - \frac{\lambda_r^2}{\eta\tau_p\lambda_r+\sigma^2}\right) \\
&\overset{(a)}{=} \frac{1}{N\beta}\left(\frac{(x+y)^2}{x+y+c} - \frac{x^2}{x+c} - \frac{y^2}{y+c}\right) \\
&= \frac{(x+y)^2(xy+c(x+y+c)) - (x^2(y+c)+y^2(x+c))(x+y+c)}{N\beta(x+y+c)(x+c)(y+c)} \\
&\overset{(b)}{=} \frac{(x+y)^2xy + 2cxy(x+y+c) - xy(x+y)(x+y+c)}{N\beta(x+y+c)(x+c)(y+c)} \\
&\overset{(c)}{=} \frac{2cxy(x+y+c) - cxy(x+y)}{N\beta(x+y+c)(x+c)(y+c)} \overset{(d)}{>} 0
\end{aligned} \tag{A.2}$$

where we have made $x = \lambda_{r-1}$, $y = \lambda_r$, and $c = \sigma^2/(\eta\tau_p)$ in (a). In (b) and (c), we have canceled the terms $(x+2+y^2)c(x+y+c)$ and $(x+Y)^2xy$ in the numerator, and (d) was obtained from the fact that x, y, and $c > 0$.

A.3 Proof from Theorem 5.1:

The signal in Eq. 5.1 can be interpreted as a discrete memoryless interference channel described in Lemma 3.3 with estimated channel $h = \mathbf{v}_k^H \mathbf{D}_k \hat{\mathbf{h}}_k$, the input $x = s_k$ and the output $y = \mathbf{v}_k^H \mathbf{D}_k \mathbf{y}^{\text{ul}}$, and realizations $u = \left\{ \mathbf{D}_k \hat{\mathbf{h}}_i : i = 1, \ldots, K \right\}$ that affects the conditional variance of the interference. For this case, all of the interference and noise is included in v, with $n = 0$, since the noise does not necessarily has to be Gaussian. The inputs power is $p = \mathbb{E}\left\{ |s_k|^2 \right\} = p_k$. The term v is written as

$$v = \sum_{\substack{i=1 \\ i \neq k}}^{K} \mathbf{v}_k^H \mathbf{D}_k \hat{\mathbf{h}}_i s_i + \sum_{i=1}^{K} \mathbf{v}_k^H \mathbf{D}_k \tilde{\mathbf{h}}_i s_i + \mathbf{v}_k^H \mathbf{D}_k \mathbf{n} \tag{A.3}$$

Now, we need to show that the requirements of Lemma 3.3 are satisfied. We can see that the realizations of h and v are known at the CPU, and also v has conditional zero mean given $(h.u)$, that is $\mathbb{E}\left\{ v | h, u \right\} = 0$, since the symbols s_i and the noise vector $\mathbf{n}$ are independent of the channel estimates and the zero mean estimation errors. Additionally, the conditional variance given (h, u) is given by

$$\begin{aligned}
p_v(h,u) &= \mathbb{E}\left\{ |v|^2 | h, u \right\} = \mathbb{E}\left\{ |v|^2 | \left\{ \mathbf{D}_k \hat{\mathbf{h}}_i \right\} \right\} \\
&= \sum_{\substack{i=1 \\ i \neq 1}}^{K} p_i \left| \mathbf{v}_k^H \mathbf{D}_k \hat{\mathbf{h}}_i \right|^2 + \sum_{i=1}^{K} p_i \mathbf{v}_k^H \mathbf{D}_k \mathbb{E}\left\{ \tilde{\mathbf{h}}_i \tilde{\mathbf{h}}_i^H \right\} \mathbf{D}_k \mathbf{v}_k \\
&+ \mathbf{v}_k^H \mathbf{D}_k \mathbb{E}\left\{ \mathbf{n}\mathbf{n}^H \right\} \mathbf{D}_k \mathbf{v}_k \\
&= \sum_{\substack{i=1 \\ i \neq 1}}^{K} p_i \left| \mathbf{v}_k^H \mathbf{D}_k \hat{\mathbf{h}}_i \right|^2 + \sum_{i=1}^{K} p_i \mathbf{v}_k^H \mathbf{D}_k \mathbf{C}_i \mathbf{D}_k \mathbf{v}_k + \sigma^2 \mathbf{v}_k^H \mathbf{D}_k \mathbf{D}_k \mathbf{v}_k \\
&= \sum_{\substack{i=1 \\ i \neq k}}^{K} p_i \left| \mathbf{v}_k^H \mathbf{D}_k \hat{\mathbf{h}}_i \right|^2 + \mathbf{v}_k^H \mathbf{Z}_k \mathbf{v}_k + \sigma^2 \left\| \mathbf{D}_k \mathbf{v}_k \right\|^2
\end{aligned} \tag{A.4}$$

Where we have assumed that the terms of v are uncorrelated , and the combining vector is independent of the channel estimation errors. Also, the input signal $x = s_k$ is conditionally uncorrelated with v given (h, u), due to the independence of the symbols and the zero mean estimation errors.

Finally, we notice that only a fraction of the transmitted symbols are actually used to convey information. Therefore the expression should be multiplied by the pre-log factor τ_u / τ_c.

A.4 Proof from Theorem 5.2:

The CPU has no information about the channel estimates (as they are delegated to the AP). So it needs to treat the average channel gain $\mathbf{a}_k^H \mathbb{E}\{\mathbf{g}_{kk}\}$ as the true channel. The signal can then be written as

$$\hat{s}_k = \mathbf{a}_k^H \mathbb{E}\{\mathbf{g}_{kk}\} s_k + v_k \tag{A.5}$$

where v_K is the interference plus noise term

$$v_k = \left(\mathbf{a_k}^H \mathbf{g}_{kk} - \mathbf{a}_k^H \mathbb{E}\{\mathbf{g}_k k\}\right) s_k + \sum_{\substack{i=1 \\ i \neq k}}^{K} \mathbf{a_k}^H \mathbf{g}_{ki} s_i + n'_k \tag{A.6}$$

The term v_k has zero mean and is uncorrelated with s_k since

$$\underbrace{\mathbb{E}\left\{\mathbf{a_k}^H \mathbf{g}_{kk} - \mathbf{a}_k^H \mathbb{E}\{\mathbf{g}_k k\}\right\}}_{=0} \mathbb{E}\left\{|s_k|^2\right\} = 0 \tag{A.7}$$

Thus, we can apply Lemma 3.3 with $h = \mathbf{a}_k^H \mathbb{E}\{\mathbf{g}_{kk}\}$, $x = s_k$, $p = p_k$, $v = v_k$, and $\sigma^2 = 0$. Also, the signals of different UEs are independent and the receiver noise at different APs are also independent. Therefore,

$$\mathbb{E}\left\{|v_k|^2\right\} = \sum_{i=1}^{K} p_i \mathbb{E}\left\{|\mathbf{a}_k^H \mathbf{g}_{ki}|^2\right\} - p_k \mathbb{E}\left\{|\mathbf{a}_k^H \mathbf{g}_{kk}|^2\right\} + \mathbf{a}_k^H \mathbf{F}_k \mathbf{a}_k. \tag{A.8}$$

Then, from Lemma 3.3 we can derive the SE expression. Finally, we notice that only a fraction of the transmitted symbols are actually used to convey information. Therefore the expression should be multiplied by the pre-log factor τ_u/τ_c.

A.5 Proof from Corollary 5.3:

We begin with the term

$$\begin{aligned}
[\mathbb{E}\{\mathbf{g}_{ki}\}]_l = \mathbb{E}\left\{\mathbf{v}_{kl}^H \mathbf{D}_{kl} \mathbf{h}_{il}\right\} &= \operatorname{tr}\left(\mathbf{D}_{kl} \mathbb{E}\left\{\hat{\mathbf{h}}_{il} \hat{\mathbf{h}}_{kl}^H\right\}\right) \\
&= \begin{cases} \sqrt{\eta_k \eta_i} \tau_p \operatorname{tr}\left(\mathbf{D}_{kl} \mathbf{R}_{il} \Psi_{t_{kl}}^{-1} \mathbf{R}_{kl}\right) & i \in \mathcal{P}_k \\ 0 & i \not\in \mathcal{P}_k \end{cases}
\end{aligned} \tag{A.9}$$

where the second equality comes from Lemma 3.4, and the independence of $\hat{\mathbf{h}}_k$ and $\hat{\mathbf{h}}_i$.

In a similar manner:

$$[\mathbf{F}_k]_{ll} = \sigma^2 \mathbb{E}\left\{\|\mathbf{D}_{kl}\mathbf{v}_{kl}\|^2\right\} = \sigma^2 \operatorname{tr}\left(\mathbf{D}_{kl}\mathbb{E}\left\{\hat{\mathbf{h}}_{kl}\hat{\mathbf{h}}_{kl}^H\right\}\right) = \sigma^2[\mathbb{E}\left\{\mathbf{g}_{kk}\right\}]_l \tag{A.10}$$

We need only to compute the elements of $\mathbb{E}\left\{\mathbf{g}_{ki}\mathbf{g}_{ki}^H\right\}$. We see that $\mathbb{E}\left\{[\mathbf{g}_{ki}]_l[\mathbf{g}_{ki}^*]_r\right\} = \mathbb{E}\left\{[\mathbf{g}_{ki}]_l\right\}\mathbb{E}\left\{[\mathbf{g}_{ki}^*]_r\right\}$ for $r \neq l$ because the channels of different APs are independent. Thus, we need only compute

$$[\mathbb{E}\left\{\mathbf{g}_{ki}\mathbf{g}_{ki}^H\right\}]_{ll} = \mathbb{E}\left\{\hat{\mathbf{h}}_{kl}\mathbf{D}_{kl}\mathbf{h}_{il}\mathbf{h}_{il}^H\mathbf{D}_{kl}\hat{\mathbf{h}}_{kl}^H\right\} = \operatorname{tr}\left(\mathbf{D}_{kl}\mathbb{E}\left\{\mathbf{h}_{il}\mathbf{h}_{il}^H\mathbf{D}_{kl}\hat{\mathbf{h}}_{kl}\hat{\mathbf{h}}_{kl}^H\right\}\right) \tag{A.11}$$

where we have again used Lemma 3.4. If $i \notin \mathcal{P}_k$ we can say that $\mathbf{h}_{il}$ and $\hat{\mathbf{h}}_{kl}$ are independent, so that

$$\begin{aligned}\operatorname{tr}\left(\mathbf{D}_{kl}\mathbb{E}\left\{\mathbf{h}_{il}\mathbf{h}_{il}^H\mathbf{D}_{kl}\hat{\mathbf{h}}_{kl}\hat{\mathbf{h}}_{kl}^H\right\}\right) &= \operatorname{tr}\left(\mathbf{D}_{kl}\mathbb{E}\left\{\mathbf{h}_{il}\mathbf{h}_{il}^H\right\}\mathbf{D}_{kl}\mathbb{E}\left\{\hat{\mathbf{h}}_{kl}\hat{\mathbf{h}}_{kl}^H\right\}\right) \\ &= \eta_k\tau_p \operatorname{tr}\left(\mathbf{D}_{kl}\mathbf{R}_{il}\mathbf{R}_{kl}\Psi_{t_kl}^{-1}\mathbf{R}_{kl}\right)\end{aligned} \tag{A.12}$$

where one term $\mathbf{D}_{kl}$ was ommited for brevity.
If $i \in \mathcal{P}_k$, then

$$\begin{aligned}&\operatorname{tr}\left(\mathbf{D}_{kl}\mathbb{E}\left\{\mathbf{h}_{il}\mathbf{h}_{il}^H\mathbf{D}_{kl}\hat{\mathbf{h}}_{kl}\hat{\mathbf{h}}_{kl}^H\right\}\right) \\ &= \operatorname{tr}\left(\mathbf{D}_{kl}\mathbb{E}\left\{\hat{\mathbf{h}}_{il}\hat{\mathbf{h}}_{il}^H\mathbf{D}_{kl}\hat{\mathbf{h}}_{kl}\hat{\mathbf{h}}_{kl}^H\right\}\right) + \operatorname{tr}\left(\mathbf{D}_{kl}\mathbb{E}\left\{\tilde{\mathbf{h}}_{il}\tilde{\mathbf{h}}_{il}^H\mathbf{D}_{kl}\hat{\mathbf{h}}_{kl}\hat{\mathbf{h}}_{kl}^H\right\}\right)\end{aligned} \tag{A.13}$$

where the equality follows from splitting the channel $\mathbf{h}_{il}$ into $\hat{\mathbf{h}}_{il}$, the estimated channel and $\tilde{\mathbf{h}}_{il}$, the estimation error. The second term becomes $\eta_k\tau_p \operatorname{tr}\left(\mathbf{D}_{kl}\mathbf{C}_{il}\mathbf{R}_{kl}\Psi_{t_kl}^{-1}\mathbf{R}_{kl}\right)$ due to the independence between the estimate and the estimation error. The first term is computed by writing the estimate as $\hat{\mathbf{h}}_{il} = \sqrt{\frac{\eta_i}{\eta_k}}\mathbf{R}_{il}\mathbf{R}_{kl}^{-1}\hat{\mathbf{h}}_{kl}$:

$$\begin{aligned}&\operatorname{tr}\left(\mathbf{D}_{kl}\mathbb{E}\left\{\mathbf{h}_{il}\mathbf{h}_{il}^H\mathbf{D}_{kl}\hat{\mathbf{h}}_{kl}\hat{\mathbf{h}}_{kl}^H\right\}\right) \\ &= \frac{\eta_i}{\eta_k}\operatorname{tr}\left(\mathbf{D}_{kl}\mathbb{E}\left\{\mathbf{R}_{il}(\mathbf{R}_{kl})^{-1}\hat{\mathbf{h}}_{kl}\hat{\mathbf{h}}_{kl}^H(\mathbf{R}_{kl})^{-1}\mathbf{R}_{il}\mathbf{D}_{kl}\hat{\mathbf{h}}_{kl}\hat{\mathbf{h}}_{kl}^H\right\}\right) \\ &= \frac{\eta_i}{\eta_k}\mathbb{E}\left\{\left|\hat{\mathbf{h}}_{kl}^H\mathbf{D}_{kl}\mathbf{R}_{il}(\mathbf{R}_{kl})^{-1}\hat{\mathbf{h}}_{kl}\right|^2\right\} \\ &= \eta_k\eta_I\tau_p^2\left|\operatorname{tr}\left(\mathbf{D}_{kl}\mathbf{R}_{il}\Psi_{t_kl}^{-1}\mathbf{R}_{kl}\right)\right|^2 + \eta_k\eta_I\tau_p^2 \operatorname{tr}\left(\mathbf{D}_{kl}(\mathbf{R}_{il}-\mathbf{C}_{il})\mathbf{R}_{kl}\Psi_{t_kl}^{-1}\mathbf{R}_{kl}\right)^2\end{aligned} \tag{A.14}$$

where the full demonstration of the last step can be found in the Appendix of [8]. By adding the terms, we conclude the proof.