Ph.D. Dissertation:

Coordination Strategies for Interference Management in MIMO Dense Cellular Networks

Author:

Sandra Lagén Morancho

Advisors:

Adrián Agustín de Dios
Josep Vidal Manzano

Abstract

The envisioned rapid and exponential increase of wireless data traffic demand in the next years imposes rethinking current wireless cellular networks due to the scarcity of the available spectrum. In this regard, three main drivers are considered to increase the capacity of today's most advanced (4G systems) and future (5G systems and beyond) cellular networks:

- use more bandwidth (more Hz) through spectral aggregation,
- enhance the spectral efficiency per base station (BS) (more bits/s/Hz/BS) by using multiple antennas at BSs and users (i.e. **MIMO** systems), and
- increase the density of BSs (more BSs/km^2) through a dense and heterogeneous deployment (known as **dense heterogeneous cellular networks**).

We focus on the last two drivers. First, the use of multi-antenna systems allows exploiting the spatial dimension for several purposes: improving the capacity of a conventional point-to-point wireless link, increasing the number of served users, and reducing unwanted emissions (interference). Second, dense heterogeneous networks are a simple and cost-effective way to boost the area spectral efficiency by densifying the network with BSs that dispose of different coverage areas and by improving the spatial re-use of the spectrum.

However, increasing the density of BSs entails two main technical challenges:

- the interference in the network increases because neighboring BSs/users are nearer, and
- the amount of data traffic, as well as the downlink (DL) and uplink (UL) traffic asymmetry, varies over space and time more drastically since the number of users per BS is reduced.

The increase of interference in the network makes the development of efficient **interference management** techniques a key enabler for MIMO dense heterogeneous networks. But, as we move towards denser networks, interference management is becoming increasingly challenging.

On the other hand, the variability of the per-BS data traffic amount and of the DL/UL traffic asymmetry convert flexible duplexing (i.e. flexible and dynamic allocation of DL/UL resources per BS, either in time or frequency domain) into a necessity for an efficient radio resource usage that meets the non-uniform and time-varying DL/UL per-BS traffic loads. Therefore, the development of **traffic-aware resource management** schemes capable of adapting to the varying traffic load, as well as interference management, become crucial.

Accordingly, this doctoral thesis focuses on:

1. the development of advanced interference management techniques to deal with inter-cell interference in MIMO dense cellular networks, and
2. the design of traffic-aware and interference-aware resource management schemes for flexible duplexing systems in asymmetric traffic conditions.

To these goals, the wide deployment of MIMO systems is capitalized to develop advanced multi-antenna signal processing techniques when full reuse of time and frequency resources among densely deployed BSs is adopted.

In the first part of this work, different statistical characterizations of the transmitted signals are analyzed so as to improve the capacity of wireless interference channels. In this regard, **advanced signaling schemes** are developed and the use of improper Gaussian signaling (IGS) is investigated, which allows taking advantage of the real and imaginary dimensions of the MIMO channels by splitting one spatial dimension into two halves. Majorization theory is exploited to demonstrate the strict superiority of IGS. Then, the benefits of IGS are applied to different MIMO interference-limited scenarios.

Another way to manage interference under full frequency reuse is through the coordination and/or cooperation of BSs. Coordination among BSs allows adjusting in a coordinated manner the transmit strategies at different BSs so as to reduce the impact of interference in the network. In contrast, cooperation among BSs allows BSs to act as a single multi-antenna transmitter and has the great advantage of converting interference into useful signal through the joint transmission of cooperative BSs towards the same user. However, cooperation comes at the cost of a tight synchronization and high backhaul capacity to share user data among cooperative BSs. For that reason, in practical implementations, it can only be achieved between a limited number of BSs (which form a cluster) and coordination among clusters is still needed to deal with interference. Both coordination and cooperation, either implemented in a centralized or decentralized fashion, require knowledge of all channel matrices in the network, which imposes stringent channel estimation requirements for interference management in dense networks.

In the second part of the Ph.D. dissertation, **transmit coordination strategies** are proposed to manage interference in extremely dense cellular networks. The focus is on the DL data transmission. The design of the BSs transmit strategies (involving design of the spatial transmit and receive filters, transmit power control, and scheduling of users) is coordinated with the objective of optimizing different network functions (as, for instance, the weighted sum of the user throughputs) while reducing the stringent requirements needed for channel estimation in dense networks. Coordination strategies for the case in which different signaling schemes (proper and improper) coexist in the network are also derived. Further, the thesis develops coordination strategies for cluster-based joint transmissions, where BSs are grouped into clusters formed by a low number of cooperative BSs and different clusters interfere to each other. In this case, the transmit strategy is jointly optimized together with the user-centric cluster formation.

Finally, we address traffic-aware and interference-aware resource management in flexible duplexing systems, where resources have to be properly distributed between DL and UL according to the traffic load and traffic asymmetries of each BSs. Under reuse of resources among densely deployed BSs, the use of flexible duplexing entails changes to the interference generated between neighbor BSs/users. As a consequence, new kinds of interference (like BS to BS) arise.

The third part of this thesis focuses on the design of **traffic-aware duplexing techniques** for resource management and interference management. In contrast to the previous parts, DL and UL data transmissions are considered for each BS. The main objective is to make a better use of the available time/frequency resources by taking into account the asymmetric traffic conditions that arise in dense networks as well as managing the new kinds of interference that come up under flexible duplexing. Short-term and long-term optimizations are investigated, being therefore the interference managed instantaneously and statistically, respectively.

Resumen

El aumento rápido y exponencial previsto para la demanda de tráfico de datos en los próximos años impone rediseñar las redes celulares inalámbricas actuales debido a la escasez del espectro radioeléctrico disponible. En este sentido, se consideran tres ejes directores para aumentar la capacidad de las redes celulares más avanzadas de hoy en día (sistemas 4G) y las del futuro (sistemas 5G y más allá):

- utilizar más ancho de banda (más Hz) a través de la agregación de espectro,

- mejorar la eficiencia espectral por estación base (BS) (más bits/s/Hz/BS) utilizando múltiples antenas en las BSs y los usuarios (sistemas **MIMO**), e

- incrementar la densidad de BSs (más BSs/km^2) mediante un despliegue denso y heterogéneo (conocido como **redes densas y heterogéneas**).

Esta tesis se centra en los dos últimos ejes directores. En primer lugar, el uso de sistemas multi-antena permite explotar la dimensión espacial con varias finalidades: mejorar la capacidad de un enlace inalámbrico convencional punto a punto, incrementar el número de usuarios servidos y reducir emisiones indeseadas (interferencias). En segundo lugar, las redes densas y heterogéneas son una manera simple y rentable de mejorar la eficiencia espectral por área a través de la densificación de la red con BSs de diferentes características y de la reutilización espacial del espectro radioeléctrico.

Sin embargo, el incremento de la densidad de BSs plantea dos principales desafíos técnicos:

- las interferencias en la red aumentan porque BSs/usuarios vecinos están más próximos y

- la cantidad de tráfico de datos, así como la asimetría del tráfico de bajada (DL) y de subida (UL), fluctúa con el tiempo y el espacio más drásticamente debido a que el número de usuarios por BS se reduce.

El aumento de interferencias en la red hace que un factor clave para las redes MIMO densas y heterogéneas sea el desarrollo de técnicas eficientes de **gestión de interferencias**. Pero, a medida que avanzamos hacia redes más densas, la gestión de interferencias se convierte cada vez en un reto más desafiante.

Por otro lado, la variabilidad de la cantidad de tráfico de datos por BS y de la asimetría del tráfico DL/UL convierten en una necesidad el duplexado flexible (es decir, asignaciones flexibles y dinámicas de recursos DL/UL por BS, ya sea en el dominio temporal o frecuencial) para conseguir un uso eficiente de los recursos radio que satisfaga las cargas de tráfico no uniformes en espacio y variantes en tiempo. Por lo tanto, se vuelve crucial el desarrollo de esquemas de **gestión de recursos** capaces de adaptarse a cargas de tráfico variable y de, a su vez, gestionar las interferencias.

En este sentido, esta tesis doctoral se centra en:

1. el desarrollo de técnicas avanzadas de gestión de interferencias para hacer frente a las interferencias entre celdas en redes celulares MIMO densas, y

2. el diseño de esquemas de gestión de recursos que tengan en cuenta el tráfico y la interferencia para sistemas de duplexado flexible bajo condiciones de tráfico asimétricas.

Para alcanzar estos objetivos, se aprovecha el amplio despliegue de sistemas MIMO con el fin de desarrollar técnicas multi-antena avanzadas de procesado de señales cuando se adopta un reúso completo de los recursos en tiempo y en frecuencia entre BSs densamente desplegadas en la red.

En la primera parte de la tesis, se analizan diferentes caracterizaciones estadísticas de las señales de transmisión para mejorar la capacidad de los canales inalámbricos interferentes. En este sentido, se desarrollan **esquemas de señalización avanzados** y se investiga el uso de la señalización Gaussiana improper (IGS), la cual permite aprovechar las dimensiones reales e imaginarias de los canales de propagación MIMO mediante la división de una dimensión espacial en dos mitades. La teoria de la majorización se explota para demostrar la superioridad estricta de IGS. Después, los beneficios de IGS se aplican a diferentes escenarios MIMO limitados por interferencia.

Otra forma de gestionar la interferencia con reuso completo de los recursos frecuenciales es mediante la coordinación y/o cooperación de BSs. La coordinación entre BSs permite ajustar de manera coordinada las estrategias de transmisión de diferentes BSs con el objetivo de reducir el impacto de las interferencias en la red. Por el contrario, la cooperación entre BSs permite que las BSs actuen como un único transmisor multi-antena y tiene la gran ventaja de que convierte la interferencia en señal útil a través de la transmisión conjunta de BSs cooperativas hacia un mismo usuario. Sin embargo, la cooperación requiere sincronización estricta y alta capacidad de backhaul para compartir datos de usuario entre BSs. Por esta razón, en implementaciones prácticas, la cooperación sólo se puede lograr entre un número reducido de BSs (las cuales forman un grupo) y la coordinación entre grupos sigue siendo necesaria para hacer frente a las interferencias. Tanto la coordinación como la cooperación, ya sean implementadas de forma centralizada o descentralizada, requieren el conocimiento de todos los canales de propagación de la red, lo cual impone requisitos estrictos en cuanto a estimación de canal para la gestión de interferencias en redes densas.

En la segunda parte de este trabajo se proponen **estrategias de transmisión coordinadas** para gestionar interferencias en las redes celulares extremadamente densas. El foco está en la transmisión DL. El diseño de las estrategias de transmisión en las BSs (incluyendo el diseño de los filtros espaciales de transmisión y recepción, el control de potencia y la selección de usuarios) es coordinado con tal de optimizar diferentes funciones de red (como, por ejemplo, la suma ponderada de las tasas de transmisión), mientras que se reducen los estrictos requisitos necesarios para estimación de canal en redes densas. También se analizan estrategias de coordinación para el caso en que diferentes esquemas de señalización (proper e improper) coexisten en la red. Además, la tesis deriva estrategias de coordinación para transmisiones conjuntas basadas en grupos, donde las BSs se agrupan en grupos formados por un número reducido de BSs cooperativas y grupos vecinos se interfieren entre sí. En este caso, la estrategia de transmisión se optimiza conjuntamente con la formación de los grupos.

Por último, se aborda la gestión de recursos en sistemas de duplexado flexible, donde los recursos tienen que ser distribuidos adecuadamente entre las transmisiones DL y UL de acuerdo con las asimetrías y la cantidad de tráfico de cada BS. Bajo una reutilización de recursos en BSs densamente desplegadas, el uso del duplexado flexible conlleva cambios en la interferencia generada entre BSs y/o usuarios vecinos. Como consecuencia, surgen nuevos tipos de interferencias (como la interferencia de BS a BS).

La tercera parte de la tesis se centra en el diseño de **técnicas de duplexado flexible** que tienen en cuenta el tráfico para la gestión de recursos y de interferencias. En contraste con las partes anteriores, se consideran transmisiones DL y UL para cada BS. El objetivo principal es hacer un mejor uso de los recursos tiempo/frecuencia disponibles, teniendo en cuenta las condiciones de tráfico asimétricas que surgen en redes densas así como la gestión de los nuevos tipos de interferencias que aparecen bajo sistemas de duplexado flexible. Se investigan optimizaciones a corto plazo y a largo plazo, siendo entonces la interferencia gestionada de manera instantánea y de manera estadística, respectivamente.

Acknowledgments

Escriure aquestes línies em provoca una barreja de sentiments contrastats perquè significa acabar no només el doctorat sinó també la meva estada a la UPC. Sento alegria pels bons moments viscuts i per tot l'aprenentatge que m'emporto però també tristesa pel fet d'acabar una gran etapa a la UPC, on em sento com a casa.

Els que han fet més fort aquest vincle han estat els meus supervisors, l'Adrián Agustín i el Josep Vidal. Tot va començar amb tu, Josep, quan, després de l'examen de PS, ens vas dir si volíem col·laborar amb el vostre grup. D'aquí va sortir el PFC al 2011. Va seguir a Roma, en un viatge del FREEDOM, quan em vau preguntar si m'agradaria fer el doctorat amb vosaltres. I així va ser.

Per mi, compartir és la clau de la vida. Compartir permet transmetre i, per tant, aprendre. Em sento molt afortunada perquè és exactament el que heu fet amb mi: compartir visions, discussions, opinions, coneixements, experiències, aprenentatges, cafès, dinars, viatges, esquiades... Vull agrair-vos la confiança que heu dipositat en mi, els dos, sempre. Gràcies per tot el coneixement, aptituds i valors que heu sabut transmetre'm. També vull agrair-vos que m'hàgiu involucrat en projectes de recerca i que m'hàgiu donat l'oportunitat d'escriure una patent i de participar en les contribucions al 3GPP. He après moltíssim.

No podia haver tingut un millor "tàndem" que vosaltres. Adri, gràcies per la teva motivació, inspiració, ajuda i dedicació, que han estat claus per completar molts punts d'aquesta tesi. M'has ensenyat a afrontar els reptes, a replantejar els problemes, a saber trobar sentit i aplicacions a les solucions. A més, m'has ajudat aportant molt bones idees, debatent les propostes, comentant cada detall, aconsellant-me quan no sabia per on tirar... Per tot això, n'estic enormement agraïda. Josep, tu vas ser el primer en confiar en mi i en donar-me els primers coneixements de processament de senyals per a comunicacions mòbils. Vull donar-te les gràcies per saber contagiar la teva motivació, inquietud i entusiasme, per fer que entendre's amb tu sigui tan fàcil, per haver-me donat un enorme aprenentatge tècnic i personal (perquè tant m'has ensenyat de codis ortogonals com de codis ètics) i, en definitiva, per ser un tutor i una persona excepcional. Tens uns valors que admiro i que marquen la diferència entre un cap i un gran líder (com tu).

També vull donar les gràcies als companys de treball del projecte TROPIC i del de Huawei: Olga Muñoz, Toni Pascual i Marc Torrellas, per les hores compartides junts. Olga y Toni, gracias por darme la oportunidad de colaborar con vosotros, de aprender de vosotros, y también os agradezco la implicación en los trabajos realizados conjuntamente. Ha sido un placer. Además, Toni, gracias de verdad por tu entusiasmo, por contar conmigo y por ayudarme a encarar el futuro. Outside UPC (but related to the TROPIC project), I want to thank Mariana Goldhamer for the long and fruitful discussions and for providing us with enough support to submit contributions to 3GPP.

Tornant a la UPC, vull agrair a tots els professors que he tingut del màster MERIT per dedicar temps de la seva vida a ensenyar (Gregori Vázquez, Javier R. Fonollosa, Miguel A. Lagunas, Alba Pagès, Ana I. Pérez, Javier Villares, Antonio Pascual i Francesc Rey). He tingut grans professors i part dels coneixements apressos m'han estat molt útils per realitzar aquesta tesi. Agraeixo especialment a aquells que estimen la ciència i saben transmetre-ho. De la UPC, també vull agrair sincerament a l'Anna Lladó per la seva amabilitat i atenció per resoldre dubtes de l'apassionant teoria de grafs i pels seus valuosos comentaris.

Els meus agraïments també van cap als companys de despatx del D5: Javier Rubio, Jaume del Olmo, Igor Jauk, Pere Giménez, Josep Font, Eva Lagunas, Màrius Caus i Adriano Pastore. En especial, Javi, quiero darte las gracias por tu generosidad, por tus consejos, por presionarme e introducirme al LaTeX, por las discusiones y por tener un momento siempre que te lo he pedido. També vull agrair al Jaume per la seva companyia en moments necessaris de descans i per la disposició a escoltar i ajudar.

I also want to thank the people from Nokia Networks and Aalborg University that made my three-month research appointment there so easy. Specially to Beatriz Soret and Klaus Pedersen, who welcomed me so kindly and organized interesting meetings during my visit. Also to the every-day colleagues in Nokia (Marta Gatnau, Andrijana Popovska, Laura Luque, Guillermo Pocovi, Víctor Fernández, and Lucas Chavarría) and to the ones from Aalborg University for the fruitful discussions (Gilberto Berardinelli and Nurul Huda). I cannot forget Jagjit Singh, who helped me to discover the town, gràcies crack!

En l'àmbit personal, vull agrair a les meves nenes (Lidia, Blanca, Marta, Judith, Alba, Laura, Judit, Sole i Marina) per saber distreure'm quan calia desconnectar. En especial a la Lidia i la Blanca, per estar sempre al meu costat i per aquesta complicitat tan especial. També vull agrair a la Mariona, per les estones de comprensió i reflexió compartides juntes. Finalment, a alguns companys de la carrera que, tot i la distància, sempre m'enrecordo de vosaltres en moments claus de la vida: Edu, Gisela, Anna i Cris.

Per últim, i molt merescut, vull agrair als meus pares, Ana i Pedro, a l'Òscar, al meu germà Víctor, i a la meva familia, Lagén i Morancho, pel seu suport incondicional, per preocupar-se per mi cada dia i per haver-me ajudat a arribar fins aquí. Tots sou part de mi.

<div align="right">

Sandra Lagén

Barcelona, 2016

</div>

Acronyms and Abbreviations

2G	2nd Generation mobile networks (GSM, CDMA 1x).
3D	3 Dimensions.
3G	3rd Generation mobile networks (WCDMA/HSPA, TD-SCDMA, CDMA EV-DO, Mobile WiMAX).
3GPP	3rd Generation Partnership Project.
4G	4th Generation mobile networks (LTE, LTE-A).
5G	5th Generation mobile networks (not yet standardized).
ABS	Almost Blank Subframes.
AWGN	Additive White Gaussian Noise.
BC	Broadcast Channel.
BCD	Block Coordinate Descent.
BF	Beamforming.
BS	Base Station.
CDF	Cumulative Distribution Function.
CDMA	Code Division Multiple Access.
CE	Calibration Error.
CoMP	Coordinated Multi-Point.
CoMP-CB	Coordinated Multi-Point Coordinated Beamforming.
CoMP-CS	Coordinated Multi-Point Coordinated Scheduling.
CoMP-DPS	Coordinated Multi-Point Dynamic Point Selection.
CoMP-JR	Coordinated Multi-Point Joint Reception.
CoMP-JT	Coordinated Multi-Point Joint Transmission.
Co-SRS	Coordinated Sounding Reference Signals.
CQI	Channel Quality Indicator.
CSCG	Circularly Symmetric Complex Gaussian.
CS-CP-CD	Coordinated Scheduling, Coordinated Precoding and Coordinated Direction.
CS-CP-uD	Coordinated Scheduling, Coordinated Precoding and uncoordinated Direction.
CSI	Channel State Information.
CSIT	Channel State Information at the Transmitter.
D2D	Device-to-Device.

D-CoP	Decentralized Coordinated Precoding.
DJCP	Decentralized Joint Clustering and Precoding.
DL	Downlink.
DM-RS	Demodulation Reference Signals.
DoF	Degrees of Freedom.
DPC	Dirty Paper Coding.
e.g.	for example.
eICIC	enhanced Inter-Cell Interference Coordination.
eIMTA	enhanced DL-UL Interference Management and Traffic Adaptation.
eNB	evolved Node B.
EVD	Eigenvalue Decomposition.
FeICIC	Further enhanced Inter-Cell Interference Coordination.
FDD	Frequency Division Duplex.
FDMA	Frequency Division Multiple Access.
FFR	Fractional Frequency Reuse.
FR	Frequency Reuse.
FTP	File Transfer Protocol.
FTP3	File Transfer Protocol model 3.
GP	Guard Period.
HCN	Heterogeneous Cellular Network.
HeNB	Home evolved Node B.
HetTX	Heterogeneous Transmitters.
IA	Interference Alignment.
IC	Interference Channel.
ICIC	Inter-Cell Interference Coordination.
i.e.	that is.
IEEE	Institute of Electrical and Electronics Engineers.
IGS	Improper Gaussian Signaling.
i.i.d.	independent and identically distributed.
INR	Interference-to-Noise Ratio.
IRC	Interference Rejection Combining.
ISIM	Interference Suppressing Interference Mitigation.
ITU	International Telecommunication Union.
IW	Iterative Water-filling.
KKT	Karush-Kuhn-Tucker.
LE	Linear Estimation.
LMMSE-IRC	Linear Minimum Mean Square Error - Interference Rejection Combining.
LOS	Line-of-Sight.

LP	Linear Precoding.
LTE	Long Term Evolution.
LTE-A	Long Term Evolution - Advanced.
MAC	Multiple Access Channel.
MC	Macro Cell.
MCS	Modulation and Coding Scheme.
MeNB	Macro evolved Node B.
MIMO	Multiple-Input Multiple-Output.
MISO	Multiple-Input Single-Output.
ML	Maximum Likelihood.
MMSE	Minimum Mean Square Error.
MMSE-IRC	Minimum Mean Square Error - Interference Rejection Combining.
MP2MP	Multiple-Point to Multiple-Point.
MRC	Maximum Ratio Combining.
MRT	Maximum Ratio Transmission.
MSE	Mean Square Error.
MTC	Machine Type Communications.
MUE	Macrocell User Equipment.
NAICS	Network Assisted Interference Cancellation and Suppression.
NIB	Non-Ideal Backhaul.
NLOS	Non-Line-of-Sight.
NP	Nondeterministic Polynomial time.
OFDM	Orthogonal Frequency Division Multiplexing.
OFDMA	Orthogonal Frequency Division Multiple Access.
OPA	Optimal Power Allocation.
P2P	Point-to-Point channel.
P2P-I	Point-to-Point channel with Interference.
PDSCH	Physical Downlink Shared Channel.
PF	Proportional Fair.
PGS	Proper Gaussian Signaling.
PMI	Precoding Matrix Indicator.
PUCCH	Physical Uplink Control Channel.
PUSCH	Physical Uplink Shared Channel.
QoS	Quality of Service.
RB	Resource Block.
REB	Range Extension Bias.
RF	Radio-Frequency.
RI	Rank Indicator.

RNTP	Relative Narrow-band Transmit Power.
RSRP	Reference Signal Received Power.
RSRQ	Reference Signal Received Quality.
RU	Resource Utilization.
RX	Receiver.
SB	Subband.
SC	Smallcell.
SCN	Smallcell Network.
SeNB	Small evolved Node B.
SFN	Subframe Number.
SIC	Successive Interference Cancellation.
SIMO	Single-Input Multiple-Output.
SINR	Signal-to-Interference-plus-Noise Ratio.
SIR	Signal-to-Interference Ratio.
SISO	Single-Input Single-Output.
SLNR	Signal-to-Leackage-and-Noise Ratio.
SNR	Signal-to-Noise Ratio.
SR	Sum-Rate.
SRS	Sounding Reference Signal.
SUE	Smallcell User Equipment.
SVD	Singular Value Decomposition.
SYNC	Synchronization.
TDD	Time Division Duplex.
TDMA	Time Division Multiple Access.
THP	Tomlinson-Harashima precoding.
TM	Transmission Mode.
TR	Technical Report.
TSG-RAN	Technical Specification Group - Radio Access Network.
TTI	Transmission Time Interval.
TX	Transmitter.
UE	User Equipment.
UL	Uplink.
UMTS	Universal Mobile Telecommunications System.
UPA	Uniform Power Allocation.
UpPTS	Uplink Pilot Time Slot.
UPT	User Packet Throughput.
uS-uP-uD	Uncoordinated Scheduling, uncoordinated Precoding and uncoordinated Direction.

UT	User Throughput.
vs.	versus.
WG	Working Group.
WiMAX	Worldwide Interoperability for Microwave Access.
WLE	Widely Linear Estimation.
WLP	Widely Linear Precoding.
WMSE	Weighted Mean Square Error.
WMMSE	Weighted Minimum Mean Square Error.
w.r.t.	with respect to.
WSR	Weighted Sum-Rate.
ZF	Zero Forcing.
Z-IC	Z-Interference Channel.

Notation

x	A scalar.
$\mathbf{x}$	A column vector.
$\mathbf{X}$	A matrix.
$\mathcal{X}$	A set.
$\mathbf{I}$	The identity matrix.
$\mathbf{J}$	The anti-identity matrix.
$\mathbf{0}$	The zero vector/matrix.
$\emptyset$	The empty set.

Matrices

$\mathbf{X}^T$	The transpose matrix of matrix $\mathbf{X}$.
$\mathbf{X}^H$	The hermitian matrix of matrix $\mathbf{X}$.
$\mathbf{X}^*$	The conjugate matrix of matrix $\mathbf{X}$.
$\mathbf{X}^{-1}$	The inverse matrix of matrix $\mathbf{X}$.
$\mathbf{X}^+$	The Moore-Penrose pseudoinverse matrix of matrix $\mathbf{X}$.
$\mathbf{X}^{\frac{1}{2}}$	The square root matrix of a positive semi-definite matrix $\mathbf{X}$, i.e. $\mathbf{X} = \mathbf{X}^{\frac{1}{2}}(\mathbf{X}^{\frac{1}{2}})^H$.
$\mathrm{Tr}(\mathbf{X})$	The trace of matrix $\mathbf{X}$.
$|\mathbf{X}|$	The determinant of matrix $\mathbf{X}$.
$\mathrm{rank}(\mathbf{X})$	The rank of matrix $\mathbf{X}$.
$\|\mathbf{X}\|_F^2$	The squared Frobenius norm of matrix $\mathbf{X}$, i.e. $\|\mathbf{X}\|_F^2 = \sum_i \sum_j |x_{i,j}|^2$.
$\mathbb{E}[\mathbf{X}]$	The expectation of matrix $\mathbf{X}$.
$\mathbb{E}[\mathbf{X}|\mathbf{Y}]$	The expectation of matrix $\mathbf{X}$ for a known matrix $\mathbf{Y}$.
$\Re\{\mathbf{X}\}$	The real part of matrix $\mathbf{X}$.
$\Im\{\mathbf{X}\}$	The imaginary part of matrix $\mathbf{X}$.
$\mathbb{1}\{\mathbf{X}\}$	The indicator function defined as $\mathbb{1}\{\mathbf{X}\} = 0$ if $\mathbf{X} = \mathbf{0}$, otherwise $\mathbb{1}\{\mathbf{X}\} = 1$.
$\frac{df(\mathbf{X})}{d\mathbf{X}}$	The derivative of function $f(\mathbf{X})$ with respect to matrix $\mathbf{X}$.
$\{\mathbf{X}_k\}$	The set of matrices $\{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_K\}$.
$\mathbf{X} \succeq \mathbf{0}$	Matrix $\mathbf{X}$ is positive semi-definite.

$\mathbf{X} = \mathrm{diag}(x_1 \ldots x_n)$	The diagonal matrix of size $n \times n$ with values $x_1, x_2, \ldots, x_n$ in its diagonal and 0 in the non-diagonal elements.
$\mathbf{x} = \mathrm{diag}(\mathbf{X})$	The vector containing the diagonal values of matrix $\mathbf{X}$.
$\mathbf{x} = \mathrm{eig}(\mathbf{X})$	The vector that stacks the eigenvalues of matrix $\mathbf{X}$ in decreasing order.
$x = \mathrm{eig}(\mathbf{X})_i$	The i-th eigenvalue of matrix $\mathbf{X}$.

Vectors

$\mathbf{x}^{-1}$	The vector that contains the inverse of each element of vector $\mathbf{x}$ in decreasing order, for a positive real-valued vector $\mathbf{x}$.
$\bar{\mathbf{x}}$	The double-sized real-valued vector of vector $\mathbf{x}$, which stacks its real and imaginary parts as: $\bar{\mathbf{x}} = \left[\Re\{\mathbf{x}\}^T \ \Im\{\mathbf{x}\}^T \right]^T$.
$\underline{\mathbf{x}}$	The double-sized complex-valued vector of vector $\mathbf{x}$, which stacks the vector and its complex conjugate as: $\underline{\mathbf{x}} = \left[\mathbf{x}^T \ (\mathbf{x}^*)^T \right]^T$.
$\|\mathbf{x}\|^2$	The squared 2-norm of vector $\mathbf{x}$, i.e. $\|\mathbf{x}\|^2 = \sum_i |x_i|^2$.
$\mathbf{x} \circ \mathbf{y}$	The Hadamard product of vectors $\mathbf{x}$ and $\mathbf{y}$, i.e. it returns a vector with the component-wise product of the vectors $\mathbf{x}$ and $\mathbf{y}$.
$\mathbf{x} \succ \mathbf{y}$	Vector $\mathbf{x}$ strongly majorizes vector $\mathbf{y}$.
$\mathbf{x} \succ_w \mathbf{y}$	Vector $\mathbf{x}$ weakly majorizes vector $\mathbf{y}$.
$\mathbf{x} \succ_{w \log} \mathbf{y}$	Vector $\mathbf{x}$ weakly log-majorizes vector $\mathbf{y}$.
$x = \min(\mathbf{x})$	The minimum value of the elements in vector $\mathbf{x}$.

Scalars

$|x|$	The complex modulus of scalar x.
$\log_2(x)$	The base-2 logarithm of real-valued scalar x.
$\ln(x)$	The natural logarithm of real-valued scalar x.
$\lfloor x \rfloor$	The floor of real-valued scalar x.
$\min(x, y)$	The minimum between real-valued scalars x and y.
$(x)^+$	The maximum between real-valued scalars x and 0.
$\Pr(x < y)$	The probability of real-valued scalar x being smaller than real-valued scalar y.
$\frac{df(x)}{dx}$	The derivative of function $f(x)$ with respect to scalar x.
$f(x)|_y$	Function $f(x)$ evaluated at $x = y$.

Sets

$|\mathcal{X}|$	The cardinality of set $\mathcal{X}$.
$\mathcal{X} \cap \mathcal{Y}$	The intersection of sets $\mathcal{X}$ and $\mathcal{Y}$.
$\mathcal{X} \cup \mathcal{Y}$	The union of sets $\mathcal{X}$ and $\mathcal{Y}$.
$\mathcal{X} \subseteq \mathcal{Y}$	Set $\mathcal{X}$ is a subset of set $\mathcal{Y}$ (i.e. set $\mathcal{X}$ is included in set $\mathcal{Y}$).

$x \in \mathcal{X}$	Scalar x belongs to set $\mathcal{X}$.
$x \notin \mathcal{X}$	Scalar x does not belong to set $\mathcal{X}$.

Other

$\triangleq$	Defined as.
$\approx$	Approximately equal to.
$\sim$	Distributed according to.
$\mathbb{C}^{m \times n}$	The m by n dimensional complex space.
$\mathbb{R}^{m \times n}$	The m by n dimensional real space.
$\mathbb{R}_+^{m \times n}$	The m by n dimensional real positive space.
$\mathcal{CN}(\mathbf{m}, \mathbf{C})$	The circularly symmetric complex normal distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$ of appropriate dimensions.
$\mathcal{TN}(\mathbf{m}, \mathbf{C}, x, y)$	The circularly symmetric real truncated normal distribution in the interval $[x, y]$ with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$ of appropriate dimensions.

Contents

Chapter 1

Introduction

With the advent of new sophisticated terminals and bandwidth-demanding services, which are changing our every-day life, data traffic demand in wireless networks is experiencing an explosive growth. Thus, system designers of cellular networks are pushed towards satisfying the strict requirements of: enhancing the area spectral efficiency [in bits/s/Hz/km^2], providing homogeneous coverage, improving the per-user data rate (or link capacity) [in bits/s], reducing the energy consumption and cost, diminishing the end-to-end latency, addressing massive connections of devices, and providing ultra-reliable and low-latency communications.

In order to address such massive data demands and to meet the strict requirements, constrained to the fact that the available spectrum is scarce, today's most advanced cellular networks (as Long-Term Evolution Advanced (LTE-A) for fourth generation (4G) systems [1, 2]) and future cellular networks (as fifth generation (5G) systems and beyond [3, 4]) consider three main drivers for wireless evolution [5, 6] (see Fig. 1.1):

1. Use more **bandwidth** (more Hz) through spectral aggregation. Spectral aggregation consists on utilizing larger portions of the spectrum in diverse bands, including the unlicensed spectrum (as the 5 GHz band used by Wi-Fi) [7] and the millimeter-wave bands (30-300 GHz) [8]. Bands over the 6 GHz have not been already licensed, but the advantage is that larger bandwidths are available there. However, the aggregation of disparate frequency bands unleashes a spate of challenges for antenna and radio-frequency transceiver design that need to be overcome to support spectral aggregation [6].

2. Enhance the **spectral efficiency** per base station (BS) (more bits/s/Hz/BS). One of the main technologies considered to improve the spectral efficiency is the use of multi-antenna terminals at BSs and users (i.e. **multiple-input multiple output** (MIMO) systems). The case where the number of antennas at the BSs is very large is receiving much attention for 5G systems, namely massive MIMO or large-scale MIMO [9]. Also, coordinated multi-point (CoMP) strategies and interference suppression/cancellation techniques are key enablers to enhance the spectral efficiency of the system [10–12].

3. Increase the **density** of BSs and devices deployed per geographical area (more loaded BSs/km^2), i.e. spatial densification. Usually, spatial densification is done through BSs of different characteristics (e.g. with different coverage areas), hence conforming the so-called **dense heterogeneous cellular networks** (HCNs) [11, 13]. Also, device-to-device (D2D) communications and machine type communications (MTC) are part of the so-called network densification [14, 15].
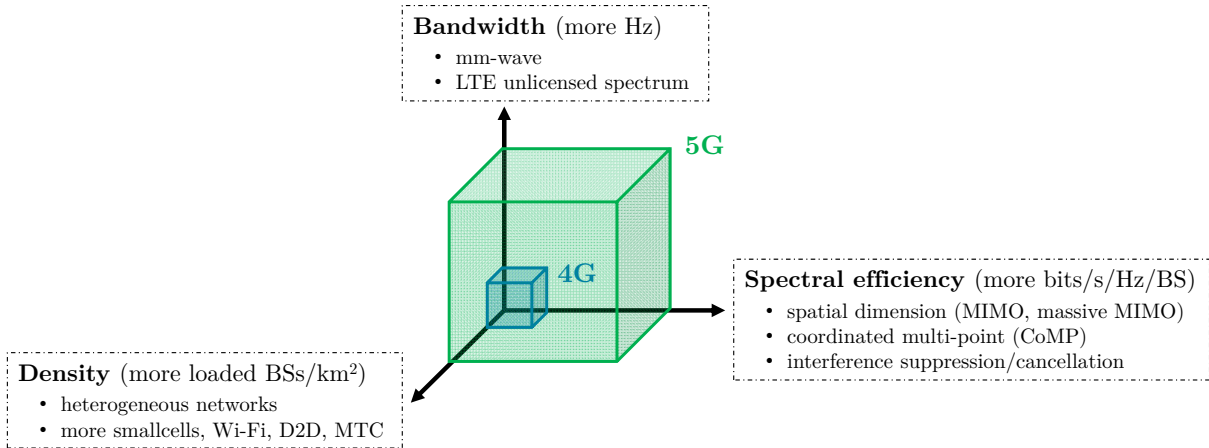
Figure 1.1: Drivers for wireless evolution.

The product of the three aforementioned drivers constitutes the network capacity in terms of bits/s/km^2. Among them, the most promising driver to significantly improve the network capacity has been shown to be the spatial densification [5] (although the three do contribute).

Remark 1.1. *This doctoral thesis focuses on spatial densification and spectral efficiency enhancement for systems with scarce bandwidth availability. Spectral aggregation to support using more bandwidth is out of the scope of the thesis.*

Increasing the density of BSs involves [6]:

- a reduction of the load factor (or number of users per BS) because user data traffic is sparsely distributed among BSs, and

- an increase of both the desired signal power and the received interference due to the fact that the serving BS as well as the interfering BSs are closer to the user.

The former contributes to the improvement of the per-link capacity because more resources can be devoted per user. Regarding the latter, the increase of the desired signal power is beneficial in terms of spectral efficiency provided that the increased level of interference is properly managed, otherwise the spectral efficiency gain vanish (and so does the per-link capacity). The gains of spatial densification are obtained at the cost of increased interference in the network [16].

⋆ *Challenge 1 of spatial densification: interference levels increase significantly.*

Consequently, **interference management** becomes a key enabler for future cellular systems. But, the denser is the network the more challenging interference management is [11]. For that reason, it has received much attention in the last years and has been extensively studied in academia, industry, and standardization bodies such as third Generation Partnership Project (3GPP)[1] [17]. On the whole, interference management involves many approaches (avoid, cancel, reduce, exploit), multiple domains (space, frequency, time, modulation, user, power, code, transmit direction), mechanisms (coordination, cooperation), and kinds of optimizations (short-term, long-term) to tackle interference (a complete review is included in Section 1.2).

---

[1]3GPP is a collaboration between groups of telecommunications associations that is in charge of the development and maintenance of: (*i*) GSM and related 2G and 2.5G standards including GPRS and EDGE, (*ii*) UMTS and related 3G standards including HSPA, (*iii*) LTE and related 4G standards, and (*iv*) 5G standards.

A classical approach to address interference is through the medium access control and medium sharing techniques, which severely compromise network performance to manage interference due to explicit time sharing over the common resources. Previous generations of cellular network standards employed orthogonal frequency reuse-n schemes, where the total bandwidth is partitioned into n orthogonal sets of frequency resources (bands) and neighboring BSs employ different bands. Hence, BSs do not interfere to each other (i.e. interference is avoided) at the expenses of reducing the area spectral efficiency of the system. Differently, Universal Mobile Telecommunications System (UMTS) (i.e. third generation (3G) systems) [18] and LTE/LTE-A (i.e. 4G systems) [2] moved from reuse-n to reuse-1 (i.e. full reuse), hence allowing a better reuse of the resources while rendering interference to its limits [10]. We follow this direction in the present work.

Remark 1.2. *This Ph.D. dissertation capitalizes on the wide deployment of MIMO systems to develop advanced multi-antenna signal processing techniques capable of managing interference when full reuse (i.e. reuse-1) of time and frequency resources among BSs densely deployed in the network is adopted.*

From a system perspective, the reduction of the load factor in dense networks raises an additional challenge. The per-BS traffic is more user dependent and thus the amount of data traffic and the downlink (DL) and uplink (UL) traffic asymmetries vary over space (i.e. BSs) and time more drastically as compared to conventional macrocell-based networks.

⋆ ***Challenge 2*** *of spatial densification: traffic is increasingly asymmetric.*

In this sense, **flexible duplexing** systems (i.e. systems that allow a flexible allocation of DL/UL resources among BSs) are being considered a key component of future cellular networks in order to satisfy asymmetric data traffic demands and allow an efficient use of the available spectrum [19]. Under flexible duplexing, new kinds of interference arise when neighboring BSs select different transmission patterns (e.g. one BS in DL and one BS in UL). Therefore, **traffic-aware and interference-aware resource management** schemes are crucial to meet the non-uniform and time-varying DL/UL per-BS traffic load while managing interference.

Remark 1.3. *This Ph.D. dissertation derives traffic-aware and interference-aware resource management schemes for flexible duplexing systems in asymmetric traffic conditions. For short-term management, the deployment of MIMO systems is leveraged to develop advanced coordinated strategies. For long-term management, graph coloring is exploited to distribute resources among densely deployed BSs.*

This chapter provides an overview of interference management for MIMO dense HCNs and presents the organization of the present document and the main results produced. First, Section 1.1 describes the dense HCN deployments, including co-channel and non-co-channel deployments, and its main impairments due to interference. Next, Section 1.2 introduces interference management, describing the available approaches, domains, mechanisms, and optimizations for interference management. Then, Section 1.3 presents the interference scenarios to be addressed in the thesis. After that, Section 1.4 contains a review of the available techniques for interference management in 3GPP standardization body [17] and Section 1.5 states the main challenges related to interference management in multi-cell scenarios with full reuse of resources. Section 1.6 specifies the mathematical tools and the knowledge required for the completion of the thesis. Finally, Section 1.7 presents the organization and contents of the Ph.D. dissertation and Section 1.8 details the research contributions that have been produced from the present work.
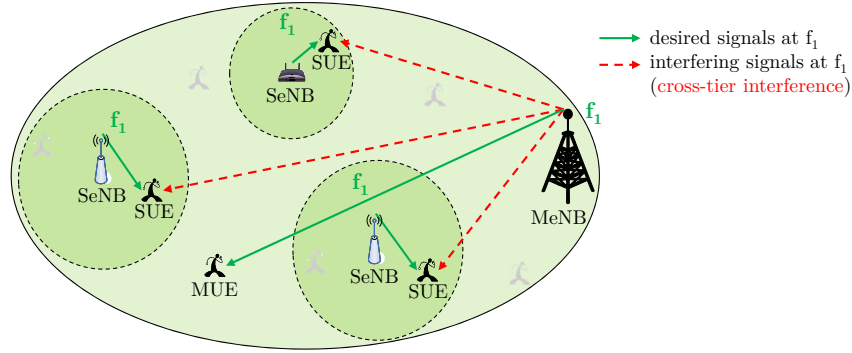
Figure 1.2: Co-channel deployment with uniformly deployed SeNBs. Major impairment: cross-tier interference.
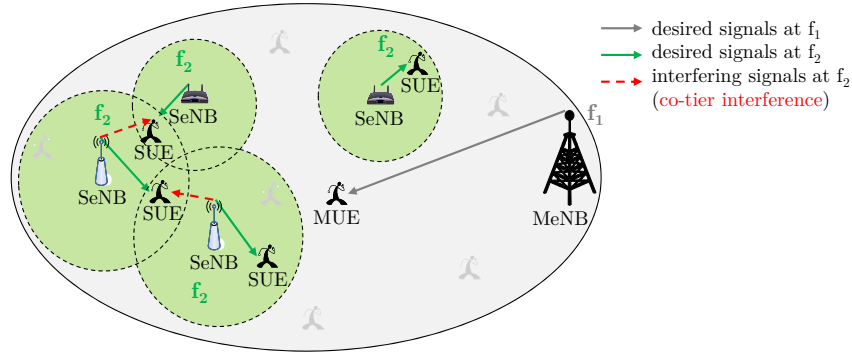
Figure 1.3: Non-co-channel deployment with densely deployed SeNBs. Major impairment: co-tier interference.

1.1 Dense Heterogeneous Cellular Networks

Dense HCNs consist of a multi-tier multi-cell deployment of macrocells and densely deployed smallcells (encompassing picocells, femtocells, and relays), which are being extensively studied in academia, industry, and standardization bodies such as 3GPP [17]. Each tier corresponds to a deployment of a different type of BSs[2]. Generally, dense HCNs correspond to a two-tier deployment if we include all smallcells in the same tier, i.e. the macro-tier and the small-tier. Macrocells are covered by macro evolved Node Bs (MeNBs), which are intended to provide general coverage and service to high mobility User Equipments (UEs). Differently, smallcells are served by small evolved Node Bs (SeNBs), which dispose of a lower transmit power, cover smaller areas than MeNBs, and are intended to serve a small number of low mobility UEs [13].

Dense HCNs consider two kinds of deployments [16]:

- **co-channel deployments**, where MeNBs and SeNBs use the same band (e.g. carrier frequency f_1 in Fig. 1.2), and

- **non-co-channel deployments**, where MeNBs and SeNBs use different bands (e.g. MeNBs work at carrier frequency f_1 and SeNBs operate at f_2, as shown in Fig. 1.3).

Each deployment has its own peculiarities.

---

[2]Through this Ph.D. dissertation, the term BSs is used to refer to general access points, encompassing MeNBs and SeNBs.

Co-channel deployments have to deal with a major impairment: the **cross-tier interference** [1], i.e. interference that is created between the macro-tier and the small-tier (e.g. interference from MeNBs towards UEs served by the SeNBs (denoted by smallcell UEs, SUEs) or interference from SeNBs towards UEs served by the MeNBs (denoted by macrocell UEs, MUEs)). Cross-tier interference can be very detrimental specially in the direction from MeNBs towards SUEs due to the difference in the transmitted power of the useful signal (from SeNB) and the interfering signal (from MeNB) (see Fig. 1.2).

In contrast, non-co-channel deployments eliminate the cross-tier interference due to the use of different bands among tiers. This comes at the cost of a possible reduction of the system spectral efficiency. Even though, they are interesting from the point of view of operators because MeNBs may keep on working as usual and traffic can be offloaded to the new underlying dense *smallcell network* (SCN) [20–22]. In non-co-channel deployments, **co-tier interference** between SeNBs can, however, become a major problem in concentrated clusters that serve hot-spot areas with high user data traffic demands [23] (see Fig. 1.3).

Traditionally, frequency division duplex (FDD) is adopted at MeNBs, i.e. DL and UL transmissions are duplexed in frequency domain. In FDD, the amount of band devoted for DL and UL is fixed and equally divided. In contrast, time division duplex (TDD), where DL and UL transmissions are duplexed in time domain, allows for asymmetric DL-UL allocations (i.e. asymmetric distribution of resources to DL and UL transmissions). Thus, TDD provides higher flexibility than FDD to adapt the system configuration to the *UL/DL traffic asymmetries* [24]. It is therefore more useful in scenarios where the amount of DL and UL traffic per cell varies and dynamic adaptation is required.

The advantage of a new band for the SCN in non-co-channel deployments is that SeNBs can operate in TDD so as to better match the UL/DL traffic asymmetries. Said traffic asymmetries arise in SCN because the number of users that are expected to be served per SeNB is reduced and hence the amount of DL and UL traffic per SeNB can vary over space and time more drastically in SCNs than in conventional macrocell-based networks [25]. In addition to the higher flexibility for UL/DL traffic asymmetries adaptation, TDD allows exploiting the *reciprocity of UL and DL propagation channels* to design advanced interference management techniques through the use of MIMO precoding. This allows improving the performance of the TDD SCN and reducing the channel feedback signaling needed as compared to the FDD case [24].

Anyway, both co-channel and non-co-channel deployments in MIMO dense HCNs have to deal with some kind of interference. Thus, advanced interference management techniques acquire a crucial role in future cellular systems.

1.2   Interference Management: Approaches and Domains

In a general multi-cell multi-user cellular system with full reuse of resources, two types of interference might appear (see Fig. 1.4):

- **intra-cell interference**: interference caused due to the simultaneous transmission (reception) of a BS towards (from) multiple users on the same time/frequency resource, and
- **inter-cell interference**: interference generated among different BSs and/or UEs associated to different BSs due to the simultaneous transmission (reception) on the same time/frequency resource.

(a) 2 BSs in DL   (b) 1 BS in DL and 1 BS in UL
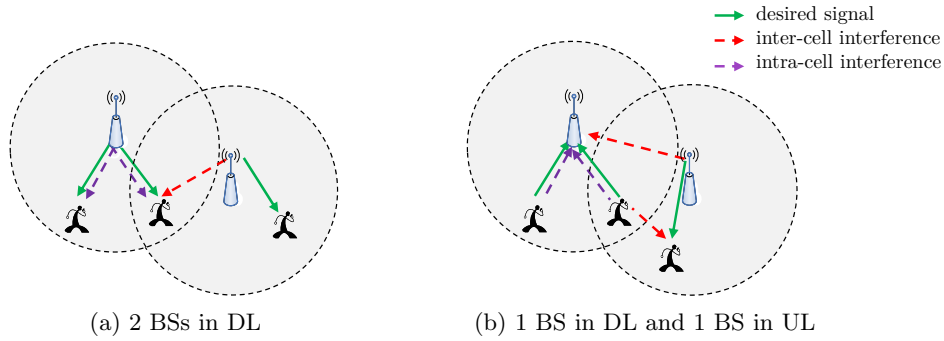
Figure 1.4: Inter-cell and intra-cell interference in multi-cell multi-user scenarios.

Remark 1.4. *The present work gives special attention to inter-cell interference management for dense BSs deployments, but intra-cell interference is also considered whenever within-cell multi-user transmissions are appointed.*

The focus on inter-cell interference management is due to the following reasons:

1. The upcoming network densification with heterogeneous BSs imposes that the main source of interference in future cellular systems is inter-cell interference. Two kinds of inter-cell interference arise in dense HCNs (as described in Section 1.1):

    - **cross-tier interference** and
    - **co-tier interference**.

2. The co-channel deployment in dense HCNs, where BSs employing different transmit powers coexist in the same band (see Fig. 1.2), introduces high-levels of inter-cell interference which predominate over intra-cell interference owing to the differences between the transmitted powers at MeNBs and SeNBs (e.g. high-level of inter-cell interference is received at SUEs from MeNBs, as MeNBs transmit with higher power than SeNBs).

3. A commonly adopted radio access technology in communication systems is orthogonal frequency division multiple access (**OFDMA**), which is employed in DL transmission of 3GPP LTE and LTE-A systems [2], in both UL and DL of IEEE 802.16m advanced WiMAX [26], and is one of the major candidates for different use cases of future 5G systems [27]. In OFDMA-based networks, the intra-cell users are assumed to be orthogonal to each other and the primary source of interference is inter-cell interference.

4. Under a flexible TDD system where each BS can decide its duplexing transmission pattern, i.e. switching instants between DL and UL transmissions, new types of inter-cell interference appear in the network (see Fig. 1.4.(b)) [19]:

    - **DL-to-UL interference**: interference created by a BS transmitting in DL towards another BS that is receiving in UL, and
    - **UL-to-DL interference**: interference created by a UE transmitting in UL towards another UE that is receiving in DL.

These inter-cell interference predominate over intra-cell interference due to the likely line-of-sight (LOS) conditions of neighboring BSs and because of the differences between the transmitted power at UEs and BSs. The most harmful interference is DL-to-UL interference, while UL-to-DL interference is statistically negligible (UEs should be in short distance to be interfered) but when it appears it can also seriously harm the transmission.

(a) Approaches   (b) Domains

Figure 1.5: Approaches and domains for interference management.

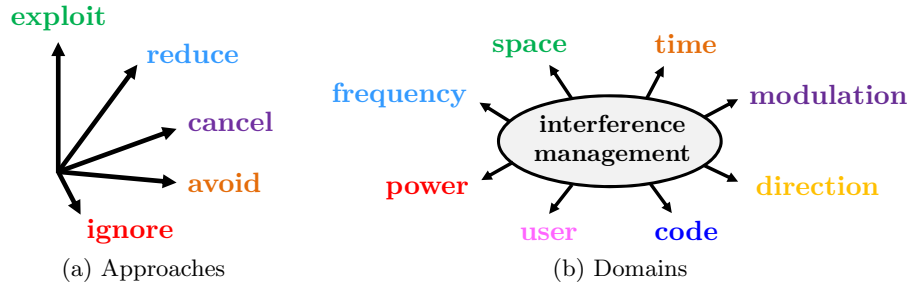In next sections we describe the approaches and domains for interference management (see Fig. 1.5)[3], as well as the mechanisms that allow them and the optimizations that can be employed. As BSs and users can act either as transmitters (TXs) or as receivers (RXs), depending if they are operating in DL or UL, we refer to general TXs/RXs.

1.2.1   Approaches

The approaches to treat interference can be classified in five groups (see Fig. 1.5.(a)):

- **Ignore**: generation/reception of interference is simply ignored. It can be adopted at TX and/or RX side. At TX side, no interference management is done (e.g. maximum-ratio transmission (MRT) beamformer [28]). At RX side, interference is simply treated as additive noise (e.g. maximum ratio combining (MRC) receiver [29]).

- **Avoid**: the TX avoids the generation of interference through the orthogonalization of the transmit resources. It is adopted at TX side and applies both to single-cell and multi-cell scenarios. It is useful when the RX is equipped with low-complexity receive filters that dispose of little or none interference rejection capabilities. For the single-cell multi-user case, resources are orthogonalized among different RXs to avoid intra-cell interference. Examples are time division multiple access (TDMA), frequency division multiple access (FDMA) and code division multiple access (CDMA), which use orthogonal time resources, frequency resources or codes, respectively. In the multi-cell case, orthogonalization of time/frequency resources among the different TXs is required to avoid inter-cell interference. So, TDMA/FDMA/CDMA among TXs do also apply. The distribution of resources among TXs can follow either static or dynamic assignments. Examples of static assignments are the frequency reuse-n (FR-n) scheme, where the total bandwidth is partitioned into n subsets and each one is assigned to one TX, or the fractional frequency reuse (FFR) scheme [30, 31], where the total bandwidth is partitioned such that (i) cell-edge users of neighbor TXs do not interfere with each other and (ii) interference received by (and created by) cell-interior users is reduced. In the dynamic assignment case, the distribution of resources among TXs is dynamically optimized [32]. Hence, it provides higher flexibility to adapt the distribution to the actual channel conditions, but coordination among TXs is required to determine the specific time/frequency resources to be used by each TX.

- **Cancel**[4]: the generated/received interference is canceled out, not through the use of different resources but through the exploitation of signal processing and multi-antenna

---

[3]This is a classification made by the author. Alternative classifications and definitions can be found elsewhere.
[4]Cancel approach is also known as 'reject' approach when referring to cancellation at the RX side.

techniques. It can be adopted at TX and/or RX side and applies both to single-cell and multi-cell scenarios. At TX side, the generated interference can be totally canceled through linear beamforming if there are enough antennas (e.g. zero-forcing (ZF) scheme, which allows setting zeros towards specific spatial directions [33,34]) or partially canceled through non-linear precoding (e.g. dirty paper coding (DPC) scheme [35] and Tomlinson-Harashima precoding (THP) [36], i.e. the practical and suboptimal DPC variant, where interference for some users are totally canceled out while others are not [37]). At RX side, the RX has capabilities to cancel out the received interference through two main approaches: direct rejection based on linear beamforming (e.g. ZF receiver or interference rejection combining (IRC) receiver [29]) or interference decoding followed by subtraction (e.g. successive interference cancellation (SIC) receiver [29]), which requires knowledge of codebook and modulations used by the interferer and achieves a partial cancellation (similarly as the DPC scheme for transmission). From a system level point of view, interference cancellation techniques require coordination among the different TXs/RXs and establish a clear trade-off between the degrees of freedom for interference cancellation and the available ones for useful signal transmission/reception. Therefore, two kinds of interference cancellation arise: total cancellation, where all the generated/received interference is canceled, or partial cancellation, where only a part of the generated/received interference is canceled while degrees of freedom are reserved for useful signal transmission/detection.

- **Reduce**: TXs coordinate their transmit strategy so as to reduce the level of received interference. It is adopted at TX side and applies to multi-cell scenarios. The TX does not avoid or cancel out interference but facilitates the rejection at the RXs. In this case it is important to maximize some system performance metric and hence optimize the interference distribution so as to minimize its impact. Examples of interference reduction are: power control (where the transmitted power is adjusted in a coordinated manner among TXs), coordinated scheduling and coordinated beamforming (where decisions on the user scheduling or transmit beamforming are done in a coordinated manner so as to maximize a system utility function and reduce interference) (known as coordinated multipoint coordinated scheduling/coordinated beamforming (CoMP CS/CB) in LTE-A) [38], signal-to-leakage-and-noise ratio (SLNR) precoding [39], Pareto optimal beamforming [40–42], and interference alignment (IA) [43], among others.

- **Exploit**: interference is converted into useful signal through the cooperation[5] (or joint transmission/reception) of TXs/RXs [44]. It can be adopted at TX and/or RX side and applies to multi-cell scenarios. When TXs perform a joint transmission to serve a user, they act as a single big TX with geographically separated antennas. Conversely, when RXs perform a joint reception to receive from a user, they act as a single big RX with geographically separated antennas. Therefore, this approach requires tight channel state information (CSI), synchronization control, as well as data sharing of useful data messages among the TXs/RXs that perform the joint transmission/reception to exploit interference. An example of transmit cooperation is network MIMO [45], which is also known as coordinated multi-point joint transmission/joint reception (CoMP-JT/JR) in LTE-A [38].

---

[5]Through this Ph.D. dissertation we make a clear distinction between transmit coordination and transmit cooperation. Transmit cooperation refers to a joint transmission (i.e. same data is transmitted from multiple BSs) while transmit coordination refers to any kind of coordination among BSs (at the control-plane but without data sharing) to adjust the BS transmit strategy. Thus, coordination can be understood as control-plane coordination among BSs while cooperation refers to control-plane plus data-plane coordination between BSs.

Table 1.1: SOME WELL-KNOWN TECHNIQUES FOR INTRA-CELL AND INTER-CELL INTERFERENCE MANAGEMENT.

Approach	Applies at	Intra-cell interference	Inter-cell interference
Ignore	TXs/RXs	MRT [28], MRC [29]	MRT [28], MRC [29]
Avoid	TXs	TDMA, FDMA, CDMA	TDMA, FDMA, CDMA, FR-n, FFR [30–32], ICIC [46]
Cancel	TXs/RXs	linear: ZF [33,34], IRC [29] non-linear: DPC [35,37], THP [36], SIC [29,37]	linear: multi-cell ZF [47], IRC [29] non-linear: multi-cell DPC [48,49], SIC [29,37]
Reduce	TXs	*not applicable*	CoMP-CS/CB [12,38,50,51], SLNR [39], Pareto optimal [40–42], IA [43,52–55], eICIC [56–58]
Exploit	TXs/RXs	*not applicable*	CoMP-JT/JR (or network-MIMO) [38,44,45,50,59,60]

Table 1.1 summarizes some existing techniques in the literature for intra-cell and inter-cell interference management, separately, classified according to the different approaches.

Depending on whether interference management approaches at TX sides and RX sides are combined or not, different levels of collaboration are obtained:

- *Isolated RX-side (or TX-side) interference management*: interference management approaches are only applied at RX sides (or TX sides).
- *Network-side collaboration*: interference management approaches at TX sides and RX sides are properly combined.

Ideally, interference management approaches should be properly combined so as to obtain a network-side collaboration, e.g. we should take into account that if a specific RX has capabilities for interference cancellation then there is no need to do so at the TX side.

1.2.2 Domains

There are multiple domains to tackle interference (see Fig. 1.5.(b)):

- **space**-domain (e.g. MIMO, beamforming),
- **frequency**-domain (e.g. frequency allocation),
- **time**-domain (e.g. time allocation),
- **user**-domain (e.g. scheduling of users),
- **power**-domain (e.g. transmit power control),
- **code**-domain (e.g. code assignation in CDMA-based systems),
- **modulation**-domain[6] (e.g. proper or improper Gaussian signaling), and
- **transmit direction**-domain (i.e. selecting transmission in DL or in UL).

In MIMO systems, the *space-domain* acquires a special role. Many leading concepts for interference management rely on it. Even though, it may be properly combined with the others.

---

[6]As modulation-domain we refer to the domain that allows treating in a different manner the signal and the complex conjugate of the signal. More details are given in Section 1.4.5.

Table 1.2: Mechanisms Required for Inter-Cell Interference Management in each Approach.

Approach	Mechanism
Ignore	-
Avoid	TXs coordination
Cancel	TXs/RXs coordination
Reduce	TXs coordination
Exploit	TXs/RXs cooperation

1.2.3   Mechanisms

In case of inter-cell interference management, there are two different mechanisms that allow the implementation of the approaches presented in Section 1.2.1:

- **Coordination**: exchange of control-plane messages between TXs/RXs. Coordination among TXs/RXs is required to perform the *avoid, cancel, reduce* and *exploit* approaches for inter-cell interference management.
- **Cooperation**: exchange of control-plane messages and data-plane information between TXs/RXs. Cooperation among TXs/RXs is needed to implement the *exploit* approach for inter-cell interference management.

Table 1.2 includes the mechanisms (coordination/cooperation) required for inter-cell interference management in each of the approaches.

1.2.4   Optimization time frame

The interference management approaches described in Section 1.2.1 can be applied by following two types of optimizations:

- **Short-term** optimization: instantaneous information of the network is available at TXs and/or RXs (like instantaneous channel conditions or instantaneous load conditions) and the optimization is performed and applied at the short-term.
- **Long-term** optimization: statistical information of the network is available at TXs and/or RXs (like channel statistics or load statistics) and the optimization is performed in the long-term[7], hence not matching the instantaneous network conditions.

Remark 1.5. *Among the previously presented approaches, domains, mechanisms, and optimizations, this doctoral thesis considers[8]:*

- cancel *approach at RXs,*
- avoid/reduce/exploit *approaches at TXs (specially, the* reduce *approach),*
- space/frequency/time/user/power/modulation/transmit direction-*domains (mostly, the* space *domain),*
- coordination/cooperation *mechanisms (with emphasis on* coordination*), and*
- short-term/long-term *optimizations (with special focus on the* short-term *optimization).*

---

[7]Long-term optimization involves optimization for several consecutive transmission time intervals (TTIs), rather than dynamically performing the optimization at each TTI.

[8]The specific chapters of this work in which they are applied are detailed in Table 1.3 of Section 1.7.

1.3   Interference Scenarios

The main scenarios considered for MIMO dense HCNs that are tackled in this Ph.D. dissertation are (see Fig. 1.6):

- single-cell MIMO (both single-user and multi-user),
- multi-cell MIMO (both single-user and multi-user), and
- cooperative multi-cell MIMO (multi-user).

Single-user and multi-user refer, respectively, to the cases in which each BS serves a single-user or has capabilities to serve multiple users simultaneously on the same time/frequency resource. In the multi-cell MIMO scenario, multiple BSs serve simultaneously one or multiple users and hence they might interfere each other (see red dashed lines in Fig. 1.6.(c)-(d)).

In the cooperative multi-cell MIMO scenario, some BSs are allowed to cooperate through a joint transmission (and hence *exploit* interference) while others generate interference among them. Two approaches for joint transmission arise depending on how the BSs cluster for transmit cooperation is performed (see Fig. 1.6.(e)-(f)): (*i*) BS-disjoint clustering, where clusters are formed from the network-side perspective and each cluster is formed by a disjoint group of BSs, and (*ii*) UE-centric clustering, where a cluster is defined per user and each BS can belong to multiple clusters (see more details in Section 1.4.2).

(a)   Single-cell single-user MIMO
(MIMO point-to-point with interference)

(b)   Particular 2-cell single-user MIMO
(MIMO Z-interference channel)

(c)   Multi-cell single-user MIMO
(K-user MIMO interference channel)

(d)   Multi-cell multi-user MIMO

(e)   Cooperative multi-cell multi-user MIMO
with BS-disjoint clustering for CoMP-JT

(f)   Cooperative multi-cell multi-user MIMO
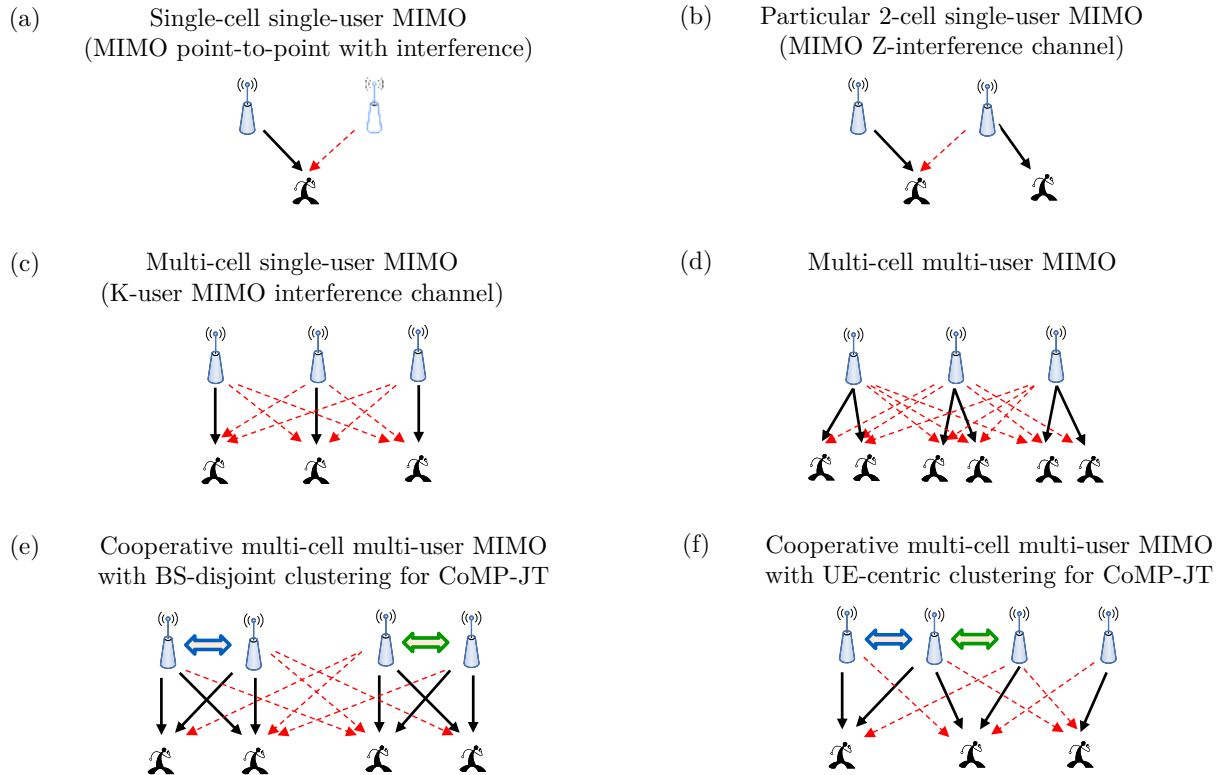with UE-centric clustering for CoMP-JT

Figure 1.6: Interference scenarios addressed in this Ph.D. dissertation.

Scenarios in Fig. 1.6 are useful to model dense HCNs and dense SCNs as follows:

- The particular 2-cell single-user MIMO (MIMO Z-interference channel) in Fig. 1.6.(b) models the major cross-tier interference generated from a MeNB (that serves its MUE) towards a SUE (served by the SeNB) in co-channel deployments of HCNs (see Fig. 1.2).
- The multi-cell single/multi-user MIMO scenarios in Fig. 1.6.(c)-(d) are useful to model dense SCNs in non-co-channel deployments (see Fig. 1.3), where significant interference arises in many spatial directions due to the dense deployment of SeNBs.
- The cooperative multi-cell multi-user MIMO scenario with BS-disjoint clustering in Fig. 1.6.(e) can be adopted in regular network deployments formed by MeNBs, where static clusters of MeNBs can be used for joint transmission.
- The cooperative multi-cell multi-user MIMO scenario with UE-centric clustering in Fig. 1.6.(f) can be employed in irregular network deployments, as the dense SCNs, where the cluster formation is not trivial and having a cluster per user might be more convenient.

Let us finally recall that under flexible duplexing, the multi-cell single/multi-user MIMO scenarios in Fig. 1.6.(c)-(d) apply but the role of TXs and RXs can correspond either to BSs or to users depending on the transmit direction (i.e. DL or UL) that is selected at BSs.

1.4   Interference Management Techniques in 3GPP

Most advanced cellular network standardization, i.e. 3GPP LTE-A [2], considers four main techniques to deal with interference in MIMO dense HCNs:

- enhanced inter-cell interference coordination (eICIC) [61],
- coordinated multi-point transmission (CoMP) [62],
- network assisted interference cancellation and suppression (NAICS) [29], and
- enhanced DL-UL interference management and traffic adaptation (eIMTA) [63].

1.4.1   eICIC

eICIC is the main technique considered in 3GPP to improve the spectral efficiency in DL transmissions (BSs to UEs) in co-channel deployments of HCNs (see Fig. 1.2). eICIC is based on the *avoid* approach and *time/user-domain* coordination. It encompasses: (*i*) almost blank subframes (ABS), subframes during which the MeNBs remain silent in the data-plane to avoid the generation of cross-tier interference, and (*ii*) biased user association, in which a bias is added over the reference signals measurements used for the user-BS association process so as to allow offloading more users to the SeNBs [56, 57].

3GPP has analyzed an improvement of eICIC, namely further eICIC (FeICIC), in order to deal with interference not only at the data-plane level but also at the control-plane level. FeICIC considers interference cancellation capabilities at UEs, and muting and reduced power control at BSs for interference coordination (so *cancel/reduce* approaches and *power/space-domain* coordination were added). However, although performance improvements are obtained with eICIC/FeICIC based on ABS, the implicit muting of the MeNB in some sub-frames can lead to a spectral efficiency loss.

Remark 1.6. *This Ph.D. dissertation uses eICIC technique for comparison purposes in co-channel HCN deployments.*

1.4.2 CoMP

CoMP is a potential technique for interference management that is based on the *reduce/exploit* approaches and *space/power/user-domain* coordination. CoMP allows dealing with inter-cell interference through the coordination of BSs [38, 50] and is proposed both for non-co-channel and co-channel deployments in HCNs. 3GPP LTE-A mainly distinguishes between the following categories of CoMP for DL transmission (i.e. BSs to UEs) [62]:

- *CoMP coordinated scheduling-coordinated beamforming* (CoMP-CS/CB): user data is only available in one BS, the so-called serving BS, but user scheduling and beamforming decisions are dynamically coordinated among the BSs in order to control and reduce the interference level between different transmissions (*reduce* approach).
- *CoMP dynamic point selection* (CoMP-DPS): user data is available at multiple BSs in the network but the serving cell is dynamically selected among them, such that the transmission is done from one BS at a time (*reduce* approach).
- *CoMP joint transmission* (CoMP-JT): user data is simultaneously transmitted from multiple BSs in the network, which cooperate in order to work as a single transmitter with geographically separated transmit antennas. CoMP-JT has the benefit that it resolves the interference created by the strongest interferers (*exploit* approach).

In the UL transmission (i.e. UEs to BSs), two categories of CoMP are considered [62]:

- *CoMP coordinated scheduling* (CoMP-CS): user scheduling and precoding selection decisions are coordinated among the BSs that act as RXs (*reduce* approach).
- *CoMP joint reception* (CoMP-JR): multiple UEs transmit simultaneously to multiple BSs and a joint reception is performed at BSs (*exploit* approach).

Remark 1.7. *This doctoral thesis investigates advanced techniques for CoMP-CS/CB (*re-duce *approach) (both for DL and UL) and CoMP-JT (*exploit *approach) (for DL).*

CoMP-CS/CB

In CoMP-CS/CB, multiple geographically separated BSs coordinate their scheduling and/or beamforming (i.e. precoding) decisions so as to maximize a global system utility function. The BSs transmissions are coupled by interference and the system can be modeled by the general **multi-cell multi-user MIMO** system (see Fig. 1.6.(d)): a system where multiple BSs, each equipped with multiple antennas, wish to simultaneously send independent messages to its served UEs while generating interference to unintended UEs.

CoMP-JT

In CoMP-JT multiple geographically separated BSs jointly transmit data to each user, such that data has to be shared among the cooperative BSs. Although large theoretical capacity gains are obtained, in practical implementations the gains of CoMP-JT are saturated with the number of cooperating BSs due to the overhead required to acquire knowledge of the channel matrices, the use of non-ideal backhaul links, and the impact of channel estimation errors [64]. For that reason, the number of BSs in which CoMP-JT can take place (called the *cluster size* or CoMP set) has to be limited. The setup can be modeled by the **cooperative multi-cell multi-user MIMO** system shown in Fig. 1.6.(e)-(f) where, due to the limited number of BSs cooperating

(a) BS-disjoint clustering   (b) UE-centric clustering
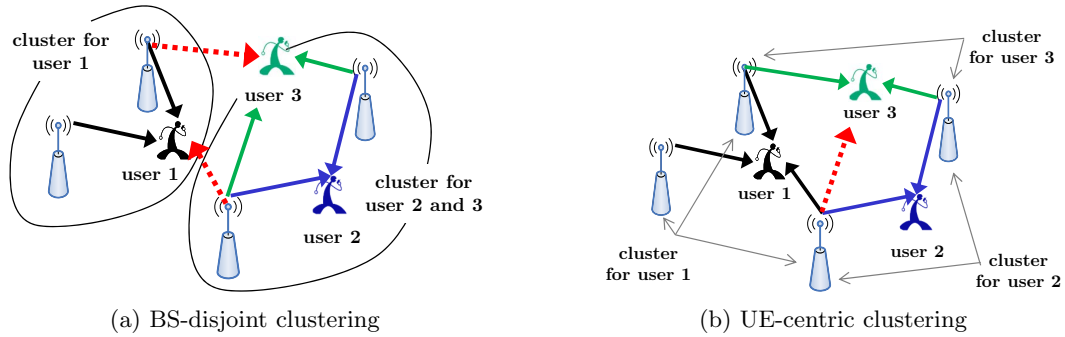
Figure 1.7: BSs clustering schemes for CoMP-JT.

for CoMP-JT and the multi-user scenario, interference is still present. In LTE-A the cluster size is 3, being suitable for classical regular deployments [22]. However, in a dense and irregular deployments of smallcells, the best BSs clustering and cluster size selection for CoMP-JT is a problem still under discussion.

There are mainly two types of BSs clustering schemes for CoMP-JT (as shown in Fig. 1.7):

- **BS-disjoint clustering**: non-overlapping clusters of BSs are formed according to the deployment and BSs in each cluster jointly serve all users within their coverage area. This way, users at the cluster-edge still suffer from considerable inter-cluster interference (see red dashed lines in Fig. 1.7.(a)) and other kinds of coordination might be needed. Although it is useful for regular and classical BS deployments, it might not be the most suitable scheme from the perspective of users (cluster-edge users still appear) and it might not be adequate for dense and irregular deployments of BSs.

- **UE-centric clustering**: a cluster of BSs is selected per user. This way, clusters associated to different users might overlap as they may share some BSs (see Fig. 1.7.(b)). In this case, the concept of cluster-edge users disappears but not the interference in the network. It is more adequate than the BS-disjoint clustering scheme from the perspective of the users and also for dense and irregular deployment of BSs.

1.4.3   NAICS

NAICS enables more effective and robust RX-side interference cancellation and/or suppression with possible network coordination. NAICS receivers are based on the *cancel* approach and *space-domain* coordination. They are classified as [29]:

- *Interference suppression receivers* (e.g. linear minimum mean square error - interference rejection combining (LMMSE-IRC), enhanced LMMSE-IRC, widely LMMSE-IRC),
- *Maximum likelihood receivers* (ML) (e.g. ML, reduced complexity ML),
- *Interference cancellation receivers* (e.g. linear code word level successive interference cancellation, symbol level interference cancellation, parallel interference cancellation).

Remark 1.8. *Through the present work, we usually assume that NAICS with interference suppression is applied at receivers. Transmit coordination strategies are derived by taking such assumption into consideration. Therefore, a* network-side collaboration *(where interference management approaches at TX sides and RX sides are combined) is obtained as a result.*

1.4.4 eIMTA

eIMTA develops mechanisms for dynamic allocation of DL and UL subframes in LTE TDD so as to match the instantaneous data traffic situation [63]. eIMTA is based on the *avoid/cancel/reduce* approaches and *time/user/power/transmit direction-domain* coordination. Four different potential interference management schemes have been investigated for eIMTA:

- *Cell clustering interference mitigation*: BSs are grouped into clusters and all BSs within each cluster use the same transmission pattern (i.e. all use either UL or DL in any subframe), hence DL-to-UL and UL-to-DL interference are avoided within each cluster.

- *Scheduling-dependent interference mitigation*: each BS adjusts the scheduling strategies (resource allocation, transmit power, transmit direction) considering the DL and UL channel quality, the DL-to-UL and UL-to-DL interference, and the traffic load.

- *Interference mitigation based on eICIC/FeICIC*: the interference mitigation schemes and procedures from eICIC/FeICIC are reused to TDD UL-DL reconfiguration based on traffic adaptation.

- *Interference suppressing interference mitigation* (ISIM): ISIM is considered for UL transmission of either smallcells or macrocells hence allowing suppression of one or more of the dominant DL-to-UL interfering signals, e.g. by LMMSE-IRC.

Remark 1.9. *This Ph.D. dissertation focuses on eIMTA with scheduling-dependent interference mitigation to develop traffic-aware and interference-aware resource management techniques for flexible duplexing systems.*

1.4.5 IGS

In addition to the previous mentioned techniques, we also consider the optimization of the *modulation*-domain or, more explicitly, the use of **improper Gaussian signaling** (IGS). IGS differs from the conventional proper Gaussian signaling (PGS) in the characterization of the second order statistics: real and imaginary parts of IGS are correlated or have unequal power, while those of PGS are uncorrelated with equal power. Therefore, IGS adds more flexibility to the transmission since it allows exploiting the real and imaginary dimensions of the MIMO channel more effectively. IGS arises naturally in communications, e.g. due to the gain imbalance between the in-phase and in-quadrature components or due to the use of specific digital modulations that exhibit improper characteristics. In this sense, IGS has also been analyzed in 3GPP LTE-A from the receiver perspective when improper constellations are employed, see [22].

In order to generate and estimate IGS with linear processing, widely linear precoding (WLP) and widely linear estimation (WLE) [65] need to be adopted at transmitter and receiver sides, respectively, instead of the commonly used linear precoding (LP) and linear estimation (LE) techniques that maintain the proper Gaussian distribution. WLE is a generalized concept used in estimation theory whenever improper noise is encountered [66] and in systems that transmit improper signal constellations [67]. On the other hand, WLP can be used to generate IGS even when departing from proper signal constellations [68,69], but there are no studies analyzing the benefits of coordinated WLP designs. This is studied in the present work.

Remark 1.10. *In this doctoral thesis we develop advanced IGS schemes, derive new insights on the use of IGS, and address the coexistence of linear and widely linear transceivers in the network.*

1.5   Relevant Challenges related to Reduce/Exploit Approaches

All interference management approaches presented in Section 1.2.1 (*avoid/cancel/reduce/exploit*) require some kind of coordination among BSs (which can either act as TXs or RXs). In addition, *exploit* approach (e.g. CoMP-JT and CoMP-JR in LTE-A) needs cooperation among BSs. Among them, *reduce* and *exploit* approaches allow to significantly enhance the system spectral efficiency and the cell-edge throughput from a theoretical point of view [50]. But, they impose additional signaling overhead related to the coordination/cooperation mechanisms (see Section 1.2.3). Therefore, new requirements are imposed to the backhaul network, especially in terms of latency, capacity, and synchronization precision between the coordinated/cooperative BSs [51]. These requirements are more stringent in the *exploit* approach.

From a practical point of view, in multi-cell environments under full reuse of resources, the main challenges related to the implementation of short-term interference management techniques based on *reduce/exploit* approaches are [50]:

- channel knowledge,
- synchronization,
- backhaul,
- clustering,
- centralized versus decentralized architecture.

1.5.1   Channel knowledge

Acquisition of channel knowledge (i.e. the channel matrices in MIMO systems) - referring to both desired channels and the interfering channels - at the BSs is usually required to derive interference management techniques based on *reduce/exploit* approaches.

Usually, in FDD systems, the channel is estimated at the UEs and then fed back to the BSs. In this sense, standard channel estimation and feedback concepts can principally be extended to estimate the channels observed at the UEs from multiple BSs and thus enable BSs coordination/cooperation [50]. However, estimating the channel matrices towards a high number of BSs is questionable in practice due to the fact that weak links cannot be estimated accurately and the involved pilot planning and feedback overhead may become prohibitive [64].

An alternative approach to acquire channel knowledge at BSs is through the exploitation of propagation channel reciprocity in TDD systems [24]: as DL and UL channels are reciprocal within the channel coherence time, the BSs can directly acquire channel knowledge from an UL pilot-based transmission. In this case, as compared to FDD, the overhead required to feed back the estimated channel at UEs and the effect of imperfect feedback links is avoided. Nevertheless, perfect propagation channel reciprocity requires perfect hardware calibration of the DL/UL radio-frequency (RF) chains at BSs and UEs [70, 71]. Effective channel reciprocity calibration techniques have been proposed and analyzed in the literature [72, 73]. Although perfect calibration can be achieved in practice at BSs (because the variability of surrounding scenario is relatively slow and hardware performances tend to be stable), calibration at UEs may be imperfect due to environmental variations caused by time, power and/or temperature [74]. Hence, non-ideal calibration might result in channel estimation errors.

In practice, channel knowledge at TXs in dense networks implies:

- additional computational cost associated to the estimation of multiple channels,
- network planning required for pilot signals,
- performance loss due to imperfect estimation of the interfering channel matrices that are estimated with a low SNR,
- performance loss due to imperfect feedback links (in case of FDD systems),
- additional overhead required to feed back the estimated channels (in case of FDD systems),
- performance loss due to imperfect calibration of the RF chains (in case of TDD systems).

To avoid performance losses due to imperfect channel knowledge (that can come due to imperfect estimation, imperfect feedback in FDD, and imperfect calibration in TDD), robust designs that consider statistical channel knowledge and statistical error characterization have been analyzed in the literature [75, 76].

Remark 1.11. *This work goes a step further than existing interference management techniques to overcome channel acquisition problems in practical implementations: we derive transmit coordination strategies that do not require estimation of the interfering channel matrices while they are still able to manage interference effectively (i.e. as a strategy with knowledge of all channel matrices could do). Thus, the impairments/requirements listed above are reduced.*

1.5.2 Synchronization

In multi-cell environments under full resource reuse, synchronization of coordinated/cooperative BSs and coordinately/cooperatively served UEs in time and frequency is required. In particular, strict requirements in clock synchronization, as well as compensation of signal timing offsets [77] and carrier frequency offsets [78] are indispensable (e.g. signals from different BSs should arrive at each UE synchronized in time and frequency) [79]. For cooperation, time synchronization at the symbol level is required, while for coordination, time synchronization at the frame level is needed. In BC (point to multi-point) and MAC (multi-point to point), time synchronization may be readily achieved with the aid of traditional time synchronization techniques, which use estimated signal time offsets to pre-compensate or post-compensate possible desynchronizations at BSs and UEs, respectively. These techniques do not readily apply to systems comprised of multiple BSs that transmit simultaneously to multiple UEs in the same resource. In this sense, pre-compensation and post-compensation algorithms have been proposed in the literature to synchronize the signals in a multi-point to multi-point communication links [80].

Remark 1.12. *Through this Ph.D. dissertation, we assume that signals are synchronized in time and frequency, which can be achieved with existing pre-compensation and post-compensation techniques (e.g. [80]). However, the impact of imperfect synchronization as well as the non-feasibility of synchronization in dense networks are topics of current interest.*

1.5.3 Backhaul

Depending on the existing infrastructure of a mobile operator, both backhaul capacity and latency requirements of CoMP-based techniques may be the main cost drivers or potential stoppers of its theoretical capacity gains [6]. For that reason, 3GPP devoted a full study item [81]

to analyze the effect of non-ideal backhaul (NIB) links for CoMP, in which results of the present work were included. The study item aimed at evaluating the performance benefits and identifying potential standardization impacts for candidate CoMP techniques involving multiple BSs with NIB. It was observed that the gain of CoMP with NIB varies according to the deployment scenario, backhaul delay, coordination scheme, resource utilization factor, and coordination size. The main limiting factor was the backhaul delay: for 5 ms backhaul delay, gains of CoMP were still observed and significant, while for 50 ms backhaul delay, the implementation of CoMP techniques was detrimental for the system performance.

Remark 1.13. *This Ph.D. dissertation overcomes backhaul requirements problems: we derive decentralized transmit coordination strategies where all the information required for interference management is obtained through the air-interface thus avoiding the information excess to be exchanged among BSs through the backhaul.*

Remark 1.14. *High-speed backhaul links are crucial to implement transmit cooperation (i.e. CoMP-JT with* exploit *approach) but, when it is adopted in this thesis, the cooperative cluster size is limited to a small number of BSs.*

1.5.4   Clustering

In case of the *exploit* approach (i.e. CoMP-JT), due to the large amount of required signaling overhead and high-speed backhaul, only a limited number of BSs can cooperate in order to keep the overhead manageable. This raises the question of which BSs should form cooperative clusters in order to exploit the advantages of transmit cooperation at limited complexity. The clusters can be either static or dynamic, conformed by disjoint BSs or overlapping BSs, formed from the point of view of either the BS deployment (BS-centric) or the specific UEs to be served (UE-centric). An overview is presented in [50]. In the literature, mainly two kinds of BSs clustering schemes are adopted: BS-disjoint clustering (static, conformed by disjoint BSs, and BS-centric) [59] and UE-centric clustering (dynamic, conformed by possibly overlapping BSs, and UE-centric) [60, 82] (see Section 1.4.2).

Remark 1.15. *When the* exploit *approach is investigated in this work, we adopt a UE-centric clustering scheme and derive procedures that jointly optimize the cluster formation and the precoding design by setting a maximum cluster size per user (which could be imposed, e.g., by the mobile network infrastructure based on the user position).*

1.5.5   Centralized versus Decentralized Architecture

Interference management can be performed either in a centralized or decentralized manner, depending on two distinct architectures.

In the centralized architecture, a network controller placed in the core network is responsible for some network functions [83]. Generally, the network controller has to: *i*) gather CSI of all users both in DL and UL and *ii*) perform the interference/resource management through the optimization of a global objective function. The centralized-based concept is being considered as a promising solution for 5G systems. However, as the number of BSs increases so does the computation required for centralized management. Therefore, parallel management techniques are still preferred to eliminate high computation loads in centralized architectures [84]. Anyway,

the main drawback of the centralized interference/resource management techniques in dense networks is the added cost of a network controller required to collect the CSI of all users and the scalability of the solution.

An alternative solution for dense networks is the decentralized architecture. In this case, each BS has its own baseband unit and there is no need for a network controller. However, so as to being able to optimize a global objective function in a parallelized manner among BSs, BSs have to exchange information (e.g. the CSI of some users). Decentralized interference/resource management has the advantages of: scalability, robustness against failures, the inherent parallelization related to decentralized implementation (and hence the associated distribution of the computational load), and the lower backhaul requirements. Hence, it might be more suitable for dense and irregular networks.

Remark 1.16. *Through this doctoral thesis, both centralized and decentralized architectures are considered. Special attention is given to the decentralized case due to its suitable application to dense networks.*

1.6 Mathematical Tools and Knowledge

The completion of this thesis has required knowledge of some important mathematical tools. Also, knowledge of standardization and patent filing aspects has been needed.

The mathematical tools that are used through the Ph.D. dissertation are:

- **Matrix algebra**: Matrix algebra allows operating with matrices, including matrix functions, matrix decompositions and factorizations, matrix structures and matrix equations. It is a tool constantly used throughout this work because matrix variables naturally appear due to the adoption of MIMO systems. A complete reference for matrix algebra can be found in [85]. Additional material is available at [86, 87].

- **Matrix calculus**: Matrix calculus refers to the tool needed to perform matrix differentials and matrix derivatives. It is constantly used in the present work because matrix derivatives are needed to solve (in closed-form) optimization problems that involve matrix variables. A complete reference for complex-valued matrices is found in [88], which extends the matrix derivatives for real-valued matrices that are introduced in [89]. An important summary that is key for developing this Ph.D. dissertation is presented in [90], where derivatives of complex-valued matrix functions with respect to a complex-valued matrix variable and the complex conjugate of this variable are reviewed.

- **Constrained optimization theory**: Most problems of practical interest can be appropriately formulated as constrained optimization problems, whereby a function is minimized subject to either one or multiple constraints. Constrained optimization theory refers to the framework that allows solving these king of problems. In some cases, possibly after some mathematical manipulations, the problems can be expressed in convex form and thus they can be optimally solved through very efficient methods [91]. A complete reference for convex optimization theory can be found in [92]. In other cases, it would be of interest to decompose the problems through decomposition methods to achieve decentralized and/or parallelized solutions that work iteratively. See overviews on how to decompose a constrained optimization problem into separable parts in [93–95].

- **Improper Gaussian random processes**: Improper Gaussian signals have special properties and second-order statistics. In wireless communications, the statistical properties of the signals are very important because they might impact on the system performance, and therefore a full understanding of these signals might be profitable. A complete characterization of improper Gaussian signals, as well as the tools needed for estimating and detecting them, can be found in [65]. Also, a new line of research has been recently proposed to generate improper Gaussian signals and exploit their benefits in interference scenarios [68]. An overview is presented in Appendix 2.A.

- **Majorization theory**: Majorization makes precise the vague notion that the components of a real-valued vector **y** are "more spread out" or "less nearly equal" than are the components of another real-valued vector **x**. In this sense, majorization theory provides a general framework that allows comparing different solutions and solving complicated non-convex constrained optimization problems. A complete reference for majorization theory can be found in [96]. Also, interesting applications in wireless communications are analyzed in [97] and [98]. An overview is provided in Appendix 2.B.

- **Graph theory**: Graph theory is a wide area of discrete mathematics that deals with graphs (i.e. a mathematical structure consisting of a finite set of vertices and a finite set of edges that connect the vertices). In includes a lot of different graph structures, analysis of graph properties and graph coloring (i.e. assignation of colors to either the vertices or the edges of a graph). To complete this Ph.D. dissertation, knowledge of graph properties, recognizing special graph structures [99,100] and graph coloring algorithms [101] have been needed. A complete reference for graph theory can be found in [102]. An overview of the graph theory topics that are needed for this thesis is given in Appendix 7.A.

In addition to the mathematical tools, the following aspects have been needed to exploit the results:

- **3GPP standardization**: knowledge required to submit contributions to 3GPP LTE-A standard and to send contributions to specific technical reports.

- **Patent filing**: knowledge required to write and file a patent and to follow the associated processes.

1.7   Organization

Figure 1.8 shows the organization of this document. The technical contributions of this thesis are included in Chapters 2-7, which are structured in three blocks:

- development of **advanced signaling schemes** (Chapter 2),
- design of **transmit coordination strategies** (Chapters 3-4-5), including transmit coordination and transmit cooperation as well as signaling coexistence, and
- analysis of **traffic-aware duplexing techniques** (Chapters 6-7).

In the advanced signaling block, i.e. **Chapter 2**, we study advanced signaling schemes (in particular, the use of IGS) and analyze its potential benefits for MIMO interference-limited scenarios (including co-channel HCNs). The employed approaches and optimized domains are: *reduce* approach at TXs, *cancel* approach at RXs, *space* domain (through MIMO precoding design) and *modulation* domain (though the use of PGS/IGS). Chapter 2 serves as an input for Chapter 5, where the coexistence of different signaling schemes is investigated.
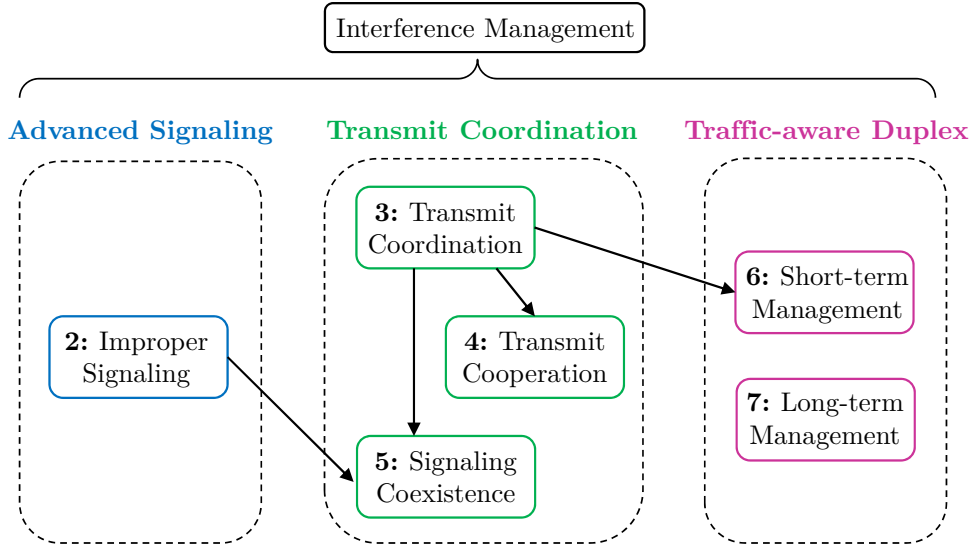
Figure 1.8: Organization of this Ph.D. dissertation.

In the transmit coordination block, the thesis develops transmit coordination and transmit cooperation strategies for interference management.

First, **Chapter 3** analyzes transmit coordination for multi-cell multi-user MIMO scenarios. It presents decentralized coordinated precoding designs assuming either perfect or imperfect channel state information, and analyzes its application to dense SCNs. The adopted approaches and optimized domains are: *reduce* approach at TXs (BSs), *cancel* approach at RXs (UEs), *space/power/user* domains (through MIMO precoding design). Chapter 3 serves as an input for Chapters 4, 5, and 6, which do all exploit the interference-cost concept for interference management presented in Chapter 3.

Second, **Chapter 4** investigates the joint implementation of transmit coordination and transmit cooperation for cooperative multi-cell multi-user MIMO scenarios. Cooperation among a limited number of BSs (which form a cluster) is assumed and coordination between clusters is adopted (for which the concepts presented in Chapter 3 are used). In this sense, the joint design of decentralized coordinated precoding and BSs clustering is analyzed by following a UE-centric clustering scheme (see Fig. (1.7)). The proposed technique can be easily adopted when using a BS-centric clustering scheme. The used approaches and optimized domains are: *reduce/exploit* approaches at TXs (BSs), *cancel* approach at RXs (UEs), *space/power/user* domains (through MIMO precoding design).

Finally, **Chapter 5** develops transmit coordination strategies to address the coexistence of linear and widely linear transceivers in multi-cell MIMO scenarios. The heterogeneous transceiver deployment arises due to the use of different signaling schemes: PGS and IGS, which are analyzed in Chapter 2. In this sense, Chapter 5 derives coordinated precoding designs and generalizes the results in Chapter 3 for an heterogeneous transceiver deployment. The addressed approaches and optimized domains are: *reduce* approach at TXs, *cancel* approach at RXs, *space/power/user* domains (through MIMO precoding design) and *modulation* domain (through the use of PGS/IGS).

In the traffic-aware duplex block, advanced traffic-aware and interference-aware resource management techniques are investigated for flexible duplexing in TDD multi-cell MIMO scenarios. We aim at making a better use of the available time/frequency resources by taking into account the DL/UL per-BS data traffic asymmetries that arise in dense networks.

Short-term resource management is included in **Chapter 6**, where advanced 5G system architectures and frame structures are assumed to implement flexible TDD. Algorithms for the joint optimization of user scheduling, transmit direction selection, and precoding design are developed assuming instantaneous knowledge of the channel conditions and average knowledge of the DL/UL traffic asymmetry conditions. The adopted approaches and optimized domains are: *reduce* approach at TXs, *cancel* approach at RXs, *space/power/user /transmit direction* domains (through MIMO precoding design) and *time* domain.

In contrast, **Chapter 7** analyzes long-term resource management, where the distribution of frequency resources among BSs is done in the long-term hence being independent of the specific users to be served but dynamic enough to follow significant variations on the DL/UL traffic load conditions per-BS. Differently from previous chapters, the proposed techniques focus on the *avoid* approach at TXs and the optimized domains are *time/frequency/transmit direction*.

Let us recall that Chapters 2 and 5 refer to general TXs/RXs. Chapters 3 and 4 focus on the DL transmission, and hence TXs are BSs and RXs are UEs. Finally, in Chapters 6 and 7 the role of TXs and RXs can be either adopted at BSs or at UEs owing to the flexible duplexing.

For each chapter of the present document where technical contributions are reported, Table 1.3 details the approaches (at TXs and at RXs), mechanisms, optimization, and domains for interference management (see detailed description in Section 1.2) that are considered therein.

Table 1.3: Approaches, Mechanisms, Optimization, and Domains for Interference Management Adopted on each Chapter of the Ph.D. Dissertation.

Chapter	TX app.	RX app.	Mechanism	Optimiz.	Domain
2: Improper Signaling	reduce	cancel	coordination	short-term	space, modulation
3: Transmit Coordination	reduce	cancel	coordination	short-term	space, power, user
4: Transmit Cooperation	reduce, exploit	cancel	coordination, cooperation	short-term	space, power, user
5: Signaling Coexistence	reduce	cancel	coordination	short-term	space, power, user, modulation
6: Short-term Management	reduce	cancel	coordination	short-term	space, time, user, power, direction
7: Long-term Management	avoid	-	coordination	long-term	frequency, time, direction

Table 1.4 contains the challenges of interference management (see Section 1.5) that are addressed on each technical chapter. The architecture considered on each one is also detailed. Let us remark that: *i*) challenges described in Section 1.5 mainly apply for the *reduce/exploit* approaches (i.e. they are not addressed in Chapter 7), *ii*) synchronization is a requirement for all the techniques presented in this doctoral thesis (hence not included in Table 1.4), and *iii*) clustering challenge only applies for the *exploit* approach (i.e. Chapter 4).

Table 1.4: Challenges of Interference Management Addressed on each Chapter of the Ph.D. Dissertation.

Chapter	Channel knowledge	Backhaul	Clustering	Architecture
2: Improper Signaling	√	√		decentralized
3: Transmit Coordination	√	√		decentralized
4: Transmit Cooperation	√	√	√	decentralized
5: Signaling Coexistence	√	√		decentralized
6: Short-term Management				centralized
7: Long-term Management				centralized

Table 1.5: Mathematical Tools Required for each Chapter of the Ph.D. Dissertation.

Chapter	Matrix algebra	Matrix calculus	Optimization	Improper signals	Majorization	Graph theory
2: Improper Signaling	√	√	√	√	√	
3: Transmit Coordination	√	√	√			
4: Transmit Cooperation	√	√	√			
5: Signaling Coexistence	√	√	√	√		
6: Short-term Management	√	√	√			
7: Long-term Management			√			√

Table 1.6: Research Contributions Produced from each Chapter of the Ph.D. Dissertation.

Chapter	Journal	Confer.	Patent	WG/TR	Project
2: Improper Signaling	1	1	-	-/-	Tropic D312
3: Transmit Coordination	1	1	1	7/1	Tropic D322 and D61
4: Transmit Cooperation	-	2	-	-/-	Tropic D322
5: Signaling Coexistence	1	1	-	-/-	Tropic D322
6: Short-term Management	1	1	-	-/-	Huawei D13
7: Long-term Management	-	1	-	-/-	Huawei D13

Table 1.5 specifies the mathematical tools (see Section 1.6) that have been required to derive the contributions on each technical chapter: matrix algebra, matrix calculus (matrix differentiation), constrained optimization theory, improper Gaussian random vectors, majorization theory, and graph theory.

Table 1.6 summarizes the research contributions that have been produced from each technical chapter in terms of publication of journal articles, publication of conference papers, filed patents, contributions to 3GPP working groups (WG) and 3GPP technical reports (TR), and contributions to projects. The specific details can be found in Section 1.8.

Finally, let us note that all the proposed techniques are evaluated in a Matlab® [103] simulator compliant with 3GPP LTE-A evaluation methodology specifications [104]. They are compared with the already existing solutions in the literature and with the techniques available in 3GPP documents.

1.8   Research Contributions

1.8.1   Publication of journal/conference papers

The research contributions to journals and conferences that have been derived from each technical chapter of this Ph.D. dissertation are detailed in the following.

Chapter 2

The main results in this chapter are regarding the demonstration of the strict superiority of IGS over conventional PGS and its application to MIMO interference scenarios. They have been published in one journal paper and one conference paper:

[J1]   **S. Lagen**, A. Agustin, J. Vidal, "On the Superiority of Improper Gaussian Signaling in Wireless Interference MIMO scenarios", *IEEE Trans. on Commun.*, vol. 64, no. 8, pp. 3350-3368, Aug. 2016. doi: 10.1109/TCOMM.2016.2584601.

[C1]   **S. Lagen**, A. Agustin, J. Vidal, "Improper Gaussian signaling for the Z-Interference Channel", *IEEE Int. Conf. on Acoustics, Speech and Signal Processing* (ICASSP), Florence (Italy), May 2014. doi: 10.1109/ICASSP.2014.6853775.

Chapter 3

The main results in this chapter involve the derivation of decentralized coordinated precoding strategies for interference management in dense multi-cell multi-user scenarios and the analysis of its practical implementation in 3GPP-based networks. They have produced one patent application, several contributions to 3GPP LTE, one journal paper and one conference paper:

[J2]   **S. Lagen**, A. Agustin, J. Vidal, "Decentralized Coordinated Precoding for Dense TDD Small Cell Networks", *IEEE Trans. on Wireless Commun.*, vol. 14, no. 8, pp. 4546 - 4561, Aug. 2015. doi: 10.1109/TWC.2015.2422704.

[C2]   **S. Lagen**, A. Agustin, J. Vidal, "Decentralized Beamforming with Coordinated Sounding for Inter-Cell Interference Management", *European Wireless 2014*, Barcelona (Spain), May 2014.

Chapter 4

The main results in this chapter refer to the development of decentralized techniques for interference coordination among clusters of BSs where transmit cooperation (i.e. CoMP-JT) is applied. As a result, two conference papers have been obtained (being one related to the BS-disjoint clustering and the other to the UE-centric clustering for CoMP-JT):

[C3]   **S. Lagen**, A. Agustin, J. Vidal, "Distributed Inter-Cluster Interference Management for CoMP-based cellular networks", *IEEE Global Commun. Conf.* (GLOBECOM), Atlanta (USA), Dec. 2013. doi: 10.1109/GLOCOM.2013.6831733.

[C4]   **S. Lagen**, A. Agustin, J. Vidal, B. Soret, K. I. Pedersen, "Distributed User-Centric Clustering and Precoding Design for CoMP Joint Transmission", *IEEE Global Commun. Conf.* (GLOBECOM), San Diego (USA), Dec. 2015. doi: 10.1109/GLOCOM.2015.7417090.

Chapter 5

The main results in this chapter include the design of transmit coordination strategies to address the coexistence of linear and widely linear transceivers (which correspond to different signaling schemes) in multi-cell scenarios. One journal paper, which addresses the coexistence of WLP and LP transmitters for sum-rate maximization, and one conference paper, which deals with the coexistence of WLE and LE receivers for sum mean square error (MSE) minimization, have been published:

[J3] **S. Lagen**, A. Agustin, J. Vidal, "Coexisting Linear and Widely Linear Transceivers in the MIMO Interference Channel", *IEEE Trans. on Signal Processing*, vol. 64, no. 3, pp. 652 - 664, Jan. 2016. doi: 10.1109/TSP.2015.2489604.

[C5] **S. Lagen**, A. Agustin, J. Vidal, "Decentralized Interference Management with Improper Gaussian Signaling for MIMO-IC", *IEEE Global Commun. Conf.* (GLOBE-COM), 3rd Int. Workshop on Emerging Technologies for 5G Wireless Cellular Networks, Austin (USA), Dec. 2014. doi: 10.1109/GLOCOMW.2014.7063526.

Chapter 6

The main results in this chapter are related to the development of traffic-aware and interference-aware resource management techniques for flexible TDD systems in dense networks, where the transmit direction selection, user scheduling and transmit precoding design are jointly optimized. They have produced one journal paper (that contains the MIMO case) and one conference paper (that considers the SISO case):

[J4] **S. Lagen**, A. Agustin, J. Vidal, "Joint User Scheduling, Precoder Design and Transmit Direction Selection in MIMO TDD Small Cell Networks", under minor revision at *IEEE Trans. on Wireless Commun.*, Oct. 2016.

[C6] **S. Lagen**, A. Agustin, J. Vidal, "Joint User Scheduling and Transmit Direction Selection in 5G TDD Dense Small Cell Networks", *IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun.* (PIMRC), Valencia (Spain), Sep. 2016.

Chapter 7

The main results in this chapter are regarding the derivation of long-term traffic-aware and interference-aware resource management schemes for dense OFDMA-based networks. They have resulted in one conference paper:

[C7] **S. Lagen**, O. Muñoz, A. Pascual-Iserte, J. Vidal, A. Agustin, "Long-term Provisioning of Radio Resources Based on their Utilization in Dense OFDMA Networks", *IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun.* (PIMRC), Valencia (Spain), Sep. 2016.

1.8.2   Other journal/conference papers

Although related but not included in the doctoral thesis, the author has contributed to the publication of the following journal papers and conference papers that refer to different research topics.

Benefits of CoMP-JT in relay-assisted transmissions

[J5]   **S. Lagen**, A. Agustin, and J. Vidal, "Network MIMO for downlink in-band relay transmissions", *EURASIP Journal on Wireless Commun. and Networking*, Jan. 2013. doi: 10.1186/1687-1499-2013-13.

[C8]   J. Vidal, A. Agustin, **S. Lagen**, E. Valera, O. Muñoz, A. G. Armada, M. Sanchez, "Network-MIMO Backhauling for QoS-Constrained Relay Transmission", *IEEE Int. Conf. on Acoustics, Speech and Signal Processing* (ICASSP), Prague (Czech Republic), May 2011. doi: 10.1109/ICASSP.2011.5947098.

[C9]   A. Agustin, J. Vidal, **S. Lagen**, and E. Valera, "Network MIMO for Downlink in-band Relay Transmissions with Relaying Phases of Fixed Duration", *19th European Signal Processing Conf.* (EUSIPCO), Barcelona (Spain), Sep. 2011.

Energy efficient communications and offloading techniques

[J6]   A. Pascual-Iserte, O. Muñoz, **S. Lagen**, and J. Vidal, "Energy Efficiency in Latency-Constrained Application Offloading for Multiple Virtual Machines", submitted to *IEEE Trans. on Vehicular Techn.*, Feb. 2016.

[C10]   A. Agustin, **S. Lagen**, and J. Vidal, "Energy Efficient Cell Load-aware Coverage Optimization for Small-Cell Networks", *IEEE Int. Conf. on Commun.* (ICC), London (United Kingdom), Jun. 2015. doi: 10.1109/ICC.2015.7248625.

Pilot design for channel estimation in dense networks

[C11]   A. Agustin, **S. Lagen**, and J. Vidal, "Channel Training Procedures for MIMO Interfering Point-to-Multipoint Channel", *IEEE Global Commun. Conf.* (GLOBECOM), Austin (USA), Dec. 2014, doi: 10.1109/GLOCOM.2014.7037450.

Time synchronization in dense networks

[C12]   A. Agustin, **S. Lagen**, and J. Vidal, "Signal-Timing-Offset Compensation in TDD SeNBs OFDM-based Networks", submitted to *IEEE Int. Conf. on Commun.* (ICC), Paris (France), May. 2017.

Massive machine type communications

[C13]   **S. Lagen**, J. García, A. Agustin, and J. Vidal, "Throughput, Delay and Energy Analysis of Finite-User Slotted Random Access with Feedback-free Collision Resolution", submitted to *IEEE Int. Conf. on Commun.* (ICC), Paris (France), May. 2017.

[J7]   **S. Lagen**, A. Agustin, J. Vidal, and J. García, "Throughput, Delay and Energy Analysis of Finite-User Slotted Random Access with Feedback-free Collision Resolution for massive MTC", in preparation, Oct. 2016.

1.8.3 Patent applications

In connection with Chapter 3 of this document, a patent application was filed in the U.S. Patent and Trademark Office on 6th August 2014, as U.S. Patent Application Serial No.: 14/452,968. The details are as follows:

[P]   S. Lagen, A. Agustin, J. Vidal, "Methods and systems for decentralized interference management in a multi-antenna wireless communication system", U.S. Patent 14/452,968, Aug. 6, 2014.

The patent has been accepted and is available at:
http://www.freepatentsonline.com/20150078186.pdf.

1.8.4 Contributions to 3GPP

As a result of the research included in Chapter 3 of this Ph.D. dissertation, several contributions to 3GPP LTE release 12 have been produced. They were addressed to the technical specification group on radio access network (RAN) - working group 1 (i.e. TSG-RAN WG1)[9]:

[Ite1]   R1-133213, Beamforming with Coordinated Sounding (BF-CoS) exploiting TDD channel reciprocity, 3GPP TSG-RAN WG1 #74, Barcelona, Spain, 19-23 August 2013.

[Ite2]   R1-133214, Coordinated Sounding for CoMP BF (CoMP BF-CoS) exploiting TDD channel reciprocity, 3GPP TSG-RAN WG1 #74, Barcelona, Spain, 19-23 August 2013.

[Ite3]   R1-134291, Coordinated Sounding for CoMP BF (CoMP BF-CoS) exploiting TDD channel reciprocity, 3GPP TSG-RAN WG1 #74bis, Guangzhou, China, 7-11 October 2013.

[Ite4]   R1-135239, Coordinated Sounding for CoMP BF (CoMP BF-CoS) including calibration information, 3GPP TSG-RAN WG1 #75, San Francisco, USA, 11-15 November 2013.

[Ite5]   R1-140510, Analysis and proposals for the enhancement of X2 signaling, 3GPP TSG-RAN WG1 #76, Prague, Czech Republic, 10-14 February 2014.

[Ite6]   R1-141234, Signaling of PDSCH and PUSCH usage for CoMP Hypothesis and Enhanced RNTP, 3GPP TSG-RAN WG1 #76bis, Shenzhen, China, 31 March - 4 April 2014.

[Ite7]   R1-141625, Signaling of SRS scheduling for CoMP Hypothesis, 3GPP TSG-RAN WG1 #76bis, Shenzhen, China, 31 March - 4 April 2014.

In addition, although related but not included in the present work, the author has participated in the following contributions to 3GPP LTE release 11 and 12 (that were also intended to TSG-RAN WG1):

[Ite8]   R1-112096, Proposal of a CoMP study focused on relay-based networks, 3GPP TSG-RAN WG1 #66, Athens, Greece, 22-26 August 2011.

[Ite9]   R1-133215, Collaborative distributed scheduling, 3GPP TSG-RAN WG1 meeting #74, Barcelona, Spain, 19-23 August 2013.

---

[9]The 3GPP specification work is done in technical specification groups (TSGs) and working groups (WGs). There are four TSGs: GERAN, RAN, SA, and CT, each of which consists of multiple WGs [17].

[lte10]   R1-134292, Collaborative distributed scheduling, 3GPP TSG-RAN WG1 #74bis, Guang-
zhou, China, 7-11 October 2013.

[lte11]   R1-135241, Collaborative distributed scheduling - updated information, 3GPP TSG-
RAN WG1 #75, San Francisco, USA, 11-15 November 2013.

[lte12]   R1-135241, Assessment of differences between centralized and knowledge-based dis-
tributed scheduling, 3GPP TSG-RAN WG1 #75, San Francisco, USA, 11-15 November
2013.

The contributions R1-135239 and R1-135241 were both included in a technical report (TR)
of 3GPP LTE release 12, whose details are:

[TR]   3GPP TR 36.874, 3rd Generation Partnership Project, Technical Specification Group
Radio Access Network, "Coordinated multi-point operation for LTE with non-ideal
backhaul (Release 12)", v.12.0.0, Dec. 2013.

The technical report can be found online at: `http://www.3gpp.org/dynareport/36874.htm`.

1.8.5   Contributions to projects

TROPIC project

The results obtained in Chapters 2-3-4-5 of this doctoral thesis have allowed the contribution
to the collaborative project ICT-318784 TROPIC: '*Distributed computing, storage and radio
resource allocation over cooperative femtocells*', which aims at exploiting the convergence of
smallcells network infrastructure and cloud computing paradigms for virtualization/distribution
of applications and services. The TROPIC project was funded by the European Comission
under the 7th framework program and had a duration from September 2012 to April 2015.

More specifically, the author has contributed to the following deliverables of the project:

- D21: Scenarios and requirements
- D312: MP2MP communication systems for LTE-A HeNB deployments
- D321: Cooperative and distributed interference estimation-detection
- D322: HeNB network coordination
- D61: Performance assessment of the femto-clouding
- D71: Dissemination report
- D72: Standardization report
- D73: IPR management report

The documents are available at: `http://ict-tropic.eu/`.

Let us remark that, within the framework of the TROPIC project, one of the techniques
developed in this thesis (in particular, in Chapter 3) was implemented in a real-time simulator
that emulates a complete protocol stack (including the LTE stack for the radio links and the IP
stack for the network links). The details can be found in deliverable D61.

Huawei-UPC Cooperation

Some of the research work presented in Chapters 6-7 of this Ph.D. dissertation have been carried out within a research and development cooperative project between Huawei and UPC, which was entitled: '*Study of methods leading to a more efficient utilization of frequency bands in asymmetric traffic conditions*'. The project was funded by Huawei and had a duration from May 2015 to May 2016.

The author has contributed to all the deliverables of the project:

- D11: Traffic models and scenarios
- D12: Advanced use of paired bands for asymmetric traffic
- D13: Interference management in a multi-cell scenario
- D14: Time Synchronization and SIC implementation for TDD SeNBs deployed in the FDD UL spectrum
- D21: IPR generation report
- D22: Dissemination report

The documents are not available owing to confidentiality reasons.

1.8.6 Keynote talks

In the context of the TROPIC project, the author gave a keynote talk at the TROPIC Training Workshop: '*Edge Cloud Empowered 5G Networks*'. The talk was held in Rome (Italy) in February 2015, and was entitled:

- *PHY layer enablers to edge cloud assisted application offloading*

In particular, the talk dealt with physical layer optimization to enable offloading of applications to the edge cloud by means of enhancing the spectral efficiency in dense smallcell networks and providing physical(PHY)-layer support to offloading techniques.

The slides can be downloaded from the project website: `http://ict-tropic.eu/`.

1.8.7 Research Appointment

From March to May 2015 the author did a 3-month research appointment at Nokia Networks and Aalborg University in Aalborg, Denmark. As a result of the collaboration, the conference paper **[C4]** related to Chapter 4 of the doctoral thesis arose.

Chapter 2

Advanced Signaling Schemes

This chapter investigates the use of advanced signaling schemes. In particular, it exploits majorization theory tools to formally quantify the gains of improper Gaussian signaling (IGS) along with widely linear transceivers as compared to conventional proper Gaussian signaling (PGS) in interference-limited MIMO scenarios. The MIMO point-to-point channel with interference (P2P-I) is analyzed, assuming that received interference can be either proper or improper. We demonstrate that the use of the optimal IGS, when received interference is improper, strictly outperforms (in terms of achievable rate and mean square error (MSE)) the use of the optimal PGS when interference is proper. Then, these results are extended to two practical situations. First, the MIMO Z-interference channel (Z-IC) is investigated, where a trade-off arises: with IGS we could increase the achievable rate of the interfered user while gracefully degrading the rate of the non-interfered user. Second, these concepts are applied to a two-tier heterogeneous cellular network (HCN) where macrocells and smallcells coexist and multiple MIMO Z-IC appear.

The technical papers related to this chapter are:

[J1]   S. Lagen, A. Agustin, J. Vidal, "On the Superiority of Improper Gaussian Signaling in Wireless Interference MIMO scenarios", *IEEE Trans. on Commun.*, vol. 64, no. 8, pp. 3350-3368, Aug. 2016. doi: 10.1109/TCOMM.2016.2584601.

[C1]   S. Lagen, A. Agustin, J. Vidal, "Improper Gaussian signaling for the Z-Interference Channel", *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Florence (Italy), May 2014. doi: 10.1109/ICASSP.2014.6853775.

This chapter is structured as follows. Section 2.1 contains the state of the art for IGS. The main contributions of the chapter are detailed in Section 2.2. Section 2.3 presents the system model for the use of IGS in the MIMO P2P-I, as well as the optimal signaling and the associated transmission scheme. In Section 2.4, the superiority of IGS in the MIMO P2P-I is demonstrated in terms of achievable rate and MSE. In Section 2.5, a practical improper-based scheme for the MIMO Z-IC is proposed and evaluated. Then, the proposed scheme is applied and evaluated in an HCN deployment in Section 2.6. Finally, concluding remarks are included in Section 2.7.

Appendix 2.A presents preliminaries for improper Gaussian random vectors. Preliminaries for majorization theory are included in Appendix 2.B.

2.1   State of the Art

It is well known that the statistical characteristics of the signals affect the maximum achievable rate of a wireless communication. In this sense, conventional PGS has been shown to be optimal in terms of capacity for the Gaussian MIMO P2P channel [105], for the Gaussian MIMO broadcast channel (BC) with DPC as the capacity achieving strategy and for the Gaussian MIMO multiple access channel (MAC) [48,49]. Due to that, PGS is usually also assumed in multi-user wireless communications. Nevertheless, recently, it has been shown that IGS, or circularly asymmetric complex Gaussian signaling, is able to improve the achievable rate in interference-limited scenarios [52,69,106].

The key difference between PGS and IGS is the characterization of the second order statistics: while a proper Gaussian random vector $\mathbf{x}$ is fully specified by the covariance matrix $\mathbf{C_x}=\mathbb{E}[\mathbf{xx}^H]$ under the zero-mean assumption, an improper Gaussian random vector $\mathbf{x}$ is characterized not only by the covariance matrix but also by the pseudo-covariance matrix $\tilde{\mathbf{C}}_{\mathbf{x}}=\mathbb{E}[\mathbf{xx}^T]$, which means that the improper Gaussian random vector $\mathbf{x}$ is correlated with its complex conjugate $\mathbf{x}^*$ [65,107]. Thus, the real and imaginary parts of an improper Gaussian random vector are correlated or have unequal power [65]. This adds more flexibility to the system, as it allows treating in a different manner the real and imaginary parts of the random vectors. IGS arises naturally in communications, e.g. due to the gain imbalance between the in-phase and in-quadrature components or due to the use of specific digital modulations that exhibit improper characteristics, such as binary phase shift keying (BPSK), Gaussian minimum shift keying (GMSK), or continuous phase modulation (CPM).

It is shown in [108,109] that optimality of PGS does not necessarily hold for the MIMO BC when transceivers are restricted to be linear and that capacity gains can be obtained by employing IGS because, in the absence of non-linear interference cancellation, the MIMO BC becomes an interference-limited scenario [110]. The same applies in the Gaussian MIMO interference channel (IC), where IGS is able to improve the known achievable rates when interference is treated as noise. For example, in [69] it is shown that for a fixed set of transmit covariance matrices, the achievable rate of a specific user could be increased by optimizing the transmit pseudo-covariance matrices. Even so, the capacity of the MIMO IC is still an open problem: it is difficult to be analyzed when linear transceivers are considered due to the non-convexity of the problem and, for that reason, performance gains with IGS in the MIMO IC have been only shown in terms of degrees of freedom at high SNR [52,106,111,112] or in terms of capacity and rate region in specific MIMO configurations [68,69,113]. In addition to the MIMO IC, IGS has also been shown to be beneficial in other interference scenarios, as underlay cognitive radio systems [114–117] or full-duplex systems [118,119].

In order to generate and estimate IGS with linear processing, widely linear precoding (WLP) [69] and widely linear estimation (WLE) [65] need to be adopted at transmitter and receiver side, respectively, instead of the commonly used linear precoding (LP) and linear estimation (LE) techniques that maintain the proper Gaussian distribution. WLE is a generalized concept used in estimation theory whenever improper noise is encountered [66] and in systems that transmit improper signal constellations [67], as is already investigated in 3GPP LTE-A [22]. On the other hand, WLP can be used to generate IGS even when departing from proper signal constellations [108]. While WLE has been extensively studied in literature, the generation of IGS through WLP to handle interference more effectively is a rather new line of research.

2.2 Contribution

In this chapter we focus on gaining more insights on the use of IGS along with widely linear transceivers for MIMO systems in interference-limited conditions. To do so we use the equivalent **composite real representation** [105], whereby real and imaginary parts of the MIMO channel are separated, and exploit majorization theory [96, 97] to formally quantify the benefits. We analyze the MIMO P2P channel with interference (MIMO P2P-I), in which a transmitter (TX) equipped with M antennas wish to send information to a receiver (RX) equipped with N antennas that is receiving noise-plus-interference.

The optimal signaling (PGS or IGS) and the associated transmit/receive scheme (linear or widely linear) for the MIMO P2P-I are well known in the literature and can be obtained from the minimization of any Schur-concave function of the MSE-matrix [98, 120], like maximization of the achievable rate or minimization of the MSE of the transmitted symbols. Their solutions are determined by the statistics of the interference, which can be modeled either as a PGS or IGS depending on the transmission strategy of the interferer: if interference is proper then the optimal signaling is PGS, while if interference is improper then the optimal signaling is IGS. In this sense, we focus on the following question:

Assuming the optimal signaling on each case, is it better to use IGS (and improper interference) or to use PGS (and proper interference) in the MIMO P2P-I?

By using majorization theory tools (see [96, 97]), we demonstrate that the use of IGS allows obtaining a strict improvement in terms of achievable rate and MSE for $M \geq N$ and full rank channel matrix. The proof sheds light on why the use of IGS is better. Then, we investigate how to exploit these properties in different interference-limited scenarios.

On the one hand, we study the **MIMO Z-IC** [121], a two-transmitter two-receiver multi-antenna interfering scenario where one of the receivers (RX_B) just observes noise while the other (RX_A) receives interference. Certain situations in cellular networks can be modeled through the MIMO Z-IC, as shown in Fig. 2.1 where RX_A is in the cell-edge. Under the condition of treating interference as noise, the optimal transmission scheme for sum-rate maximization in the Z-IC is only known for the SISO case [122], case for which the optimal rate region boundary has been derived in [123]. In both cases, a set of five possible solutions that subsume PGS and IGS depending on the channel conditions is obtained. The MIMO Z-IC is interesting because TX_B should apply conventional PGS for optimally transmitting to its intended receiver (RX_B) but, according to the properties derived in the sequel for the MIMO P2P-I, it will be beneficial for RX_A that TX_B employs IGS such that the interference injected onto RX_A is improper and TX_A can design WLP to improve the performance of RX_A (see Fig. 2.1). Clearly this comes at the cost of reducing the performance of RX_B, such that a clear trade-off in the rates of the two receivers arises. In this regard, we analyze the trade-off and propose a simple improper-based scheme for TX_B that allows improving the fairness and controlling the sum-rate performance.
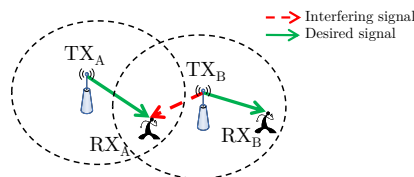

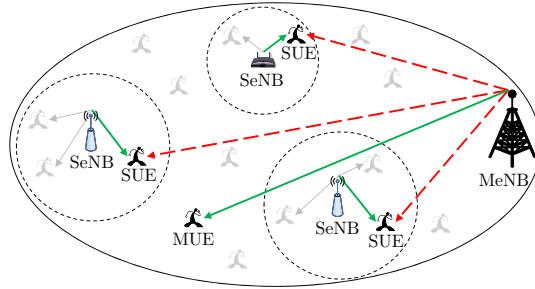Figure 2.1: Example of the MIMO Z-IC.

Figure 2.2: Predominant cross-tier interference (red dashed lines) in HCNs, which lead to multiple MIMO Z-IC.

On the other hand, we apply the aforementioned properties to MIMO **HCNs** [6], a promising deployment for future cellular systems. HCNs consist of a multi-tier deployment of MeNBs and SeNBs that allow boosting the spectral efficiency of the system thanks to the network densification and the spatial re-use of the spectrum among MeNBs and SeNBs (see description in Section 1.1). However, HCNs have to deal with a major impairment: the *cross-tier interference*, which can be very detrimental specially in the direction from MeNBs towards the users served by SeNBs with a lower transmit power [1][10], as illustrated in Fig. 2.2. Interestingly, the interfering channel that models the major cross-tier interference impairment in HCNs is the MIMO Z-IC. Furthermore, there will appear as many MIMO Z-IC as the number of SeNBs deployed within the MeNB coverage area, see Fig. 2.2. Hence, the derived improper-based scheme for the MIMO Z-IC could be adopted at the MeNB (corresponding to TX_B in Fig. 2.1) and a performance improvement of all the users served by SeNBs would be guaranteed thanks to the properties derived for the MIMO P2P-I.

To summarize, the main contributions of this chapter are:

- In the MIMO P2P-I with $M \geq N$, when uniform power allocation (UPA) is adopted, we demonstrate that the use of optimal IGS when received interference is improper strictly outperforms in terms of achievable rate and MSE the use of optimal PGS when interference is proper. If optimal power allocation (OPA) is used, then the achievable rate is shown to be equal or larger with IGS, and strict superiority is shown for a particular case.

- In the MIMO Z-IC, we propose a simple improper-based scheme without any claim about optimality that allows improving the achievable rate of the most impaired user and controlling the rate loss of the non-interfered user. The scheme is characterized by a parameter that provides two extreme solutions: PGS or maximal IGS. If such parameter is fixed (e.g. maximal IGS is used), then its implementation does not require knowledge of the interfering channel either at TXs or at RXs.

- The proposed improper-based scheme is evaluated in an HCN deployment compliant with 3GPP LTE-A specifications [2]. The scheme is compared with eICIC technique (used in 3GPP LTE-A for HCNs [1, 56]), time-sharing solutions, and proper-based schemes with full-reuse of time and frequency resources at MeNB and SeNBs. Simulations show significant gains in terms of 5%-tile and mean user throughput, with a reduced implementation complexity and a simpler coordination, as compared to conventional interference coordination techniques.

---

[10]Note that if SeNBs were deployed in concentrated areas then the co-tier interference would become relevant and additional techniques would be required, see [**J2**]. We explicitly focus on the dense SCN in Chapters 3-4-6.
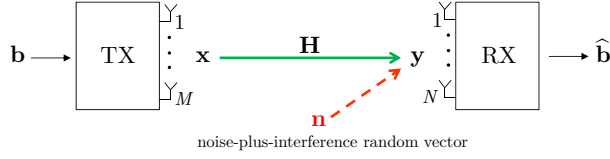
Figure 2.3: MIMO P2P channel with interference (MIMO P2P-I).

2.3 System Model

Consider a MIMO P2P-I between one TX equipped with M antennas and one RX with N receive antenna elements, as shown in Fig. 2.3. The MIMO channel is described by a matrix $\mathbf{H}\in\mathbb{C}^{N\times M}$ containing complex-valued channel gains of the different antenna-pairs. In the following we assume $M\geq N$ and full rank channel matrix $\mathbf{H}$. Hence, assuming narrow-band transmissions, the equivalent baseband signal observed at the receiver $\mathbf{y}\in\mathbb{C}^{N\times 1}$ is expressed as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \ \ \mathbf{n} = \mathbf{v} + \mathbf{i}, \tag{2.1}$$

where $\mathbf{x}\in\mathbb{C}^{M\times 1}$ is the complex-valued transmitted vector and $\mathbf{n}\in\mathbb{C}^{N\times 1}$ denotes the noise-plus-interference received vector (see Fig. 2.3), which contains a proper Gaussian noise $\mathbf{v}\sim\mathcal{CN}(\mathbf{0},\sigma^2\mathbf{I})$ and a Gaussian interference component $\mathbf{i}$ that is caused by an interfering TX. Different from the conventional transmission setup where PGS is assumed (i.e. $\mathbf{x}\sim\mathcal{CN}(\mathbf{0},\mathbf{C_x})$), in this chapter the more general IGS is adopted for $\mathbf{x}$.

Any improper Gaussian signal $\mathbf{x}$ can be generated from a proper Gaussian information-bearing signal $\mathbf{b}\in\mathbb{C}^{N\times 1}$ with unitary power (i.e. $\mathbf{b}\sim\mathcal{CN}(\mathbf{0},\mathbf{I})$)) and the use of WLP (see Lemma 2.6 in Appendix 2.A) as follows:

$$\mathbf{x} = \mathbf{T}_1\mathbf{b} + \mathbf{T}_2\mathbf{b}^*, \tag{2.2}$$

where matrices $\mathbf{T}_1$ and $\mathbf{T}_2\in\mathbb{C}^{M\times N}$ denote the linear transmit precoders for the information-bearing signal $\mathbf{b}$ and its complex conjugate $\mathbf{b}^*$. The conventional LP scheme is a special case of WLP in which $\mathbf{T}_2=\mathbf{0}$ such that $\mathbf{x}$ in (2.2) is PGS. It is assumed that each user has N parallel data streams, although some of the streams can have a rate of zero, and let us recall that $\mathbb{E}[\mathbf{b}\mathbf{b}^H]=\mathbf{I}$ and $\mathbb{E}[\mathbf{b}\mathbf{b}^T]=\mathbf{0}$. Then, the covariance matrix $\mathbf{C_x}=\mathbb{E}[\mathbf{x}\mathbf{x}^H]$ and the pseudo-covariance matrix $\tilde{\mathbf{C}}_\mathbf{x}=\mathbb{E}[\mathbf{x}\mathbf{x}^T]$ of $\mathbf{x}$ in (2.2) are:

$$\mathbf{C_x} = \mathbf{T}_1\mathbf{T}_1^H + \mathbf{T}_2\mathbf{T}_2^H, \quad \tilde{\mathbf{C}}_\mathbf{x} = \mathbf{T}_1\mathbf{T}_2^T + \mathbf{T}_2\mathbf{T}_1^T. \tag{2.3}$$

Similarly, since the Gaussian interference component $\mathbf{i}$ in (2.1) can be either proper or improper depending on the transmission scheme adopted by the interferer, we denote by $\mathbf{C_n}=\mathbb{E}[\mathbf{n}\mathbf{n}^H]$ and $\tilde{\mathbf{C}}_\mathbf{n}=\mathbb{E}[\mathbf{n}\mathbf{n}^T]$ the covariance matrix and the pseudo-covariance matrix of the noise-plus-interference vector $\mathbf{n}=\mathbf{v}+\mathbf{i}$ in (2.1), respectively.

In order to access the information contained in the received signal $\mathbf{y}$, WLE should be applied at the RX [65]. Hence, the information-bearing signal $\mathbf{b}$ is linearly estimated according to:

$$\hat{\mathbf{b}} = \mathbf{R}_1^H\mathbf{y} + \mathbf{R}_2^H\mathbf{y}^*, \tag{2.4}$$

where $\mathbf{R}_1$ and $\mathbf{R}_2\in\mathbb{C}^{N\times N}$ are the linear receive filters applied over $\mathbf{y}$ and $\mathbf{y}^*$, respectively. The well-known LE is a particular case of WLE in which $\mathbf{R}_2=\mathbf{0}$.

2.3.1　Composite real formulation

The signal model for the MIMO P2P-I can be equivalently expressed using the *composite real representation* [105], in which real and imaginary parts of the signals are separated. Accordingly, for any complex vector $\mathbf{a}$, let us denote by $\bar{\mathbf{a}}$ to the double-sized real-valued vector that stacks its real and imaginary parts as: $\bar{\mathbf{a}} = \left[\Re\{\mathbf{a}\}^T \ \Im\{\mathbf{a}\}^T \right]^T$. The composite real formulation is useful to derive the optimal signaling and the associated transmission scheme in (2.2). This way, the input-output relation in (2.1) can be equivalently written as:

$$\bar{\mathbf{y}} = \left[\begin{array}{c} \Re\{\mathbf{y}\} \\ \Im\{\mathbf{y}\} \end{array} \right] = \bar{\mathbf{H}}\bar{\mathbf{x}} + \bar{\mathbf{n}}, \quad \bar{\mathbf{H}} = \left[\begin{array}{cc} \Re\{\mathbf{H}\} & -\Im\{\mathbf{H}\} \\ \Im\{\mathbf{H}\} & \Re\{\mathbf{H}\} \end{array} \right], \quad (2.5)$$

where $\bar{\mathbf{y}} \in \mathbb{R}^{2N \times 1}$, $\bar{\mathbf{H}} \in \mathbb{R}^{2N \times 2M}$, $\bar{\mathbf{n}} \in \mathbb{R}^{2N \times 1}$, and $\bar{\mathbf{x}} \in \mathbb{R}^{2M \times 1}$ (following WLP in (2.2)) is given by:

$$\bar{\mathbf{x}} = \left[\begin{array}{c} \Re\{\mathbf{x}\} \\ \Im\{\mathbf{x}\} \end{array} \right] = \bar{\mathbf{T}}\bar{\mathbf{b}}, \quad \bar{\mathbf{T}} = \left[\begin{array}{cc} \Re\{\mathbf{T}_1 + \mathbf{T}_2\} & -\Im\{\mathbf{T}_1 - \mathbf{T}_2\} \\ \Im\{\mathbf{T}_1 + \mathbf{T}_2\} & \Re\{\mathbf{T}_1 - \mathbf{T}_2\} \end{array} \right]. \quad (2.6)$$

This way, the covariance matrix of $\bar{\mathbf{x}}$ in (2.6) is related to $\mathbf{C}_{\mathbf{x}}$ and $\tilde{\mathbf{C}}_{\mathbf{x}}$ in (2.3) as:

$$\mathbf{C}_{\bar{\mathbf{x}}} = \mathbb{E}[\bar{\mathbf{x}}\bar{\mathbf{x}}^T] = \tfrac{1}{2} \left[\begin{array}{cc} \Re\{\mathbf{C}_{\mathbf{x}} + \tilde{\mathbf{C}}_{\mathbf{x}}\} & -\Im\{\mathbf{C}_{\mathbf{x}} - \tilde{\mathbf{C}}_{\mathbf{x}}\} \\ \Im\{\mathbf{C}_{\mathbf{x}} + \tilde{\mathbf{C}}_{\mathbf{x}}\} & \Re\{\mathbf{C}_{\mathbf{x}} - \tilde{\mathbf{C}}_{\mathbf{x}}\} \end{array} \right]. \quad (2.7)$$

Finally, following WLE in (2.4), the double-sized real-valued vector of the estimated information-bearing signal $\bar{\hat{\mathbf{b}}} \in \mathbb{R}^{2N \times 1}$ is estimated at the receiver from $\bar{\mathbf{y}}$ in (2.5) as:

$$\bar{\hat{\mathbf{b}}} = \left[\begin{array}{c} \Re\{\hat{\mathbf{b}}\} \\ \Im\{\hat{\mathbf{b}}\} \end{array} \right] = \bar{\mathbf{R}}^T\bar{\mathbf{y}}, \quad \bar{\mathbf{R}}^T = \left[\begin{array}{cc} \Re\{\mathbf{R}_1^H + \mathbf{R}_2^H\} & -\Im\{\mathbf{R}_1^H - \mathbf{R}_2^H\} \\ \Im\{\mathbf{R}_1^H + \mathbf{R}_2^H\} & \Re\{\mathbf{R}_1^H - \mathbf{R}_2^H\} \end{array} \right]. \quad (2.8)$$

2.3.2　Achievable rate and MSE-matrix

The achievable rate (R) with IGS is given by [107]:

$$R = I(\mathbf{x}; \mathbf{y}) = I(\bar{\mathbf{x}}; \bar{\mathbf{y}}) = h(\bar{\mathbf{y}}) - h(\bar{\mathbf{y}}/\bar{\mathbf{x}}) = h(\bar{\mathbf{y}}) - h(\bar{\mathbf{n}}) = \tfrac{1}{2}\log_2|\mathbf{I}_{2N} + \tfrac{1}{2}\bar{\mathbf{H}}\bar{\mathbf{T}}\bar{\mathbf{T}}^T\bar{\mathbf{H}}^T\mathbf{C}_{\bar{\mathbf{n}}}^{-1}|, \quad (2.9)$$

where $I(\bar{\mathbf{x}}; \bar{\mathbf{y}})$ denotes the mutual information between real random vectors $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$, $h(\bar{\mathbf{x}})$ refers to the entropy of $\bar{\mathbf{x}}$ [107], and the $\tfrac{1}{2}$ factor inside the determinant comes from $\mathbb{E}[\bar{\mathbf{b}}\bar{\mathbf{b}}^H] = \tfrac{1}{2}\mathbf{I}$. $\mathbf{C}_{\bar{\mathbf{n}}} = \mathbb{E}[\bar{\mathbf{n}}\bar{\mathbf{n}}^T] \in \mathbb{R}^{2N \times 2N}$ accounts for the covariance matrix of $\bar{\mathbf{n}}$ in (2.5), which can be written in terms of $\mathbf{C}_{\mathbf{n}}$ and $\tilde{\mathbf{C}}_{\mathbf{n}}$ through a transformation matrix $\mathbf{Z} \in \mathbb{C}^{2N \times 2N}$ as (see details in Lemma 2.5 of Appendix 2.A):

$$\mathbf{C}_{\bar{\mathbf{n}}} = \mathbb{E}[\bar{\mathbf{n}}\bar{\mathbf{n}}^T] = \mathbf{Z} \left[\begin{array}{cc} \mathbf{C}_{\mathbf{n}} & \tilde{\mathbf{C}}_{\mathbf{n}} \\ \tilde{\mathbf{C}}_{\mathbf{n}}^* & \mathbf{C}_{\mathbf{n}}^* \end{array} \right] \mathbf{Z}^H, \quad \mathbf{Z} = \tfrac{1}{2} \left[\begin{array}{cc} \mathbf{I} & \mathbf{I} \\ -j\mathbf{I} & j\mathbf{I} \end{array} \right]. \quad (2.10)$$

The MSE of the transmitted symbols is expressed as: $\epsilon = \mathrm{Tr}(\mathbf{E})$, being $\mathbf{E}$ the MSE-matrix $\mathbf{E} = \mathbb{E}[(\mathbf{b} - \hat{\mathbf{b}})(\mathbf{b} - \hat{\mathbf{b}})^H] \in \mathbb{C}^{N \times N}$. It can be similarly obtained from the MSE-matrix in the composite real formulation as:

$$\epsilon = \mathrm{Tr}(\bar{\mathbf{E}}) = \mathrm{Tr}(\mathbb{E}[(\bar{\mathbf{b}} - \bar{\hat{\mathbf{b}}})(\bar{\mathbf{b}} - \bar{\hat{\mathbf{b}}})^T]). \quad (2.11)$$

Assuming optimum WLE for $\bar{\mathbf{R}}$ in (2.8) (see details in [65]), the MSE-matrix results:

$$\bar{\mathbf{E}} = \tfrac{1}{2}\left(\mathbf{I} + \tfrac{1}{2}\bar{\mathbf{T}}^T\bar{\mathbf{H}}^T\mathbf{C}_{\bar{\mathbf{n}}}^{-1}\bar{\mathbf{H}}\bar{\mathbf{T}}\right)^{-1}. \tag{2.12}$$

Equations (2.9) and (2.12) are directly related. Moreover, the relation between the MSE-matrices ($\bar{\mathbf{E}}$ and $\mathbf{E}$) and the achievable rate R can be extracted from the composite real representation. The result is given in the following Lemma.

Lemma 2.1. *When IGS is adopted in a MIMO P2P-I, the achievable rate R is related to the MSE-matrix in the composite real form $\bar{\mathbf{E}}$ in (2.12) and to the MSE-matrix $\mathbf{E}$ through:*

$$R = -\tfrac{1}{2}\log_2|2\bar{\mathbf{E}}| = -\tfrac{1}{2}\log_2|\mathbf{E}| - \tfrac{1}{2}\log_2|\mathbf{E}^* - \tilde{\mathbf{E}}^*\mathbf{E}^{-1}\tilde{\mathbf{E}}|, \tag{2.13}$$

where $\tilde{\mathbf{E}}=\mathbb{E}[(\mathbf{b}-\hat{\mathbf{b}})(\mathbf{b}-\hat{\mathbf{b}})^T]$. Further, using the Schur complement of $|\bar{\mathbf{E}}|$ and applying the improved Fischer determinant inequality [124], the achievable rate in (2.13) is lower bounded by:

$$R \underset{(a)}{\geq} -\log_2|\mathbf{E}| + \tfrac{1}{2}\log_2\left(1+\tfrac{1}{2}|\tilde{\mathbf{E}}||\tilde{\mathbf{E}}^*||\bar{\mathbf{E}}|^{-1}\right) \underset{(b)}{\geq} -\log_2|\mathbf{E}|, \tag{2.14}$$

where (a) is satisfied with equality if $\mathbf{E}$ is a scalar (i.e. for SISO, MISO, or SIMO systems) and (b) is satisfied with equality for the PGS case.

Proof. See Appendix 2.C. ∎

Note that in case of PGS (i.e. $\tilde{\mathbf{E}}=\mathbf{0}$), the conventional relation between the MSE-matrix $\mathbf{E}$ and the achievable rate R is obtained from (2.13): $R=-\log_2|\mathbf{E}|$ [125]. The lower bound in (2.14) indicates that, for a given MSE-matrix $\mathbf{E}$, the rate achieved with IGS is equal to or larger than the rate obtained with PGS. Moreover, it also shows that both $\mathbf{E}$ and $\tilde{\mathbf{E}}$ are required to determine the transmission rate, and not only $|\mathbf{E}|$ as in the PGS case.

2.3.3 MSE-based optimal scheme

This section presents the optimal design for transmit precoder $\bar{\mathbf{T}}$ in (2.6), a result that is already known in the literature but that is needed to demonstrate the superiority of IGS in Section 2.4. In the most general MSE-based designs (including minimum MSE and maximum achievable rate designs) [98, 120], the optimal transmit precoder $\bar{\mathbf{T}}$ in (2.6) is obtained from the minimization of a Schur-concave function of the MSE-matrix $\bar{\mathbf{E}}$ in (2.12) subject to a maximum transmit power constraint:

$$(\mathrm{P}_{2,1}): \quad \underset{\bar{\mathbf{T}}}{\text{minimize}} \quad f(\bar{\mathbf{E}}) \tag{2.15}$$

$$\text{subject to} \quad \tfrac{1}{2}\mathrm{Tr}(\bar{\mathbf{T}}\bar{\mathbf{T}}^T) = P^{\mathrm{max}},$$

where $f(.)$ is any Schur-concave function of $\bar{\mathbf{E}}$ in (2.12) and P^{max} is the maximum available power at the TX. For instance:

- $f(\bar{\mathbf{E}})=\epsilon=\mathrm{Tr}(\bar{\mathbf{E}})$ defines the minimum MSE problem, while
- $f(\bar{\mathbf{E}})=-R=\tfrac{1}{2}\log_2|2\bar{\mathbf{E}}|$ defines the maximum achievable rate problem.

Consider the following eigenvalue decomposition (EVD):

$$\tfrac{1}{2}\bar{\mathbf{H}}^T\mathbf{C}_{\bar{\mathbf{n}}}^{-1}\bar{\mathbf{H}} = \bar{\mathbf{Q}}\bar{\mathbf{\Lambda}}\bar{\mathbf{Q}}^H, \tag{2.16}$$

where $\bar{\mathbf{\Lambda}}=\text{diag}(\bar{\lambda}_1\ldots\bar{\lambda}_{2N})\in\mathbb{R}^{2M\times2N}$ is a diagonal matrix containing the positive eigenvalues of $\tfrac{1}{2}\bar{\mathbf{H}}^T\mathbf{C}_{\bar{\mathbf{n}}}^{-1}\bar{\mathbf{H}}$ and $\bar{\mathbf{Q}}\in\mathbb{R}^{2M\times2N}$ corresponds to a unitary matrix that contains the associated eigenvectors stacked in columns. Recall that the rank of $\tfrac{1}{2}\bar{\mathbf{H}}^T\mathbf{C}_{\bar{\mathbf{n}}}^{-1}\bar{\mathbf{H}}$ (i.e. the number of positive eigenvalues) is given by $2N$ and does not depend on the interference statistics, since $\text{rank}(\mathbf{C}_{\bar{\mathbf{n}}}^{-1})=2N$, $\bar{\mathbf{H}}$ is a full row rank matrix, and hence $\text{rank}(\tfrac{1}{2}\bar{\mathbf{H}}^T\mathbf{C}_{\bar{\mathbf{n}}}^{-1}\bar{\mathbf{H}})=\text{rank}(\mathbf{C}_{\bar{\mathbf{n}}}^{-1})=2N$.

Then, the optimal transmit precoder $\bar{\mathbf{T}}^{\text{opt}}$ to problem $(P_{2,1})$ in (2.15) presents the following structure [120]:

$$\bar{\mathbf{T}}^{\text{opt}} = \bar{\mathbf{Q}}\bar{\mathbf{P}}^{\frac{1}{2}}, \tag{2.17}$$

where $\bar{\mathbf{P}}=\text{diag}(\bar{p}_1\ldots\bar{p}_{2N})\in\mathbb{R}^{2M\times2N}$ is a diagonal matrix that describes the power allocation per stream, satisfies $\tfrac{1}{2}\text{Tr}(\bar{\mathbf{P}})=P^{\max}$ (see problem $(P_{2,1})$ in (2.15)), and depends on the optimization criterion (i.e. $f(\bar{\mathbf{E}})$ in (2.15) [126]) and on the channel knowledge we dispose of (see Sections 2.4.1 and 2.4.2). The optimal transmit precoder in (2.17) allows diagonalizing the MSE-matrix in (2.12):

$$\bar{\mathbf{E}}^{\text{opt}} = \tfrac{1}{2}\big(\mathbf{I}+\bar{\mathbf{P}}\bar{\mathbf{\Lambda}}\big)^{-1}. \tag{2.18}$$

So, the optimal achievable rate and the optimal MSE become, respectively:

$$R^{\text{opt}}(\bar{\mathbf{p}}\circ\bar{\mathbf{\lambda}}) = -\tfrac{1}{2}\log_2|2\bar{\mathbf{E}}^{\text{opt}}| = \tfrac{1}{2}\sum_{i=1}^{2N}\log_2\big(1+\bar{p}_i\bar{\lambda}_i\big), \tag{2.19}$$

$$\epsilon^{\text{opt}}(\bar{\mathbf{p}}\circ\bar{\mathbf{\lambda}}) = \text{Tr}(\bar{\mathbf{E}}^{\text{opt}}) = \tfrac{1}{2}\sum_{i=1}^{2N}\frac{1}{1+\bar{p}_i\bar{\lambda}_i}, \tag{2.20}$$

where $\bar{\mathbf{p}}\circ\bar{\mathbf{\lambda}}$ denotes the Hadamard product (component-wise product) of vectors $\bar{\mathbf{p}}=\text{diag}(\bar{\mathbf{P}})=[\bar{p}_1\ldots\bar{p}_{2N}]^T$ and $\bar{\mathbf{\lambda}}=\text{diag}(\bar{\mathbf{\Lambda}})=[\bar{\lambda}_1\ldots\bar{\lambda}_{2N}]^T$. We occasionally use the notations $R^{\text{opt}}(\bar{\mathbf{p}}\circ\bar{\mathbf{\lambda}})$ and $\epsilon^{\text{opt}}(\bar{\mathbf{p}}\circ\bar{\mathbf{\lambda}})$ to make their dependencies on $\bar{\mathbf{p}}\circ\bar{\mathbf{\lambda}}$ explicit.

The key aspect of the optimal solution in (2.17) is the EVD in (2.16), for which two cases are differentiated:

- proper interference ($\tilde{\mathbf{C}}_{\mathbf{n}}=\mathbf{0}$) or
- improper interference ($\tilde{\mathbf{C}}_{\mathbf{n}}\neq\mathbf{0}$).

It is well known that if received interference is proper the optimum signaling is PGS (i.e. $\tilde{\mathbf{C}}_{\mathbf{x}}=\mathbf{0}$), while if received interference is improper the optimum signaling is IGS (i.e. $\tilde{\mathbf{C}}_{\mathbf{x}}\neq\mathbf{0}$).

Proper interference

When received interference is proper Gaussian distributed (i.e. $\tilde{\mathbf{C}}_{\mathbf{n}}=\mathbf{0}$), the solution for $\bar{\mathbf{Q}}$ in (2.16) can be obtained without making use of the equivalent double-sized real-valued decomposition. By developing $\tfrac{1}{2}\bar{\mathbf{H}}^T\mathbf{C}_{\bar{\mathbf{n}}}^{-1}\bar{\mathbf{H}}$ with the structure of $\bar{\mathbf{H}}$ in (2.5) and with the structure of $\mathbf{C}_{\bar{\mathbf{n}}}$ in (2.10) where $\tilde{\mathbf{C}}_{\mathbf{n}}=\mathbf{0}$, we get:

$$\tfrac{1}{2}\bar{\mathbf{H}}^T\mathbf{C}_{\bar{\mathbf{n}}}^{-1}\bar{\mathbf{H}} = \begin{bmatrix} \Re\{\mathbf{H}^H\mathbf{C}_{\mathbf{n}}^{-1}\mathbf{H}\} & -\Im\{\mathbf{H}^H\mathbf{C}_{\mathbf{n}}^{-1}\mathbf{H}\} \\ \Im\{\mathbf{H}^H\mathbf{C}_{\mathbf{n}}^{-1}\mathbf{H}\} & \Re\{\mathbf{H}^H\mathbf{C}_{\mathbf{n}}^{-1}\mathbf{H}\} \end{bmatrix}. \tag{2.21}$$

Consequently, with some matrix manipulations, it can be shown that the solution for $\bar{\mathbf{Q}}$ in (2.17) is equivalently obtained from the eigenvectors of $\mathbf{H}^H\mathbf{C_n}^{-1}\mathbf{H}{=}\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^H$ stacked as follows:

$$\bar{\mathbf{Q}} = \left[\begin{array}{cc} \Re\{\mathbf{Q}\} & -\Im\{\mathbf{Q}\} \\ \Im\{\mathbf{Q}\} & \Re\{\mathbf{Q}\} \end{array}\right]. \tag{2.22}$$

By comparing (2.22) and (2.6), the following transmit precoders in (2.17) are obtained for PGS: $\mathbf{T}_1^{\text{opt}}{=}\mathbf{Q}\mathbf{P}^{\frac{1}{2}}$, $\mathbf{T}_2^{\text{opt}}{=}\mathbf{0}$, and hence $\tilde{\mathbf{C}}_{\mathbf{x}}{=}\mathbf{0}$ in (2.3) such that PGS is optimum.

Improper interference

When received interference is improper Gaussian distributed (i.e. $\tilde{\mathbf{C}}_{\mathbf{n}}{\neq}\mathbf{0}$), relation in (2.21) is not further satisfied and the optimal solution has to be obtained from the EVD in (2.16). Thus, the optimum transmit strategy leads to a precoder $\bar{\mathbf{T}}^{\text{opt}}$ in (2.17) with non-equal block diagonal matrices for $\mathbf{C}_{\bar{\mathbf{x}}}$ in (2.7) and hence $\tilde{\mathbf{C}}_{\mathbf{x}}{\neq}\mathbf{0}$, i.e. the optimal transmit strategy is given by IGS as it is the only one able to diagonalize the MSE-matrix $\bar{\mathbf{E}}$ in (2.12).

2.4 Superiority of IGS for the MIMO P2P-I

In this section we focus on comparing the optimal schemes when interference is either proper or improper so as to determine if IGS is beneficial (in terms of achievable rate and MSE) or not even in the simple MIMO P2P-I. The comparison reduces to relate the eigenvalues (see (2.19)-(2.20)) of the following two matrices (which are obtained from (2.16) by using the structure of $\mathbf{C}_{\bar{\mathbf{n}}}$ in (2.10)) for a fixed $\mathbf{C_n}$:

$$\begin{aligned} \text{Proper:} &\qquad \bar{\boldsymbol{\lambda}}_I = \text{eig}\Big(\tfrac{1}{2}\bar{\mathbf{H}}^T\mathbf{Z}^{-H}\left[\begin{array}{cc} \mathbf{C_n} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_n^*} \end{array}\right]^{-1}\mathbf{Z}^{-1}\bar{\mathbf{H}}\Big), \\[2mm] \text{Improper:} &\qquad \bar{\boldsymbol{\lambda}}_P = \text{eig}\Big(\tfrac{1}{2}\bar{\mathbf{H}}^T\mathbf{Z}^{-H}\left[\begin{array}{cc} \mathbf{C_n} & \tilde{\mathbf{C}}_{\mathbf{n}} \\ \tilde{\mathbf{C}}_{\mathbf{n}}^* & \mathbf{C_n^*} \end{array}\right]^{-1}\mathbf{Z}^{-1}\bar{\mathbf{H}}\Big), \end{aligned} \tag{2.23}$$

where $\bar{\boldsymbol{\lambda}}_I{\in}\mathbb{R}_+^{2N\times1}$ and $\bar{\boldsymbol{\lambda}}_P{\in}\mathbb{R}_+^{2N\times1}$ denote the vectors of positive eigenvalues in decreasing order under the reception of improper and proper Gaussian interference, respectively. The comparison in the sequel is based on fixing the same $\mathbf{C_n}$, so the level of received interference-plus-noise power is the same but the difference comes from the proper or improper statistics of the interference (i.e. $\tilde{\mathbf{C}}_{\mathbf{n}}{=}\mathbf{0}$ or $\tilde{\mathbf{C}}_{\mathbf{n}}{\neq}\mathbf{0}$, see (2.23)) and the corresponding optimal signaling (i.e. PGS or IGS).

In Lemma 2.2 we show how the eigenvalues in (2.23) are related and, afterwards, how their relation impacts on the optimal achievable rate in (2.19) and the optimal MSE in (2.20).

Lemma 2.2. *For $M{\geq}N$ and full rank channel matrix $\mathbf{H}$, the eigenvalues in (2.23) are related by the following majorization relations:*

$$\bar{\boldsymbol{\lambda}}_I^{-1}{\succ}\bar{\boldsymbol{\lambda}}_P^{-1} \quad \text{and} \quad \bar{\boldsymbol{\lambda}}_I{\succ}_{w\log}\bar{\boldsymbol{\lambda}}_P, \tag{2.24}$$

where $\succ$ refers to strong majorization, $\succ_{w\log}$ denotes weak log-majorization, and $\bar{\boldsymbol{\lambda}}_I^{-1}{\in}\mathbb{R}_+^{2N\times1}$ and $\bar{\boldsymbol{\lambda}}_P^{-1}{\in}\mathbb{R}_+^{2N\times1}$ refer to vectors where each component is obtained from the inverse of the components in $\bar{\boldsymbol{\lambda}}_I$ and $\bar{\boldsymbol{\lambda}}_P$, respectively, in decreasing order.

Proof. See Appendix 2.D. ∎

The strong majorization relation in (2.24) accounts for (see Definition 2.5 in Appendix 2.A):

$$\sum_{i=1}^{n} \frac{1}{\bar{\lambda}_{I,i}} \leq \sum_{i=1}^{n} \frac{1}{\bar{\lambda}_{P,i}} \quad n = 1, ..., 2N - 1 \quad \text{and} \quad \sum_{i=1}^{2N} \frac{1}{\bar{\lambda}_{I,i}} = \sum_{i=1}^{2N} \frac{1}{\bar{\lambda}_{P,i}}. \quad (2.25)$$

On the other hand, the weak log-majorization relation in (2.24) is equivalent to (see Definition 2.7 in Appendix 2.A):

$$\prod_{i=1}^{n} \bar{\lambda}_{I,i} \geq \prod_{i=1}^{n} \bar{\lambda}_{P,i} \quad n = 1, ..., 2N \quad \text{and} \quad \sum_{i=1}^{n} \bar{\lambda}_{I,i} \geq \sum_{i=1}^{n} \bar{\lambda}_{P,i} \quad n = 1, ..., 2N. \quad (2.26)$$

Therefore, when receiving improper interference and using IGS in the MIMO P2P-I, the eigenvalues of the equivalent channel are more spread out having an equal or larger sum, an equal or larger product, and an equal sum of the inverses, as compared to receiving proper interference and using PGS. In other words, the arithmetic mean and the geometric mean of the eigenvalues are equal or larger, while the harmonic mean of the eigenvalues is equal.

The proof of the strong majorization result in Lemma 2.2 (i.e. $\bar{\boldsymbol{\lambda}}_I^{-1} \succ \bar{\boldsymbol{\lambda}}_P^{-1}$) is only valid for $M \geq N$ and the channel matrix $\mathbf{H}$ being full rank. Otherwise, the inverse in (2.80) (see Appendix 2.D) does not exist. This is due to the fact that if the desired signal subspace at the receiver has less than N dimensions (which happens either if $M < N$ or if the channel is rank deficient), then the strong majorization result is not valid since the properties of the noise-plus-interference in the noise-plus-interference subspace can be chosen arbitrarily without influencing the desired signal space. In case that either $M < N$ or the channel matrix $\mathbf{H}$ is rank deficient, then the weak-log majorization result in Lemma 2.2 is satisfied, i.e. $\bar{\boldsymbol{\lambda}}_I \succ_{w \log} \bar{\boldsymbol{\lambda}}_P$ (see [C1]), but the strong majorization result in Lemma 2.2 (which is key to demonstrate the strict superiority of IGS in what follows) is not further valid.

Now we show how Lemma 2.2 allows determining the superiority of IGS in terms of achievable rate in (2.19) and MSE in (2.20). The strong majorization result in (2.24) allows us to set the superiority in terms of achievable rate and MSE when uniform power allocation (UPA) is adopted, while the weak-log majorization result in (2.24) allows deriving the superiority in terms of achievable rate when optimal power allocation (OPA) is used.

2.4.1   Superiority for uniform power allocation

When adopting a UPA strategy, the solution for $\bar{\mathbf{P}} = \text{diag}(\bar{p}_1 \ldots \bar{p}_{2N})$ in (2.17) is given by:

$$\bar{p}_i = \bar{P} = P^{\max}/N, \forall i, \quad (2.27)$$

such that $\bar{\mathbf{p}} \circ \bar{\boldsymbol{\lambda}} = \bar{P} \bar{\boldsymbol{\lambda}}$. Thus, the optimal achievable rate in (2.19) and the optimal MSE in (2.20) become, respectively:

$$R^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}) = \frac{1}{2} \sum_{i=1}^{2N} \log_2 \left(1 + \bar{P}\bar{\lambda}_i\right), \quad (2.28)$$

$$\epsilon^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}) = \frac{1}{2} \sum_{i=1}^{2N} \frac{1}{1 + \bar{P}\bar{\lambda}_i}. \quad (2.29)$$

In the following theorems majorization theory is exploited over the relations in Lemma 2.2 to strictly quantify the achievable rate and MSE improvement under UPA strategy.

Theorem 2.1. *Assume a MIMO P2P-I, $M{\geq}N$, full rank channel matrix $\mathbf{H}$, receiving noise-plus-interference with a given covariance matrix $\mathbf{C_n}$. When applying the optimal signaling and the associated transmission scheme with UPA, the achievable rate R^{opt} in (2.19) is strictly increased in the improper interference scenario (i.e. $\tilde{\mathbf{C}}_{\mathbf{n}}{\neq}\mathbf{0}$) as compared to the proper interference scenario (i.e. $\tilde{\mathbf{C}}_{\mathbf{n}}{=}\mathbf{0}$):*

$$R^{\mathrm{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_I) - R^{\mathrm{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_P) \geq c_R^{\mathrm{upa}}\left(\|\bar{\boldsymbol{\lambda}}_I^{-1}\|^2 - \|\bar{\boldsymbol{\lambda}}_P^{-1}\|^2\right), \tag{2.30}$$

with a positive constant c_R^{upa}:

$$c_R^{\mathrm{upa}} = \frac{\bar{P}\lambda_{\min}^3\left(1 + 0.5\bar{P}\lambda_{\min}\right)}{2\ln(2)\left(1 + \bar{P}\lambda_{\min}\right)^2} > 0, \tag{2.31}$$

where $\lambda_{\min}$ denotes the minimum eigenvalue within the vector $\bar{\boldsymbol{\lambda}}_I$ and $\bar{P}$ is defined in (2.27). Since $\|\bar{\boldsymbol{\lambda}}_I^{-1}\|^2{\geq}\|\bar{\boldsymbol{\lambda}}_P^{-1}\|^2$, the rate gap (i.e. right-hand side in (2.30)) is strictly positive provided that $\bar{\boldsymbol{\lambda}}_I{\neq}\bar{\boldsymbol{\lambda}}_P$.

Proof. See Appendix 2.E. The proof departs from Lemma 2.2 and exploits majorization theory on strongly Schur-convex functions (see Definition 2.9 in Appendix 2.B). ∎

The rate gap increases as the difference among the squared 2-norm of the inverses of the eigenvalues increases (i.e. as the impropriety of the interference increases or, equivalently, as $\tilde{\mathbf{C}}_{\mathbf{n}}$ "increases", see (2.23)). Further, we can determine the rate gap behavior when varying the desired signal power (i.e. $\bar{P}$) for a fixed interference-plus-noise power level or, equivalently, when varying the signal-to-interference-plus-noise ratio (SINR):

- At high SINR, c_R^{upa} in (2.31) scales as $c_R^{\mathrm{upa}}{\sim}\frac{1}{4\ln(2)}\lambda_{\min}^2$ such that the rate gap in (2.30) is constant as the SINR increases. This means that, at the high SINR regime, the rates grow with the same slope with PGS and IGS but there is a constant difference among said achievable rates such that the use of IGS is always beneficial.
- At low SINR, c_R^{upa} in (2.31) scales as $c_R^{\mathrm{upa}}{\sim}\frac{1}{2\ln(2)}\bar{P}\lambda_{\min}^3$ and hence the rate gap in (2.30) increases with the SINR.

Theorem 2.2. *Assume a MIMO P2P-I, $M{\geq}N$, full rank channel matrix $\mathbf{H}$, receiving noise-plus-interference with a given covariance matrix $\mathbf{C_n}$. When applying the optimal signaling and the associated transmission scheme with UPA, the MSE ϵ^{opt} in (2.20) is strictly reduced in the improper interference scenario (i.e. $\tilde{\mathbf{C}}_{\mathbf{n}}{\neq}\mathbf{0}$) as compared to the proper interference scenario (i.e. $\tilde{\mathbf{C}}_{\mathbf{n}}{=}\mathbf{0}$):*

$$\epsilon^{\mathrm{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_P) - \epsilon^{\mathrm{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_I) \geq c_\epsilon^{\mathrm{upa}}\left(\|\bar{\boldsymbol{\lambda}}_I^{-1}\|^2 - \|\bar{\boldsymbol{\lambda}}_P^{-1}\|^2\right), \tag{2.32}$$

with a positive constant $c_\epsilon^{\mathrm{upa}}$:

$$c_\epsilon^{\mathrm{upa}} = \frac{\bar{P}\lambda_{\min}^3}{2\left(1 + \bar{P}\lambda_{\min}\right)^3} > 0. \tag{2.33}$$

The error gap (i.e. right-hand side in (2.32)) is strictly positive provided that $\bar{\boldsymbol{\lambda}}_I{\neq}\bar{\boldsymbol{\lambda}}_P$.

Proof. See Appendix 2.F. The proof departs from Lemma 2.2 and exploits majorization theory on strongly Schur-concave functions (see Definition 2.9 in Appendix 2.B). ∎

Similarly as in Theorem 2.1, the error gap increases as the difference among the squared 2-norm of the inverses of the eigenvalues increases (i.e. as the impropriety of the interference increases). The error gap behavior with respect to the SINR is as follows:

- At high SINR, the term $\frac{c_E}{\bar{P}^2}$ in (2.33) scales as $\frac{c_E}{\bar{P}^2} \sim \frac{1}{2}\bar{P}^{-2}$ such that the error gap in (2.32) decreases as the SINR increases. This is related to the fact that the rate gap in (2.30) is constant at high SINR, because due to the convex rate-MSE relation in (2.13) the rate gap being constant implies the error gap being reduced as the SINR increases.

- At low SINR, the term $\frac{c_E}{\bar{P}^2}$ in (2.33) scales as $\frac{c_E}{\bar{P}^2} \sim \frac{1}{2}\bar{P}\lambda_{\min}^3$ and therefore the error gap in (2.32) increases as the SINR increases.

Corollary 2.1. *If UPA is used in a MIMO P2P-I such that $M \geq N$ and $\mathbf{H}$ is full rank, the achievable rate R^{opt} in (2.19) and the MSE ϵ^{opt} in (2.20) are strictly outperformed when received interference is improper Gaussian distributed (i.e. $\tilde{\mathbf{C}}_{\mathbf{n}} \neq \mathbf{0}$).*

Proof. through Theorem 2.1 and Theorem 2.2. ∎

2.4.2   Superiority for optimal power allocation

When adopting an OPA strategy to maximize the achievable rate (i.e. $f(\bar{\mathbf{E}}) = -R = \frac{1}{2}\log_2|2\bar{\mathbf{E}}|$ in (2.15)), the optimal solution for $\bar{\mathbf{P}} = \mathrm{diag}(\bar{p}_1 \ldots \bar{p}_{2N})$ in (2.17) is given by [126]:

$$\bar{p}_i = \left(\mu - \bar{\lambda}_i^{-1}\right)^+, \forall i, \ \mu = \frac{1}{k}\left(2P^{\max} + \sum\nolimits_{i=1}^{k}\bar{\lambda}_i^{-1}\right), \tag{2.34}$$

where $k \leq 2N$ is the number of active streams after the water-filling solution in (2.34). Accordingly, let us denote by $\bar{\mathbf{p}}_P$ and $\bar{\mathbf{p}}_I$ the vectors with power allocation in (2.34) (ordered in decreasing order) for the proper and improper interference cases, respectively.

Due to the water-filling solution in (2.34), some streams can have a power equal to 0 and the number of active streams in the proper and improper interference cases might differ. Thus, as the power allocation in (2.34) depends on the eigenvalues, we cannot exploit majorization theory on strongly Schur-convex functions to guarantee a strict rate improvement with OPA.

Theorem 2.3. *Assume a MIMO P2P-I, $M \geq N$, full rank channel matrix $\mathbf{H}$, receiving noise-plus-interference with a given covariance matrix $\mathbf{C}_{\mathbf{n}}$. When applying the optimal signaling and the associated transmission scheme with OPA in (2.34) for maximum achievable rate, the achievable rate R^{opt} in (2.19) is equal or larger in the improper interference scenario (i.e. $\tilde{\mathbf{C}}_{\mathbf{n}} \neq \mathbf{0}$) as compared to the proper interference scenario (i.e. $\tilde{\mathbf{C}}_{\mathbf{n}} = \mathbf{0}$):*

$$R^{\mathrm{opt}}(\bar{\mathbf{p}}_I \circ \bar{\boldsymbol{\lambda}}_I) \geq R^{\mathrm{opt}}(\bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_P). \tag{2.35}$$

The equality in (2.35) is obtained if and only if $\bar{\boldsymbol{\lambda}}_I = \bar{\boldsymbol{\lambda}}_P$, i.e. $\tilde{\mathbf{C}}_{\mathbf{n}} = \mathbf{0}$, such that $\bar{\mathbf{p}}_I = \bar{\mathbf{p}}_P$ (see (2.34)) and hence $\bar{\mathbf{p}}_I \circ \bar{\boldsymbol{\lambda}}_I = \bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_P$.

Proof. See Appendix 2.G. The proof departs from Lemma 2.2 and is valid even if the number of streams with OPA in (2.34) is different in the proper and improper interference cases. ∎

There is, however, a particular case in which the strict superiority of IGS over PGS under OPA strategy can be demonstrated through the application of majorization theory tools. In

case that all steams are active both with IGS and PGS, i.e. $k_I=k_P=2N$ in (2.34), then the water-level μ in (2.34) is equal with IGS and PGS: $\mu=\mu_I=\mu_P$ because $\sum_{i=1}^{2N} \bar\lambda_{I,i}^{-1} = \sum_{i=1}^{2N} \bar\lambda_{P,i}^{-1}$ (see (2.25)). In this case, the optimal achievable rate in (2.19) becomes:

$$R^{\text{opt}}(\bar{\mathbf{p}} \circ \bar{\boldsymbol{\lambda}}) = \tfrac{1}{2} \sum_{i=1}^{2N} \log_2\left(\mu\bar\lambda_i\right). \tag{2.36}$$

Theorem 2.4. *Assume a MIMO P2P-I, $M \geq N$, full rank channel matrix $\mathbf{H}$, receiving noise-plus-interference with a given covariance matrix $\mathbf{C_n}$. When applying the optimal signaling and the associated transmission scheme with OPA in (2.34) for maximum achievable rate and all streams are active for IGS and PGS, the achievable rate R^{opt} in (2.19) is strictly increased in the improper interference scenario (i.e. $\tilde{\mathbf{C}}_{\mathbf{n}} \neq \mathbf{0}$) as compared to the proper interference scenario (i.e. $\tilde{\mathbf{C}}_{\mathbf{n}} = \mathbf{0}$):*

$$R^{\text{opt}}(\bar{\mathbf{p}}_I \circ \bar{\boldsymbol{\lambda}}_I) - R^{\text{opt}}(\bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_P) \geq c_R^{\text{opa}}\left(\|\bar{\boldsymbol{\lambda}}_I^{-1}\|^2 - \|\bar{\boldsymbol{\lambda}}_P^{-1}\|^2\right), \tag{2.37}$$

with a positive constant c_R^{opa}:

$$c_R^{\text{opa}} = \frac{\lambda_{\min}^2}{4\ln(2)} > 0. \tag{2.38}$$

The rate gap (i.e. right-hand side in (2.37)) is strictly positive provided that $\bar{\boldsymbol{\lambda}}_I \neq \bar{\boldsymbol{\lambda}}_P$.

Proof. See Appendix 2.H. The proof departs from Lemma 2.2 and exploits majorization theory on strongly Schur-convex functions (see Definition 2.9 in Appendix 2.B). ∎

Similarly as in Theorem 2.1, the rate gap increases as the difference among the squared 2-norm of the inverses of the eigenvalues increases (i.e. as the impropriety of the interference increases). However, result in Theorem 2.4 is only valid when the number of active streams under IGS and PGS is equal to $2N$, which occurs at medium/high SINR regimes. The rate gap under OPA strategy in (2.38) coincides with the one derived for UPA in (2.31) at high SINR.

2.4.3   Simulation results

First, let us show through simulations the strict rate and MSE improvement of IGS over PGS derived from Theorems 2.1 and 2.2 for UPA strategy. We evaluate the gains in the MIMO P2P-I for a single channel realization, when interference is either proper or improper Gaussian distributed. Fig. 2.4.(a) and Fig. 2.4.(b) depict the actual rate gap and the actual MSE gap versus the SINR, respectively, for an interference-to-noise ratio (INR) of INR=20 dB and antenna configuration $M=N=2$ (2×2). The figures also display the lower bound of the gaps presented in (2.30) for rate and in (2.32) for MSE. Further, the actual rate gap and the lower bound of gap derived in Theorem 2.4 under OPA strategy are included in Fig. 2.4.(a).

With regard to the rate (see Fig. 2.4.(a)), it can be observed that the rate gap is constant at high SINR and that it is increasing with the SINR at low SINR. Differently, in terms of the MSE (see Fig. 2.4.(b)), the error gap is decreasing with the SINR at high SINR and it is increasing with the SINR at low SINR. This corroborates the conclusions drawn from Theorems 2.1 and 2.2. The closer the lower bound of the gap is to the actual gap depends on the individual channel realizations, but it is observed to be tighter in the MSE than in the rate. Also, it is observed through simulations that the lower bound of the rate gap derived for UPA strategy is
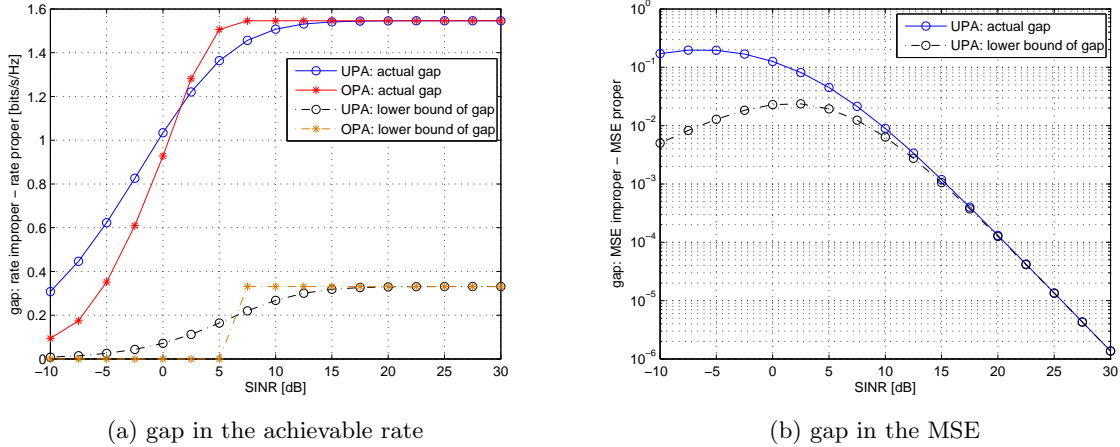
(a) gap in the achievable rate

(b) gap in the MSE

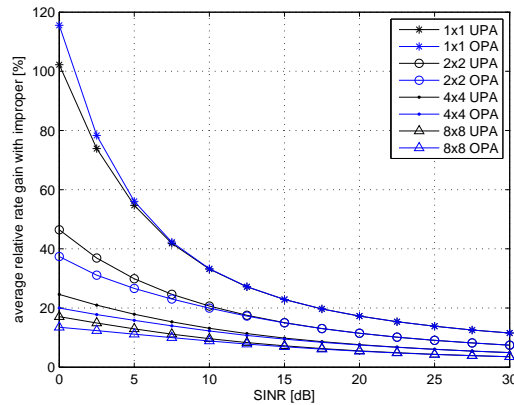Figure 2.4: Gap in the achievable rate and in the MSE vs. SINR of a specific MIMO P2P-I. 2×2, INR=20 dB. UPA and OPA strategies.

Figure 2.5: Average relative rate gain (in %) of IGS over PGS vs. SINR for MIMO P2P-I with different antenna configurations: 1×1, 2×2, 4×4, and 8×8. INR=20 dB. UPA and OPA strategies.

also valid for OPA strategy in all SINR regimes and antenna configurations (although this has been mathematically demonstrated only for medium/high SINR regimes, see Theorem 2.4).

Second, we show how the rate improvement of IGS over PGS scales as the number of transmit/receive antennas increases. 1000 channel realizations of the MIMO P2P-I are used to take statistic results. Fig. 2.5 displays the average of the relative rate gain (in %) of IGS over PGS versus the SINR for INR=20 dB and different antenna configurations: 1×1, 2×2, 4×4, and 8×8. The relative rate gains decrease as the number of transmit/receive antennas increase. This is due to the fact that IGS provides flexibility by splitting one dimension into two halves, which is more useful when the number of transmit/receive antennas is low, otherwise extra dimensions are already added to the system by adding antennas. Even though, it is important to recall that for the 8×8 antenna case, relative rate gains are still obtained for all SINR regimes (see Fig. 2.5). In addition, in Fig. 2.5 we can observe that the relative rate gains are larger for low SINR regimes whereby the MIMO P2P-I is highly limited by interference. However, as the SINR increases, rate gains are still obtained due to the use and reception of IGS.
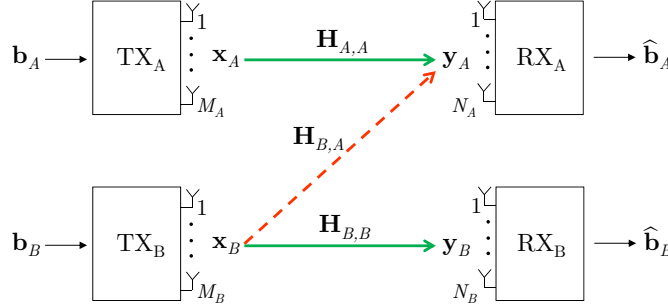
Figure 2.6: MIMO Z interference channel (MIMO Z-IC).

2.5 Application to the MIMO Z-IC

In this section we show how to exploit the benefits of using IGS for the MIMO Z-IC through a simple WLP design. The MIMO Z-IC is shown in Fig. 2.6, where one of the receivers (RX_B) just observes noise while the other (RX_A) receives interference. The signal model detailed in Section 2.1 applies for each of the two MIMO P2P-I links that appear in the MIMO Z-IC. In this regard, we add the subindex $(.)_A$ and $(.)_B$ to refer to the corresponding link, see Fig. 2.6.

Concerning RX_B that is receiving only proper noise, the optimal scheme at TX_B is a proper scheme (i.e. LP) given by $\mathbf{T}_{B,1}=\mathbf{T}_{B,1}^{\mathrm{opt}}$ and $\mathbf{T}_{B,2}=\mathbf{0}$. Such proper-based scheme would generate proper interference onto RX_A. On the other hand, RX_A could benefit in terms of achievable rate from the reception of improper interference from TX_B, as is demonstrated in Section 2.4, but the fact that TX_B transmits an improper Gaussian signal implies a degradation of the rate of RX_B due to the sub-optimality of the signaling scheme [105]. Therefore, a clear trade-off arises: with IGS the rate of RX_A could be increased but the rate of RX_B would be decreased.

Assuming that we can tolerate a certain achievable rate loss at RX_B, performance gains at RX_A would be guaranteed with IGS provided that TX_B uses the same transmit covariance matrix as in the optimum proper scheme for RX_B (i.e. $\mathbf{C}_{\mathbf{x}_B}=\mathbf{T}_{B,1}^{\mathrm{opt}}\mathbf{T}_{B,1}^{\mathrm{opt}H}$), such that the noise-plus-interference covariance matrix at RX_A ($\mathbf{C}_{\mathbf{n}_A}=\mathbf{H}_{B,A}\mathbf{C}_{\mathbf{x}_B}\mathbf{H}_{B,A}^H+\sigma_A^2\mathbf{I}$) is the same as in the case of proper interference. Therefore, due to the use of WLP with $\tilde{\mathbf{C}}_{\mathbf{x}_B}\neq\mathbf{0}$ (and hence $\tilde{\mathbf{C}}_{\mathbf{n}_A}=\mathbf{H}_{B,A}\tilde{\mathbf{C}}_{\mathbf{x}_B}\mathbf{H}_{B,A}^T\neq\mathbf{0}$), the achievable rate of RX_A would be increased by Theorem 2.1 for UPA and by Theorem 2.3 for OPA.

2.5.1 WLP scheme

In order to guarantee a rate improvement of RX_A we need to construct an improper-based scheme at TX_B (characterized by transmit precoders $\mathbf{T}_{B,1}$ and $\mathbf{T}_{B,2}$) that maintains the transmit covariance matrix used in the optimal proper-based scheme for RX_B (i.e. $\mathbf{C}_{\mathbf{x}_B}=\mathbf{T}_{B,1}^{\mathrm{opt}}\mathbf{T}_{B,1}^{\mathrm{opt}H}$). We propose deriving a practical WLP scheme at TX_B that does not degrade too much the performance of RX_B and generates improper interference onto RX_A by right-multiply the transmit precoder $\mathbf{T}_{B,1}^{\mathrm{opt}}$ by two scaling factors $\sqrt{\alpha}$ and $\sqrt{1-\alpha}$ and two unitary matrices $\mathbf{U}_1\in\mathbb{C}^{N\times N}$ and $\mathbf{U}_2\in\mathbb{C}^{N\times N}$. The WLP scheme is then created as follows:

$$\mathbf{T}_{B,1} = \sqrt{1-\alpha}\,\mathbf{T}_{B,1}^{\mathrm{opt}}\mathbf{U}_1, \qquad \mathbf{T}_{B,2} = \sqrt{\alpha}\,\mathbf{T}_{B,1}^{\mathrm{opt}}\mathbf{U}_2, \tag{2.39}$$

where $\alpha\in[0,0.5]$ is a measure of impropriety.

The improper-based scheme in (2.39) for TX_B has the same transmit covariance matrix than the optimal proper-based scheme for RX_B provided that $\mathbf{U}_1$ and $\mathbf{U}_2$ are unitary:

$$\mathbf{C}_{\mathbf{x}_B} = \mathbf{T}_{B,1}\mathbf{T}_{B,1}^H + \mathbf{T}_{B,2}\mathbf{T}_{B,2}^H = \mathbf{T}_{B,1}^{\text{opt}}\mathbf{T}_{B,1}^{\text{opt}H}, \qquad (2.40)$$

but the transmit pseudo-covariance matrix does not vanish to $\mathbf{0}$ and can be tunned with α:

$$\tilde{\mathbf{C}}_{\mathbf{x}_B} = \sqrt{\alpha(1-\alpha)}\mathbf{T}_{B,1}^{\text{opt}}(\mathbf{U}_1\mathbf{U}_2^T + \mathbf{U}_2\mathbf{U}_1^T)\mathbf{T}_{B,1}^{\text{opt}T}. \qquad (2.41)$$

Hence, by performing WLP in (2.39) at TX_B, the achievable rate of RX_B is degraded while the achievable rate of RX_A is ensured to be increased as compared to the use of $\mathbf{T}_{B,1}{=}\mathbf{T}_{B,1}^{\text{opt}}$ and $\mathbf{T}_{B,2}{=}\mathbf{0}$. This is possible because the WLP scheme in (2.39) allows maintaining the transmit covariance matrix $\mathbf{C}_{\mathbf{x}_B}$ independently of the values of α, $\mathbf{U}_1$, and $\mathbf{U}_2$, such that the properties in Section 2.4 apply.

Impropriety measure α

The range $\alpha{\in}[0, 0.5]$ is considered in (2.39) because $\tilde{\mathbf{C}}_{\mathbf{x}_B}$ has a symmetric shape with respect to α that is centered on $\alpha{=}0.5$, i.e. the system performance of the MIMO Z-IC is equivalent when selecting α or $\alpha'{=}1{-}\alpha$. The suitable selection of parameter α allows controlling the level of *impropriety*: if $\alpha{=}0$ the generated signal at TX_B is proper, while if $\alpha{>}0$ the generated signal at TX_B is improper and thus improper interference is generated towards RX_A in Fig. 2.6. When $\alpha{=}0.5$, the generated signal at TX_B achieves the maximum level of impropriety. Therefore, α is a measure of impropriety[11].

Spatial unitary matrices $\mathbf{U}_1$ and $\mathbf{U}_2$

Matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ are included in (2.39) so as to get different MIMO spatial structures of $\mathbf{T}_{B,1}$ and $\mathbf{T}_{B,2}$, although $\mathbf{U}_1{=}\mathbf{U}_2{=}\mathbf{I}$ is also an option. It is important to realize from (2.41) that the performance of the MIMO Z-IC is affected by the term $(\mathbf{U}_1\mathbf{U}_2^T{+}\mathbf{U}_2\mathbf{U}_1^T)$ rather than by the specific selection of $\mathbf{U}_1$ and $\mathbf{U}_2$. So, if we design $\mathbf{U}_1$ and $\mathbf{U}_2$ such that $\mathbf{U}_1\mathbf{U}_2^T{=}\mathbf{U}_2\mathbf{U}_1^T$, then a single choice would control the performance of the MIMO Z-IC. For $N{=}2$, [**C1**] proposes a design for $\mathbf{U}_1$ and $\mathbf{U}_2$ that satisfies such condition and depends on one parameter. For the general MIMO case, $\mathbf{U}_1$ and $\mathbf{U}_2$ are designed in order to get two totally opposite behaviors:

$$\text{Option 1 (identity):} \quad \mathbf{U}_1\mathbf{U}_2^T = \mathbf{U}_2\mathbf{U}_1^T = \mathbf{I} \quad \rightarrow \quad \mathbf{U}_1 = \mathbf{U}_2 = \mathbf{I}, \qquad (2.42)$$

$$\text{Option 2 (anti-identity):} \quad \mathbf{U}_1\mathbf{U}_2^T = \mathbf{U}_2\mathbf{U}_1^T = \mathbf{J} \quad \rightarrow \quad \mathbf{U}_1 = \mathbf{J}, \ \mathbf{U}_2 = \mathbf{I}, \qquad (2.43)$$

being $\mathbf{J}$ the anti-identity matrix[12] of size N. Both options have a trivial solution for unitary matrices $\mathbf{U}_1$ and $\mathbf{U}_2$, as shown in (2.42)-(2.43).

---

[11]The impropriety measure α defined in this work differs from the circularity coefficients defined in [65, Sect. 3.2], which are also a measure or impropriety.

[12]We refer with anti-identity matrix (also known as reflection matrix) to a square matrix where all the entries are 0 except those on the anti-diagonal that are equal to 1 (going from the lower left corner to the upper right corner).

2.5.2 Trade-off

Let us now show that the impropriety measure α allows trading-off on the achievable rates of the MIMO Z-IC.

Rate improvement at $\mathbf{RX}_A$

The achievable rate improvement of RX_A is guaranteed by Theorem 2.1 for UPA or by Theorem 2.3 for OPA. The exact rate improvement expression cannot be extracted, however, it is given by how much the eigenvalues in the improper interference case majorize the eigenvalues in the proper interference case, i.e. the difference in the majorization relations in (2.24). Said difference becomes larger as $\tilde{\mathbf{C}}_{\mathbf{n}_A}$ in (2.23) "increases" because the stronger the off-diagonal blocks in (2.23) are the more spread out the eigenvalues become (see Lemma 2.7 in Appendix 2.B).

Accordingly, the optimum value of the impropriety measure α to maximize the achievable rate of RX_A (R_A) is $\alpha{=}0.5$ (see (2.41)), as it allows "increasing" $\tilde{\mathbf{C}}_{\mathbf{x}_B}$ and, as consequence, "increasing" the pseudo-covariance matrix of the interference-plus-noise received at RX_A: $\tilde{\mathbf{C}}_{\mathbf{n}_A}{=}\mathbf{H}_{B,A}\tilde{\mathbf{C}}_{\mathbf{x}_B}\mathbf{H}_{B,A}^T$.

Rate degradation at $\mathbf{RX}_B$

Due to the use of WLP at TX_B, the achievable rate of RX_B is degraded as RX_B only receives proper noise (i.e. $\mathbf{C}_{\mathbf{n}_B}{=}\sigma_B^2\mathbf{I}$ and $\tilde{\mathbf{C}}_{\mathbf{n}_B}{=}\mathbf{0}$). When using the WLP scheme in (2.39) at TX_B, the achievable rate of RX_B (R_B) can be decomposed as (derived from [69]):

$$R_B = \underbrace{\log_2\left|\mathbf{I} + \mathbf{C}_{\mathbf{n}_B}^{-1}\mathbf{H}_{B,B}\mathbf{C}_{\mathbf{x}_B}\mathbf{H}_{B,B}^H\right|}_{R_B^{\mathrm{opt}}} + \underbrace{\frac{1}{2}\log_2\left|\mathbf{I} - \mathbf{C}_{\mathbf{y}_B}^{-1}\tilde{\mathbf{C}}_{\mathbf{y}_B}\mathbf{C}_{\mathbf{y}_B}^{-T}\tilde{\mathbf{C}}_{\mathbf{y}_B}^H\right|}_{R_B^{\mathrm{penalty}}\leq 0}, \qquad (2.44)$$

where $\mathbf{C}_{\mathbf{y}_B}{=}\mathbf{H}_{B,B}\mathbf{C}_{\mathbf{x}_B}\mathbf{H}_{B,B}^H{+}\sigma_B^2\mathbf{I}$ and $\tilde{\mathbf{C}}_{\mathbf{y}_B}{=}\mathbf{H}_{B,B}\tilde{\mathbf{C}}_{\mathbf{x}_B}\mathbf{H}_{B,B}^T$. The first term in (2.44) corresponds to the achievable rate of RX_B if the transmission scheme adopted at TX_B was PGS with linear precoding (the optimum one for RX_B, with $\tilde{\mathbf{C}}_{\mathbf{x}_B}{=}\mathbf{0}$). However, due to the use of an improper-based scheme in which $\tilde{\mathbf{C}}_{\mathbf{x}_B}{\neq}\mathbf{0}$, an achievable rate penalty is obtained at RX_B, which is given by the second term in (2.44) and is strictly negative. The fact that $R_B^{\mathrm{penalty}}{\leq}0$ can be shown by using the Fischer's inequality [85, Sect. 8.2] applied over the partitioned matrix $[\mathbf{C}_{\mathbf{y}_B}\ \tilde{\mathbf{C}}_{\mathbf{y}_B}; \mathbf{C}_{\mathbf{y}_B}^*\ \tilde{\mathbf{C}}_{\mathbf{y}_B}^*]$ that is positive semidefinite (see Theorem 2.5 in Appendix 2.A), from which it is concluded that $\left|\mathbf{I}{-}\mathbf{C}_{\mathbf{y}_B}^{-1}\tilde{\mathbf{C}}_{\mathbf{y}_B}\mathbf{C}_{\mathbf{y}_B}^{-T}\tilde{\mathbf{C}}_{\mathbf{y}_B}^H\right|{\leq}1$.

Consequently, the closer the impropriety measure α is to 0.5 (i.e. $\tilde{\mathbf{C}}_{\mathbf{y}_B}$ in (2.44) "increases", see (2.41)) the more the achievable rate of RX_B is reduced.

2.5.3 Sum-rate performance

The selection of the impropriety measure α has a clear trade-off on the sum-rate performance: R_B in (2.44) is decreasing while R_A is increasing with $\alpha{\in}[0,0.5]$. Let us consider $\tilde{\alpha}{=}\sqrt{\alpha(1-\alpha)}$ ${\in}[0,0.5]$, being $\tilde{\alpha}$ the parameter that impacts on the sum-rate performance of the MIMO Z-IC (see (2.41)). In this sense, R_B is decreasing and R_A is increasing with respect to $\tilde{\alpha}{\in}[0,0.5]$. Let us draw their rate expressions.

The rate expression R_B in (2.44) (using (2.40) and (2.41)) as a function of $\tilde{\alpha}$ is:

$$R_B(\tilde{\alpha}) = R_B^{\text{opt}} + \tfrac{1}{2}\log_2\left|\mathbf{I} - \tilde{\alpha}^2\mathbf{X}\right|, \tag{2.45}$$

where $\mathbf{X}{=}\mathbf{C}_{\mathbf{y}_B}^{-1}\mathbf{H}_{B,B}\mathbf{T}_{B,1}^{\text{opt}}\mathbf{U}\mathbf{T}_{B,1}^{\text{opt}T}\mathbf{H}_{B,B}^T\mathbf{C}_{\mathbf{y}_B}^{-T}\mathbf{H}_{B,B}^*\mathbf{T}_{B,1}^{\text{opt}*}\mathbf{U}^H\mathbf{T}_{B,1}^{\text{opt}H}\mathbf{H}_{B,B}^H$, $\mathbf{U}{=}\mathbf{U}_1\mathbf{U}_2^T{+}\mathbf{U}_2\mathbf{U}_1^T$.

On the other hand, the rate expression R_A (that is derived from (2.19) and (2.23)) as a function of $\tilde{\alpha}$ is:

$$R_A(\tilde{\alpha}) = \tfrac{1}{2}\sum_{i=1}^{2N}\log_2(1 + \bar{p}_i\bar{\lambda}_i), \tag{2.46}$$

$$\bar{\boldsymbol{\lambda}} = \text{eig}\left(\tfrac{1}{2}\bar{\mathbf{H}}_{A,A}^T\mathbf{Z}^{-H}\begin{bmatrix} \mathbf{C}_{\mathbf{n}_A} & \tilde{\alpha}\mathbf{Y} \\ \tilde{\alpha}\mathbf{Y}^* & \mathbf{C}_{\mathbf{n}_A}^* \end{bmatrix}^{-1}\mathbf{Z}^{-1}\bar{\mathbf{H}}_{A,A}\right), \tag{2.47}$$

where $\mathbf{Y}{=}\mathbf{H}_{B,A}\mathbf{T}_{B,1}^{\text{opt}}\mathbf{U}\mathbf{T}_{B,1}^{\text{opt}T}\mathbf{H}_{B,A}^T$ and $\mathbf{C}_{\mathbf{n}_A}{=}\mathbf{H}_{B,A}\mathbf{T}_{B,1}^{\text{opt}}\mathbf{T}_{B,1}^{\text{opt}H}\mathbf{H}_{B,A}^H{+}\sigma_A^2\mathbf{I}$.

Then, constrained to the WLP design proposed in Section 2.5.1, the optimum $\tilde{\alpha}{=}\sqrt{\alpha(1-\alpha)}$ in order to maximize the sum-rate performance of the MIMO Z-IC is obtained from the following optimization problem:

$$(\text{P}_{2,2}) : \quad \underset{\tilde{\alpha}}{\text{maximize}} \quad R_A(\tilde{\alpha}) + R_B(\tilde{\alpha}) \tag{2.48}$$
$$\text{subject to} \quad 0 \le \tilde{\alpha} \le 0.5.$$

The optimization problem $(\text{P}_{2,2})$ in (2.48) can be solved by setting the first order derivative to zero, and selecting the value of $\tilde{\alpha}$ from the candidate points given by the extreme of the domain (0 and 0.5) and the positive real roots of the first order derivative that lie within the domain and have a negative second order derivative (i.e. are maximum points).

Remark 2.1. *Through extensive numerical evaluations, we have observed that the optimal value of $\tilde{\alpha}$ always lies in the extremes of the domain ($\tilde{\alpha}^{\text{opt}}{=}0$ or $\tilde{\alpha}^{\text{opt}}{=}0.5$), which correspond to an impropriety measure $\alpha^{\text{opt}}{=}0$ (i.e. PGS) or $\alpha^{\text{opt}}{=}0.5$ (i.e. IGS with maximum impropriety level).*

Intuitively, for low interference regimes the optimal value is $\alpha^{\text{opt}}{=}0$ because increasing α leads to an increase of the rate of RX_A lower than the reduction of the rate of RX_B. On the contrary, for high interference regimes the optimal value is $\alpha^{\text{opt}}{=}0.5$ since reducing α leads to a reduction of the rate of RX_A larger than the increase of the rate of RX_B.

For the MIMO Z-IC, the optimal value of the transmitted power at TX_A corresponds always to the maximum one, as R_A is increasing with the available power and no interference is generated onto RX_B [122]. Therefore, the operating points of the proposed scheme are obtained by varying the values of the impropriety measure α from 0 to 0.5 (at maximum power of TX_B) and then varying the transmitted power at TX_B from maximum power to 0.

2.5.4   Simulation results

The proposed improper-based scheme is evaluated for a MIMO Z-IC scenario (as the one displayed in Fig. 2.6). It is assumed that both TXs have the same available power P^{max}. Signal-to-noise ratio is defined as $\text{SNR}{=}P^{\text{max}}/\sigma^2$ and signal-to-interference ratio by $\text{SIR}{=}1/\eta$, where factor $\eta{\geq}0$ denotes the average ratio between interfering and direct channel attenuations. Channels are modeled through a Rayleigh distribution, such that $\mathbf{H}_{A,A} \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$, $\mathbf{H}_{B,B}{\sim}\mathcal{CN}(\mathbf{0},\mathbf{I})$, and $\mathbf{H}_{B,A}{\sim}\mathcal{CN}(\mathbf{0},\eta\mathbf{I})$.

Achievable rate region

The figures below show the achievable rate region of the MIMO Z-IC for a specific channel realization with SNR=10 dB and SNR=20 dB, $\eta=1$, $M=N=2$ (2×2). The MIMO channel matrices are:

$$\mathbf{H}_{A,A} = \begin{bmatrix} 1.01e^{-j174.9} & 0.74e^{j152.8} \\ 0.86e^{-j55.5} & 0.82e^{j166.7} \end{bmatrix},$$

$$\mathbf{H}_{B,B} = \begin{bmatrix} 0.49e^{j162.3} & 1.30e^{-j101.0} \\ 0.70e^{-j43.8} & 0.46e^{j9.8} \end{bmatrix},$$

$$\mathbf{H}_{B,A} = \begin{bmatrix} 1.16e^{-j132.6} & 0.90e^{-j88.5} \\ 0.93e^{-j141.5} & 0.70e^{j67.1} \end{bmatrix}. \tag{2.49}$$

The following transmission schemes are evaluated:

- **proper (LP)**: TXs use LP,
- **improper (WLP) id: varying α**: TXs use WLP, whereby TX_B uses the improper-based scheme in (2.39) with different values of the impropriety measure α and the identity solution for $\mathbf{U}_1$ and $\mathbf{U}_2$ in (2.42),
- **improper (WLP) anti-id: varying α**: TXs use WLP, and TX_B uses the improper-based scheme in (2.39) with different values of the impropriety measure α and the anti-identity solution for $\mathbf{U}_1$ and $\mathbf{U}_2$ in (2.43).

Fig. 2.7 displays the achievable rate region when UPA in (2.27) and OPA in (2.34) are used. Another way to achieve a similar behavior than the one given by the proposed improper-based scheme (i.e. reduce rate of RX_B to increase rate of RX_A) is by reducing the power used by TX_B ($P \le P^{\max}$), which can be applied both for LP and WLP cases. Such performance results are depicted in the figures with the label '*varying P*'. Note that the optimal solution for TX_A corresponds to always using the maximum power $P^{\max}$, as it does not interfere RX_B.
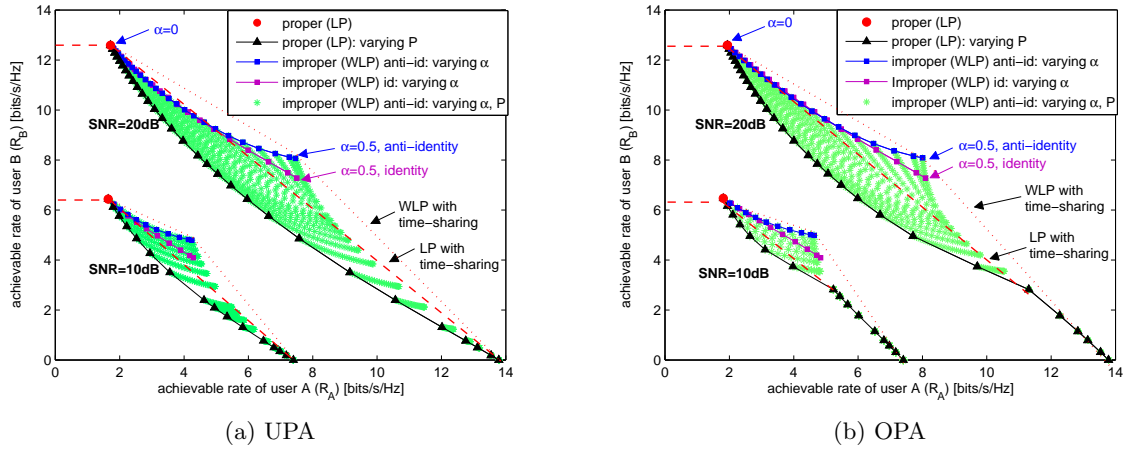

(a) UPA   (b) OPA

Figure 2.7: Achievable rate region of a MIMO Z-IC with $\mathbf{H}_{A,A}$, $\mathbf{H}_{B,B}$, $\mathbf{H}_{B,A}$ in (2.49). 2×2, SNR=10 dB and SNR=20 dB.

By comparing 'improper (WLP) varying α' with respect to 'proper (LP)' at $P=P^{\max}$, it is verified that the rate of RX_A is always increased when received interference is improper (as is demonstrated by Theorems 2.1 and 2.3) at the expenses of a reduced rate for RX_B. Accordingly, the proposed WLP scheme is beneficial when interference is non-negligible, such that the gain at RX_A is significant. The choice of $\mathbf{U}_1$ and $\mathbf{U}_2$ (i.e. identity or anti-identity) does not have a significant impact on the rate of RX_A, however it affects the rate of RX_B, being the anti-identity (Option 2 in (2.43)) shown to be the best solution. Therefore, in interference-limited scenarios, the proposed improper-based scheme allows trading in transmission fairness mainly through the impropriety measure α without adjusting the transmitted power. Even with time-sharing, the improper-based scheme outperforms the proper-based scheme, as is shown with dashed lines in Fig. 2.7. To conclude, it can be observed that the achievable rate region is enlarged when using IGS through WLP and, additionally, both system sum-rate (i.e. R_A+R_B) and system fairness (i.e. $\min(R_A, R_B)$) are improved for this channel realization.

**Averaged sum-rate and min-rate performance**

The figures in the following display simulation results averaged over 1000 random channel realizations when varying η for SNR=10 dB, OPA, and different antenna configurations. The schemes that are evaluated are:

- **proper (LP)**: TXs use LP (i.e. $\alpha=0$ at TX_B in (2.39)) and maximum power.
- **improper (WLP) $\alpha=0.5$**: TXs use WLP and maximum power, whereby for TX_B: $\alpha=0.5$ (maximum impropriety level) and the anti-identity solution for $\mathbf{U}_1$ and $\mathbf{U}_2$ (i.e. Option 2 in (2.43)) are used in (2.39).
- **only TX_A (LP)**: TX_A transmits through optimal LP design, and TX_B is turned off.
- **optimal scheme SISO**: optimal sum-rate scheme presented in [122] for the SISO Z-IC. It is valid for $M=N=1$ and encompasses PGS and IGS solutions.
- **WMMSE**: weighted minimum MSE algorithm presented in [127] for sum-rate maximization in MIMO IC. To adopt IGS, the complex-valued MIMO Z-IC is transformed into an equivalent double-sized real-valued MIMO Z-IC where the WMMSE algorithm applies. It can be employed for any antenna configuration, encompasses PGS and IGS solutions, and converges to a local optimum.

Fig. 2.8 displays the sum-rate (i.e. R_A+R_B) and the min-rate (i.e. $\min(R_A, R_B)$) versus η for 1×1. Fig. 2.9 and Fig. 2.10 show the same for antenna configurations 2×2 and 4×4, respectively, so as to see how the system performance scales with the number of transmit/receive antennas.

Note that 'optimal scheme SISO' and 'WMMSE' schemes are used as benchmarks in terms of sum-rate. They involve larger complexity, coordination among TXs and knowledge of all channel matrices. On the other hand, 'proper (LP)', 'improper (WLP) $\alpha=0.5$', and 'only TXA' do only require knowledge of the direct channels and no coordination among TXs is required.

Let us compare the proposed 'improper (WLP) $\alpha=0.5$' with 'proper (LP)'. In the 1×1 antenna case (see Fig. 2.8), it can be observed that the average sum-rate of the system is increased with IGS for values of $\eta\geq1.25$, but for values of $\eta\geq0.5$ the use of IGS starts to be important in some channel realizations (as shown by 'optimal scheme SISO'). On the other hand, the average fairness of the system is increased with IGS for values of $\eta\geq0.5$. In the 2×2 antenna case (see Fig. 2.9), the average sum-rate is improved with IGS for $\eta\geq1.5$ and the average fairness
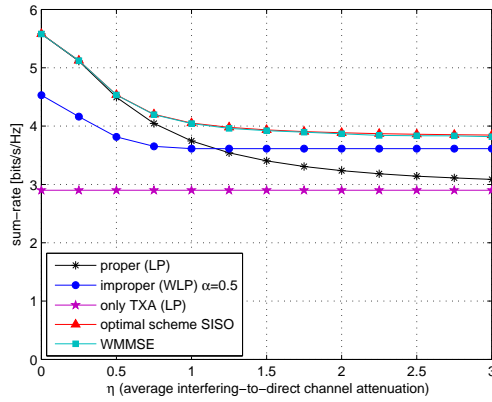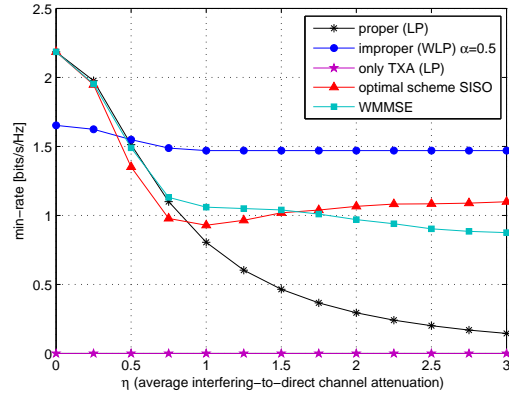
(a) sum-rate   (b) min-rate

Figure 2.8: Average sum-rate and min-rate (in bits/s/Hz) for MIMO Z-IC vs. η. 1×1, OPA, SNR=10 dB.
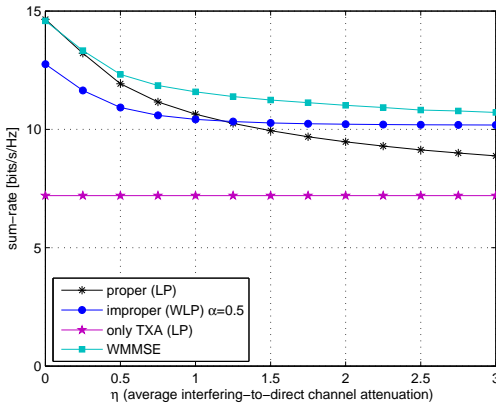

(a) sum-rate   (b) min-rate
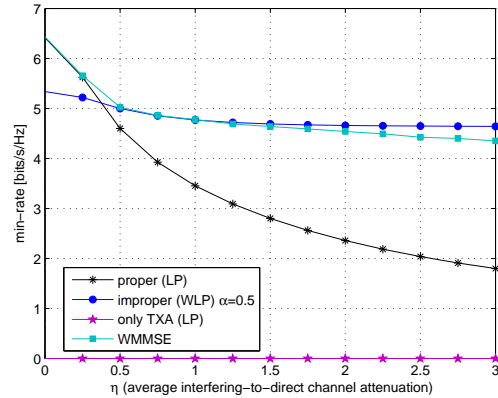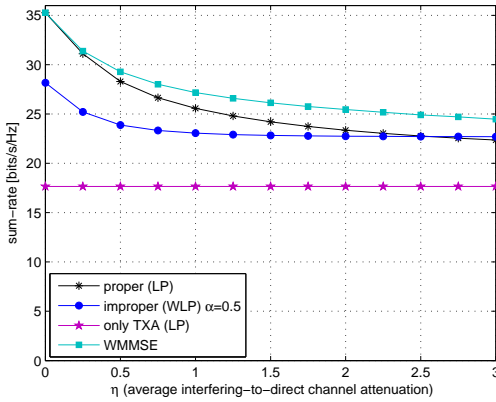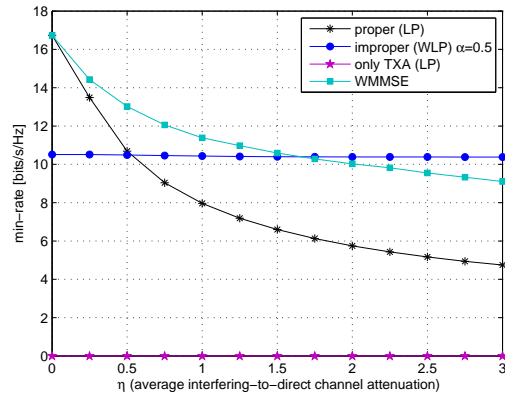
Figure 2.9: Average sum-rate and min-rate (in bits/s/Hz) for MIMO Z-IC vs. η. 2×2, OPA, SNR=10 dB.

(a) sum-rate   (b) min-rate

Figure 2.10: Average sum-rate and min-rate (in bits/s/Hz) for MIMO Z-IC vs. η. 4×4, OPA, SNR=10 dB.

for $\eta \geq 0.5$. Finally, in the 4×4 antenna case (see Fig. 2.10), the average sum-rate is improved for $\eta \geq 2.75$ and the average fairness for $\eta \geq 0.75$.

Therefore, the use of IGS is beneficial when interference exceeds a certain threshold, where the proposed 'improper (WLP) α=0.5' allows trading in transmission fairness and system sum-rate. Such threshold increases as the number of transmit/receive antennas increase, due to the fact that the larger is the number of transmit/receive antennas the larger are the dimensions of the system and the lower are the gains of the flexibility provided by IGS.

In the SISO case (1×1), the sum-rate performance of the proposed 'improper (WLP) α=0.5' is close to the optimal sum-rate performance of 'optimal scheme SISO' (see Fig. 2.8.(a)), while the average min-rate is improved (see Fig. 2.8.(b)). Recall that the proposed scheme has much less complexity when α is fixed, as it does not require knowledge of the interfering channel. Further, it can be adopted in any MIMO system such that $M \geq N$.

2.6   Application to MIMO HCNs

In this section we show how to exploit the benefits of using IGS in HCNs through the simple WLP design presented in Section 2.5. Let us focus on the deployment shown in Fig. 2.2 with one MeNB and multiple SeNBs. We assume that on a given time/frequency resource the MeNB serves a single user (called MUE) and each SeNB serves a single user (called SUE). Then, the interference channel towards each SUE can be modeled by the MIMO Z-IC. More specifically, the HCN in Fig. 2.2 is related to the MIMO Z-IC in Fig. 2.6 as follows:

- the MUE (that just receives noise) corresponds to RX_B in Fig. 2.6, while
- each SUE (that might receive interference due to the active transmission of the MeNB on the same time/frequency resource) corresponds to RX_A in Fig. 2.6.

Moreover, due to the deployment of multiple SeNBs within the MeNB coverage area (see Fig. 2.2), there will appear as many MIMO Z-IC as the number of SUEs being served.

The interesting part is that the proposed improper-based scheme with WLP in Section 2.5.1 for TX_B (i.e. the MeNB) when using maximal IGS (i.e. α=0.5 in (2.39)) can be applied independently of the interfering channel, as it is based on generating improper interference towards the SUEs (i.e. RX_A) but does not depend on the specific SUE that is selected. Accordingly, the proposed improper-based scheme can be easily applied to multi-tier HCN deployments defined in 3GPP LTE-A [17] so as to increase the achievable rate of the SUEs. Furthermore, as multiple MIMO Z-IC may appear, large gains could be obtained with the use of the proposed improper-based scheme because all SUEs would be guaranteed to increase their rate according to Theorem 2.1 for UPA and to Theorem 2.3 for OPA.

2.6.1   Practical implementation in 3GPP LTE-A compliant HCNs

The proposed WLP design can be easily applied to multi-tier HCN deployments defined in 3GPP LTE-A [2]. We have to take into account that SeNBs resources are often underutilized, in which case the MeNB should normally transmit with PGS or the rate of its MUE would be degraded. Accordingly, a procedure to indicate the activation of IGS usage in each frame is required.
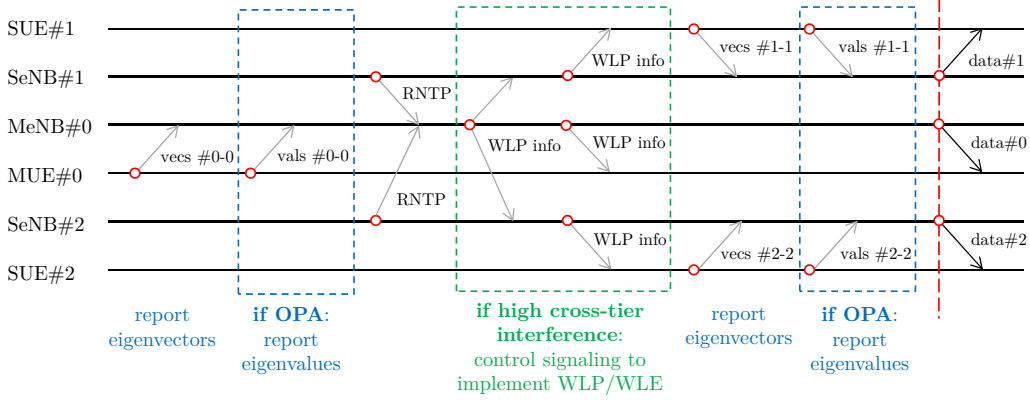
Figure 2.11: Example of control signaling for IGS activation in a HCN consisting of 1 MeNB and 2 SeNBs. 2 parallel MIMO Z-IC appear.

The procedure for activation of IGS usage is done through the control-plane, as illustrated in Fig. 2.11. First of all, the MeNB acquires knowledge of the eigenvectors (and eigenvalues for OPA) of the equivalent MIMO channel towards its serving MUEs (i.e. $\mathbf{H}_{B,B}^H \mathbf{C}_{\mathbf{n}_B}^{-1} \mathbf{H}_{B,B}$). This can be done using the precoding matrix indicator (PMI), the rank indicator (RI) and channel quality indicator (CQI) in LTE-A [22]. Then, relative narrow-band transmit power (RNTP)[13] is sent from the SeNBs towards the MeNB, such that the MeNB is aware of the resources employed by the SeNB and high cross-tier interference will be generated. Based on this knowledge, the MeNB changes the transmission scheme on such resources where any of the SeNB will transmit. The scheme is changed to WLP in (2.39) by selecting α such that the rate requirements of the serving MUE on such resource are satisfied. If the selected α is larger than 0 (improper is selected) on a given resource, the neighbor SeNBs can transmit with WLP in order to improve the rate of its associated SUEs, and all UEs should implement WLE for data decoding. But before data decoding, the SeNBs should acquire knowledge of the eigenvectors (and eigenvalues for OPA) of the equivalent real-valued MIMO channel towards its serving SUEs taking into account the interference received from the MeNB (i.e. $\bar{\mathbf{H}}_{A,A}^H \mathbf{C}_{\bar{\mathbf{n}}_A}^{-1} \bar{\mathbf{H}}_{A,A}$).

With this procedure, the signaling among TXs is minimum and no further coordination is required. Only local information is required at MeNB/SeNBs and it is not needed to acquire explicit knowledge about the interfering channel matrices either at BSs or users. The local information required at MeNB and SeNBs to implement the proposed WLP design is:

- *MeNB*: associated MUE have to report the eigenvectors of $\mathbf{H}_{B,B}^H \mathbf{C}_{\mathbf{n}_B}^{-1} \mathbf{H}_{B,B}$ for UPA, or the eigenvectors and the eigenvalues of $\mathbf{H}_{B,B}^H \mathbf{C}_{\mathbf{n}_B}^{-1} \mathbf{H}_{B,B}$ for OPA. From such reporting, $\mathbf{T}_{B,1}^{\text{opt}}$ can be selected at MeNB, and WLP scheme in (2.39) can be used to get $\mathbf{T}_{B,1}$ and $\mathbf{T}_{B,2}$.

- *SeNB*: associated SUE have to report the eigenvectors of $\bar{\mathbf{H}}_{A,A}^T \mathbf{C}_{\bar{\mathbf{n}}_A}^{-1} \bar{\mathbf{H}}_{A,A}$ for UPA, or the eigenvectors and the eigenvalues of $\bar{\mathbf{H}}_{A,A}^T \mathbf{C}_{\bar{\mathbf{n}}_A}^{-1} \bar{\mathbf{H}}_{A,A}$ for OPA. From such reporting, $\bar{\mathbf{T}}_A^{\text{opt}}$ in (2.17) can be constructed at SeNB to obtain the equivalent widely linear precoders $\mathbf{T}_{A,1}^{\text{opt}}$ and $\mathbf{T}_{A,2}^{\text{opt}}$ using (2.6). This has to be done at all SeNBs.

---

[13]RNTP is an indicator per resource block (RB) signaled to neighboring eNBs, indicating the maximum anticipated DL transmit power level per RB [2].

2.6.2   Simulation results

The proposed improper-based scheme is evaluated through Monte Carlo simulations in a simulator compliant with 3GPP LTE-A specifications in [104]. The network consists of a multi-tier deployment where MeNBs and SeNBs use the same carrier frequency of 2 GHz with 10 MHz bandwidth. It is used the Small Cell Scenario 1, following deployment and simulation parameters specified in [104]. The deployment consists of 1 MeNB and 4 SeNBs that are uniformly distributed within the macrocell area. 60 UEs are deployed per macrocell area, being 2/3 of them placed near the SeNBs, and the remaining UEs are uniformly distributed within the MeNB coverage area. All UEs are placed outdoor. ITU Urban Macro and ITU Urban Micro models with 3D distance are used for path loss and shadowing modeling for MeNB-UE and SeNB-UE links, respectively. For frequency-selective fading modeling, the typical urban model is used. Maximum transmit power at MeNB is $P_{\text{MeNB}}^{\max}$=46 dBm. At SeNBs we use two different maximum power values for simulation: either $P_{\text{SeNB}}^{\max}$=24 dBm or $P_{\text{SeNB}}^{\max}$=15 dBm. Antenna gains are 17 dBi at MeNB, 5 dBi at SeNB, and 0 dBi at UE. Noise spectral density is -174 dBm/Hz. The number of antennas is $M_{\text{MeNB}}=M_{\text{SeNB}}$=2, and N_{UE}=2 for all UEs.

Cell selection at each UE is based on best downlink reference signal receive power (RSRP) [104]. A range extension bias (REB) is added at the RSRP received from each SeNB in order to expand its cell-range and offload more UEs to the SeNBs [23]. REB=10 dB is used for $P_{\text{SeNB}}^{\max}$=24 dBm and REB=15 dB is employed for $P_{\text{SeNB}}^{\max}$=15 dBm, such that around 73% of the UEs (in mean over different deployments) are offloaded to the SeNBs in both cases.

It is assumed that the LTE-A frame is composed of 8 downlink subframes and 2 uplink subframes [17], but only downlink is evaluated in the sequel. Full-load traffic model is used, where all UEs in the network have packets to be received. For each frame, the UEs associated to the same MeNB or SeNB are uniformly distributed in frequency domain among the available resource blocks (RBs), such that intra-cell interference is removed and only inter-cell interference remains (both cross-tier and co-tier interference). The power available at each MeNB or SeNB is uniformly distributed among the RBs where the UEs have been scheduled.

The following techniques are evaluated on each RB:

- **time-sharing**: time-sharing solution among MeNB and SeNBs. Different muting ratios are used, represented by $X/8$ in the figures, which means that all SeNBs transmit in X subframes, while MeNB transmits in the $(8-X)$ subsequent subframes.

- **eICIC ABS**: eICIC technique defined in LTE-A with time-division muting based on almost blank subframes (ABS) [56]. Different muting ratios are used, represented by $X/8$ in the figures, which means that the MeNB is muted X subframes where the SeNBs transmit, while in the $(8-X)$ subsequent subframes all MeNB and SeNBs transmit.

- **FR proper (LP)**: full-reuse of the frequency band and time slots for MeNB and SeNBs, using proper-based schemes (i.e. LP) at MeNB and SeNBs with OPA in (2.34).

- **FR improper (WLP)**: full-reuse of the frequency band and time slots for MeNB and SeNBs, using improper-based schemes (i.e. WLP) at MeNB and SeNBs with OPA in (2.34). The proposed scheme in (2.39) is employed at MeNB with α=0.5 and the anti-identity solution for $\mathbf{U}_1$ and $\mathbf{U}_2$ in (2.43).

Fig. 2.12 displays the evaluated transmission schemes in a frequency/time grid corresponding to 1 RB and 8 downlink subframes, showing the subframes in which MeNB and SeNBs are allowed to transmit.

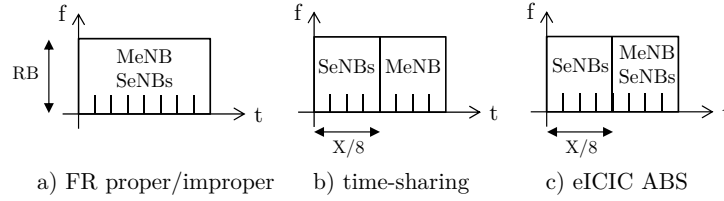a) FR proper/improper   b) time-sharing   c) eICIC ABS

Figure 2.12: Transmission schemes considered for HCNs, shown in a frequency/time grid corresponding to 1 RB and 8 downlink subframes.

The performance indicator is the user throughput (UT) measured in Mbits/s. Let us remark that all the evaluated schemes have as main objective dealing with the cross-tier interference, either with time-sharing or with an efficient spatial precoding design. However, the intra-tier interference can degrade the system performance in some cases depending on the deployment and the system parameters. In the case of a full-reuse, the cross-tier interference from the SeNB to the MUEs can also be prejudicial in some deployments.

Fig. 2.13 displays the cumulative distribution function (CDF) of the achievable rates per RB of the MUEs and the SUEs (in bits/s/Hz) separately for the case of $P_{\text{SeNB}}^{\max}$=24 dBm. In the 'eICIC ABS' technique, as the muting ratio ($X/8$) increases, the achievable rate of the MUEs is degraded while the achievable rate of the SUEs is improved because the cross-tier interference is eliminated. As compared to the baseline 'FR proper (LP)', the proposed 'FR improper (WLP)' has a similar behavior than the 'eICIC ABS': it degrades the achievable rate of the MUEs so as to improve the achievable rate of the SUEs (as is also shown in Section 2.5.4). However, it can be observed that, among all techniques, 'FR improper (WLP)' achieves the fairest performance when considering all the MUEs and the SUEs.

Fig. 2.14 displays the 5%-tile UT vs. the mean UT when considering all UEs for the case of (a) $P_{\text{SeNB}}^{\max}$=24 dBm and (b) $P_{\text{SeNB}}^{\max}$=15 dBm. The use of 'time-sharing' or 'eICIC ABS' define an oval area where the system can operate based on the variation of the muting ratio ($X/8$). As the muting ratio increases, the mean UT is increased because the SeNBs have more subframes
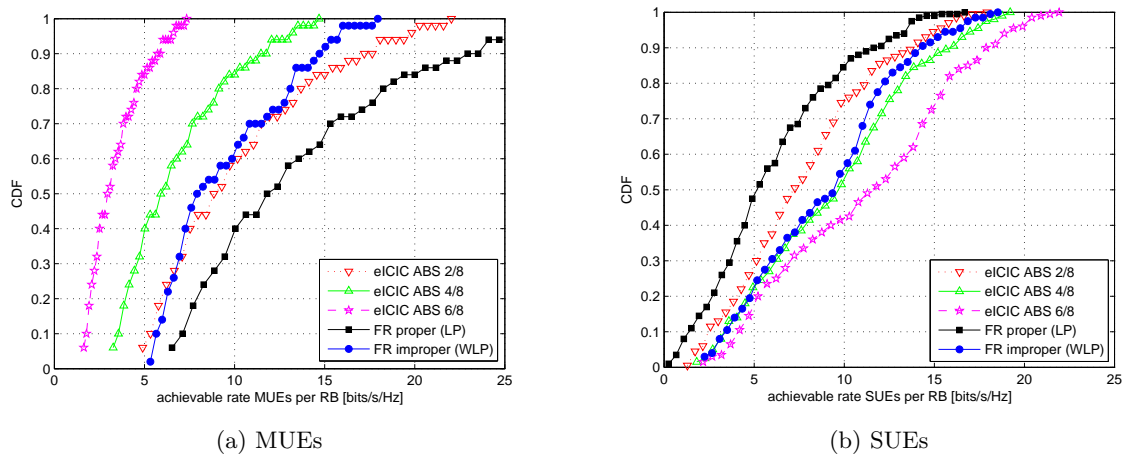

(a) MUEs   (b) SUEs

Figure 2.13: CDF of the achievable rates (in bits/s/Hz) per RB of the MUEs and the SUEs, separately, for $P_{\text{SeNB}}^{\max}$=24 dBm.

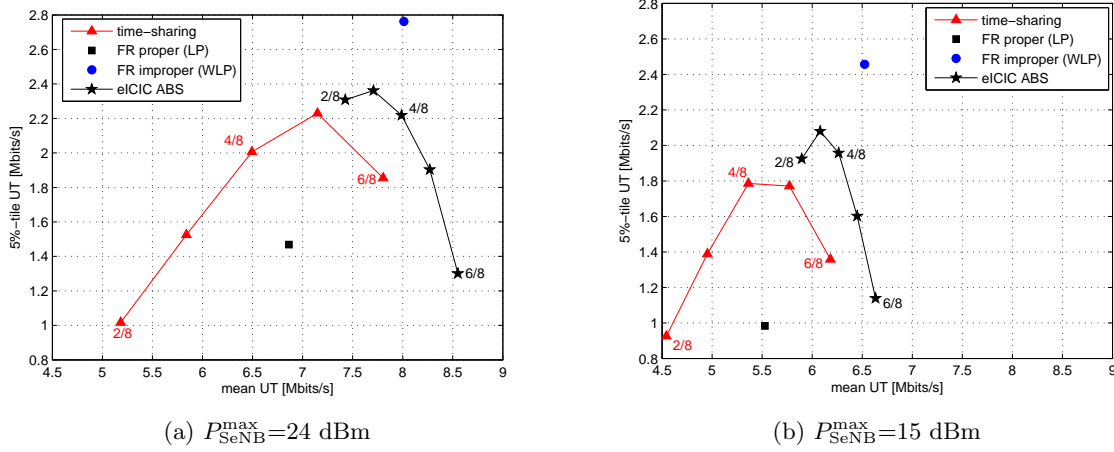(a) $P_{\text{SeNB}}^{\max}$=24 dBm

(b) $P_{\text{SeNB}}^{\max}$=15 dBm

Figure 2.14: 5%-tile UT vs. mean UT (in Mbits/s) for different SeNB maximum transmit power ($P_{\text{SeNB}}^{\max}$). Different muting ratios are displayed for time-sharing and eICIC solutions.

available with less interference (see Fig. 2.12). On the contrary, there is a limit in terms of fairness (or 5%-tile UT) as when the muting ratio increases the resources devoted for the MUEs decrease and there is a switching point in which the fairness passes from being limited by SUEs to being constrained by the MUEs. However, it is important to realize that, although the 'FR proper (LP)' lies inside the oval areas delimited by the 'time-sharing' or 'eICIC ABS' solutions, the proposed 'FR improper (WLP)' allows to go out of these areas and improve both the system fairness and the system sum-rate performance.

When reducing the SeNB transmit power to $P_{\text{SeNB}}^{\max}$=15 dBm (see Fig. 2.14.(b)), similar conclusions can be extracted as for $P_{\text{SeNB}}^{\max}$=24 dBm. However, the 5%-tile and the mean UT gains of 'FR improper (WLP)' are larger than in the case of $P_{\text{SeNB}}^{\max}$=24 dBm, owing to the fact that when reducing the power at SeNBs the interfering-to-direct ratio for the SUEs increases (same interfering power from MeNB, but less direct power from SeNB) and hence the use of IGS provides larger system performance gains.

In addition to the system fairness and performance gains, 'FR improper (WLP)' has the following advantages as compared to 'time-sharing' or 'eICIC ABS' solutions:

- there is no need to optimize the muting ratio, and
- the on/off switching of the MeNB does not need to be tackled.

Furthermore, the proposed 'FR improper (WLP)' does not add any complexity in terms of implementability, either in overhead or in additional information required.

2.7   Conclusions

This chapter exploits majorization theory to formally quantify the benefits of improper Gaussian signaling in the MIMO P2P-I, and then applies the derived concepts to the MIMO Z-IC and to co-channel heterogeneous cellular networks through an efficient design of widely linear transceivers.

First, for the MIMO P2P-I, we show that the use of IGS allows:

- strictly improving the achievable rate and the mean square error when uniform power allocation is performed and
- improving the achievable rate when optimal power allocation for maximum rate is adopted, whereby the strict superiority is proved for a particular case.

Second, for the MIMO Z-IC, we propose a practical improper-based signaling scheme through a simple WLP design, which allows:

- improving the achievable rate of the most impaired user and
- improving the system fairness and controlling the sum-rate performance through a single parameter (the impropriety measure) in interference-limited conditions.

The proposed improper-based scheme is useful provided that the interference level in the MIMO Z-IC is high, in which case the performance gains are significant.

Finally, such benefits are applied to co-channel HCNs where multiple MIMO Z-IC appear with high interference levels. In this scenario, 3GPP-compliant simulations show that the 5%-tile user throughput and the mean user throughput can be improved with the proposed improper-based scheme as compared to conventional time-sharing solutions, proper-based schemes, and the well-known eICIC technique, due to the fact that the proposed scheme can cope with a full reuse of the frequency bands and time slots while providing enough flexibility to combat the predominant cross-tier interference in co-channel HCNs.

Appendices

2.A Preliminaries for Improper Gaussian Random Vectors

In this section, some basic notions of improper random vectors are introduced. See [65] for more details.

Definition 2.1 ([65]). *Given a zero-mean complex random vector $\mathbf{x} \in \mathbb{C}^{n \times 1}$, $\mathbf{C_x} = \mathbb{E}[\mathbf{x}\mathbf{x}^H]$ denotes the covariance matrix of $\mathbf{x}$ and $\tilde{\mathbf{C}}_\mathbf{x} = \mathbb{E}[\mathbf{x}\mathbf{x}^T]$ refers to the pseudo-covariance matrix of $\mathbf{x}$.*

By definition, it is easy to check that the covariance matrix $\mathbf{C_x}$ is Hermitian and positive semidefinite while the pseudo-covariance matrix $\tilde{\mathbf{C}}_\mathbf{x}$ is symmetric.

Definition 2.2 ([128] - proper). *A complex random vector $\mathbf{x}$ is called proper if its pseudo-covariance matrix $\tilde{\mathbf{C}}_\mathbf{x}$ vanishes to a zero matrix, otherwise it is called improper.*

A more restrictive definition than properness is known as circularly symmetric.

Definition 2.3 ([65] - circularly symmetric). *A complex random vector $\mathbf{x}$ is circularly symmetric if its distribution is rotationally invariant, i.e. if $\mathbf{x}$ and $\hat{\mathbf{x}} = \mathbf{x}e^{j\alpha}$ have the same distribution for any real value α.*

For a circularly symmetric random vector $\mathbf{x}$, we have $\tilde{\mathbf{C}}_\mathbf{x} = \tilde{\mathbf{C}}_{\hat{\mathbf{x}}} = \mathbb{E}[\hat{\mathbf{x}}\hat{\mathbf{x}}^T] = e^{j2\alpha}\tilde{\mathbf{C}}_\mathbf{x}$, which implies $\tilde{\mathbf{C}}_\mathbf{x} = \mathbf{0}$. Thus, circularity implies properness, but the converse is not true in general. Nevertheless, if $\mathbf{x}$ is a zero-mean Gaussian random vector, then properness and circularity are equivalent [65], as given by the following lemma.

Lemma 2.3 ([65]). *A complex zero-mean Gaussian random vector $\mathbf{x}$ is circularly symmetric if and only if it is proper.*

For example, the commonly adopted assumption that the noise vector is zero-mean circularly symmetric complex Gaussian (CSCG) is equivalent to say that the noise vector is a proper Gaussian random vector.

For an arbitrary zero-mean complex random vector $\mathbf{x} \in \mathbb{C}^{n \times 1}$, two mathematical tools have been used in the recent literature to represent them: the *composite real representation*, whereby the real part and the imaginary part of $\mathbf{x}$ are separated: $\bar{\mathbf{x}} = \left[\Re\{\mathbf{x}\}^T \; \Im\{\mathbf{x}\}^T\right]^T \in \mathbb{R}^{2n \times 1}$, and the *augmented complex representation*, which works with the complex vector and its complex conjugate: $\underline{\mathbf{x}} = \left[\mathbf{x}^T \; (\mathbf{x}^*)^T\right]^T \in \mathbb{C}^{2n \times 1}$ [129]. Both representations are mathematically equivalent

in the sense that it is equivalent to work with one or the other, since they are related by the following bijective transformation:

$$\bar{\mathbf{x}} = \left[\begin{array}{c} \Re\{\mathbf{x}\} \\ \Im\{\mathbf{x}\} \end{array} \right] = \mathbf{Z} \left[\begin{array}{c} \mathbf{x} \\ \mathbf{x}^* \end{array} \right] = \mathbf{Z}\underline{\mathbf{x}}, \tag{2.50}$$

where

$$\mathbf{Z} = \tfrac{1}{2} \left[\begin{array}{cc} \mathbf{I} & \mathbf{I} \\ -j\mathbf{I} & j\mathbf{I} \end{array} \right]. \tag{2.51}$$

It is very important to recall that although $\bar{\mathbf{x}}$ is a real-valued vector, $\mathbf{Z}$ and $\underline{\mathbf{x}}$ are complex-valued. Thus, $\bar{\mathbf{x}}^T = \underline{\mathbf{x}}^H \mathbf{Z}^H$.

Lemma 2.4 ([65]). *Any arbitrary zero-mean complex random vector* $\mathbf{x} \in \mathbb{C}^{n \times 1}$ *is characterized by the covariance matrix of the augmented vector* $\underline{\mathbf{x}}$*, which is given by:*

$$\mathbf{C}_{\underline{\mathbf{x}}} = \mathbb{E}[\underline{\mathbf{x}}\underline{\mathbf{x}}^H] = \left[\begin{array}{cc} \mathbf{C}_{\mathbf{x}} & \tilde{\mathbf{C}}_{\mathbf{x}} \\ \tilde{\mathbf{C}}_{\mathbf{x}}^* & \mathbf{C}_{\mathbf{x}}^* \end{array} \right]. \tag{2.52}$$

The augmented covariance matrix $\mathbf{C}_{\underline{\mathbf{x}}}$ has some built-in redundancy for the second-order characterization of $\mathbf{x}$. However, it is useful as shown in the following theorems.

Theorem 2.5 ([65]). $\mathbf{C}_{\mathbf{x}}$ *and* $\tilde{\mathbf{C}}_{\mathbf{x}}$ *are a valid set of covariance and pseudo-covariance matrices, i.e., there exists a random vector* $\mathbf{x}$ *with covariance and pseudo-covariance matrices given by* $\mathbf{C}_{\mathbf{x}}$ *and* $\tilde{\mathbf{C}}_{\mathbf{x}}$*, respectively, if and only if the augmented covariance matrix* $\mathbf{C}_{\underline{\mathbf{x}}}$ *is positive semidefinite, i.e.* $\mathbf{C}_{\underline{\mathbf{x}}} \succeq \mathbf{0}$.

The conditions of the covariance matrix $\mathbf{C}_{\mathbf{x}}$ being Hermitian and positive semidefinite and the pseudo-covariance matrix $\tilde{\mathbf{C}}_{\mathbf{x}}$ being symmetric are already implied by the augmented covariance matrix $\mathbf{C}_{\underline{\mathbf{x}}}$ being positive semidefinite [65].

Furthermore, for the improper complex Gaussian random vectors, the differential entropy is in general a function of both the covariance and pseudo-covariance matrices, which can be expressed in terms of $\mathbf{C}_{\underline{\mathbf{x}}}$ as shown by the following theorem.

Theorem 2.6 ([65]). *The entropy of a complex Gaussian random vector* $\mathbf{x}$ *of length* n *with augmented covariance matrix* $\mathbf{C}_{\underline{\mathbf{x}}}$ *is:*

$$h(\mathbf{x}) = \frac{1}{2} \log_2 \left((\pi e)^{2n} \left| \mathbf{C}_{\underline{\mathbf{x}}} \right| \right) \tag{2.53}$$

Theorem 2.6 generalizes the entropy result for proper Gaussian random vectors. If $\tilde{\mathbf{C}}_{\mathbf{x}} = \mathbf{0}$, (2.53) reduces to the well-known entropy expression for proper Gaussian random vectors [128]:

$$h(\mathbf{x}) = \log_2 \left((\pi e)^n \left| \mathbf{C}_{\mathbf{x}} \right| \right) \tag{2.54}$$

The relation between the covariance matrix of the augmented vector $\underline{\mathbf{x}}$ (i.e. $\mathbf{C}_{\underline{\mathbf{x}}}$ in (2.52)) and the covariance matrix of the real-valued vector $\bar{\mathbf{x}}$ (i.e. $\mathbf{C}_{\bar{\mathbf{x}}}$) can be derived from (2.50). It is included in the following lemma.

Lemma 2.5. *The covariance matrix of the real-valued vector $\bar{\mathbf{x}}$ is related to the covariance matrix of the augmented vector $\underline{\mathbf{x}}$ as:*

$$\mathbf{C}_{\bar{\mathbf{x}}} = \mathbb{E}[\bar{\mathbf{x}}\bar{\mathbf{x}}^T] = \mathbb{E}[\mathbf{Z}\underline{\mathbf{x}}\underline{\mathbf{x}}^H\mathbf{Z}^H] = \mathbf{Z}\mathbf{C}_{\underline{\mathbf{x}}}\mathbf{Z}^H = \mathbf{Z}\begin{bmatrix} \mathbf{C}_{\mathbf{x}} & \tilde{\mathbf{C}}_{\mathbf{x}} \\ \tilde{\mathbf{C}}_{\mathbf{x}}^* & \mathbf{C}_{\mathbf{x}}^* \end{bmatrix}\mathbf{Z}^H, \tag{2.55}$$

being matrix $\mathbf{Z}$ the one defined in (2.51). Due to the bijective transformation in (2.50) and Lemma 2.4, $\mathbf{C}_{\bar{\mathbf{x}}}$ in (2.55) characterizes any arbitrary zero-mean complex random vector $\mathbf{x}$.

In wireless communications, the information-bearing signals are usually selected from a CSCG (i.e. proper Gaussian) codebook. Hence, the problem of how efficiently generate an improper Gaussian signal $\mathbf{x}\in\mathbb{C}^{n\times 1}$ from a proper Gaussian signal $\mathbf{b}\in\mathbb{C}^{m\times 1}$ selected from a CSCG codebook arises. Without loss of generality, assume that $\mathbf{b}\sim\mathcal{CN}(\mathbf{0},\mathbf{I})$, i.e. $\mathbf{C}_{\mathbf{b}}=\mathbf{I}$ and $\tilde{\mathbf{C}}_{\mathbf{b}}=\mathbf{0}$. Note that the conventional linear precoding, i.e. $\mathbf{x}=\mathbf{Tb}$ where $\mathbf{T}\in\mathbb{C}^{n\times m}$ is the precoding matrix, it not able to map the proper Gaussian signal $\mathbf{b}$ into an improper Gaussian signal $\mathbf{x}$, since the pseudo-covariance matrix of $\mathbf{x}$ vanishes $\tilde{\mathbf{C}}_{\mathbf{x}}=\mathbf{T}\tilde{\mathbf{C}}_{\mathbf{b}}\mathbf{T}^T=\mathbf{0}$. However, widely linear precoding is able to do it [69]. The main result is summarized in the following lemma.

Lemma 2.6 ([69]). *Given a proper Gaussian information-bearing signal $\mathbf{b}\in\mathbb{C}^{m\times 1}$ characterized by $\mathbf{C}_{\mathbf{b}}=\mathbf{I}$ and $\tilde{\mathbf{C}}_{\mathbf{b}}=\mathbf{0}$, an improper Gaussian signal $\mathbf{x}\in\mathbb{C}^{n\times 1}$ can be obtained by applying the following widely linear precoding (WLP) to the proper Gaussian signal $\mathbf{b}$:*

$$\mathbf{x} = \mathbf{T}_1\mathbf{b} + \mathbf{T}_2\mathbf{b}^*, \tag{2.56}$$

where $\mathbf{T}_1\in\mathbb{C}^{n\times m}$ and $\mathbf{T}_2\in\mathbb{C}^{n\times m}$ are the precoding matrices corresponding to the blocks of $\mathbf{C}_{\underline{\mathbf{x}}}$ in (2.52):

$$\mathbf{C}_{\underline{\mathbf{x}}}^{\frac{1}{2}} = \begin{bmatrix} \mathbf{T}_1 & \mathbf{T}_2 \\ \mathbf{T}_2^* & \mathbf{T}_1^* \end{bmatrix}. \tag{2.57}$$

Therefore, as $\mathbf{C}_{\underline{\mathbf{x}}}$ characterizes any arbitrary zero-mean complex random vector, any improper Gaussian signal can be generated from a proper Gaussian signal through the use of WLP [69, Sect. II.C].

In other words, Lemma 2.6 states the relation between the covariance and pseudo-covariance matrices of signal $\mathbf{x}$ ($\mathbf{C}_{\mathbf{x}}$ and $\tilde{\mathbf{C}}_{\mathbf{x}}$) and the precoding matrices $\mathbf{T}_1$ and $\mathbf{T}_2$:

$$\mathbf{C}_{\mathbf{x}} = \mathbf{T}_1\mathbf{T}_1^H + \mathbf{T}_2\mathbf{T}_2^H \tag{2.58}$$
$$\tilde{\mathbf{C}}_{\mathbf{x}} = \mathbf{T}_1\mathbf{T}_2^T + \mathbf{T}_2\mathbf{T}_1^T \tag{2.59}$$

Note that if $\mathbf{T}_2=\mathbf{0}$, which is the case when $\mathbf{C}_{\underline{\mathbf{x}}}$ is block-diagonal (i.e. $\tilde{\mathbf{C}}_{\mathbf{x}}=\mathbf{0}$), then (2.56) reduces to the conventional linear precoding used for proper Gaussian signaling: $\mathbf{x}=\mathbf{T}_1\mathbf{d}$.

2.B   Preliminaries for Majorization Theory

In this section, some basic notions of majorization theory are presented. See [96] for a complete reference and [97] for its applications in wireless communications. Majorization theory makes precise the vague notion that the components of a real-valued vector $\mathbf{y}$ are "more spread out" or "less nearly equal" than are the components of a real-valued vector $\mathbf{x}$.

Definition 2.4. *For any real-valued vector* $\mathbf{x} \in \mathbb{R}^{n \times 1}$*, let*

$$\mathbf{x}_{[1]} \geq \cdots \geq \mathbf{x}_{[n]} \tag{2.60}$$

denote the components of vector $\mathbf{x}$ *in decreasing order. Similarly, let*

$$\mathbf{x}_{(1)} \leq \cdots \leq \mathbf{x}_{(n)} \tag{2.61}$$

denote the components of vector $\mathbf{x}$ *in increasing order.*

Under this ordering of vectors, different kinds of majorization relations when comparing two vectors arise: strong majorization, weak majorization, and weak log-majorization, defined in the following.

Definition 2.5 ([96] - strong majorization). *For any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n \times 1}$*,* $\mathbf{y}$ *majorizes* $\mathbf{x}$ *(or* $\mathbf{x}$ *is majorized by* $\mathbf{y}$*), written as* $\mathbf{y} \succ \mathbf{x}$ *(or, equivalently,* $\mathbf{x} \prec \mathbf{y}$*), if:*

$$\sum_{i=1}^{m} \mathbf{y}_{[i]} \geq \sum_{i=1}^{m} \mathbf{x}_{[i]} \quad m = 1, ..., n-1 \quad and \quad \sum_{i=1}^{n} \mathbf{y}_{[i]} = \sum_{i=1}^{n} \mathbf{x}_{[i]}. \tag{2.62}$$

Note that the conditions (2.62) are equivalent to:

$$\sum_{i=1}^{m} \mathbf{y}_{(i)} \leq \sum_{i=1}^{m} \mathbf{x}_{(i)} \quad m = 1, ..., n-1 \quad and \quad \sum_{i=1}^{n} \mathbf{y}_{(i)} = \sum_{i=1}^{n} \mathbf{x}_{(i)}. \tag{2.63}$$

Definition 2.6 ([96] - weak majorization). *For any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n \times 1}$*,* $\mathbf{y}$ *weakly majorizes* $\mathbf{x}$ *(or* $\mathbf{x}$ *is weakly majorized by* $\mathbf{y}$*), written as* $\mathbf{y} \succ_w \mathbf{x}$ *(or, equivalently,* $\mathbf{x} \prec_w \mathbf{y}$*), if:*

$$\sum_{i=1}^{m} \mathbf{y}_{[i]} \geq \sum_{i=1}^{m} \mathbf{x}_{[i]} \quad m = 1, ..., n. \tag{2.64}$$

Note that $\mathbf{y} \succ \mathbf{x}$ implies $\mathbf{y} \succ_w \mathbf{x}$, so strong majorization is a more restrictive definition than weak majorization.

Definition 2.7 ([96] - weak log-majorization). *For any* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n \times 1}$*,* $\mathbf{y}$ *weakly log-majorizes* $\mathbf{x}$ *(or* $\mathbf{x}$ *is weakly log-majorized by* $\mathbf{y}$*), written as* $\mathbf{y} \succ_{w \log} \mathbf{x}$ *(or, equivalently,* $\mathbf{x} \prec_{w \log} \mathbf{y}$*), if:*

$$\prod_{i=1}^{m} \mathbf{y}_{[i]} \geq \prod_{i=1}^{m} \mathbf{x}_{[i]} \quad m = 1, ..., n. \tag{2.65}$$

Note that $\mathbf{y} \succ_{w \log} \mathbf{x}$ implies $\mathbf{y} \succ_w \mathbf{x}$, so weak log-majorization is a stronger definition than weak majorization.

Functions that preserve the ordering of majorization are said to be Schur-convex (or Schur-concave if the order is reverted), as defined in the following.

Definition 2.8 ([96] - Schur-convex and Schur-concave functions). *A real-valued function* ϕ *defined on a set* $\mathcal{A} \in \mathbb{R}^{n \times 1}$ *is said to be Schur-convex on* $\mathcal{A}$ *if:*

$$\mathbf{y} \succ \mathbf{x} \quad on \ \mathcal{A} \quad \Rightarrow \quad \phi(\mathbf{y}) \geq \phi(\mathbf{x}). \tag{2.66}$$

Similarly, ϕ *is said to be Schur-concave on* $\mathcal{A}$ *if:*

$$\mathbf{y} \succ \mathbf{x} \quad on \ \mathcal{A} \quad \Rightarrow \quad \phi(\mathbf{y}) \leq \phi(\mathbf{x}). \tag{2.67}$$

The Schur-convex (Schur-concave) property of a function can be identified as follows.

Proposition 2.1 ([96]). *If* $\mathbb{I}\in\mathbb{R}$ *is an interval and* $g : \mathbb{I}\to\mathbb{R}$ *is convex, then:*

$$\phi(\mathbf{x}) = \sum\nolimits_{i=1}^{n} g(x_{[i]}) \tag{2.68}$$

is Schur-convex on $\mathbb{I}^{n\times 1}$. *Similarly, if* $g : \mathbb{I}\to\mathbb{R}$ *is concave, then:* $\phi(\mathbf{x})=\sum_{i=1}^{n} g(x_{[i]})$ *is Schur-concave on* $\mathbb{I}^{n\times 1}$.

Furthermore, there is a stronger version of Schur-convexity (Schur-concavity) for functions that preserve (revert) the ordering of majorization with a positive increment, as introduced in the following.

Definition 2.9 ([130] - Strongly Schur-convex and Strongly Schur-concave functions). *A real-valued function* ϕ *defined on a set* $\mathcal{A}\in\mathbb{R}^n$ *is said to be strongly Schur-convex with modulus* $c>0$ *on* $\mathcal{A}$ *if:*

$$\mathbf{y} \succ \mathbf{x} \quad on \quad \mathcal{A} \quad \Rightarrow \quad \phi(\mathbf{y}) \geq \phi(\mathbf{x}) + c(||\mathbf{y}||^2 - ||\mathbf{x}||^2). \tag{2.69}$$

Similarly, ϕ *is said to be strongly Schur-concave with modulus* $c>0$ *on* $\mathcal{A}$ *if:*

$$\mathbf{y} \succ \mathbf{x} \quad on \quad \mathcal{A} \quad \Rightarrow \quad \phi(\mathbf{y}) \leq \phi(\mathbf{x}) - c(||\mathbf{y}||^2 - ||\mathbf{x}||^2). \tag{2.70}$$

Note that the usual Schur-convexity and Schur-concavity correspond to the case $c=0$.

The strongly Schur-convex (strongly Schur-concave) property of a function can be identified as follows.

Proposition 2.2 ([130]). *If* $\mathbb{I}\in\mathbb{R}$ *is an interval and* $g:\mathbb{I}\to\mathbb{R}$ *is strongly convex with modulus* $c>0$, *then:*

$$\phi(\mathbf{x}) = \sum\nolimits_{i=1}^{n} g(x_{[i]}) \tag{2.71}$$

is strongly Schur-convex with modulus $c>0$ *on* $\mathbb{I}^{n\times 1}$. *Similarly, if* $g:\mathbb{I}\to\mathbb{R}$ *is strongly concave with modulus* $c>0$, *then:* $\phi(\mathbf{x})=\sum_{i=1}^{n} g(x_{[i]})$ *is strongly Schur-concave with modulus* $c>0$ *on* $\mathbb{I}^{n\times 1}$.

Finally, let us present an important majorization result for partitioned Hermitian matrices which states that, for two Hermitian matrices with equal diagonal blocks, the stronger the off-diagonal blocks are the more spread out the eigenvalues become, see [65, Result A3.7]. This result can also be derived from the pinching inequality, see [86].

Lemma 2.7 ([65] - partitioned Hermitian matrices). *For Hermitian matrices* $\mathbf{A}$, $\mathbf{B}$ *in the form:*

$$\mathbf{A} = \left[\begin{array}{cc} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{12}^H & \mathbf{A}_{22} \end{array} \right], \quad \mathbf{B} = \left[\begin{array}{cc} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{array} \right], \tag{2.72}$$

the following strong majorization result is fulfilled:

$$\mathrm{eig}(\mathbf{A}) \succ \mathrm{eig}(\mathbf{B}). \tag{2.73}$$

2.C   Proof of Lemma 2.1

The relation among R in (2.9) and $\bar{\mathbf{E}}$ in (2.12) can be easily obtained as: $R=-\frac{1}{2}\log_2|2\bar{\mathbf{E}}|$. Further, the MSE-matrix in the double-sized real-valued form $\bar{\mathbf{E}}$ can be related to the MSE-matrix of the transmitted symbols (i.e. $\mathbf{E}=\mathbb{E}[(\mathbf{b}-\hat{\mathbf{b}})(\mathbf{b}-\hat{\mathbf{b}})^H]$) as:

$$\bar{\mathbf{E}} = \mathbf{Z}\bar{\bar{\mathbf{E}}}\mathbf{Z}^H, \quad \bar{\bar{\mathbf{E}}}=\left[\begin{array}{cc} \mathbf{E} & \tilde{\mathbf{E}} \\ \tilde{\mathbf{E}}^* & \mathbf{E}^* \end{array}\right], \tag{2.74}$$

where $\tilde{\mathbf{E}}=\mathbb{E}[(\mathbf{b}-\hat{\mathbf{b}})(\mathbf{b}-\hat{\mathbf{b}})^T]$ and transformation matrix $\mathbf{Z}$ is defined in (2.10). So, by plugging (2.74) into the relation in (2.13) and applying the properties of the determinant and the fact that $\mathbf{Z}^{-1}\mathbf{Z}^{-H}=2\mathbf{I}$, we obtain the second equality of equation (2.13):

$$R = -\tfrac{1}{2}\log_2|2\bar{\mathbf{E}}| = -\tfrac{1}{2}\log_2|\bar{\bar{\mathbf{E}}}| = -\tfrac{1}{2}\log_2|\mathbf{E}| - \tfrac{1}{2}\log_2|\mathbf{E}^* - \tilde{\mathbf{E}}^*\mathbf{E}^{-1}\tilde{\mathbf{E}}|. \tag{2.75}$$

Finally, R can be lower bounded by using the Fischer improved determinant inequality [124] which states that: $|\bar{\bar{\mathbf{E}}}|\leq|\mathbf{E}|^2-|\tilde{\mathbf{E}}||\tilde{\mathbf{E}}^*|$. Hence, applying logarithms to the inequality, we get:

$$-\tfrac{1}{2}\log_2|\bar{\bar{\mathbf{E}}}| + \tfrac{1}{2}\log_2|\mathbf{E}|^2 \geq \tfrac{1}{2}\log_2(1 + |\tilde{\mathbf{E}}||\tilde{\mathbf{E}}^*||\bar{\bar{\mathbf{E}}}|^{-1}). \tag{2.76}$$

So, by using $|\bar{\bar{\mathbf{E}}}|=|2\bar{\mathbf{E}}|$, inequality in (2.14) is derived from (2.76). Further, as the right-hand in (2.14) is strictly positive, we can conclude:

$$R \geq -\log_2|\mathbf{E}| + \tfrac{1}{2}\log_2\big(1 + \tfrac{1}{2}|\tilde{\mathbf{E}}||\tilde{\mathbf{E}}^*||\bar{\mathbf{E}}|^{-1}\big) \geq -\log_2|\mathbf{E}|. \tag{2.77}$$

2.D   Proof of Lemma 2.2

Assume that the channel coefficients of $\mathbf{H}\in\mathbb{C}^{N\times M}$ in (2.1) follow a Rayleigh distribution (i.e. $\mathbf{H}$ is full rank) and $M\geq N$. Then, let us write the vector of eigenvalues in (2.23) as follows:

$$\begin{aligned} \bar{\boldsymbol{\lambda}}_P &= \mathrm{eig}\Big(\left[\begin{array}{cc} \mathbf{C_n} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_n^*} \end{array}\right]^{-1}\tfrac{1}{2}\mathbf{Z}^{-1}\bar{\mathbf{H}}\bar{\mathbf{H}}^T\mathbf{Z}^{-H}\Big), \\ \bar{\boldsymbol{\lambda}}_I &= \mathrm{eig}\Big(\left[\begin{array}{cc} \mathbf{C_n} & \tilde{\mathbf{C}}_\mathbf{n} \\ \tilde{\mathbf{C}}_\mathbf{n}^* & \mathbf{C_n^*} \end{array}\right]^{-1}\tfrac{1}{2}\mathbf{Z}^{-1}\bar{\mathbf{H}}\bar{\mathbf{H}}^T\mathbf{Z}^{-H}\Big), \end{aligned} \tag{2.78}$$

in which we have used the fact that $\mathrm{eig}(\mathbf{XY})=\mathrm{eig}(\mathbf{YX})$ for any matrix $\mathbf{X}$ and $\mathbf{Y}$. Note that the block-matrices $[\mathbf{C_n}\ \mathbf{0};\mathbf{0}\ \mathbf{C_n^*}]$ and $[\mathbf{C_n}\ \tilde{\mathbf{C}}_\mathbf{n};\tilde{\mathbf{C}}_\mathbf{n}^*\ \mathbf{C_n^*}]$ in (2.78) are full rank, so their inverses exist. By using the structure of $\bar{\mathbf{H}}$ in (2.5), the structure of $\mathbf{Z}$ in (2.10), and some matrix manipulations, it can be shown that $\frac{1}{2}\mathbf{Z}^{-1}\bar{\mathbf{H}}\bar{\mathbf{H}}^T\mathbf{Z}^{-H}$ in (2.78) is equal to:

$$\tfrac{1}{2}\mathbf{Z}^{-1}\bar{\mathbf{H}}\bar{\mathbf{H}}^T\mathbf{Z}^{-H} = \left[\begin{array}{cc} \mathbf{H}\mathbf{H}^H & \mathbf{0} \\ \mathbf{0} & \mathbf{H}^*\mathbf{H}^T \end{array}\right]. \tag{2.79}$$

This relation is key for the proof. Note that $\frac{1}{2}\mathbf{Z}^{-1}\bar{\mathbf{H}}\bar{\mathbf{H}}^T\mathbf{Z}^{-H}$ in (2.79) is full rank for $M\geq N$ and $\mathbf{H}$ full rank, so its inverse exists and it is equal to $(\frac{1}{2}\mathbf{Z}^{-1}\bar{\mathbf{H}}\bar{\mathbf{H}}^T\mathbf{Z}^{-H})^{-1}=[(\mathbf{H}\mathbf{H}^H)^{-1}\ \mathbf{0};\mathbf{0}\ (\mathbf{H}\mathbf{H}^H)^{-1}]$.

The inverses of the positive eigenvalues in (2.78) (i.e. $\bar{\boldsymbol{\lambda}}_P^{-1}$ and $\bar{\boldsymbol{\lambda}}_I^{-1}$) can be obtained from the positive eigenvalues of the inverse matrix. Then, using the relation in (2.79), we get:

$$
\begin{aligned}
\bar{\boldsymbol{\lambda}}_P^{-1} &= \mathrm{eig}\left(\begin{bmatrix} \mathbf{C_n} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_n^*} \end{bmatrix} \begin{bmatrix} (\mathbf{HH}^H)^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{H}^*\mathbf{H}^T)^{-1} \end{bmatrix} \right), \\
\bar{\boldsymbol{\lambda}}_I^{-1} &= \mathrm{eig}\left(\begin{bmatrix} \mathbf{C_n} & \tilde{\mathbf{C}}_\mathbf{n} \\ \tilde{\mathbf{C}}_\mathbf{n}^* & \mathbf{C_n^*} \end{bmatrix} \begin{bmatrix} (\mathbf{HH}^H)^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{H}^*\mathbf{H}^T)^{-1} \end{bmatrix} \right),
\end{aligned}
\tag{2.80}
$$

such that, multiplying the block-matrices, we have:

$$
\begin{aligned}
\bar{\boldsymbol{\lambda}}_P^{-1} &= \mathrm{eig}\left(\begin{bmatrix} \mathbf{C_n}(\mathbf{HH}^H)^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_n^*}(\mathbf{H}^*\mathbf{H}^T)^{-1} \end{bmatrix} \right), \\
\bar{\boldsymbol{\lambda}}_I^{-1} &= \mathrm{eig}\left(\begin{bmatrix} \mathbf{C_n}(\mathbf{HH}^H)^{-1} & \tilde{\mathbf{C}}_\mathbf{n}(\mathbf{H}^*\mathbf{H}^T)^{-1} \\ \tilde{\mathbf{C}}_\mathbf{n}^*(\mathbf{HH}^H)^{-1} & \mathbf{C_n^*}(\mathbf{H}^*\mathbf{H}^T)^{-1} \end{bmatrix} \right).
\end{aligned}
\tag{2.81}
$$

Therefore, as the diagonal blocks of the partitioned Hermitian matrices in (2.81) are equal, we can make use of Lemma 2.7 in Appendix 2.B and hence, from (2.81), obtain:

$$
\bar{\boldsymbol{\lambda}}_I^{-1} \succ \bar{\boldsymbol{\lambda}}_P^{-1},
\tag{2.82}
$$

which demonstrates the strong majorization result in (2.24) of Lemma 2.2.

The weak log-majorization result in (2.24) of Lemma 2.2 is demonstrated in continuation by using some useful majorization theory properties that can be derived from (2.82). The first useful majorization theory property is [96, Sect. 5.A.1.d]: if $\mathrm{eig}(\mathbf{A}) \succ \mathrm{eig}(\mathbf{B})$, then $\prod_{i=k}^{K} \mathrm{eig}(\mathbf{A})_i \le \prod_{i=k}^{K} \mathrm{eig}(\mathbf{B})_i$, $k = 1, ..., K$, being $\mathrm{eig}(\mathbf{A})_i$ the i-th eigenvalue of $\mathbf{A}$ and K the rank of $\mathbf{A}$. Due to the ordering of eigenvalues, the last components of the vectors in (2.82) correspond to the first components of the eigenvalues in (2.78), such that as an implication of (2.82):

$$
\prod_{i=1}^{n} \frac{1}{\bar{\lambda}_{I,i}} \le \prod_{i=1}^{n} \frac{1}{\bar{\lambda}_{P,i}} \quad n = 1, ..., 2N,
\tag{2.83}
$$

and hence:

$$
\prod_{i=1}^{n} \bar{\lambda}_{I,i} \ge \prod_{i=1}^{n} \bar{\lambda}_{P,i} \quad n = 1, ..., 2N.
\tag{2.84}
$$

The second useful majorization theory property is [96, Sect. 5.A.1]: if $\mathrm{eig}(\mathbf{A}) \succ \mathrm{eig}(\mathbf{B})$, then we can apply a convex function $g(.)$ over each component of the vectors and the following is satisfied: $g(\mathrm{eig}(\mathbf{A})) \succ_w g(\mathrm{eig}(\mathbf{B}))$. As $g(x) = \frac{1}{x}$ is a convex function, it follows from (2.82) that:

$$
\sum_{i=1}^{n} \bar{\lambda}_{I,i} \ge \sum_{i=1}^{n} \bar{\lambda}_{P,i} \quad n = 1, ..., 2N.
\tag{2.85}
$$

Again, the ordering of eigenvalues is very important here. Finally, from (2.84) and (2.85), the weak log-majorization result in (2.24) is demonstrated (see Definition 2.7 in Appendix 2.B).

2.E   Proof of Theorem 2.1

The function $\frac{1}{2}\log_2\left(1 + \frac{\bar{P}}{x}\right)$ is convex w.r.t. $x \ge 0$ for $\bar{P} > 0$. Thus, as convex functions generate Schur-convex sums (see Proposition 2.1 in Appendix 2.B), the function $\sum_i \frac{1}{2}\log_2\left(1 + \frac{\bar{P}}{x_i}\right)$ is a

Schur-convex function. The achievable rate in (2.28) can be written in such a form with $x_i = \frac{1}{\lambda_i}$: $R^{\text{opt}} = \sum_{i=1}^{2N} \frac{1}{2} \log_2 \left(1 + \frac{\bar{P}}{1/\lambda_i}\right)$, so it is a Schur-convex function on $\bar{\boldsymbol{\lambda}}^{-1}$ for $\bar{P} > 0$. Therefore, due to Lemma 2.2 ($\bar{\boldsymbol{\lambda}}_I^{-1} \succ \bar{\boldsymbol{\lambda}}_P^{-1}$) and as the achievable rate is a Schur-convex function on $\bar{\boldsymbol{\lambda}}^{-1}$, by majorization theory on Schur-convex functions (see Definition 2.8 in Appendix 2.B) we get:

$$R^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_I) \geq R^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_P). \tag{2.86}$$

To further extend this inequality, we make use of the results from strong Schur-convexity (see Definition 2.9 in Appendix 2.B) [130]. Inequality in (2.30) of Theorem 2.1 is obtained by showing that the rate expression $R^{\text{opt}} = \sum_{i=1}^{2N} \frac{1}{2} \log_2 \left(1 + \frac{\bar{P}}{x_i}\right)$ with $x_i = \frac{1}{\lambda_i}$ is a *strongly Schur-convex* function with modulus c_R^{upa}. As strongly convex functions generate strongly Schur-convex sums (see Proposition 2.2 in Appendix 2.B), we need to prove that $\frac{1}{2} \log_2 \left(1 + \frac{\bar{P}}{x}\right)$ is strongly convex[14] with modulus c_R^{upa} on interval $x \in [0, x_{\max}]$ or, equivalently, to show that $\frac{1}{2} \log_2 \left(1 + \frac{\bar{P}}{x}\right) - c_R^{\text{upa}} x^2$ is convex on the interval $x \in [0, x_{\max}]$ [130]. By checking the second order derivative, it is easy to show that $\frac{1}{2} \log_2 (1 + \frac{\bar{P}}{x}) - c_R^{\text{upa}} x^2$ is convex for $c_R^{\text{upa}} \leq \frac{\bar{P}(x + 0.5\bar{P})}{2 \ln(2) x^2 (x + \bar{P})^2}$. Therefore, as the values of $x_i = \frac{1}{\lambda_i}$ are upper bounded by the minimum positive eigenvalue $x_{\max} = \frac{1}{\min(\bar{\lambda}_i)}$, there exists an interval $x \in [0, x_{\max}]$ in which $\frac{1}{2} \log_2 \left(1 + \frac{\bar{P}}{x}\right)$ is strongly convex with modulus:

$$c_R^{\text{upa}} = \frac{\bar{P}(x_{\max} + 0.5\bar{P})}{2 \ln(2) x_{\max}^2 (x_{\max} + \bar{P})^2} > 0. \tag{2.87}$$

Consequently, due to Lemma 2.2 ($\bar{\boldsymbol{\lambda}}_I^{-1} \succ \bar{\boldsymbol{\lambda}}_P^{-1}$) and as the achievable rate is a strongly Schur-convex function on $\bar{\boldsymbol{\lambda}}^{-1}$ with modulus c_R^{upa} (see Definition 2.9 in Appendix 2.B):

$$R^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_I) \geq R^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_P) + c_R^{\text{upa}} \left(\|\bar{\boldsymbol{\lambda}}_I^{-1}\|^2 - \|\bar{\boldsymbol{\lambda}}_P^{-1}\|^2\right). \tag{2.88}$$

Note that due to the strong majorization result in Lemma 2.2: $\frac{1}{\min(\bar{\boldsymbol{\lambda}}_P)} \leq \frac{1}{\min(\bar{\boldsymbol{\lambda}}_I)}$, such that the interval $[0, x_{\max}]$ is determined by the IGS case. Therefore, by setting $x_{\max} = \frac{1}{\min(\bar{\boldsymbol{\lambda}}_I)}$ in (2.87), c_R^{upa} in (2.31) is derived. Finally, as the squared 2-norm function is a Schur-convex function then $\left(\|\bar{\boldsymbol{\lambda}}_I^{-1}\|^2 - \|\bar{\boldsymbol{\lambda}}_P^{-1}\|^2\right) \geq 0$ (with equality if and only if $\bar{\boldsymbol{\lambda}}_I = \bar{\boldsymbol{\lambda}}_P$) [130]. So, the rate gap in (2.30) is strictly positive provided that $\bar{\boldsymbol{\lambda}}_I \neq \bar{\boldsymbol{\lambda}}_P$.

2.F   Proof of Theorem 2.2

The function $\frac{1}{2} \frac{x}{(\bar{P}+x)}$ is a concave function on $x \geq 0$ for $\bar{P} > 0$. Then, as concave functions generate Schur-concave sums (see Proposition 2.1 in Appendix 2.B), the function $\sum_i \frac{1}{2} \frac{x_i}{(1+x_i)}$ is a *Schur-concave* function. The MSE in (2.29) can be written in such a form if we consider $x_i = \frac{1}{\lambda_i}$: $\epsilon^{\text{opt}} = \sum_{i=1}^{2N} \frac{1}{2} \frac{1/\bar{\lambda}_i}{(\bar{P}+1/\bar{\lambda}_i)}$, so it is a Schur-concave function on $\bar{\boldsymbol{\lambda}}^{-1}$ for $\bar{P} > 0$. Therefore, due to Lemma 2.2 ($\bar{\boldsymbol{\lambda}}_I^{-1} \succ \bar{\boldsymbol{\lambda}}_P^{-1}$) and as the MSE is a Schur-concave function on $\bar{\boldsymbol{\lambda}}^{-1}$, by majorization

---

[14]A twice continuously differentiable function $f : (a, b) \to \mathbb{R}$ is strongly convex with modulus m if and only if $f''(x) \geq m > 0$ for $x \in (a, b)$.

theory on Schur-concave functions (see Definition 2.8 in Appendix 2.B) we have:

$$\epsilon^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_I) \leq \epsilon^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_P). \tag{2.89}$$

To further extend this inequality, we make use of the results from strong Schur-concavity (see Definition 2.9 in Appendix 2.B) [130]. Inequality in (2.32) of Theorem 2.2 is obtained by showing that the error expression $\epsilon^{\text{opt}}=\sum_{i=1}^{2N} \frac{1}{2} \frac{x_i}{(\bar{P}+x_i)}$ with $x_i=\frac{1}{\lambda_i}$ is a *strongly Schur-concave* function with modulus c_ϵ^{upa}. As strongly concave functions generate strongly Schur-concave sums (see Proposition 2.2 in Appendix 2.B), we need to prove that $\frac{1}{2} \frac{x}{(\bar{P}+x)}$ is strongly concave[15] with modulus c_ϵ^{upa} on interval $x\in[0, x_{\max}]$ or, equivalently, to show that $\frac{1}{2} \frac{x}{(\bar{P}+x)}+c_\epsilon^{\text{upa}}x^2$ is concave on interval $x\in[0, x_{\max}]$ [130]. By checking the second order derivative, one can show that $\frac{1}{2} \frac{x}{(\bar{P}+x)}+c_\epsilon^{\text{upa}}x^2$ is concave for $c_\epsilon^{\text{upa}}\leq \frac{\bar{P}}{2(x+\bar{P})^3}$. Therefore, as the values of $x_i=\frac{1}{\lambda_i}$ are upper bounded by the minimum positive eigenvalue $x_{\max}=\frac{1}{\min(\tilde{\lambda}_i)}$, there exists an interval $x\in[0, x_{\max}]$ in which the function $\frac{1}{2} \frac{x}{(\bar{P}+x)}$ is strongly concave with modulus:

$$c_\epsilon^{\text{upa}} = \frac{\bar{P}}{2(x_{\max} + \bar{P})^3} > 0. \tag{2.90}$$

Consequently, due to Lemma 2.2 $(\bar{\boldsymbol{\lambda}}_I^{-1}\succ\bar{\boldsymbol{\lambda}}_P^{-1})$ and as the MSE is a strongly Schur-concave function on $\bar{\boldsymbol{\lambda}}^{-1}$ with modulus c_ϵ^{upa} (see Definition 2.9 in Appendix 2.B):

$$\epsilon^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_I) \leq \epsilon^{\text{opt}}(\bar{P}\bar{\boldsymbol{\lambda}}_P) - c_\epsilon^{\text{upa}}\big(\|\bar{\boldsymbol{\lambda}}_I^{-1}\|^2 - \|\bar{\boldsymbol{\lambda}}_P^{-1}\|^2\big). \tag{2.91}$$

By setting $x_{\max}=\frac{1}{\min(\bar{\boldsymbol{\lambda}}_I)}$ in (2.90), c_ϵ^{upa} in (2.33) is derived. Also, as $\big(\|\bar{\boldsymbol{\lambda}}_I^{-1}\|^2-\|\bar{\boldsymbol{\lambda}}_P^{-1}\|^2\big)\geq0$, the error gap in (2.32) is strictly positive provided that $\bar{\boldsymbol{\lambda}}_I\neq\bar{\boldsymbol{\lambda}}_P$.

2.G Proof of Theorem 2.3

In order to demonstrate an achievable rate improvement with the use of IGS when OPA in (2.34) is implemented, let us focus on proving the following inequalities:

$$R^{\text{opt}}(\bar{\mathbf{p}}_I \circ \bar{\boldsymbol{\lambda}}_I) \geq R(\bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_I) \geq R^{\text{opt}}(\bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_P), \tag{2.92}$$

where $R(\bar{\mathbf{p}}_P\circ\bar{\boldsymbol{\lambda}}_I)$ refers to the achievable rate in the improper interference case when the power allocation derived from the proper interference case is used (which is not the optimum, but a valid power allocation). Regarding the second inequality in (2.92), in the following we introduce some interesting properties that would allow us to demonstrate it.

- Property 1: [96, Prop. 3.H.3.b] states that if $\mathbf{a}\succ_w\mathbf{b}$, then $\mathbf{a}\circ\mathbf{u}\succ_w\mathbf{b}\circ\mathbf{u}$ for any $\mathbf{u}\in\mathcal{D}_+$, where $\mathcal{D}_+$ denotes the set of vectors of length L such that $\{(u_1, ..., u_L) : u_1\geq\ldots\geq u_L\geq0\}$. The extension of this proposition to the weakly logarithm majorization is straightforward by realizing the properties of the product operation. So we can state that: if $\mathbf{a}\succ_{w\log}\mathbf{b}$, then

---

[15]A twice continuously differentiable function $f : (a, b) \to \mathbb{R}$ is strongly concave with modulus m if and only if $f''(x)\leq m < 0$ for $x\in(a, b)$.

$\mathbf{a} \circ \mathbf{u} \succ_{w\log} \mathbf{b} \circ \mathbf{u}$ for any $\mathbf{u} \in \mathcal{D}_+$. Accordingly, as $\bar{\mathbf{p}}_P \in \mathcal{D}_+$ and due to Lemma 2.2 $(\bar{\boldsymbol{\lambda}}_I \succ_{w\log} \bar{\boldsymbol{\lambda}}_P)$, we can conclude that when using the power allocation derived of the proper interference case (i.e. $\bar{\mathbf{p}}_P$), the following relation is satisfied:

$$\bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_I \succ_{w\log} \bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_P. \tag{2.93}$$

- Property 2: [131, Prop. 1.3] claims that if $\mathbf{a} \succ_{w\log} \mathbf{b}$, then $\sum_i \log_2(1+a_i) \geq \sum_i \log_2(1+b_i)$, as $\log_2(1+x)$ is an increasing function on $x \in [0, \infty)$ and $\log_2(1+e^x)$ is a convex function w.r.t. $x \in [0, \infty)$. Accordingly, as the achievable rate in (2.19) has such a form and owing to Property 1 $(\bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_I \succ_{w\log} \bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_P)$:

$$R(\bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_I) \geq R^{\text{opt}}(\bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_P). \tag{2.94}$$

Hence, the second inequality in (2.92) is demonstrated. The first inequality in (2.92) is intrinsic of the water-filling solution in (2.34), as for a given $\bar{\boldsymbol{\lambda}}_I$ the optimal power allocation in terms of achievable rate is given by $\bar{\mathbf{p}}_I$, so: $R^{\text{opt}}(\bar{\mathbf{p}}_I \circ \bar{\boldsymbol{\lambda}}_I) \geq R(\bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_I)$ and the proof is completed.

2.H   Proof of Theorem 2.4

The function $\frac{1}{2}\log_2\left(\frac{\mu}{x}\right)$ is a convex function w.r.t. $x \geq 0$ for $\mu > 0$. Thus, as convex functions generate Schur-convex sums (see Proposition 2.1 in Appendix 2.B), the function $\sum_i \frac{1}{2}\log_2\left(\frac{\mu}{x_i}\right)$ is a *Schur-convex* function. The achievable rate in (2.36) can be written in such a form with $x_i = \frac{1}{\lambda_i}$: $R^{\text{opt}} = \sum_{i=1}^{2N} \frac{1}{2}\log_2\left(\frac{\mu}{1/\lambda_i}\right)$, so it is a Schur-convex function on $\bar{\boldsymbol{\lambda}}^{-1}$ for $\mu > 0$. Furthermore, by checking the second order derivative and following similar rationale as in Appendix 2.E, it can be observed that the rate expression $R^{\text{opt}} = \sum_{i=1}^{2N} \frac{1}{2}\log_2\left(\frac{\mu}{x_i}\right)$ is a *strongly Schur-convex* function with modulus c_R^{opa} (i.e. $\frac{1}{2}\log_2\left(\frac{\mu}{x}\right)$ is strongly convex with modulus c_R^{opa} on interval $x \in [0, x_{\max}]$ or, equivalently, $\frac{1}{2}\log_2\left(\frac{\mu}{x}\right) - c_R^{\text{opa}} x^2$ is convex on the interval $x \in [0, x_{\max}]$ [130]):

$$c_R^{\text{opa}} = \frac{1}{4\ln(2)x_{\max}^2} > 0. \tag{2.95}$$

Consequently, due to Lemma 2.2 $(\bar{\boldsymbol{\lambda}}_I^{-1} \succ \bar{\boldsymbol{\lambda}}_P^{-1})$ and as the achievable rate is a strongly Schur-convex function on $\bar{\boldsymbol{\lambda}}^{-1}$ with modulus c_R^{opa} (see Definition 2.9 in Appendix 2.B):

$$R^{\text{opt}}(\bar{\mathbf{p}}_I \circ \bar{\boldsymbol{\lambda}}_I) \geq R^{\text{opt}}(\bar{\mathbf{p}}_P \circ \bar{\boldsymbol{\lambda}}_P) + c_R^{\text{opa}}\left(\|\bar{\boldsymbol{\lambda}}_I^{-1}\|^2 - \|\bar{\boldsymbol{\lambda}}_P^{-1}\|^2\right). \tag{2.96}$$

By setting $x_{\max} = \frac{1}{\min(\bar{\boldsymbol{\lambda}}_I)}$ in (2.95), c_R^{opa} in (2.38) is derived. Finally, as $\left(\|\bar{\boldsymbol{\lambda}}_I^{-1}\|^2 - \|\bar{\boldsymbol{\lambda}}_P^{-1}\|^2\right) \geq 0$, the rate gap in (2.37) is strictly positive provided that $\bar{\boldsymbol{\lambda}}_I \neq \bar{\boldsymbol{\lambda}}_P$.

Chapter 3

Transmit Coordination Strategies

This chapter investigates transmit coordination strategies for weighted sum-rate (WSR) maximization in the downlink (DL) transmission of MIMO multi-cell multi-user systems. In this sense, we focus on dense MIMO TDD smallcell networks (SCNs) and propose decentralized coordinated precoding (D-CoP) designs. Each BS designs its own precoding matrices based on channel state informacion (CSI) of the served users and knowledge of the interference-cost matrix, which plays the role of an interference tax and allows managing interference towards unintended users. A protocol is proposed to acquire the interference-cost matrix by processing the covariance matrix of the uplink (UL) received signal provided that: (i) channel reciprocity can be assumed and (ii) all users participating in DL can transmit in UL with an adequate transmit filter. In contrast to existing transmit coordination techniques, D-CoP is fully scalable, avoids estimation of the interfering channels, and does not require information exchange between BSs. In case all parameters are perfectly acquired, an iterative algorithm is presented with demonstrated monotonic convergence when all BSs update their transmit precoders simultaneously. Further, the problem is reformulated to derive a robust D-CoP under imperfect CSI conditions. Finally, simulations in 3GPP LTE-A SCNs show significant user packet throughput gains without increasing the complexity associated to transmit coordination. Robustness to imperfect CSI and non-ideal channel reciprocity is shown.

The technical papers and research contributions related to this chapter are:

[J2]   S. Lagen, A. Agustin, J. Vidal, "Decentralized Coordinated Precoding for Dense TDD Small Cell Networks", *IEEE Trans. on Wireless Commun.*, vol. 14, no. 8, pp. 4546 - 4561, Aug. 2015. doi: 10.1109/TWC.2015.2422704.

[C2]   S. Lagen, A. Agustin, J. Vidal, "Decentralized Beamforming with Coordinated Sounding for Inter-Cell Interference Management", *European Wireless 2014*, Barcelona (Spain), May 2014.

[P]   S. Lagen, A. Agustin, J. Vidal, "Methods and systems for decentralized interference management in a multi-antenna wireless communication system", U.S. Patent 14/452,968, Aug. 6, 2014.

[lte1-7]   7 contributions to 3GPP LTE-A release 12 (see details in Section 1.8.4).

[TR]   3GPP TR 36.874, 3rd Generation Partnership Project, Technical Specification Group Radio Access Network, "Coordinated multi-point operation for LTE with non-ideal backhaul (Release 12)", v.12.0.0, Dec. 2013.

This chapter is structured as follows. Section 3.1 contains the state of the art for WSR maximization in MIMO interference channels. The main contributions of the chapter are detailed in Section 3.2. In Section 3.3 the system model is presented, including the signal model for channel

estimation and DL and UL transmissions. In Section 3.4, the maximum WSR problem is formulated. In Section 3.5 the proposed D-CoP is derived assuming either perfect CSI or imperfect CSI conditions for the direct links, and how to acquire the required parameters for decentralized transmit filters design at BS is detailed. The iterative algorithm for D-CoP is also presented there (including convergence and complexity considerations) and practical impairments are analyzed. Section 3.6 investigates the practical implementation of D-CoP in a 3GPP-based network and different methods are proposed. In Section 3.7 the proposed procedure is evaluated using a 3GPP LTE-A SCN. Finally, concluding remarks are included in Section 3.8.

3.1   State of the Art

SCNs arise due to the network densification and the spatial re-use of the spectrum considered to improve the capacity of future cellular systems [6] (see Section 1.1). Differently from MeNB, which are intended to provide general coverage and service to high mobility users (UEs), BSs dispose of a lower transmit power, cover smaller areas, and are intended to serve low mobility UEs so as to avoid frequent handovers. This allows obtaining a reliable estimation of the propagation channel from the UE towards the serving BS due to the long channel coherence time [1].

An interesting type of deployment contemplated in 3GPP LTE-A networks is the non-co-channel HCNs deployment, where MeNBs and SeNBs use different carrier frequencies (see Fig. 1.3). This way, MeNBs may keep on working as usual and traffic can be offloaded to the new underlying SCN [22]. Although cross-tier interference is eliminated, the co-tier interference between SeNBs becomes a major problem that can significantly reduce the system throughput if SeNBs are densely deployed to serve hot-spot areas with high user traffic demands [104]. One advantage of the SCN is that SeNBs are supposed to operate in TDD so as to better match the UL:DL traffic asymmetry, which allows exploiting the reciprocity of UL and DL propagation channels to design advanced precoding techniques and hence improving the performance of the TDD SCN and reducing the channel feedback signaling needed as compared to FDD [24].

The dense TDD SCN can be modeled as a multi-cell multi-user MIMO system, a generic model for multi-user cellular communication systems where multiple BSs, each equipped with multiple antennas, wish to simultaneously send independent messages to its served users while generating interference to unintended users.

Unfortunately, the optimal transmit/receive strategy with linear filters that maximizes the WSR of said system is not known. From an optimization theory perspective, the maximum WSR problem is non-convex and NP-hard even in the single-antenna case [132], but several approaches that reach a local optima have been proposed in the literature. Some of them are centralized methods, whose main drawback is the added cost of a central unit required to collect the CSI of all users and the scalability of the solution. This has sparked a great interest in developing decentralized approaches that work iteratively and are obtained by decomposing the problem into separable parts, see [93–95].

Iterative methods in [133, 134] (and references therein) are based on the concept of the so-called **interference-cost**, whereby each BS maximizes its own utility function minus the interference-cost which reflects the interference created by the BS towards unintended UEs. Methods are available to ensure the convergence of these algorithms by slightly reformulating the problem (see for example [134]), nevertheless, all channel matrices (i.e. direct and interfering channel matrices) have to be estimated and reported along with the interference-cost.

An alternative approach for WSR maximization is presented in [127]. It relies on a **iterative minimization of the sum of weighted mean square errors** (WMSE) [135], which was initially introduced in [136] for the broadcast channel. The solution is achieved by iteratively updating transmit filters at BSs and receive filters and weighting matrices at UEs. Although monotonic convergence is demonstrated, its decentralized implementation in multi-cell systems requires again estimation of all channel matrices and the existence of feedback links from each UE towards all BSs is needed to report the updated receive filters and weighting matrices (see [127]).

Therefore, the main drawbacks of existing approaches to solve the maximization of the WSR of multi-cell multi-user MIMO systems in the literature ([127, 133]) are:

- the estimation of all channel matrices (as well as the associated computational cost and required network planning for channel estimation),
- the impact of channel estimation errors of the interfering channel matrices that are estimated with a low SNR, and
- the use of non-ideal backhaul and/or feedback links to exchange information among different terminals (BSs/UEs).

All has a detrimental effect on the overall potential performance gains of transmit coordination.

3.2   Contribution

In this chapter we exploit the fact that propagation channel reciprocity and a reliable estimation of the direct channels are available in TDD SCNs in order to propose a decentralized, scalable, and coordinated interference management procedure for maximizing the WSR in DL with linear transmit/receive filters. The WMSE formulation is used to face the maximization of the WSR, but different from [127] we decentralize the minimum WMSE problem by following the interference-cost concept such that a different transmit filters design at BSs is derived. Each SeNB designs its own transmit filters (precoders) based on the knowledge of the direct channel matrices towards the served UEs, the covariance matrix of the DL inter-cell interference plus noise that can be acquired from a parameter reported by the served UEs, and the interference-cost matrix that allows managing interference towards unintended users.

We propose that each BS acquires the interference-cost matrix over-the-air by using the UL received signal, provided that all UEs currently participating in DL can transmit simultaneously a specific pilot sequence in UL (as is shown in Fig. 3.1). By doing so, and in contrast to previous works, the estimation of the interfering channel matrices is avoided, hence significantly reducing its associated complexity in terms of computational cost and network planning for pilot signals, and also reducing the performance degradation due to imperfect estimation of the interfering channel matrices. Furthermore, information exchange among BSs is not required, which alleviates the impact of non-ideal backhaul links, and the number of feedback links and amount of information from UEs to BSs are significantly reduced as compared to [127].

The contributions of this chapter are:

- A decentralized coordinated precoding (D-CoP) for DL WSR maximization is proposed, in which each BS solves its own subproblem to design transmit filters.
- The subproblem is reformulated to tackle imperfect CSI conditions and a robust D-CoP is derived to overcome estimation errors of the direct channel matrices at BSs.

(a) DL desired signals (solid) and
interfering signals (dashed)

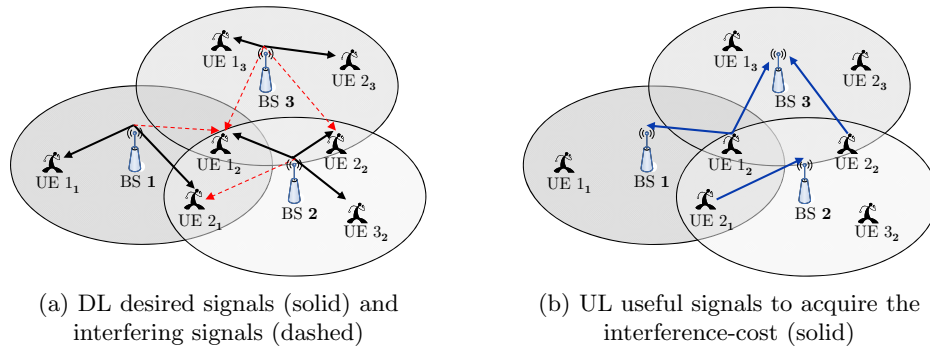(b) UL useful signals to acquire the
interference-cost (solid)

Figure 3.1: Acquisition of the interference-cost matrix from the UL transmission for DL transmit coordination in dense TDD smallcell networks.

- A new protocol is proposed to estimate the interference-cost matrix at each BS over-the-air from an UL pilot-based transmission.

- An iterative algorithm for D-CoP is presented, subsuming the acquisition of the required parameters at BSs and the simultaneous per-BS transmit filters design.

- Monotonic convergence of the algorithm for D-CoP is demonstrated if all parameters are perfectly acquired.

- The proposed D-CoP is evaluated in a 3GPP LTE-A smallcell scenario [104], showing large DL performance gains even when only 1 iteration is implemented.

Although the proposed D-CoP is applied to TDD SCNs in this chapter, it can be used to control interference in the general multi-cell multi-user MIMO TDD systems.

3.3   System Model

Consider a synchronized TDD SCN composed of a set of $\mathcal{K}\triangleq\{1,\dots,K\}$ BSs equipped with M_k antennas each $(k=1,\dots,K)$. Each k-th BS serves a set of $\mathcal{I}_k\triangleq\{1_k,\dots,I_k\}$ UEs. Let us define i_k to be the i-th UE served by the k-th BS and N_{i_k} the receive antenna elements at the i_k-th UE $(i=1,\dots,I_k)$. The total set of users in the system is denoted by $\mathcal{I} = \bigcup_{k\in\mathcal{K}} \mathcal{I}_k$. An example is shown in Fig. 3.1 for $|\mathcal{K}|=3$ and $|\mathcal{I}|=7$ $(|\mathcal{I}_1|=2, |\mathcal{I}_2|=3, |\mathcal{I}_2|=2)$. Through this chapter, we use i, j to indicate the UE index and k, l to indicate the BS index. All BSs present in the cellular network transmit information simultaneously to its served UEs, such that severe interference is created. Our focus is on DL interference management, including intra-cell and inter-cell interference management.

Low mobility UEs are assumed to be connected to the synchronized TDD SCN. During the channel coherence time, the communication is split into 2 phases (as shown in Fig. 3.2):

- **Phase 1 (channel estimation):** Phase 1 is devoted to acquire knowledge at each k-th BS of the direct channel matrix towards its served i_k-th UE, $\forall i_k\in\mathcal{I}_k$, which is denoted by $\mathbf{H}_{k,i_k}\in\mathbb{C}^{N_{i_k}\times M_k}$ and contains the complex-valued gains of the different antenna-pairs.

- **Phase 2 (D-CoP algorithm):** In phase 2, UL and DL transmissions are duplexed. The proposed algorithm iterates between UL and DL: in UL transmissions parameters like the DL inter-cell interference plus noise covariance matrix of the served i_k-th UE $(\mathbf{N}_{i_k}, \forall i_k\in\mathcal{I}_k)$
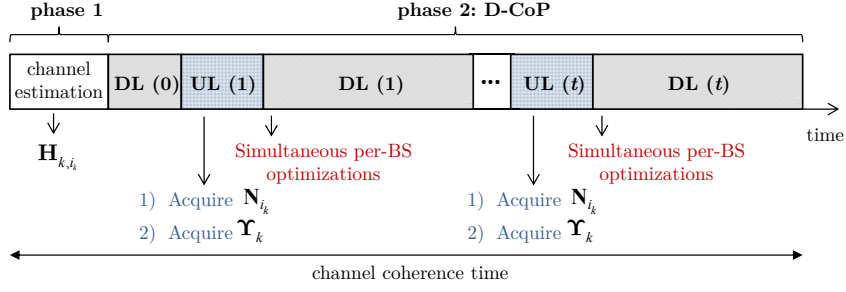
Figure 3.2: Procedure for D-CoP in TDD smallcell networks.

and the interference-cost matrix for the k-th BS ($\boldsymbol{\Upsilon}_k$) are acquired at each k-th BS, which are used for decentralized transmit filters design in the subsequent DL transmission. See details in Section 3.5.

In TDD systems, channel estimation can be done either at the BS or at the UE. In the later case, a feedback link is needed to report the channel matrix to the serving BS. For the ease of exposition, we consider that the direct channel matrix is estimated at the BS from an UL pilot-based transmission assuming perfect UL-DL propagation channel reciprocity (i.e. the channel matrix in UL $\overleftarrow{\mathbf{H}}_{k,i_k} \in \mathbb{C}^{M_k \times N_{i_k}}$ is given by the transpose of the channel matrix in DL: $\overleftarrow{\mathbf{H}}_{k,i_k} = \mathbf{H}_{k,i_k}^T$, $\forall k, \forall i_k$). See details in Section 3.3.1.

Different from previous works ([127, 133]), the proposed D-CoP scheme does not need to know the interfering channel matrices (i.e. $\mathbf{H}_{k,j_l} \in \mathbb{C}^{N_{j_l} \times M_k}$, between the k-th BS and the j_l-th UE served by l-th BS, $\forall l \neq k$) either at BSs or at UEs. Consequently, we avoid:

- the computational cost associated to channel estimation,
- the network planning required for pilot signals, and
- the performance loss due to imperfect estimation of the interfering channel matrices that are estimated with a low SNR.

In addition, all the required information for interference management is obtained through the air-interface without BS information exchange, such that the strict backhaul requirements usually needed for transmit coordination are strongly reduced.

3.3.1 Signal model for channel estimation (phase 1)

For channel estimation at BSs we use a well-established orthogonal training-based scheme [137], so interference between neighbor UEs for CSI acquisition is avoided. Hence, the signal received by the k-th BS at the M_k receive antennas during $T \geq N_{i_k}$ channel uses $\mathbf{S}_k \in \mathbb{C}^{M_k \times T}$ is:

$$\mathbf{S}_k = \mathbf{H}_{k,i_k}^T \mathbf{P}_{i_k} + \mathbf{V}_k, \tag{3.1}$$

where $\mathbf{P}_{i_k} \in \mathbb{C}^{N_{i_k} \times T}$ is the known matrix of training sequences associated to the i_k-th UE, and $\mathbf{V}_k \in \mathbb{C}^{M_k \times T}$ is the collection of channel noise vectors, composed of Gaussian independent and identically distributed (*i.i.d.*) components with distribution $\mathcal{CN}(0, \sigma_v^2)$. Let us denote P_T to the total power for channel estimation, i.e. $P_T = \mathrm{Tr}(\mathbf{P}_{i_k} \mathbf{P}_{i_k}^H)$. Channel matrix $\mathbf{H}_{k,i_k}$ is modeled by Gaussian *i.i.d.* components with distribution $\mathcal{CN}(0, \delta_{k,i_k})$, where δ_{k,i_k} refers to the gain introduced by path-loss and shadowing between the k-th BS and the i_k-th UE.

The channel estimation model is described by:

$$\mathbf{H}_{k,i_k} = \hat{\mathbf{H}}_{k,i_k} + \tilde{\mathbf{H}}_{k,i_k}, \tag{3.2}$$

where $\hat{\mathbf{H}}_{k,i_k}$ is the estimated channel matrix and $\tilde{\mathbf{H}}_{k,i_k}$ is the channel estimation error matrix. We assume that the channel matrix is estimated at BSs by applying a linear minimum mean square error (MMSE) estimator [138], i.e.

$$\hat{\mathbf{H}}_{k,i_k}^T = \mathbb{E}\left[\mathbf{H}_{k,i_k}^T | \mathbf{S}_k\right] = \mathbf{S}_k \mathbf{P}_{i_k}^H \left(\mathbf{P}_{i_k}\mathbf{P}_{i_k}^H + \frac{\sigma_v^2}{\delta_{k,i_k}}\mathbf{I}\right)^{-1}. \tag{3.3}$$

Then, assuming mutually white and orthogonal training sequences (i.e. $\mathbf{P}_{i_k}$ is a unitary matrix scaled by $\sqrt{P_T/N_{i_k}}$), the components of $\tilde{\mathbf{H}}_{k,i_k}$ in (3.2) turn out to be Gaussian and mutually uncorrelated with a variance for each channel estimation error component given by [138]:

$$\bar{J}_{k,i_k} = \frac{1}{M_k N_{i_k}} \operatorname{Tr}\left(\mathbb{E}\left[\tilde{\mathbf{H}}_{k,i_k}^T \tilde{\mathbf{H}}_{k,i_k}^*\right]\right) = \left(\frac{1}{\delta_{k,i_k}} + \frac{P_T}{N_{i_k}\sigma_v^2}\right)^{-1}. \tag{3.4}$$

For simplicity, it is assumed that noise power in the channel estimation phase is equal for all links, such that the different variances in the estimation errors in (3.4) come from the path-loss and shadowing variations of the different links.

3.3.2   Signal model for downlink (phase 2)

The signal transmitted by the k-th BS in DL is given by:

$$\mathbf{x}_k = \sum_{i_k \in \mathcal{I}_k} \mathbf{T}_{i_k}\mathbf{b}_{i_k}, \tag{3.5}$$

where $\mathbf{b}_{i_k} \in \mathbb{C}^{m_{i_k} \times 1}$ contains the unit power independent Gaussian symbols for the i_k-th UE (i.e. $\mathbf{b}_{i_k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$) and $\mathbf{T}_{i_k} \in \mathbb{C}^{M_k \times m_{i_k}}$ denotes its associated transmit filter (or precoder), being $m_{i_k} \leq \min(N_{i_k}, M_k)$ the number of streams. The total power spent at the k-th BS is:

$$P_k = \sum_{i_k \in \mathcal{I}_k} \operatorname{Tr}\left(\mathbf{T}_{i_k}\mathbf{T}_{i_k}^H\right). \tag{3.6}$$

Under narrow-band transmissions, the baseband signal observed at the i_k-th UE is:

$$\mathbf{y}_{i_k} = \mathbf{H}_{k,i_k}\mathbf{T}_{i_k}\mathbf{b}_{i_k} + \underbrace{\sum_{j_k \in \mathcal{I}_k, j_k \neq i_k} \mathbf{H}_{k,i_k}\mathbf{T}_{j_k}\mathbf{b}_{j_k}}_{\text{intra-cell interference}} + \underbrace{\sum_{l \in \mathcal{K}, l \neq k} \sum_{j_l \in \mathcal{I}_l} \mathbf{H}_{l,i_k}\mathbf{T}_{j_l}\mathbf{b}_{j_l}}_{\text{inter-cell interference}} + \mathbf{v}_{i_k}, \tag{3.7}$$

where $\mathbf{H}_{l,i_k} \in \mathbb{C}^{N_{i_k} \times M_l}$ is the channel matrix between the transmitting antennas at the l-th BS and the receiving antennas at the i_k-th UE. The second term in (3.7) contains the intra-cell interference, the third term denotes the inter-cell interference, and the last term refers to the additive zero-mean white Gaussian noise with distribution $\mathbf{v}_{i_k} \sim \mathcal{CN}(\mathbf{0}, \sigma_{i_k}^2\mathbf{I})$. Hence, under the independence assumption of $\{\mathbf{b}_{i_k}\}_{\forall i_k, k}$ and $\mathbf{v}_{i_k}$, the covariance matrix of the received signal at

the i_k-th UE is:

$$\mathbf{C}_{\mathbf{y}_{i_k}} = \mathbb{E}\left[\mathbf{y}_{i_k}\mathbf{y}_{i_k}^H\right] = \mathbf{H}_{k,i_k}\mathbf{T}_{i_k}\mathbf{T}_{i_k}^H\mathbf{H}_{k,i_k}^H + \sum_{j_k \in \mathcal{I}_k, j_k \neq i_k} \mathbf{H}_{k,i_k}\mathbf{T}_{j_k}\mathbf{T}_{j_k}^H\mathbf{H}_{k,i_k}^H + \mathbf{N}_{i_k}, \qquad (3.8)$$

where $\mathbf{N}_{i_k}$ is the covariance matrix of the inter-cell interference plus noise at the i_k-th UE:

$$\mathbf{N}_{i_k} = \sum_{l \in \mathcal{K}, l \neq k} \sum_{j_l \in \mathcal{I}_l} \mathbf{H}_{l,i_k}\mathbf{T}_{j_l}\mathbf{T}_{j_l}^H\mathbf{H}_{l,i_k}^H + \sigma_{i_k}^2\mathbf{I}. \qquad (3.9)$$

The symbols are estimated at the i_k-th UE assuming that interference is treated as noise and that a linear receive filter $\mathbf{R}_{i_k} \in \mathbb{C}^{N_{i_k} \times m_{i_k}}$ is applied at UE:

$$\hat{\mathbf{b}}_{i_k} = \mathbf{R}_{i_k}^H\mathbf{y}_{i_k}. \qquad (3.10)$$

Metrics for perfect CSI

The mean square error (MSE) for the symbols transmitted towards the i_k-th UE can be expressed through the so-called MSE-matrix: $\mathbf{E}_{i_k} = \mathbb{E}\left[(\hat{\mathbf{b}}_{i_k} - \mathbf{b}_{i_k})(\hat{\mathbf{b}}_{i_k} - \mathbf{b}_{i_k})^H\right]$, which can be expressed in terms of $\mathbf{R}_{i_k}$ in (3.10) and $\{\mathbf{T}_{i_k}\}$ in (3.5):

$$\mathbf{E}_{i_k}(\mathbf{R}_{i_k}, \{\mathbf{T}_{i_k}\}) = \mathbf{I} + \mathbf{R}_{i_k}^H\left(\sum_{j_k \in \mathcal{I}_k} \mathbf{H}_{k,i_k}\mathbf{T}_{j_k}\mathbf{T}_{j_k}^H\mathbf{H}_{k,i_k}^H + \mathbf{N}_{i_k}\right)\mathbf{R}_{i_k} - \mathbf{R}_{i_k}^H\mathbf{H}_{k,i_k}\mathbf{T}_{i_k} - \mathbf{T}_{i_k}^H\mathbf{H}_{k,i_k}^H\mathbf{R}_{i_k}.$$
$$(3.11)$$

The achievable rate of the i_k-th UE is understood as the well-known "log-det" capacity formula that depends on the set of transmit filters:

$$R_{i_k}(\{\mathbf{T}_{i_k}\}) = \log_2\left|\mathbf{I} + \mathbf{H}_{k,i_k}\mathbf{T}_{i_k}\mathbf{T}_{i_k}^H\mathbf{H}_{k,i_k}^H\left(\sum_{j_k \in \mathcal{I}_k, j_k \neq i_k} \mathbf{H}_{k,i_k}\mathbf{T}_{j_k}\mathbf{T}_{j_k}^H\mathbf{H}_{k,i_k}^H + \mathbf{N}_{i_k}\right)^{-1}\right|. \qquad (3.12)$$

Let us emphasize that the optimal MMSE receive filter[16] for the i_k-th UE is [136]:

$$\mathbf{R}_{i_k}^{\text{mmse}} = \mathbf{C}_{\mathbf{y}_{i_k}}^{-1}\mathbf{H}_{k,i_k}\mathbf{T}_{i_k} = \left(\sum_{j_k \in \mathcal{I}_k} \mathbf{H}_{k,i_k}\mathbf{T}_{j_k}\mathbf{T}_{j_k}^H\mathbf{H}_{k,i_k}^H + \mathbf{N}_{i_k}\right)^{-1}\mathbf{H}_{k,i_k}\mathbf{T}_{i_k}. \qquad (3.13)$$

Then, by using the optimal MMSE receiver filter and applying the matrix inversion lemma, the MSE-matrix in (3.11) results:

$$\mathbf{E}_{i_k}^{\text{mmse}}(\{\mathbf{T}_{i_k}\}) = \left(\mathbf{I} + \mathbf{T}_{i_k}^H\mathbf{H}_{k,i_k}^H\left(\sum_{j_k \in \mathcal{I}_k, j_k \neq i_k} \mathbf{H}_{k,i_k}\mathbf{T}_{j_k}\mathbf{T}_{j_k}^H\mathbf{H}_{k,i_k}^H + \mathbf{N}_{i_k}\right)^{-1}\mathbf{H}_{k,i_k}\mathbf{T}_{i_k}\right)^{-1}. \qquad (3.14)$$

Accordingly, the expression that relates the achievable rate in (3.12) and the MSE-matrix in (3.14) (when optimal MMSE receiver filter is used)[17] is easily obtained [136]:

$$R_{i_k}(\{\mathbf{T}_{i_k}\}) = -\log_2\left|\mathbf{E}_{i_k}^{\text{mmse}}(\{\mathbf{T}_{i_k}\})\right|. \qquad (3.15)$$

---

[16]The optimal MMSE receive filter is the one that minimizes the MSE, i.e. $\text{Tr}(\mathbf{E}_{i_k})$.
[17]Note that the MMSE receive filter is such that the achievable rate is preserved.

Metrics for imperfect CSI

In case of imperfect CSI of the direct channel matrix $\mathbf{H}_{k,i_k}$, we have to use $\mathbf{H}_{k,i_k}{=}\hat{\mathbf{H}}_{k,i_k}{+}\tilde{\mathbf{H}}_{k,i_k}$ (see (3.2)) and the MSE-matrix in (3.11) has to be averaged with respect to the conditional probability density function of the channel estimation errors $\tilde{\mathbf{H}}_{k,i_k}$ given the channel estimate $\hat{\mathbf{H}}_{k,i_k}$ [75], i.e. $\bar{\mathbf{E}}_{i_k}{=}\mathbb{E}\big[(\hat{\mathbf{b}}_{i_k}{-}\mathbf{b}_{i_k})(\hat{\mathbf{b}}_{i_k}{-}\mathbf{b}_{i_k})^H|\hat{\mathbf{H}}_{k,i_k}\big]$. Therefore, by using the channel estimation model in (3.2) and the channel estimation error characterization in (3.4), the averaged MSE-matrix results:

$$\bar{\mathbf{E}}_{i_k}(\mathbf{R}_{i_k},\{\mathbf{T}_{i_k}\}) = \mathbf{I}{+}\mathbf{R}_{i_k}^H\Big(\sum_{j_k\in\mathcal{I}_k}\hat{\mathbf{H}}_{k,i_k}\mathbf{T}_{j_k}\mathbf{T}_{j_k}^H\hat{\mathbf{H}}_{k,i_k}^H{+}\mathbf{N}_{i_k}\Big)\mathbf{R}_{i_k}$$
$$-\mathbf{R}_{i_k}^H\hat{\mathbf{H}}_{k,i_k}\mathbf{T}_{i_k}{-}\mathbf{T}_{i_k}^H\hat{\mathbf{H}}_{k,i_k}^H\mathbf{R}_{i_k}{+}\bar{J}_{k,i_k}\sum_{j_k\in\mathcal{I}_k}\text{Tr}\left(\mathbf{T}_{j_k}\mathbf{T}_{j_k}^H\right)\mathbf{R}_{i_k}^H\mathbf{R}_{i_k}, \quad (3.16)$$

being $\bar{J}_{k,i_k}$ shown in (3.4). In the imperfect CSI case, the optimal MMSE receiver filter[18] for the i_k-th UE derives as:

$$\bar{\mathbf{R}}_{i_k}^{\text{mmse}} = \Big(\sum_{j_k\in\mathcal{I}_k}\left(\hat{\mathbf{H}}_{k,i_k}\mathbf{T}_{j_k}\mathbf{T}_{j_k}^H\hat{\mathbf{H}}_{k,i_k}^H{+}\bar{J}_{k,i_k}\,\text{Tr}\left(\mathbf{T}_{j_k}\mathbf{T}_{j_k}^H\right)\mathbf{I}\right){+}\mathbf{N}_{i_k}\Big)^{-1}\hat{\mathbf{H}}_{k,i_k}\mathbf{T}_{i_k}. \quad (3.17)$$

So, the averaged MSE-matrix in (3.16) with the MMSE receive filter in (3.17) results:

$$\bar{\mathbf{E}}_{i_k}^{\text{mmse}}(\{\mathbf{T}_{i_k}\}){=}\Big(\mathbf{I}{+}\mathbf{T}_{i_k}^H\hat{\mathbf{H}}_{k,i_k}^H\Big(\sum_{\substack{j_k\in\mathcal{I}_k\\j_k\neq i_k}}\hat{\mathbf{H}}_{k,i_k}\mathbf{T}_{j_k}\mathbf{T}_{j_k}^H\hat{\mathbf{H}}_{k,i_k}^H{+}\bar{J}_{k,i_k}\sum_{j_k\in\mathcal{I}_k}\text{Tr}(\mathbf{T}_{j_k}\mathbf{T}_{j_k}^H)\mathbf{I}{+}\mathbf{N}_{i_k}\Big)^{-1}\hat{\mathbf{H}}_{k,i_k}\mathbf{T}_{i_k}\Big)^{-1}.$$
$$(3.18)$$

Under imperfect CSI, a lower bound for the achievable rate is derived in [137], whereby channel estimation errors appear as Gaussian noise. Note that the MMSE channel estimation for $\hat{\mathbf{H}}_{k,i_k}$ (see (3.3)) is the only channel estimate that satisfies that the equivalent additive noise (including noise, interference, and the residual channel estimation error) and the desired signal are uncorrelated [137]. This fact is exploited to derive the achievable rate lower bound in [137]. The lower bound for the i_k-th UE is:

$$\bar{R}_{i_k}^{\text{LB}}(\{\mathbf{T}_{i_k}\}){=}\log_2\Big|\mathbf{I}{+}\hat{\mathbf{H}}_{k,i_k}\mathbf{T}_{i_k}\mathbf{T}_{i_k}^H\hat{\mathbf{H}}_{k,i_k}^H\Big(\sum_{\substack{j_k\in\mathcal{I}_k\\j_k\neq i_k}}\hat{\mathbf{H}}_{k,i_k}\mathbf{T}_{j_k}\mathbf{T}_{j_k}^H\hat{\mathbf{H}}_{k,i_k}^H{+}\bar{J}_{k,i_k}\sum_{j_k\in\mathcal{I}_k}\text{Tr}(\mathbf{T}_{j_k}\mathbf{T}_{j_k}^H)\mathbf{I}{+}\mathbf{N}_{i_k}\Big)^{-1}\Big|.$$
$$(3.19)$$

At this point, we can realize that the achievable rate lower bound in (3.19) and the averaged MSE-matrix (under MMSE receive filter usage) in (3.18) are related as follows:

$$\bar{R}_{i_k}^{\text{LB}}\left(\{\mathbf{T}_{i_k}\}\right) = -\log_2\left|\bar{\mathbf{E}}_{i_k}^{\text{mmse}}(\{\mathbf{T}_{i_k}\})\right|. \quad (3.20)$$

Therefore, the same relation between the achievable rate lower bound and the averaged MSE-matrix as in the perfect CSI case for the achievable rate and the MSE-matrix is obtained (see (3.15)).

---

[18]The optimal MMSE receive filter under imperfect CSI is the one that minimizes the averaged MSE, i.e. $\text{Tr}(\bar{\mathbf{E}}_{i_k})$.

3.3.3 Signal model for uplink (phase 2)

Assuming perfect UL-DL propagation channel reciprocity (i.e. $\overleftarrow{\mathbf{H}}_{k,j_l} = \mathbf{H}_{k,j_l}^T$, $\forall k, \forall j_l, \forall l$), the received signal at the k-th BS in the UL transmission is given by:

$$\overleftarrow{\mathbf{y}}_k = \sum_{i_k \in \mathcal{I}_k} \mathbf{H}_{k,i_k}^T \overleftarrow{\mathbf{T}}_{i_k} \overleftarrow{\mathbf{s}}_{i_k} + \underbrace{\sum_{l \in \mathcal{K}, l \neq k} \sum_{j_l \in \mathcal{I}_l} \mathbf{H}_{k,j_l}^T \overleftarrow{\mathbf{T}}_{j_l} \overleftarrow{\mathbf{s}}_{j_l}}_{\text{interference}} + \overleftarrow{\mathbf{v}}_k, \tag{3.21}$$

where $\overleftarrow{\mathbf{T}}_{i_k}$ denotes the transmit filter used at the i_k-th UE, $\overleftarrow{\mathbf{s}}_{i_k}$ the stream of independent symbols sent by the i_k-th UE, and $\overleftarrow{\mathbf{v}}_k$ denotes the UL noise with distribution $\overleftarrow{\mathbf{v}}_k \sim \mathcal{CN}(\mathbf{0}, \overleftarrow{\sigma}_k^2 \mathbf{I})$. Thus, the covariance matrix of the received signal at the k-th BS $\mathbf{C}_{\overleftarrow{\mathbf{y}}_k}$ is given by:

$$\mathbf{C}_{\overleftarrow{\mathbf{y}}_k} = \mathbb{E}\left[\overleftarrow{\mathbf{y}}_k \overleftarrow{\mathbf{y}}_k^H\right] = \sum_{i_k \in \mathcal{I}_k} \mathbf{H}_{k,i_k}^T \overleftarrow{\mathbf{T}}_{i_k} \overleftarrow{\mathbf{T}}_{i_k}^H \mathbf{H}_{k,i_k}^* + \overleftarrow{\mathbf{N}}_k, \tag{3.22}$$

where $\overleftarrow{\mathbf{N}}_k$ denotes the covariance matrix of the received inter-cell interference plus noise at the k-th BS:

$$\overleftarrow{\mathbf{N}}_k = \sum_{l \in \mathcal{K}, l \neq k} \sum_{j_l \in \mathcal{I}_l} \mathbf{H}_{k,j_l}^T \overleftarrow{\mathbf{T}}_{j_l} \overleftarrow{\mathbf{T}}_{j_l}^H \mathbf{H}_{k,j_l}^* + \overleftarrow{\sigma}_k^2 \mathbf{I}. \tag{3.23}$$

3.4 Problem Formulation

With the objective of maximizing the total DL WSR of the system with a maximum transmitted power constraint per BS, the transmit filters at BSs (or, equivalently, the transmit precoding matrices) are designed as the solution to the following optimization problem:

$$(\text{P}_{3,1}): \underset{\{\mathbf{T}_{i_k}\}_{\forall k, \forall i_k}}{\text{maximize}} \quad \sum_{k \in \mathcal{K}} \sum_{i_k \in \mathcal{I}_k} \mu_{i_k} R_{i_k}(\{\mathbf{T}_{i_k}\}) \tag{3.24}$$

$$\text{subject to} \quad \sum_{i_k \in \mathcal{I}_k} \text{Tr}\left(\mathbf{T}_{i_k} \mathbf{T}_{i_k}^H\right) \leq P_k^{\max} \quad \forall k,$$

where μ_{i_k} is a weighting coefficient associated to the priority of the i_k-th UE, $R_{i_k}(\{\mathbf{T}_{i_k}\})$ is the achievable rate in (3.12), and $P_k^{\max}$ is the available transmit power at the k-th BS.

Due to interference, problem $(\text{P}_{3,1})$ in (3.24) is not convex w.r.t. $\{\mathbf{T}_{i_k}\}$ and the optimal solution cannot be guaranteed. Nevertheless, it is shown in [127] that one solution for transmit filters that attains a local optima of the maximum WSR problem $(\text{P}_{3,1})$ in (3.24) can be obtained by solving the following optimization problem that considers minimization of the total sum of WMSE:

$$(\text{P}_{3,2}): \underset{\substack{\{\mathbf{T}_{i_k}\}, \{\mathbf{R}_{i_k}\}, \\ \{\mathbf{W}_{i_k}\}_{\forall k, \forall i_k}}}{\text{minimize}} \quad \sum_{k \in \mathcal{K}} \sum_{i_k \in \mathcal{I}_k} \left(\text{Tr}\left(\mathbf{W}_{i_k} \mathbf{E}_{i_k}\left(\mathbf{R}_{i_k}, \{\mathbf{T}_{i_k}\}\right)\right) - \mu_{i_k} \log_2 \left| \frac{\ln(2)}{\mu_{i_k}} \mathbf{W}_{i_k} \right| \right) \tag{3.25}$$

$$\text{subject to} \quad \sum_{i_k \in \mathcal{I}_k} \text{Tr}\left(\mathbf{T}_{i_k} \mathbf{T}_{i_k}^H\right) \leq P_k^{\max} \quad \forall k,$$

where $\mathbf{W}_{i_k}$ is a weighting matrix associated to the i_k-th UE and $\mathbf{E}_{i_k}(\mathbf{R}_{i_k}, \{\mathbf{T}_{i_k}\})$ corresponds

to the MSE-matrix in (3.11). The main idea behind the equivalence of the maximum WSR problem ($P_{3,1}$) in (3.24) and the minimum sum WMSE problem ($P_{3,2}$) in (3.25)) relies on the relation between the rate and the MSE-matrix shown in (3.15).

Problem ($P_{3,2}$) in (3.25) is not jointly convex w.r.t. all sets of optimization variables, but it turns out to be convex w.r.t. each set of unknowns ($\{\mathbf{T}_{i_k}\}$, $\{\mathbf{R}_{i_k}\}$, and $\{\mathbf{W}_{i_k}\}$) separately. Furthermore, each of them can be derived analytically assuming that the other two sets of variables are fixed. Therefore, a block coordinate descent (BCD) approach [139, 140] can be followed to find a local optimum of the problem ($P_{3,2}$) in (3.25) by alternate optimization of transmit filters $\{\mathbf{T}_{i_k}\}$, receive filters $\{\mathbf{R}_{i_k}\}$, and weighting matrices $\{\mathbf{W}_{i_k}\}$. The attained solution is a local optimal solution of the maximum WSR problem ($P_{3,1}$) in (3.24) (see more details in [127] on the equivalence among the maximum WSR problem ($P_{3,1}$) in (3.24) and the minimum WMSE problem ($P_{3,2}$) in (3.25)).

In case a centralized procedure was implemented, it is required that all channel matrices from all BSs to all UEs (i.e. $\{\mathbf{H}_{k,j_l}\}$, $\forall k$, $\forall j_l$, $\forall l$) are collected in a central processor node, see [127]. The decentralized implementation proposed in [127], where $\{\mathbf{T}_{i_k}\}$ are updated at BSs while $\{\mathbf{R}_{i_k}\}$ and $\{\mathbf{W}_{i_k}\}$ are updated at UEs, also requires that each k-th BS knows the channel matrices towards all UEs in the network (i.e. $\{\mathbf{H}_{k,j_l}\}$, $\forall j_l$, $\forall l$) and that feedback links are available from each UEs towards all BSs to report the updated $\mathbf{R}_{i_k}$ and $\mathbf{W}_{i_k}$.

3.5   Decentralized Coordinated Precoding

Different from [127,133], in this section a decentralized coordinated precoding (D-CoP) design for maximizing the WSR in (3.24) with linear transmit/receive filters is presented. First, through the use of the minimum WMSE problem ($P_{3,2}$) in (3.25) the maximum WSR problem is decomposed into parallel subproblems (one per BS) and a decentralized solution for DL transmit filters design at BSs is derived in Section 3.5.1, either assuming perfect CSI or imperfect CSI of the direct channel matrices. Section 3.5.2 details the receive filter to be used at each UE. Second, how to acquire the required parameters for transmit filters design at each k-th BS is detailed in Section 3.5.3, where it is shown how to acquire the inter-cell interference plus noise covariance matrices ($\{\mathbf{N}_{i_k}\}$, $\forall i_k \in \mathcal{I}_k$) and how to acquire the interference-cost matrix ($\mathbf{\Upsilon}_k$) by exploiting an UL transmission. Finally, the iterative algorithm for D-CoP is presented in Section 3.5.4, subsuming the acquisition of the required parameters at each BS from UL and the simultaneous per-BS optimizations for DL, as shown in Fig. 3.2.

Let us define the following matrix that will allow us to decompose problem ($P_{3,2}$) in (3.25):

$$\mathbf{\Upsilon}_k = \sum_{l \in \mathcal{K}, l \neq k} \sum_{j_l \in \mathcal{I}_l} \mathbf{H}_{k,j_l}^H \mathbf{R}_{j_l} \mathbf{W}_{j_l} \mathbf{R}_{j_l}^H \mathbf{H}_{k,j_l}, \qquad (3.26)$$

which is referred to as the *interference-cost matrix* because it reflects the interference that can be created by the k-th BS towards the unintended UEs (i.e. $j_l, l \neq k$) and it can be included as a penalizing term for the design of the transmit filters at the k-th BS (see next decomposed problem in (3.27)).

Proposition 3.1. *Problem* $(P_{3,2})$ *in* (3.25) *can be decomposed into* K *parallel optimization problems (one per BS). The problem to be solved at the* k*-th BS for fixed* $\mathbf{\Upsilon}_k$ *and* $\{\mathbf{N}_{i_k}\}$, $\forall i_k \in \mathcal{I}_k$, *is:*

$$(P_{3,3}^k) : \underset{\substack{\{\mathbf{T}_{i_k}\},\{\mathbf{R}_{i_k}\}, \\ \{\mathbf{W}_{i_k}\}\forall i_k}}{\text{minimize}} \sum_{i_k \in \mathcal{I}_k} \left(\text{Tr}\left(\mathbf{W}_{i_k}\mathbf{E}_{i_k}\left(\mathbf{R}_{i_k}, \{\mathbf{T}_{i_k}\}, \mathbf{N}_{i_k}\right)\right) - \mu_{i_k}\log_2\left|\frac{\ln(2)}{\mu_{i_k}}\mathbf{W}_{i_k}\right| + \text{Tr}\left(\mathbf{\Upsilon}_k\mathbf{T}_{i_k}\mathbf{T}_{i_k}^H\right)\right)$$

$$(3.27)$$

$$\text{subject to} \quad \sum_{i_k \in \mathcal{I}_k} \text{Tr}\left(\mathbf{T}_{i_k}\mathbf{T}_{i_k}^H\right) \leq P_k^{\max},$$

where $\mathbf{E}_{i_k}(\mathbf{R}_{i_k}, \{\mathbf{T}_{i_k}\}, \mathbf{N}_{i_k})$ *is the MSE-matrix in* (3.11) *assuming that* $\mathbf{N}_{i_k}$ *is fixed.*

Proof. See Appendix 3.A, where it is shown that the gradients of problem $(P_{3,2})$ in (3.25) and problem $(P_{3,3}^k)$ in (3.27) with respect to $\{\mathbf{T}_{i_k}\}$, $\{\mathbf{R}_{i_k}\}$, $\{\mathbf{W}_{i_k}\}$, $\forall i_k$, are the same if $\mathbf{\Upsilon}_k$ and $\{\mathbf{N}_{i_k}\}$, $\forall i_k$, are fixed. ∎

3.5.1 Optimization at each BS

Perfect CSI case

Problem $(P_{3,3}^k)$ in (3.27) is convex w.r.t. each set of variables separately, and each of them can be derived analytically assuming that the other two sets are fixed. Therefore, each k-th BS can solve problem $(P_{3,3}^k)$ in (3.27) for given $\mathbf{\Upsilon}_k$ and $\{\mathbf{N}_{i_k}\}$, $\forall i_k \in \mathcal{I}_k$, with alternate optimization among the following three sets of variables:

$$
\begin{aligned}
\text{i}: \quad & \mathbf{R}_{i_k}^{(n)} = \left(\mathbf{A}_{i_k}^{(n-1)} + \mathbf{N}_{i_k}\right)^{-1}\mathbf{H}_{k,i_k}\mathbf{T}_{i_k}^{(n-1)}, \forall i_k, \\
\text{ii}: \quad & \mathbf{W}_{i_k}^{(n)} = \frac{\mu_{i_k}}{\ln(2)}\mathbf{E}_{i_k}^{-1}\left(\mathbf{R}_{i_k}^{(n)}, \{\mathbf{T}_{i_k}^{(n-1)}\}, \mathbf{N}_{i_k}\right), \forall i_k, \\
\text{iii}: \quad & \mathbf{T}_{i_k}^{(n)} = \left(\mathbf{B}_k^{(n)} + \mathbf{\Upsilon}_k + \lambda_k^{(n)}\mathbf{I}\right)^{-1}\mathbf{H}_{k,i_k}^H\mathbf{R}_{i_k}^{(n)}\mathbf{W}_{i_k}^{(n)}, \forall i_k,
\end{aligned}
\tag{3.28}
$$

where

$$\mathbf{A}_{i_k}^{(n-1)} = \sum_{j_k \in \mathcal{I}_k}\mathbf{H}_{k,i_k}\mathbf{T}_{j_k}^{(n-1)}\mathbf{T}_{j_k}^{(n-1)H}\mathbf{H}_{k,i_k}^H, \qquad \mathbf{B}_k^{(n)} = \sum_{j_k \in \mathcal{I}_k}\mathbf{H}_{k,j_k}^H\mathbf{R}_{j_k}^{(n)}\mathbf{W}_{j_k}^{(n)}\mathbf{R}_{j_k}^{(n)H}\mathbf{H}_{k,j_k},$$

n is the iteration number, and $\lambda_k^{(n)}$ denotes a non-negative dual variable associated to the per-BS power constraint in (3.27). At each iteration, $\lambda_k^{(n)}$ has to be optimized to meet the power constraint, which can be efficiently solved using convex optimization techniques (e.g. subgradient method) [92]. Hence, monotonic convergence of the decentralized problem $(P_{3,3}^k)$ in (3.27) is ensured for given $\mathbf{\Upsilon}_k$ and $\{\mathbf{N}_{i_k}\}$, $\forall i_k \in \mathcal{I}_k$. However, due to inter-cell interference, so far we cannot guarantee that if each k-th BS solves its own problem $(P_{3,3}^k)$ in (3.27) and all of them do so simultaneously, convergence of $(P_{3,2})$ in (3.25) would be achieved.

The proposed solution for DL transmit filters design in (3.28) differs from [127], where each BS updates the transmit filters $\{\mathbf{T}_{i_k}\}$, $\forall i_k$, by following the last equation in (3.28) and evaluating the expression of $\mathbf{\Upsilon}_k$ in (3.26) that requires knowledge of all channel matrices, all weighting matrices, and all receive filters used at all UEs (i.e. $\mathbf{H}_{k,j_l}$, $\mathbf{W}_{j_l}$ and $\mathbf{R}_{j_l}$, $\forall j_l$, $\forall l$).

Imperfect CSI case

When the direct channel matrices towards the served UEs (i.e. $\{\mathbf{H}_{k,i_k}\}$, $\forall i_k$) are acquired with errors at the k-th BS, a stochastic (or Bayesian) robust design [75] can be used to overcome the impact of channel estimation errors. In our case, so as to implement a robust DL transmit filters design, the averaged MSE-matrix $\bar{\mathbf{E}}_{i_k}(\mathbf{R}_{i_k}, \{\mathbf{T}_{i_k}\}, \mathbf{N}_{i_k})$ in (3.16) (assuming that $\mathbf{N}_{i_k}$ is fixed) has to be considered under the assumption that $\hat{\mathbf{H}}_{k,i_k}$ in (3.3) and $\bar{J}_{k,i_k}$ in (3.4), $\forall i_k \in \mathcal{I}_k$, are known at the k-th BS. Note that in this case the focus lies on the maximization of a weighted sum of the achievable rate lower bounds (see (3.20)).

Therefore, the decentralized problem to be solved at the k-th BS is $(\mathrm{P}_{3,3}^k)$ in (3.27) with $\bar{\mathbf{E}}_{i_k}(\mathbf{R}_{i_k}, \{\mathbf{T}_{i_k}\}, \mathbf{N}_{i_k})$ instead of $\mathbf{E}_{i_k}(\mathbf{R}_{i_k}, \{\mathbf{T}_{i_k}\}, \mathbf{N}_{i_k})$. Similarly as for the perfect CSI case in (3.27), the problem is convex w.r.t. each set of variables separately, and each of them can be derived analytically assuming that the other two sets are fixed. Therefore, each k-th BS can find a robust transmit filters design for given $\mathbf{\Upsilon}_k$ and $\{\mathbf{N}_{i_k}\}$, $\forall i_k \in \mathcal{I}_k$, with alternate optimization among the following three sets of variables:

$$
\begin{aligned}
&\text{i}: &&\mathbf{R}_{i_k}^{(n)} = \left(\mathbf{A}_{i_k}^{(n-1)} + \mathbf{N}_{i_k}\right)^{-1}\hat{\mathbf{H}}_{k,i_k}\mathbf{T}_{i_k}^{(n-1)}, \forall i_k, \\
&\text{ii}: &&\mathbf{W}_{i_k}^{(n)} = \tfrac{\mu_{i_k}}{\ln(2)}\bar{\mathbf{E}}_{i_k}^{-1}\left(\mathbf{R}_{i_k}^{(n)}, \{\mathbf{T}_{i_k}^{(n-1)}\}, \mathbf{N}_{i_k}\right), \forall i_k, \\
&\text{iii}: &&\mathbf{T}_{i_k}^{(n)} = \left(\mathbf{B}_k^{(n)} + \mathbf{\Upsilon}_k + \lambda_k^{(n)}\mathbf{I}\right)^{-1}\hat{\mathbf{H}}_{k,i_k}^H\mathbf{R}_{i_k}^{(n)}\mathbf{W}_{i_k}^{(n)}, \forall i_k,
\end{aligned}
\tag{3.29}
$$

where

$$
\begin{aligned}
\mathbf{A}_{i_k}^{(n-1)} &= \sum_{j_k \in \mathcal{I}_k}\left(\hat{\mathbf{H}}_{k,i_k}\mathbf{T}_{j_k}^{(n-1)}\mathbf{T}_{j_k}^{(n-1)H}\hat{\mathbf{H}}_{k,i_k}^H + \bar{J}_{k,i_k}\,\mathrm{Tr}\left(\mathbf{T}_{j_k}^{(n-1)}\mathbf{T}_{j_k}^{(n-1)H}\right)\mathbf{I}\right), \\
\mathbf{B}_k^{(n)} &= \sum_{j_k \in \mathcal{I}_k}\left(\hat{\mathbf{H}}_{k,j_k}^H\mathbf{R}_{j_k}^{(n)}\mathbf{W}_{j_k}^{(n)}\mathbf{R}_{j_k}^{(n)H}\hat{\mathbf{H}}_{k,j_k} + \bar{J}_{k,j_k}\,\mathrm{Tr}\left(\mathbf{R}_{j_k}^{(n)}\mathbf{W}_{j_k}^{(n)}\mathbf{R}_{j_k}^{(n)H}\right)\mathbf{I}\right),
\end{aligned}
$$

and $\bar{\mathbf{E}}_{i_k}(\mathbf{R}_{i_k}, \{\mathbf{T}_{i_k}\}, \mathbf{N}_{i_k})$ is the averaged MSE-matrix in (3.16) assuming that $\mathbf{N}_{i_k}$ is fixed and $\lambda_k^{(n)}$ denotes the non-negative dual variable associated to the per-BS power constraint in (3.27) (to be optimized at every n-th iteration).

3.5.2   Optimization at each UE

Given the transmit filters, $\{\mathbf{T}_{i_k}\}$, the optimal DL receive filters for each i_k-th UE are given by the MMSE receivers [136] (see (3.13)):

$$
\mathbf{R}_{i_k} = \mathbf{C}_{\mathbf{y}_{i_k}}^{-1}\mathbf{H}_{k,i_k}\mathbf{T}_{i_k},
\tag{3.30}
$$

where $\mathbf{C}_{\mathbf{y}_{i_k}}$ is defined in (3.8). As it is done in real deployments [141], each UE can compute $\mathbf{R}_{i_k}$ in (3.30) based on the estimation of the equivalent channel $\mathbf{H}_{k,i_k}\mathbf{T}_{i_k}$ and the covariance matrix of the received signal $\mathbf{C}_{\mathbf{y}_{i_k}}$. Then, each UE can also compute the weighting matrix $\mathbf{W}_{i_k} = \tfrac{\mu_{i_k}}{\ln(2)}\mathbf{E}_{i_k}^{-1}$ based on $\mathbf{C}_{\mathbf{y}_{i_k}}$, $\mathbf{H}_{k,i_k}\mathbf{T}_{i_k}$, and $\mathbf{R}_{i_k}$ using (3.11). In case of imperfect CSI, $\mathbf{H}_{k,i_k}\mathbf{T}_{i_k}$ in (3.30) has to be replaced by the equivalent channel estimate $\hat{\mathbf{H}}_{k,i_k}\mathbf{T}_{i_k}$ (see (3.17)).

3.5.3   Acquisition of parameters at each BS

Acquisition of $\mathbf{N}_{i_k}$, $\forall i_k \in \mathcal{I}_k$

We assume that $\mathbf{C}_{\mathbf{y}_{i_k}}$ is reported from the UE towards the serving BS through an UL feedback link (in Phase 2 - UL in Fig 3.2), such that the BS collects $\mathbf{C}_{\mathbf{y}_{i_k}}$, $\forall i_k \in \mathcal{I}_k$, from its served UEs. Let us recall that, as compared to [127] where feedback links are required from each UE towards all BSs in the network in order to report $\mathbf{R}_{i_k}$ and $\mathbf{W}_{i_k}$, in our case only one feedback link is needed from each UE towards the serving BS through which only $\mathbf{C}_{\mathbf{y}_{i_k}}$ is reported. To reduce the communication overhead, the fact that $\mathbf{C}_{\mathbf{y}_{i_k}}$ is a positive semidefinite matrix (and hence hermitian) can be exploited. Then, inter-cell interference plus noise covariance matrices $\{\mathbf{N}_{i_k}\}$, $\forall i_k \in \mathcal{I}_k$, can be easily estimated at the k-th BS based on $\{\mathbf{H}_{k,i_k}\}$, $\{\mathbf{T}_{i_k}\}$, and $\{\mathbf{C}_{\mathbf{y}_{i_k}}\}$, $\forall i_k \in \mathcal{I}_k$, by following (3.8):

$$\mathbf{N}_{i_k} = \mathbf{C}_{\mathbf{y}_{i_k}} - \sum_{j_k \in \mathcal{I}_k} \mathbf{H}_{k,i_k} \mathbf{T}_{j_k} \mathbf{T}_{j_k}^H \mathbf{H}_{k,i_k}^H. \tag{3.31}$$

So it is not needed to estimate the interfering channel matrices (i.e. $\mathbf{H}_{l,i_k}, \forall l \neq k$) to compute $\mathbf{N}_{i_k}$ in (3.9) at the k-th BS. Note, however, that acquisition of $\mathbf{N}_{i_k}$, $\forall i_k \in \mathcal{I}_k$ would not be required in case each k-th BS was in charge of updating only the transmit filters $\mathbf{T}_{i_k}$, $\forall i_k \in \mathcal{I}_k$, but in that case feedback of $\mathbf{R}_{i_k}$ and $\mathbf{W}_{i_k}$, $\forall i_k \in \mathcal{I}_k$, would be needed.

Acquisition of $\mathbf{\Upsilon}_k$

In order to avoid the complex task of estimating the most harmful interfering channel matrices that are needed to compute the interference-cost matrix $\mathbf{\Upsilon}_k$ in (3.26) (as is assumed in [127]), we propose obtaining an estimate of $\mathbf{\Upsilon}_k$ from the covariance matrix of the UL interference plus noise (i.e. $\overleftarrow{\mathbf{N}}_k$ in (3.23)) when properly designing UL transmit filters $\overleftarrow{\mathbf{T}}_{j_l}$ (see Phase 2 - UL in Fig 3.2). By doing so, estimation of interfering channel matrices is avoided as $\overleftarrow{\mathbf{N}}_k$ can be computed by subtracting the desired signals to the covariance matrix of the UL received signal at the k-th BS (see (3.22)).

Notice that $\overleftarrow{\mathbf{N}}_k^*$ in (3.23) and $\mathbf{\Upsilon}_k$ in (3.26) differ just in the noise term in case UEs employ as UL transmit filter $\overleftarrow{\mathbf{T}}_{j_l} = \mathbf{R}_{j_l}^* \mathbf{W}_{j_l}^{\frac{1}{2}*}$, $\forall j_l$, $\forall l$, where $\mathbf{W}_{j_l} = \mathbf{W}_{j_l}^{\frac{1}{2}} \mathbf{W}_{j_l}^{\frac{1}{2}H}$. Therefore, in case the available power at UEs is not a limiting factor, by using a precoded UL pilot-based transmission, we could have a biased estimate of $\mathbf{\Upsilon}_k$ as: $\hat{\mathbf{\Upsilon}}_k = \overleftarrow{\mathbf{N}}_k^* = \mathbf{\Upsilon}_k + \overleftarrow{\sigma}_k^2 \mathbf{I}$.

Nevertheless, usually UL transmit filters cannot be applied as such unless we take into account the maximum transmit power constraints for UL, i.e.

$$\mathrm{Tr}\left(\overleftarrow{\mathbf{T}}_{j_l} \overleftarrow{\mathbf{T}}_{j_l}^H\right) \leq P_{j_l}^{\mathrm{UE}}, \tag{3.32}$$

where $P_{j_l}^{\mathrm{UE}}$ is the available power at the j_l-th UE. The proposed solution is to scale the DL receive filters in (3.30) by a common scaling cell-wide factor $F < 1$ as: $\overleftarrow{\mathbf{T}}_{j_l} = \sqrt{F} \mathbf{R}_{j_l}^* \mathbf{W}_{j_l}^{\frac{1}{2}*}$. The scaling cell-wide factor F would be designed beforehand in order to ensure that the $\epsilon\%$ of the UEs fulfill the transmit power constraint in (3.32), i.e. $\mathrm{Pr}\left(\mathrm{Tr}\left(F\mathbf{R}_{j_l}^* \mathbf{W}_{j_l}^* \mathbf{R}_{j_l}^T\right) \leq P_{j_l}^{\mathrm{UE}}\right) = \epsilon$. However, it is assumed that those UEs not satisfying the constraint in (3.32) with the a priori selected F will transmit at maximum power by scaling the DL receive filters with a per-user factor $f_{j_l} < F$.

Therefore, the proposed UL transmit filter is:

$$\overleftarrow{\mathbf{T}}_{j_l} = \sqrt{f_{j_l}} \mathbf{R}_{j_l}^* \mathbf{W}_{j_l}^{\frac{1}{2}*}, \tag{3.33}$$

where

$$f_{j_l} = \begin{cases} F\,, & \text{if } \mathrm{Tr}\left(F\mathbf{R}_{j_l}^*\mathbf{W}_{j_l}^*\mathbf{R}_{j_l}^T\right) \le P_{j_l}^{\mathrm{UE}}, \\ \dfrac{P_{j_l}^{\mathrm{UE}}}{\mathrm{Tr}\left(\mathbf{R}_{j_l}^*\mathbf{W}_{j_l}^*\mathbf{R}_{j_l}^T\right)}, & \text{otherwise.} \end{cases} \tag{3.34}$$

By using this approach, we can obtain a biased estimate of $\mathbf{\Upsilon}_k$ in (3.26) from the covariance matrix of the UL interference plus noise $\overleftarrow{\mathbf{N}}_k$ in (3.23). Its expected value is:

$$\hat{\mathbf{\Upsilon}}_k = \frac{1}{F}\overleftarrow{\mathbf{N}}_k^* = \sum_{l \in \mathcal{K}, l \ne k} \sum_{j_l \in \mathcal{I}_l} \frac{f_{j_l}}{F} \mathbf{H}_{k,j_l}^H \mathbf{R}_{j_l} \mathbf{W}_{j_l} \mathbf{R}_{j_l}^H \mathbf{H}_{k,j_l} + \frac{\overleftarrow{\sigma}_k^2}{F}\mathbf{I} = \mathbf{\Upsilon}_k + \tilde{\mathbf{\Upsilon}}_k + \frac{1}{F}\overleftarrow{\sigma}_k^2\mathbf{I}, \tag{3.35}$$

where $\tilde{\mathbf{\Upsilon}}_k$ describes the error introduced by those UEs served by neighbor BSs ($l \ne k$) that are transmitting at maximum power (i.e. $f_{j_l} < F$):

$$\tilde{\mathbf{\Upsilon}}_k = \sum_{l \in \mathcal{K}, l \ne k} \sum_{j_l \in \mathcal{I}_l} \frac{1}{F}\left(f_{j_l} - F\right)\mathbf{H}_{k,j_l}^H \mathbf{R}_{j_l} \mathbf{W}_{j_l} \mathbf{R}_{j_l}^H \mathbf{H}_{k,j_l}. \tag{3.36}$$

The proposed approach presents a bias that depends on the noise power increased by F^{-1} and an additional matrix associated to the non-serving UEs that are transmitting at maximum power (see (3.35)). So, there is an evident trade-off when the UE transmit power is a limiting factor: if F is small, the errors in the estimation of $\mathbf{\Upsilon}_k$ in (3.35) come due to the increased noise power, while if F is large the errors come due to $\tilde{\mathbf{\Upsilon}}_k$ in (3.36) as the major part of the UEs will use a per-user scaling factor f_{j_l} lower than F (see (3.33)).

However, as is shown in [**C3**], the estimation errors in $\hat{\mathbf{\Upsilon}}_k$ are negligible and do not affect the system performance when properly selecting the scaling cell-wide factor F in (3.33) in such a way that a certain percentage of the UEs (which depends on the deployment) fulfill the transmit power constraint in (3.32) with $f_{j_l} = F$. The intuitive explanation for a SISO case comes in the following: the UEs highly interfered in the DL have a lower absolute value of the receive filter in the DL due to the interference (see (3.28)) such that they would use $f_{j_l} = F$ as scaling factor in (3.33). This way, no errors would appear from the contributions of the highly interfered users to (3.35) and, as they are the users which contribute more to (3.35), the errors from the other users would be negligible. The key point is the selection of the scaling cell-wide factor F to allow a group of UEs to fulfill the constraint, while not making it too small such that the contribution of the noise power predominates in (3.35).

Note that all what is needed to obtain $\hat{\mathbf{\Upsilon}}_k$ in (3.35) from the UL transmission is the received signal covariance matrix $\overleftarrow{\mathbf{N}}_k^*$ such that we can directly obtain the interference-cost matrix estimate as: $\hat{\mathbf{\Upsilon}}_k = \frac{1}{F}\overleftarrow{\mathbf{N}}_k^*$. $\overleftarrow{\mathbf{N}}_k^*$ can be obtained from the coordinated use of UL pilots.

We would like to remark that the proposed scheme for estimating $\mathbf{\Upsilon}_k$ avoids the estimation errors of the interfering channel matrices that comes up in the procedure proposed in [127]. In such a case, since every single channel matrix has to be estimated, the robust transmit filters

Table 3.1: INTERFERENCE-COST MATRIX ACQUISITION.

Technique	Perfect CSI	Imperfect CSI
D-WMMSE in [127]	$\mathbf{\Upsilon}_k$ in (3.26)	$\mathbf{\Phi}_k$ in (3.37)
proposed **D-CoP**	$\hat{\mathbf{\Upsilon}}_k$ in (3.35)	$\hat{\mathbf{\Upsilon}}_k$ in (3.35)

design in Section 3.5.1 should use the following matrix $\mathbf{\Phi}_k$ instead of $\mathbf{\Upsilon}_k$ in (3.29):

$$\mathbf{\Phi}_k = \sum_{l\in\mathcal{K}, l\neq k} \sum_{j_l\in\mathcal{I}_l} \left(\hat{\mathbf{H}}_{k,j_l}^H \mathbf{R}_{j_l} \mathbf{W}_{j_l} \mathbf{R}_{j_l}^H \hat{\mathbf{H}}_{k,j_l} + \bar{J}_{k,j_l} \operatorname{Tr}\left(\mathbf{R}_{j_l} \mathbf{W}_{j_l} \mathbf{R}_{j_l}^H \right) \mathbf{I} \right), \qquad (3.37)$$

which includes the estimation errors of the interfering channel matrices from the k-th BS towards all the unintended UEs (i.e. j_l, $l\neq k$). Nevertheless, in the proposed D-CoP an estimate of $\mathbf{\Upsilon}_k$ can be obtained from the UL transmission as a function of the real channel matrices (see (3.35)), hence avoiding channel estimation errors of the interfering links and resulting in enhanced robustness.

In this regard, Table 3.1 summarizes the interference-cost matrix that has to be used for transmit coordination depending on whether the interfering channels are estimated (as in the decentralized procedure in [127]) or not (as in the proposed D-CoP). Further, in case a centralized approach was used then $\mathbf{N}_{i_k}$ in (3.29) should also include the estimation errors from all BSs ($l\neq k$) towards the i_k-th UE, resulting even in a worse performance.

3.5.4 Algorithm for D-CoP

Algorithm 3.1 summarizes the iterative procedure to solve problem (P$_{3,2}$) in (3.25) in a decentralized manner. The procedure follows the two phases shown in Fig. 3.2. During the channel estimation phase (Phase 1), the direct channel matrices $\mathbf{H}_{k,i_k}$, $\forall i_k\in\mathcal{I}_k$, are acquired at each k-th BS, which remain constant during the optimization (Phase 2) where the iterative algorithm for D-CoP is implemented and DL/UL transmissions are duplexed. The algorithm for D-CoP (Phase 2) includes the acquisition of the required parameters in UL (i.e. $\mathbf{N}_{i_k}$ as is detailed in Section 3.5.3, $\forall i_k\in\mathcal{I}_k$, and $\mathbf{\Upsilon}_k$ as is shown in Section 3.5.3) and the simultaneous transmit filters designs that are performed at each BS for DL (i.e. (3.28) in Section 3.5.1 for perfect CSI or (3.29) in Section 3.5.1 for imperfect CSI).

N_{iter} denotes the number of iterations allowed, and t the iteration index. The algorithm starts from an initialization of the transmit filters $\{\mathbf{T}_{i_k}^{(0)}\}$, $\forall i_k\in\mathcal{I}_k$, at each k-th BS that satisfy the per-BS power constraint in (3.27). For simulation purposes, a suitable initialization of $\{\mathbf{T}_{i_k}^{(0)}\}$, $\forall i_k\in\mathcal{I}_k$, can be obtained by solving (P$_{3,3}^k$) in (3.27) using $\mathbf{\Upsilon}_k^{(0)}=\mathbf{0}$. Then, a DL transmission is carried out using $\{\mathbf{T}_{i_k}^{(0)}\}$ (Phase 2 - DL(0)), where UEs can evaluate the covariance matrix of the received signal $\mathbf{C}_{\mathbf{y}_{i_k}}$ and update the receive filter $\mathbf{R}_{i_k}^{(0)}$ and the weighting matrix $\mathbf{W}_{i_k}^{(0)}$ using (3.30) (line 3). See details in Section 3.5.2.

Then, at each iteration $t=1,...,N_{\text{iter}}$, the following steps are followed:

- *Feedback of* $\mathbf{C}_{\mathbf{y}_{i_k}}^{(t-1)}$ *to acquire* $\mathbf{N}_{i_k}^{(t-1)}$ *at BS* (Phase 2 - UL(t)): Each UE reports $\mathbf{C}_{\mathbf{y}_{i_k}}^{(t-1)}$ to the serving BS (line 8), and hence the BS can acquire the inter-cell interference plus noise covariance matrix $\mathbf{N}_{i_k}^{(t-1)}$ of served UEs using (3.31), $\forall i_k\in\mathcal{I}_k$ (line 9). See Section 3.5.3.

Algorithm 3.1 D-CoP to solve $(P_{3,2})$ in (3.25) with simultaneous per-BS optimizations

---

1: *# Phase 1: Channel estimation*
2: All BSs $(\forall k)$: estimate direct channel matrices $\{\mathbf{H}_{k,i_k}\}$, $\forall i_k \in \mathcal{I}_k$
3: *# Phase 2 - DL(0): DL transmission*
4: All BSs $(\forall k)$: initialize $\{\mathbf{T}_{i_k}^{(0)}\}$, $\forall i_k \in \mathcal{I}_k$, and transmit with $\{\mathbf{T}_{i_k}^{(0)}\}$
5: All UEs $(\forall i_k)$: compute $\mathbf{R}_{i_k}^{(0)}$ and $\mathbf{W}_{i_k}^{(0)}$ using (3.30) with $\mathbf{C}_{\mathbf{y}_{i_k}}$, $\mathbf{T}_{i_k}^{(0)}$
6: **for** $t = 1, ..., N_{\text{iter}}$ **do**
7:     *# Phase 2 - UL(t): Feedback of $C_{\mathbf{y}_{i_k}}^{(t-1)}$ to acquire $\mathbf{N}_{i_k}^{(t-1)}$ at BS*
8:     All UEs $(\forall i_k)$: report $\mathbf{C}_{\mathbf{y}_{i_k}}^{(t-1)}$ to the serving k-th BS
9:     All BSs $(\forall k)$: compute $\mathbf{N}_{i_k}^{(t-1)}$, $\forall i_k \in \mathcal{I}_k$, using (3.31) with $\mathbf{C}_{\mathbf{y}_{i_k}}^{(t-1)}$, $\mathbf{T}_{i_k}^{(t-1)}$
10:    *# Phase 2 - UL(t): UL transmission to acquire $\hat{\mathbf{\Upsilon}}_k^{(t)}$ at BS*
11:    All UEs $(\forall i_k)$: transmit a pilot signal with $\overleftarrow{\mathbf{T}}_{i_k}^{(t)}$ in (3.33) computed from $\mathbf{R}_{i_k}^{(t-1)}$, $\mathbf{W}_{i_k}^{(t-1)}$
12:    All BSs $(\forall k)$: estimate $\hat{\mathbf{\Upsilon}}_k^{(t)}$ using (3.35)
13:    *# Reproduction of variables and initialization of $\{\mathbf{T}_{i_k}\}$ for per-BS optimization*
14:    All BSs $(\forall k)$:
15:        - Reproduce $\mathbf{R}_{i_k}^{(t-1)}$ and $\mathbf{W}_{i_k}^{(t-1)}$, $\forall i_k$, using (3.30) with $\{\mathbf{T}_{i_k}^{(t-1)}\}$, $\forall i_k \in \mathcal{I}_k$, $\mathbf{N}_{i_k}^{(t-1)}$
16:        - Compute $\mathbf{T}_{i_k}^{\text{aux}}$, $\forall i_k \in \mathcal{I}_k$, using (3.28) if perfect CSI or (3.29) if imperfect CSI with $\mathbf{R}_{i_k}^{(t-1)}$, $\mathbf{W}_{i_k}^{(t-1)}$, $\forall i_k \in \mathcal{I}_k$, $\hat{\mathbf{\Upsilon}}_k^{(t)}$
17:    *# Simultaneous per-BS optimizations*
18:    All BSs $(\forall k,$ simultaneously$)$: solve $(P_{3,3}^k)$ in (3.27) for a fixed $\hat{\mathbf{\Upsilon}}_k^{(t)}$ and $\{\mathbf{N}_{i_k}^{(t-1)}\}$, $\forall i_k \in \mathcal{I}_k$, departing from $\mathbf{T}_{i_k}^{\text{aux}}$, $\forall i_k \in \mathcal{I}_k$, with alternate optimization of $\overline{\mathbf{R}}_{i_k}^{(n)}$, $\overline{\mathbf{W}}_{i_k}^{(n)}$, and $\overline{\mathbf{T}}_{i_k}^{(n)}$, $\forall i_k \in \mathcal{I}_k$, in (3.28) if perfect CSI or $\overline{\mathbf{R}}_{i_k}^{(n)}$, $\overline{\mathbf{W}}_{i_k}^{(n)}$, and $\overline{\mathbf{T}}_{i_k}^{(n)}$, $\forall i_k \in \mathcal{I}_k$, in (3.29) if imperfect CSI, to finally get $\mathbf{T}_{i_k}^{(t)}$
19:    *# Phase 2 - DL(t): DL transmission*
20:    All BSs $(\forall k)$; transmit with $\{\mathbf{T}_{i_k}^{(t)}\}$, $\forall i_k \in \mathcal{I}_k$
21:    All UEs $(\forall i_k)$: compute $\mathbf{R}_{i_k}^{(t)}$ and $\mathbf{W}_{i_k}^{(t)}$ using (3.30) with $\mathbf{C}_{\mathbf{y}_{i_k}}^{(t)}$, $\mathbf{T}_{i_k}^{(t)}$
22: **end for**

- *UL transmission to acquire $\hat{\mathbf{\Upsilon}}_k^{(t)}$ at BS* (Phase 2 - UL(t)): An UL transmission is carried out, where all UEs transmit with $\overleftarrow{\mathbf{T}}_{i_k}^{(t)}$ in (3.33) computed from $\mathbf{R}_{i_k}^{(t-1)}$ and $\mathbf{W}_{i_k}^{(t-1)}$ (line 11), such that each BS can acquire the interference-cost matrix $\hat{\mathbf{\Upsilon}}_k^{(t)}$ using (3.35) (line 12). See details in Section 3.5.3.

- *Reproduction of variables at BS and initialization of $\{\mathbf{T}_{i_k}\}$ for per-BS optimization*: Each BS reproduces $\{\mathbf{R}_{i_k}^{(t-1)}\}$ and $\{\mathbf{W}_{i_k}^{(t-1)}\}$, $\forall i_k \in \mathcal{I}_k$, based on the acquired $\{\mathbf{N}_{i_k}^{(t-1)}\}$, $\forall i_k \in \mathcal{I}_k$, and the transmit filters used in the previous iteration $\{\mathbf{T}_{i_k}^{(t-1)}\}$, $\forall i_k \in \mathcal{I}_k$ (line 15). Note that $\{\mathbf{R}_{i_k}^{(t-1)}\}$ and $\{\mathbf{W}_{i_k}^{(t-1)}\}$ coincide with the ones computed at the served UEs in the previous iteration if $\mathbf{C}_{\mathbf{y}_{i_k}}^{(t-1)}$ is perfectly reported. Then, with the reproduced $\{\mathbf{R}_{i_k}^{(t-1)}\}$ and $\{\mathbf{W}_{i_k}^{(t-1)}\}$, $\forall i_k \in \mathcal{I}_k$, each BS can update the transmit filters $\{\mathbf{T}_{i_k}\}$, $\forall i_k \in \mathcal{I}_k$, (denoted by $\{\mathbf{T}_{i_k}^{\text{aux}}\}$, $\forall i_k \in \mathcal{I}_k$) by using last equation in (3.28) if perfect CSI or in (3.29) if imperfect CSI (line 16). $\{\mathbf{T}_{i_k}^{\text{aux}}\}$ will be used as initialization for the optimization at BS.

- *Simultaneous per-BS optimizations*: By starting with $\{\mathbf{T}_{i_k}^{\text{aux}}\}$, $\forall i_k \in \mathcal{I}_k$, an alternate optimization is performed at each BS for fixed $\{\mathbf{N}_{i_k}^{(t-1)}\}$, $\forall i_k \in \mathcal{I}_k$, and $\hat{\mathbf{\Upsilon}}_k^{(t)}$ that have been acquired in the previous steps. Receive filters $\{\mathbf{R}_{i_k}\}$, weighting matrices $\{\mathbf{W}_{i_k}\}$, and trans-

mit filters $\{\mathbf{T}_{i_k}\}$, $\forall i_k \in \mathcal{I}_k$, are iteratively computed by using (3.28) if perfect CSI or by using (3.29) if imperfect CSI (line 18). See Section 3.5.1. The obtained transmit filters design are denoted by $\{\mathbf{T}_{i_k}^{(t)}\}$.

- *DL transmission* (Phase 2 - DL(t)): DL transmission is carried out using $\{\mathbf{T}_{i_k}^{(t)}\}$ (line 20), where UEs can evaluate the covariance matrix of the received signal $\mathbf{C}_{\mathbf{y}_{i_k}}^{(t)}$ and update the receive filter $\mathbf{R}_{i_k}^{(t)}$ and the weighting matrix $\mathbf{W}_{i_k}^{(t)}$ using (3.30) (line 21). See details in Section 3.5.2.

Let us emphasize the following key points about the proposed Algorithm 3.1 for D-CoP:

- it is not needed to feed back either the weighting matrix $\mathbf{W}_{i_k}^{(t-1)}$ or the updated receive filter $\mathbf{R}_{i_k}^{(t-1)}$ from the UE to the serving BS. They can be reproduced at the BS based on the acquired $\mathbf{N}_{i_k}^{(t-1)}$ and the transmit filter used in the previous iteration ($\mathbf{T}_{i_k}^{(t-1)}$) which are already known at the BS (see line 15). This way, the required feedback links are reduced as compared to [127], where it was assumed that $\mathbf{W}_{i_k}$ and $\mathbf{R}_{i_k}$ were fed back from each UE to all the BSs in the network, while we only need to feed back $\mathbf{C}_{\mathbf{y}_{i_k}}$ from each UE to the serving BS.

- it is not needed to report the updated transmit filter $\mathbf{T}_{i_k}^{(t)}$ from the BS to the served UE. In practical implementations, the UE estimates the equivalent channel $\mathbf{H}_{k,i_k} \mathbf{T}_{i_k}^{(t)}$ every time a DL data transmission is carried out, which is enough to compute $\mathbf{R}_{i_k}^{(t)}$ in (3.30) [141].

3.5.5 Convergence

Algorithm 3.1 ensures convergence (for N_{iter} sufficiently large) to a stationary point of problem (P$_{3,2}$) in (3.25) (which is also a stationary point of problem (P$_{3,1}$) in (3.24) [127]) if $\{\mathbf{\Upsilon}_k\}$, $\{\mathbf{N}_{i_k}\}$, and $\{\mathbf{H}_{k,i_k}\}$, $\forall i_k \in \mathcal{I}_k$, are perfectly acquired at each k-th BS. Furthermore, the convergence of Algorithm 3.1 for D-CoP is proven to be monotonic (i.e. at each iteration t, the WSR is increased).

Theorem 3.1. *If $\mathbf{\Upsilon}_k$, $\{\mathbf{N}_{i_k}\}$, and $\{\mathbf{H}_{k,i_k}\}$, $\forall i_k \in \mathcal{I}_k$, are obtained without errors at each k-th BS and all BSs solve simultaneously its decentralized problem (P$_{3,3}^k$) in (3.27), Algorithm 3.1 converges in a finite number of iterations to a stationary point of the WSR problem (P$_{3,1}$) in (3.24) and the convergence is monotonic.*

Proof. See Appendix 3.B. ∎

The monotonic convergence proof with simultaneous per-BS optimizations is highly relevant for practical implementation issues, as when only one iteration (or a few) of the algorithm can be performed such proof allows ensuring a performance improvement with parallel and independent per-BS optimizations.

Remark 3.1. *At the best of our knowledge, this is the first proof of monotonic convergence of an algorithm where all BSs update simultaneously their transmit filters to maximize the system WSR in (3.24) in an interfering multi-cell multi-user scenario with no information exchange among non-associated terminals.*

   The key point is that we exploit the convexity of the problem w.r.t. each variable and then, even if we perform parallel and simultaneous per-BS optimizations, by updating first all $\{\mathbf{N}_{i_k}\}$ and $\{\mathbf{R}_{i_k}\}$ (i.e. line 9 in Algorithm 3.1) and after performing the update of all $\{\boldsymbol{\Upsilon}_k\}$ and $\{\mathbf{T}_{i_k}\}$ (i.e. line 12 in Algorithm 3.1), convergence can be proven. Many proofs of convergence (but not monotonic) with simultaneous per-BS optimizations in the literature are based on the insertion of an additional term in the objective function of the problem that allows to linearize the objective function with respect to the optimization variables (see for example [134]), which is not needed in our case.

   If errors appear in the estimation of $\boldsymbol{\Upsilon}_k$, $\mathbf{H}_{k,i_k}$, or $\mathbf{N}_{i_k}$, the update of the DL transmit/receive filters at each k-th BS can be controlled by slightly reformulating the problem $(\mathrm{P}_{3,3}^k)$ in (3.27) so as to include a proximal point term [139] that controls large deviations due to the estimation errors, similarly as is done in [95]. However, as is pointed out before, the errors in the estimation of $\boldsymbol{\Upsilon}_k$ from the UL transmission are very low provided that the scaling cell-wide factor F in (3.33) is properly designed, and we consistently observe convergence of the proposed Algorithm 3.1 without including the proximal point term.

3.5.6   Complexity

The complexity of the proposed algorithm is related to the number of iterations (N_{iter}), as each iteration requires an UL and a DL transmission and they have to be carried out during the channel coherence time (see Fig. 3.2). In LTE-A TDD networks, UL and DL transmissions are duplexed according to a set of predefined patterns [61], where in a frame of 10 ms up to 2 DL and 2 UL transmissions can be duplexed. Hence, 2 iterations of the algorithm could be performed every 10 ms. In a low mobility scenario (mean user speed of 3 Km/h), the channel coherence time at the 3.5 GHz band (band devoted to the SCN [104]) results approximately of 25 ms. This way, during the channel coherence time, up to 5 iterations of the algorithm could be implemented. However, in many practical implementations only 1 iteration is allowed, and for that reason in the evaluations N_{iter}=1 is used in many cases (see Section 3.7), which allows improving the WSR due to the monotonic convergence proof in Theorem 3.1.

3.5.7   Errors in the estimation of the interference-cost matrix

Two different sources of error arise in the practical estimation of the interference-cost matrix $\hat{\boldsymbol{\Upsilon}}_k$ in (3.35):

- the use of non-orthogonal pilot sequences and
- non-ideal channel reciprocity conditions.

   The transmitted sequences should be orthogonal among neighbor UEs in order to properly estimate $\hat{\boldsymbol{\Upsilon}}_k$. However, so as to avoid such coordination, pilot sequences composed of random symbols can be used, such that if the sequence length is large enough the orthogonality of the sequences is nearly achieved. In Section 3.7.4 we evaluate the performance loss due to the use of non-orthogonal UL pilot signals to estimate $\hat{\boldsymbol{\Upsilon}}_k$.

   On the other hand, the acquisition of $\hat{\boldsymbol{\Upsilon}}_k$ in (3.35) assumes perfect UL-DL propagation channel reciprocity, which requires perfect hardware calibration of the radio-frequency chains in TDD systems [70, 71, 142]. Although perfect calibration can be achieved at BSs because the variability of surrounding scenario is relatively slow and hardware performances tend to be

stable, calibration at UEs may be imperfect due to environmental variations caused by time, power and/or temperature [74]. In this case, the UL and DL propagation channel matrices from the k-th BS to the j_l-th UE are related by [70]:

$$\overleftarrow{\mathbf{H}}_{k,j_l} = \mathbf{H}_{k,j_l}^T \mathbf{C}_{j_l}, \tag{3.38}$$

where $\mathbf{C}_{j_l}$ is a diagonal matrix containing calibration errors at the j_l-th UE. Therefore, calibration errors at UEs may affect the estimation of $\hat{\mathbf{\Upsilon}}_k$. In Section 3.7.4 we evaluate the performance loss due to non-ideal propagation channel reciprocity conditions to estimate $\hat{\mathbf{\Upsilon}}_k$.

3.6 Practical Implementation in 3GPP LTE-A

The proposed D-CoP defines that inter-cell interference management can be done by means of sensing the UL transmission and processing the received signal, provided that propagation channel reciprocity is available. In order that each BS gets the desired information for interference mitigation without disturbing the UL data transmission (i.e. $\hat{\mathbf{\Upsilon}}_k$ in (3.35)), as all what is needed is the UL received signal but not the decoded symbols, we propose to use the already defined sounding reference signals (SRS) in 3GPP LTE-A standard [17, 22] and coordinate its transmission. SRS can be used both to get channel knowledge $\mathbf{H}_{k,i_k}$ and to get knowledge of the interference-cost matrix $\hat{\mathbf{\Upsilon}}_k$ at BS. In the latter case, we will refer to the SRS as Co-SRS (coordinated SRS).

3.6.1 Basic principles

The basic principles of the proposed D-CoP with Co-SRS are:

- Data for a UE is only available at and transmitted from one BS.
- Interference-cost matrix is acquired at BS by exploiting the UL Co-SRS transmission.
- At each BS, precoding selection is performed to manage interference in a decentralized manner by using channel knowledge $\mathbf{H}_{k,i_k}$ of the serving UEs, the covariance matrix of the received interference $\mathbf{N}_{i_k}$ (reported by serving UEs) and the interference cost matrix $\hat{\mathbf{\Upsilon}}_k$ (estimated at BS from the UL transmission).
- For the selected precoding, each BS conducts autonomous scheduling and modulation and coding scheme (MCS) selection based on most recent CSI and interference information.

3.6.2 Procedure for D-CoP with Co-SRS

Fig. 3.3 shows the steps for D-CoP and the communication needed between a BS and its serving UE, as described below.

(a) First of all, BS acquires the channel matrix $\mathbf{H}_{k,i_k}$ using SRS transmitted by UE in the UL.

(b) Based on the channel knowledge $\mathbf{H}_{k,i_k}$, the BS selects a downlink filter and transmits precoded demodulation reference signals (DM-RS) and precoded data through the physical DL shared channel (PDSCH), while the UE estimates the equivalent precoded channel using DM-RS and implements an MMSE-IRC receiver for data demodulation.

(c) UL transmission is carried out, in which the active UEs transmit simultaneously Co-SRS so that the BS receives the signal not only from its served UE but also from UEs attached
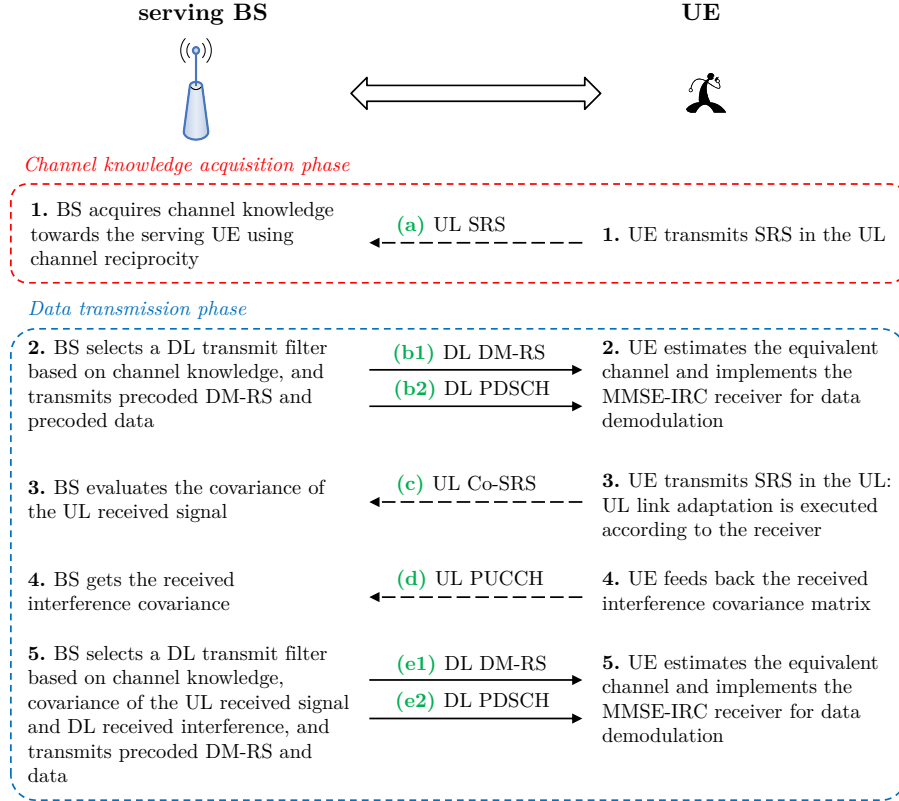
Figure 3.3: Communication between serving BS and its associated UE for D-CoP based on coordinated sounding.

to neighbor BSs. Each UE executes the link adaptation as a function of the MMSE-IRC receiver. Thanks to the use of this UL link adaptation and channel reciprocity, the covariance matrix of the received signal at the SeNB contains information about how the BS will interfere to unintended UEs present in neighbor BSs and $\hat{\mathbf{\Upsilon}}_k$ can be estimated.

(d) UE feeds back the received interference $\mathbf{N}_{i_k}$ to the serving BS through the physical UL control channel (PUCCH).

(e) Based on the channel knowledge $\mathbf{H}_{k,i_k}$ (step (a)), the covariance matrix of the received signal in the UL transmission (or, equivalently, $\hat{\mathbf{\Upsilon}}_k$) (step (c)) and the received interference $\mathbf{N}_{i_k}$ (step (d)), the BS designs a new DL transmit filter following a WSR criterion (by solving optimization problem $(\mathrm{P}_{3,3}^k)$ in (3.27)) and the DL data transmission is carried out through the PDSCH.

If steps (c), (d) and (e) were repeated, the system performance could be improved (see convergence considerations in Section 3.5.5). However, in practical implementations a single acquisition of $\hat{\mathbf{\Upsilon}}_k$ and report of $\mathbf{N}_{i_k}$ is usually allowed (as it has been represented in Fig. 3.3), which is shown through simulations to be enough to significantly improve the system throughput.

Co-SRS are useful to get knowledge of $\hat{\mathbf{\Upsilon}}_k$ in (3.35) through the air interface provided that UEs use an adequate link adaptation and that UL Co-SRS transmissions are simultaneous in time for all the active UEs, which requires coordination between cells. Hence, UE behavior and coordination/configuration of the Co-SRS transmissions are key points for the proposed D-CoP.

In the following sections we review both aspects, and provide different methods for practical implementation based on the modifications that the 3GPP LTE-A standard [17] could admit or not. Let us recall that in a downlink LTE-A system, the proposed D-CoP has to be applied to each subband (SB)[19] because the DL interference to be managed varies among SBs.

3.6.3   Limitations in LTE-A for UL sounding

The limitations in the LTE-A release 11-12 [22] are:

- Component SB specific UL power control is not available. So, for a given time instant, the user can only do the UL sounding adjusting UL link adaptation in a specific SB. Otherwise, a user could be assigned to multiple SBs simultaneously in time.

- The baseline for SRS operation in LTE-A is non-precoded and antenna-specific, with a minimum length equivalent to 4 resource blocks (RBs). This fact implies that if a UE is equipped with multiple antennas then multiple SRS transmissions multiplexed in time are needed to get the desired UL signal. If new precoded SRS were defined, only one SRS transmission would be required and the SRS length could be set to 1 RB such that more sounding granularity would be obtained to adapt the design according to channel variations in frequency.

3.6.4   UE behavior

In the UL transmission, precoded Co-SRS could be used so as to get the desired information $\hat{\mathbf{\Upsilon}}_k$ without disturbing the UL data transmission. As it is shown in (3.33), the UL transmit filter should be computed from the DL MMSE-IRC receiver. However, if non-precoded Co-SRS are used, multiple time instants (the same number as the number of UE receive antennas) are required to get the desired UL information $\hat{\mathbf{\Upsilon}}_k$: at each time instant Co-SRS should be transmitted on a different antenna. This fact entails a noise power increase of 3 dB for each additional antenna element in the estimation of the interference cost matrix $\hat{\mathbf{\Upsilon}}_k$.

3.6.5   Co-SRS configuration

For the proposed D-CoP, the proper configuration of the Co-SRS involves frequency-hopping mode [22], in which for a given time instant the Co-SRS are transmitted on a specific SB. This mode is needed because the sounding has to be different for each SB where the user is scheduled, as the DL interference varies. The periodicity of the Co-SRS and the hopping scheduling depend on the specific method adopted (which will be specified in next Section 3.6.6).

SRS are allocated to the last OFDM symbols of the Uplink Pilot Time Slot (UpPTS) in the synchronization (SYNC) subframe of a TDD system [22]. Hence, the minimum periodicity of the SRS is 5 ms, which corresponds to the DL-to-UL switch-point periodicity. Further, let us recall that as UpPTS is devoted for channel sounding and random access, the utilization of Co-SRS for D-CoP does not imply any additional overhead for coordination because UpPTS cannot be used for data transmission even if they are not used for channel sounding.

---

[19]In LTE-A, a subband is a collection of n adjacent physical resource blocks where the value of n can be 2, 3, 4, 6, or 8 depending on the channel bandwidth and the CQI feedback mode.

3.6.6   Methods for UL sounding

In the following, four different methods are described based on the modifications that LTE-A could admit or not. Methods A and B do not require any modification in LTE-A release 11 [22]. In Method A, the sounding is done in all SBs simultaneously in time, so that each user can only be scheduled to a single SB as the UE is not able to apply SB specific UL power control. In Method B, the sounding for each SB is done in different time instants, hence implying that a specific user could be scheduled to multiple SBs but the number of SBs is limited by the channel coherence time.

Method C uses component SB specific UL power control. Hence, a specific user can be assigned to multiple SBs simultaneously in time and do the UL sounding appropriately on each SB. Such procedure becomes independent of the number of SBs and allows more flexibility in the user scheduling process and Co-SRS periodicity than methods A and B.

Method D uses component SB specific UL power control and a new type of Co-SRS that include precoding and a minimum length equivalent to 1 RB. Then, a specific user can be assigned to multiple SBs simultaneously in time and do the UL sounding appropriately on each RB, and only one SRS transmission would be required. Method D is independent of the number of SBs, the number of antennas at the UE and inter-subband channel variations, hence allowing more flexibility in the user scheduling process, Co-SRS periodicity and sounding granularity than previous methods.

For each method, assumptions, Co-SRS configuration, requirements and implications are detailed in Table 3.2. Small Cell Scenario 2a in [23] and a low mobility scenario are used, such that the channel coherence time is 25 ms at the 3.5 GHz band and the maximum user speed is assumed to be 3 Km/h. The maximum Co-SRS periodicity is set to 20 ms, conditioned by the channel coherence time. The total bandwidth is described by the number of RBs, B, and S refers to the number of SBs in which the B RBs are spread. Requirements in Table 3.2 refer to:

- req1: Component SB specific UL power control is available.
- req2: Precoded and non-antenna specific SRS are available with a minimum length equivalent to 1 RB.

3.6.7   Signaling for Co-SRS

One example of the signaling required among BSs to implement D-CoP with Co-SRS is shown in Fig. 3.4, whereby UE_1 is served by BS_1, UE_2 is served by BS_2, and UE_1 might suffer strong interference from BS_2.

- Step 1: In subframe number 1 (SFN_1) each BS serves its associated UE and, based on the long-term channel information, each BS decides which cells should coordinate the SRS transmissions and sends semi-static signaling (containing SRS configurations) to them. In the example, BS_2 should be coordinated with BS_1 to avoid interference towards UE_1, so semi-static signaling (containing UE_1 SRS configuration) is sent to BS_2.
- Step 2: In SFN_{1+n} (corresponding to a SYNC subframe) each BS has all the relevant SRS configurations, so that the interference-cost matrix can be acquired by taking advantage of channel reciprocity. In the example, BS_2 gets the interference cost from UE_1.
- Step 3: In the specified SFN_{1+m} for DL transmission, all BSs start transmitting data to its attached UE meanwhile mitigating the inter-cell interference towards victim UEs.

Table 3.2: UL Sounding Methods using Co-SRS for $B=16$ RBs.

	Method A	Method B	Method C	Method D
Assumptions:				
1) user scheduling	for each BS, a single UE per SB is scheduled			
2) SB length	4 RBs	4 RBs	4 RBs	1 RB
3) number of SBs: S	4 SBs	4 SBs	4 SBs	16 SBs
4) max. SBs per UE	1	S	S	S
Co-SRS configuration:				
1) Co-SRS periodicity	20 ms	5 ms	20 ms	20 ms
2) Co-SRS mode	frequency-hopping			
Requirements:	-	-	req1	req1&req2
Implications:				
1) SRS granularity	4 RBs	4 RBs	4 RBs	1 RB
2) max. RBs per UE	4 RBs	16 RBs	16 RBs	16 RB
3) number of OFDM symbols per SRS	N_{i_k}	N_{i_k}	N_{i_k}	1
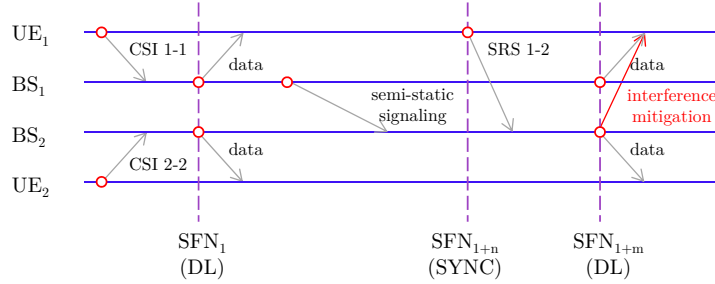
Figure 3.4: Inter-BS signaling for D-CoP with Co-SRS in 3GPP LTE-A systems. Only one interfering BS is considered.

3.7 Simulation Results

The evaluation of the proposed approaches is done on a LTE-A non-co-channel HCNs deployment [104], where macrocells and smallcells are not using the same carrier frequency: MeNBs operate at 2 GHz band and SeNBs use the 3.5 GHz band, both with 10 MHz bandwidth conformed of 50 RBs each. It is used the Small Cell Scenario 2a, following deployment and simulation parameters specified in [104]. The deployment consists of an hexagonal grid with 21 macrocell areas (each covered by a MeNB), as shown in Fig. 3.5. There are 7 macro-sites, each with 3 sector MeNBs. One cluster of SeNBs is deployed per macrocell area, each consisting of 4 or 10 SeNBs/cluster. 60 users are deployed per macrocell area, being 2/3 of them placed inside the cluster and the remaining UEs are uniformly distributed within the macrocell area (see Fig. 3.5). 80% of users are indoor and 20% of the users are placed outdoor. ITU Urban Macro and ITU Urban Micro models with 3D distance are used for path loss and shadowing modeling for MeNB-UE and SeNB-UE links, respectively. For fast fading modeling, the typical urban model is used. Transmit powers are 46 dBm, 30 dBm and 23 dBm at MeNB, SeNB and UE, respectively, and antenna gain of 17 dBi at MeNB, 5 dBi at SeNB and 0 dBi at UE. Noise spectral density is -174 dBm/Hz. The number of antennas is $M_{\text{MeNB}}=2$, $M_{\text{SeNB}}=2$ and $N_{\text{UE}}=2$ (such that $m_{i_k}=2, \forall i_k, \forall k$, in (3.5)).

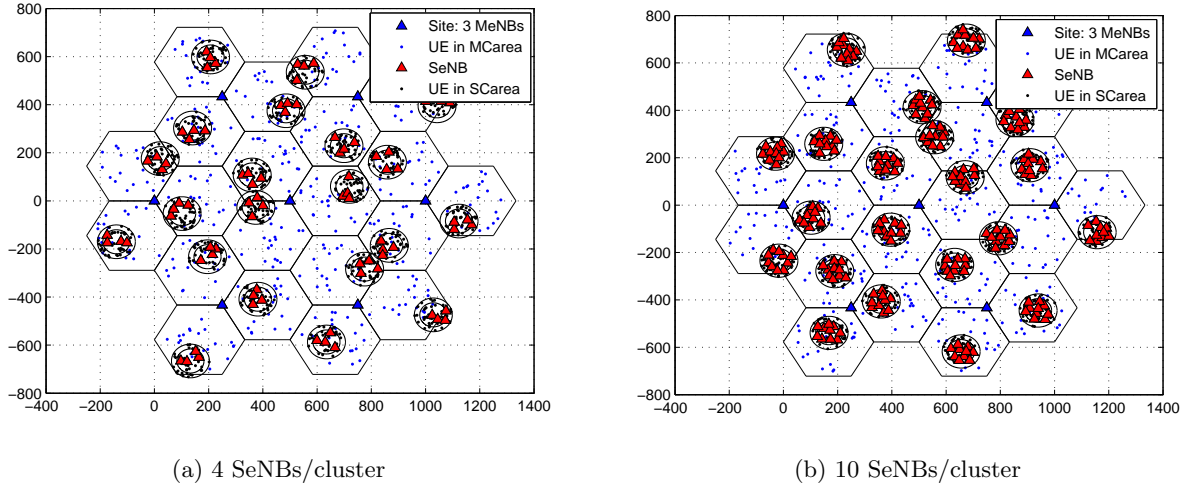(a) 4 SeNBs/cluster

(b) 10 SeNBs/cluster

Figure 3.5: Cluster, MeNB, SeNB and UE deployment for different layout configurations: 4 and 10 SeNBs/cluster.

Cell selection at each UE is based on reference-signal-received-quality (RSRQ) for inter-frequency selection and reference-signal-received-power (RSRP) for intra-frequency selection [104]. After the cell selection we focus on the SCN, i.e. the UEs that have been associated to SeNBs. The number (or percentage) of UEs offloaded to the SCN for each layout configuration is: 68% for 4 SeNBs/cluster and 75% for 10 SeNBs/cluster [143] (i.e. 40 and 45 UEs/cluster, respectively).

Two traffic models are implemented:

- *Full-load traffic model*, where all UEs in the network have packets to be transmitted.
- *FTP3 traffic model*, with various packet arrival rates (λ, in packets/s) that lead to different traffic load conditions [104]. In FTP3 traffic model, packets for the same UE arrive according to a Poisson process with arrival rate λ. The packet size is 0.5 Mbytes.

The active UEs associated to the same SeNB are uniformly distributed among the available RBs such that intra-cell interference is removed and only inter-cell interference remains. Hence, on each RB each SeNB serves a single UE (i.e. $|\mathcal{I}_k|=1, \forall k$, so $|\mathcal{I}|=|\mathcal{K}|$). The power available at each SeNB is uniformly distributed among the RBs where the active UEs have been scheduled.

The following techniques are evaluated at every RB (in which the system model in Section 3.3 applies):

- **D-IW**: decentralized procedure where each SeNB designs its transmit filter based on the channel knowledge $\mathbf{H}_{k,i_k}$ and acquired $\mathbf{N}_{i_k}$ from the served UE, but $\boldsymbol{\Upsilon}_k=\mathbf{0}$. It is equivalent to perform simultaneous MIMO iterative water-filling (IW) algorithms [144].
- **D-CoP**: decentralized interference management technique detailed in Section 3.5 and Algorithm 3.1 to solve problem ($P_{3,2}$) in (3.25), where each SeNB designs its transmit filter based on the knowledge of $\mathbf{H}_{k,i_k}$, $\mathbf{N}_{i_k}$ (acquired from its served UE), and $\hat{\boldsymbol{\Upsilon}}_k$ (estimated from an UL transmission as in (3.35)).
- **D-CoP ideal**: similar to *D-CoP*, but using the ideal value of $\boldsymbol{\Upsilon}_k$ in (3.26) at each SeNB.
- **D-WMMSE**: decentralized procedure presented in [127] to solve problem ($P_{3,2}$) in (3.25), where channel matrices from each SeNB to all UEs in the network are known at each k-th

SeNB (i.e. $\mathbf{H}_{k,j_l}, \forall j_l, \forall l$). Its performance is equal to *C-WMMSE*.

- **C-WMMSE**: centralized interference management procedure to solve problem $(\mathrm{P}_{3,2})$ in (3.25), in which case a central unit processor is assumed to collect channel matrices from all SeNBs to all UEs (i.e. $\mathbf{H}_{k,j_l}, \forall k, \forall j_l, \forall l$).

For all techniques, two optimization strategies are evaluated:

- *minWMSE*: solves $(\mathrm{P}_{3,2})$ in (3.25) with $\mathbf{W}_{i_k}=\mathbf{I}$, $\forall i_k$, $\forall k$,
- *maxWSR*: solves $(\mathrm{P}_{3,2})$ in (3.25) with $\mathbf{W}_{i_k}$ properly optimized so as to achieve the maximum WSR. $\mu_{i_k}=1$, $\forall i_k, \forall k$, is used.

The number of iterations is set to $N_{\mathrm{iter}}=1$, except for the convergence evaluations in Section 3.7.1, and in some cases $N_{\mathrm{iter}}=10$ is used for comparison purposes.

The performance indicator is *user packet throughput* (UPT) measured in Mbits/s and defined as the amount of data over the time needed to download data, without including the packet waiting time in the buffer. Maximum modulation and coding schemes from LTE-A and 3 dB of coding losses are used. For some purposes we evaluate the *per-user achievable rate* on a specific RB, according to (3.12) and measured in bits/s/Hz.

3.7.1   Convergence

The monotonic convergence of the proposed D-CoP is demonstrated on a specific RB without considering channel estimation errors, for the configuration of 4 and 10 SeNBs/cluster, when using the *full-load traffic model*. Fig. 3.6.(a) shows the convergence in terms of per-user MSE for *minWMSE* strategy (let us recall that $m_{i_k}=2$ is used for all UEs in the problem set up). Fig. 3.6.(b) displays the convergence in terms of per-user achievable rate (in bits/s/Hz) for *maxWSR* strategy. In both figures, the centralized solution is displayed with a dotted line.

It can be observed that D-CoP ideal has a monotonic convergence, while the proposed D-CoP based on propagation channel reciprocity also converges due to the fact that estimation errors in the acquisition of $\mathbf{\Upsilon}_k$ from UL are not relevant (see justification in Section 3.5.3). The convergence speed of D-CoP is faster than D-WMMSE [127] because with the proposed approach an alternate optimization can be performed at each SeNB on each iteration of the algorithm, hence achieving a faster convergence. C-WMMSE and D-WMMSE [127] approaches have the same performance result when a large number of iterations is used, because the equations employed to achieve the solution are equivalent.

The proposed D-CoP achieves a system performance very similar to C-WMMSE and D-WMMSE [127] approaches, but with much less overhead and complexity (in fact, the complexity of D-CoP is comparable to D-IW [144]). As the problem is non-convex w.r.t. all variables, all the interference management approaches only achieve a local minimum, so in some channel realizations they lead to the same solution (as in the *minWMSE* case in Fig. 3.6.(a)) while in others the attained locally optimal solution is different (as in the *maxWSR* case in Fig. 3.6.(b)). As compared to the baseline D-IW scheme [144], the reduction in the per-user MSE and the increase in the per-user achievable rate are larger in the layout of 10 SeNBs/cluster, as more interference is present in the SCN. Furthermore, it is important to recall that when performing only $N_{\mathrm{iter}}=1$ iteration, the proposed D-CoP already outperforms the baseline D-IW [144] scheme, which might not happen with D-WMMSE [127] because the decentralized optimizations at SeNBs with D-CoP allow converging faster to a stable point.
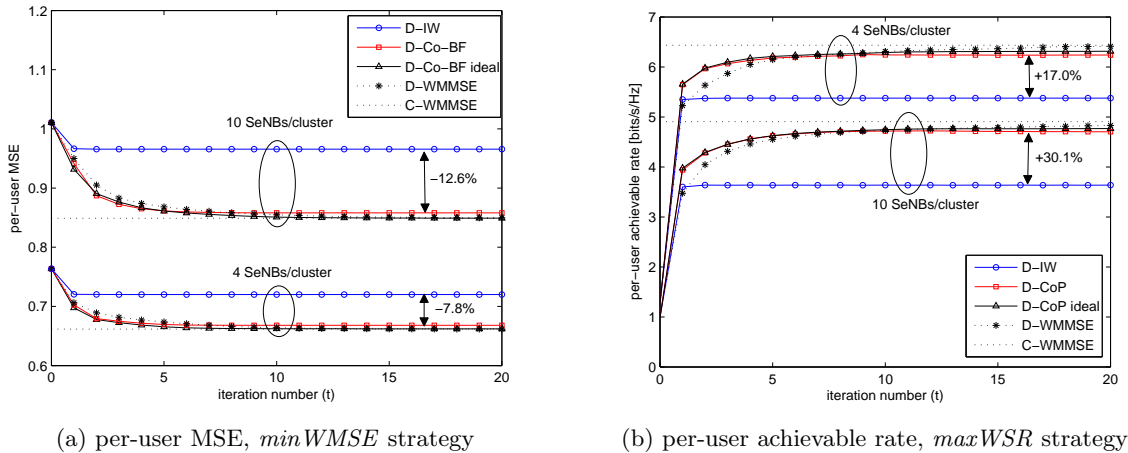
(a) per-user MSE, *minWMSE* strategy

(b) per-user achievable rate, *maxWSR* strategy

Figure 3.6: Convergence in terms of per-user MSE for *minWMSE* strategy and per-user achievable rate (in bits/s/Hz) for *maxWSR* strategy on a specific RB. Layout configuration: 4 and 10 SeNBs/cluster.

3.7.2   Evaluations under perfect CSI conditions

Fig. 3.7 displays the 5%-tile UPT vs. the mean UPT for the configuration of 4 and 10 SeNBs/cluster, when using the *full-load traffic model* and *minWMSE* or *maxWSR* strategy. The performance of D-CoP and D-WMMSE [127] are included for N_{iter}=1 and N_{iter}=10. On the one side, the *minWMSE* strategy allows improving specially the 5%-tile UPT, and N_{iter}=1 iteration is shown to be enough with D-CoP to get close to the performance given by the ideal centralized approach. On the other side, the *maxWSR* strategy provides larger gains in the mean UPT, but as the number of iterations increases the 5%-tile UPT is degraded. In this case the number of iterations is relevant in some cases, as with 10 iterations the mean UPT is enlarged specially for the layout of 10 SeNBs/cluster where more interference is present. However, the larger mean UPT gain with D-CoP comes in the first iteration, as is shown in Fig. 3.6.(b), because the decentralized optimization proposed at SeNBs with D-CoP allows a faster convergence (and it does not happen with D-WMMSE [127]).

Fig. 3.8 shows the 5%-tile UPT vs. the mean UPT for different values of the packet arrival rate (λ=0.5, 1, 1.5, 2 packets/s) of the Poisson distribution used in *FTP3 traffic model*, for the configuration of 4 and 10 SeNBs/cluster. *maxWSR* optimization strategy and N_{iter}=1 are used. D-CoP is compared to D-IW [144] that requires a similar complexity, and also to D-WMMSE [127] that requires more channels to be estimated and more feedback links. As expected, both the 5%-tile UPT and the mean UPT decrease as the offered load (i.e. λ) increases for all simulation cases. The UPT gains of D-CoP are larger for the 10 SeNBs/cluster deployment and they are also larger when λ increases, as more interference is present in the SCN either due to a denser deployment of SeNBs or due to a traffic load increase. The UPT gains provided with $N_{\text{iter}} = 1$ iteration are larger for D-CoP than for D-WMMSE [127], as was already observed in Fig. 3.6.(b).

In general, both in 5%-tile and mean UPT gains are obtained when using the *maxWSR* strategy due to the fact that each SeNB has 50 RBs available where the served UEs are uniformly distributed. Then, as each user is allocated to multiple RBs, the probability of being in outage in all the assigned RBs is very low and, even if we maximize the system WSR on each RB,
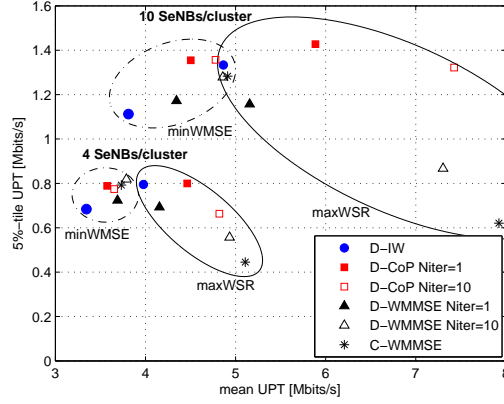
Figure 3.7: 5%-tile UPT (Mbits/s) vs. mean UPT (Mbits/s) in the full-load traffic model. Layout configuration: 4 and 10 SeNBs/cluster. *maxWSR* and *minWMSE* optimization strategies. N_{iter}=1 and N_{iter}=10.
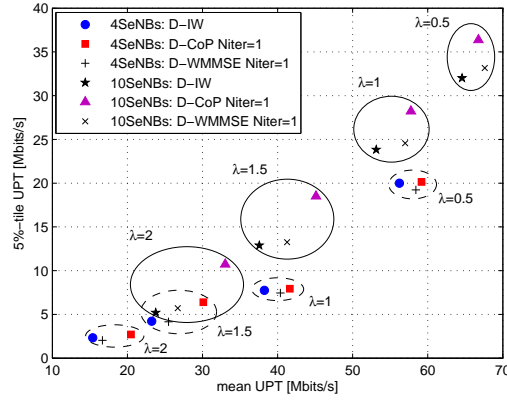

Figure 3.8: 5%-tile UPT (Mbits/s) vs. mean UPT (Mbits/s) for different packet arrival rates (λ in packets/s) in the FTP3 traffic model. Layout configuration: 4 and 10 SeNBs/cluster. *maxWSR* optimization strategy. N_{iter}=1.

we get an increase in the 5%-tile UPT. This effect is more remarkable in the 10 SeNBs/cluster deployment and for the *FTP3 traffic model*, where more resources are available.

To summarize, a significant UPT performance gain of the proposed D-CoP and N_{iter}=1 with respect to the baseline D-IW [144] scheme and with respect to D-WMMSE [127] is observed in all layout configurations and traffic simulation conditions. The relative gains are larger for medium-to-high traffic loads (up to a certain limit) and for denser deployments of SeNBs, due to the fact that D-CoP allows a full reuse of subbands and subframes while preemptively managing interference at UEs and, especially, at cell-edge UEs with a low complexity implementation. So it can be concluded that D-CoP is a suitable approach for interference limited scenarios and dense networks. It is important to recall that large UPT gains are obtained even if only one iteration of the proposed D-CoP is performed and that no additional overhead is needed for D-CoP if we use an UL pilot-based transmission to get the interference-cost matrix, which makes the approach amenable for practical system implementations.
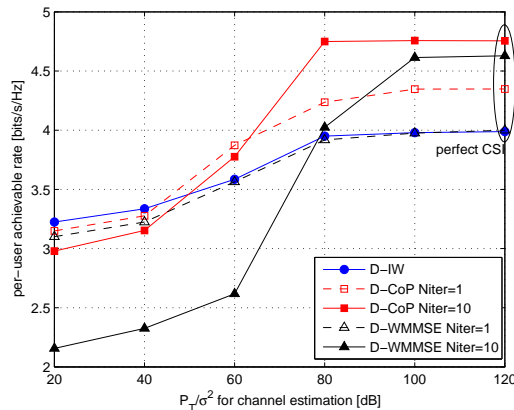
Figure 3.9: Per-user achievable rate on a specific RB (bits/s/Hz) vs. P_T/σ_v^2 used for channel estimation (dB) in the full-load traffic model. Layout configuration: 4 SeNBs/cluster. *maxWSR* optimization strategy. $N_{\text{iter}}=1$ and $N_{\text{iter}}=10$.

3.7.3   Evaluations under imperfect CSI conditions

As is shown in Table 3.1 and in Section 3.5.1, robust precoders are designed according to the adopted interference management procedure (D-WMMSE [127] or D-CoP). Fig. 3.9 displays the per-user achievable rate on a specific RB (in bits/s/Hz) vs. P_T/σ_v^2 used during the training phase (see definition in (3.4)) for the configuration of 4 SeNBs/cluster, when using the *full-load traffic model*. Let us recall that the results with $P_T/\sigma_v^2=120$ dB are nearly equal to the results with perfect CSI. For low values of P_T/σ_v^2, the performance loss due to imperfect CSI is larger for $N_{\text{iter}}=10$ than for $N_{\text{iter}}=1$ due to the propagation of errors in the iterative algorithms for D-WMMSE and D-CoP.

It can be observed that as P_T/σ_v^2 is reduced, the performance of the D-WMMSE [127] decreases drastically with $N_{\text{iter}}=10$ due to the errors in the estimation of interfering channels (see $\boldsymbol{\Phi}_k$ in (3.37)). However, the proposed D-CoP and the baseline D-IW [144] schemes are shown to be more robust to imperfect CSI because they only involve estimation errors in the direct channel matrix towards the serving SeNB. This shows that UPT gains can be obtained with the proposed D-CoP technique even if the channels are estimated with a low SNR, which is not the case with conventional transmit coordination techniques that require estimation of all the channel matrices. Let us recall that in case C-WMMSE approach was used, then the performance would be even worse than D-WMMSE [127] because $\mathbf{N}_{i_k}$ in (3.29) would also include the estimation errors of the interfering channel matrices.

3.7.4   Evaluations under real impairments to acquire the interference-cost

In this Section we evaluate the effect of real impairments on the acquisition of the interference-cost matrices $\hat{\boldsymbol{\Upsilon}}_k$ in (3.35) for the proposed D-CoP, as discussed in Section 3.5.7.

First, we consider the case in which non-orthogonal UL pilot sequences are used to estimate $\hat{\boldsymbol{\Upsilon}}_k$. Sequences composed of $\{+1, -1\}$ random symbols are used, with a length equal to S. Fig. 3.10 shows the per-user achievable rate (in bits/s/Hz) on a specific RB vs. S when using the *full-load traffic model* for the configuration of 10 SeNBs/cluster. $N_{\text{iter}}=1$ is used for D-CoP. It can be observed that if non-orthogonal random sequences are used to get $\hat{\boldsymbol{\Upsilon}}_k$, even with a low
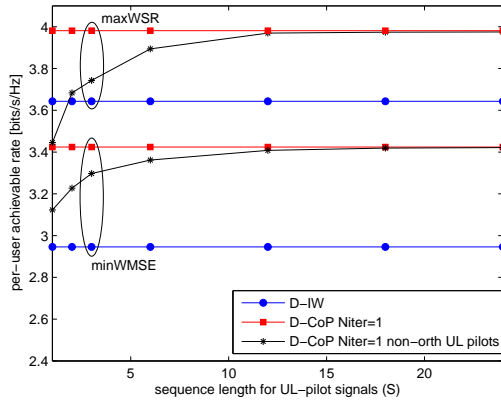
Figure 3.10: Per-user achievable rate (bits/s/Hz) on a specific RB vs. UL pilot sequence length (S) in the full-load traffic model. Layout configuration: 10 SeNBs/cluster. *minWMSE* and *maxWSR* strategy. N_{iter}=1.
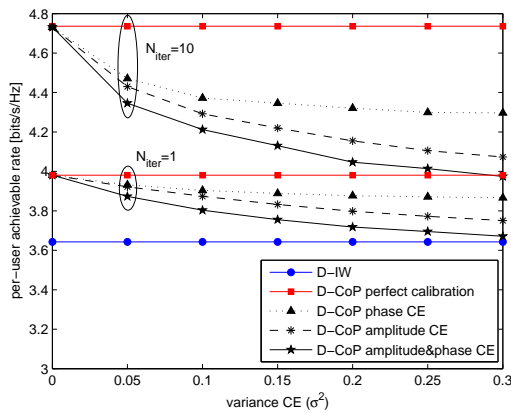

Figure 3.11: Per-user achievable rate (bits/s/Hz) on a specific RB vs. σ^2 (referring to the variance in phase and/or amplitude CE) in the full-load traffic model. Layout configuration: 10 SeNBs/cluster. *maxWSR* strategy. N_{iter}=1 and N_{iter}=10.

number of symbols (S), the performance is near to the one with orthogonal sequences among all UEs. Let us recall that in practical LTE-A implementations S=12 can be used within a RB, such that the effect of non-orthogonal UL pilot sequences would be negligible.

Second, we evaluate the impact of non-ideal channel reciprocity conditions to acquire $\hat{\mathbf{\Upsilon}}_k$. The diagonal entries of the calibration error matrix $\mathbf{C}_{j_l}$ in (3.38) are modeled by: $(1+a_{j_l})e^{j\phi_{j_l}}$, being $a_{j_l}{\sim}\mathcal{TN}(0,\sigma_a^2,-1,1)$ and $\phi_{j_l}{\sim}\mathcal{TN}(0,\sigma_\phi^2,-\pi/6,\pi/6)$ random variables with a truncated normal distribution that include the calibration error (CE) in amplitude and phase, respectively [70]. The variance for amplitude and phase CE at UEs (σ_a^2 and σ_ϕ^2) are assumed to be equal for all UEs, and are taken as an input for the simulations. Fig. 3.11 shows the per-user achievable rate (in bits/s/Hz) on a specific RB vs. σ^2 (which either refers to the variance of the amplitude CE, σ_a^2, and/or to the variance of the phase CE, σ_ϕ^2, as is indicated in the legend). We have used the configuration of 10 SeNBs/cluster, the *full-load traffic model* and *maxWSR* strategy. The performance of D-CoP is included for N_{iter}=1 and N_{iter}=10. It can be observed that the relative performance loss as compared to the perfect calibration case is larger for N_{iter}=10 than for N_{iter}=1, due to the propagation of errors in the iterative algorithm: at each iteration, the

interference-cost matrix is acquired with errors which impacts on the transmit/receive filters design and, on its turn, on the interference-cost matrix. As more iterations are done, higher impact of these errors on performance is observed. The impact of amplitude CE is a bit more harmful than phase CE, but in MIMO systems phase calibration at UEs is also important. However, even with high variances of the CE (note that, for example [71] considers values of $\sigma_a^2=\sigma_\phi^2=0.01$, while we are displaying values up to $\sigma_a^2=\sigma_\phi^2=0.3$), the performance of D-CoP outperforms the baseline D-IW [144] scheme.

3.8   Conclusions

This chapter presents a decentralized coordinated precoding (D-CoP) design for weighted sum-rate maximization in dense MIMO TDD smallcell networks. Precoding decisions are done at each BS based on the knowledge of the propagation channel towards its served UEs, the acquired DL interference, and the received signal in the UL. The received signal in the UL allows to estimate the interference-cost matrix, which informs about how the BS is interfering to unintended UEs. This way, it is not needed to estimate the interfering channels either at UEs or at BSs, no back-haul traffic between BSs is required, and minimal reporting of information from the UEs to the serving BS is needed. Then, an iterative algorithm for D-CoP is presented, which subsumes the acquisition of the required parameters and the simultaneous per-BS optimizations. Monotonic convergence of the algorithm is demonstrated when all BSs update its transmit precoders in parallel. Both the monotonic convergence and the parallel optimizations are highly relevant for practical implementation issues. Furthermore, a robust D-CoP is presented to tackle imperfect CSI conditions, and the impact of non-ideal propagation channel reciprocity is analyzed.

Significant gains in terms of user packet throughput, and specially 5%-tile UPT, are observed in 3GPP LTE-A SCNs for all layout configurations and all traffic simulation conditions, as compared to a baseline scheme with comparable complexity. The UPT gains are comparable to well-known centralized and decentralized interference management techniques, which involve much more complexity. The use of only one iteration of the algorithm allows improving the UPT and the relative gains are larger for medium-to-high traffic loads and for denser deployments of BSs, which demonstrates the suitable applicability of the proposed D-CoP in interference-limited scenarios. Further, the proposed scheme is shown through simulations to be robust to imperfect CSI conditions as well as real impairments in the estimation of the interference-cost matrix, like the use of non-orthogonal UL pilot sequences at UEs and the non-ideal propagation channel reciprocity conditions.

Appendices

3.A   Proof of Proposition 3.1

On the one hand, the Lagrangian function ($\mathcal{L}$) of the problem (P$_{3,2}$) in (3.25) is [92]:

$$
\begin{aligned}
\mathcal{L} \;=\; & \sum_{k\in\mathcal{K}}\sum_{i_k\in\mathcal{I}_k}\left(\mathrm{Tr}\left(\mathbf{W}_{i_k}\mathbf{E}_{i_k}(\mathbf{R}_{i_k},\{\mathbf{T}_{i_k}\})\right)-\mu_{i_k}\log_2\left|\tfrac{\ln(2)}{\mu_{i_k}}\mathbf{W}_{i_k}\right|\right) \\
& +\sum_{k\in\mathcal{K}}\lambda_k\left(\sum_{i_k\in\mathcal{I}_k}\mathrm{Tr}\left(\mathbf{T}_{i_k}\mathbf{T}_{i_k}^H\right)-P_k^{\max}\right),
\end{aligned}
\tag{3.39}
$$

being $\mathbf{E}_{i_k}(\mathbf{R}_{i_k},\{\mathbf{T}_{i_k}\})$ defined in (3.25).

On the other hand, the Lagrangian function ($\mathcal{L}_k$) of the decentralized problem (P$_{3,3}^k$) in (3.27) for fixed $\boldsymbol{\Upsilon}_k$ and $\mathbf{N}_{i_k}$, $\forall i_k$, is:

$$
\begin{aligned}
\mathcal{L}_k \;=\; & \sum_{i_k\in\mathcal{I}_k}\left(\mathrm{Tr}\left(\mathbf{W}_{i_k}\mathbf{E}_{i_k}(\mathbf{R}_{i_k},\{\mathbf{T}_{i_k}\},\mathbf{N}_{i_k})\right)-\mu_{i_k}\log_2\left|\tfrac{\ln(2)}{\mu_{i_k}}\mathbf{W}_{i_k}\right|+\mathrm{Tr}\left(\boldsymbol{\Upsilon}_k\mathbf{T}_{i_k}\mathbf{T}_{i_k}^H\right)\right) \\
& +\lambda_k\left(\sum_{i_k\in\mathcal{I}_k}\mathrm{Tr}\left(\mathbf{T}_{i_k}\mathbf{T}_{i_k}^H\right)-P_k^{\max}\right),
\end{aligned}
\tag{3.40}
$$

being $\mathbf{E}_{i_k}(\mathbf{R}_{i_k},\{\mathbf{T}_{i_k}\},\mathbf{N}_{i_k})$ the MSE-matrix in (3.25) assuming that $\mathbf{N}_{i_k}$ is fixed. Note that the Lagrangian function $\mathcal{L}_k$ and other $\mathcal{L}_l$, $\forall l\neq k$, are not coupled through any variable as the matrices containing inter-cell interference have been fixed (i.e. $\boldsymbol{\Upsilon}_k$ and $\mathbf{N}_{i_k}$, $\forall i_k$).

It is easy to check that the derivative of the Lagrangian functions in (3.39) and (3.40) with respect to $\mathbf{W}_{i_k}$ coincide, $\forall i_k$, $\forall k$, when $\boldsymbol{\Upsilon}_k$ is fixed. Further, it is also easy to check that the derivatives with respect to $\mathbf{R}_{i_k}$ are equal when $\boldsymbol{\Upsilon}_k$ is fixed, $\forall i_k$, $\forall k$, as $\mathbf{R}_{i_k}$ is only included in $\mathbf{E}_{i_k}$. So, let us focus on checking that the gradients coincide when we derive with respect to $\mathbf{T}_{i_k}$.

The derivative of $\mathcal{L}$ in (3.39) with respect to $\mathbf{T}_{i_k}$ is:

$$
\begin{aligned}
\frac{d\mathcal{L}}{d\mathbf{T}_{i_k}} \;=\; & \left(\sum_{j_k\in\mathcal{I}_k}\mathbf{H}_{k,j_k}^H\mathbf{R}_{j_k}\mathbf{W}_{j_k}\mathbf{R}_{j_k}^H\mathbf{H}_{k,j_k}\right)\mathbf{T}_{i_k}-\mathbf{H}_{k,i_k}^H\mathbf{R}_{i_k}\mathbf{W}_{i_k} \\
& +\left(\sum_{l\in\mathcal{K},l\neq k}\sum_{j_l\in\mathcal{I}_l}\mathbf{H}_{k,j_l}^H\mathbf{R}_{j_l}\mathbf{W}_{j_l}\mathbf{R}_{j_l}^H\mathbf{H}_{k,j_l}\right)\mathbf{T}_{i_k}+\lambda_k\mathbf{T}_{i_k},
\end{aligned}
\tag{3.41}
$$

where we have taken into account that $\mathbf{T}_{i_k}$ affects not only $\mathbf{E}_{i_k}$, but also the MSE-matrices of the remaining UEs (i.e. $\mathbf{E}_{j_l}$, $\forall j_l$, $\forall l$).

The derivative of $\mathcal{L}_k$ in (3.40) with respect to $\mathbf{T}_{i_k}$, when $\mathbf{\Upsilon}_k$ and $\mathbf{N}_{i_k}$, $\forall i_k$, are fixed, is:

$$\frac{d\mathcal{L}_k}{d\mathbf{T}_{i_k}} = \Big(\sum_{j_k \in \mathcal{I}_k} \mathbf{H}_{k,j_k}^H \mathbf{R}_{j_k} \mathbf{W}_{j_k} \mathbf{R}_{j_k}^H \mathbf{H}_{k,j_k} \Big) \mathbf{T}_{i_k} - \mathbf{H}_{k,i_k}^H \mathbf{R}_{i_k} \mathbf{W}_{i_k} + \mathbf{\Upsilon}_k \mathbf{T}_{i_k} + \lambda_k \mathbf{T}_{i_k}, \qquad (3.42)$$

where it is important to recall that in this case $\mathbf{T}_{i_k}$ only affects the MSE-matrices of the UEs served by the k-th BS (i.e. $\mathbf{E}_{j_k}$, $\forall j_k$), as $\mathbf{N}_{i_k}$, $\forall i_k$, have been fixed. Therefore, the gradients in (3.41) and in (3.42) are equal if $\mathbf{\Upsilon}_k = \sum_{l \in \mathcal{K}, l \neq k} \sum_{j_l \in \mathcal{I}_l} \mathbf{H}_{k,j_l}^H \mathbf{R}_{j_l} \mathbf{W}_{j_l} \mathbf{R}_{j_l}^H \mathbf{H}_{k,j_l}$ is set.

3.B   Proof of Theorem 3.1

Convergence of Algorithm 3.1 into a stationary point of the WSR problem (P$_{3,1}$) in (3.24) can be proved by grace of monotonic convergence of the objective function of the WMSE problem (P$_{3,2}$) in (3.25) [127]. Let us express the objective function of the WMSE problem (P$_{3,2}$) in (3.25) as follows:

$$f_o \left(\{\mathbf{T}_{i_k}\}, \{\mathbf{R}_{i_k}\}, \{\mathbf{W}_{i_k}\} \right) = \sum_{k \in \mathcal{K}} \sum_{i_k \in \mathcal{I}_k} \Big(\mathrm{Tr} \left(\mathbf{W}_{i_k} \mathbf{E}_{i_k} \left(\mathbf{R}_{i_k}, \{\mathbf{T}_{i_k}\} \right) \right) - \chi_{i_k} \Big), \qquad (3.43)$$

where $\chi_{i_k} = \mu_{i_k} \log_2 \left| \frac{\ln(2)}{\mu_{i_k}} \mathbf{W}_{i_k} \right|$.

$f_o(.)$ in (3.43) is a convex function w.r.t. each set of optimization variables separately (i.e. $\{\mathbf{T}_{i_k}\}$, $\{\mathbf{R}_{i_k}\}$ and $\{\mathbf{W}_{i_k}\}$). In particular, for fixed weighting matrices, $f_o(.)$ is convex w.r.t. transmit (receive) filters for a set of given receive (transmit) filters. In our proposed D-CoP procedure, we need to demonstrate a reduction of $f_o(.)$ in (3.43) whenever any of the three sets of optimization variables is updated. In particular, when performing lines 16, 18 and 21 in Algorithm 3.1. These updates are summarized in four steps:

- **Step 1** (line 16) (for all BSs): Update of the transmit filters $\{\mathbf{T}_{i_k}\}$, $\forall i_k \in \mathcal{I}_k$, (denoted by $\{\mathbf{T}_{i_k}^{\mathrm{aux}}\}$, $\forall i_k \in \mathcal{I}_k$) based on the acquired interference-cost matrix $\mathbf{\Upsilon}_k$ at each BS.
- **Step 2** (line 18) (for all BSs): Alternate optimization of the receive filters $\mathbf{R}_{i_k}$, weighting matrices $\mathbf{W}_{i_k}$, and transmit filters $\mathbf{T}_{i_k}$, $\forall i_k \in \mathcal{I}_k$, at each BS for a fixed $\{\mathbf{N}_{i_k}\}$, $\forall i_k \in \mathcal{I}_k$, and $\mathbf{\Upsilon}_k$.
- **Step 3** (line 21) (for all UEs): Update of the receive filter $\mathbf{R}_{i_k}$ based on the actual covariance matrix of the received inter-cell interference plus noise per UE $\mathbf{N}_{i_k}$.
- **Step 4** (line 21) (for all UEs): Update of the weighting matrix per UE $\mathbf{W}_{i_k}$.

In the following proof it is assumed that $\mathbf{\Upsilon}_k$, $\{\mathbf{N}_{i_k}\}$, and $\{\mathbf{H}_{k,i_k}\}$, $\forall i_k \in \mathcal{I}_k$, are acquired without errors at each k-th BS (so we focus on the perfect CSI case in Section 3.5.1).

When solving the optimization problem at each BS, DL transmit filters are designed at each k-th BS following the expressions in (3.28) for fixed $\{\mathbf{N}_{i_k}\}$, $\forall i_k \in \mathcal{I}_k$, in (3.9) and $\mathbf{\Upsilon}_k$ in (3.26). As a consequence, $f_o(.)$ in (3.43) can be written as:

$$f_o \big(\{\mathbf{T}_{i_k}(\mathbf{\Upsilon}_k)\}, \{\mathbf{R}_{i_k}(\mathbf{N}_{i_k})\}, \{\mathbf{W}_{i_k}\} \big) = \sum_{k \in \mathcal{K}} \sum_{i_k \in \mathcal{I}_k} \Big(\mathrm{Tr} \left(\mathbf{W}_{i_k} \mathbf{E}_{i_k} (\mathbf{R}_{i_k}(\mathbf{N}_{i_k}), \{\mathbf{T}_{i_k}(\mathbf{\Upsilon}_k)\}) \right) - \chi_{i_k} \Big).$$

$$(3.44)$$

The key point in the expression in (3.44) is that, for fixed weighting matrices $\{\mathbf{W}_{i_k}\}$, each transmit filter $\mathbf{T}_{i_k}$ depends on the receive filters $\{\mathbf{R}_{i_k}\}$ and on the interference-cost matrix $\mathbf{\Upsilon}_k$ that only depends on receive filters $\{\mathbf{R}_{i_k}\}$. Similarly, for fixed weighting matrices $\{\mathbf{W}_{i_k}\}$, each receive filter $\mathbf{R}_{i_k}$ depends on the transmit filter $\{\mathbf{T}_{i_k}\}$ and on the covariance matrix of the received inter-cell interference plus noise $\mathbf{N}_{i_k}$ that only depends on transmit filters $\{\mathbf{T}_{i_k}\}$. Now we are ready to prove convergence on each step of Algorithm 3.1. In the sequel, t denotes the iteration number.

Proof of objective function reduction in Step 1

Step 1 consists on updating $\mathbf{\Upsilon}_k^{(t)}$ and $\mathbf{T}_{i_k}^{\mathrm{aux}}, \forall i_k, \forall k$. If we use the fact that for the set of fixed $\{\mathbf{R}_{i_k}^{(t-1)}\}$ and $\{\mathbf{W}_{i_k}^{(t-1)}\}$, since $f_0(.)$ is convex w.r.t. $\{\mathbf{T}_{i_k}\}$, we can update sequentially: (a) all $\{\mathbf{\Upsilon}_k^{(t)}\}$, (b) all $\{\mathbf{T}_{i_k}^{\mathrm{aux}}\}$ following (3.28), and then convergence is guaranteed because transmit filters given by (3.28) are the optimum for a given set of $\{\mathbf{R}_{i_k}^{(t-1)}\}$ and $\{\mathbf{W}_{i_k}^{(t-1)}\}$ if $\{\mathbf{\Upsilon}_k^{(t)}\}$ is computed from the same set of $\{\mathbf{R}_{i_k}^{(t-1)}\}$ and $\{\mathbf{W}_{i_k}^{(t-1)}\}$. In other words:

$$f_o\big(\{\mathbf{T}_{i_k}^{(t-1)}\}, \{\mathbf{R}_{i_k}^{(t-1)}(\mathbf{N}_{i_k}^{(t-1)})\}, \{\mathbf{W}_{i_k}^{(t-1)}\}\big) \geq f_o\big(\{\mathbf{T}_{i_k}^{\mathrm{aux}}(\mathbf{\Upsilon}_k^{(t)})\}, \{\mathbf{R}_{i_k}^{(t-1)}(\mathbf{N}_{i_k}^{(t-1)})\}, \{\mathbf{W}_{i_k}^{(t-1)}\}\big).$$
(3.45)

Proof of objective function reduction in Step 2

Precoding design at each BS follows a BCD method [139] with alternate optimization between receive filters, weighting matrices, and transmit filters in (3.28). Therefore, an objective reduction for a fixed $\{\mathbf{N}_{i_k}^{(t-1)}\}$, $\forall i_k$, and $\mathbf{\Upsilon}_k^{(t)}$ at each k-th BS is guaranteed if the alternate optimization starts from $\{\mathbf{T}_{i_k}^{\mathrm{aux}}\}$. The variables as a result of the alternate optimization are denoted by: $\{\mathbf{T}_{i_k}^{(t)}\}$, $\{\mathbf{R}_{i_k}^{\mathrm{aux}}\}$, $\{\mathbf{W}_{i_k}^{\mathrm{aux}}\}$. As each subproblem converges, convergence of the sum is guaranteed for a fixed $\{\mathbf{N}_{i_k}^{(t-1)}\}$ and $\mathbf{\Upsilon}_k^{(t)}$, $\forall i_k, \forall k$:

$$f_o\big(\{\mathbf{T}_{i_k}^{\mathrm{aux}}(\mathbf{\Upsilon}_k^{(t)})\}, \{\mathbf{R}_{i_k}^{(t-1)}(\mathbf{N}_{i_k}^{(t-1)})\}, \{\mathbf{W}_{i_k}^{(t-1)}\}\big) \geq f_o\big(\{\mathbf{T}_{i_k}^{(t)}(\mathbf{\Upsilon}_k^{(t)})\}, \{\mathbf{R}_{i_k}^{\mathrm{aux}}(\mathbf{N}_{i_k}^{(t-1)})\}, \{\mathbf{W}_{i_k}^{\mathrm{aux}}\}\big).$$
(3.46)

Proof of objective function reduction in Step 3

Step 3 consists on updating $\mathbf{N}_{i_k}^{(t)}$ and $\mathbf{R}_{i_k}^{(t)}$, $\forall i_k, \forall k$. As $f_0(.)$ is convex w.r.t. $\{\mathbf{R}_{i_k}\}$ for a set of fixed $\{\mathbf{T}_{i_k}^{(t)}\}$ and $\{\mathbf{W}_{i_k}^{\mathrm{aux}}\}$, we can update sequentially: (a) all $\{\mathbf{N}_{i_k}^{(t)}\}$, (b) all $\{\mathbf{R}_{i_k}^{(t)}\}$ following (3.30), and thereby convergence is guaranteed since receivers given by (3.30) are the optimum for a given set of $\{\mathbf{T}_{i_k}^{(t)}\}$ if the used $\{\mathbf{N}_{i_k}^{(t)}\}$ is computed from the same set of transmit filters $\{\mathbf{T}_{i_k}^{(t)}\}$. This way:

$$f_o\big(\{\mathbf{T}_{i_k}^{(t)}(\mathbf{\Upsilon}_k^{(t)})\}, \{\mathbf{R}_{i_k}^{\mathrm{aux}}(\mathbf{N}_{i_k}^{(t-1)})\}, \{\mathbf{W}_{i_k}^{\mathrm{aux}}\}\big) \geq f_o\big(\{\mathbf{T}_{i_k}^{(t)}(\mathbf{\Upsilon}_k^{(t)})\}, \{\mathbf{R}_{i_k}^{(t)}(\mathbf{N}_{i_k}^{(t)})\}, \{\mathbf{W}_{i_k}^{\mathrm{aux}}\}\big).$$
(3.47)

Proof of objective function reduction in Step 4

Step 4 consists on updating $\mathbf{W}_{i_k}^{(t)}$, $\forall i_k, \forall k$. As $f_o(.)$ is convex w.r.t. $\{\mathbf{W}_{i_k}\}$, we can update them all by following (3.28) for a set of fixed $\{\mathbf{T}_{i_k}^{(t)}\}$ and $\{\mathbf{R}_{i_k}^{(t)}\}$, and thereby a reduction in the objective function in (3.43) is guaranteed, i.e.

$$f_o\big(\{\mathbf{T}_{i_k}^{(t)}(\mathbf{\Upsilon}_k^{(t)})\}, \{\mathbf{R}_{i_k}^{(t)}(\mathbf{N}_{i_k}^{(t)})\}, \{\mathbf{W}_{i_k}^{\mathrm{aux}}\}\big) \geq f_o\big(\{\mathbf{T}_{i_k}^{(t)}(\mathbf{\Upsilon}_k^{(t)})\}, \{\mathbf{R}_{i_k}^{(t)}(\mathbf{N}_{i_k}^{(t)})\}, \{\mathbf{W}_{i_k}^{(t)}\}\big). \qquad (3.48)$$

Objective function reduction at each iteration

Therefore, with the proof of an objective function reduction in these 4 consecutive steps, monotonic convergence of Algorithm 3.1 for D-CoP is demonstrated at each iteration t:

$$f_o\big(\{\mathbf{T}_{i_k}^{(t-1)}\}, \{\mathbf{R}_{i_k}^{(t-1)}(\mathbf{N}_{i_k}^{(t-1)})\}, \{\mathbf{W}_{i_k}^{(t-1)}\}\big) \geq f_o\big(\{\mathbf{T}_{i_k}^{(t)}(\mathbf{\Upsilon}_k^{(t)})\}, \{\mathbf{R}_{i_k}^{(t)}(\mathbf{N}_{i_k}^{(t)})\}, \{\mathbf{W}_{i_k}^{(t)}\}\big). \qquad (3.49)$$

Steps 3 and 4-1 could be interchanged and convergence would also be guaranteed. However, it is indispensable to update $\{\mathbf{N}_{i_k}\}$ and $\{\mathbf{\Upsilon}_k\}$ in different steps to guarantee convergence.

Chapter 4

Transmit Coordination for Cluster-based Joint Transmissions

This chapter investigates decentralized interference management in downlink (DL) of multi-cell MIMO TDD cellular networks where transmit cooperation is allowed at BSs. Each user (UE) is associated with a UE-centric cluster of BSs, which cooperatively serve the user through a joint transmission (or coordinated multi-point joint transmission (CoMP-JT)). Clusters of different users possibly share some BSs such that they may overlap, being coupled by interference and transmit power constraints at each BS. Our objective is the design of BSs clustering and precoding matrices per-user in order to maximize the weighted sum-rate of the system by controlling the interference and the power spent at BSs. In contrast to previous works where all channel matrices in the system are needed, we propose a decentralized procedure whereby only channel matrices towards a limited number of candidate BSs per user are required while interference is still controlled by using the signal received from an uplink transmission. For an LTE-compliant dense deployment of 2×2 MIMO BSs/users, results show gains of 6-16% in terms of sum-rate and 49-84% in terms of 5%-tile user rate (depending on the maximum cluster size) as compared to decentralized BS-disjoint clustering schemes.

The technical papers related to this topic are:

[C3]   S. Lagen, A. Agustin, J. Vidal, "Distributed Inter-Cluster Interference Management for CoMP-based cellular networks", *IEEE Global Commun. Conf.*, Atlanta (USA), Dec. 2013. doi: 10.1109/GLOCOM.2013.6831733.

[C4]   S. Lagen, A. Agustin, J. Vidal, B. Soret, K. I. Pedersen, "Distributed User-Centric Clustering and Precoding Design for CoMP Joint Transmission", *IEEE Global Commun. Conf.*, San Diego (USA), Dec. 2015. doi: 10.1109/GLOCOMW.2014.7063526.

This chapter contains results published in **[C4]** that deal with the joint design of clustering and precoding for CoMP-JT under a UE-centric clustering scheme. Differently, **[C3]** addresses the coordinated precoding design for CoMP-JT under a BS-centric clustering scheme, but it can be derived from **[C4]** with slight modifications when setting as fixed and disjoints the clusters for CoMP-JT. Both approaches are compared in this chapter.

This chapter is structured as follows. Section 4.1 contains the state of the art for CoMP-JT. The main contributions of the chapter are detailed in Section 4.2. In Section 4.3 the system model is presented. The problem formulation is set in Section 4.4. In Section 4.5 the decentralized joint clustering and precoding for cluster-based joint transmission is proposed. Simulation results are

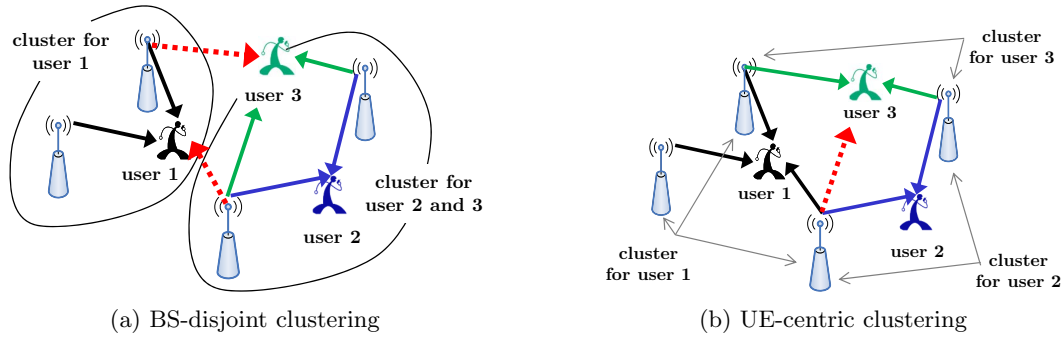(a) BS-disjoint clustering   (b) UE-centric clustering

Figure 4.1: BSs clustering schemes for CoMP-JT.

included in Section 4.6 for a 3GPP LTE SCN, whereby the BS-centric clustering and the UE-centric clustering schemes are compared. Section 4.7 analyzes the complementarities among CoMP-JT and the use of multiple antennas at BSs (known as large-scale MIMO systems [9]). Finally, concluding remarks are included in Section 4.8.

4.1   State of the Art

In CoMP-JT multiple geographically separated BSs transmit data to each user, such that data has to be shared among the coordinated BSs that will jointly transmit. Although large theoretical capacity gains are obtained, in practical implementations the observed gains of CoMP-JT become saturated with the number of cooperating BSs due to [64]:

- the overhead required to acquire knowledge of the channel matrices and
- the impact of channel estimation errors.

For that reason, the number of BSs in which CoMP-JT can take place (called the **cluster size** or CoMP set) has to be limited. In LTE-A the maximum cluster size is 3, see [38], being suitable for classical regular deployments. However, in a dense deployment of smallcells how to form the BSs clustering may not be trivial and challenging.

As it has been introduced in Section 1.4.2, there are two types of BSs clustering schemes for CoMP-JT: BS-disjoint clustering and UE-centric clustering (see Fig. 4.1). In the BS-disjoint clustering scheme, non-overlapping clusters of BSs are formed according to the deployment and BSs in each cluster jointly serve all users within their coverage area. This way, users at the cluster-edge still suffer from considerable interference and other kinds of coordination might be needed (see [**C3**]). Therefore, it is not the most suitable scheme from the users perspective and for dense and irregular deployments. In contrast, in the **UE-centric clustering** scheme, a cluster of BSs is selected per user such that different clusters of different users might overlap. In this case, the concept of cluster-edge users disappears but not the interference in the network.

Previous works in the literature have addressed the problem of coordinated precoding for CoMP-JT with BS-disjoint clustering scheme, either in a centralized or decentralized manner. For example, in [59] transmit beamformers filters are designed following an interference pre-cancellation technique based on zero-forcing, in which each cluster uses some degrees of freedom to pre-cancel interference to users in neighbor clusters. Recent procedures consider the exchange of control-plane messages among transmitters (interference prices) to manage the interference

in a decentralized manner [133] by selecting linear precoders, decoders and transmit powers. Methods are available to ensure the convergence of those algorithms (see, e.g., [144]), nevertheless, their main drawback is that interfering channels have to be estimated and reported to interfering transmitters along with the generated interference price. Hence, channel estimation errors on the interfering links and the overhead associated to the reporting of channel gains have a detrimental effect on the overall potential gains.

Previous works in the literature have addressed the problem of joint BSs clustering and precoding for CoMP-JT with UE-centric clustering scheme assuming that all BSs in a given area are candidates to form the cluster of each user [60, 82]. In [60], the problem is faced by introducing sparsity over the precoding matrices and penalizing the objective function so as to avoid larger cluster sizes. In [82], per-BS backhaul rate constraints are included in the problem in order to limit the cluster size according to the data sharing condition among BSs involved for CoMP-JT. However, in both cases the problem is solved in a centralized manner assuming that knowledge of all channel matrices from all BSs towards all users is available. In [60] decentralized implementation is discussed, but it requires again knowledge of all channel matrices at BSs.

4.2 Contribution

In this chapter we propose a decentralized procedure for joint BSs clustering and precoding (DJCP) that avoids estimation of all channel matrices. Our objective is the maximization of the DL WSR under per-BS power constraints (per-antenna power constraints could be imposed easily). We assume that each user selects a limited number of **candidate** BSs for its cluster and that one of them adopts the role of the BS **master** of the cluster. We decompose the maximum WSR problem into parallel subproblems to be solved at each master, which decide the BSs clustering and precoding for CoMP-JT in a coordinated manner with the rest of neighboring masters (see Fig. 4.2). The decomposition takes into account both the inter-cluster interference and the coupling per-BS power constraints, as one BS can be cooperating in the transmission towards different clusters of different users. To tackle interference in a decentralized manner we use the **interference-cost** concept [133] (presented in Chapter 3) and propose a procedure to acquire the interference-cost by exploiting an uplink (UL) pilot-based transmission in TDD, such that estimation of the interfering channel matrices is not needed while interference in the network can be managed. Finally, to further reduce the cluster size, a penalizing term is introduced in the formulation as a weighted sum of the power spent by candidate BSs in the cluster and we propose a rule to update the weights iteratively so as to take out of the cluster those BSs using too low power.

Different from previous works, [17, 60], in the proposed DJCP each master only requires knowledge of the channel matrices between the user and the candidate BSs in the cluster and estimation of the interfering channel matrices is not needed either at BSs or at users. Consequently, we avoid:

- the computational cost associated to channel estimation,
- the network planning required for pilot signals, and
- the performance loss due to imperfect estimation of channel matrices that are estimated with a low signal-to-noise ratio.
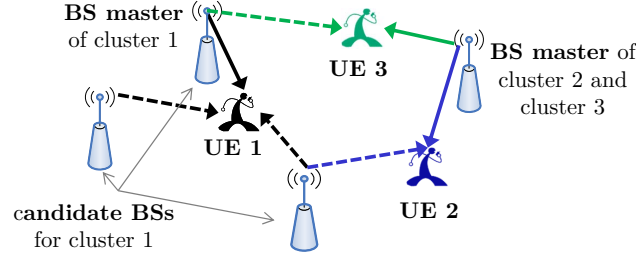
Figure 4.2: Example of UE-centric clustering for CoMP-JT in a cellular network composed of $|\mathcal{K}|$=4 BSs and $|\mathcal{I}|$=3 users. For decentralized implementation, each user selects a subset of candidate BSs to form its cluster. Candidate BSs will be controlled by the BS master of the cluster.

4.3   System Model

Consider a DL multi-cell network composed of a set of $\mathcal{K} \triangleq \{1, \ldots, K\}$ TDD BSs equipped with M_k antennas each (k=1, \ldots, K$) and a set of $\mathcal{I} \triangleq \{1, \ldots, I\}$ users with N_i antenna elements each (i=1, \ldots, I$). An example is shown in Fig. 4.2 for $|\mathcal{K}|$=4 BSs and $|\mathcal{I}|$=3 users.

Assume that each i-th user selects a subset of potential *candidate* BSs to form its UE-centric cluster, denoted by $\mathcal{C}_i \subseteq \mathcal{K}$. The subset of candidate BSs can be selected with different criteria: a maximum number per user can be established; or each user can choose the BSs from which the received power from the reference signals exceeds a certain threshold; or even a combination of both criteria could be adopted. Assume that one of the candidate BSs for the i-th user (e.g. the nearest BS) adopts the role of the BS *master* for the i-th cluster (see Fig. 4.2). The master is assumed to be connected through high-speed backhaul links with the candidate BSs in the cluster, and it will decide the BSs clustering and precoding design to serve the i-th user through CoMP-JT in coordination with the rest of masters. The final cluster of BSs for the i-th user is denoted by $\mathcal{B}_i \subseteq \mathcal{C}_i$.

To represent the BSs clustering and precoding in a compact form, we use network-wide matrices defined in the following. Let $\mathbf{H}_{k,i} \in \mathbb{C}^{N_i \times M_k}$ denote the channel matrix between the transmitting antennas at the k-th BS and the receiving antennas at the i-th user. Let $\mathbf{T}_{k,i} \in \mathbb{C}^{M_k \times m_i}$ denote the transmit precoder that the k-th BS uses to transmit m_i streams towards the i-th user. Define:

$$\mathbf{H}_i \triangleq [\mathbf{H}_{1,i}, \ldots, \mathbf{H}_{K,i}] \in \mathbb{C}^{N_i \times M_T} \tag{4.1}$$

as the channel matrix between the transmitting antennas at all BSs and the receiving antennas at the i-th user, being $M_T = \sum_{k \in \mathcal{K}} M_k$. Similarly, define:

$$\mathbf{T}_i \triangleq [\mathbf{T}_{1,i}^T, \ldots, \mathbf{T}_{K,i}^T]^T \in \mathbb{C}^{M_T \times m_i} \tag{4.2}$$

as the collection of all transmit precoders intended for the i-th user. Note that precoders corresponding to non-candidate BSs for the i-th user are directly set to $\mathbf{0}$: $\mathbf{T}_{k,i}=\mathbf{0}$, $\forall k \notin \mathcal{C}_i$. Furthermore, the BSs clustering for the i-th user (i.e. $\mathcal{B}_i$) will be obtained after optimization with the number of precoders $\mathbf{T}_{k,i}$ among $k \in \mathcal{C}_i$ that are different from $\mathbf{0}$.

We assume that for each i-th user only the channel matrices towards the candidate BSs (i.e. $\mathbf{H}_{k,i}, \forall k \in \mathcal{C}_i$) are known at the master of its cluster. Note that $\mathbf{H}_i$ in (4.1) contains both the candidate channels and the interfering channels (i.e. $\mathbf{H}_{k,i}, \forall k \notin \mathcal{C}_i$).

Let $\mathbf{x}$ denote the collection of transmitted signals of all BSs: $\mathbf{x} \triangleq [\mathbf{x}_1^T, \ldots, \mathbf{x}_K^T]^T \in \mathbb{C}^{M_T \times 1}$, being $\mathbf{x}_k \in \mathbb{C}^{M_k \times 1}$ the signal transmitted by the k-th BS, i.e.

$$\mathbf{x} = \sum_{i \in \mathcal{I}} \mathbf{T}_i \mathbf{b}_i, \quad \mathbf{x}_k = \sum_{i \in \mathcal{I}} \mathbf{T}_{k,i} \mathbf{b}_i, \tag{4.3}$$

where $\mathbf{b}_i \in \mathbb{C}^{m_i \times 1}$ contains the unit power independent Gaussian symbols of the i-th user (i.e. $\mathbf{b}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$). The total power spent at the k-th BS is:

$$P_k = \sum_{i \in \mathcal{I}} \text{Tr}\left(\mathbf{T}_{k,i} \mathbf{T}_{k,i}^H\right). \tag{4.4}$$

Assuming narrow-band transmissions, the equivalent baseband signal $\mathbf{y}_i \in \mathbb{C}^{N_i \times 1}$ observed at the i-th user is:

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{T}_i \mathbf{b}_i + \sum_{j \in \mathcal{I}, j \neq i} \mathbf{H}_i \mathbf{T}_j \mathbf{b}_j + \mathbf{v}_i, \tag{4.5}$$

where $\mathbf{v}_i$ refers to the additive zero-mean white Gaussian noise with distribution $\mathbf{v}_i \sim \mathcal{CN}(\mathbf{0}, \sigma_i^2 \mathbf{I})$. The symbols are estimated at the i-th user assuming that interference is treated as noise and that a linear receive filter $\mathbf{R}_i \in \mathbb{C}^{N_i \times m_i}$ is applied:

$$\hat{\mathbf{b}}_i = \mathbf{R}_i^H \mathbf{y}_i. \tag{4.6}$$

The MSE for the symbols transmitted towards the i-th user can be expressed through the so-called MSE-matrix $\mathbf{E}_i = \mathbb{E}\left[(\hat{\mathbf{b}}_i - \mathbf{b}_i)(\hat{\mathbf{b}}_i - \mathbf{b}_i)^H\right] \in \mathbb{C}^{m_i \times m_i}$. Under the independence assumption of $\{\mathbf{b}_i\}_{\forall i}$ and $\mathbf{v}_i$, the MSE-matrix results:

$$\mathbf{E}_i = \mathbf{I} + \mathbf{R}_i^H \mathbf{C}_{\mathbf{y}_i} \mathbf{R}_i - \mathbf{R}_i^H \mathbf{H}_i \mathbf{T}_i - \mathbf{T}_i^H \mathbf{H}_i^H \mathbf{R}_i, \tag{4.7}$$

where $\mathbf{C}_{\mathbf{y}_i} = \mathbb{E}\left[\mathbf{y}_i \mathbf{y}_i^H\right] \in \mathbb{C}^{N_i \times N_i}$ is the covariance matrix of the received signal at the i-th user:

$$\mathbf{C}_{\mathbf{y}_i} = \mathbf{H}_i \mathbf{T}_i \mathbf{T}_i^H \mathbf{H}_i^H + \mathbf{N}_i, \tag{4.8}$$

where $\mathbf{N}_i$ denotes the interference-plus-noise covariance matrix for the i-th user:

$$\mathbf{N}_i = \sum_{j \in \mathcal{I}, j \neq i} \mathbf{H}_i \mathbf{T}_j \mathbf{T}_j^H \mathbf{H}_i^H + \sigma_i^2 \mathbf{I}. \tag{4.9}$$

The receive filter that minimizes the MSE, i.e. $\text{Tr}(\mathbf{E}_i)$, is the MMSE receiver [127]:

$$\mathbf{R}_i^{\text{mmse}} = \mathbf{C}_{\mathbf{y}_i}^{-1} \mathbf{H}_i \mathbf{T}_i, \tag{4.10}$$

such that the MSE-matrix in (4.7) with $\mathbf{R}_i^{\text{mmse}}$ in (4.10) results:

$$\mathbf{E}_i^{\text{mmse}} = \mathbf{I} - \mathbf{T}_i^H \mathbf{H}_i^H \mathbf{C}_{\mathbf{y}_i}^{-1} \mathbf{H}_i \mathbf{T}_i = \left(\mathbf{I} + \mathbf{T}_i^H \mathbf{H}_i^H \mathbf{N}_i^{-1} \mathbf{H}_i \mathbf{T}_i\right)^{-1}. \tag{4.11}$$

The achievable rate of the i-th user is:

$$R_i = \log_2 \left|\mathbf{I} + \mathbf{H}_i \mathbf{T}_i \mathbf{T}_i^H \mathbf{H}_i^H \mathbf{N}_i^{-1}\right| = -\log_2 \left|\mathbf{E}_i^{\text{mmse}}\right|. \tag{4.12}$$

4.4   Problem Formulation

With the objective of maximizing the total DL WSR of the system with a maximum power constraint per BS as well as further reducing the cluster size if desirable, the joint BSs clustering and precoding are obtained as the solution to the following problem (inspired by [60]):

$$(\text{P}_{4,1}) : \underset{\{\mathbf{T}_i\}_{\forall i}}{\text{maximize}} \sum_{i\in\mathcal{I}} \left(\mu_i R_i - \alpha \sum_{k\in\mathcal{C}_i} \gamma_{k,i} \operatorname{Tr}\left(\mathbf{T}_{k,i}\mathbf{T}_{k,i}^H\right) \right) \tag{4.13}$$

$$\text{subject to} \sum_{i\in\mathcal{I}} \operatorname{Tr}\left(\mathbf{T}_{k,i}\mathbf{T}_{k,i}^H\right) \le P_k^{\max} \quad \forall k,$$

$$\mathbf{T}_{k,i} = \mathbf{0} \quad \forall k \notin \mathcal{C}_i, \forall i,$$

where μ_i is a weighting coefficient associated to the priority of the i-th user, R_i is the achievable rate shown in (4.12), and $P_k^{\max}$ is the available transmit power at the k-th BS. $\alpha \ge 0$ is a parameter that trades-off between maximum WSR performance and minimum weighted power spent at BSs:

- if $\alpha=0$ it reduces to plane maximum WSR, while
- if $\alpha>0$ the power spent at the BSs is taken into account with the goal of reducing the cluster size.

In addition to [60], we introduce the parameter $\gamma_{k,i}$ (per user and per BS) in (4.13), which will allow taking out of the cluster of the i-th user those BSs devoting little power towards that user (see details in Section 4.5.4).

Due to interference, problem $(\text{P}_{4,1})$ in (4.13) is not convex w.r.t. $\{\mathbf{T}_i\}_{\forall i}$ and the optimal solution cannot be guaranteed. Nevertheless, one solution attaining a local optimum of $(\text{P}_{4,1})$ in (4.13) can be obtained by solving the minimization of the total weighted sum of MSEs [127] while still keeping the minimization of the cluster size in the problem:

$$(\text{P}_{4,2}) : \underset{\{\mathbf{T}_i\},\{\mathbf{R}_i\},\{\mathbf{W}_i\}_{\forall i}}{\text{minimize}} \sum_{i\in\mathcal{I}} \left(\operatorname{Tr}\left(\mathbf{W}_i\mathbf{E}_i\right) - \mu_i \log_2\left|\tfrac{\ln(2)}{\mu_i}\mathbf{W}_i\right| + \alpha \sum_{k\in\mathcal{C}_i} \gamma_{k,i} \operatorname{Tr}\left(\mathbf{T}_{k,i}\mathbf{T}_{k,i}^H\right) \right) \tag{4.14}$$

$$\text{subject to} \quad \sum_{i\in\mathcal{I}} \operatorname{Tr}\left(\mathbf{T}_{k,i}\mathbf{T}_{k,i}^H\right) \le P_k^{\max} \quad \forall k,$$

$$\mathbf{T}_{k,i} = \mathbf{0} \quad \forall k \notin \mathcal{C}_i, \forall i,$$

where $\mathbf{W}_i \in \mathbb{C}^{m_i\times m_i}$ is a weighting matrix associated to the i-th user and $\mathbf{E}_i$ corresponds to the MSE-matrix shown in (4.7).

Problem $(\text{P}_{4,2})$ in (4.14) is convex w.r.t. each set of variables ($\{\mathbf{T}_i\}_{\forall i}$, $\{\mathbf{R}_i\}_{\forall i}$, and $\{\mathbf{W}_i\}_{\forall i}$) separately. Therefore, a block coordinate descent (BCD) approach [139] can be followed to find a local optimum by alternating the optimization between $\{\mathbf{T}_i\}_{\forall i}$, $\{\mathbf{R}_i\}_{\forall i}$, and $\{\mathbf{W}_i\}_{\forall i}$ if all channel matrices are known. The attained solution is a local optimal solution to problem $(\text{P}_{4,1})$ in (4.13), see [60] [127]. However, this procedure requires all channel matrices from all BSs to all users to be collected.

4.5   Decentralized Joint Clustering and Precoding

We focus on solving $(P_{4,2})$ in (4.14) in a decentralized manner. To do so, we split the problem in such a way that each master solves a subproblem in coordination with other masters. More specifically, first, problem $(P_{4,2})$ in (4.14) is decomposed into two problems:

- $(P_{4,3}^i)$ to be solved at the master of the i-th cluster so as to find the optimal BSs clustering and precoding (included in $\mathbf{T}_i$) when $\mathbf{R}_i$ and $\mathbf{W}_i$ are fixed, and
- $(P_{4,4}^i)$ to be solved at the i-th user so as to find $\mathbf{R}_i$ and $\mathbf{W}_i$ when $\mathbf{T}_i$ is fixed.

Then, how to acquire the required parameters for decentralized design is detailed. Finally, the iterative algorithm is presented, which subsumes the acquisition of the required parameters at the master of each i-th cluster and the simultaneous optimizations.

Let us define the following parameters that will allow decomposing problem $(P_{4,2})$ in (4.14).

- The interference-cost matrix is defined as:

$$\mathbf{\Upsilon}_i = \sum_{j \in \mathcal{I}, j \neq i} \mathbf{H}_j^H \mathbf{R}_j \mathbf{W}_j \mathbf{R}_j^H \mathbf{H}_j, \tag{4.15}$$

which reflects the DL interference that could be created by the i-th cluster towards unintended users (i.e. $j \neq i$, $j \in \mathcal{I}$) and it is seen as a penalizing term for the design of BSs clustering and precoding at the i-th cluster (see next problem in (4.20)). Let us partition $\mathbf{\Upsilon}_i$ in (4.15) as follows:

$$\mathbf{\Upsilon}_i = \begin{bmatrix} \mathbf{\Upsilon}_i[1,1] & \cdots & \mathbf{\Upsilon}_i[1,K] \\ \vdots & \ddots & \vdots \\ \mathbf{\Upsilon}_i[K,1] & \cdots & \mathbf{\Upsilon}_i[K,K] \end{bmatrix}, \tag{4.16}$$

such that (as $\mathbf{H}_j = [\mathbf{H}_{1,j}, \ldots, \mathbf{H}_{K,j}]$):

$$\mathbf{\Upsilon}_i[k,l] = \sum_{j \in \mathcal{I}, j \neq i} \mathbf{H}_{k,j}^H \mathbf{R}_j \mathbf{W}_j \mathbf{R}_j^H \mathbf{H}_{l,j}. \tag{4.17}$$

In Section 4.5.1 we show that not all blocks in (4.17) are needed to control interference, and Section 4.5.3 explains how to obtain the needed ones without having to estimate every single interfering channel matrix.

- The power spent by the k-th BS involved in the i-th cluster towards users in other clusters ($j \neq i$, $j \in \mathcal{I}$) is:

$$P_{k,-i} = \sum_{j \in \mathcal{I}, j \neq i} \mathrm{Tr}\left(\mathbf{T}_{k,j}(\mathbf{T}_{k,j})^H\right), \tag{4.18}$$

such that the per-BS power constraint in (4.14) can be decoupled as:

$$\mathrm{Tr}\left(\mathbf{T}_{k,i}\mathbf{T}_{k,i}^H\right) + P_{k,-i} \leq P_k^{\mathrm{max}}. \tag{4.19}$$

Differently from Chapter 3 where the precoding design at each BSs was coupled with neighboring BSs designs due to the interference-cost matrices (i.e. $\mathbf{\Upsilon}_i$ in (4.16)), a new coupling arises owing to the per-BS power constraints because a specific BS can belong to different clusters.

4.5.1   Optimization at the master of each cluster

Given $\{P_{k,-i}\}_{\forall i,\forall k\in\mathcal{C}_i}$, $\{\mathbf{R}_i\}_{\forall i}$, and $\{\mathbf{W}_i\}_{\forall i}$, problem (P$_{4,2}$) in (4.14) can be decomposed into I parallel optimization problems (one per master). The problem to be solved at the master of the i-th cluster (considering only terms in (4.14) that are affected by $\mathbf{T}_i$) for given $\{P_{-i}^k\}_{\forall k\in\mathcal{C}_i}$, $\{\mathbf{R}_i\}_{\forall i}$, and $\{\mathbf{W}_i\}_{\forall i}$, is:

$$(\text{P}_{4,3}^i) : \underset{\mathbf{T}_i}{\text{minimize}} \quad f_i + g_{-i} \tag{4.20}$$
$$\text{subject to } \text{Tr}\left(\mathbf{T}_{k,i}\mathbf{T}_{k,i}^H\right) + P_{k,-i} \leq P_k^{\max} \quad \forall k \in \mathcal{C}_i,$$
$$\mathbf{T}_{k,i} = \mathbf{0} \quad \forall k \notin \mathcal{C}_i,$$

where f_i measures the impact over the i-th user:

$$f_i = \text{Tr}\left(\mathbf{W}_i\mathbf{R}_i^H\mathbf{H}_i\mathbf{T}_i\mathbf{T}_i^H\mathbf{H}_i^H\mathbf{R}_i\right) - \text{Tr}\left(\mathbf{W}_i\mathbf{R}_i^H\mathbf{H}_i\mathbf{T}_i\right) - \text{Tr}\left(\mathbf{W}_i\mathbf{T}_i^H\mathbf{H}_i^H\mathbf{R}_i\right)$$
$$+\alpha \sum_{k\in\mathcal{C}_i} \gamma_{k,i}\,\text{Tr}\left(\mathbf{T}_{k,i}\mathbf{T}_{k,i}^H\right), \tag{4.21}$$

and g_{-i} considers the generated interference towards unintended users ($j\neq i$, see (4.15)):

$$g_{-i} = \text{Tr}\left(\mathbf{\Upsilon}_i\mathbf{T}_i\mathbf{T}_i^H\right). \tag{4.22}$$

The main part of the algorithm is to find the optimal BSs clustering and precoding (both included in $\mathbf{T}_i$) from (P$_{4,3}^i$). Problem (P$_{4,3}^i$) in (4.20) is convex w.r.t. $\mathbf{T}_i$ and the optimal structure for $\mathbf{T}_i$ could be directly obtained by deriving its Lagrangian function. However, as we are interested in controlling the cluster size, we have to work with the blocks of the network-wide matrix $\mathbf{T}_i = [\mathbf{T}_{1,i}^H, \ldots, \mathbf{T}_{K,i}^H]^H$, where only the blocks corresponding to candidate BSs selected by the i-th user have to be optimized (i.e. $\mathbf{T}_{k,i}, \forall k\in\mathcal{C}_i$). Accordingly, as only these block matrices of $\mathbf{T}_i$ can be different from $\mathbf{0}$, the blocks of $\mathbf{\Upsilon}_i$ in (4.16) that are needed for decentralized optimization at the master of the i-th cluster are those $\mathbf{\Upsilon}_i[k,l]$ such that $k\in\mathcal{C}_i$ and $l\in\mathcal{C}_i$ (see (4.22)). Note that problem (P$_{4,3}^i$) in (4.20) is formulated in a way such that only knowledge of $\mathbf{\Upsilon}_i[k,l]_{\forall k,l\in\mathcal{C}_i}$, $\{P_{k,-i}\}_{\forall k\in\mathcal{C}_i}$, $\mathbf{R}_i$, and $\mathbf{W}_i$ is required. So from now on, let us assume they are given at the master of the i-th cluster (which is equivalent but less restrictive to assume that $\{P_{-i}^k\}_{\forall k\in\mathcal{C}_i}$, $\{\mathbf{R}_i\}_{\forall i}$, and $\{\mathbf{W}_i\}_{\forall i}$ are given).

Interestingly, problem (P$_{4,3}^i$) in (4.20) is separable among the precoders of different BSs, such that a BCD method [139] can be applied with $\{\mathbf{T}_{k,i}\}_{\forall k\in\mathcal{C}_i}$ as block variables. To do so, we follow similar steps as in [60] but extended to the multi-stream case per user and applied to solve problem (P$_{4,3}^i$) in (4.20).

Define the following two sets of variables:

$$\mathbf{J}_i \triangleq \mathbf{H}_i^H\mathbf{R}_i\mathbf{W}_i\mathbf{R}_i^H\mathbf{H}_i, \quad \mathbf{D}_i \triangleq \mathbf{H}_i^H\mathbf{R}_i\mathbf{W}_i. \tag{4.23}$$

Partition $\mathbf{J}_i\in\mathbb{C}^{M_T\times M_T}$ and $\mathbf{D}_i\in\mathbb{C}^{M_T\times m_i}$ in (4.23) as:

$$\mathbf{J}_i = \begin{bmatrix} \mathbf{J}_i[1,1] & \cdots & \mathbf{J}_i[1,K] \\ \vdots & \ddots & \vdots \\ \mathbf{J}_i[K,1] & \cdots & \mathbf{J}_i[K,K] \end{bmatrix}, \quad \mathbf{D}_i = \begin{bmatrix} \mathbf{D}_i[1] \\ \vdots \\ \mathbf{D}_i[K] \end{bmatrix}. \tag{4.24}$$

It is important to emphasize here that:

$$\mathbf{J}_i[k,l] = (\mathbf{H}_i^k)^H \mathbf{R}_i \mathbf{W}_i \mathbf{R}_i^H \mathbf{H}_i^l, \quad \mathbf{D}_i[k] = (\mathbf{H}_i^k)^H \mathbf{R}_i \mathbf{W}_i. \tag{4.25}$$

According to the definitions and partitions in (4.24) and (4.16), the gradient of the Lagrangian function of problem ($\mathrm{P}_{4,3}^i$) in (4.20) ($\mathcal{L}$) with respect to $\mathbf{T}_{k,i}$ for $k \in \mathcal{C}_i$ is:

$$\nabla_{\mathbf{T}_{k,i}} \mathcal{L} = \mathbf{J}_i\,[k,k]\,\mathbf{T}_{k,i} + \sum_{l \in \mathcal{C}_i, l \neq k} \mathbf{J}_i\,[k,l]\,\mathbf{T}_{l,i} - \mathbf{D}_i\,[k] + \alpha\gamma_{k,i}\mathbf{T}_{k,i}$$
$$+ \boldsymbol{\Upsilon}_i\,[k,k]\,\mathbf{T}_{k,i} + \sum_{l \in \mathcal{C}_i, l \neq k} \boldsymbol{\Upsilon}_i\,[k,l]\,\mathbf{T}_{l,i} + \lambda_k \mathbf{T}_{k,i}, \tag{4.26}$$

where λ_k is a non-negative dual variable associated to the k-th per-BS power constraint in (4.20). Then, by equating (4.26) to $\mathbf{0}$ we obtain the precoding structure $\mathbf{T}_{k,i}$ for $k \in \mathcal{C}_i$ (being $\mathbf{0}$ otherwise):

$$\mathbf{T}_{k,i} = \begin{cases} \left(\mathbf{J}_i[k,k] + \boldsymbol{\Upsilon}_i[k,k] + (\lambda_k + \alpha\gamma_{k,i})\mathbf{I} \right)^{-1} \mathbf{F}_{k,i}, & \text{if } k \in \mathcal{C}_i, \\ \mathbf{T}_{k,i} = \mathbf{0}, & \text{if } k \notin \mathcal{C}_i. \end{cases} \tag{4.27}$$

where

$$\mathbf{F}_{k,i} = \mathbf{D}_i[k] - \sum_{l \in \mathcal{C}_i, l \neq k} \mathbf{J}_i\,[k,l]\,\mathbf{T}_{l,i} - \sum_{l \in \mathcal{C}_i, l \neq k} \boldsymbol{\Upsilon}_i\,[k,l]\,\mathbf{T}_{l,i}. \tag{4.28}$$

The precoder in (4.27) is coupled with the precoders of BSs in the i-th cluster (i.e. $\{\mathbf{T}_{l,i}\}_{\forall l \neq k, l \in \mathcal{C}_i}$). So the solution for joint BSs clustering and precoding is achieved at the master of the cluster of the i-th user by applying a BCD method among block variables $\{\mathbf{T}_{k,i}\}_{\forall k \in \mathcal{C}_i}$ until a stop condition (e.g. convergence or maximum number of iterations achieved) for given $\mathbf{R}_i$, $\mathbf{W}_i$, $\boldsymbol{\Upsilon}_i[k,l]_{\forall k,l \in \mathcal{C}_i}$ and $\{P_{k,-i}\}_{\forall k \in \mathcal{C}_i}$. Algorithm 4.1 summarizes the BCD method among block variables $\{\mathbf{T}_{k,i}\}_{\forall k \in \mathcal{C}_i}$. Section 4.5.3 details how to acquire the required parameters at each master.

As it can be observed in (4.27)-(4.28) and (4.25), for given $\boldsymbol{\Upsilon}_i[k,l], \forall k,l \in \mathcal{C}_i$, the BSs clustering and precoding design can be performed in a decentralized manner at each master by having knowledge only of the channel matrices towards the candidate BSs selected by the user (i.e. $\mathbf{H}_{k,i}, \forall k \in \mathcal{C}_i$).

Algorithm 4.1 Clustering and precoding design at the master of the i-th cluster for given $\{\boldsymbol{\Upsilon}_i[k,l]\}_{\forall k,l \in \mathcal{C}_i}$, $\{P_{k,-i}\}_{\forall k \in \mathcal{C}_i}$, $\mathbf{R}_i$, and $\mathbf{W}_i$

1: Compute $\mathbf{J}_i[k,l]$ and $\mathbf{D}_i[k]$ using (4.25) $\forall k,l \in \mathcal{C}_i$
2: **initialize**: $\mathbf{T}_{k,i}, \forall k \in \mathcal{C}_i$
3: **repeat**(cyclically pick a candidate BS $k \in \mathcal{C}_i$)
4:     Compute $\mathbf{T}_{k,i}$ using (4.27) with $\mathbf{J}_i[k,l], \forall l \in \mathcal{C}_i$, $\mathbf{D}_i[k]$, $\mathbf{R}_i$, $\mathbf{W}_i$, and $\boldsymbol{\Upsilon}_i[k,l], \forall l \in \mathcal{C}_i$ (compute λ_k with the bisection method [92] such that $\mathrm{Tr}(\mathbf{T}_{k,i}\mathbf{T}_{k,i}^H) + P_{k,-i} \leq P_{\max}^k$)
5: **until** stop condition

4.5.2   Optimization at each user

Given all the precoding matrices, $\{\mathbf{T}_i\}_{\forall i}$, problem $(\text{P}_{4,2})$ in (4.14) can be easily decomposed into I parallel optimization problems (one per user), where the optimal receive filter $(\mathbf{R}_i)$ and the optimal weighting matrix $(\mathbf{W}_i)$ for the i-th user are obtained as the solution to:

$$(\text{P}_{4,4}^i) : \underset{\mathbf{R}_i, \mathbf{W}_i}{\text{minimize}} \ \ \text{Tr}\left(\mathbf{W}_i \mathbf{E}_i\right) - \mu_i \log_2 \left| \tfrac{\ln(2)}{\mu_i} \mathbf{W}_i \right|. \tag{4.29}$$

The optimal receive filter $\mathbf{R}_i$ to $(\text{P}_{4,4}^i)$ in (4.29) is given by the MMSE receive filter $\mathbf{R}_i^{\text{mmse}}$ in (4.10). As it is done in real deployments [141], each user can compute $\mathbf{R}_i^{\text{mmse}}$ in (4.10) based on the estimation of the equivalent channel $\mathbf{H}_i \mathbf{T}_i$ and $\mathbf{C}_{\mathbf{y}_i}$. It is not needed to estimate the interfering channels to get $\mathbf{C}_{\mathbf{y}_i}$ in (4.8), as it can be evaluated by averaging $\mathbf{y}_i \mathbf{y}_i^H$ [141].

Once the receive filter $\mathbf{R}_i$ is designed, the optimal weighting matrix $\mathbf{W}_i$ to $(\text{P}_{4,4}^i)$ in (4.29) is given by [127]:

$$\mathbf{W}_i = \tfrac{\mu_i}{\ln(2)} (\mathbf{E}_i^{\text{mmse}})^{-1}. \tag{4.30}$$

So each user can compute $\mathbf{W}_i$ in (4.30) based on $\mathbf{H}_i \mathbf{T}_i$ and $\mathbf{C}_{\mathbf{y}_i}$, i.e. the same information needed to compute $\mathbf{R}_i^{\text{mmse}}$ in (4.10).

4.5.3   Acquisition of parameters at the master of each cluster

Acquisition of the interference-cost matrix $\boldsymbol{\Upsilon}_i$

Similarly as in Chapter 3 (see Section 3.5.3), we can get an estimate of the interference-cost matrix for the UE-centric clustering scheme by using an UL pilot-based transmission. More specifically, the blocks needed for decentralized optimization from $\boldsymbol{\Upsilon}_i$ in (4.17) can be estimated from the covariance matrix of the interference-plus-noise received signal in UL when: i) channel reciprocity is assumed (as in TDD systems) and ii) users in UL transmit with a specific pilot signal that is precoded as a function of the receive filter $\mathbf{R}_i$ and the weighting matrix $\mathbf{W}_i$. This way, we avoid the complex task associated to the estimation of the most harmful interfering channel matrices that would be needed to compute $\boldsymbol{\Upsilon}_i$ by following (4.15) and we avoid also the reporting of all receive filters and weighting matrices to non-serving BSs (i.e. $\mathbf{R}_j$ and $\mathbf{W}_j$, $\forall j \neq i$, see (4.15)). As pointed out before, the blocks of $\boldsymbol{\Upsilon}_i$ in (4.16) that are needed for decentralized optimization at the master of the i-th cluster are those such that $\boldsymbol{\Upsilon}_i[k,l], \forall k, l \in \mathcal{C}_i$ (see (4.22)). The procedure to obtain them is described in the following.

From an UL pilot-based transmission and thanks to channel reciprocity in TDD, the covariance matrix of the received interference-plus-noise signal at the BSs ($\forall k \in \mathcal{C}_i$) of the i-th cluster in UL $\overleftarrow{\mathbf{N}}_i \in \mathbb{C}^{M_{U_i} \times M_{U_i}}$ (being $M_{U_i} = \sum_{k \in \mathcal{C}_i} M_k$) is:

$$\overleftarrow{\mathbf{N}}_i = \sum_{j \in \mathcal{I}, j \neq i} \bar{\mathbf{H}}_{i,j} \overleftarrow{\mathbf{T}}_j \overleftarrow{\mathbf{T}}_j^H (\bar{\mathbf{H}}_{i,j})^H + \overleftarrow{\sigma}_i^2 \mathbf{I}, \tag{4.31}$$

where $\overleftarrow{\mathbf{T}}_j \in \mathbb{C}^{N_j \times m_j}$ denotes the UL precoder used at the j-th user, $\overleftarrow{\sigma}_i^2$ the UL noise power, and $\bar{\mathbf{H}}_{i,j}$ contains the channel matrices from the j-th user towards the BSs in the i-th cluster stacked as: $\bar{\mathbf{H}}_{i,j} = [\mathbf{H}_{k,j}, \ldots, \mathbf{H}_{l,j}]^T$, $\forall k, l \in \mathcal{C}_i$. Let us partition $\overleftarrow{\mathbf{N}}_i$ as we did for the interference-cost matrix $\boldsymbol{\Upsilon}_i$ in (4.16), but the partition of $\overleftarrow{\mathbf{N}}_i$ only contains the blocks of those candidate BSs

in the cluster of the i-th user. Let us maintain the index of the BSs for referencing the blocks, such that $(\forall k, l \in \mathcal{C}_i)$:

$$\overleftarrow{\mathbf{N}}_i[k,l] = \sum_{j \in \mathcal{I}, j \neq i} \mathbf{H}_{k,j}^T \overleftarrow{\mathbf{T}}_j \overleftarrow{\mathbf{T}}_j^H \mathbf{H}_{l,j}^* + \overleftarrow{\sigma}_i^2 \mathbf{I}. \tag{4.32}$$

If the UL precoder is designed according to:

$$\overleftarrow{\mathbf{T}}_j = \sqrt{F} \mathbf{R}_j^* (\mathbf{W}_j^{\frac{1}{2}})^*, \tag{4.33}$$

being $F<1$ a scaling cell-wide factor that allows meeting the UL power constraint (see further details in [**C3**]) and $\mathbf{W}_j = \mathbf{W}_j^{\frac{1}{2}} (\mathbf{W}_j^{\frac{1}{2}})^H$, then the blocks of $\mathbf{\Upsilon}_i$ in (4.16) that are needed for decentralized optimization at the master of the i-th cluster (i.e. $\mathbf{\Upsilon}_i[k,l], \forall k, l \in \mathcal{C}_i$) can be estimated as:

$$\hat{\mathbf{\Upsilon}}_i[k,l] = F^{-1}(\overleftarrow{\mathbf{N}}_i[k,l])^*. \tag{4.34}$$

As it is shown in [**C3**], the estimation errors in $\hat{\mathbf{\Upsilon}}_i$ are negligible when properly selecting the scaling factor F in (4.33). Note that it is not needed to estimate the interfering channels to get $\overleftarrow{\mathbf{N}}_i$ in (4.31) because it can be estimated by averaging the UL received signal if high-speed backhaul links connect the BSs within each cluster. Further, it is not needed to decode the UL transmitted symbols, so an UL pilot-based transmission is enough to get $\hat{\mathbf{\Upsilon}}_i$. In LTE-A we could use the already defined sounding reference signals [62] with UL power control.

Acquisition of the power spent by candidate BSs in the cluster towards other clusters $\{P_{k,-i}\}_{\forall k \in \mathcal{C}_i}$

We assume that the master of each i-th cluster collects from the candidate BSs that form the i-th cluster the power that they use towards other clusters, i.e. $P_{k,-i}, \forall k \in \mathcal{C}_i$. So, only exchange of control information with neighboring BSs is required.

Acquisition of $\mathbf{R}_i$ and $\mathbf{W}_i$

We assume that $\mathbf{R}_i$ and $\mathbf{W}_i$ are reported from the i-th user towards the master of its i-th cluster through an UL feedback link. As compared to [60] where feedback links are required from each user towards all BSs in the network to report $\mathbf{R}_i$ and $\mathbf{W}_i$, in our case only one feedback link is needed per-user towards its BS master.

4.5.4 Design of the weighting power coefficients

The introduction of the weighting power coefficients $\gamma_{k,i}$ in $(P_{4,1})$ in (4.13) aims at the reduction of the cluster size by taking out of the cluster those candidate BSs that use too low power. So $\gamma_{k,i}$ should decrease with the power spent by the k-th BS towards the i-th user. A suitable selection is:

$$\gamma_{k,i} = \frac{1}{\tau + \text{Tr}(\bar{\mathbf{T}}_{k,i} \bar{\mathbf{T}}_{k,i}^H)}, \tag{4.35}$$

where $\tau>0$ is a small constant factor to avoid $\gamma_{k,i} \to \infty$ at any point in the iterations and $\text{Tr}(\bar{\mathbf{T}}_{k,i} \bar{\mathbf{T}}_{k,i}^H)$ is the power spent by the k-th BS to the i-th user in the previous iteration. Note that those candidate BSs using low power towards a specific user would have associated a high

value for $\gamma_{k,i}$. Therefore, at the next iteration, the power at that BS would tend to 0 due to the problem formulation, see (P$_{4,1}$) in (4.13), and thus the cluster size of the i-th user would be reduced.

4.5.5　Algorithm

Algorithm 4.2 summarizes the iterative procedure to solve (P$_{4,2}$) in (4.14) in a decentralized manner. First of all, each i-th user selects the candidate BSs to form its cluster (i.e. $\mathcal{C}_i$) and among them the BS being the master of its cluster (line 1), which acquires the channel matrices from candidate BSs in the cluster towards the user (i.e. $\{\mathbf{H}_{k,i}\}_{\forall k \in \mathcal{C}_i}$) (line 2). Then, the algorithm starts from an initialization of the precoders $\{\mathbf{T}_{k,i}\}_{\forall k \in \mathcal{C}_i}$, $\forall i$, that satisfies the per-BS power constraints in (4.20) (line 3). For simulation purposes, a suitable initialization is obtained by solving (P$_{4,3}^i$) in (4.20) using $\mathbf{\Upsilon}_i = \mathbf{0}$. Then, a DL transmission is carried out using $\{\mathbf{T}_{k,i}\}$ (line 3) where users can evaluate the covariance matrix of the received signal $\mathbf{C}_{\mathbf{y}_i}$ and update $\mathbf{R}_i$ and $\mathbf{W}_i$ as shown in Section 4.5.2 (line 4).

After the initialization, the iterative procedure is implemented assuming a TDD synchronized system and alternate UL/DL transmissions. First, the required parameters for decentralized optimization are acquired at each master, as detailed in Section 4.5.3 (i.e. $\{\mathbf{\Upsilon}_i[k,l]\}_{\forall k,l \in \mathcal{C}_i}$ from the UL, $\{P_{-i}^k\}_{\forall k \in \mathcal{C}_i}$ through inter-BS signaling, and $\mathbf{R}_i$ and $\mathbf{W}_i$ through feedback) (lines 7-9). Also, the weighting power coefficients $\{\gamma_{k,i}\}_{\forall k \in \mathcal{C}_i}$ are updated (line 10). Next, the simultaneous BSs clustering and precoding designs are performed at the master of each i-th cluster for DL transmission, as shown in Section 4.5.1 (line 12). Let us recall that after the per-cluster optimization the precoders have to be scaled so as to strictly satisfy the per-BS power constraint as, due to the uncoupling of the per-BS transmit power, in a given iteration the constraint could be violated. Finally, DL transmission is performed (line 13) such that the optimization at users can be done as shown in Section 4.5.2 (line 15).

Algorithm 4.2 DJCP to solve (P$_{4,2}$) in (4.14)

1: All users ($\forall i$): select set of candidate BSs, $\mathcal{C}_i$, one of them being the master of the i-th cluster
2: All masters ($\forall i$): acquire knowledge of the channel matrix from candidate BSs towards the i-th user, $\{\mathbf{H}_{k,i}\}_{\forall k \in \mathcal{C}_i}$
3: All masters ($\forall i$): initialize $\{\mathbf{T}_{k,i}\}_{\forall k \in \mathcal{C}_i}$ and transmit in DL
4: All users ($\forall i$): compute $\mathbf{R}_i$ in (4.10) and $\mathbf{W}_i$ in (4.30)
5: **repeat**
6:　　 # *Acquisition of parameters*
7:　　 All users ($\forall i$): transmit in UL a pilot signal properly precoded as in (4.33) using $\mathbf{R}_i$ and $\mathbf{W}_i$, such that each master acquires $\mathbf{\Upsilon}_i[k,l], \forall k,l \in \mathcal{C}_i$ as in (4.34)
8:　　 All masters ($\forall i$): acquire power spent by BSs in the cluster towards other clusters, $P_{k,-i}, \forall k \in \mathcal{C}_i$
9:　　 All users ($\forall i$): report $\mathbf{R}_i$ and $\mathbf{W}_i$ to the master of the i-th cluster
10:　　 All masters ($\forall i$): update weights $\{\gamma_{k,i}\}_{\forall k \in \mathcal{C}_i}$ as in (4.35)
11:　　 # *Simultaneous optimizations at each master*
12:　　 All masters ($\forall i$): do **Algorithm 4.1** for fixed $\{\mathbf{\Upsilon}_i[k,l]\}_{\forall k,l \in \mathcal{C}_i}$, $\{P_{k,-i}\}_{\forall k \in \mathcal{C}_i}$, $\mathbf{R}_i$, and $\mathbf{W}_i$
13:　　 All BSs ($\forall k$): transmit in DL with $\{\mathbf{T}_{k,i}\}_{\forall i}$
14:　　 # *Simultaneous optimizations at each user*
15:　　 All users ($\forall i$): compute $\mathbf{R}_i$ in (4.10) and $\mathbf{W}_i$ in (4.30)
16: **until** stop condition

Monotonic convergence of Algorithm 4.2 can be proven when all parameters are perfectly acquired, $\gamma_{k,i}$ is fixed, $\forall i, \forall k$, and the optimizations at each cluster are performed in a sequential manner. This can be demonstrated thanks to the special convex properties of problem $(P_{4,2})$ in (4.14) and by following a similar rationale as in Chapter 3 (see Section 3.5.5) (omitted here due to space limit). If all clusters perform the optimization simultaneously, monotonic convergence cannot be guaranteed due to the coupling of the per-BS power constraints among clusters. However, even with simultaneous optimizations, convergence is consistently observed in simulations. Convergence is also observed when including the rule to update $\gamma_{k,i}$ in (4.35).

4.6 Simulation Results

The network scenario consists of a dense synchronized TDD deployment of $|\mathcal{K}|$ outdoor SeNBs (that act as BSs) in a concentrated area, following specifications for Scenario 2a in [104]. BSs are randomly placed within a circle of 50 m radius with a minimum distance of 20 m among them, and $|\mathcal{I}|$ users are randomly placed in a concentric 70 m radius circle. All BSs operate on the same carrier frequency at 3.5 GHz with 10 MHz bandwidth. ITU Urban Micro model with 3D distance is used for path loss and shadowing modeling. The antenna pattern is omnidirectional and the transmit power is 24 dBm. Noise spectral density is -174 dBm/Hz. The number of antennas is $M_k=2, \forall k$, and $N_i=2, \forall i$ (such that $m_i=2, \forall i$). Full-load traffic model is adopted.

Two different criteria are used:

- sum-rate (SR) with $\mu_i=1, \forall i$, in (4.13) and
- proportional fair (PF) where $\mathbf{W}_i=\frac{1}{\ln(2)R_i}(\mathbf{E}_i^{\mathrm{mmse}})^{-1}$ is used in (4.30) (see [60]).

Results are averaged over 100 random deployments.

4.6.1 Performance versus the number of candidate BSs

First, we consider a network of $|\mathcal{K}|=8$ BSs and $|\mathcal{I}|=8$ users and study the performance of the proposed DJCP when varying the number of candidate BSs selected by the users (i.e. $|\mathcal{C}_i|$) for $\alpha=0$ in (4.13). Note that the larger is the value of $|\mathcal{C}_i|$, the more channel matrices have to be estimated. $K=8$ BSs is used so as to compare DJCP with the decentralized 'BS-disjoint clustering' scheme in [C3] using disjoint clusters of either 2 BSs or 4 BSs, in which inter-cluster interference management is adopted and only channel matrices towards the BSs in the cluster are needed (as in the proposed DJCP).

Fig. 4.3 shows the convergence of DJCP in a given random deployment for different values of $|\mathcal{C}_i|$ (1, 2, 3, 4, 8) and SR criterion. Convergence is consistently achieved.

Fig. 4.4.(a) and Fig. 4.4.(b) display the sum-rate performance and the 5%-tile of the per-user rates, respectively, versus the number of candidate BSs for SR and PF criteria. As benchmark we use the 'WMMSE' algorithm [127] for the broadcast channel, where the 8 BSs form a giant virtual transmitter that serves all UEs, including per-BS power constraints and knowledge of all channel matrices. By comparing the decentralized schemes with the same number of estimated channel matrices (2 BSs or 4 BSs, marked with gray arrows in figures), it can be observed that DJCP outperforms in terms of sum-rate and 5%-tile rate the decentralized BS-disjoint clustering for both SR and PF criteria. The gains are: 8% (PF-4BSs), 16% (PF-2BSs), 6% (SR-4BSs), 10% (SR-2BSs) in sum-rate and 84% (PF-4BSs), 49% (PF-2BSs) in 5%-tile rate. The gain is specially
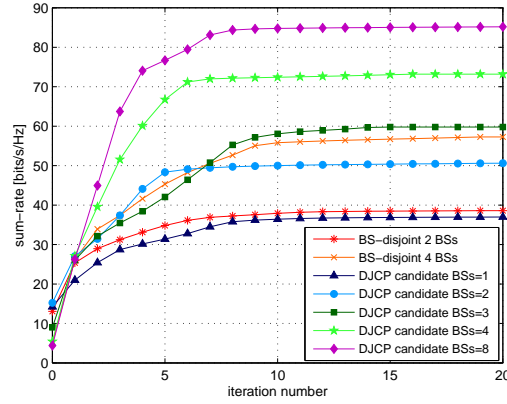
Figure 4.3: Convergence of DJCP in terms of sum-rate for different number of candidate BSs selected per user ($|\mathcal{C}_i|$=1, 2, 3, or 4). $|\mathcal{K}|$=8, $|\mathcal{I}|$=8. α=0. SR as utility function.

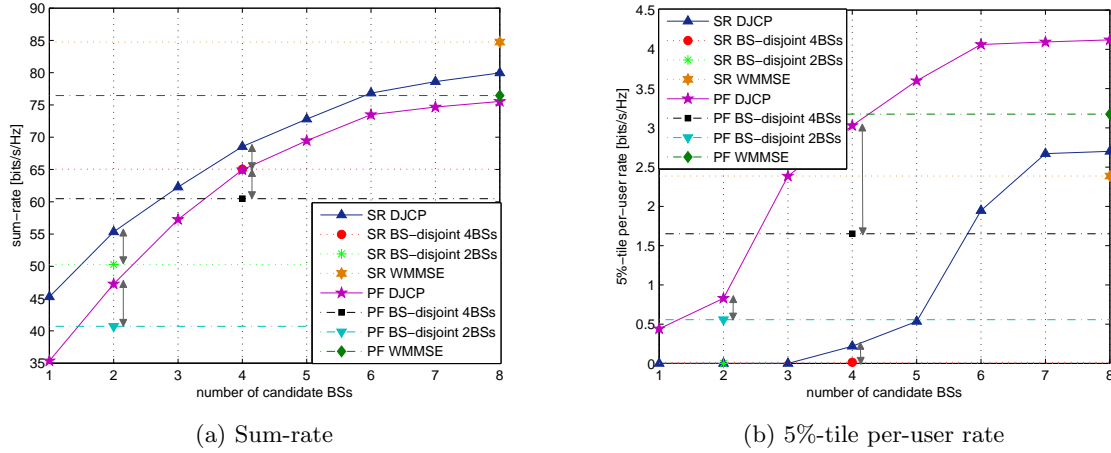(a) Sum-rate   (b) 5%-tile per-user rate

Figure 4.4: Sum-rate (in bits/s/Hz) and 5%-tile per-user rate (in bits/s/Hz) vs. the number of candidate BSs selected per user ($|\mathcal{C}_i|$). $|\mathcal{K}|$=8, $|\mathcal{I}|$=8. α=0. SR and PF as utility function.

remarkable in the 5%-tile rate, where 'PF DJCP' with 3 candidate BSs already outperforms 'PF BS-disjoint 4BSs' and, in the case of 4 BSs in the cluster, 'SR DJCP' gets a 5%-tile rate of 0.218 bits/s/Hz while BS-disjoint clustering gets 0.013 bits/s/Hz (see Fig. 4.4.(b)). This is due to the high flexibility that DJCP offers for dense and irregular deployments of BSs.

4.6.2   Performance versus α

Second, we consider a network of $|\mathcal{K}|$=10 BSs (as specified in [104]) and $|\mathcal{I}|$=10 users and study the performance of DJCP for different values of α in (4.13) when the number of candidate BSs selected by the users is $|\mathcal{C}_i|$=4. When α>0 we simulate two cases: $\gamma_{k,i}$=1, $\forall i, k$, and $\gamma_{k,i}$ updated as in (4.35) using τ=0.1. We use an α in (4.13) that is normalized by the maximum power at BSs (P).
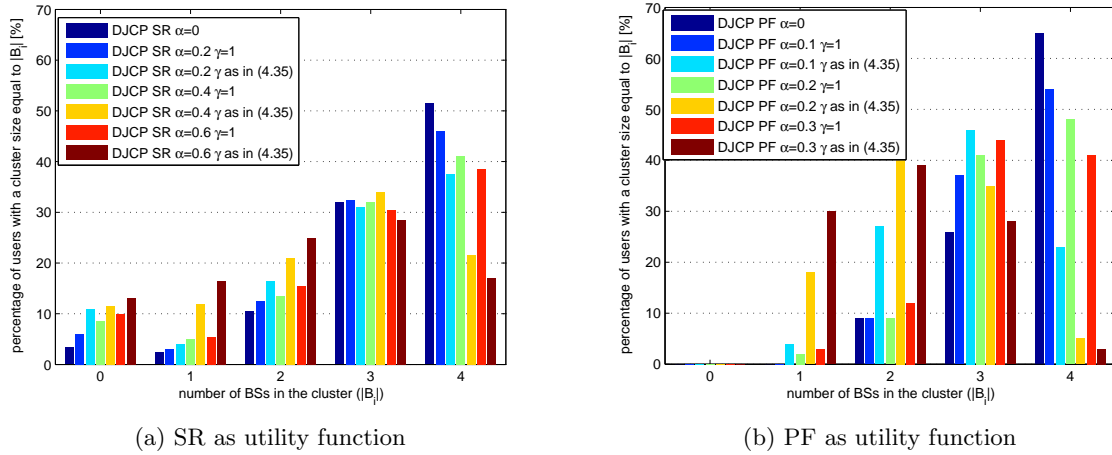
(a) SR as utility function

(b) PF as utility function

Figure 4.5: Percentage of users (%) having a cluster size ($|\mathcal{B}_i|$) of 0, 1, 2, 3, or 4 BSs, for different DJCP optimization cases and SR or PF criterion. $|\mathcal{K}|$=10, $|\mathcal{I}|$=10. $|\mathcal{C}_i|$=4.
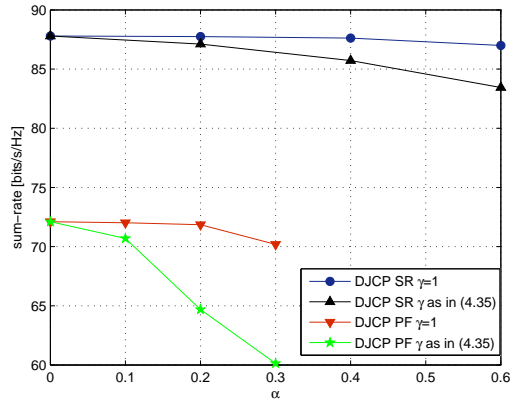

Figure 4.6: Sum-rate vs. α. $|\mathcal{K}|$=10, $|\mathcal{I}|$=10. $|\mathcal{C}_i|$=4. SR and PF as utility function.

Fig. 4.5 illustrates the percentage of users having a cluster size of 0, 1, 2, 3, or 4 BSs for SR (Fig. 4.5.(a)) and PF (Fig. 4.5.(b)) criteria. Results are shown when using different values of α for ($P_{4,1}$) in (4.13): α={0, 0.2, 0.4, 0.6} for SR and α={0, 0.1, 0.2, 0.3} for PF. Different values of α are considered for SR and PF criteria as the utility function is not of the same magnitude. It can be observed that increasing α reduces the cluster size. The proposed rule to update the parameter $\gamma_{k,i}$ in (4.35) per user and per BS allows further reduction of the cluster size as compared to $\gamma_{k,i}$=1 both for SR and PF criteria. In the case of SR, some users are assigned to a cluster size of 0 BSs, which means they are not scheduled. This effect does not appear with PF criterion (see Fig. 4.5.(b)).

Fig. 4.6 shows the sum-rate versus α when using $\gamma_{k,i}$=1 or $\gamma_{k,i}$ updated as in (4.35) and either SR or PF criterion. In the case of SR, the performance degradation is very mild, while the number of BSs per cluster is significantly reduced with the proposed design of $\gamma_{k,i}$ in (4.35) (see Fig. 4.5.(a)). In the case of PF, the sum-rate is very sensitive to α when $\gamma_{k,i}$ is designed as in (4.35). However, it is worth comparing the case α=0.3 and $\gamma_{k,i}$=1 with respect to α=0.1 and $\gamma_{k,i}$ as in (4.35): both lead to similar sum-rate while the number of BSs per cluster is significantly reduced in the later case (see Fig. 4.5.(b)).

4.7　Complementarities of CoMP-JT and multi-antenna Systems

In this section we aim at determining the possible complementarities of CoMP-JT and the use of multiple antennas at BSs, as well as to determine which approach can provide larger gains in theory depending on the considered scenario and channel conditions. **CoMP-JT** allows that different BSs act as a single transmitter with geographically separated antennas. This provides diversity and allows converting interfering signals into useful signals, but requires a very tight synchronization among the BSs cooperating for CoMP-JT. On the other hand, the use of multiple antennas at BSs is recently receiving a lot of attention (specially when the number of antennas is very large, i.e. **large-scale MIMO**, also known as massive-MIMO [9]) because it requires simple linear signal processing approaches and it allows combating the small-scale fading. The adoption of multiple antennas at BSs provides additional degrees of freedom which can be used to simultaneously serve multiple users or to null interference towards unintended users.

In order to perform the analysis, we use the DJCP technique presented in Section 4.5. The DJCP algorithm decides about the BSs clustering for CoMP-JT and the adequate precoding in a user-centric manner, i.e. the cluster and precoding are decided per user. Due to the per-UE centric clustering scheme, clusters for CoMP-JT might be overlapped (i.e. they may share some BSs). In this sense, the main goal of DJCP technique is to deal with all the coupling issues arising in the system: the inter-cluster interference and the coupling power constraint per BS. With such approach we can play with the maximum number of candidate BSs in which CoMP-JT can take place. This parameter will be an input for the simulations and will be fixed to be equal for all users (i.e. the maximum CoMP-JT cluster size for all users will be the same).

In the present analysis we focus on the evaluation of DJCP when varying either:

- the number of candidate BSs for CoMP-JT or
- the number of available transmit antennas at BSs.

4.7.1　Simulation results in simplified scenarios

We consider a deployment of $|\mathcal{K}|$ BSs and $|\mathcal{I}|=|\mathcal{K}|$ users. Each BSs is equipped with M antennas, and each user has N receive antenna elements. The noise is assumed to be white proper Gaussian with distribution $\sim\mathcal{CN}(\mathbf{0},\sigma^2\mathbf{I})$. It is assumed that all BSs have the same available power P. Signal-to-noise ratio is defined as $\text{SNR}=P/\sigma^2$.

To perform the analysis we either vary the number of antennas at BSs M or allow the use of CoMP-JT among BSs by increasing the number of candidate BSs per user. All the remaining parameters are kept constant: i.e. same transmit power P, same SNR, same number of BSs and users K, and same number of receive antennas N. We evaluate two different homogeneous scenarios (note that the same maximum power at all BSs is used) depending on the propagation conditions:

- **Symmetric scenario**: pathloss from all BSs to all users is of the same magnitude. Propagation channels are modeled by a Rayleigh distribution: $\mathbf{H}_{k,i}\sim\mathcal{CN}(\mathbf{0},\mathbf{I}),\forall k,i$.
- **Non-symmetric scenario**: each user has a stronger (or predominant) BS. Propagation channels are modeled through a Rayleigh distribution as: $\mathbf{H}_{k,k}\sim\mathcal{CN}(\mathbf{0},\mathbf{I}),\forall k$ (predominant BS), and $\mathbf{H}_{k,i}\sim\mathcal{CN}(\mathbf{0},\eta_{k,i}\mathbf{I}),\forall i\neq k$ (for all remaining BSs), where factor $\eta_{k,i}$ is modeled as an uniformly distributed random variable that takes values between 0 and 1 (hence it introduces the common asymmetry present in wireless communications).

Due to the use of multiple antennas at BSs, we add **spatial transmit correlation** in the channel modeling. So the equivalent channel is modeled as:

$$\mathbf{H}_{k,i}^{\text{eq}} = \mathbf{H}_{k,i}\mathbf{R}_T^{\frac{1}{2}}, \forall k, \forall i \tag{4.36}$$

In order to model the correlation among transmit antenna elements we use a common, simple and scalable model usually adopted in the literature [138], where each component of the transmit correlation matrix $\mathbf{R}_T$ is given by:

$$\mathbf{R}_T[m,n] = \rho^{|m-n|}, \forall m, \forall n \tag{4.37}$$

where $\mathbf{R}_T = \mathbf{R}_T^{\frac{T}{2}}\mathbf{R}_T^{\frac{1}{2}}$, $[m,n]$ refers to the m-th row n-th column element, and $0 \leq \rho < 1$ is the correlation factor. This model is totally characterized by the value of the correlation factor ρ: the larger values of ρ are related to higher correlation between the transmit antenna elements. For simulation purposes we use different cases:

- $\rho = 0$ (no correlation),
- $\rho = 0.9$ (medium correlation),
- $\rho = 0.99$ (high correlation).

In the following, the performance of DJCP is evaluated through Montecarlo simulations by allowing CoMP-JT among a multiple BSs (i.e. by varying the number of candidate BSs per user) and by using multiple antennas at BSs (i.e. by varying M). PF criterion is used. A deployment of $|\mathcal{K}| = 8$ BSs and $|\mathcal{I}| = 8$ UEs is considered. SNR=10dB is used. 100 different channel realizations are averaged. Antenna configurations are depicted in figures as $M \times N$.

Rate versus number of candidate BSs and number of transmit antennas

The following figures display the average user rate (in bits/s/Hz) versus the number of candidate BSs for CoMP-JT for different antenna configurations: 2×2, 4×2, 8×2, 16×2, and 32×2. The symmetric scenario and the non-symmetric scenario are evaluated with the different antenna correlation cases: no correlation, $\rho = 0.9$, and $\rho = 0.99$.

Fig. 4.7 displays the results for the symmetric scenario and Fig. 4.8 shows the results for the non-symmetric scenario. From these figures several conclusions can be extracted.

First, both CoMP-JT with a larger cluster size and the use of a larger number of transmit antennas provide significant rate gains. Further, its complementary can be observed in all the considered scenarios (symmetric and non-symmetric) and in all channel conditions (correlated and non-correlated). So we can conclude that the techniques are complementary. However, the gains are not so large in the following cases:

- The gains of increasing the cluster size for CoMP-JT are lower in the non-symmetric scenario when a large number of transmit antennas is used (**observation 1**), see Fig. 4.8.(a), where the rate tends to be saturated. This is due to the fact that if the number of antennas is high then the major interference in the network can be suppressed and all the users can already receive two streams, such that the gains of CoMP-JT are not much significant (for example, for the 16×2 configuration case, gains are not significant by increasing the cluster size more than 3 BSs).

(a) no correlation
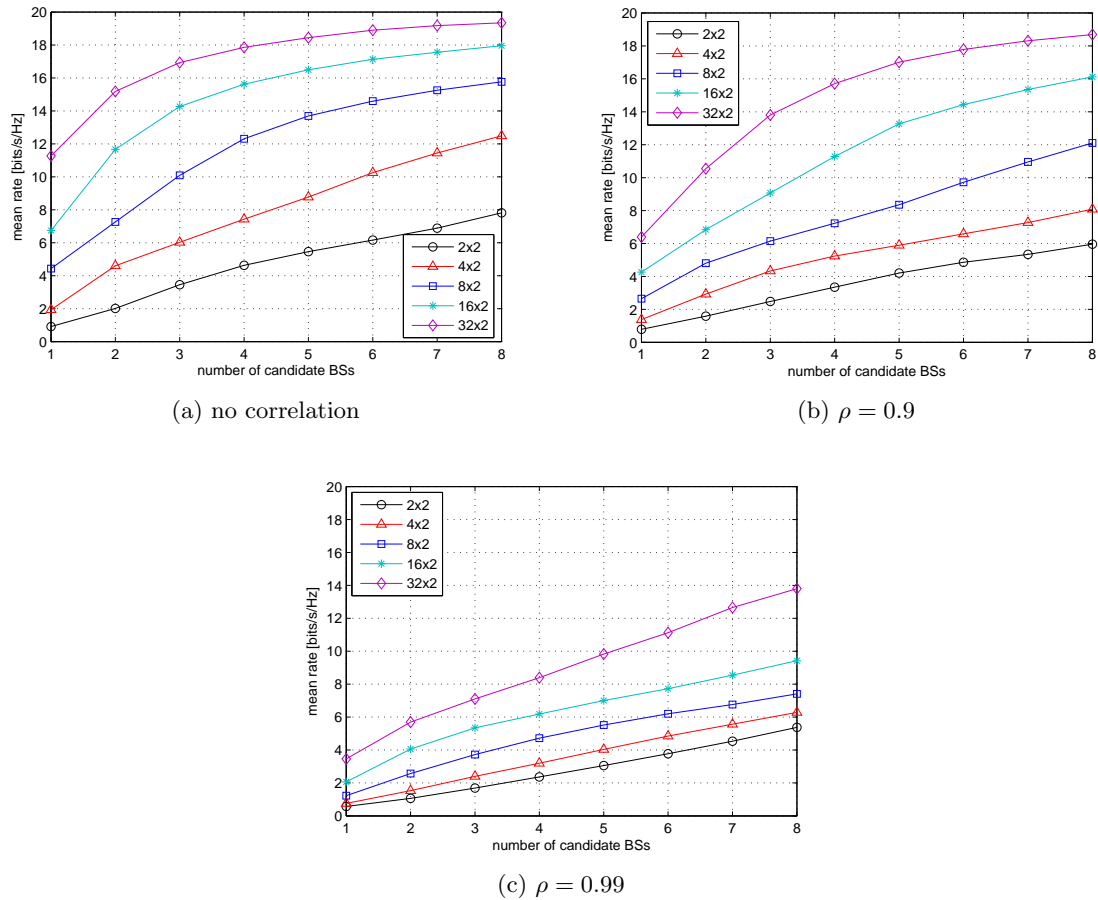
(b) $\rho = 0.9$

(c) $\rho = 0.99$

Figure 4.7: Mean rate (bits/s/Hz) vs. the number of candidate BSs for CoMP-JT for different antenna configurations. Symmetric scenario.

- The gains of increasing the number of transmit antennas are lower in both symmetric and non-symmetric scenarios for high transmit correlation see Fig. 4.7.(c) and Fig. 4.8.(c). This is due to the high correlation of the transmit antenna elements. In these cases, it is better to employ CoMP-JT as non-correlated antennas can be used to serve a user.

Second, as the power per BS and the number of users in the system are kept constant in the simulations, we can focus on comparing the use of CoMP-JT and the use of a larger number of antennas at BSs for the configurations in which the number of transmit antennas available to serve a given user is the same. For example, we can compare the case of 2×2 with cooperation of 4 BSs with the cases of: 4×2 with cooperation of 2 BSs, and 8×2 with no cooperation. Similarly we can compare the case of 2×2 with cooperation of 8 BSs with the cases of: 4×2 with cooperation of 4 BSs, 8×2 with cooperation of 2 BSs, and 16×2 with no cooperation. We can also compare the case of 4×2 with cooperation of 8 BSs with the cases of: 8×2 with cooperation of 4 BSs, 16×2 with cooperation of 2 BSs, and 32×2 with no cooperation. And so on. Under this kind of comparison, depending on the scenario and the channel conditions we can decide which technique (i.e. CoMP-JT or multiple transmit antennas at BSs) provides larger gains, as detailed in the following.

(a) no correlation

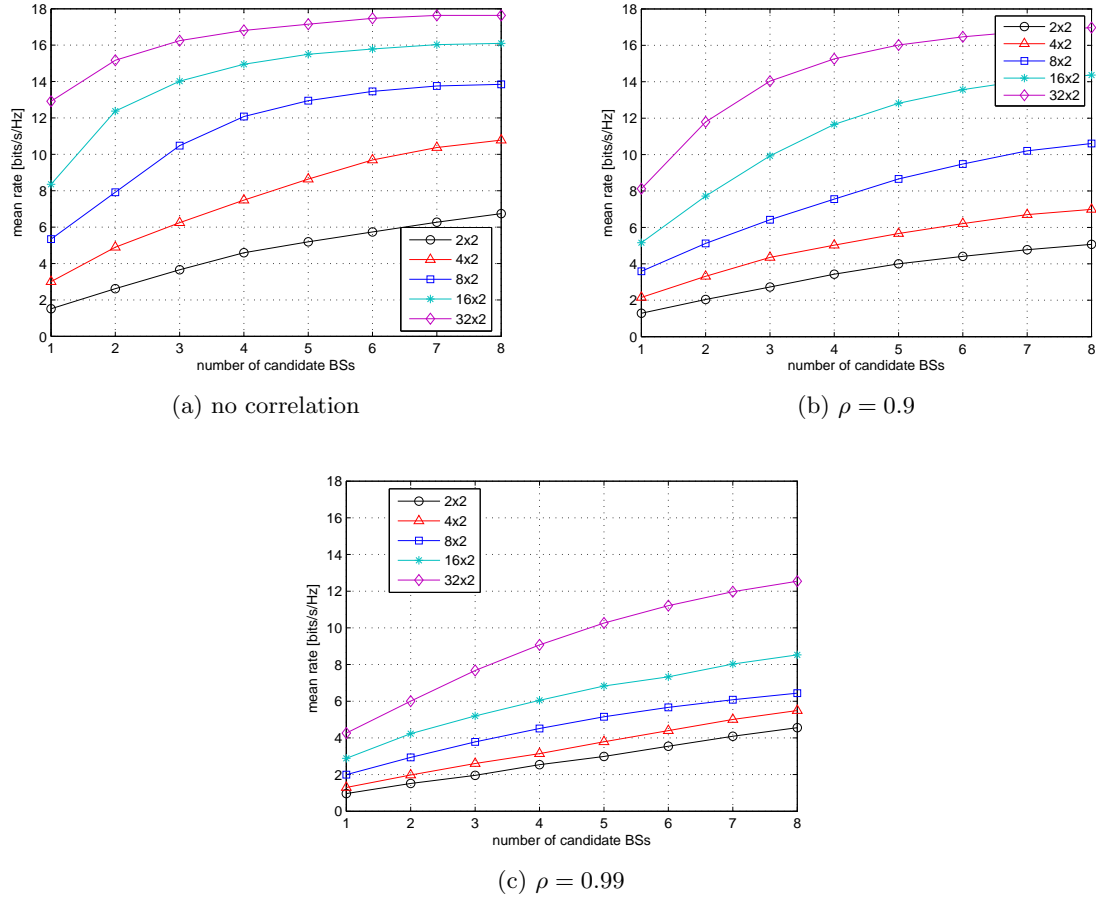(b) $\rho = 0.9$

(c) $\rho = 0.99$

Figure 4.8: Mean rate (bits/s/Hz) vs. the number of candidate BSs for CoMP-JT for different antenna configurations. Non-symmetric scenario.

- Symmetric scenario:
  - if no correlation or medium correlation (see Fig. 4.7.(a) and Fig. 4.7.(b)): similar gains are obtained by increasing the number of antennas at the BSs than by allowing the use of CoMP-JT. This occurs because the channels are symmetric and it is approximately the same in terms of performance if the antennas are distributed or not.
  - if high correlation (see Fig. 4.7.(c)): rate gains are obtained with CoMP-JT as compared to adding more transmit antennas. This is due to the fact that CoMP-JT allows having non-correlated antennas per user.
- Non-symmetric scenario:
  - if no correlation (see Fig. 4.8.(a)): gains are obtained by adding antennas at the BSs. This is owing to the fact that the scenario is non-symmetric and CoMP-JT is implemented from BSs with different path-loss values, so if there is no correlation among the antenna elements it is better to have them all at the strongest BS.
  - if medium correlation (see Fig. 4.8.(b)): similar gains are obtained by increasing the number of antennas at the BSs or by allowing the use of CoMP-JT.

– if high correlation (see Fig. 4.8.(c)): rate gain are obtained with CoMP-JT as compared to adding transmit antennas. Similarly as in the symmetric scenario, this happens because CoMP-JT allows having non-correlated antennas per user. So, even if the scenario is non-symmetric when the correlation is very high then CoMP-JT provides the required diversity to increase the rate.

Let us recall that the results for the homogeneous non-symmetric scenario are valid for a smallcell deployment where smallcells are deployed in clusters and CoMP-JT can be applied among them.

Accordingly, the larger gains can be obtained by increasing the cluster size for CoMP-JT or by using additional transmit antennas at BSs depending on the scenario and channel conditions. However, it is not worth recalling that the complementarities of using CoMP-JT and adding multiple antennas per BS are significant in all cases.

Justification of observation 1

In order to illustrate **observation 1** with numerical results, we display in the following figures the number of streams per user as the number of antennas increases or the number of candidate BSs for CoMP-JT increases when using the non-symmetric scenario with high transmit correlation among the antenna elements (corresponding to the rate performance shown in Fig. 4.8.(a)).

Fig. 4.9 shows the percentage of users having 1 stream and the percentage of users having 2 streams as the number of BSs for CoMP-JT increases (x-axis) and as the number of antennas increases (shown in the different figures). As it is stated before, these results confirm that in the 16×2 antenna configuration case (see Fig. 4.9.(d)), most part of the users are already served with 2 streams when no cooperation is used. Further, in the same configuration (16×2), when the number of candidate BSs for CoMP-JT is equal to 2 then all the users are served with 2 streams, which indicates that the maximum rate might be obtained and interference is properly suppressed. So in this scenario and channel conditions (non-symmetric scenario with no transmit correlation), the gains of CoMP-JT are appreciable with a small number of cooperative BSs. Such number is lower as the number of transmit antennas increase. However even in the 16×2 antenna configuration case, cooperation of 2 BSs is still useful in terms of rate (see Fig. 4.8.(a)).

Finally, from Fig. 4.9 we can conclude that the DJCP technique works as it was expected and that the use of more antennas at BSs or the use of CoMP-JT have similar outcomes in terms of number of streams: both techniques allow increasing the number of streams per user as the number of antennas increases or as CoMP-JT is performed among a larger number of BSs.

4.7.2   Simulation results in 3GPP cluster-based scenarios

In this section we evaluate the performance of DJCP through Montecarlo simulations by allowing CoMP-JT among a multiple BSs (i.e. by varying the number of candidate BSs per user) and by using multiple antennas at BSs in a 3GPP-based scenario. The network scenario consists of a dense synchronized TDD deployment of $|\mathcal{K}|$ SeNBs (that act as BSs) in a concentrated area, following specifications for Scenario 2a in [104]. BSs are randomly placed within a circle of 50 m radius with a minimum distance of 20 m among them, and $|\mathcal{I}|=|\mathcal{K}|$ users are randomly placed in a concentric 70 m radius circle. All BSs operate on the same carrier frequency at 3.5 GHz with 10 MHz bandwidth. ITU Urban Micro model with 3D distance is used for path loss and

(a) 2×2

(b) 4×2
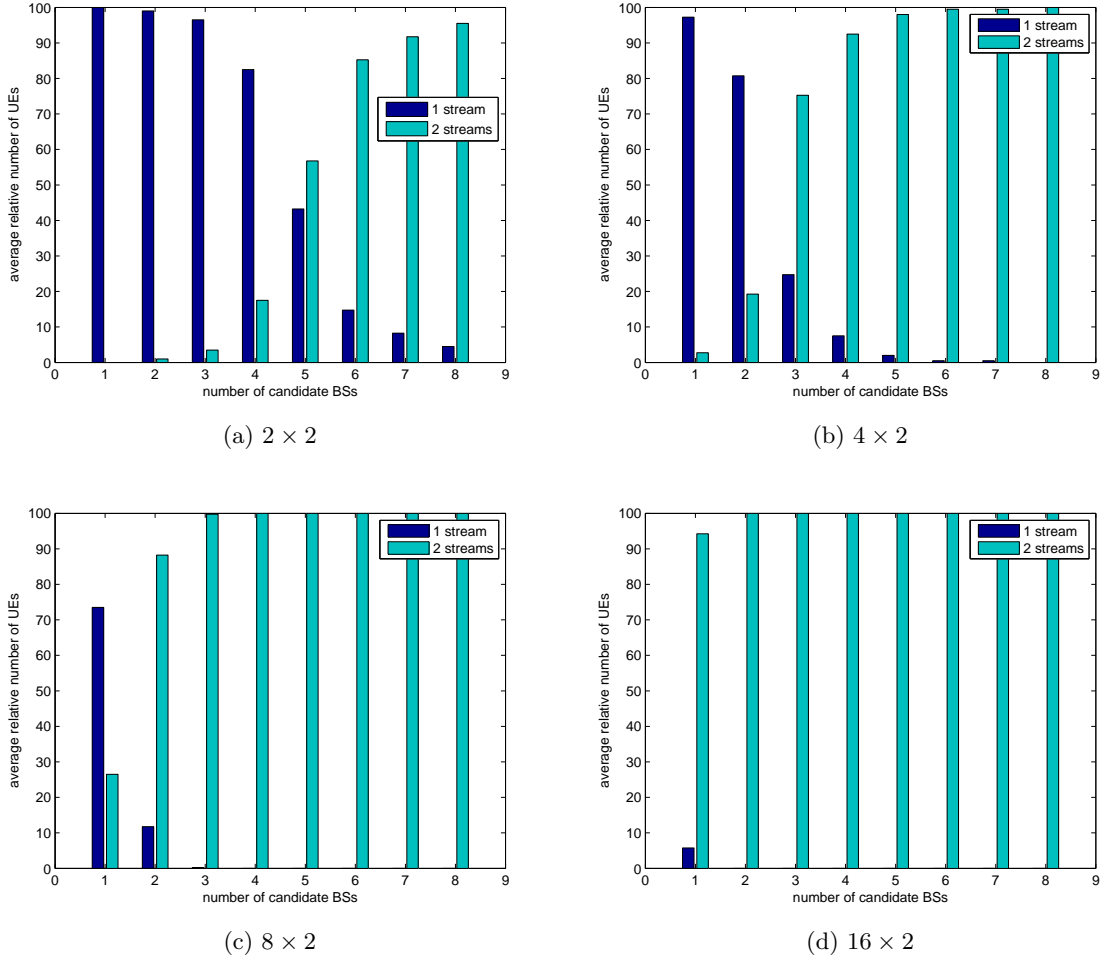

(c) 8×2

(d) 16×2

Figure 4.9: Percentage of users having 1 stream and 2 streams. Non-symmetric scenario. No correlation case.

shadowing modeling. The antenna pattern is omnidirectional and the transmit power is 24 dBm. Noise spectral density is -174 dBm/Hz. Transmit antenna correlation is used at BSs according to the model detailed in Section 4.7.1. Full-load traffic model is adopted. PF criterion is used. A deployment of $|\mathcal{K}|$=8 BSs and $|\mathcal{I}|$=8 UEs is considered. 100 different channel realizations are averaged. Antenna configurations are depicted in figures as $M \times N$.

Rate versus number of candidate BSs and number of transmit antennas

Fig. 4.10 shows the average user rate (in bits/s/Hz) versus the number of candidate BSs for CoMP-JT for different antenna configurations: 2×2, 4×2, 8×2, 16×2, and 32×2. Two antenna correlation cases detailed in Section 4.7.1 are evaluated (i.e. no correlation and ρ=0.9).

In this network deployment we can conclude that CoMP-JT and adding more transmit antennas are complementary techniques. The conclusions are similar to the ones for the homogeneous non-symmetric scenario in Section 4.7.1 as the cluster-based scenario is non-symmetric. The mean rate is larger because the mean SNR is larger than 10dB for the strongest BSs (in general with up to 4 BSs). However, the mean SNR with the remaining BSs is much lower than 10dB,

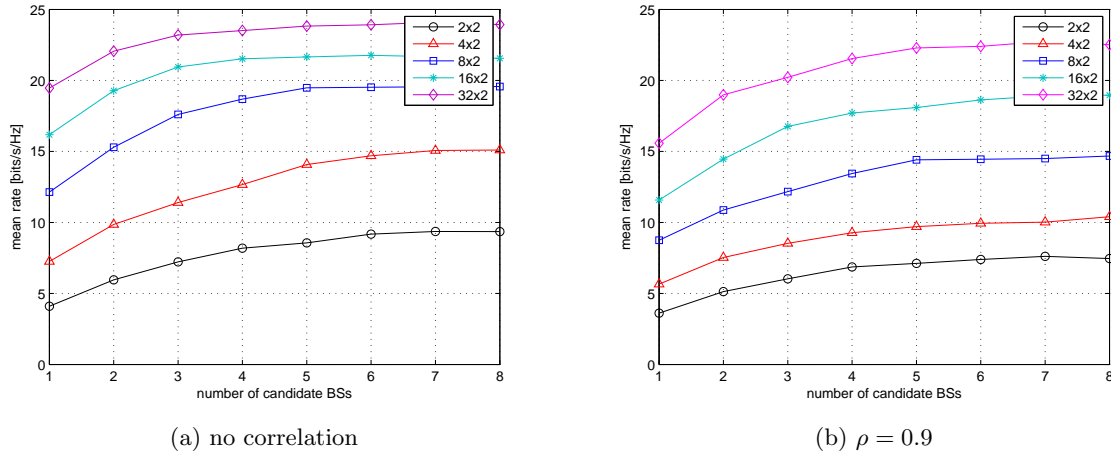(a) no correlation   (b) $\rho = 0.9$

Figure 4.10: Mean rate (bits/s/Hz) vs. the number of candidate BSs for CoMP-JT for different antenna configurations. 3GPP cluster-based scenario.

and in most of the cases lower than 0dB (**observation 2**). Note that 10dB was used in Section 4.7.1 as maximum, and only in the 10% of the cases a SNR lower than 0dB was observed by the user. Therefore, the interference in the network is lower than in the non-symmetric scenario considered in Section 4.7.1. Also, due to that, the gains of BSs transmit cooperation get saturated earlier when increasing the number of BSs for CoMP-JT and/or the number of transmit antennas at BSs.

We can conclude that **4 cooperative BSs is enough**, because increasing the cluster size larger than 4 BSs leads to a small rate improvement. On the contrary, increasing the number of transmit antennas at BSs always leads to significant improvements.

If we perform the comparison in Section 4.7.1, i.e. compare the use of CoMP-JT and the use of a larger number of antennas at BSs for the configurations in which the number of transmit antennas available to serve a user is the same (note that the power per BS and the number of users in the system are not modified in the simulations), we can conclude that in this cluster-based scenario and for all antenna correlation cases (see Fig. 4.10) larger rate gains are obtained by adding antennas at the BSs than by performing CoMP-JT. This is owing to the fact that the scenario is non-symmetric and CoMP-JT is implemented from BSs with different path-loss values, so it is better to have the antennas at the strongest BS even if the correlation is high.

Justification of observation 2

In order to illustrate **observation 2** with numerical results, Fig. 4.11 displays the cumulative distribution function (CDF) of the SNR (in dB) with the different candidate BSs as observed by each user in a specific random deployment. It can be shown that a SNR in mean larger than 10dB is obtained with up to 4 candidate BSs. However, the 5th candidate BSs is observed in general with a mean SNR of 5dB. Finally, the 6th, 7th, and 8th candidate BSs are observed with a SNR lower than 0dB. For that reason, in the cluster-based scenario, appreciable rate gains are obtained when including up to 4 BSs in the cluster for CoMP-JT.
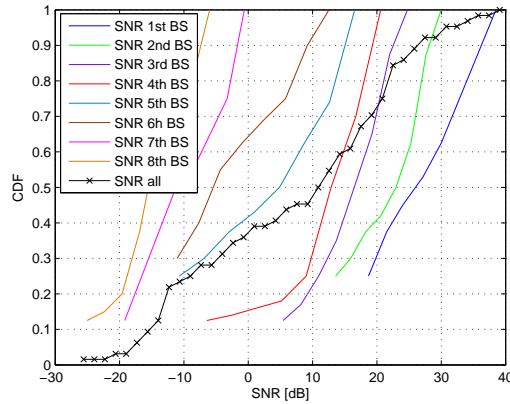
Figure 4.11: CDF of the SNR (in dB) as observed with the different candidate BSs.

4.8 Conclusions

This chapter proposes a decentralized joint BSs clustering and precoding for cooperative transmission (CoMP-JT) in DL of multi-cell MIMO TDD systems. A UE-centric clustering scheme is used, being the cluster size limited to a maximum number of candidate BSs selected by the user. The problem is distributed among clusters by uncoupling the per-BS power constraints and using the UL received signal as a way of reporting the impact of interference. This way, only channel matrices towards candidate BSs need to be reported. Further, an additional term is included in the problem so as to reduce the cluster size by taking out of the cluster those candidate BSs that use too low power. Results in dense deployments of BSs show significant gains as compared to distributed BS-disjoint clustering schemes and an effective reduction of the cluster size with the proposed design.

Finally, the possible complementarities of CoMP-JT through a UE-centric clustering and large-scale MIMO systems are analyzed. The complementarities are investigated in different simplified interference scenarios (symmetric and non-symmetric) and under different antenna correlation cases. Then, they are evaluated in dense SCN with realistic deployments and channel conditions. Simulation results in 3GPP-compliant dense SCN scenarios show that 4 BSs are enough for transmit cooperation in most of the correlation cases, otherwise increasing the cluster size provides not enough performance benefits. It is also observed that potential gains are obtained from the complementarity of using CoMP-JT and adding multiple antennas at BSs in all cases.

Chapter 5

Transmit Coordination for Coexisting Signaling Schemes

Motivated by a transitional scenario where legacy linear transmitters coexist with widely linear ones, this chapter investigates the general K-user MIMO interference channel (IC) in a heterogeneous (linear and widely linear) transmitter deployment. Thus, this chapter generalizes Chapter 3 with an heterogeneous deployment that arises due to the use of different signaling schemes: proper Gaussian signaling (PGS) and improper Gaussian signaling (IGS), which have been previously introduced in Chapter 2. In particular, we address the maximization of the weighted sum-rate (WSR) for (widely) linear transmit filters design. To do so, we use of the complex-valued formulation because it allows facing the problem. Since the maximum WSR problem is non-convex, and thus difficult to be solved, we formulate an equivalent minimum weighted mean square error (WMSE) problem that allows deriving closed-form expressions for (widely) linear transceivers. Then an iterative procedure is proposed, which is proven to reach a stationary point of the maximum WSR problem. Simulations show that the proposed procedure allows increasing the sum-rate as compared to coordinated linear transceiver schemes. The gains are larger and significant in two different non-exclusive conditions: as the interference level increases or when the number of antennas is low.

The technical papers related to this topic are:

[J3]   S. Lagen, A. Agustin, J. Vidal, "Coexisting Linear and Widely Linear Transceivers in the MIMO Interference Channel", *IEEE Trans. on Signal Processing*, vol. 64, no. 3, pp. 652 - 664, Jan. 2016. doi: 10.1109/TSP.2015.2489604.

[C5]   S. Lagen, A. Agustin, J. Vidal, "Decentralized Interference Management with Improper Gaussian Signaling for MIMO-IC", *IEEE Global Commun. Conf.*, Workshop on Emerging Technologies for 5G Wireless Cellular Networks, Austin (USA), Dec. 2014. doi: 10.1109/GLOCOMW.2014.7063526.

This chapter contains results presented in **[J3]**, which deal with the coexistence of WLP and LP transmitters for WSR maximization. Differently, **[C5]** addresses the coexistence of WLE and LE receivers in terms of MSE minimization. Note that the coexistence of heterogeneous receivers does not have an impact in terms of WSR, although it does in terms of MSE.

This chapter is structured as follows. Section 5.1 contains the state of the art for IGS and WLP operations. The main contributions of the chapter are detailed in Section 5.2. In Section 5.3 the system model for the MIMO IC with heterogeneous transmitters is presented

and the signal model using the complex-valued channel model is detailed. In Section 5.4, the maximum WSR problem for mixed transceiver design is formulated. An equivalent minimum WMSE problem is proposed in Section 5.5. Then, in Section 5.6, closed-form expressions for mixed transceiver design are derived, a centralized algorithm based on alternate optimization is presented and decentralized implementations are detailed. Section 5.7 shows the simulation results. Finally, concluding remarks are included in Section 5.8.

Preliminaries for improper Gaussian random vectors and widely linear processing operations were presented in Appendix 2.A.

5.1   State of the Art

The K-user MIMO IC is a generic model for cellular communication systems that consists of K transmitter-receiver pairs, each equipped with multiple antennas. All transmitters wish to send independent streams to its intended receiver simultaneously, such that interference is generated towards unintended receivers. Unfortunately, the optimal transmit/receive strategy with linear filters that maximizes the WSR of the system is not known. From an optimization theory perspective, the problem is non-convex and NP-hard even in the single-antenna case [132]. Even so, there are two main approaches to find a stationary point to the maximum WSR problem.

On the one hand, strategies in [133, 134] (and references therein) focus on the coordination among transmitters based on the interference-cost concept, where each transmitter maximizes its own utility function minus the interference-cost, hence managing interference in a decentralized manner. On the other hand, convergence to a stationary point can be obtained by iteratively minimizing the weighted sum of WMSE, see [127], where transmit filters, receive filters, and weighting matrices are alternatively optimized, being the weighting matrices chosen according to the inherent relation between the achievable rate and the MSE [136]. Decentralized implementations of the *WMSE approach* are analyzed in [127] and Chapter 3 (**[J2]**). In [127] it is shown that the WMSE approach and the strategies based on the interference-cost ([133, 134]) yield almost the same WSR performance, but with less complexity in the WMSE case. In Chapter 3 (see **[J2]**), the WMSE approach and the interference-cost concept have been combined and, by exploiting the channel reciprocity property available in TDD systems, almost the same WSR results are obtained but with less stringent requirements for channel estimation and information reporting than [127, 133, 134].

In all these works and related ones it is assumed that transmitted signals are proper (or circularly symmetric complex) Gaussian distributed [128]. PGS has been shown to be optimal in terms of capacity for the MIMO P2P channel [105], and also for the MIMO BC with DPC as the capacity achieving strategy [48, 49]. However, recent results have shown that optimality of PGS does not necessarily hold for the MIMO BC if the transmit strategy is restricted to widely linear transceivers [108, 109]. It is shown there that capacity gains can be obtained by employing **improper (or circularly asymmetric complex) Gaussian signaling** because, in the absence of non-linear interference cancellation, the MIMO BC becomes interference-limited [110]. The same happens in the MIMO IC, where IGS through the use of widely linear precoding is able to improve the known achievable rates when interference is treated as Gaussian noise [68, 69, 113] and is also shown to obtain larger degrees of freedom (DoF), i.e. slope of the sum-rate as a function of the SNR at the high SNR regime, [52, 106, 111].

An effective way to generate and estimate IGS is by using WLP [69] and WLE [65], respectively, instead of the commonly used LP and LE techniques that maintain the proper Gaussian distribution of the signals. WLE is a generalized concept used in communication channels that encounter improper noise [66] and in systems that transmit improper signal constellations [67]. The later case is already under investigation in 3GPP LTE-A [1], see [29]. Recently, strategies for WLP design in the MIMO BC are analyzed in [109], however, little has been done regarding the coordinated design of WLP at transmitters for the MIMO IC.

There are two main approaches in the recent literature on IGS and widely linear operations for the MIMO IC [129]:

- the composite real formulation, whereby real and imaginary parts of the MIMO channel matrix are separated, and
- the complex-valued formulation.

Many works in the literature, e.g. [68, 106, 111], tackle the SISO IC with IGS through the composite real formulation by reformulating the problem into an equivalent double-sized real-valued MIMO IC. This way, most of the approaches already developed for the MIMO IC with PGS can be applied. Further, this formulation has been used to state why IGS is better at handling interference (and under which conditions) in the SISO Z-IC (a special IC setting) [122] and in the MIMO Z-IC (see Chapter 2, [**J1**]). Nevertheless, in [69] it is shown that some new insights on the use of IGS can be gained thanks to the use of the **complex-valued formulation**. They show that for any given set of covariance matrices the achievable rates for the MIMO IC can be improved with IGS by optimizing the transmit pseudo-covariance matrices. In this regard, transmit covariance and pseudo-covariance optimization techniques are proposed for the two-user SISO IC in [69] and for the K-user MISO IC in [113], but they are only valid for the single-antenna receiver case.

5.2 Contribution

In this chapter we address the maximization of the WSR in the general K-user MIMO IC and we focus on covering a backwards compatibility-oriented scenario where different types of linear and widely linear transmitters coexist. This heterogeneous scenario is of relevance in legacy cellular networks (as 3GPP LTE-A [1]), where user terminals of different releases coexist and different transmission modes (TM) can be adopted at transmitters [145]. In the downlink transmission, LTE-A considers different TM based on non-codebook-based precoding at the base stations, which have been introduced in different releases [145]: TM7 in release 8, TM8 in release 9, TM9 in release 10, and TM10 in release 11. Therefore, the heterogeneous transmitter scenario can appear in the DL due to the use of TMs of different releases. In the UL transmission, where transmitters are the UEs, the heterogeneous transmitter scenario naturally arises due to the coexistence of UEs from different releases. To cover said situations, we assume that some transmitters are constrained to employ conventional LP while other transmitters can apply WLP. The scenario with mixed transmitters is shown in Fig. 5.1. As an example, transmitter 2 (TX 2) is restricted to use LP and the remaining transmitters can adopt WLP if required.

If all transmitters used WLP then the maximum WSR problem could be addressed with the composite real formulation and conventional schemes developed for the MIMO IC could be applied. However, the backwards compatibility-oriented scenario with mixed transmitters
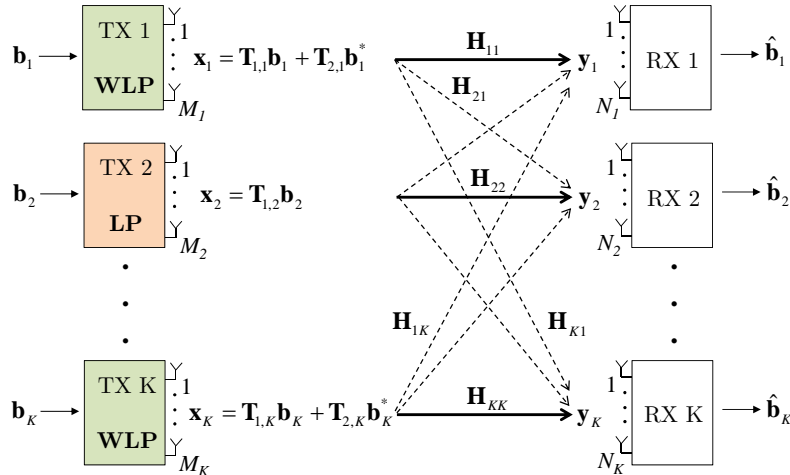
Figure 5.1: K-user MIMO IC for a backwards compatibility-oriented scenario with mixed transmitters. Solid lines represent desired signals and dashed lines denote interfering signals. In this particular scenario, TX 2 is restricted to employ LP while the other transmitters can adopt WLP if required.

cannot be tackled in a straightforward manner with the composite real formulation: working with double-sized matrices and imposing particular structures over transmit filters (which are needed for LP) entails a cumbersome formulation. On the contrary, the complex-valued formulation allows facing the problem for transceiver design with mixed (linear and widely linear) transmitters in a unified and more candid way.

In **[C5]** we use the complex-valued formulation to address the minimization of the sum of MSEs in the K-user MIMO IC with heterogeneous receivers. In this chapter, as in **[J3]**, we use the complex-valued formulation to face the maximization of the WSR in the K-user MIMO IC with an heterogeneous deployment of transmitters and assuming optimal receivers.

To summarize, the contributions of this chapter are:

- The relation between the mutual information and the MSE matrices is derived when the transmitted signal and the interference-plus-noise signal are improper Gaussian distributed.

- The maximum WSR (maxWSR) problem for (widely) linear transceiver design in the most general K-user MIMO IC is formulated in such a way to cover a backwards compatibility-oriented scenario.

- An equivalent minimum WMSE (minWMSE) problem is proposed to solve the maxWSR problem thanks to the use of the complex-valued channel model, and closed-form expressions for mixed transceiver design are derived.

- An iterative algorithm with alternate optimization between (widely) linear transmit filters (WLP-LP), widely linear receive filters (WLE), and weighting matrices is presented. Convergence of the algorithm to a stationary point of the maxWSR problem is demonstrated. Decentralized implementations are described.

We show through simulations that the sum-rate is enhanced with the proposed transceiver design (that includes transmit coordination and IGS) as compared to interference coordination procedures with PGS. In this regard, simulation results allow identifying the scenarios where the gains are largest.

5.3 System Model

Consider a MIMO IC composed of a set of $\mathcal{K} \triangleq \{1, \ldots, K\}$ transmitter-receiver pairs, where each k-th transmitter is equipped with M_k antennas and each k-th receiver has N_k receive antenna elements ($k=1, \ldots, K$). On a given time/frequency resource, each k-th transmitter serves its associated k-th receiver and, therefore, interference is generated among non-associated terminals, as shown in Fig. 5.1. Assume an heterogeneous scenario of linear and widely linear transmitters (for example, in Fig. 5.1 TX2 is restricted to use LP while TX1 and TXK can adopt WLP). Define the set of WLP transmitters as $\mathcal{S}^{\text{WLP}}$ and the set of LP transmitters as $\mathcal{S}^{\text{LP}}$, such that $\mathcal{S}^{\text{WLP}} \cap \mathcal{S}^{\text{LP}} = \emptyset$ and $\mathcal{S}^{\text{WLP}} \cup \mathcal{S}^{\text{LP}} = \mathcal{K}$.

Without loss of generality, we assume that each k-th transmitter intends to send a proper Gaussian information-bearing signal with unitary power towards its intended k-th receiver, i.e. $\mathbf{b}_k \in \mathbb{C}^{m_k \times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, where $m_k = \min(N_k, M_k)$ denotes the maximum number of streams towards the k-th receiver. Unlike conventional transmit linear processing where the transmitted signal is assumed to be proper Gaussian distributed (i.e. $\mathbf{x}_k \in \mathbb{C}^{M_k \times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\mathbf{x}_k})$, being $\mathbf{C}_{\mathbf{x}_k}$ the covariance matrix of the transmitted signal), in this chapter the general IGS is used. Hence, the second order statistics of the transmitted signal $\mathbf{x}_k$ are not only given by the covariance matrix $\mathbf{C}_{\mathbf{x}_k}$ but also by the pseudo-covariance matrix $\tilde{\mathbf{C}}_{\mathbf{x}_k}$ [65].

Any improper Gaussian signal $\mathbf{x}_k$ can be generated from a proper Gaussian information-bearing signal $\mathbf{b}_k$ by using WLP (see [69, Sect. II.C]). We assume that legacy transmitters (i.e. $k \in \mathcal{S}^{\text{LP}}$) will use the conventional LP scheme. Therefore, the transmitted signal at the k-th transmitter is given by:

$$\mathbf{x}_k = \begin{cases} \mathbf{T}_{1,k} \mathbf{b}_k + \mathbf{T}_{2,k} \mathbf{b}_k^* & \text{if } k \in \mathcal{S}^{\text{WLP}} \\ \mathbf{T}_{1,k} \mathbf{b}_k & \text{if } k \in \mathcal{S}^{\text{LP}} \end{cases}, \tag{5.1}$$

where matrices $\mathbf{T}_{1,k}$ and $\mathbf{T}_{2,k} \in \mathbb{C}^{M_k \times m_k}$ are the transmit linear precoders for the information-bearing signal $\mathbf{b}_k$ and its complex conjugate $\mathbf{b}_k^*$, respectively. This way, the covariance matrix $\mathbf{C}_{\mathbf{x}_k}$ and the pseudo-covariance matrix $\tilde{\mathbf{C}}_{\mathbf{x}_k}$ of the transmitted signal $\mathbf{x}_k$ in (5.1) result [65]:

$$\mathbf{C}_{\mathbf{x}_k} = \mathbb{E}[\mathbf{x}_k \mathbf{x}_k^H] = \mathbf{T}_{1,k} \mathbf{T}_{1,k}^H + \mathbf{T}_{2,k} \mathbf{T}_{2,k}^H, \tag{5.2}$$

$$\tilde{\mathbf{C}}_{\mathbf{x}_k} = \mathbb{E}[\mathbf{x}_k \mathbf{x}_k^T] = \mathbf{T}_{1,k} \mathbf{T}_{2,k}^T + \mathbf{T}_{2,k} \mathbf{T}_{1,k}^T. \tag{5.3}$$

The conventional LP scheme is a special case of WLP in which $\mathbf{T}_{2,k} = \mathbf{0}$ in (5.1), such that $\mathbf{x}_k$ is a proper Gaussian signal (i.e. $\tilde{\mathbf{C}}_{\mathbf{x}_k} = \mathbf{0}$ in (5.3)). The power radiated by the k-th transmitter is given by $\text{Tr}(\mathbf{C}_{\mathbf{x}_k}) = \text{Tr}(\mathbf{T}_{1,k} \mathbf{T}_{1,k}^H + \mathbf{T}_{2,k} \mathbf{T}_{2,k}^H)$.

Assuming narrow-band transmissions, the equivalent baseband signal observed at the k-th receiver is expressed as:

$$\mathbf{y}_k = \mathbf{H}_{k,k} \mathbf{x}_k + \mathbf{n}_k = \mathbf{H}_{k,k} \mathbf{x}_k + \sum_{j \in \mathcal{K}, j \neq k} \mathbf{H}_{j,k} \mathbf{x}_j + \mathbf{v}_k, \tag{5.4}$$

where $\mathbf{H}_{j,k} \in \mathbb{C}^{N_k \times M_j}$ denotes the MIMO channel matrix between the j-th transmitter and the k-th receiver (containing the complex-valued channel gains of the different antenna-pairs) and $\mathbf{n}_k \in \mathbb{C}^{N_k \times 1}$ refers to the received interference-plus-noise at the k-th receiver. $\mathbf{n}_k$ is composed by an interference component plus circularly symmetric complex (i.e. proper) Gaussian

noise with distribution $\mathbf{v}_k \sim \mathcal{CN}(\mathbf{0}, \sigma_k \mathbf{I})$. Hence, under the independence assumption of $\mathbf{v}_k$ and $\{\mathbf{b}_k\} = \{\mathbf{b}_1, \ldots, \mathbf{b}_K\}$, the covariance matrix $\mathbf{C}_{\mathbf{y}_k}$ and the pseudo-covariance matrix $\tilde{\mathbf{C}}_{\mathbf{y}_k}$ of the received signal $\mathbf{y}_k$ in (5.4) are:

$$\mathbf{C}_{\mathbf{y}_k} = \mathbb{E}[\mathbf{y}_k \mathbf{y}_k^H] = \mathbf{H}_{k,k}\mathbf{C}_{\mathbf{x}_k}\mathbf{H}_{k,k}^H + \mathbf{C}_{\mathbf{n}_k}, \quad \mathbf{C}_{\mathbf{n}_k} = \sum_{j \in \mathcal{K}, j \neq k} \mathbf{H}_{j,k}\mathbf{C}_{\mathbf{x}_j}\mathbf{H}_{j,k}^H + \sigma_k \mathbf{I}, \quad (5.5)$$

$$\tilde{\mathbf{C}}_{\mathbf{y}_k} = \mathbb{E}[\mathbf{y}_k \mathbf{y}_k^T] = \mathbf{H}_{k,k}\tilde{\mathbf{C}}_{\mathbf{x}_k}\mathbf{H}_{k,k}^T + \tilde{\mathbf{C}}_{\mathbf{n}_k}, \quad \tilde{\mathbf{C}}_{\mathbf{n}_k} = \sum_{j \in \mathcal{K}, j \neq k} \mathbf{H}_{j,k}\tilde{\mathbf{C}}_{\mathbf{x}_j}\mathbf{H}_{j,k}^T, \quad (5.6)$$

where $\mathbf{C}_{\mathbf{n}_k}$ and $\tilde{\mathbf{C}}_{\mathbf{n}_k}$ denote the covariance matrix and the pseudo-covariance matrix of the received interference-plus-noise signal $\mathbf{n}_k$ in (5.4), respectively.

In order to access the information contained in the received signal $\mathbf{y}_k$ in (5.4), WLE should be applied at the receiver [65]. Therefore, the information-bearing signal $\mathbf{b}_k$ is estimated at the receiver side according to:

$$\hat{\mathbf{b}}_k = \mathbf{R}_{1,k}^H \mathbf{y}_k + \mathbf{R}_{2,k}^H \mathbf{y}_k^*, \quad (5.7)$$

where $\mathbf{R}_{1,k}$ and $\mathbf{R}_{2,k} \in \mathbb{C}^{N_k \times m_k}$ are the linear receive filters.

The MSE for the symbols transmitted towards the k-th receiver can be expressed through the so-called MSE-matrix $\mathbf{E}_k = \mathbb{E}[(\mathbf{b}_k - \hat{\mathbf{b}}_k)(\mathbf{b}_k - \hat{\mathbf{b}}_k)^H]$. As the different information-bearing signals $\{\mathbf{b}_k\}$ correspond to independent and proper Gaussian random vectors (and hence uncorrelated), the *MSE-matrix* can be developed in terms of the transmit and receive filters in (5.1) and (5.7) as:

$$\mathbf{E}_k = \mathbf{I} - \mathbf{R}_{1,k}^H \mathbf{H}_{k,k}\mathbf{T}_{1,k} - \mathbf{T}_{1,k}^H \mathbf{H}_{k,k}^H \mathbf{R}_{1,k} - \mathbf{R}_{2,k}^H \mathbf{H}_{k,k}^* \mathbf{T}_{2,k}^* - \mathbf{T}_{2,k}^T \mathbf{H}_{k,k}^T \mathbf{R}_{2,k} + \mathbf{R}_{1,k}^H \mathbf{C}_{\mathbf{y}_k}\mathbf{R}_{1,k}$$
$$+ \mathbf{R}_{2,k}^H \mathbf{C}_{\mathbf{y}_k}^* \mathbf{R}_{2,k} + \mathbf{R}_{1,k}^H \tilde{\mathbf{C}}_{\mathbf{y}_k}\mathbf{R}_{2,k} + \mathbf{R}_{2,k}^H \tilde{\mathbf{C}}_{\mathbf{y}_k}^* \mathbf{R}_{1,k} \ . \quad (5.8)$$

Further, an additional matrix is of relevance when working with WLE receivers, which we call the pseudo-MSE-matrix. The *pseudo-MSE-matrix* is defined as $\tilde{\mathbf{E}}_k = \mathbb{E}[(\mathbf{b}_k - \hat{\mathbf{b}}_k)(\mathbf{b}_k - \hat{\mathbf{b}}_k)^T]$. Similarly as for the MSE-matrix in (5.8), we can develop $\tilde{\mathbf{E}}_k$ as a function of the transmit and receive filters in (5.1) and (5.7):

$$\tilde{\mathbf{E}}_k = -\mathbf{R}_{1,k}^H \mathbf{H}_{k,k}\mathbf{T}_{2,k} - \mathbf{T}_{1,k}^H \mathbf{H}_{k,k}^H \mathbf{R}_{2,k}^* - \mathbf{R}_{2,k}^H \mathbf{H}_{k,k}^* \mathbf{T}_{1,k}^* - \mathbf{T}_{2,k}^T \mathbf{H}_{k,k}^T \mathbf{R}_{1,k}^* + \mathbf{R}_{1,k}^H \tilde{\mathbf{C}}_{\mathbf{y}_k}\mathbf{R}_{1,k}^*$$
$$+ \mathbf{R}_{2,k}^H \tilde{\mathbf{C}}_{\mathbf{y}_k}^* \mathbf{R}_{2,k}^* + \mathbf{R}_{1,k}^H \mathbf{C}_{\mathbf{y}_k}\mathbf{R}_{2,k}^* + \mathbf{R}_{2,k}^H \mathbf{C}_{\mathbf{y}_k}^H \mathbf{R}_{1,k}^* \ . \quad (5.9)$$

Fixed all the transmit filters, $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$, the optimal widely linear receive filters for the k-th receiver are well known in the literature. They are obtained from the minimization of the MSE [65], i.e. $\mathrm{Tr}(\mathbf{E}_k)$, and are given by:

$$\mathbf{R}_{1,k} = \left(\mathbf{C}_{\mathbf{y}_k} - \tilde{\mathbf{C}}_{\mathbf{y}_k}\mathbf{C}_{\mathbf{y}_k}^{-*}\tilde{\mathbf{C}}_{\mathbf{y}_k}^*\right)^{-1}(\mathbf{H}_{k,k}\mathbf{T}_{1,k} - \tilde{\mathbf{C}}_{\mathbf{y}_k}\mathbf{C}_{\mathbf{y}_k}^{-*}\mathbf{H}_{k,k}^*\mathbf{T}_{2,k}^*), \quad (5.10)$$

$$\mathbf{R}_{2,k} = \left(\mathbf{C}_{\mathbf{y}_k} - \tilde{\mathbf{C}}_{\mathbf{y}_k}\mathbf{C}_{\mathbf{y}_k}^{-*}\tilde{\mathbf{C}}_{\mathbf{y}_k}^*\right)^{-1}(\mathbf{H}_{k,k}^*\mathbf{T}_{2,k}^* - \tilde{\mathbf{C}}_{\mathbf{y}_k}^*\mathbf{C}_{\mathbf{y}_k}^{-1}\mathbf{H}_{k,k}\mathbf{T}_{1,k}). \quad (5.11)$$

Under the assumption that interference is treated as Gaussian noise, the achievable rate for the k-th receiver is given by [69]:

$$R_k = \log_2 \frac{|\mathbf{C}_{\mathbf{y}_k}|}{|\mathbf{C}_{\mathbf{n}_k}|} + \tfrac{1}{2}\log_2 \left|\mathbf{I} - \mathbf{C}_{\mathbf{y}_k}^{-1}\tilde{\mathbf{C}}_{\mathbf{y}_k}\mathbf{C}_{\mathbf{y}_k}^{-T}\tilde{\mathbf{C}}_{\mathbf{y}_k}^H\right| - \tfrac{1}{2}\log_2 \left|\mathbf{I} - \mathbf{C}_{\mathbf{n}_k}^{-1}\tilde{\mathbf{C}}_{\mathbf{n}_k}\mathbf{C}_{\mathbf{n}_k}^{-T}\tilde{\mathbf{C}}_{\mathbf{n}_k}^H\right| \ . \quad (5.12)$$

Further, it is shown in Chapter 2 (see (2.13)) that the achievable rate R_k in (5.12) is related to the MSE-matrix $\mathbf{E}_k$ in (5.8) and to the pseudo-MSE-matrix $\tilde{\mathbf{E}}_k$ in (5.9) through:

$$R_k = \underset{\mathbf{R}_{1,k},\mathbf{R}_{2,k}}{\text{maximize}} -\tfrac{1}{2}\log_2 |\mathbf{E}_k\mathbf{F}_k^*|, \qquad (5.13)$$

where

$$\mathbf{F}_k = \mathbf{E}_k - \tilde{\mathbf{E}}_k\mathbf{E}_k^{-*}\tilde{\mathbf{E}}_k^*. \qquad (5.14)$$

The optimal receive filters $\mathbf{R}_{1,k}$, $\mathbf{R}_{2,k}$ maximizing the expression in (5.13) are those in (5.10)-(5.11), which can be demonstrated either by using the composite real formulation [65][**C1**] or by plugging the optimal structures of $\mathbf{R}_{1,k}$, $\mathbf{R}_{1,k}$ in (5.10)-(5.11) into $\mathbf{E}_k$ in (5.8) and $\tilde{\mathbf{E}}_k$ in (5.9). Therefore, the receive filters in (5.10)-(5.11) preserve the achievable rate in (5.12).

5.4 Problem Formulation

When adopting a maximum WSR (maxWSR) criterion, the problem of interest is to find (widely) linear transmit filters (or precoders) such that the WSR is maximized while the power budget of each transmitter is respected. Hence, (widely) linear transmit filters $\{\mathbf{T}_{1,k}\}, \{\mathbf{T}_{2,k}\}$ are obtained as the solution to the following maxWSR problem:

$$(\mathrm{P}_{5,1}): \underset{\{\mathbf{T}_{1,k}\},\{\mathbf{T}_{2,k}\}}{\text{maximize}} \quad \sum_{k\in\mathcal{K}} \mu_k R_k \qquad (5.15)$$

$$\text{subject to} \quad \begin{cases} \mathrm{Tr}(\mathbf{T}_{1,k}\mathbf{T}_{1,k}^H + \mathbf{T}_{2,k}\mathbf{T}_{2,k}^H) \le P_k^{\max} \quad \forall k \\ \mathbf{T}_{2,k} = \mathbf{0} \quad \forall k\in\mathcal{S}^{\mathrm{LP}} \end{cases}$$

where μ_k is a weighting coefficient associated to the priority of the k-th receiver (thus controlling the quality-of-service of each receiver), R_k is the achievable rate for the k-th receiver given in (5.12), and $P_k^{\max}$ is the maximum power available at the k-th transmitter. The design of the parameter μ_k is not addressed in the chapter, but in Section 5.7.4 we analyze an alternative design that allows guaranteeing fairness in the system. The second constraint in (5.15) imposes $\mathbf{T}_{2,k}=\mathbf{0}$ for those transmitters that are restricted to use LP (i.e. $k\in\mathcal{S}^{\mathrm{LP}}$). Due to interference, maxWSR problem $(\mathrm{P}_{5,1})$ in (5.15) is not convex w.r.t. $\{\mathbf{T}_{1,k}\}, \{\mathbf{T}_{2,k}\}$ (either jointly or separately), so the optimal solution cannot be guaranteed.

By using the relation among the achievable rate and the MSE-matrices in (5.13), the maxWSR problem $(\mathrm{P}_{5,1})$ in (5.15) can be equivalently written as follows [127, Sect. II.B]:

$$(\mathrm{P}_{5,2}): \underset{\substack{\{\mathbf{T}_{1,k}\},\{\mathbf{T}_{2,k}\} \\ \{\mathbf{R}_{1,k}\},\{\mathbf{R}_{2,k}\}}}{\text{minimize}} \quad \sum_{k\in\mathcal{K}} \tfrac{\mu_k}{2}\log_2|\mathbf{E}_k\mathbf{F}_k^*| \qquad (5.16)$$

$$\text{subject to} \quad \begin{cases} \mathrm{Tr}(\mathbf{T}_{1,k}\mathbf{T}_{1,k}^H + \mathbf{T}_{2,k}\mathbf{T}_{2,k}^H) \le P_k^{\max} \quad \forall k \\ \mathbf{T}_{2,k} = \mathbf{0} \quad \forall k\in\mathcal{S}^{\mathrm{LP}} \end{cases}$$

But again, due to interference, the maxWSR problem $(\mathrm{P}_{5,2})$ in (5.16) is not convex w.r.t. $\{\mathbf{T}_{1,k}\}$, $\{\mathbf{T}_{2,k}\}$, $\{\mathbf{R}_{1,k}\}$, $\{\mathbf{R}_{2,k}\}$. The equivalence among the maxWSR problems $(\mathrm{P}_{5,1})$ in (5.15) and $(\mathrm{P}_{5,2})$ in (5.16) is in the sense that the global optimal solution $\{\mathbf{T}_{1,k}^\star, \mathbf{T}_{2,k}^\star\}$ for the two problems is identical. Further, if $\{\mathbf{T}_{1,k}^\star, \mathbf{T}_{2,k}^\star, \mathbf{R}_{1,k}^\star, \mathbf{R}_{2,k}^\star\}$ is a stationary point of $(\mathrm{P}_{5,2})$ then $\{\mathbf{T}_{1,k}^\star, \mathbf{T}_{2,k}^\star\}$

is a stationary point of ($P_{5,1}$) (and the converse) [127].

5.5   Equivalent WMSE Formulation

In this section we propose an equivalent minimum weighted mean square error (minWMSE) problem that will allow tackling the maxWSR problem ($P_{5,2}$) presented in (5.16) for the case of improper Gaussian signaling with mixed transceivers.

Let us introduce auxiliary weighting matrices $\mathbf{W}_{1,k}$ and $\mathbf{W}_{2,k}$ for the k-th receiver, being $\mathbf{W}_{1,k}$ an hermitian positive semidefinite matrix and $\mathbf{W}_{2,k}$ a symmetric matrix. Then, the following result establishes the equivalence between the maxWSR problem ($P_{5,2}$) in (5.16) and the minWMSE problem ($P_{5,3}$) proposed in (5.17).

Theorem 5.1. *The maxWSR problem* ($P_{5,2}$) *in* (5.16) *is equivalent to the following minWMSE problem* ($P_{5,3}$):

$$(P_{5,3}): \min_{\substack{\{\mathbf{T}_{1,k}\},\{\mathbf{T}_{2,k}\} \\ \{\mathbf{R}_{1,k}\},\{\mathbf{R}_{2,k}\} \\ \{\mathbf{W}_{1,k}\},\{\mathbf{W}_{2,k}\}}} \sum_{k\in\mathcal{K}} \left(\tfrac{1}{2}\operatorname{Tr}\left(\mathbf{W}_{1,k}\mathbf{E}_k + \mathbf{W}_{1,k}^*\mathbf{E}_k^* + \mathbf{W}_{2,k}^*\tilde{\mathbf{E}}_k + \mathbf{W}_{2,k}\tilde{\mathbf{E}}_k^*\right) - \tfrac{\mu_k}{2}\log_2\left|\tfrac{\ln(2)}{\mu_k}\overline{\mathbf{W}}_k\right| \right)$$

(5.17)

$$\text{subject to} \quad \begin{cases} \operatorname{Tr}(\mathbf{T}_{1,k}\mathbf{T}_{1,k}^H + \mathbf{T}_{2,k}\mathbf{T}_{2,k}^H) \leq P_k^{\max} \quad \forall k \\ \mathbf{T}_{2,k} = \mathbf{0} \quad \forall k \in \mathcal{S}^{LP} \end{cases}$$

where $\mathbf{E}_k$ *and* $\tilde{\mathbf{E}}_k$ *correspond to the MSE-matrix and the pseudo-MSE-matrix for the k-th receiver detailed in* (5.8) *and* (5.9), *respectively, and* $\overline{\mathbf{W}}_k$ *is a block matrix given by:*

$$\overline{\mathbf{W}}_k = \begin{bmatrix} \mathbf{W}_{1,k} & \mathbf{W}_{2,k} \\ \mathbf{W}_{2,k}^* & \mathbf{W}_{1,k}^* \end{bmatrix},$$

(5.18)

which is hermitian positive semidefinite by construction.

The equivalence among the maxWSR problem ($P_{5,2}$) *in* (5.16) *and the minWMSE problem* ($P_{5,3}$) *in* (5.17) *is in the sense that the global optimal solution* $\{\mathbf{T}_{1,k}^\star, \mathbf{T}_{2,k}^\star, \mathbf{R}_{1,k}^\star, \mathbf{R}_{2,k}^\star\}$ *for the two problems is identical. Furthermore, if* $\{\mathbf{T}_{1,k}^\star, \mathbf{T}_{2,k}^\star, \mathbf{R}_{1,k}^\star, \mathbf{R}_{2,k}^\star, \mathbf{W}_{1,k}^\star, \mathbf{W}_{2,k}^\star\}$ *is a stationary point of* ($P_{5,3}$) *then* $\{\mathbf{T}_{1,k}^\star, \mathbf{T}_{2,k}^\star, \mathbf{R}_{1,k}^\star, \mathbf{R}_{2,k}^\star\}$ *is a stationary point of* ($P_{5,2}$) *(and the converse), as they satisfy the first-order optimality conditions of both problems.*

Proof. See Appendix 5.A.   ∎

The advantage of problem ($P_{5,3}$) in (5.17) as compared to ($P_{5,1}$) in (5.15) and ($P_{5,2}$) in (5.16) is that it is convex w.r.t. each set of variables separately (i.e. it is convex w.r.t. $\{\mathbf{T}_{1,k}\}$ given $\{\mathbf{T}_{2,k}\}$, $\{\mathbf{R}_{1,k}\}$, $\{\mathbf{R}_{2,k}\}$, $\{\mathbf{W}_{1,k}\}$, and $\{\mathbf{W}_{2,k}\}$, and so on). This property suggests that a block coordinate descent (BCD) method [139] can be used to find a stationary point to ($P_{5,3}$) in (5.17).

Let us emphasize that problem ($P_{5,3}$) in (5.17) is also valid to cover the minimization of the sum of MSE in the K-user MIMO IC if we set $\mathbf{W}_{2,k}=\mathbf{0}$ and $\mathbf{W}_{1,k}=\mathbf{I}$, $\forall k$, such that the objective function in (5.17) simply results: $\sum_{k\in\mathcal{K}}\operatorname{Tr}(\mathbf{E}_k)$. In this case, the use of heterogeneous (linear and widely linear) receivers could be included into the problem formulation, see [**C5**], as the use of LE or WLE affects the MSE but might not impact on the maximum achievable rate.

5.6   Centralized Coordinated Precoding

In this section we exploit the convex properties of the minWMSE problem ($P_{5,3}$) in (5.17) so as to find a stationary point solution. First, by checking the first-order optimality conditions of ($P_{5,3}$) in (5.17) and manipulating the obtained equalities, we derive analytical expressions for each paired-set of variables (i.e. $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$, $\{\mathbf{W}_{1,k}, \mathbf{W}_{2,k}\}$, and $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$) assuming that the remaining paired-sets are fixed. The results are given in the following propositions. Then, in Section 5.6.1 we propose an algorithm based on the BCD method with alternate optimization among the paired-sets of variables that is shown to reach a stationary point to ($P_{5,3}$) in (5.17).

The optimal widely linear receive filters $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$ to the maxWSR problem ($P_{5,2}$) in (5.16) for given transmit filters $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$ are those in (5.10)-(5.11), as the objective function can be uncoupled for each receiver when considering the variables $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$ and the receive filters in (5.10)-(5.11) are known to preserve the achievable rate (see (5.13)). Accordingly, $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$ in (5.10)-(5.11) are the optimal widely linear receive filters to the minWMSE problem ($P_{5,3}$) in (5.17) for given transmit filters $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$ and weighting matrices $\{\mathbf{W}_{1,k}, \mathbf{W}_{2,k}\}$.

The result for the optimal weighting matrices $\{\mathbf{W}_{1,k}, \mathbf{W}_{2,k}\}$ when $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$ and $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$ are fixed is detailed in the following corollary (as they were already derived within the proof of Theorem 5.1). See derivation in Appendix 5.A.

Corollary 5.1. *For given transmit filters $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$ and receive filters $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$, the optimal weighting matrices $\{\mathbf{W}_{1,k}, \mathbf{W}_{2,k}\}$ to the minWMSE problem ($P_{5,3}$) in (5.17) are:*

$$\mathbf{W}_{1,k} = \tfrac{\mu_k}{\ln(2)}\mathbf{F}_k^{-1}, \tag{5.19}$$

$$\mathbf{W}_{2,k} = -\tfrac{\mu_k}{\ln(2)}\mathbf{E}_k^{-1}\tilde{\mathbf{E}}_k\mathbf{F}_k^{-*}, \tag{5.20}$$

where $\mathbf{F}_k$ is the one defined in (5.14).

In order to compact the (widely) linear transmit filter design, let us define the following matrices:

$$\mathbf{A}_k = \mathbf{H}_{k,k}^H(\mathbf{R}_{1,k}\mathbf{W}_{1,k}\mathbf{R}_{1,k}^H + \mathbf{R}_{2,k}^*\mathbf{W}_{1,k}^*\mathbf{R}_{2,k}^T)\mathbf{H}_{k,k} + \boldsymbol{\Upsilon}_k, \tag{5.21}$$

$$\boldsymbol{\Upsilon}_k = \sum_{j \in \mathcal{K}, j \neq k} \mathbf{H}_{k,j}^H(\mathbf{R}_{1,j}\mathbf{W}_{1,j}\mathbf{R}_{1,j}^H + \mathbf{R}_{2,j}^*\mathbf{W}_{1,j}^*\mathbf{R}_{2,j}^T)\mathbf{H}_{k,j}, \tag{5.22}$$

$$\mathbf{B}_k = \mathbf{H}_{k,k}^H(\mathbf{R}_{1,k}\mathbf{W}_{1,k}\mathbf{R}_{2,k}^H + \mathbf{R}_{2,k}^*\mathbf{W}_{1,k}^*\mathbf{R}_{1,k}^T)\mathbf{H}_{k,k}^* + \boldsymbol{\Gamma}_k, \tag{5.23}$$

$$\boldsymbol{\Gamma}_k = \sum_{j \in \mathcal{K}, j \neq k} \mathbf{H}_{k,j}^H(\mathbf{R}_{1,j}\mathbf{W}_{1,j}\mathbf{R}_{2,j}^H + \mathbf{R}_{2,j}^*\mathbf{W}_{1,j}^*\mathbf{R}_{1,j}^T)\mathbf{H}_{k,j}^*, \tag{5.24}$$

$$\mathbf{C}_k = \mathbf{H}_{k,k}^H(\mathbf{R}_{1,k}\mathbf{W}_{2,k}\mathbf{R}_{2,k}^T + \mathbf{R}_{2,k}^*\mathbf{W}_{2,k}^*\mathbf{R}_{1,k}^H)\mathbf{H}_{k,k} + \boldsymbol{\Phi}_k, \tag{5.25}$$

$$\boldsymbol{\Phi}_k = \sum_{j \in \mathcal{K}, j \neq k} \mathbf{H}_{k,j}^H(\mathbf{R}_{1,j}\mathbf{W}_{2,j}\mathbf{R}_{2,j}^T + \mathbf{R}_{2,j}^*\mathbf{W}_{2,j}^*\mathbf{R}_{1,j}^H)\mathbf{H}_{k,j}, \tag{5.26}$$

$$\mathbf{D}_k = \mathbf{H}_{k,k}^H(\mathbf{R}_{1,k}\mathbf{W}_{2,k}\mathbf{R}_{1,k}^T + \mathbf{R}_{2,k}^*\mathbf{W}_{2,k}^*\mathbf{R}_{2,k}^H)\mathbf{H}_{k,k}^* + \boldsymbol{\Psi}_k, \tag{5.27}$$

$$\boldsymbol{\Psi}_k = \sum_{j \in \mathcal{K}, j \neq k} \mathbf{H}_{k,j}^H(\mathbf{R}_{1,j}\mathbf{W}_{2,j}\mathbf{R}_{1,j}^T + \mathbf{R}_{2,j}^*\mathbf{W}_{2,j}^*\mathbf{R}_{2,j}^H)\mathbf{H}_{k,j}^*. \tag{5.28}$$

The optimal (widely) linear transmit filters $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$ for fixed $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$ and fixed $\{\mathbf{W}_{1,k}, \mathbf{W}_{2,k}\}$ are detailed in the following proposition.

Proposition 5.1. *For given receive filters* $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$ *and weighting matrices* $\{\mathbf{W}_{1,k}, \mathbf{W}_{2,k}\}$, *the optimal transmit filters* $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$ *to the minWMSE problem* $(P_{5,3})$ *in* (5.17) *are given by:*

$$
\mathbf{T}_{1,k} = \begin{cases}
\mathbf{G}_k^{-1}\Big(\mathbf{H}_{k,k}^H\mathbf{R}_{1,k}\mathbf{W}_{1,k} + \mathbf{H}_{k,k}^H\mathbf{R}_{2,k}^*\mathbf{W}_{2,k}^* - \mathbf{J}_k(\mathbf{H}_{k,k}^T\mathbf{R}_{2,k}\mathbf{W}_{1,k} + \mathbf{H}_{k,k}^T\mathbf{R}_{1,k}^*\mathbf{W}_{2,k}^*)\Big) & \forall k \in \mathcal{S}^{WLP} \\
(\mathbf{A}_k + \mathbf{C}_k + \lambda_k\mathbf{I})^{-1}\Big(\mathbf{H}_{k,k}^H\mathbf{R}_{1,k}\mathbf{W}_{1,k} + \mathbf{H}_{k,k}^H\mathbf{R}_{2,k}^*\,\mathbf{W}_{2,k}^*\Big) & \forall k \in \mathcal{S}^{LP}
\end{cases},
$$

$$(5.29)$$

$$
\mathbf{T}_{2,k} = \begin{cases}
\mathbf{G}_k^{-1}\Big(\mathbf{H}_{k,k}^H\mathbf{R}_{2,k}^*\mathbf{W}_{1,k}^* + \mathbf{H}_{k,k}^H\mathbf{R}_{1,k}\mathbf{W}_{2,k} - \mathbf{J}_k(\mathbf{H}_{k,k}^T\mathbf{R}_{1,k}^*\mathbf{W}_{1,k}^* + \mathbf{H}_{k,k}^T\mathbf{R}_{2,k}\mathbf{W}_{2,k})\Big) & \forall k \in \mathcal{S}^{WLP} \\
\mathbf{0} & \forall k \in \mathcal{S}^{LP}
\end{cases},
$$

$$(5.30)$$

where

$$\mathbf{G}_k = \mathbf{A}_k + \mathbf{C}_k + \lambda_k\mathbf{I} - (\mathbf{B}_k + \mathbf{D}_k)(\mathbf{A}_k + \mathbf{C}_k + \lambda_k\mathbf{I})^{-*}(\mathbf{B}_k^* + \mathbf{D}_k^*), \tag{5.31}$$

$$\mathbf{J}_k = (\mathbf{B}_k + \mathbf{D}_k)(\mathbf{A}_k + \mathbf{C}_k + \lambda_k\mathbf{I})^{-*}, \tag{5.32}$$

and λ_k *is a non-negative dual variable associated to the k-th transmit power constraint in* (5.17). *Since* $(\mathbf{A}_k + \mathbf{C}_k + \lambda_k\mathbf{I})$ *is a full rank matrix, it can be proven by means of the Schur complement property* [92] *that the inverse of* $\mathbf{G}_k$ *in* (5.31) *exists.*

Proof. See Appendix 5.B.   ∎

5.6.1   Algorithm

In order to solve $(P_{5,3})$ in (5.17) we use the BCD method [139] with alternate optimization among the three different paired-sets of variables (i.e. $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$, $\{\mathbf{W}_{1,k}, \mathbf{W}_{2,k}\}$, and $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$). The alternate optimization is detailed in Algorithm 5.1. By departing from an initialization of the (widely) linear transmit filters that satisfies the per-transmitter power constraints in (5.17), then we alternatively update: *i)* receive filters $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$ using (5.10)-(5.11), *ii)* weighting matrices $\{\mathbf{W}_{1,k}, \mathbf{W}_{2,k}\}$ using (5.19)-(5.20), and *iii)* transmit filters $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$ using (5.29)-(5.30). The procedure is iterated until a stop condition (e.g. convergence is achieved, tolerance criterion is met, or maximum number of iterations is reached).

Note that if all transmit filters were initialized with LP (i.e. $\mathbf{T}_{2,k} = \mathbf{0}$, $\forall k$), then all receivers would get $\mathbf{R}_{2,k} = \mathbf{0}$ in (5.11) since $\mathbf{T}_{2,k} = \mathbf{0}$ and $\tilde{\mathbf{C}}_{\mathbf{y}_k} = \mathbf{0}$, $\forall k$, and the proposed iterative procedure would get stuck to a LP-LE scheme (i.e. proper-based solution). A similar observation was found in [109] with iterative algorithms for the MIMO BC: it was shown that a WLP initialization was required to get an improper-based solution. In Algorithm 5.1, so as to avoid getting stuck to a proper-based solution, it is required that at least one transmitter is initialized with WLP.

The procedure detailed in Algorithm 5.1 can be implemented either in a centralized or decentralized manner. In the centralized implementation, a central processing node gathers all the channel matrices in the system (i.e. $\mathbf{H}_{j,k}, \forall k, \forall j$) and then performs the iterative procedure in Algorithm 5.1. After that, transmitters are informed of the transmit filters to be used and each receiver updates its receive filters based on the received signal and an estimation of the equivalent channel (as it is done in practical implementations [29, 141]). On the other hand,

---

Algorithm 5.1 Mixed transceiver design to solve minWMSE problem ($\text{P}_{5,3}$) in (5.17)

1: **initialize**: $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$, $\forall k$
2: **repeat**
3:      Compute $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$, $\forall k$, in (5.10)-(5.11) given $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$.
4:      Compute $\{\mathbf{W}_{1,k}, \mathbf{W}_{2,k}\}$, $\forall k$, in (5.19)-(5.20) given $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$, $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$.
5:      Compute $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$, $\forall k$, in (5.29)-(5.30) given $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$, $\{\mathbf{W}_{1,k}, \mathbf{W}_{2,k}\}$ (update λ_k through the bisection method [92] such that $\text{Tr}(\mathbf{T}_{1,k}\mathbf{T}_{1,k}^H + \mathbf{T}_{2,k}\mathbf{T}_{2,k}^H) \leq P_k^{\max}$)
6: **until** stop condition

---

in the decentralized implementation, the alternate optimization in Algorithm 5.1 is combined among transmitters and receivers: receivers perform lines 3 and 4 (i.e. update of receive filters and weighting matrices), while transmitters execute line 5 (i.e. update of transmit filters). To carry out such updates it is required that each k-th transmitter has available the channel matrices towards all receivers in the network (i.e. $\mathbf{H}_{k,j}, \forall j$). Further, it is also required that each receiver has an additional link to feedback the updated receive filters and weighting matrices to all transmitters in the network at each iteration (similar as in the linear transceiver case in [127]). In order to avoid such large requirements for channel knowledge and feedback links, in Section 5.6.4 we describe how the minWMSE problem ($\text{P}_{5,3}$) in (5.17) could be decomposed to obtain alternative decentralized procedures.

If Algorithm 5.1 is implemented in a central processing node, then LP transmitters do not need to be aware that IGS is being used in the network. However, for decentralized implementations it is required that LP transmitters know that WLE is being applied at receivers.

5.6.2 Convergence

The minWMSE problem ($\text{P}_{5,3}$) in (5.17) is convex w.r.t. each set of variables separately and closed-form expressions have been derived. This ensures that if we update the sets of variables one-by-one then a monotonic reduction of the objective function of ($\text{P}_{5,3}$) is obtained. Further, by checking the first-order optimality conditions we have arrived to a unique optimal solution for each paired-set of variables given the other two paired-sets of variables (e.g. $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$ given $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$ and $\{\mathbf{W}_{1,k}, \mathbf{W}_{2,k}\}$). This way, the two sets on each paired-set can be updated simultaneously while maintaining the monotonic reduction properties of the BCD method applied over the minWMSE problem ($\text{P}_{5,3}$). Accordingly, with the alternating minimization process in Algorithm 5.1, the objective function of ($\text{P}_{5,3}$) in (5.17) decreases monotonically at each iteration.

Theorem 5.2. *Algorithm 5.1 converges to a stationary point of problem* ($\text{P}_{5,1}$) *in* (5.15).

Proof. Due to the equivalent relationships among the maxWSR problems ($\text{P}_{5,1}$) in (5.15) and ($\text{P}_{5,2}$) in (5.16), and among the maxWSR problem ($\text{P}_{5,2}$) in (5.16) and the minWMSE problem ($\text{P}_{5,3}$) in (5.17), it is sufficient to show that Algorithm 5.1 converges to a stationary point of ($\text{P}_{5,3}$). When we fix two of the paired-sets of variables (e.g. $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$ and $\{\mathbf{W}_{1,k}, \mathbf{W}_{2,k}\}$), a unique optimal solution is obtained for the remaining paired-set of variables (e.g. $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$). Accordingly, the alternate optimization in Algorithm 5.1 ensures a monotonic reduction of the objective function of ($\text{P}_{5,3}$) at each iteration and, as the objective function is bounded by the optimal value, convergence in terms of the objective function is achieved. Furthermore, since

problem ($P_{5,3}$) has a differentiable objective function, the constraint set is separable among the paired-sets of variables, and the problem has a unique minimum point along any coordinate direction, then it follows from general optimization theory (see [146, Sect. 8.9]) that Algorithm 5.1 (that is based on the BCD method) converges to a stationary point of ($P_{5,3}$). Finally, due to Theorem 5.1, the obtained solution is also a stationary point of the maxWSR problem ($P_{5,2}$) in (5.16) (and, equivalently, of problem ($P_{5,1}$) in (5.15)).   ∎

5.6.3   Complexity of centralized implementation

The complexity of the proposed Algorithm 5.1 is related to the number of iterations. Simulations show that around 15-20 iterations are enough to reach a stationary point of the maxWSR problem. However, the larger improvement is obtained in the first iterations (see Section 5.7.1). For decentralized implementations, the number of iterations is relevant and should be limited. In this case, Theorem 5.2 demonstrates that at each iteration the objective function of ($P_{5,3}$) in (5.17) is monotonically reduced, which in the general case leads to an improvement of the objective function of ($P_{5,3}$) in (5.15) (although a monotonic WSR improvement cannot be guaranteed). It is important to recall that the proposed Algorithm 5.1 already outperforms the conventional coordinated linear transceiver schemes with a low number of iterations (e.g. 5, see Section 5.7.1).

   At each iteration of Algorithm 5.1, three different updates are performed (i.e. one update for each paired-set of variables). For simplicity of the complexity analysis, let K be the total number of transmit-receive pairs, let M, N denote the number of antennas at each transmitter and receiver, respectively, and let m be the number of streams per receiver. We assume that all transceivers are widely linear to compute an upper bound of the total complexity. For matrices $\mathbf{X} \in \mathbb{C}^{a \times b}$, $\mathbf{Y} \in \mathbb{C}^{b \times c}$, $\mathbf{Z} \in \mathbb{C}^{d \times d}$, the complexity of the product $\mathbf{XY}$ is $\mathcal{O}(abc)$, the complexity of the sum $\mathbf{X}+\mathbf{X}$ is $\mathcal{O}(ab)$, and the complexity of the inverse $\mathbf{Z}^{-1}$ is $\mathcal{O}(d^3)$ [87]. Then, the complexity at each step in one iteration of Algorithm 5.1 is:

- updating all receive filters (line 3): $\mathcal{O}(K^2MNm+K^2N^2m+KN^3)$,
- updating all weighting matrices (line 4): $\mathcal{O}(KN^2m+Km^2N+Km^3)$,
- updating all transmit filters (line 5): $\mathcal{O}(K^2MNm+K^2M^2m+K^2Mm^2+KM^3)$.

If we use the upper bound $m < M$, the total complexity per iteration of the proposed Algorithm 5.1 is upper bounded by: $\mathcal{O}(K^2M^3+K^2M^2N+K^2N^2M+KN^3)$. The order of complexity is the same as the one associated to the algorithm proposed in [127] for the linear transceiver case.

5.6.4   Decentralized formulation

By taking into account the decompositions of matrices $\mathbf{A}_k$ in (5.21), $\mathbf{B}_k$ in (5.23), $\mathbf{C}_k$ in (5.25), and $\mathbf{D}_k$ in (5.27), decentralized approaches could be derived if each k-th transmitter had knowledge of $\mathbf{\Upsilon}_k$ in (5.22), $\mathbf{\Gamma}_k$ in (5.24), $\mathbf{\Phi}_k$ in (5.26), and $\mathbf{\Psi}_k$ in (5.28), $\mathbf{C}_{\mathbf{n}_k}$ in (5.5), and $\tilde{\mathbf{C}}_{\mathbf{n}_k}$ in (5.6). $\mathbf{\Upsilon}_k$, $\mathbf{\Gamma}_k$, $\mathbf{\Phi}_k$, and $\mathbf{\Psi}_k$ are the terms that require knowledge of information not available either at the k-th transmitter or at the k-th receiver, while $\mathbf{C}_{\mathbf{n}_k}$ and $\tilde{\mathbf{C}}_{\mathbf{n}_k}$ could be reported by the intended k-th receiver.

Proposition 5.2. *The minWMSE problem* (P$_{5,3}$) *in* (5.17) *can be decomposed into* K *parallel optimization problems (one per transmitter). The problem to be solved at the* k-*th transmitter for a fixed* $\boldsymbol{\Upsilon}_k$, $\boldsymbol{\Gamma}_k$, $\boldsymbol{\Phi}_k$, $\boldsymbol{\Psi}_k$, $\mathbf{C}_{\mathbf{n}_k}$, *and* $\tilde{\mathbf{C}}_{\mathbf{n}_k}$, *is:*

$$(\text{P}_{5,4}^k) : \min_{\substack{\mathbf{T}_{1,k},\mathbf{T}_{2,k},\mathbf{R}_{1,k},\mathbf{R}_{2,k}, \\ \mathbf{W}_{1,k},\mathbf{W}_{2,k}}} \quad f_k + g_k \tag{5.33}$$

$$\text{subject to} \quad \begin{cases} \text{Tr}(\mathbf{T}_{1,k}\mathbf{T}_{1,k}^H + \mathbf{T}_{2,k}\mathbf{T}_{2,k}^H) \leq P_k^{\max} \\ \mathbf{T}_{2,k} = \mathbf{0} \quad \forall k \in \mathcal{S}^{LP} \end{cases}$$

where f_k *accounts for the impact over the intended receiver:*

$$f_k = \tfrac{1}{2}\,\text{Tr}\left(\mathbf{W}_{1,k}\mathbf{E}_k + \mathbf{W}_{1,k}^*\mathbf{E}_k^* + \mathbf{W}_{2,k}^*\tilde{\mathbf{E}}_k + \mathbf{W}_{2,k}\tilde{\mathbf{E}}_k^*\right) - \tfrac{\mu_k}{2}\log_2\left|\tfrac{\ln(2)}{\mu_k}\overline{\mathbf{W}}_k\right|, \tag{5.34}$$

while g_k *reflects the impact over the unintended receivers:*

$$g_k = \text{Tr}\left((\boldsymbol{\Upsilon}_k + \boldsymbol{\Phi}_k)(\mathbf{T}_{1,k}\mathbf{T}_{1,k}^H + \mathbf{T}_{2,k}\mathbf{T}_{2,k}^H)\right) + \text{Tr}\left((\boldsymbol{\Gamma}_k + \boldsymbol{\Psi}_k)(\mathbf{T}_{2,k}^*\mathbf{T}_{1,k}^H + \mathbf{T}_{1,k}\mathbf{T}_{2,k}^T)\right) . \tag{5.35}$$

Proof. The first-order optimality conditions of the minWMSE problem (P$_{5,3}$) in (5.17) and the decomposed problem (P$_{5,4}^k$) in (5.33) are the same for a fixed $\boldsymbol{\Upsilon}_k$, $\boldsymbol{\Gamma}_k$, $\boldsymbol{\Phi}_k$, $\boldsymbol{\Psi}_k$, $\mathbf{C}_{\mathbf{n}_k}$, and $\tilde{\mathbf{C}}_{\mathbf{n}_k}$. See the conditions in Appendices 5.A and 5.B. ∎

Problem (P$_{5,4}^k$) in (5.33) is convex w.r.t. each variable separately and it leads to solutions in (5.10)-(5.11), (5.19)-(5.20), and (5.29)-(5.30). Then, each k-th transmitter can solve problem (P$_{5,4}^k$) in (5.33) with alternate optimization between receive filters $\mathbf{R}_{1,k}$, $\mathbf{R}_{2,k}$, weighting matrices $\mathbf{W}_{1,k}$, $\mathbf{W}_{2,k}$, and transmit filters $\mathbf{T}_{1,k}$, $\mathbf{T}_{2,k}$.

A possible mechanism to obtain the matrices $\boldsymbol{\Upsilon}_k$, $\boldsymbol{\Phi}_k$, $\boldsymbol{\Gamma}_k$, and $\boldsymbol{\Psi}_k$ (also called the interference-cost and pseudo-interference cost matrices [**C5**]) is by exchanging control-plane messages among transmitters, as $\boldsymbol{\Upsilon}_k$, $\boldsymbol{\Phi}_k$, $\boldsymbol{\Gamma}_k$, and $\boldsymbol{\Psi}_k$ can be seen as the sum of information from neighboring transmitters. See details for the linear transceiver case in Chapter 3. However, such approach requires knowledge of all the interfering channel matrices to compute $\boldsymbol{\Upsilon}_k$, $\boldsymbol{\Phi}_k$, $\boldsymbol{\Gamma}_k$, and $\boldsymbol{\Psi}_k$.

An alternative procedure to obtain matrices $\boldsymbol{\Upsilon}_k$, $\boldsymbol{\Phi}_k$, $\boldsymbol{\Gamma}_k$, and $\boldsymbol{\Psi}_k$ is by exploiting UL-DL propagation channel reciprocity, as is done in the proper signaling-based procedure in Chapter 3. In that case, if we focus on DL interference coordination, matrices $\boldsymbol{\Upsilon}_k$, $\boldsymbol{\Gamma}_k$, $\boldsymbol{\Phi}_k$, and $\boldsymbol{\Psi}_k$, could be obtained from a UL pilot-based transmission provided that receivers use a specific pilot precoder for UL transmission that should be a function of the DL receive filters. By doing so, estimation of the interfering channels is not required and information exchange among transmitters is avoided. We omit the signal model here. See more details for the linear transceiver case in Chapter 3 ([**J2**][**C2**]), and for the widely linear transceiver case with minimum MSE criterion in [**C5**].

5.6.5 Multi-user MIMO extension

The minWMSE problem (P$_{5,3}$) in (5.17) (as well as its decomposed problem (P$_{5,4}^k$) in (5.33)) can easily be extended to the case in which each transmitter serves multiple receivers simultaneously on the same time/frequency resource (multi-cell multi-user MIMO system). In this case, the power constraint associated to each transmitter couples the design of the (widely) linear transmit filters to be used to serve the associated receivers, but the solution can be derived by following the approach in Section 5.5. See extension in Chapter 3 ([**J2**]) for the linear transceiver case.

5.7   Simulation Results

In this section we evaluate through Montecarlo simulations the performance of the proposed Algorithm 5.1 in Section 5.5. The procedure is compared to the one in [127] that uses linear transceivers and proper Gaussian signaling (i.e. LP-LE). A MIMO IC composed of $|\mathcal{K}|$ transmitter-receiver pairs is considered. Channels are modeled through a Rayleigh distribution: $\mathbf{H}_{k,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}), \forall k$, $\mathbf{H}_{k,j} \sim \mathcal{CN}(\mathbf{0}, \eta\mathbf{I}), \forall j \neq k$, where factor $\eta \geq 0$ models the relative average interference strength. The noise is assumed to be white proper Gaussian, i.e. $\mathbf{v}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$, $\forall k$. All transmitters are assumed to have the same available power $P_k = P$, $\forall k$, such that the signal-to-noise ratio is defined by $\mathrm{SNR} = P/\sigma^2$. All transmitters (or receivers) are assumed to have the same number of antennas, which is denoted by M (or N). Antenna configurations are depicted in figures as $M \times N$. 500 channel realizations are averaged.

The following techniques are evaluated:

- **LP-LE**: interference coordination technique in [127] with linear transceivers (LP-LE).

- **LP-LE IW**: iterative water-filling (IW) algorithm [144] with linear transceivers (LP-LE). It is equivalent to iteratively solve ($\mathrm{P}_{5,4}^k$) in (5.33) at each k-th transmitter with $g_k = 0$ (i.e. no coordination) and LP-LE.

- **WLP-WLE**: interference coordination technique proposed in [127] but using the composite real formulation (i.e. the equivalent double-sized real-valued MIMO IC) such that widely linear transceivers are obtained (WLP-WLE)[20]. The proposed scheme in this chapter would provide the same solution.

- **WLP-WLE IW**: IW algorithm with widely linear transceivers (WLP-WLE). It is equivalent to iteratively solve ($\mathrm{P}_{5,4}^k$) in (5.33) at each k-th transmitter with $g_k = 0$ (i.e. no coordination) and WLP-WLE.

- **HetTX(l)-WLE**: interference coordination technique proposed in this chapter for a scenario of heterogeneous transmitters, whereby $\lfloor l \rfloor$ transmitters employ LP while the remaining ($|\mathcal{K}| - \lfloor l \rfloor$) transmitters can adopt WLP. $l = 1$ and $l = |\mathcal{K}|/2$ are used for simulations.

The number of iterations of all algorithms is set to 50, and the best result among 3 random transmit initializations is used. For the improper-based optimizations, a WLP initialization is required at least in some transmitters (i.e. random $\{\mathbf{T}_{1,k}, \mathbf{T}_{2,k}\}$ such that the power budget at each k-th transmitter is respected), otherwise the proposed procedure would lead to a proper Gaussian signaling solution (see discussion about initialization in Section 5.6.1).

The performance indicator is the sum of achievable rates measured in bits/s/Hz. $\mu_k = 1$ is used in (5.15), $\forall k$, except for Section 5.7.4.

In Section 5.7.1 we show the convergence of Algorithm 5.1. The sum-rate performance is evaluated in Section 5.7.2 for different antenna configurations when varying the number of transmitter-receiver pairs ($|\mathcal{K}|$) and the interference strength (η). In Section 5.7.3, the proposed technique is compared with interference alignment schemes at different SNR regimes. Finally, in Section 5.7.4, a slightly modified version of Algorithm 5.1 is presented to guarantee fairness in the system and the performance is shown both in terms of sum-rate and 5%-tile rate.

---

[20]Recall that any approach (as the coordinated technique proposed in [127]) can be adopted to get WLP-WLE solutions by using the equivalent double-sized real-valued MIMO IC provided that there are no constraints on the linear and widely linear operation of the transceivers in the network.
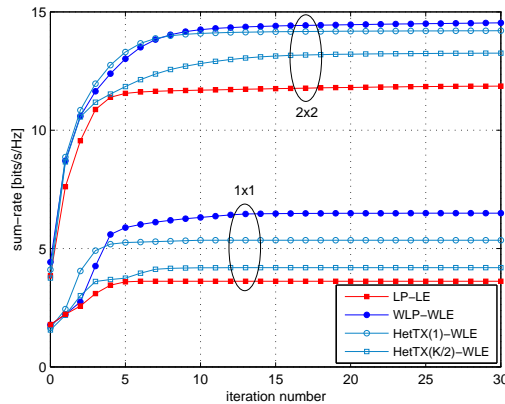
Figure 5.2: Sum-rate (in bits/s/Hz) vs. iteration number in a random channel realization of the K-user MIMO IC. $|\mathcal{K}|$=5, SNR=10 dB, η=1. Antenna configuration: 2×2 and 1×1.

5.7.1 Convergence

In this section we verify the convergence of the proposed Algorithm 5.1 for mixed transceiver design. The performance is compared to the one obtained for the linear transceiver case (i.e. LP-LE) in [127].

Fig. 5.2 shows the sum-rate versus the iteration number for a specific channel realization with antenna configurations 2×2 and 1×1, $|\mathcal{K}|$=5, SNR=10 dB, and η=1. Convergence in terms of sum-rate is observed. Further, monotonic convergence is obtained in this channel realization. Note that monotonic convergence is guaranteed in terms of the objective function of problem $(P_{5,3})$ in (5.17) but not necessarily in terms of sum-rate, although this is also often the case. The convergence speed is a bit slower for 'WLP-WLE' and 'HetTX-WLE' as compared to 'LP-LE', although it varies depending on the individual channel realizations.

5.7.2 Impact of antenna configuration and interference strength

In this section we evaluate the performance of the proposed Algorithm 5.1 for different antenna configurations ($M{\times}N$), different $|\mathcal{K}|$ values, and different η values. SNR=10 dB is used.

Fig. 5.3 displays the sum-rate versus $|\mathcal{K}|$ for SNR=10 dB, η=1, and different antenna configurations: (a) for 1×1, (b) for 2×2, and (c) for 4×4. As it is expected, the sum-rate is increased as $|\mathcal{K}|$ increases when transmit coordination is implemented. However, for the cases in which transmit coordination is not used (i.e. IW solutions) the tendency is on the contrary because all transmitters use the maximum transmit power, such that increasing $|\mathcal{K}|$ implies a significant increase of the interference in the network and, consequently, a degradation of the system performance.

Figure 5.4 depicts the sum-rate versus η for $|\mathcal{K}|$=5, SNR=10 dB, and different antenna configurations: (a) for 1×1 and (b) for 2×2.
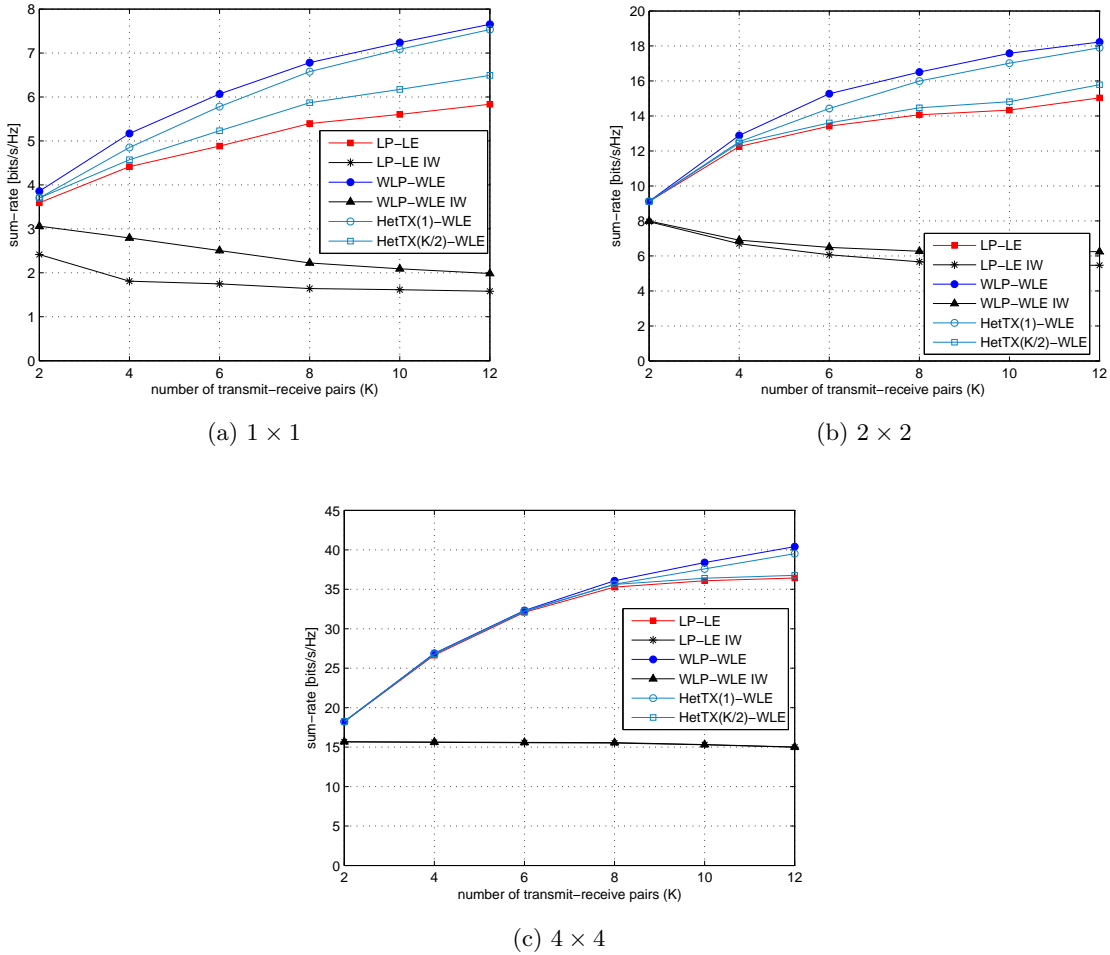
(a) 1×1

(b) 2×2

(c) 4×4

Figure 5.3: Sum-rate (in bits/s/Hz) vs. $|\mathcal{K}|$ for the K-user MIMO IC. SNR=10 dB, η=1.
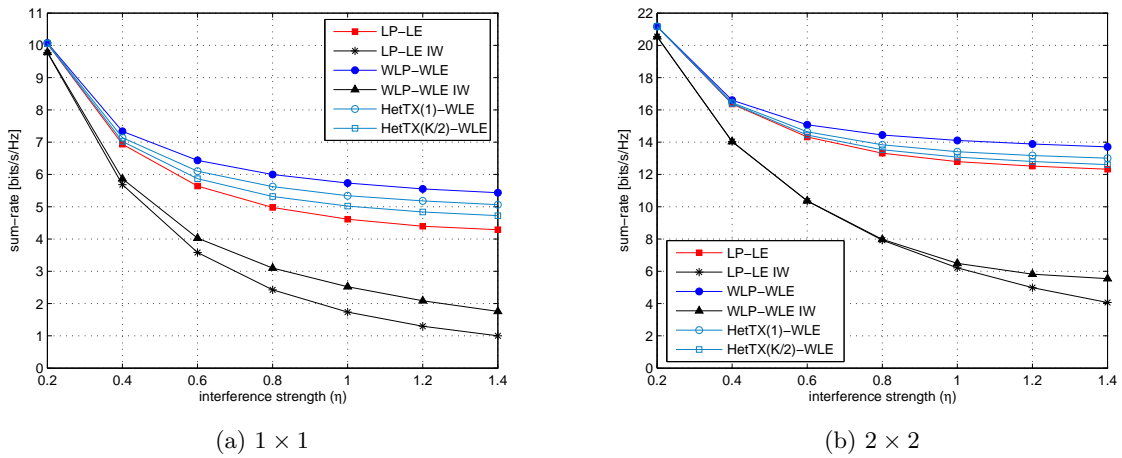

(a) 1×1

(b) 2×2

Figure 5.4: Sum-rate (in bits/s/Hz) vs. η (interference strength) for the K-user MIMO IC. $|\mathcal{K}|$=5, SNR=10 dB.

From Fig. 5.3 and Fig. 5.4 several conclusions can be extracted:

- *Proper (LP-LE) vs. improper (WLP-WLE)*: In all cases it can be observed that 'WLP-WLE IW' outperforms 'LP-LE IW'. Also 'WLP-WLE' and 'HetTX-WLE' outperform 'LP-LE', as the proper-based optimization is a special case of the improper-based optimization (but not the other way round).

- *IW vs. interference coordination*: The sum-rate gain obtained with interference coordination techniques is significant for all antenna configurations, all numbers of interfering transmitters, and all interference levels.

- *Number of transmitter-receiver pairs ($|\mathcal{K}|$)*: The sum-rate increment provided by the use of improper Gaussian signaling is larger as $|\mathcal{K}|$ increases due to the fact that the interference to be managed is stronger, see Fig. 5.3. The $|\mathcal{K}|$ value for which the gains start to appear depends on the antenna configuration, as detailed in next bullet.

- *Antenna configuration ($M \times N$)*: The sum-rate gain obtained with the use of improper Gaussian signaling is larger for the 1×1 and 2×2 cases rather than for the 4×4 case, because the use of improper Gaussian signaling provides flexibility by splitting one dimension into two halves. This is more useful when the number of transmit/receive antennas is low compared to the number of users. Otherwise, by adding antennas, extra dimensions are already added to the system. For that reason, the gains of interference coordination and improper Gaussian signaling are appreciable for $|\mathcal{K}| \geq 4$ in the 2×2 case and for $|\mathcal{K}| \geq 8$ in the 4×4 case, while in the 1×1 case they are substantial for $|\mathcal{K}| \geq 2$ (i.e. even for a scenario with a single interferer), see Fig. 5.3.

- *Interference strength (η)*: The sum-rate gain provided by improper Gaussian signaling is larger as the interference level increases (i.e. for larger η), see Fig. 5.4.

- *Heterogeneous scenarios*: The sum-rate gains of improper Gaussian signaling in heterogeneous scenarios are not proportional to the number of widely linear transmitters in the network. The largest gains are obtained when most of the transmitters are widely linear (i.e. lower l in 'HetTX(l)-WLE').

5.7.3   Performance versus SNR and comparison with IA

In this section we evaluate the performance of the proposed Algorithm 5.1 at different SNR regimes. We compare the proposed technique with two algorithms based on interference alignment (IA) that are presented in [43]:

- **IA**: distributed iterative IA algorithm in [43].
- **Max SINR**: distributed Max-SINR algorithm in [43].

These two algorithms are optimal in terms of DoF (i.e. slope of the sum-rate at the high SNR regime) for the MIMO IC with $|\mathcal{K}|=3$ users and 2×2 antenna configuration, but not for other configurations.

Figure 5.5 shows the sum-rate versus SNR for $\eta=1$, 2×2 antenna configuration, and different $|\mathcal{K}|$ setups: (a) $|\mathcal{K}|=3$ and (b) $|\mathcal{K}|=5$. For $|\mathcal{K}|=3$ and $M=N=2$, 'IA' and 'Max SINR' algorithms attain the optimal DoF (i.e. 3 DoF), as shown in Fig. 5.5.(a). In contrast, the proposed scheme attains only 2 DoF but it is able to provide significant sum-rate values at low-medium SNR regimes. For $|\mathcal{K}|=5$ and $M=N=2$, 'Max SINR' and 'IA' algorithms tend to become saturated

(a) $2 \times 2, |\mathcal{K}| = 3$ (b) $2 \times 2, |\mathcal{K}| = 5$
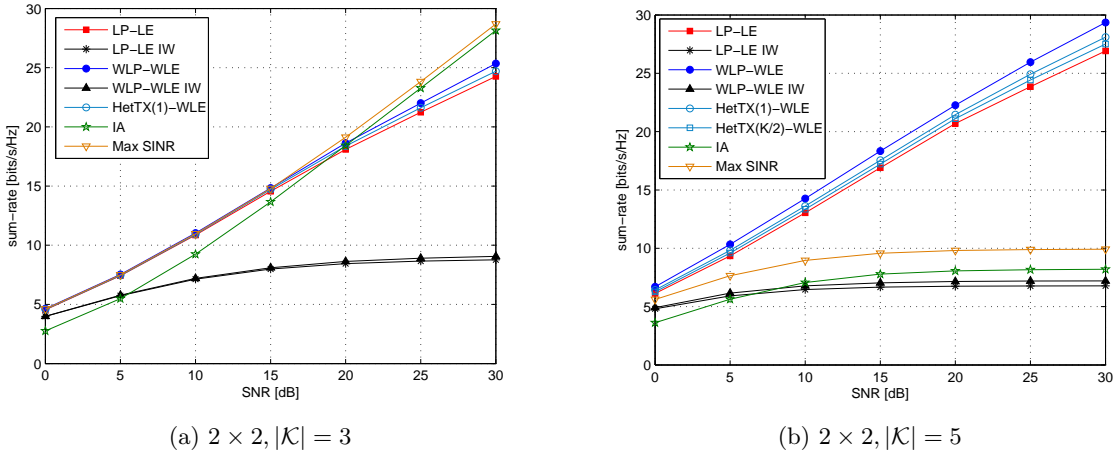
Figure 5.5: Sum-rate (in bits/s/Hz) vs. SNR (in dB) for the K-user MIMO IC. η=1, 2×2.

in terms of sum-rate (thus 0 DoF), see Fig. 5.5.(b), while the proposed technique is able to get 2.2 DoF.

Many of the existing schemes for IA with improper Gaussian signaling are only valid for specific antenna configurations and number of users. On the contrary, the proposed scheme can be applied in all possible antenna/users configurations and still get some positive DoF.

For $|\mathcal{K}|$=5, we can conclude that the gains of IGS are appreciable for all SNR regimes. $|\mathcal{K}|$=3 has been used for comparison with the IA algorithms, but it was already shown in Fig. 5.3.(b) that in the 2×2 case we should use $|\mathcal{K}| \geq 4$ to get appreciable gains with IGS over 'LP-LE'.

5.7.4   Performance with fairness utility

In this section we evaluate the performance of a slightly modified version of the proposed Algorithm 5.1. In [127] it is shown that general utility functions can be accommodated in the WMSE problem formulation and only the design of the weighting matrices turns out to be affected. For example, in the linear transceiver design case, if a proportional fair utility function is adopted then the weighting matrices should be updated as: $\mathbf{W}_{1,k} = \frac{1}{R_k \ln(2)} \mathbf{E}_k^{-1}$ (see [127]). Similarly, in order to guarantee fairness in the system with mixed transceivers, we update the weighting matrices in Algorithm 5.1 as:

$$\mathbf{W}_{1,k} = \frac{1}{R_k \ln(2)} \mathbf{F}_k^{-1}, \quad \mathbf{W}_{2,k} = -\frac{1}{R_k \ln(2)} \mathbf{E}_k^{-1} \tilde{\mathbf{E}}_k \mathbf{F}_k^{-*}, \quad (5.36)$$

being R_k the achievable rate obtained in the previous iteration.

Figure 5.6 shows the performance in terms of sum-rate and 5%-tile rate versus $|\mathcal{K}|$ for SNR=10 dB, η=1, and antenna configuration 2×2. It can be observed that 'WLP-WLE' and 'HetTX-WLE' techniques allow increasing the system fairness (as shown in terms of the 5%-tile rate in Fig. 5.6.(b)) while large gains in terms of sum-rate are still obtained (see Fig. 5.6.(a)).
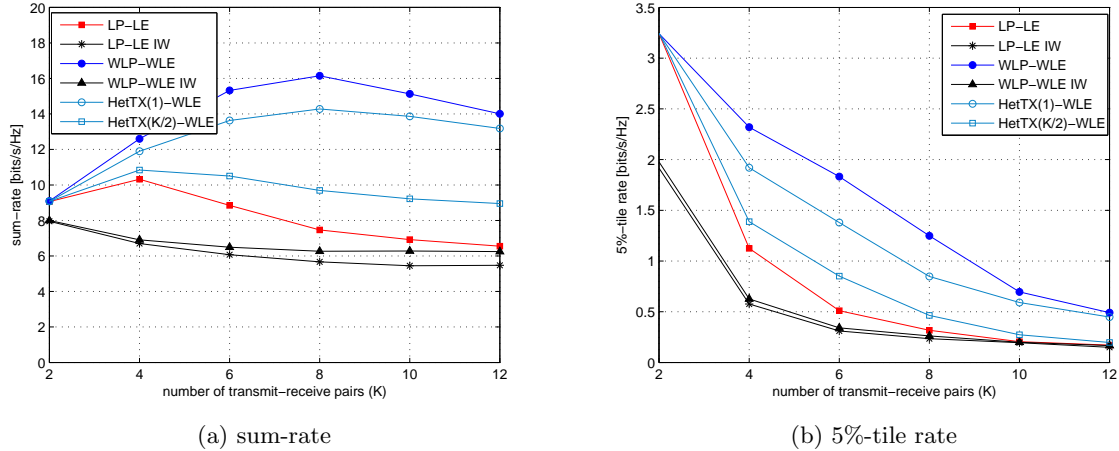
(a) sum-rate   (b) 5%-tile rate

Figure 5.6: Performance vs. $|\mathcal{K}|$ for the K-user MIMO IC. SNR=10 dB, η=1, 2×2. Fairness utility.

5.8 Conclusions

In this chapter, transceiver design with improper Gaussian signaling for weighted sum-rate maximization in the K-user MIMO interference channel has been investigated. Maximization of the WSR is formulated using the complex-valued channel model, which allows covering a scenario where different types of transmitters (linear and widely linear) coexist. The initial maximum WSR problem is solved through the minimization of an equivalent WMSE problem and closed-form expressions for mixed transceiver design are derived. In this regard, an iterative algorithm is presented which allows reaching a stationary point of the maximum WSR problem.

The proposed transceiver design (including transmit coordination and the use of improper Gaussian signaling) provides gains in terms of sum-rate as compared to conventional transmit coordination techniques with proper Gaussian signaling. The largest gains are observed in the following non-exclusive situations:

- when the level of interference is high (either due to a small number of strong interferers or owing to multiple interfering nodes), or
- when the number of transmit/receive antennas is low.

If the interference level is not significant or new dimensions are provided by adding multiple antennas at transmitters and receivers, the proposed coordinated transceiver scheme leads to a proper Gaussian signaling solution (i.e. solution in Chapter 3).

Appendices

5.A Proof of Theorem 5.1

The equivalence of the minWMSE problem ($P_{5,3}$) in (5.17) and the maxWSR problem ($P_{5,2}$) in (5.16) is shown in the following by deriving the optimal weighting matrices $\{\mathbf{W}_{1,k}\}$, $\{\mathbf{W}_{2,k}\}$, for ($P_{5,3}$), plugging them into the objective function of ($P_{5,3}$), and then showing that the resulting optimization problem is exactly ($P_{5,2}$).

The Lagrangian function ($\mathcal{L}$) [92] of the minWMSE problem ($P_{5,3}$) in (5.17) is given by:

$$
\begin{aligned}
\mathcal{L} \;=\; & \sum_{k\in\mathcal{K}} \left(\tfrac{1}{2}\,\mathrm{Tr}\left(\mathbf{W}_{1,k}\mathbf{E}_k + \mathbf{W}_{1,k}^*\mathbf{E}_k^* + \mathbf{W}_{2,k}^*\tilde{\mathbf{E}}_k + \mathbf{W}_{2,k}\tilde{\mathbf{E}}_k^*\right) - \tfrac{\mu_k}{2}\log_2\left|\tfrac{\ln(2)}{\mu_k}\overline{\mathbf{W}}_k\right| \right) \\
& + \sum_{k\in\mathcal{K}} \lambda_k\left(\,\mathrm{Tr}(\mathbf{T}_{1,k}\mathbf{T}_{1,k}^H + \mathbf{T}_{2,k}\mathbf{T}_{2,k}^H) - P_k^{\max}\right),
\end{aligned}
\tag{5.37}
$$

where λ_k denotes the non-negative dual variable associated to the k-th transmit power constraint in (5.17). Recall that the determinant of $\frac{\ln(2)}{\mu_k}\overline{\mathbf{W}}_k$ in (5.37) (through the use of the structure in (5.18)) is given by:

$$
\left|\tfrac{\ln(2)}{\mu_k}\overline{\mathbf{W}}_k\right| = \left|\tfrac{\ln(2)}{\mu_k}\mathbf{W}_{1,k}\right|\left|\tfrac{\ln(2)}{\mu_k}\left(\mathbf{W}_{1,k}^* - \mathbf{W}_{2,k}^*\mathbf{W}_{1,k}^{-1}\mathbf{W}_{2,k}\right)\right|.
\tag{5.38}
$$

The first-order optimality conditions for $\mathbf{W}_{1,k}$ and $\mathbf{W}_{2,k}$ lead to:

$$
\frac{d\mathcal{L}}{d\mathbf{W}_{1,k}^*} \;=\; \tfrac{1}{2}\mathbf{E}_k^* - \tfrac{\mu_k}{2\ln(2)}\left(\mathbf{W}_{1,k}^* - \mathbf{W}_{2,k}^*\mathbf{W}_{1,k}^{-1}\mathbf{W}_{2,k}\right)^{-1} = \mathbf{0},
\tag{5.39}
$$

$$
\frac{d\mathcal{L}}{d\mathbf{W}_{2,k}^*} \;=\; \tfrac{1}{2}\tilde{\mathbf{E}}_k + \tfrac{\mu_k}{2\ln(2)}\mathbf{W}_{1,k}^{-1}\mathbf{W}_{2,k}\left(\mathbf{W}_{1,k}^* - \mathbf{W}_{2,k}^*\mathbf{W}_{1,k}^{-1}\mathbf{W}_{2,k}\right)^{-1} = \mathbf{0},
\tag{5.40}
$$

By identifying the expression in (5.39) inside (5.40) we obtain: $\tilde{\mathbf{E}}_k = -\mathbf{W}_{1,k}^{-1}\mathbf{W}_{2,k}\mathbf{E}_k^*$, such that: $\mathbf{W}_{2,k} = -\mathbf{W}_{1,k}\tilde{\mathbf{E}}_k\mathbf{E}_k^{-*}$. Then, by including this relation into (5.39) we get a closed-form expression for $\mathbf{W}_{1,k}$ as a function of $\{\mathbf{T}_{1,k},\mathbf{T}_{2,k}\}$, $\{\mathbf{R}_{1,k},\mathbf{R}_{2,k}\}$. Similarly, by including $\mathbf{W}_{1,k} = -\mathbf{W}_{2,k}\mathbf{E}_k^*\tilde{\mathbf{E}}_k^{-1}$ into (5.39) we obtain a closed-form expression for $\mathbf{W}_{2,k}$ as a function of $\{\mathbf{T}_{1,k},\mathbf{T}_{2,k}\}$, $\{\mathbf{R}_{1,k},\mathbf{R}_{2,k}\}$. The optimal structures for $\mathbf{W}_{1,k}$ and $\mathbf{W}_{2,k}$ given all $\{\mathbf{T}_{1,k},\mathbf{T}_{2,k}\}$, $\{\mathbf{R}_{1,k},\mathbf{R}_{2,k}\}$ are:

$$
\mathbf{W}_{1,k} = \tfrac{\mu_k}{\ln(2)}\left(\mathbf{F}_k\right)^{-1}, \quad \mathbf{W}_{2,k} = -\tfrac{\mu_k}{\ln(2)}\mathbf{E}_k^{-1}\tilde{\mathbf{E}}_k\mathbf{F}_k^{-*},
\tag{5.41}
$$

where $\mathbf{F}_k = \mathbf{E}_k - \tilde{\mathbf{E}}_k\mathbf{E}_k^{-*}\tilde{\mathbf{E}}_k^*$.

Finally, by plugging the optimal $\mathbf{W}_{1,k}$ and $\mathbf{W}_{2,k}$ in (5.41), $\forall k$, into the minWMSE problem ($\mathrm{P}_{5,3}$) in (5.17), and using the determinant relation in (5.38) and the matrix inversion lemma, we have the following equivalent optimization problem:

$$(\mathrm{P}_{5,5}) : \underset{\substack{\{\mathbf{T}_{1,k},\mathbf{T}_{2,k}\} \\ \{\mathbf{R}_{1,k},\mathbf{R}_{2,k}\}}}{\text{minimize}} \quad \sum_{k\in\mathcal{K}} \tfrac{\mu_k}{2} \log_2 |\mathbf{E}_k\mathbf{F}_k^*| \tag{5.42}$$

$$\text{subject to} \quad \begin{cases} \mathrm{Tr}(\mathbf{T}_{1,k}\mathbf{T}_{1,k}^H + \mathbf{T}_{2,k}\mathbf{T}_{2,k}^H) \leq P_k^{\max} \quad \forall k \\ \mathbf{T}_{2,k} = \mathbf{0} \quad \forall k \in \mathcal{S}^{\mathrm{LP}} \end{cases}$$

which is equal to ($\mathrm{P}_{5,2}$) in (5.16) and completes the proof. This means that ($\mathrm{P}_{5,2}$) in (5.16) and ($\mathrm{P}_{5,3}$) in (5.17) have the same global optimal solution $\{\mathbf{T}_{1,k}^\star, \mathbf{T}_{2,k}^\star, \mathbf{R}_{1,k}^\star, \mathbf{R}_{2,k}^\star\}$. Further, if $\{\mathbf{T}_{1,k}^\star, \mathbf{T}_{2,k}^\star, \mathbf{R}_{1,k}^\star, \mathbf{R}_{2,k}^\star, \mathbf{W}_{1,k}^\star, \mathbf{W}_{2,k}^\star\}$ is a stationary point of ($\mathrm{P}_{5,3}$) then $\{\mathbf{T}_{1,k}^\star, \mathbf{T}_{2,k}^\star, \mathbf{R}_{1,k}^\star, \mathbf{R}_{2,k}^\star\}$ is a stationary point of ($\mathrm{P}_{5,2}$) as they satisfy the first-order optimality conditions of both problems. The converse also holds (i.e. if $\{\mathbf{T}_{1,k}^\star, \mathbf{T}_{2,k}^\star, \mathbf{R}_{1,k}^\star, \mathbf{R}_{2,k}^\star\}$ is a stationary point of ($\mathrm{P}_{5,2}$) then $\{\mathbf{T}_{1,k}^\star, \mathbf{T}_{2,k}^\star, \mathbf{R}_{1,k}^\star, \mathbf{R}_{2,k}^\star, \mathbf{W}_{1,k}^\star, \mathbf{W}_{2,k}^\star\}$ is a stationary point of ($\mathrm{P}_{5,3}$)), being $\{\mathbf{W}_{1,k}^\star, \mathbf{W}_{2,k}^\star\}$ the ones in (5.41), because for given $\{\mathbf{T}_{1,k}^\star, \mathbf{T}_{2,k}^\star, \mathbf{R}_{1,k}^\star, \mathbf{R}_{2,k}^\star\}$ then the optimal weighting matrices $\{\mathbf{W}_{1,k}^\star, \mathbf{W}_{2,k}^\star\}$ are those in (5.41).

To give more details about how we arrive at (5.42), let us show it for a specific receiver. First, by using the matrix inversion lemma, it can be shown that when including the structures of $\mathbf{W}_{1,k}$ and $\mathbf{W}_{2,k}$ in (5.41) into $\frac{1}{2} \mathrm{Tr}\left(\mathbf{W}_{1,k}\mathbf{E}_k + \mathbf{W}_{1,k}^*\mathbf{E}_k^* + \mathbf{W}_{2,k}^*\tilde{\mathbf{E}}_k + \mathbf{W}_{2,k}\tilde{\mathbf{E}}_k^*\right)$ in (5.17), then this term is a constant and can be taken out of the optimization. So let us focus on developing the term $-\frac{\mu_k}{2} \log_2 \left|\frac{\ln(2)}{\mu_k}\overline{\mathbf{W}}_k\right|$ in (5.17) with the structures of $\mathbf{W}_{1,k}$ and $\mathbf{W}_{2,k}$ in (5.41). By using the determinant relation in (5.38) we have:

$$
\begin{aligned}
-\tfrac{\mu_k}{2} \log_2 \left|\tfrac{\ln(2)}{\mu_k}\overline{\mathbf{W}}_k\right| &= -\tfrac{\mu_k}{2} \log_2 \left|\mathbf{F}_k^{-1}\left(\mathbf{F}_k^{-*} - \mathbf{E}_k^{-*}\tilde{\mathbf{E}}_k^*\mathbf{E}_k^{-1}\tilde{\mathbf{E}}_k\mathbf{F}_k^{-*}\right)\right| \\
&= -\tfrac{\mu_k}{2} \log_2 \left|\mathbf{F}_k^{-1}\mathbf{E}_k^{-*}\left(\mathbf{E}_k^* - \tilde{\mathbf{E}}_k^*\mathbf{E}_k^{-1}\tilde{\mathbf{E}}_k\right)\mathbf{F}_k^{-*}\right| = -\tfrac{\mu_k}{2} \log_2 \left|\mathbf{F}_k^{-1}\mathbf{E}_k^{-*}\right| \\
&= -\tfrac{\mu_k}{2} \log_2 \left|\mathbf{F}_k^{-*}\mathbf{E}_k^{-1}\right| = \tfrac{\mu_k}{2} \log_2 |\mathbf{E}_k\mathbf{F}_k^*|, \tag{5.43}
\end{aligned}
$$

such that the expression in (5.42) is obtained.

5.B   Proof of Proposition 5.1

The optimal transmit filters, when keeping the remaining sets of variables fixed, are obtained by equaling the derivatives with respect to $\mathbf{T}_{1,k}^*$ and $\mathbf{T}_{2,k}^*$ of the Lagrangian function of the minWMSE problem ($\mathrm{P}_{5,3}$) in (5.17) to zero. The Lagrangian function ($\mathcal{L}$) is shown in (5.37). The derivatives result:

$$\frac{d\mathcal{L}}{d\mathbf{T}_{1,k}^*} = -\mathbf{H}_{k,k}^H\mathbf{R}_{1,k}\mathbf{W}_{1,k} - \mathbf{H}_{k,k}^H\mathbf{R}_{2,k}^*\mathbf{W}_{2,k}^* + (\mathbf{B}_k+\mathbf{D}_k)\mathbf{T}_{2,k}^* + (\mathbf{A}_k + \mathbf{C}_k + \lambda_k\mathbf{I})\mathbf{T}_{1,k} = \mathbf{0}, \tag{5.44}$$

$$\frac{d\mathcal{L}}{d\mathbf{T}_{2,k}^*} = -\mathbf{H}_{k,k}^H\mathbf{R}_{2,k}^*\mathbf{W}_{1,k}^* - \mathbf{H}_{k,k}^H\mathbf{R}_{1,k}\mathbf{W}_{2,k} + (\mathbf{B}_k+\mathbf{D}_k)\mathbf{T}_{1,k}^* + (\mathbf{A}_k+\mathbf{C}_k+\lambda_k\mathbf{I})\mathbf{T}_{2,k} = \mathbf{0}, \tag{5.45}$$

where $\mathbf{A}_k$, $\mathbf{B}_k$, $\mathbf{C}_k$, and $\mathbf{D}_k$, are defined in (5.21), (5.23), (5.25), and (5.27), respectively.

This way, if we isolate $\mathbf{T}_{1,k}$ from (5.44) and substitute it to (5.45), we get an expression for $\mathbf{T}_{2,k}$ only as a function of $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$, $\{\mathbf{W}_{1,k}, \mathbf{W}_{2,k}\}$. Similarly, if we isolate $\mathbf{T}_{2,k}$ from (5.45) and substitute it to (5.44), we obtain $\mathbf{T}_{1,k}$ as a function of $\{\mathbf{R}_{1,k}, \mathbf{R}_{2,k}\}$, $\{\mathbf{W}_{1,k}, \mathbf{W}_{2,k}\}$. The derived expressions for $\mathbf{T}_{1,k}$ and $\mathbf{T}_{2,k}$ are those shown in (5.29)-(5.30).

In case the k-th transmitter is constrained to use LP, the solution for the transmit filter $\mathbf{T}_{1,k}$ is directly obtained from (5.44): $\mathbf{T}_{1,k} = (\mathbf{A}_k + \mathbf{C}_k + \lambda_k \mathbf{I})^{-1} \left(\mathbf{H}_{k,k}^H \mathbf{R}_{1,k} \mathbf{W}_{1,k} + \mathbf{H}_{k,k}^H \mathbf{R}_{2,k}^* \mathbf{W}_{2,k}^* \right)$.

Chapter 6

Short-Term Traffic-Aware Resource Management

This chapter investigates flexible duplex techniques for short-term traffic-aware interference management. We focus on 5G TDD dense networks, where new short-length single-direction frame structures are envisioned. Thus, a single transmit direction (i.e. either downlink (DL) or uplink (UL)) can be independently chosen at each cell in every frame. This provides high flexibility to match the per-cell DL/UL traffic asymmetries and hence allows full exploitation of dynamic TDD. As a downside, interference coordination becomes crucial. In this sense, this chapter proposes a joint user scheduling, precoding design and transmit direction selection procedure for dynamic TDD in dense MIMO smallcell networks (SCNs), where the transmit direction selected per BS is dynamically optimized together with the user scheduling and transmit precoding. We focus on the maximization of a general utility function that takes into account the DL/UL traffic asymmetries of each user and the interference conditions in the network. Although the problem is non-convex and involves non-continuous constraints due to the user scheduling and transmit direction selection, the problem it is decomposed thanks to the interference-cost concept and then efficiently solved in parallel through semi-closed form expressions for MIMO systems and closed form expressions for SISO systems. Simulation results show significant gains in DL and UL average rates for different traffic asymmetries and network densities as compared to existing schemes for dynamic TDD thanks to the joint optimization of the transmit direction and the conventional allocation policies.

The technical papers related to this topic are:

[J4]   S. Lagen, A. Agustin, J. Vidal, "Joint User Scheduling, Precoder Design and Transmit Direction Selection in MIMO TDD Small Cell Networks", under minor revision at *IEEE Trans. on Wireless Commun.*, Oct. 2016.

[C6]   S. Lagen, A. Agustin, J. Vidal, "Joint User Scheduling and Transmit Direction Selection in 5G TDD Dense Small Cell Networks", *IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun.*, Valencia (Spain), Sep. 2016.

This chapter contains results presented in **[J4]** that address the joint optimization of user scheduling, precoding design and transmit direction selection for MIMO dense networks. **[C6]** shows the potential benefits of a joint user scheduling and transmit direction selection, but particularized for the SISO case.

This chapter is structured as follows. Section 6.1 contains the state of the art for dynamic TDD and presents the new short-length single-direction frame structures envisioned for 5G systems. The main contributions of the chapter are included in Section 6.2. In Section 6.3 the system model is presented. The problem for joint user scheduling, precoding design and transmit direction selection is formulated and decomposed through frames in Section 6.4. Then, in Section 6.5 the per-frame problem is parallelized into multiple subproblems (one per BS), the optimal solution for each subproblem is derived, and the global iterative algorithm is presented. Section 6.6 shows the simulation results. Finally, conclusions are included in Section 6.7.

6.1   State of the Art

Dense MIMO SCNs are considered a key technology for 5G systems as a result of their cost-effectiveness in boosting the area spectral efficiency of cellular networks through densification of the network with SeNBs [6]. SeNBs transmit low power and provide short range coverage, so the expected number of users served per SeNB is reduced. As a consequence, the amount of DL and UL traffic per cell can vary over space and time more drastically in SCNs than in conventional macrocell-based networks [25].

Differently from LTE FDD systems where the amount of band devoted for DL and UL is fixed and equally divided, LTE TDD systems allow for asymmetric DL-UL allocations by providing seven different semi-statically configured UL-DL configurations [61]. The predefined UL-DL configurations differ in the switching points between a DL and an UL transmission within an LTE frame (composed of 10 subframes), hence providing DL-UL allocation ratios that vary from 4:6 to 9:1 (DL:UL). Usually, the UL-DL configuration is the same for all cells and is determined at the network level based on long-term traffic statistics, which might not match the instantaneous per-cell traffic asymmetries.

In this regard, the new emerging **dynamic TDD** technique [19,63] offers the possibility of a dynamic UL-DL reconfiguration so as to adapt the DL-UL allocation ratio to the instantaneous traffic asymmetry at each cell. This higher flexibility is specially suited for dense MIMO SCNs. As a downside, it introduces new types of interference in the system (i.e. DL-to-UL and UL-to-DL interference). So, under these conditions, interference management procedures are key enablers for dynamic TDD.

One may find works in the literature on dynamic TDD, which mainly focus on optimizing the UL-DL configuration ([25,147,148]) or the DL-to-UL switching point decision within an LTE frame ([149–151]) per BS or per group of BSs, in many cases constrained to the frame patterns predefined in LTE TDD [63]. These works can be classified according to where and how such decision is taken. References like [25, 147, 148, 152] consider a centralized and coordinated (i.e. *cluster-specific*) decision, where the deployed BSs are divided into isolated groups of BSs (or clusters) and the same UL-DL configuration is used within the cluster. This way, DL-to-UL and UL-to-DL interference are not created inside the cluster but the flexibility of adapting to the per-BS traffic asymmetries is reduced. On the other hand, decentralized solutions are investigated in [149–151]. In [149], the decision on the DL-to-UL switching point is performed at each BS in coordination with the neighboring BSs thanks to the exchange of backhaul control plane messages (i.e. prices) that take into account the traffic asymmetry of the serving users but also how such decision affects to the users associated to neighboring BSs. In [150], the problem is formulated as a non-cooperative game so as to minimize the overall UL and DL delay in each cell.
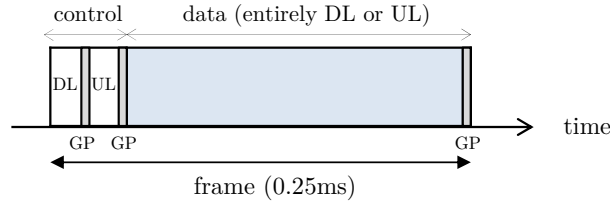
Figure 6.1: Frame structure suggested for dynamic TDD in 5G.

In [151], decentralized and uncoordinated (i.e. *BS-specific*) solutions are evaluated, where each BS performs its own decision based on the traffic asymmetry of the serving users. Results in [151] show that BS-specific decisions are sufficient if interference mitigation techniques are used at reception. A similar conclusion is obtained in [147] when interference management techniques are used at transmission, showing that BS-specific decisions with CoMP-CS/CB achieve better performance than cluster-specific decisions with CoMP-CS-CB.

All these previous works on dynamic TDD keep the order of the transmit directions fixed (i.e. first DL and then UL) and optimize the DL-to-UL switching point or the UL-DL configuration. On the contrary, [153] shows the positive benefits in terms of interference reduction of optimizing also the order in which UL and DL transmissions are performed at each cell. They present a frame structure composed of two consecutive time slots, whereby in the 1st slot the transmission is carried out in one direction and the reverse direction is used for transmission in the 2nd slot (e.g. UL in 1st slot and DL in 2nd slot, or the other way round). In this sense, the optimization of the order of transmit directions is investigated in conjunction with the design of linear precoding and equalization for MIMO systems. However, the main drawbacks of the proposed scheme are that the frame structure is equally partitioned between DL and UL and that traffic asymmetries among cells are not incorporated.

In this sense, new **short-length single-direction frame** structures are envisioned for 5G systems to meet the strict latency requirement of 1ms delay, which cannot be met with the current frame structure in LTE. Low round trip times are required by some envisioned applications for 5G such as tactile Internet. Under the frame structure suggested in [154], which is also proposed in [155], the data part on each frame is assigned to a single transmit direction and each cell can determine if is used either for DL or for UL. Such new frame structure is very powerful for scenarios where the traffic varies drastically (as SCNs) and, due to its single transmit direction per frame, avoids interference variability within the frame. The key point is that, under these new frame structures envisioned for 5G systems, full exploitation of dynamic TDD is possible without resorting to one of the predefined UL-DL configurations in LTE TDD or focusing on the DL-to-UL switching point decision. To that end, new interference management procedures being able to determine the transmit direction in a per-frame basis are needed.

A promising solution in 5G systems to perform interference management in dense deployments is the centralized-based concept, where a network controller is responsible for some network functions [83]. Then, centralized interference management techniques can be performed. However, as the number of BSs increases so does the computation required for centralized interference management. Therefore, parallel interference management techniques are still preferred to eliminate high computation loads in centralized architectures [84].

6.2 Contribution

In this chapter we assume a short-length single-direction frame structure for TDD systems, similar to the one proposed in [154,155] and as shown in Fig. 6.1, where each BS is associated to a transmit direction (DL or UL) at every frame. In this context, we propose a dynamic procedure for joint user scheduling, precoding design and transmit direction selection in interfering MIMO multi-cell scenarios that works on a per-frame basis (see Fig. 6.2). A network controller is in charge of performing the optimization procedure for a set of short-length frames during which the channel conditions do not vary.

We face the maximization of a general utility function that takes into account the DL/UL traffic asymmetries of each user and the interference conditions in the network. First, the initial problem is easily decoupled through frames. Then, the per-frame problem is decomposed per BS by using the interference cost concept and a direct solution is obtained: it is obtained in semi-closed form for MIMO systems and in closed-form for SISO systems. Thus, the per-frame problem can be efficiently solved in a parallel way at the network controller.

6.3 System Model

Consider a synchronized TDD SCN[21] composed of a set of $\mathcal{K}\triangleq\{1,\ldots,K\}$ BSs equipped with M_k antennas each. Every k-th BS ($k\in\mathcal{K}$) has a set of $\mathcal{I}_k\triangleq\{1,\ldots,I_k\}$ associated users with N_{i_k} antennas each. Let i_k denote the i-th user associated to the k-th BS ($i_k\in\mathcal{I}_k$). The total set of users in the system is denoted by $\mathcal{I}=\bigcup_{k\in\mathcal{K}}\mathcal{I}_k$. Subindexes $\{i_k,j_l\}$ and $\{k,l\}$ are used through the chapter to denote users and BSs, respectively. An example of the scenario is shown in Fig. 4.2 for $|\mathcal{K}|=3$ BSs and $|\mathcal{I}|=5$ users ($|\mathcal{I}_1|=2$, $|\mathcal{I}_2|=2$, $|\mathcal{I}_3|=1$).

The short-length frame structure in [154] is assumed. Control and data planes are separated in time, as shown in Fig. 6.1. A short guard period (GP) is inserted between every (possible) switch of the transmit direction. The data part is entirely devoted either for UL or for DL transmission, and the transmit direction in the data part (i.e. DL or UL) can vary at every frame and every BS. As a consequence, we have a totally dynamic TDD system, as illustrated in Fig. 6.2. Through the chapter, let supraindex d denote the transmit direction selected at a BS in a given frame: $d=D$ refers to DL and $d=U$ refers to UL transmission.

Assume that CSI of all users remain constant over a set of short-length frames $\mathcal{S}\triangleq\{1,\ldots,S\}$. Let $\mathbf{H}_{i_k,l}\in\mathbb{C}^{M_l\times N_{i_k}}$, $\mathbf{H}_{l,i_k}\in\mathbb{C}^{N_{i_k}\times M_l}$, $\mathbf{H}_{i_k,j_l}\in\mathbb{C}^{N_{j_l}\times N_{i_k}}$, and $\mathbf{H}_{k,l}\in\mathbb{C}^{M_l\times M_k}$ denote the channel matrix (including pathloss and shadowing) between the i_k-th user and the l-th BS, the l-th BS and the i_k-th user, the i_k-th user and the j_l-th user, and the k-th BS and the l-th BS, respectively.

Assume a centralized architecture, in which there is a network controller (see Fig. 6.2) that has two main functions: *a*) gathers CSI of all users both in DL and UL transmit directions (which are valid for the whole set of short-length frames $\mathcal{S}$) and *b*) is in charge of performing the *dynamic resource management*, which involves the joint optimization of the user scheduling, precoding design and transmit direction selection per frame. So, the time scale is as follows: the optimization is done at each frame, while acquisition of CSI is done every S frames.

It is considered that, at each s-th frame ($s=1,\ldots,S$), at most one user is scheduled in one transmit direction (either DL or UL) at every k-th BS, see Fig. 6.2.

---

[21]Synchronization in time domain among BSs and users is assumed for the considered multi-cell scenario.
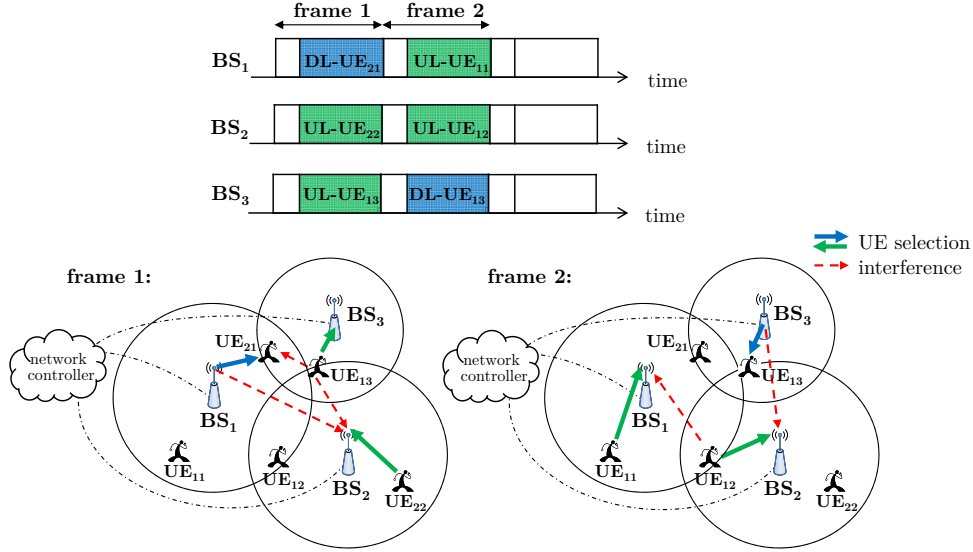
Figure 6.2: Dynamic TDD in SCNs. At each frame every BS schedules one user in one transmit direction (DL or UL). An example of the user scheduling and transmit direction selection is shown for frame 1 and frame 2.

Under narrow-band transmissions, the equivalent baseband signal observed at the i_k-th user in DL in the s-th frame (assuming DL is selected at the k-th BS and no other user is served by the k-th BS at that frame) is expressed as:

$$\mathbf{y}_{i_k,s}^D = \mathbf{H}_{k,i_k}\mathbf{x}_{i_k,s}^D + \underbrace{\sum_{l\in\mathcal{K},l\neq k}\sum_{j_l\in\mathcal{I}_l}\mathbf{H}_{l,i_k}\mathbf{x}_{j_l,s}^D}_{\text{DL-to-DL interference}} + \underbrace{\sum_{l\in\mathcal{K},l\neq k}\sum_{j_l\in\mathcal{I}_l}\mathbf{H}_{j_l,i_k}\mathbf{x}_{j_l,s}^U}_{\text{UL-to-DL interference}} + \mathbf{v}_{i_k,s}^D, \tag{6.1}$$

where $\mathbf{x}_{i_k,s}^D\in\mathbb{C}^{M_k\times 1}$, $\mathbf{x}_{j_l,s}^D\in\mathbb{C}^{M_l\times 1}$, $\mathbf{x}_{j_l,s}^U\in\mathbb{C}^{N_{j_l}\times 1}$ denote the transmitted signal at the k-th BS in DL towards the i_k-th user, the l-th BS in DL towards the j_l-th user, and the j_l-th user in UL towards the l-th BS, respectively, and $\mathbf{v}_{i_k,s}^D$ refers to the received noise vector at the i_k-th user in DL in the s-th frame. Similarly, the equivalent baseband signal observed at the k-th BS in UL in the s-th frame (assuming UL is selected at the k-th BS and only the i_k-th user transmits) is:

$$\mathbf{y}_{i_k,s}^U = \mathbf{H}_{i_k,k}\mathbf{x}_{i_k,s}^U + \underbrace{\sum_{l\in\mathcal{K},l\neq k}\sum_{j_l\in\mathcal{I}_l}\mathbf{H}_{l,k}\mathbf{x}_{j_l,s}^D}_{\text{DL-to-UL interference}} + \underbrace{\sum_{l\in\mathcal{K},l\neq k}\sum_{j_l\in\mathcal{I}_l}\mathbf{H}_{j_l,k}\mathbf{x}_{j_l,s}^U}_{\text{UL-to-UL interference}} + \mathbf{v}_{k,s}^U, \tag{6.2}$$

where $\mathbf{x}_{i_k,s}^U\in\mathbb{C}^{N_{i_k}\times 1}$ denotes the transmitted signal at the i_k-th user in UL towards the k-th BS and $\mathbf{v}_{k,s}^U$ refers to the received noise vector at the k-th BS in UL in the s-th frame. It is assumed that $\mathbf{v}_{k,s}^U\sim\mathcal{CN}(0,\sigma_k^2)$, $\mathbf{v}_{i_k,s}^D\sim\mathcal{CN}(0,\sigma_{i_k}^2)$, $\forall s$, and that all $\{\mathbf{v}_{k,s}^U\}$ and $\{\mathbf{v}_{i_k,s}^D\}$ are independent.

We assume independent encoding across different BSs and users and that a Gaussian codebook is used at each BS and user, i.e.:

$$\mathbf{x}_{i_k,s}^D \sim \mathcal{CN}(\mathbf{0},\mathbf{Q}_{i_k,s}^D), \quad \mathbf{x}_{i_k,s}^U \sim \mathcal{CN}(\mathbf{0},\mathbf{Q}_{i_k,s}^U), \tag{6.3}$$

where $\mathbf{Q}_{i_k,s}^D\succeq\mathbf{0}\in\mathbb{C}^{M_k\times M_k}$ denotes the transmit covariance matrix for the k-th BS to serve the i_k-th user in DL in the s-th frame, and $\mathbf{Q}_{i_k,s}^U\succeq\mathbf{0}\in\mathbb{C}^{N_{i_k}\times N_{i_k}}$ refers to the transmit covariance matrix for the i_k-th user to transmit towards the k-th BS in UL in the s-th frame.

The optimization variables are the transmit covariance matrices $\{\mathbf{Q}^D_{i_k,s}, \mathbf{Q}^U_{i_k,s}\}_{\forall i_k,k,s}$, which contain information of the precoding design (including power control), user scheduling, and transmit direction selection, as follows:

- the precoder design is directly obtained through the eigenvalue decomposition of the transmit covariances matrices, i.e. $\mathbf{Q}^d_{i_k,s} = \mathbf{U}^d_{i_k,s}\mathbf{P}^d_{i_k,s}(\mathbf{U}^d_{i_k,s})^{-1}$, where $\mathbf{U}^d_{i_k,s}$ is the unitary transmit precoding matrix.

- the power employed by the k-th BS in DL to serve the i_k-th user in the s-th frame is given by: $\mathrm{Tr}\left(\mathbf{Q}^D_{i_k,s}\right) = \mathrm{Tr}\left(\mathbf{P}^D_{i_k,s}\right)$, and the power used by the i_k-th user to transmit in UL towards the k-th BS in the s-th frame is: $\mathrm{Tr}\left(\mathbf{Q}^U_{i_k,s}\right) = \mathrm{Tr}\left(\mathbf{P}^U_{i_k,s}\right)$.

- the user scheduling and transmit direction selection are contained in $\{\mathbf{Q}^D_{i_k,s}, \mathbf{Q}^U_{i_k,s}\}_{\forall i_k,k,s}$ as detailed next. If $\mathrm{Tr}\left(\mathbf{Q}^d_{i_k,s}\right) > 0$, then the i_k-th user is scheduled and the d-th transmit direction is selected at the k-th BS in the s-th frame, while if $\mathrm{Tr}\left(\mathbf{Q}^d_{i_k,s}\right) = 0$ then the i_k-th user is not scheduled in the d-th transmit direction at the k-th BS in the s-th frame.

Said intrinsic information is valid if we impose certain constraints over the transmit precoding matrices $\{\mathbf{Q}^D_{i_k,s}, \mathbf{Q}^U_{i_k,s}\}_{\forall i_k,k,s}$, as shown in what follows.

The constraint that at most one user in one transmit direction is selected per BS and per frame can be imposed directly over optimization variables $\{\mathbf{Q}^D_{i_k,s}, \mathbf{Q}^U_{i_k,s}\}_{\forall i_k,k,s}$ as:

$$\sum_{i_k \in \mathcal{I}_k} \left(\mathbb{1}\{\mathbf{Q}^D_{i_k,s}\} + \mathbb{1}\{\mathbf{Q}^U_{i_k,s}\}\right) \leq 1 \quad \forall k, s, \tag{6.4}$$

where $\mathbb{1}\{\mathbf{Q}^d_{i_k,s}\}$ denotes the indicator function over transmit covariance matrix $\mathbf{Q}^d_{i_k,s}$:

$$\mathbb{1}\{\mathbf{Q}^d_{i_k,s}\} = \begin{cases} 1 & \text{if } \mathbf{Q}^d_{i_k,s} \neq \mathbf{0} \\ 0 & \text{if } \mathbf{Q}^d_{i_k,s} = \mathbf{0} \end{cases}. \tag{6.5}$$

Although the constraints in (6.4) are non-continuous and non-convex w.r.t. $\{\mathbf{Q}^D_{i_k,s}, \mathbf{Q}^U_{i_k,s}\}_{\forall i_k,k,s}$, said constraints do not couple frames and BSs. Constraint in (6.4) is required to be consistent with the formulation presented in (6.1)-(6.2). Otherwise, in case multiple users could be simultaneously served at a frame and within the same frequency resource, intra-cell interference should be considered in the signal model.

Usually, in the literature, in addition to optimizing precoding matrices, designers need to deal with integer variables that indicate the user scheduling and transmit direction selection (see [153]). Differently, thanks to the proposed formulation, a single set of variables (i.e. $\{\mathbf{Q}^D_{i_k,s}, \mathbf{Q}^U_{i_k,s}\}_{\forall i_k,k,s}$) need to be optimized.

Under this setting, assuming that interference is treated as Gaussian noise at receivers, the achievable rates in DL and UL transmissions, respectively, of the i_k-th user in the s-th frame, $R^D_{i_k,s}$, $R^U_{i_k,s}$, are given by:

$$R^D_{i_k,s} = \log_2\left|\mathbf{I} + \mathbf{H}_{k,i_k}\mathbf{Q}^D_{i_k,s}\mathbf{H}^H_{k,i_k}\left(\mathbf{N}^D_{i_k,s}\right)^{-1}\right|, \quad R^U_{i_k,s} = \log_2\left|\mathbf{I} + \mathbf{H}_{i_k,k}\mathbf{Q}^U_{i_k,s}\mathbf{H}^H_{i_k,k}\left(\mathbf{N}^U_{i_k,s}\right)^{-1}\right|, \tag{6.6}$$

where $\mathbf{N}^D_{i_k,s}$ is the covariance matrix of the noise-plus-interference received in DL at the i_k-th user when the k-th BS transmits at the s-th frame and $\mathbf{N}^U_{i_k,s}$ denotes the covariance matrix of the noise-plus-interference received in UL at the k-th BS when the i_k-th user transmits at the

s-th frame:

$$\mathbf{N}_{i_k,s}^{D} = \sigma_{i_k}^{2} \mathbf{I} + \sum_{l\in\mathcal{K},l\neq k} \sum_{j_l\in\mathcal{I}_l} \mathbf{H}_{l,i_k} \mathbf{Q}_{j_l,s}^{D} \mathbf{H}_{l,i_k}^{H} + \mathbf{H}_{j_l,i_k} \mathbf{Q}_{j_l,s}^{U} \mathbf{H}_{j_l,i_k}^{H}, \tag{6.7}$$

$$\mathbf{N}_{i_k,s}^{U} = \sigma_{k}^{2} \mathbf{I} + \sum_{l\in\mathcal{K},l\neq k} \sum_{j_l\in\mathcal{I}_l} \mathbf{H}_{l,k} \mathbf{Q}_{j_l,s}^{D} \mathbf{H}_{l,k}^{H} + \mathbf{H}_{j_l,k} \mathbf{Q}_{j_l,s}^{U} \mathbf{H}_{j_l,k}^{H}. \tag{6.8}$$

Therefore, as it can be observed in (6.7)-(6.8), interference received in DL and UL depend on the user scheduling, precoding design, and transmit direction selected at neighbor BSs in the *s*-th frame.

6.4   Problem Formulation

The problem for joint user scheduling, precoding design and transmit direction selection is formulated by following the maximization of a general utility function that takes into account the traffic asymmetries and interference conditions in the network under the constraints that at most one user in one transmit direction is selected at each BS in every frame (see (6.4)) and that a maximum power is available for DL and UL transmission at BSs and users, respectively:

$$(\mathrm{P}_{6,1}): \quad \underset{\{\mathbf{Q}_{i_k,s}^{d} \succeq \mathbf{0}\}_{\forall i_k,k,s,d}}{\operatorname{maximize}} \sum_{k\in\mathcal{K}} \sum_{i_k\in\mathcal{I}_k} \left(a_{i_k} u\left(\bar{R}_{i_k}^{D}\right) + (1-a_{i_k}) u\left(\bar{R}_{i_k}^{U}\right) \right) \tag{6.9}$$

$$\text{subject to} \quad \begin{cases} \sum_{i_k\in\mathcal{I}_k} \left(\mathbb{1}\{\mathbf{Q}_{i_k,s}^{D}\} + \mathbb{1}\{\mathbf{Q}_{i_k,s}^{U}\} \right) \leq 1 \ \forall k,s \\ \mathrm{Tr}\left(\mathbf{Q}_{i_k,s}^{D}\right) \leq P_{\mathrm{BS}}^{\max} \quad \forall i_k,k,s \\ \mathrm{Tr}\left(\mathbf{Q}_{i_k,s}^{U}\right) \leq P_{\mathrm{UE}}^{\max} \quad \forall i_k,k,s \end{cases}$$

where $0 \leq a_{i_k} \leq 1$ is related to the DL-UL data traffic asymmetry of the i_k-th user, $u(z)$ is a concave and monotonically increasing function on the interval $z \in [0,\infty)$, $P_{\mathrm{BS}}^{\max}$ is the available power at BSs, $P_{\mathrm{UE}}^{\max}$ is the available power at users, and $\bar{R}_{i_k}^{d}$ is the average rate of the i_k-th user in the *d*-th transmit direction ($d = \{D, U\}$) over the set of frames $\mathcal{S}$:

$$\bar{R}_{i_k}^{d} = \frac{1}{|\mathcal{S}|} \sum_{s\in\mathcal{S}} R_{i_k,s}^{d}, \tag{6.10}$$

being $R_{i_k,s}^{d}$ the achievable rate shown in (6.6). A classical choice for the utility function in (6.9) is $u(z) = \ln(z)$, which guarantees proportional fairness in the system [156]. Even though, alternative choices that are suitable for different types of applications can be found in [157] (e.g. $u(z) = z$ for sum rate maximization or $u(z) = -1/z$ for harmonic mean fairness).

Practical scheduling algorithms welcome on-line solutions that are capable of adapting weights to obtain desirable DL/UL user fairness [158]. In this sense, and without loss of generality, the average rate of the i_k-th user in the *d*-th transmit direction can be updated at every frame using standard stochastic approximation recursions (see [159]):

$$\bar{R}_{i_k,s+1}^{d} = \bar{R}_{i_k,s}^{d} + \alpha_s \left(R_{i_k,s}^{d} - \bar{R}_{i_k,s}^{d} \right), \tag{6.11}$$

where α_s is a step-size that can be either asymptotically vanishing (e.g. $\alpha_s = 1/s$) or constant ($\alpha_s = 1/|\mathcal{S}|$), $\bar{R}_{i_k,s}^{d}$ is the average rate of the i_k-th user in the *d*-th transmit direction in previous

frames, and $R^d_{i_k,s}$ is the rate given to the i_k-th user in the d-th transmit direction at the s-th frame (see (6.6)). Using Taylor's expansion with step-size α_s sufficiently small over the on-line averaging performed in the recursion in (6.11), we have [158]:

$$u\big(\bar{R}^d_{i_k,s+1}\big) \approx u\big(\bar{R}^d_{i_k,s}\big) + \frac{\delta u\big(\bar{R}^d_{i_k,s}\big)}{\delta \bar{R}^d_{i_k,s}}\alpha_s\big(R^d_{i_k,s} - \bar{R}^d_{i_k,s}\big). \tag{6.12}$$

Therefore, since $\bar{R}^d_{i_k,s}$, $u(\bar{R}^d_{i_k,s})$, and $\frac{\delta u(\bar{R}^d_{i_k,s})}{\delta \bar{R}^d_{i_k,s}}$ are available at the s-th frame, maximizing $(P_{6,1})$ in (6.9) with $u(\bar{R}^d_{i_k,s+1})$ in (6.12) instead of $u(\bar{R}^d_{i_k})$ in (6.9) reduces to the following maximization problem that has to be solved for every frame and corresponds to a weighted sum of the achievable rates. The on-line solution for the s-th frame is obtained from:

$$(P^s_{6,2}) : \quad \underset{\{\mathbf{Q}^d_{i_k,s}\succeq\mathbf{0}\}_{\forall i_k,k,d}}{\text{maximize}} \sum_{k\in\mathcal{K}} \sum_{i_k\in\mathcal{I}_k} \big(\mu^D_{i_k,s}R^D_{i_k,s}+\mu^U_{i_k,s}R^U_{i_k,s}\big) \tag{6.13}$$

$$\text{subject to} \quad \begin{cases} \displaystyle\sum_{i_k\in\mathcal{I}_k} \big(\mathbb{1}\{\mathbf{Q}^D_{i_k,s}\} + \mathbb{1}\{\mathbf{Q}^U_{i_k,s}\}\big) \leq 1 \ \forall k \\ \mathrm{Tr}\big(\mathbf{Q}^D_{i_k,s}\big) \leq P^{\max}_{\mathrm{BS}} \quad \forall i_k, k \\ \mathrm{Tr}\big(\mathbf{Q}^U_{i_k,s}\big) \leq P^{\max}_{\mathrm{UE}} \quad \forall i_k, k \end{cases}$$

where $\mu^d_{i_k,s}$ is a fixed weight associated to the i_k-th user and d-th transmit direction in the s-th frame that depends on the traffic asymmetry a_{i_k} (see (6.9)) and on the derivative of the utility function $u(z)$ adopted for problem $(P_{6,1})$ in (6.9) (see (6.12)):

$$\mu^D_{i_k,s} = a_{i_k} \frac{\delta u\big(\bar{R}^D_{i_k,s}\big)}{\delta \bar{R}^D_{i_k,s}}, \quad \mu^U_{i_k,s} = \big(1 - a_{i_k}\big)\frac{\delta u\big(\bar{R}^U_{i_k,s}\big)}{\delta \bar{R}^U_{i_k,s}}. \tag{6.14}$$

For example, in case $u(z)=\ln(z)$, then $\frac{\delta u\big(\bar{R}^d_{i_k,s}\big)}{\delta \bar{R}^d_{i_k,s}}=1/\bar{R}^d_{i_k,s}$ such that a modified version of the well-known proportional fair criterion [156] would be obtained. Note that the modification comes from the inclusion of the traffic asymmetry conditions, see (6.14). However, the formulation of problem $(P_{6,1})$ in (6.9) allows accommodating general utility functions that correspond to different traffic types and quality-of-service requirements (as best effort, non-real-time, and real-time services) whereas only the design of the weights per user and transmit direction $\mu^d_{i_k,s}$ for problem $(P^s_{6,2})$ in (6.13) are affected (see details in [158]).

The objective function of problem $(P^s_{6,2})$ in (6.13) is non-convex due to interference (see (6.6)) and some constraints are non-continuous. Therefore, finding the global optimum is a challenging task. In this regard, Section 6.5 proposes an efficient algorithm that yields a local optimum solution to problem $(P^s_{6,2})$ in (6.13).

After optimizing problem $(P^s_{6,2})$ in (6.13) for a given frame, the weights per user and transmit direction, $\mu^D_{i_k,s+1}$, $\mu^U_{i_k,s+1}$, have to be updated and problem $(P^{s+1}_{6,2})$ in (6.13) can be subsequently solved for the $(s+1)$-th frame. This provides an on-line solution, as the user scheduling, pre-coding design and transmit direction selection at the s-th frame might impact on the weights used in the subsequent frame $(s+1)$. Note that the whole set of per-frame problems $(P^s_{6,2})$ in (6.13), $\forall s\in\mathcal{S}$, can be solved at every frame or all together at the beginning of the set of frames if channel coherence time allows it.
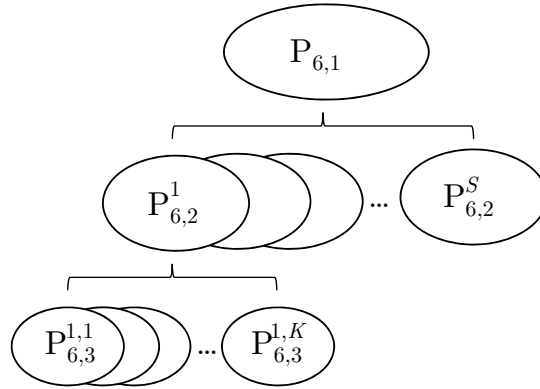
Figure 6.3: Decoupling of the global problem ($P_{6,1}$) in (6.9) into S problems, one per frame ($P_{6,2}^s$) in (6.13). Then, problem ($P_{6,2}^s$) is decomposed into multiple subproblems, one per BS ($P_{6,3}^{s,k}$) in (6.20), to be solved in parallel in the s-th frame at the network controller.

The sequence of average rates $\bar{R}_{i_k,s}^d$ obtained with the on-line solution converges as $s \to \infty$ to the sequence of average rates $\bar{R}_{i_k}^d$ that solves problem ($P_{6,1}$) in (6.9), see [160] and [161]. When the number of frames is finite, convergence of both sequences to the same solution cannot be guaranteed. Anyway, on-line solutions are desirable for practical scheduling algorithms. So, from now on, we concentrate on solving the per-frame problem ($P_{6,2}^s$) in (6.13).

6.5 Joint User Scheduling, Precoding Design and Transmit Direction Selection

In this section we propose an algorithm to solve the per-frame problem ($P_{6,2}^s$) in (6.13) and hence obtaining the user scheduling, precoding design and transmit direction selection for each BS at the s-th frame. Although the per-frame problem($P_{6,2}^s$) in (6.13) is non-convex, we exploit decomposition techniques and the interference cost concept presented in [133] to solve it. To do so, the per-frame problem ($P_{6,2}^s$) in (6.13) is decomposed into $|\mathcal{K}|$ subproblems (one subproblem per BS) and the solution is obtained by solving them iteratively at the cluster controller. The subproblems are obtained by replacing the interference that renders the objective function of problem ($P_{6,2}^s$) in (6.13) non-convex by linear approximations, as it is done in [133, 162, 163]. Fig. 6.3 shows the whole decomposition.

6.5.1 Decomposition

As the constraints of problem ($P_{6,2}^s$) in (6.13) can be directly decomposed per BS, we focus on the objective function. To isolate the interference that renders the objective function of problem ($P_{6,2}^s$) in (6.13) non-convex, define the weighted sum-rate of the links different than those related to the k-th BS as:

$$f_{-k}\big(\mathbf{Q}_{k,s}, \mathbf{Q}_{-k,s}\big) = \sum_{l \neq k, j_l \in \mathcal{I}_l} \Big(\mu_{j_l,s}^D R_{j_l,s}^D + \mu_{j_l,s}^U R_{j_l,s}^U \Big), \tag{6.15}$$

where $\mathbf{Q}_{k,s}$ refers to the set of transmit covariance matrices of all users and transmit directions related to the k-th BS and $\mathbf{Q}_{-k,s}$ denotes the set of transmit covariance matrices of all users

and transmit directions not related to the k-th BS, i.e. $\mathbf{Q}_{k,s} \triangleq \{\mathbf{Q}_{1_k,s}^D, \ldots, \mathbf{Q}_{I_k,s}^D, \mathbf{Q}_{1_k,s}^U, \ldots, \mathbf{Q}_{I_k,s}^U\}$ and $\mathbf{Q}_{-k,s} \triangleq \{\mathbf{Q}_{1,s}, \ldots, \mathbf{Q}_{k-1,s}, \mathbf{Q}_{k+1,s}, \ldots, \mathbf{Q}_{K,s}\}$.

Using first order Taylor's expansion of $f_{-k}(\mathbf{Q}_{k,s}, \mathbf{Q}_{-k,s})$ in (6.15) around $\bar{\mathbf{Q}}_{k,s}$ (i.e. $\bar{\mathbf{Q}}_{i_k,s}^D$ and $\bar{\mathbf{Q}}_{i_k,s}^U$, $\forall i_k \in \mathcal{I}_k$), we have:

$$
f_{-k}(\mathbf{Q}_{k,s}, \mathbf{Q}_{-k,s}) \approx f_{-k}(\bar{\mathbf{Q}}_{k,s}, \mathbf{Q}_{-k,s}) + \sum_{i_k \in \mathcal{I}_k} \mathrm{Tr}\left((\mathbf{Q}_{i_k,s}^D - \bar{\mathbf{Q}}_{i_k,s}^D) \left. \frac{\delta f_{-k}(\mathbf{Q}_{k,s}, \mathbf{Q}_{-k,s})}{\delta \mathbf{Q}_{i_k,s}^D} \right|_{\bar{\mathbf{Q}}_{i_k,s}^D} \right)
$$
$$
+ \sum_{i_k \in \mathcal{I}_k} \mathrm{Tr}\left((\mathbf{Q}_{i_k,s}^U - \bar{\mathbf{Q}}_{i_k,s}^U) \left. \frac{\delta f_{-k}(\mathbf{Q}_{k,s}, \mathbf{Q}_{-k,s})}{\delta \mathbf{Q}_{i_k,s}^U} \right|_{\bar{\mathbf{Q}}_{i_k,s}^U} \right). \tag{6.16}
$$

The negative partial derivatives of $f_{-k}(\mathbf{Q}_{k,s}, \mathbf{Q}_{-k,s})$ in (6.15) with respect to each of the matrices composing the set $\mathbf{Q}_{k,s}$ evaluated at $\bar{\mathbf{Q}}_{k,s}$ (i.e. $\bar{\mathbf{Q}}_{i_k,s}^D$ and $\bar{\mathbf{Q}}_{i_k,s}^U$, $\forall i_k \in \mathcal{I}_k$, see (6.16)) are given by:

$$
\boldsymbol{\Upsilon}_{i_k,s}^D = - \left. \frac{\delta f_{-k}(\mathbf{Q}_{k,s}, \mathbf{Q}_{-k,s})}{\delta \mathbf{Q}_{i_k,s}^D} \right|_{\bar{\mathbf{Q}}_{i_k,s}^D} = - \sum_{\substack{l \neq k \\ j_l \in \mathcal{I}_l}} \left. \left(\mu_{j_l,s}^D \frac{\delta R_{j_l,s}^D}{\delta \mathbf{Q}_{i_k,s}^D} + \mu_{j_l,s}^U \frac{\delta R_{j_l,s}^U}{\delta \mathbf{Q}_{i_k,s}^D} \right) \right|_{\bar{\mathbf{Q}}_{i_k,s}^D}, \tag{6.17}
$$

$$
\boldsymbol{\Upsilon}_{i_k,s}^U = - \left. \frac{\delta f_{-k}(\mathbf{Q}_{k,s}, \mathbf{Q}_{-k,s})}{\delta \mathbf{Q}_{i_k,s}^U} \right|_{\bar{\mathbf{Q}}_{i_k,s}^U} = - \sum_{\substack{l \neq k \\ j_l \in \mathcal{I}_l}} \left. \left(\mu_{j_l,s}^D \frac{\delta R_{j_l,s}^D}{\delta \mathbf{Q}_{i_k,s}^U} + \mu_{j_l,s}^U \frac{\delta R_{j_l,s}^U}{\delta \mathbf{Q}_{i_k,s}^U} \right) \right|_{\bar{\mathbf{Q}}_{i_k,s}^U}, \tag{6.18}
$$

where $\frac{\delta R_{j_l,s}^D}{\delta \mathbf{Q}_{i_k,s}^D}$ denotes a matrix corresponding to the derivative of scalar function $R_{j_l,s}^D$ with respect to transmit covariance matrix $\mathbf{Q}_{i_k,s}^D$. The derivatives can be obtained using the framework developed in [88] for complex-valued matrix differentiation. For example, $\frac{\delta R_{j_l,s}^D}{\delta \mathbf{Q}_{i_k,s}^D}$ and $\frac{\delta R_{j_l,s}^D}{\delta \mathbf{Q}_{i_k,s}^U}$ are given by:

$$
\frac{\delta R_{j_l,s}^D}{\delta \mathbf{Q}_{i_k,s}^D} = \frac{1}{\ln(2)} \mathbf{H}_{k,j_l}^H \mathbf{Z}_{j_l,s}^D \mathbf{H}_{k,j_l}, \qquad \frac{\delta R_{j_l,s}^D}{\delta \mathbf{Q}_{i_k,s}^U} = \frac{1}{\ln(2)} \mathbf{H}_{i_k,j_l}^H \mathbf{Z}_{j_l,s}^D \mathbf{H}_{i_k,j_l}, \tag{6.19}
$$

with

$$
\mathbf{Z}_{j_l,s}^D = \left(\mathbf{N}_{j_l,s}^D + \mathbf{H}_{l,j_l} \mathbf{Q}_{j_l,s}^D \mathbf{H}_{l,j_l}^H \right)^{-1} - \left(\mathbf{N}_{j_l,s}^D \right)^{-1}.
$$

The remaining derivatives $\frac{\delta R_{j_l,s}^U}{\delta \mathbf{Q}_{i_k,s}^D}$ and $\frac{\delta R_{j_l,s}^U}{\delta \mathbf{Q}_{i_k,s}^U}$ can be similarly obtained, but are omitted for brevity. Matrices $\boldsymbol{\Upsilon}_{i_k,s}^D$ in (6.17) and $\boldsymbol{\Upsilon}_{i_k,s}^U$ in (6.18) are known as cost (or price) matrices because they are related to the impact on neighboring BSs/users of selecting user i_k at BS k in transmit direction $d=D$ and $d=U$, respectively.

Therefore, using Taylor's expansion of $f_{-k}(\mathbf{Q}_{k,s}, \mathbf{Q}_{-k,s})$ in (6.16) and discarding irrelevant constant terms, it is possible to approximate the per-frame problem $(\mathrm{P}_{6,2}^s)$ in (6.13) by a set of $|\mathcal{K}|$ subproblems (one per BS).

The subproblem corresponding to the k-th BS is:

$$(\text{P}_{6,3}^{s,k}): \quad \underset{\{\mathbf{Q}_{i_k,s}^d \succeq \mathbf{0}\}_{\forall i_k \in \mathcal{I}_k, d}}{\text{maximize}} \sum_{i_k \in \mathcal{I}_k} \left(\mu_{i_k,s}^D R_{i_k,s}^D \left(\mathbf{Q}_{i_k,s}^D \right) + \mu_{i_k,s}^U R_{i_k,s}^U \left(\mathbf{Q}_{i_k,s}^U \right) \right)$$

$$- \sum_{i_k \in \mathcal{I}_k} \left(\text{Tr} \left(\mathbf{Q}_{i_k,s}^D \mathbf{\Upsilon}_{i_k,s}^D \right) + \text{Tr} \left(\mathbf{Q}_{i_k,s}^U \mathbf{\Upsilon}_{i_k,s}^U \right) \right) \qquad (6.20)$$

$$\text{subject to} \quad \begin{cases} \sum_{i_k \in \mathcal{I}_k} \left(\mathbb{1}\{\mathbf{Q}_{i_k,s}^D\} + \mathbb{1}\{\mathbf{Q}_{i_k,s}^U\} \right) \leq 1 \\ \text{Tr} \left(\mathbf{Q}_{i_k,s}^D \right) \leq P_{\text{BS}}^{\max} \qquad \forall i_k \in \mathcal{I}_k \\ \text{Tr} \left(\mathbf{Q}_{i_k,s}^U \right) \leq P_{\text{UE}}^{\max} \qquad \forall i_k \in \mathcal{I}_k \end{cases}$$

where $R_{i_k,s}^D(\mathbf{Q}_{i_k,s}^D)$ and $R_{i_k,s}^U(\mathbf{Q}_{i_k,s}^U)$ are the achievable rates in (6.6) for fixed interference-plus-noise covariance matrices $\mathbf{N}_{i_k,s}^D$ and $\mathbf{N}_{i_k,s}^U$. $\bar{\mathbf{Q}}_{k,s}$ is needed to compute $\mathbf{\Upsilon}_{i_k,s}^D, \mathbf{\Upsilon}_{i_k,s}^U, \forall i_k \in \mathcal{I}_k$, in (6.17)-(6.18), so $\bar{\mathbf{Q}}_{k,s}$ can be obtained from the previous iteration. The trace terms in the objective function of (6.20) play the role of an interference tax, discouraging selfish behavior of the i_k-th user in the d-th transmit direction. In case $\mathbf{\Upsilon}_{i_k,s}^d = \mathbf{0}$, the i_k-th user in the d-th transmit direction would otherwise just want to maximize its own achievable rate $R_{i_k,s}^d(\mathbf{Q}_{i_k,s}^d)$.

Note that $R_{i_k,s}^d(\mathbf{Q}_{i_k,s}^d)$ in (6.20) is concave w.r.t. $\mathbf{Q}_{i_k,s}^d$ and $\text{Tr}(\mathbf{Q}_{i_k,s}^d \mathbf{\Upsilon}_{i_k,s}^d)$ is linear. Then, as the objective function of subproblem $(\text{P}_{6,3}^{s,k})$ in (6.20) is separable among users and transmit directions, the objective function of subproblem $(\text{P}_{6,3}^{s,k})$ in (6.20) is concave w.r.t. $\{\mathbf{Q}_{i_k,s}^d\}_{\forall i_k \in \mathcal{I}_k, d}$. Thus, after the whole decomposition steps (see Fig. 6.3), we only have to deal with the non-continuous constraint (i.e. first constraint in (6.20)).

6.5.2 Solution to (6.20)

In case the first constraint in (6.20) was removed, subproblem $(\text{P}_{6,3}^{s,k})$ in (6.20) would be a convex optimization problem and an optimal solution for $\{\mathbf{Q}_{i_k,s}^d\}_{\forall i_k \in \mathcal{I}_k, d}$ would exist. Even though, the first constraint of subproblem $(\text{P}_{6,3}^{s,k})$ in (6.20) only imposes that at most a user in one transmit direction is selected at the k-th BS and the s-th frame. Therefore, as the objective function of subproblem $(\text{P}_{6,3}^{s,k})$ in (6.20) is separable among users and transmit directions, the optimal solution to subproblem $(\text{P}_{6,3}^{s,k})$ in (6.20) (even if non-convex) is either to allocate 0 power to all users and transmit directions (i.e. $\mathbf{Q}_{i_k,s}^d = \mathbf{0}, \forall i_k \in \mathcal{I}_k, d$) or to allocate power only to the user and transmit direction that provides a larger value of its best contribution to the objective function in (6.20). As best contribution of a specific user in a specific transmit direction, we refer to the transmit precoding matrix that maximizes the objective function in (6.20) satisfying the associated power constraint. Note that, as the objective function is concave w.r.t. $\mathbf{Q}_{i_k,s}^d$ and the power constraint is linear, a single best contribution exists. So, we can optimally solve subproblem $(\text{P}_{6,3}^{s,k})$ in (6.20) (even though its non-convexity) through a two-step procedure that:

- **Step 1**: computes all the individual best contributions of the users and transmit directions to the objective function in (6.20) (denoted by $\bar{\mathbf{Q}}_{i_k,s}^d$ in what follows) and then
- **Step 2**: selects the one that contributes more to the objective function in (6.20) (as imposed by the first constraint in (6.20)) and sets the remaining precoding matrices equal to $\mathbf{0}$. This corresponds to the optimal solution to subproblem $(\text{P}_{6,3}^{s,k})$ in (6.20) and is denoted by $\mathbf{Q}_{i_k,s}^{d\star}$ in what follows.

Step 1

The individual best contributions of each user and transmit direction to the objective function of subproblem $(\text{P}_{6,3}^{s,k})$ in (6.20) are obtained by removing the first constraint in (6.20). Focusing on the contribution of the i_k-th user in the d-th transmit direction, we obtain it from the following optimization problem:

$$\underset{\mathbf{Q}_{i_k,s}^d \succeq \mathbf{0}}{\text{maximize}} \quad \mu_{i_k,s}^d \log_2 \left| \mathbf{I} + \mathbf{H}_{i_k}^d \mathbf{Q}_{i_k,s}^d \mathbf{H}_{i_k}^{dH} \left(\mathbf{N}_{i_k,s}^d \right)^{-1} \right| - \text{Tr} \left(\mathbf{Q}_{i_k,s}^d \mathbf{\Upsilon}_{i_k,s}^d \right) \tag{6.21}$$

$$\text{subject to} \quad \text{Tr} \left(\mathbf{Q}_{i_k,s}^d \right) \leq P^{\text{max}}$$

where for the shake of presentation we have unified parts of the specific nomenclature for DL and UL as:

$$\mathbf{H}_{i_k}^d = \begin{cases} \mathbf{H}_{k,i_k} & \text{if } d = D \\ \mathbf{H}_{i_k,k} & \text{if } d = U \end{cases}, \quad P^{\text{max}} = \begin{cases} P_{\text{BS}}^{\text{max}} & \text{if } d = D \\ P_{\text{UE}}^{\text{max}} & \text{if } d = U \end{cases}. \tag{6.22}$$

Problem in (6.21) is a convex optimization problem w.r.t. $\mathbf{Q}_{i_k,s}^d$, and thus it can be solved by the standard Lagrange duality method [92]. The result is included in Proposition 6.1 (and particularized for SISO systems in Proposition 6.1).

Proposition 6.1. *At the k-th BS, the best contribution of the i_k-th user in the d-th transmit direction to the objective function of $(\text{P}_{6,3}^{s,k})$ in (6.20) (i.e. the optimal solution for the transmit covariance matrix $\bar{\mathbf{Q}}_{i_k,s}^d$ in case the i_k-th user in the d-th transmit direction was selected at the k-th BS in the s-th frame) is given by:*

$$\bar{\mathbf{Q}}_{i_k,s}^d = \left(\mathbf{B}_{i_k,s}^d \right)^{-\frac{H}{2}} \mathbf{V}_{i_k,s}^d \mathbf{\Sigma}_{i_k,s}^d \left(\mathbf{V}_{i_k,s}^d \right)^H \left(\mathbf{B}_{i_k,s}^d \right)^{-\frac{1}{2}}, \tag{6.23}$$

where $\mathbf{\Sigma}_{i_k,s}^d = \text{diag}(\sigma_{i_k,s}^d(1), \dots, \sigma_{i_k,s}^d(\tilde{M}))$ is a diagonal matrix of size $\tilde{M} = M_k$ if $d = D$ and $\tilde{M} = N_{i_k}$ if $d = U$, and

$$\mathbf{B}_{i_k,s}^d = \mathbf{\Upsilon}_{i_k,s}^d + \lambda_{i_k,s}^d \mathbf{I}, \tag{6.24}$$

$$\mathbf{V}_{i_k,s}^d \mathbf{Z}_{i_k,s}^d (\mathbf{V}_{i_k,s}^d)^H = \left(\mathbf{B}_{i_k,s}^d \right)^{-\frac{1}{2}} \mathbf{H}_{i_k}^{dH} \left(\mathbf{N}_{i_k,s}^d \right)^{-1} \mathbf{H}_{i_k}^d \left(\mathbf{B}_{i_k,s}^d \right)^{-\frac{H}{2}}, \tag{6.25}$$

$$\sigma_{i_k,s}^d(n) = \left(\frac{\mu_{i_k,s}^d}{\ln(2)} - \frac{1}{z_{i_k,s}^d(n)} \right)^+, \quad n = 1, \dots, \tilde{M}, \tag{6.26}$$

being $\lambda_{i_k,s}^d$ the Lagrange multiplier associated to the power constraint in (6.21). (6.25) denotes the eigenvalue decomposition of the matrix in the right-hand side, being $\mathbf{V}_{i_k,s}^d \in \mathbb{C}^{\tilde{M} \times \tilde{M}}$ a unitary matrix and $\mathbf{Z}_{i_k,s}^d = \text{diag}(z_{i_k,s}^d(1), \dots, z_{i_k,s}^d(\tilde{M}))$ a diagonal matrix.

Proof. See Appendix 6.A. ∎

Therefore, for general MIMO systems, the solution in (6.23) depends on a single parameter, i.e. the Lagrange multiplier $\lambda_{i_k,s}^d$ associated to the transmit power constraint. The optimal value for $\lambda_{i_k,s}^d$ can be efficiently obtained using, for instance, the bisection method or the ellipsoid method [92] (see Appendix 6.A).

For SISO systems (i.e. $M_k=N_{i_k}=1, \forall i_k \in \mathcal{I}_k, \forall k \in \mathcal{K}$), where all parameters are complex scalars and the optimization variables are the transmit power allocated to each user and transmit direction (which is denoted by $P_{i_k,s}^d$), the solution in (6.23) can be obtained in closed-form. The solution is derived from Proposition 6.1 and included in Corollary 6.1. Note also that this solution corresponds to the distributed solution for power allocation in SISO interference channels [164].

Corollary 6.1. *For SISO systems, where all parameters are (complex) scalars and the optimiza-tion variables are the powers allocated to each user and transmit direction (denoted by $P_{i_k,s}^d$); at the k-th BS, the best contribution of the i_k-th user in the d-th transmit direction to the objective function of $(P_{6,3}^{s,k})$ in (6.20) (i.e. the optimal solution for the transmit power $\bar{P}_{i_k,s}^d$ in case the i_k-th user in the d-th transmit direction was selected at the k-th BS in the s-th frame) is given by:*

$$\bar{P}_{i_k,s}^d = \min\left(\left(\frac{\mu_{i_k,s}^d}{\ln(2)\Upsilon_{i_k,s}^d} - \frac{N_{i_k,s}^d}{|H_{i_k}^d|^2}\right)^+, P^{\max}\right). \tag{6.27}$$

Proof. The proof departs from Proposition 6.1 and is included in Appendix 6.B. ∎

The SISO case allows a better understanding of the obtained solution in (6.27). If the cost $\Upsilon_{i_k,s}^d$ tends to infinity (i.e. selecting the i_k-th user in the d-th transmit direction has a detrimental impact in terms of achievable rate over neighboring BSs/users), then the associated variable $\bar{P}_{i_k,s}^d$ tends to 0. On the contrary, if the cost $\Upsilon_{i_k,s}^d$ goes to 0 (i.e. selecting the i_k-th user in the d-th transmit direction has not a detrimental impact in terms of achievable rate over neighboring BSs/users) then the associated variable $\bar{P}_{i_k,s}^d$ is given by the maximum power $(P_{\text{BS}}^{\max}, P_{\text{UE}}^{\max})$. So, the solution in (6.27) is coherent.

Step 2

Once obtained the individual best contribution of each user and transmit direction to the ob-jective function of subproblem $(P_{6,3}^{s,k})$ in (6.20), $\bar{\mathbf{Q}}_{i_k,s}^d, \forall i_k \in \mathcal{I}_k, \forall d$, as at most one user in a single transmit direction can be scheduled per-BS, the optimal solution to subproblem $(P_{6,3}^{s,k})$ in (6.20) corresponds either to select the user $i_k \in \mathcal{I}_k$ and the transmit direction $d \in \{D,U\}$ at the k-th BS and s-th frame providing a larger value of $\mu_{i_k,s}^d R_{i_k,s}^d(\bar{\mathbf{Q}}_{i_k,s}^d) - \text{Tr}(\bar{\mathbf{Q}}_{i_k,s}^d \Upsilon_{i_k,s}^d)$ if the largest value of $\mu_{i_k,s}^d R_{i_k,s}^d(\bar{\mathbf{Q}}_{i_k,s}^d) - \text{Tr}(\bar{\mathbf{Q}}_{i_k,s}^d \Upsilon_{i_k,s}^d)$ is positive or to select none (any user in any transmit direction) in case that the largest value of $\mu_{i_k,s}^d R_{i_k,s}^d(\bar{\mathbf{Q}}_{i_k,s}^d) - \text{Tr}(\bar{\mathbf{Q}}_{i_k,s}^d \Upsilon_{i_k,s}^d)$ is lower or equal to 0 (see objective function of subproblem $(P_{6,3}^{s,k})$ in (6.20)).

The two-step procedure is summarized in Algorithm 6.1. For the k-th BS, we have to check all contributions of its associated users (i.e. search on the set $\mathcal{I}_k$) in two transmit directions (i.e. DL and UL), which renders a search among $2|\mathcal{I}_k|$ possible values. But, let us recall that in SCNs the expected number of users that are associated to each BS is much lower than in conventional macrocell-based networks (i.e. low $|\mathcal{I}_k|$). Therefore, the complexity associated to Algorithm 6.1 is limited and, more important, it is not scaled with the network density as the number of users per BS is reducing with the BS density.

Remark 6.1. *Algorithm 6.1 provides the global optimum solution to subproblem $(P_{6,3}^{s,k})$ in (6.20) for the k-th BS at the s-th frame.*

Algorithm 6.1 Procedure to optimally solve subproblem $(P_{6,3}^{s,k})$ in (6.20) at the k-th BS and the s-th frame

---

1: maxVal $= 0$
2: *# compute best contributions:*
3: **for** $i_k = 1, ..., I_k$ **do**
4:     **for** $d = D, U$ **do**
5:         Compute $\bar{\mathbf{Q}}_{i_k,s}^d$ as in (6.23)
6:         val $= \mu_{i_k,s}^d R_{i_k,s}^d(\bar{\mathbf{Q}}_{i_k,s}^d) - \text{Tr}(\bar{\mathbf{Q}}_{i_k,s}^d \mathbf{\Upsilon}_{i_k,s}^d)$
7:         **if** val $>$ maxVal **then**
8:             user $= i_k$
9:             tx $= d$
10:             maxVal $=$ val
11:         **end if**
12:     **end for**
13: **end for**
14: *# select the user and transmit direction that contributes more to the objective function:*
15: **if** maxVal > 0 **then**
16:     $\mathbf{Q}_{i_k,s}^{d\star} = \bar{\mathbf{Q}}_{i_k,s}^d$, if $i_k =$ user and $d =$ tx
17:     $\mathbf{Q}_{i_k,s}^{d\star} = \mathbf{0}$, otherwise
18: **else**
19:     $\mathbf{Q}_{i_k,s}^{d\star} = \mathbf{0}, \forall i_k \in \mathcal{I}_k, \forall d$
20: **end if**

6.5.3   Algorithm to solve (6.13)

The overall algorithm to solve problem $(P_{6,2}^s)$ in (6.13) essentially solves subproblems $(P_{6,3}^{s,k})$ in (6.20) iteratively until convergence. To that end, the Gauss-Seidel iteration (sequential optimizations), the Jacobi iteration (sequential optimizations), or an entirely asynchronous iteration (simultaneous optimizations) could be adopted, see [165].

Algorithm 6.2 details the procedure to solve problem $(P_{6,2}^s)$ in (6.13) at the s-th frame by following a Gauss-Seidel iteration. It starts from an initialization of the transmit covariance matrices $\{\mathbf{Q}_{i_k,s}^d\}$ that meets constraints in (6.13) (line 1). A suitable initialization is to select (for each BS k) the user $(i_k^\star)$ and transmit direction $(d^\star)$ with largest $\mu_{i_k,s}^d R_{i_k,s}^d$ (no cost) when using the achievable rates in (6.6) as a function of the useful signal power (no interference), maximum power, and precoding matrices that diagonalize the equivalent channel. After that, the iterative algorithm is performed in which subproblems $(P_{6,3}^{s,k})$ in (6.20) are solved sequentially for all BSs. For each k-th BS: *i*) the cost matrices in (6.17)-(6.18) and the interference-plus-noise covariance matrices in (6.7)-(6.8) are computed (lines 5-6), and *ii*) the optimization in Algorithm 6.1 is performed to obtain the transmit covariances matrices (including user scheduling, precoding design and transmit direction selection) for the k-th BS, i.e. $\{\mathbf{Q}_{i_k,s}^{d\star}\}$ (line 8). The procedure is iterated until convergence is reached and provides $\{\mathbf{Q}_{i_k,s}^{d\star}\}$ as output.

Proposition 6.2. *Algorithm 6.2 converges to a limit point satisfying the Karush-Kuhn-Tucker (KKT) conditions of problem* $(P_{6,2}^s)$ *in* (6.13).

Proof. See Appendix 6.C.   ∎

Convergence of Algorithm 6.2 when the Gauss-Seidel iteration is applied can be proved by following similar steps as in [166] since the objective function of problem $(P_{6,2}^s)$ in (6.13) is

---

Algorithm 6.2 Procedure to solve problem $(P_{6,2}^s)$ in (6.13) at the s-th frame

---

1: Initialize $\{\mathbf{Q}_{i_k,s}^d\}, \forall i_k, k, d$
2: **repeat**
3:   **for** $k = 1, \ldots, K$ **do**
4:     *# update variables:*
5:     Compute cost matrices $\mathbf{\Upsilon}_{i_k,s}^d$ in (6.17)-(6.18), $\forall i_k \in \mathcal{I}_k, d$, for given $\{\mathbf{Q}_{j_l,s}^d\}, \forall j_l, l, d$
6:     Compute interference-plus-noise covariance matrices $\mathbf{N}_{i_k,s}^d$ in (6.7)-(6.8), $\forall i_k \in \mathcal{I}_k, d$, for given $\{\mathbf{Q}_{j_l,s}^d\}, \forall j_l, l, d$
7:     *# perform optimization for BS k:*
8:     Solve subproblem $(P_{6,3}^{s,k})$ in (6.20) through **Algorithm 6.1** to select transmit covariance matrices, i.e. $\{\mathbf{Q}_{i_k,s}^{d\star}\}, \forall i_k \in \mathcal{I}_k, d$
9:     Set $\mathbf{Q}_{i_k,s}^d = \mathbf{Q}_{i_k,s}^{d\star}, \forall i_k \in \mathcal{I}_k, d$
10:    **end for**
11: **until** convergence

---

not decreased at every iteration when sequential updates are performed and, in addition, the objective function is bounded from the above.

Recall also that with an entirely asynchronous iteration (simultaneous optimizations) convergence could not be guaranteed because the objective function (weighted sum rate) of problem $(P_{6,2}^s)$ in (6.13) could oscillate when the cost matrices were updated. In that case, some works have shown through simulations that convergence is achieved by performing simultaneous optimizations provided that a memory in the cost matrices is included (for instance, through the use of a low pass filter), see [162, 167].

Remark 6.2. *Algorithm 6.2 is executed at the network controller, which disposes of CSI of all links. Alternatively, it could be performed in a distributed manner at each BS provided that CSI was available at each BS and information exchange was used iteratively among BSs to get the cost matrices $\mathbf{\Upsilon}_{i_k,s}^d$ in (6.17)-(6.18). But, due to the short-length of the frame, distributed implementation might not be feasible as it would entail longer delays.*

Remark 6.3. *The proposed Algorithm 6.2 can be used to determine the user scheduling, precoding design and power control even if the transmit direction per BS is fixed at the s-th frame simply by not optimizing the transmit covariance matrices $\mathbf{Q}_{i_k,s}^d$ of transmit directions not allowed at BSs.*

Remark 6.4. *For SISO systems, the proposed Algorithm 6.2 can be used to determine the user scheduling and transmit direction selection when binary power control is adopted. In this case, the computation of the optimal values for the power allocation in (6.27) is not needed. We should simply replace the optimization rule in line 5 of Algorithm 6.1 by $\bar{P}_{i_k,s}^D = P_{\mathrm{BS}}^{\max}$ and $\bar{P}_{i_k,s}^U = P_{\mathrm{UE}}^{\max}$.*

Remark 6.5. *Under Orthogonal Frequency Division Multiple Access (OFDMA)[22] with a power spectral mask, where users for the same BS are assigned to orthogonal frequency resources and the primary source of interference is inter-cell interference, the proposed Algorithm 6.2 could be employed to determine the user scheduling and precoding design on each frequency subband.*

---

[22]OFDMA is the radio access technology employed in downlink transmission of 3GPP LTE and LTE-A systems [2], as well as in both uplink and downlink of IEEE 802.16m advanced WiMAX [26]. Furthermore, it is one of the major candidates for different use cases of future 5G systems [27].

However, we have to impose that the same transmit direction is adopted for the whole set of subbands on a given frame s at each BS, such that Algorithm 6.2 should be slightly redefined to impose it: each BS should check all subbands in DL and all in UL and then chose the best.

6.5.4   Algorithm to solve (6.9)

Finally, and for completion, Algorithm 6.3 includes the dynamic user scheduling, precoding design and transmit direction selection to solve the global problem ($P_{6,1}$) in (6.9). At every s-th frame, Algorithm 6.2 is executed to design the final $\{\mathbf{Q}_{i_k,s}^{d\star}\}$ (line 3) and, then, the weights $\mu_{i_k,s+1}^d$ to be used in the subsequent frame ($s+1$) are updated (line 6).

Algorithm 6.3 Dynamic procedure for joint user scheduling, precoding design and transmit direction selection, which aims at solving the global problem ($P_{6,1}$) in (6.9)

1: **for** $s = 1, ..., S$ **do**
2:      *# perform user scheduling, precoding design, transmit direction selection:*
3:      Solve the per-frame problem ($P_{6,2}^s$) in (6.13) by performing **Algorithm 6.2** to select $\{\mathbf{Q}_{i_k,s}^{d\star}\}$, $\forall i_k, k, d$
4:      *# update dynamic variables:*
5:      Compute achievable rates $R_{i_k,s}^d$ in (6.6), $\forall i_k, k, d$, with the selected $\{\mathbf{Q}_{i_k,s}^{d\star}\}$
6:      Update equivalent weights $\mu_{i_k,s+1}^d$ in (6.14), $\forall i_k, k, d$
7: **end for**

6.6   Simulation Results

The simulation scenario consists of a synchronized TDD deployment of $|\mathcal{K}|$ outdoor SeNBs (that act as BSs), which are randomly placed within a circular area of 100 m radius with a minimum distance of 40 m among them. The $|\mathcal{K}|$ BSs are managed by the network controller. $|\mathcal{I}_k|$ users are randomly placed around each BS k in a concentric 40 m radius circle. The evaluation methodology including system parameters, propagation characteristics and traffic modeling designed by 3GPP for outdoor pico scenario in [63] is adopted. All BSs/users operate on the same carrier frequency at 2 GHz. Path loss and shadowing models follow specifications in [63] for multi-cell pico scenario. The antenna pattern is omnidirectional and the transmit power is 24 dBm at BS and 23 dBm at user. Noise power is -84 dBm. We assume the number of antennas at BSs and UEs, respectively, is the same: $M_k=M$ and $N_{i_k}=N$, $\forall i_k, \forall k$. Full load traffic model is adopted.

For simulation purposes the same traffic asymmetry (a_{i_k}) is used for all users, $\forall i_k, \forall k$. Different values of the traffic asymmetries (a_{i_k}), network densities ($|\mathcal{K}|$), and user densities ($|\mathcal{I}_k|$) are used for simulations. $|\mathcal{S}|=100$ frames are used.

A proportional fair (PF) criterion for problem ($P_{6,1}$) in (6.9) is adopted, i.e. $u(z)=\ln(z)$. Therefore, the objective function of problem ($P_{6,1}$) in (6.9) corresponds to a weighted sum of the log of average rates:

$$f_{\text{global}} = \sum_{\substack{k\in\mathcal{K} \\ i_k\in\mathcal{I}_k}} \Big(a_{i_k} \ln\big(\bar{R}_{i_k}^D\big) + (1 - a_{i_k}) \ln\big(\bar{R}_{i_k}^U\big)\Big). \tag{6.28}$$

The aim of the present simulations is to show the benefits of jointly optimizing the transmit direction selection and the conventional allocation policies (including user scheduling and precoding design) through the use of the proposed Algorithm 6.2 as compared to schemes where the transmit direction is set at each BS according to the traffic asymmetries of the associated users while the conventional allocation policies are coordinated among BSs.

Different MIMO antenna configurations are evaluated: $M=N=1$, $M=N=2$, and $M=N=4$ (depicted in figures as 1×1, 2×2, and 4×4, respectively). We compare the following schemes:

- **CS-CP-CD**: proposed scheme in this work where the user scheduling, precoding design and transmit direction selection are dynamically optimized at every frame by following Algorithm 6.2 under a PF criterion. Thus, in addition to conventional coordinated scheduling and coordinated precoding (CS-CP), coordinated transmit direction (CD) is included.
- **CS-CP-uD**: BS-specific TDD scheme with coordinated scheduling and coordinated precoding (CS-CP) but uncoordinated transmit direction (uD) [63, 147]. The transmit direction per BS is set according to the traffic asymmetries of the associated users (e.g. if $a_{i_k}=0.7$ then 7 frames are used for DL and the consecutive 3 frames are used for UL) while the user scheduling and transmit precoding matrices are dynamically optimized in a coordinated manner among BSs at every frame under a PF criterion.
- **uS-uP-uD**: BS-specific TDD scheme without coordination. The transmit direction per BS is set according to the traffic asymmetries of the associated users (similar to 'CS-CP-uD') while user scheduling and transmit precoding are selected at each BS independently (in a selfish manner at each frame) by following a PF criterion.

Recall that, as stated in the introduction, BS-specific decisions combined with interference coordination techniques (i.e. 'CS-CP-uD' described above) achieve the best performance among existing dynamic TDD schemes [147,151]. Also note that as the same traffic asymmetry (a_{i_k}) is considered for all users, then all BSs use the same DL-UL pattern under 'CS-CP-uD' and 'uS-uP-uD' schemes (but possibly not under 'CS-CP-CD').

6.6.1 Convergence of Algorithm 6.2

First, we demonstrate convergence of the proposed Algorithm 6.2 to solve the per-frame problem $(\mathrm{P}_{6,2}^s)$ in (6.13). It corresponds to 'CS-CP-CD' for a specific frame and a concrete deployment. Fig. 6.4 shows the evolution of the weighted sum-rate (in bits/s/Hz) for $|\mathcal{K}|=6$, $|\mathcal{I}_k|=2$, and different antenna configurations (1×1, 2×2, and 4×4). For this specific evaluation, the weighting coefficients for $(\mathrm{P}_{6,2}^s)$ in (6.13) are set to $\mu_{i_k,s}^d=0.5, \forall i_k, k, d$. It can be observed that the weighted sum-rate converges quite fast. Also, the monotonic increase of the weighted sum-rate is verified.

6.6.2 Evolution of on-line solutions

Second, let us show how the proposed Algorithm 6.3 that solves the problem $(\mathrm{P}_{6,1})$ in (5.15) evolves in time for a concrete deployment. Fig. 6.5 displays the evolution of the global objective function f_{global} in (6.28) versus the frame number (i.e. $s=1,\ldots,S$) for $|\mathcal{K}|=6$, $|\mathcal{I}_k|=1$, $a_{i_k}=0.7$, and two different antenna configurations (1×1 and 4×4). We can observe that the weighted sum of the log average rates increases with the frames and then gets stabilized for all schemes, such that the on-line solutions converge. The values in which each scheme gets stabilized will be compared in what follows through simulation results averaged over different deployments.
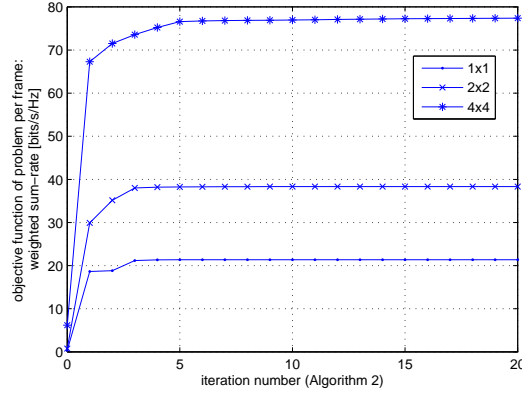
Figure 6.4: Convergence of Algorithm 6.2. Weighted sum-rate (objective function of problem $(\mathrm{P}_{6,2}^s)$ in (6.13)) vs. iteration number. $|\mathcal{K}|$=6, $|\mathcal{I}_k|$=2, $\mu_{i_k,s}^d$=0.5.
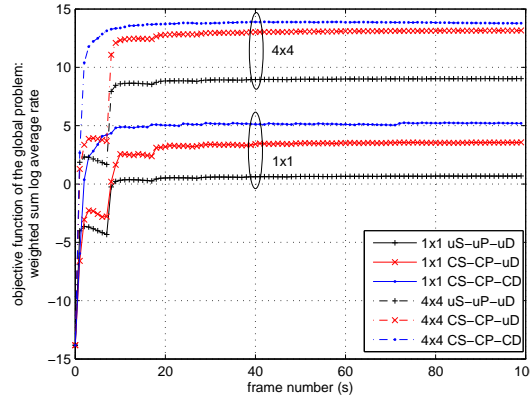

Figure 6.5: Evolution of Algorithm 6.3. Weighted sum of the log average rate (objective function f_{global} of problem $(\mathrm{P}_{6,1})$ in (6.9)) vs. frame number. $|\mathcal{K}|$=6, $|\mathcal{I}_k|$=1, a_{i_k}=0.7, $|\mathcal{S}|$=100.

6.6.3   Results for different network/user densities and asymmetries

In this section we show results averaged over 500 different random deployments.

Fig. 6.6 depicts the mean of the objective function of the global problem $(\mathrm{P}_{6,1})$ in (6.9) (i.e. f_{global} in (6.28)) versus the number of BSs ($|\mathcal{K}|$) for $|\mathcal{I}_k|$=1, a_{i_k}=0.7, $|\mathcal{S}|$=100, and different antenna configurations (1×1, 2×2, and 4×4).

Fig. 6.7 shows the mean of the objective function of the global problem $(\mathrm{P}_{6,1})$ in (6.9) (i.e. f_{global} in (6.28)) versus the number of users per BS ($|\mathcal{I}_k|$) for $|\mathcal{K}|$=4, a_{i_k}=0.7, $|\mathcal{S}|$=100, and different antenna configurations (1×1, 2×2, and 4×4).

Fig. 6.8 displays the mean of the objective function of the global problem $(\mathrm{P}_{6,1})$ in (6.9) (i.e. f_{global} in (6.28)) versus the traffic asymmetry (a_{i_k}) for $|\mathcal{K}|$=4, $|\mathcal{I}_k|$=1, $|\mathcal{S}|$=100, and different antenna configurations (1×1, 2×2, and 4×4).

We can observe in Fig. 6.6-6.7-6.8 that the proposed 'CS-CP-CD' outperforms 'CS-CP-uD' for all antenna configurations, traffic asymmetries, network densities and user densities. This is thanks to the optimization of the transmit direction selection jointly with the conventional
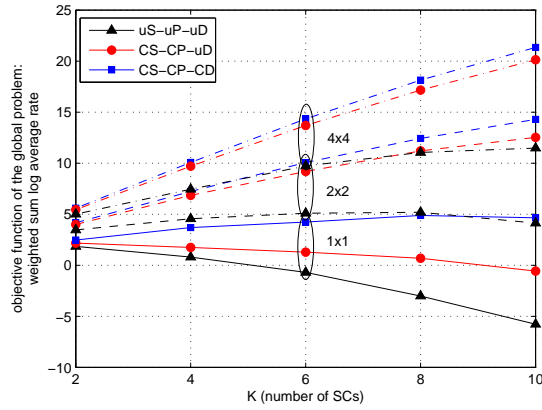
Figure 6.6: Mean weighted sum of the log average rate (objective function f_{global} of problem $(P_{6,1})$ in (6.9)) vs. number of BSs $(|\mathcal{K}|)$. $|\mathcal{I}_k|=1$, $a_{i_k}=0.7$, $|\mathcal{S}|=100$.
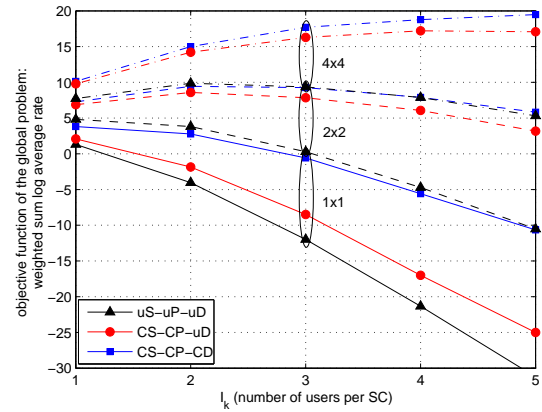

Figure 6.7: Mean weighted sum of the log average rate (objective function f_{global} of problem $(P_{6,1})$ in (6.9)) vs. number of users per BS $(|\mathcal{I}_k|)$. $|\mathcal{K}|=4$, $a_{i_k}=0.7$, $|\mathcal{S}|=100$.
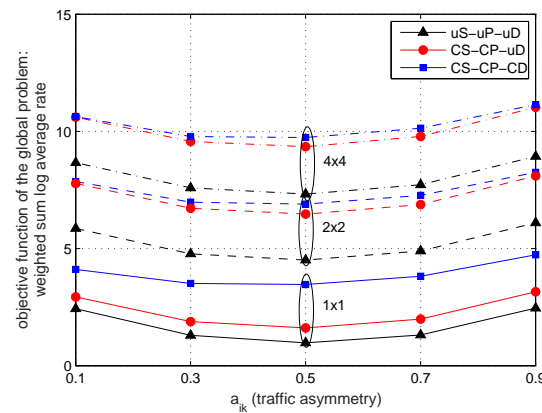

Figure 6.8: Mean weighted sum of the log average rate (objective function f_{global} of problem $(P_{6,1})$ in (6.9)) vs. traffic asymmetry (a_{i_k}). $|\mathcal{K}|=4$, $|\mathcal{I}_k|=1$, $|\mathcal{S}|=100$.

allocation policies. The gains of 'CS-CP-CD' as compared to 'CS-CP-uD' are larger for 1×1 than for 2×2 and 4×4 because when increasing the number of antennas more interference can be already managed through CS-CP and thus 'CS-CP-uD' does already provide significant gains over 'uS-uP-uD'. The gains are not much dependent on the traffic asymmetry condition (see Fig. 6.8) but, however, they do vary with the network density and the user density.

As the network density increases (i.e. as $|\mathcal{K}|$ increases, see Fig. 6.6), significant gains are obtained with the coordinated strategies ('CS-CP-uD' and 'CS-CP-CD') because more interference is present in the network and thus more important becomes the coordination of the transmit strategies in the network. In addition, the gains of the proposed 'CS-CP-uD' technique as compared to 'CS-CP-CD' also increase with the network density. So, the joint optimization of the transmit direction and the allocation policies becomes important for interference management in dense networks.

As the user density increases (i.e. as $|\mathcal{I}_k|$ increases, see Fig. 6.7), large gains are obtained with the coordinated strategies ('CS-CP-uD' and 'CS-CP-CD') because there is more flexibility to manage interference (i.e. there are more users to schedule). The largest gain from 'CS-CP-CD' over 'CS-CP-uD' appear however in the 1×1 antenna configuration, owing to the fact that in the 2×2 and 4×4 flexibility is already provided with the number of antennas and hence the gains of 'CS-CP-CD' over 'CS-CP-uD' are not as remarkable as those of 'CS-CP-uD' over 'uS-uP-uD'.

6.6.4   Results of the average rate

Figure 6.9 depicts the cumulative distribution function (CDF) of the average rates obtained in DL and in UL, separately, for $|\mathcal{K}|=4$, $|\mathcal{I}_k|=2$, $|\mathcal{S}|=100$, $a_{i_k}=0.7$, and different antenna configurations: 1×1, 2×2 and 4×4. Results are averaged over 500 different random deployments. Table 6.1 summarizes the relative gains in the mean of the average rates of 'CS-CP-uD' over 'uS-uP-uD' and of 'CS-CP-CD' over 'CS-CP-uD' for the different antenna configurations.

As it is expected, for $a_{i_k}=0.7$ (i.e. DL traffic is larger than UL traffic), DL average rates are larger than UL average rates (see Fig. 6.9). For 1×1 (see Fig. 6.9.(a)), 'CS-CP-CD' provides a significant improvement of both the outage and the mean of the average rate in DL and UL. The mean gains of 'CS-CP-CD' over 'CS-CP-uD' are 49-60% (see Table 6.1). For 2×2 and 4×4 cases (see Fig. 6.9.(b)-(c)), 'CS-CP-uD' does already provide an important enlargement of the outage rates and improves the mean average rate around 42-95% over 'uS-uP-uD' (see Table 6.1), thus leaving less improvement for 'CS-CP-CD'. However, 'CS-CP-CD' allows still improving the rates in both transmit directions as compared to 'CS-CP-uD' and provides gains between 10-16% (see Table 6.1). So the gains of 'CS-CP-CD' over 'CS-CP-uD' are reduced with the number of antennas, but not depreciable.

Table 6.1: RELATIVE GAINS OF 'CS-CP-uD' OVER 'uS-uP-uD' AND OF 'CS-CP-CD' OVER 'CS-CP-uD' IN THE MEAN OF THE AVERAGE RATES IN DL AND UL, SEPARATELY. $|\mathcal{K}|=4$, $|\mathcal{I}_k|=2$, $|\mathcal{S}|=100$, $a_{i_k}=0.7$.

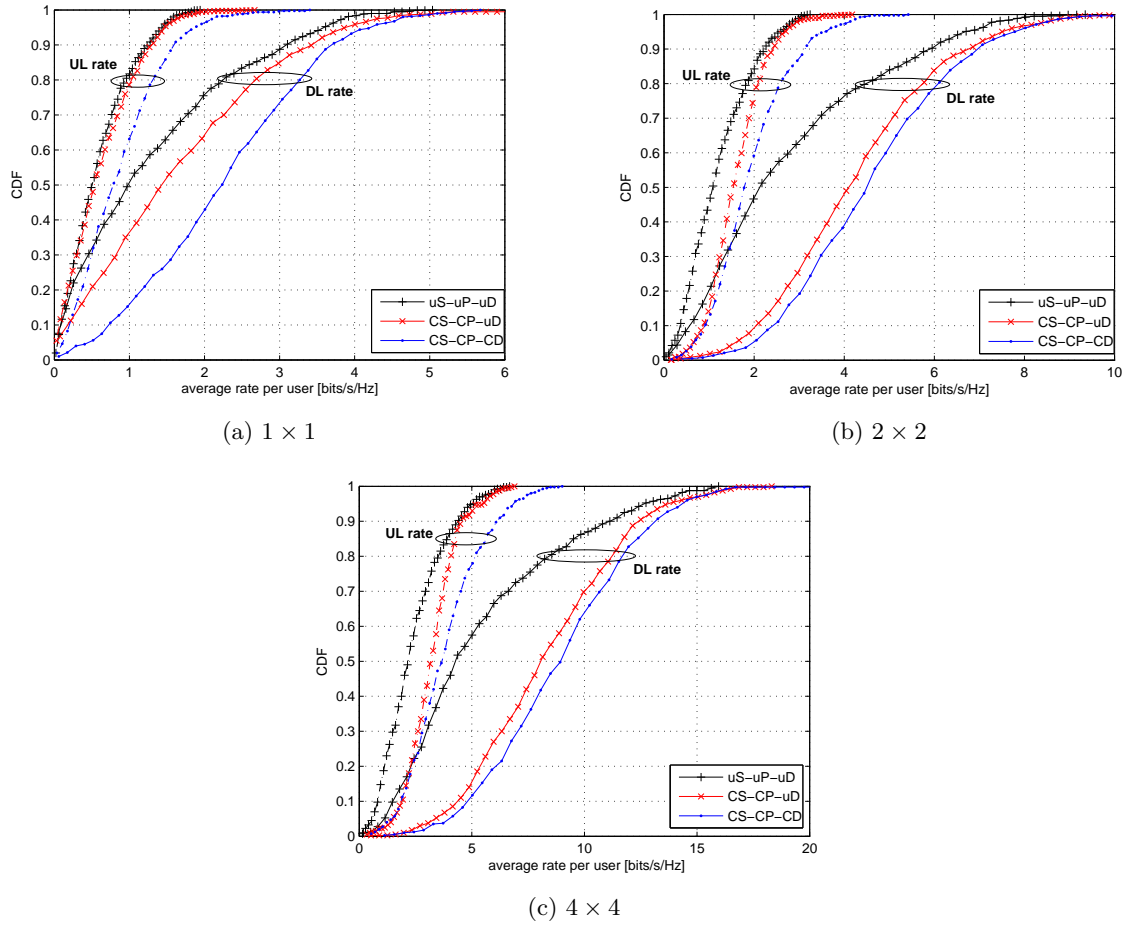|	1×1		2×2		4×4	
	DL	UL	DL	UL	DL	UL
CS-CP-uD over uS-uP-uD	46%	7%	95%	45%	86%	42%
CS-CP-CD over CS-CP-uD	60%	49%	11%	16%	10%	15%

(a) 1×1

(b) 2×2

(c) 4×4

Figure 6.9: CDF of the average rate (in bits/s/Hz) in DL and in UL. $|\mathcal{K}|$=4, $|\mathcal{I}_k|$=2, $|\mathcal{S}|$=100, a_{i_k}=0.7.

To conclude, from Fig. 6.6-6.7-6.8-6.9, we infer that:

- larger gains are obtained for 1×1 than for 2×2 and 4×4, because the 1×1 setup has less flexibility for interference management and is therefore more benefited from the optimization of the transmit direction. Nevertheless, positive gains are still reported when increasing the number of transmit/receive antennas.

- the gains of 'CS-CP-CD' increase as the network density increases (i.e. as $|\mathcal{K}|$ increases), due to the fact that as $|\mathcal{K}|$ increases then more interference is present and the optimization of the transmit direction becomes important for interference management.

- the gains of 'CS-CP-CD' increase as the user density increases (i.e. as $|\mathcal{I}_k|$ increases), because by increasing $|\mathcal{I}_k|$ more flexibility is available to select the user and transmit direction.

Therefore, the joint optimization of the transmit direction and the conventional allocation policies becomes key for interference management and it is specially important for interference-limited systems with a low number of antennas.

6.7   Conclusions

This chapter presents a dynamic procedure for joint user scheduling, precoding design and transmit direction selection in dynamic TDD MIMO smallcell networks. Differently from previous works, the transmit direction is optimized at every frame jointly with the conventional allocation policies that include user scheduling, precoding design and power control. Hence, a high adaptability to the instantaneous traffic and interference conditions in the network is achieved. The concepts for interference management presented in Chapter 3 have been used, but the whole problem formulation has been modified to include DL/UL traffic asymmetry conditions among BSs and to address the new types of interference that appear under flexible duplexing. To solve the general problem, decomposition methods (including stochastic approximation recursions and the interference-cost concept) have been applied.

Simulation results show gains in DL and UL average rates for different traffic asymmetries, network densities and user densities as compared to existing schemes for dynamic TDD. The gains are larger for systems with a low number of antennas because they have less flexibility for interference management through the coordination of transmit/receive spatial filters and hence larger gains are obtained by optimizing the transmit direction.

Appendices

6.A   Proof of Proposition 6.1

Let $\lambda_{i_k,s}^d$ denote the Lagrange multiplier associated to the power constraint in (6.21). Thus, the Lagrange function for problem in (6.21) can be written as:

$$
\begin{aligned}
\mathcal{L}_{i_k,s}^d \left(\mathbf{Q}_{i_k,s}^d, \lambda_{i_k,s}^d \right) &= \mu_{i_k,s}^d \log_2 \left| \mathbf{I} + \mathbf{H}_{i_k}^d \mathbf{Q}_{i_k,s}^d \mathbf{H}_{i_k}^{dH} \left(\mathbf{N}_{i_k,s}^d \right)^{-1} \right| - \mathrm{Tr} \left(\mathbf{Q}_{i_k,s}^d \mathbf{\Upsilon}_{i_k,s}^d \right) \\
&\quad - \lambda_{i_k,s}^d \left(\mathrm{Tr} \left(\mathbf{Q}_{i_k,s}^d \right) - P^{\max} \right).
\end{aligned}
\tag{6.29}
$$

The dual function for problem in (6.21) is given by:

$$
g\left(\lambda_{i_k,s}^d \right) = \max_{\mathbf{Q}_{i_k,s}^d \succeq \mathbf{0}} \; \mathcal{L}_{i_k,s}^d \left(\mathbf{Q}_{i_k,s}^d, \lambda_{i_k,s}^d \right).
\tag{6.30}
$$

Then, the dual problem of (6.21) is defined as [92]:

$$
\underset{\lambda_{i_k,s}^d \geq 0}{\text{minimize}} \; g\left(\lambda_{i_k,s}^d \right).
\tag{6.31}
$$

Since problem in (6.21) is convex with strictly feasible points [92], the duality gap between its optimal value and that of the dual problem in (6.31) is zero.

Therefore, problem in (6.21) can be solved equivalently by solving its dual problem in (6.31). In order to solve the dual problem, we need to obtain the dual function $g\left(\lambda_{i_k,s}^d\right)$ in (6.30) for any given $\lambda_{i_k,s}^d \geq 0$. This can be done by solving the maximization problem given in (6.30), which can be explicitly written (by discarding irrelevant constant terms in (6.29)) as:

$$
\underset{\mathbf{Q}_{i_k,s}^d \succeq \mathbf{0}}{\text{maximize}} \; \mu_{i_k,s}^d \log_2 \left| \mathbf{I} + \mathbf{H}_{i_k}^d \mathbf{Q}_{i_k,s}^d \mathbf{H}_{i_k}^{dH} \left(\mathbf{N}_{i_k,s}^d \right)^{-1} \right| - \mathrm{Tr} \left(\mathbf{B}_{i_k,s}^d \mathbf{Q}_{i_k,s}^d \right),
\tag{6.32}
$$

where $\mathbf{B}_{i_k,s}^d = \mathbf{\Upsilon}_{i_k,s}^d + \lambda_{i_k,s}^d \mathbf{I}$. $\mathbf{B}_{i_k,s}^d$ is a full rank matrix and hence its inverse $\left(\mathbf{B}_{i_k,s}^d \right)^{-1}$ exists.

Similarly as in [41], define variable $\tilde{\mathbf{Q}}_{i_k,s}^d$ as:

$$
\tilde{\mathbf{Q}}_{i_k,s}^d = \left(\mathbf{B}_{i_k,s}^d \right)^{\frac{H}{2}} \mathbf{Q}_{i_k,s}^d \left(\mathbf{B}_{i_k,s}^d \right)^{\frac{1}{2}}.
\tag{6.33}
$$

Substituting (6.33) into (6.32) leads to:

$$\underset{\tilde{\mathbf{Q}}^d_{i_k,s} \succeq \mathbf{0}}{\text{maximize}} \ \mu^d_{i_k,s} \log_2 \left| \mathbf{I} + \mathbf{H}^d_{i_k} \left(\mathbf{B}^d_{i_k,s} \right)^{-\frac{H}{2}} \tilde{\mathbf{Q}}^d_{i_k,s} \left(\mathbf{B}^d_{i_k,s} \right)^{-\frac{1}{2}} \mathbf{H}^{dH}_{i_k} \left(\mathbf{N}^d_{i_k,s} \right)^{-1} \right| - \text{Tr}(\tilde{\mathbf{Q}}^d_{i_k,s}). \quad (6.34)$$

Without loss of generality, define the following eigenvalue decomposition (EVD):

$$\left(\mathbf{B}^d_{i_k,s} \right)^{-\frac{1}{2}} \mathbf{H}^{dH}_{i_k} \left(\mathbf{N}^d_{i_k,s} \right)^{-1} \mathbf{H}^d_{i_k} \left(\mathbf{B}^d_{i_k,s} \right)^{-\frac{H}{2}} = \mathbf{V}^d_{i_k,s} \mathbf{Z}^d_{i_k,s} (\mathbf{V}^d_{i_k,s})^H, \quad (6.35)$$

where $\mathbf{V}^d_{i_k,s} \in \mathbb{C}^{\tilde{M} \times \tilde{M}}$ is a unitary matrix and $\mathbf{Z}^d_{i_k,s} = \text{diag}(z^d_{i_k,s}(1), \dots, z^d_{i_k,s}(\tilde{M})) \in \mathbb{C}^{\tilde{M} \times \tilde{M}}$ is a diagonal matrix, being $\tilde{M} = M$ if $d = D$ and $\tilde{M} = N$ if $d = U$. Substituting the above EVD in (6.35) into (6.34) and applying the Hadamard's inequality (e.g., see [168]), the optimal solution for $\tilde{\mathbf{Q}}^d_{i_k,s}$ in (6.34) is found as:

$$\tilde{\mathbf{Q}}^d_{i_k,s} = \mathbf{V}^d_{i_k,s} \mathbf{\Sigma}^d_{i_k,s} (\mathbf{V}^d_{i_k,s})^H, \quad (6.36)$$

being $\mathbf{\Sigma}^d_{i_k,s} = \text{diag}(\sigma^d_{i_k,s}(1), \dots, \sigma^d_{i_k,s}(\tilde{M})) \in \mathbb{C}^{\tilde{M} \times \tilde{M}}$ a diagonal matrix whereby each diagonal element $\sigma^d_{i_k,s}(n)$ is obtained applying the standard water-filling algorithm [168]:

$$\sigma^d_{i_k,s}(n) = \left(\frac{\mu^d_{i_k,s}}{\ln(2)} - \frac{1}{z^d_{i_k,s}(n)} \right)^+. \quad (6.37)$$

Finally, the optimal solution for $\mathbf{Q}^d_{i_k,s}$ is obtained by including (6.36) into (6.33):

$$\mathbf{Q}^d_{i_k,s} = \left(\mathbf{B}^d_{i_k,s} \right)^{-\frac{H}{2}} \tilde{\mathbf{Q}}^d_{i_k,s} \left(\mathbf{B}^d_{i_k,s} \right)^{-\frac{1}{2}} = \left(\mathbf{B}^d_{i_k,s} \right)^{-\frac{H}{2}} \mathbf{V}^d_{i_k,s} \mathbf{\Sigma}^d_{i_k,s} (\mathbf{V}^d_{i_k,s})^H \left(\mathbf{B}^d_{i_k,s} \right)^{-\frac{1}{2}}. \quad (6.38)$$

With the obtained dual function $g(\lambda^d_{i_k,s})$ for any given $\lambda^d_{i_k,s}$, the dual problem (6.31) can be solved by searching over $\lambda^d_{i_k,s} \geq 0$ to minimize $g(\lambda^d_{i_k,s})$. This can be done, for example, through the bisection method or the ellipsoid method [92]. When $\lambda^d_{i_k,s}$ converges to the optimal solution for the dual problem, the corresponding $\mathbf{Q}^d_{i_k,s}$ becomes the optimal solution for problem in (6.21).

6.B Proof of Proposition 6.1

For SISO systems, i.e. $M_k = N_{i_k} = 1$, $\forall i_k, \forall k$, all parameters in the system are scalars. Let $H^D_{i_k} = H_{k,i_k}$ and $H^U_{i_k} = H_{i_k,k}$ denote the DL and UL complex channels, respectively. Let $N^D_{i_k,s}$ and $N^U_{i_k,s}$ denote the interference-plus-noise received power in DL and UL, respectively (from (6.7)-(6.8)). Let $\Upsilon^D_{i_k,s}$ and $\Upsilon^U_{i_k,s}$ denote the cost in DL and UL, respectively (from (6.17)-(6.18)). Now, the optimization variables are the transmit power in DL ($P^D_{i_k,s}$) and in UL ($P^U_{i_k,s}$), instead of transmit covariance matrices (see (6.3)). Furthermore, the design of the Lagrange multiplier in (6.23) can be omitted in the SISO case, as the constraint of the maximum transmit power can be directly imposed over optimization variables $P^d_{i_k,s} \leq P^{\max}$, see (6.21).

So let us show how the solution for the SISO case in (6.27) is derived from the general MIMO solution in (6.23). Use $\lambda^d_{i_k,s} = 0$ in (6.24) such that: $B^d_{i_k,s} = \Upsilon^d_{i_k,s}$. The EVD in (6.25)

does not need to be performed in the SISO case, and we should simply set: $V_{i_k,s}^d=1$ and $Z_{i_k,s}^d=|H_{i_k}^d|^2/(\Upsilon_{i_k,s}^d N_{i_k,s}^d)$. This way, substituting said values into (6.26) we have:

$$\sigma_{i_k,s}^d = \left(\frac{\mu_{i_k,s}^d}{\ln(2)} - \frac{\Upsilon_{i_k,s}^d N_{i_k,s}^d}{|H_{i_k}^d|^2}\right)^+,\tag{6.39}$$

and finally from (6.23):

$$P_{i_k,s}^d = \frac{\sigma_{i_k,s}^d}{\Upsilon_{i_k,s}^d} = \left(\frac{\mu_{i_k,s}^d}{\ln(2)\Upsilon_{i_k,s}^d} - \frac{N_{i_k,s}^d}{|H_{i_k}^d|^2}\right)^+.\tag{6.40}$$

Therefore, the optimal solution for the power allocation $(\bar{P}_{i_k,s}^d)$ in (6.27) is directly obtained by constraining the value in (6.40) between 0 and the maximum available power ($P^{\max}$).

6.C   Proof of Proposition 6.2

If Algorithm 6.2 converges, then it clearly converges to a limit point satisfying the set of KKT conditions of subproblems $(P_{6,3}^{s,k})$ in (6.20) (see line 8 in Algorithm 6.2). In addition, it can be easily shown that the set of KKT conditions of subproblems $(P_{6,3}^{s,k})$ in (6.20) constitute precisely the KKT conditions of problem $(P_{6,2}^s)$ in (6.13). Therefore, if convergence is achieved, the limit point will also satisfy the KKT conditions of problem $(P_{6,2}^s)$ in (6.13).

So, let us now show that Algorithm 6.2 converges. Note first that the function $f_{-k}(\mathbf{Q}_{k,s},\mathbf{Q}_{-k,s})$ in (6.15) is jointly convex w.r.t. the set of matrices composing $\mathbf{Q}_{k,s}$ (i.e. $\mathbf{Q}_{i_k,s}^d$, $\forall i_k\in\mathcal{I}_k, d$), since $f_{-k}(\mathbf{Q}_{k,s},\mathbf{Q}_{-k,s})$ can be seen as the composition of a convex function and a linear function giving then as result a convex function [92]. Then, in the following we show that after solving problem $(P_{6,3}^{s,k})$ in (6.20) for the k-th BS, the objective function of problem $(P_{6,2}^s)$ in (6.13) is not decreased. Let $\mathbf{Q}_{k,s}^\star$ be the optimal solution to $(P_{6,3}^{s,k})$ in (6.20), i.e. $\mathbf{Q}_{k,s}^\star\triangleq\{\mathbf{Q}_{1_k,s}^{D\star},\ldots,\mathbf{Q}_{I_k,s}^{D\star},\mathbf{Q}_{1_k,s}^{U\star},\ldots,\mathbf{Q}_{I_k,s}^{U\star}\}$. Similarly as in (6.15), define the weighted sum-rate of the links related to the k-th BS as:

$$f_k(\mathbf{Q}_{k,s},\mathbf{Q}_{-k,s}) = \sum_{i_k\in\mathcal{I}_k}\left(\mu_{i_k,s}^D R_{i_k,s}^D + \mu_{i_k,s}^U R_{i_k,s}^U\right),\tag{6.41}$$

such that the objective function of $(P_{6,2}^s)$ in (6.13) is $\sum_{k\in\mathcal{K}}f_k(\mathbf{Q}_{k,s},\mathbf{Q}_{-k,s})$. Hence,

$$\sum_{k\in\mathcal{K}}f_k(\mathbf{Q}_{k,s}^\star,\mathbf{Q}_{-k,s}) = f_k(\mathbf{Q}_{k,s}^\star,\mathbf{Q}_{-k,s}) + f_{-k}(\mathbf{Q}_{k,s}^\star,\mathbf{Q}_{-k,s})$$
$$\geq f_k(\mathbf{Q}_{k,s}^\star,\mathbf{Q}_{-k,s}) + f_{-k}(\bar{\mathbf{Q}}_{k,s},\mathbf{Q}_{-k,s}) - \sum_{i_k\in\mathcal{I}_k}\left(\mathrm{Tr}((\mathbf{Q}_{i_k,s}^{D\star}-\bar{\mathbf{Q}}_{i_k,s}^D)\Upsilon_{i_k,s}^D)+\mathrm{Tr}((\mathbf{Q}_{i_k,s}^{U\star}-\bar{\mathbf{Q}}_{i_k,s}^U)\Upsilon_{i_k,s}^U)\right)$$
$$\geq f_k(\bar{\mathbf{Q}}_{k,s},\mathbf{Q}_{-k,s}) + f_{-k}(\bar{\mathbf{Q}}_{k,s},\mathbf{Q}_{-k,s}) - \sum_{i_k\in\mathcal{I}_k}\left(\mathrm{Tr}((\bar{\mathbf{Q}}_{i_k,s}^D-\bar{\mathbf{Q}}_{i_k,s}^D)\Upsilon_{i_k,s}^D)+\mathrm{Tr}((\bar{\mathbf{Q}}_{i_k,s}^U-\bar{\mathbf{Q}}_{i_k,s}^U)\Upsilon_{i_k,s}^U)\right)$$
$$= f_k(\bar{\mathbf{Q}}_{k,s},\mathbf{Q}_{-k,s}) + f_{-k}(\bar{\mathbf{Q}}_{k,s},\mathbf{Q}_{-k,s}) = \sum_{k\in\mathcal{K}}f_k(\bar{\mathbf{Q}}_{k,s},\mathbf{Q}_{-k,s}),$$
$$\tag{6.42}$$

where the first inequality is a consequence of the function $f_{-k}(\mathbf{Q}_{k,s},\mathbf{Q}_{-k,s})$ being jointly convex

w.r.t. the set of matrices composing $\mathbf{Q}_{k,s}$ and the fact that the cost matrices $\mathbf{\Upsilon}_{i_k,s}^d$ in (6.17)-(6.18) are hermitian matrices[23], and the second inequality holds since $\mathbf{Q}_{k,s}^{\star}$ (i.e. $\{\mathbf{Q}_{i_k,s}^{d\star}\}$, $\forall i_k \in \mathcal{I}_k, d$) is the optimal solution to problem $(\mathrm{P}_{6,3}^{s,k})$ in (6.20). Finally, as the objective function of problem $(\mathrm{P}_{6,2}^{s})$ in (6.13) is bounded from above due to the maximum power constraints, the algorithm must converge to (at least) a local optimum.

---

[23]For a differentiable convex function $f(\mathbf{A})$, $f(\mathbf{A}) \geq f(\mathbf{B}) + \mathrm{Tr}\big(\big(\frac{\delta f(\mathbf{A})}{\delta \mathbf{A}}\big)^H \big|_{\mathbf{B}} (\mathbf{A} - \mathbf{B})\big)$ holds, where the right-hand side term corresponds to the first order Taylor expansion of $f(\mathbf{A})$ evaluated at $\mathbf{B}$.

Chapter 7

Long-Term Traffic-Aware Resource Management

This chapter investigates how to reuse a fraction of spectrum by multiple TDD BSs in dense OFDMA-based networks, where multiple BSs with possibly overlapping coverage areas compete for the same set of resources. We propose procedures to distribute the frequency spectrum in the long-term by taking into account the per-BS traffic loads in DL and UL as well as their spectral efficiencies. The optimization is done over the long term, being independent of the specific users connected to each BS but dynamic enough to follow significant variations of the traffic load. In this sense, we focus on the average resource utilization (RU) and propose schemes to minimize either the maximum RU or the weighted sum of RU of all BSs. Firstly, we consider orthogonal resource usage among BSs. Optimal closed-form expressions for the long-term resource provisioning are derived for this case. Secondly, we assume that resources can be reused at non-overlapping BSs. In this case, the resource provisioning is solved in two steps: i) the number of resources required per BS is obtained by discretizing the optimal solution of a convex problem, and then ii) the specific resources to be utilized by each BS are determined by using graph coloring. In contrast to previous works, graph coloring can be applied to get an implementable solution for any condition of the BS loads. Simulation results show a significant reduction of the maximum RU, which translates into an increase of the served traffic and a reduction of the packet delay, as compared to static resource provisioning schemes.

The technical paper related to this chapter is:

[C7]   S. Lagen, O. Muñoz, A. Pascual-Iserte, J. Vidal, A. Agustin, "Long-term Provisioning of Radio Resources Based on their Utilization in Dense OFDMA Networks", *IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun.*, Valencia (Spain), Sep. 2016.

This chapter is structured as follows. Section 7.1 contains the state of the art for long-term resource provisioning in OFDMA-based networks. The main contributions of the chapter are detailed in Section 7.2. In Section 7.3 the system model is presented and the RU is defined, detailing how all the required parameters to estimate the RU are computed. In Section 7.4, the problem is formulated and solved in closed-form under orthogonal resource usage. In Section 7.5, the problem is formulated and solved with graph coloring when reuse of resources among non-overlapping BSs is permitted. Section 7.6 presents the simulation results. Finally, concluding remarks are included in Section 7.7.

Preliminaries for graph theory are included in Appendix 7.A.

7.1 State of the Art

OFDMA is the radio access technology employed in downlink transmission of 3GPP LTE and LTE-A systems [2], as well as in both uplink and downlink of IEEE 802.16m advanced WiMAX [26]. Furthermore, it is one of the major candidates for different uses cases of future 5G systems [27]. In OFDMA-based networks, the intra-cell users are assumed to be orthogonal to each other and the primary source of interference is inter-cell interference [46]. For that reason, due to the upcoming network densification and scarcity of spectrum, efficient provisioning of radio resources in multi-cell scenarios is crucial. Radio resource provisioning schemes should improve spatial reuse while avoiding dominant inter-cell interference. In addition, for simplicity of operation and implementability, they should be performed at the **long-term**[24], i.e. should be independent of the users to be served at a specific time instant.

The simplest resource provisioning scheme is frequency reuse-n, where the total bandwidth is partitioned in n bands and different bands are assigned to neighbor BSs to avoid dominant inter-cell interference. As an example, Fig. 7.1 shows a frequency reuse-2. To further improve the spatial reuse while minimizing inter-cell interference impact, fractional frequency reuse (FFR) was proposed [30], where the total bandwidth is partitioned such that *i*) cell-edge users of neighbor BSs do not interfere with each other and *ii*) interference received by (and created by) cell-interior users is reduced. Many different FFR schemes are analyzed in [31]. However, theses schemes are static and independent of the traffic loads of the BSs. Hence, they might not be suitable for dense and irregular deployments of BSs where the traffic load can vary drastically over the space and time domains as compared to conventional macrocell-based networks (e.g. in Fig. 7.1, BS_2 has a higher number of users and could require larger bandwidth than BS_1 and BS_3).
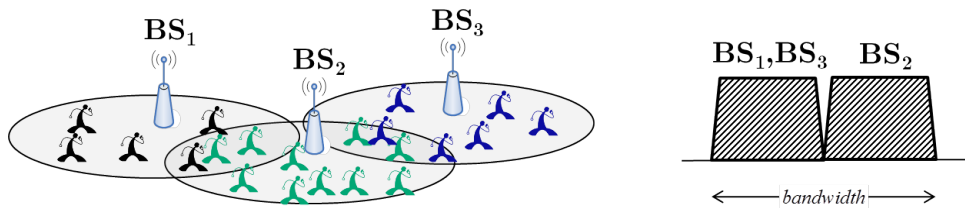

Figure 7.1: Example of frequency reuse-2 for long-term resource provisioning.

Mathematically speaking, resource provisioning in multi-cell scenarios (also known as the dynamic channel assignment problem [169]) is a combinatorial optimization task that can be mapped into a **graph coloring** problem and is therefore NP-hard. In this line, graph-based approaches are proposed in [32] to allow dynamic FFR and distribute the resources in the long-term according to the per-BS traffic loads. Similarly, [170] exploits graph coloring to perform the resource provisioning at the short-term by adapting the resource allocations to the instantaneous load of the specific users to be served. The procedure is summarized as follows, and shown through an example in Fig. 7.2. First, each BS computes the number of resources required to

---

[24]Long-term resource provisioning involves provisioning of frequency resources for several consecutive transmission time intervals (TTIs). Rather than dynamically changing the resource provisioning at each TTI, long-term resource provisioning is preferable for simplicity of operation/implementation. The periodicity might be therefore of the order of minutes.
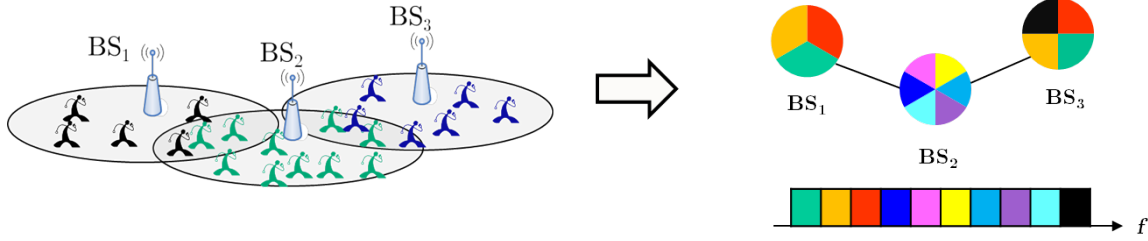
Figure 7.2: Dynamic long-term resource provisioning based on graph coloring.

meet the user demands assuming interference free conditions. Then, a graph is set according to the demands of each BS and any two BSs in an interference condition are connected through an edge. Finally, the graph is colored through graph coloring algorithms, as it is shown in Fig. 7.2 where each color represents a frequency resource.

The drawback of all existing schemes based on graph coloring is that they employ a graph (which varies according to the network deployment and per-BS traffic loads, see Fig. 7.2) which might not be colorable with the number of colors (radio resources) that are available. This means that proper coloring[25] with the number of available resources may not be possible or, equivalently, that an implementable solution satisfying the coloring constraints may not be found. For example, the graph in Fig. 7.2 could not have been colored with less than 10 colors.

7.2 Contribution

In this chapter, we present a novel approach for long-term resource provisioning that is based on optimizing the **resource utilization**, which measures the occupancy of a BS and is given by the division among the amount of data traffic and the effective capacity per BS. We formulate the resource provisioning strategy as a convex problem so as to minimize a function of the RU, such as the maximum RU among BSs or the weighted sum of RU. Two different approaches are considered:

- orthogonal resource provisioning (orthogonal access among BSs is adopted) and
- graph-based resource provisioining (reuse of resources among BSs is allowed).

When orthogonal resource usage among BSs is considered, closed-form expressions are derived. When reuse of resources among non-overlapping cells is permitted, we propose an scheme that works as follows. First, the optimization problem is formulated according to the interference graph (wherein we impose that neighbor BSs must use different resources to avoid strong intercell interference). The optimal resource provisioning per BS is obtained by solving the problem. Then, such repartition is mapped into real resources by using graph coloring over a new extended graph.

In contrast to previous works [32,170], we first solve an optimization problem which ensures that proper coloring over the extended graph will be possible and, hence, that an implementable solution will be found.

---

[25]A proper coloring of a graph is an assignment of colors to the vertices of the graph such that no two adjacent vertices have the same color.

7.3   System Model

Assume a TDD cellular network composed of a set of $\mathcal{K}=\{1,\ldots,K\}$ randomly and densely distributed BSs. Due to the random and dense geographical distribution, the coverage area of some BSs may overlap, i.e. some users might be in a location covered by multiple BSs. So, if these BSs use the same frequency resource, interference will be generated.

Assume OFDMA, as in LTE-A DL [2]. There are N effective frequency sub-channels, also called RBs in LTE-A, available in the system each with a bandwidth of B. No power control is assumed such that the transmit power per resource (or RB) is fixed or, equivalently, a spectral power mask is adopted.

Our objective is to provide a method to decide the frequency resources assigned to each BS in the *long term*, i.e.:

- it should not be adapted to a concrete set of users, as this would require to change the amount of resources whenever a user appears o disappears in the BS, but
- it should be dynamic enough to adapt to significant changes in the average traffic load.

So the optimization will be performed every T time resources, which will be duplexed at each TDD BS among DL and UL transmissions.

In addition, the proposed method must seek to:

- use the spectrum resources efficiently by avoiding over-provisioning of resources,
- avoid situations of high resource occupancy that will lead to high packet delays, and
- provide quality-of-service (QoS) to the users of the different BSs.

A suitable measure that captures all these requirements is the RU.

The strategy for long-term provisioning proposed in this chapter is performed at a central controller that controls the $|\mathcal{K}|$ BSs and disposes of long-term information and network topology information. As long-term information, only a single parameter per BS is required. As network topology information we refer to knowledge of the interference graph [32].

The *interference graph* is constructed as follows:

- every BS in the network defines a vertex and
- any two BSs (represented through a vertex) in an interference situation are connected with an edge.

Interference situations appear whenever the transmission of one BS could potentially interfere the transmission of another BS if the same resource is used. For example, in one-tier networks, an interference situation could be established when two BSs were closer than a threshold distance. For multi-tier networks, where the coverage areas of BSs belonging to different tiers differ, a threshold distance could be defined per tier and an interference situation would appear whenever two BSs were closer than at least one of the threshold distances imposed by the respective BSs.

7.3.1   Estimation of the system parameters

The RU is a measure widely used in 3GPP evaluations to report the percentage of resources employed by a BS [2]. It is given by the ratio between the total number of resources used over the total number of resources available for data traffic. Therefore, we differentiate between the RU for DL and UL transmissions of each BS.

The RU of the k-th BS in the DL and UL transmit directions (ρ_k^D for DL and ρ_k^U for UL) can be estimated as the average traffic load of the BS on each transmit direction (in bits/s) over the amount of traffic that the BS can serve on said transmit direction (i.e. the effective capacity in DL or UL, in bits/s). It is also referred to as the normalized load of a BS in DL and UL transmit directions, respectively, see [171]:

$$\rho_k^D = \frac{\lambda_k^D L_k^D}{x_k \gamma_k C_k^D} = \frac{\alpha_k^D}{x_k \gamma_k}, \tag{7.1}$$

$$\rho_k^U = \frac{\lambda_k^U L_k^U}{x_k (T - \gamma_k) C_k^U} = \frac{\alpha_k^U}{x_k (T - \gamma_k)}, \tag{7.2}$$

where

$$\alpha_k^d = \frac{\lambda_k^d L_k^d}{C_k^d}, \tag{7.3}$$

λ_k^d is the mean packet arrival rate in the d-th transmit direction (in packets/s), L_k^d denotes the mean packet length in the d-th transmit direction (in bits/packet), x_k refers to the number of frequency resources per BS, γ_k denotes the number of time resources used for DL, $(T - \gamma_k)$ the number of time resources used for UL being T the total number of time resources[26], and C_k^d is the average spectral efficiency of the k-th BS in the d-th transmit direction(in bits/s/resource).

It can be observed in (7.1)-(7.2) that increasing the amount of frequency resources per BS (x_k) leads to low ρ_k^d and hence an inefficient usage of resources (as they could be used for other purposes). On the contrary, reducing the amount of frequency resources (x_k) leads to high ρ_k^d, which implicitly increases the packet delay and reduces the QoS of the associated users due to the high resource occupancy (see next (7.6)).

Note that, although the RU is a measure bounded between 0 and 1, the normalized load of the BS on each transmit direction ρ_k^d in (7.1)-(7.2) could be by definition larger than 1 if the traffic load is very high. However, systems should be properly designed so as not to "blow up" the queues, i.e. the resource provisioning strategy should assure that $\rho_k^d \leq 1$. Otherwise, if $\rho_k^d > 1$, the system would be unstable, inducing large queue sizes, losses of packets, and unacceptable packet delays. From now on we assume that the system will operate under this condition ($\rho_k^d \leq 1$), in which the normalized load ρ_k^d is equivalent to the RU of the BS on the d-th transmit direction.

In the non-full buffer FTP traffic models used in 3GPP evaluations [61], the packet arrival rate and the packet length are fixed (although they can be varied to emulate different load conditions corresponding to different times of the day). This greatly simplifies the computation of the RU in (7.1)-(7.2). Recall also that in FTP3 traffic model, see [61], the packet arrival rate is defined per user and, consequently, λ_k^d in (7.1)-(7.2) would be the product of the per-user packet arrival rate on the d-th transmit direction and the number of users in the k-th BS.

---

[26]Among the whole set of time resources T, γ_k are used for DL and the remaining $T - \gamma_k$ are used for UL. Within each time resource, the x_k frequency resources are used per BS on the associated transmit direction. Thus, the total amount of time and frequency resources for DL is $x_k \gamma_k$ and the total amount of time and frequency resources for UL is $x_k (T - \gamma_k)$, see (7.1)-(7.2).

Parameters $\alpha_k^d, \forall k \in \mathcal{K}, d \in \{D,U\}$ in (7.3) correspond to the long-term information required at the central controller to perform the optimization. α_k^d can be estimated based on the average traffic load $(\lambda_k^d L_k^d)$ and the average spectral efficiency (C_k^d) of the k-th BS in the d-th transmit direction. The later, C_k^d, can be estimated based on the statistics of the previously served transmission rates. There are two main approaches in the literature to estimate C_k^d depending on the scheduling strategy that is adopted. If a round-robin scheduler is used, C_k^d is given by:

$$C_k^d = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} R_{i,k}^d, \tag{7.4}$$

where $R_{i,k}^d$ is the average transmission rate given to the i-th user at the k-th BS in the d-th transmit direction (in bits/s/resource) and $\mathcal{I}$ denotes the set of users among which the statistics are computed. Alternatively, if the scheduler is such that gives the same rate to all users, C_k^d can be estimated as [172,173]:

$$C_k^d = |\mathcal{I}| \left(\sum_{i \in \mathcal{I}} \frac{1}{R_{i,k}^d} \right)^{-1}. \tag{7.5}$$

These estimations are valid if a spectral power mask is used (i.e. the transmit power per resource is limited, as in LTE-A DL). Otherwise, C_k^d would depend on the power control and hence on the number of resources assigned to each BS.

If packet arrival instants are modeled as a Poisson process, for $\rho_k^d \leq 1$, the average number of bits in the queue of the k-th BS in the d-th transmit direction (W_k^d) is related to ρ_k^d in (7.1)-(7.2) as follows [171]:

$$W_k^d = \frac{\rho_k^d}{1 - \rho_k^d} \frac{l_k^d}{2L_k^d}, \tag{7.6}$$

where l_k^d denotes the mean of the squared packet length. Clearly, longer queue sizes imply higher packet delays. Therefore, in a single-cell system minimizing the RU is equivalent to minimize the average number of bits in the system (or to minimize the average packet delay), see (7.6).

In a multi-cell scenario, where multiple BSs compete for the same set of frequency resources, an efficient resource provisioning is such that the frequency resources are distributed among all the BSs in a balanced way trying to avoid very different occupancies for different BSs. In this sense, a suitable optimization criterion is the minimization of the maximum RU (max RU) among the BSs, such that resources are fairly distributed and more resources are given to those BSs with larger traffic loads and/or those BSs experiencing greater delays (see (7.1)-(7.2)).

7.3.2   Time duplexing

For a given BS k, the optimal repartition of the time resources (γ_k for DL, $(T-\gamma_k)$ for UL) in terms of RU is such that $\rho_k^D = \rho_k^U$ (i.e. same occupancy for DL and UL), which leads to a closed-form repartition:

$$\gamma_k = \frac{\alpha_k^D T}{\alpha_k^D + \alpha_k^U}. \tag{7.7}$$

This means that traffic asymmetries at each BS are taken into account to distribute the time resources among DL and UL transmissions at each BS independently. Differently, total traffic load at each BS will be considered to distribute frequency resources among BSs.

Accordingly, γ_k can be determined per BS. So let us work with the occupancy per BS (ρ_k), which will be equal for DL and UL transmit directions ($\rho_k = \rho_k^D = \rho_k^U$):

$$\rho_k = \frac{\alpha_k}{x_k}, \quad \alpha_k = \frac{\alpha_k^D}{\gamma_k} = \frac{\alpha_k^U}{(T - \gamma_k)} = \frac{\alpha_k^D + \alpha_k^U}{T}. \tag{7.8}$$

Under this setting, we focus on optimizing the allocation of frequency resources ($\{x_k\}$) in the multi-cell scenario by working with RU variables per cell ρ_k defined in (7.8). Recall that the resource provisioning in multi-cell environments is a combinatorial optimization problem that involves high complexity [169]. For that reason, we focus on solving a relaxed version of the optimization problem with continuous variables (corresponding to the distribution of resources, $\{x_k\}$) and then the obtained result is discretized.

7.4 Orthogonal Resource Provisioning

In this section, we derive a long-term resource provisioning scheme when orthogonal resource usage among BSs is assumed (i.e. the total frequency resources are split into disjoint sets, and each one is assigned to one BS). As optimization criteria we adopt the maximum RU and the weighted sum of RU.

7.4.1 Minimization of the maximum RU

The minimization of the max RU ($\rho_k = \frac{\alpha_k}{x_k}$) subject to the constraint that the sum of all resources must be lower or equal to the total number of frequency resources is formulated as:

$$(\text{P}_{7,1}): \quad \underset{\{x_k\}}{\text{minimize}} \quad \max_k \left(\frac{\alpha_k}{x_k} \right) \tag{7.9}$$
$$\text{subject to} \quad \sum_{k \in \mathcal{K}} x_k \leq N,$$

where N is the total number of frequency resources. Problem ($\text{P}_{7,1}$) in (7.9) is convex w.r.t. $\{x_k\}$. In the following we derive the optimal solution in closed-form.

Problem ($\text{P}_{7,1}$) in (7.9) may be equivalently written as:

$$(\text{P}_{7,2}): \quad \underset{\{x_k\},t}{\text{minimize}} \quad t \tag{7.10}$$
$$\text{subject to} \quad \begin{cases} \frac{\alpha_k}{x_k} \leq t \quad \forall n \\ \sum_{k \in \mathcal{K}} x_k \leq N. \end{cases}$$

Problem ($\text{P}_{7,2}$) in (7.10) is jointly convex w.r.t. $\{x_k\}$ and t. The optimal solution to ($\text{P}_{7,2}$) in (7.10) is such that all constraints are satisfied with equality [92]. This means that all RU are equal: $\frac{\alpha_k}{x_k} = \rho$. Therefore, the optimal solution for $\{x_k\}$ in ($\text{P}_{7,2}$) (and, consequently, in ($\text{P}_{7,1}$)) has the following structure:

$$x_k = \frac{\alpha_k}{\rho}. \tag{7.11}$$

By including (7.11) into the last constraint of problem ($P_{7,2}$) in (7.10) we have: $\sum_{k \in \mathcal{K}} \frac{\alpha_k}{\rho} = N$, from which we can isolate ρ:

$$\rho = \frac{1}{N} \sum_{k \in \mathcal{K}} \alpha_k. \tag{7.12}$$

Then, by combining (7.11) and (7.12), the optimal solution for the resource provisioning under orthogonal resource usage and minimum max RU is:

$$x_k^* = \frac{\alpha_k}{\sum_{l \in \mathcal{K}} \alpha_l} N. \tag{7.13}$$

The optimal solution in (7.13) gives more frequency resources to those BSs experiencing higher ratios among the average traffic load and the average spectral efficiency in DL and UL, i.e. higher α_k (see (7.8)).

7.4.2   Minimization of the weighted sum of RU

The minimization of the weighted sum of RU ($\rho_k = \frac{\alpha_k}{x_k}$) subject to the constraint that the sum of all resources must be lower or equal to the total number of frequency resources is given by:

$$(\text{P}_{7,3}): \quad \underset{\{x_k\}}{\text{minimize}} \quad \sum_{k \in \mathcal{K}} \mu_k \frac{\alpha_k}{x_k} \tag{7.14}$$
$$\text{subject to} \quad \sum_{k \in \mathcal{K}} x_k \leq N,$$

where N is the total number of frequency resources and μ_k is the weighting coefficient associated to the priority of the k-th BS. Problem ($P_{7,3}$) in (7.14) is convex w.r.t. $\{x_k\}$. Further, we can solve it in closed-form as follows.

Let us denote by λ to the Lagrange multiplier associated to the constraint in (7.14). Then, the Lagrangian function of problem ($P_{7,3}$) in (7.14) is given by [92]:

$$\mathcal{L} = \sum_{k \in \mathcal{K}} \mu_k \frac{\alpha_k}{x_k} + \lambda \left(\sum_{k \in \mathcal{K}} x_k - N \right). \tag{7.15}$$

The derivative of the Lagrangian function $\mathcal{L}$ in (7.15) equal to 0 leads to:

$$\frac{\delta \mathcal{L}}{\delta x_k} = -\frac{\mu_k \alpha_k}{x_k^2} + \lambda = 0, \tag{7.16}$$

such that the solution for x_k is:

$$x_k = \sqrt{\frac{\mu_k \alpha_k}{\lambda}}. \tag{7.17}$$

By including (7.17) into the constraint of problem ($P_{7,3}$) in (7.14), and setting the equality, we obtain the optimal value for λ:

$$\lambda = \left(\frac{\sum_{k \in \mathcal{K}} \sqrt{\mu_k \alpha_k}}{N} \right)^2. \tag{7.18}$$

Therefore, by combining (7.17) and (7.18), the optimal solution for the resource provisioning under orthogonal resource usage and minimum weighted sum of RU is:

$$x_k^* = \frac{\sqrt{\mu_k \alpha_k}}{\sum_{l \in \mathcal{K}} \sqrt{\mu_l \alpha_l}} N. \tag{7.19}$$

Similarly as in the minimum max RU case (see (7.13)), the optimal solution in (7.19) gives more frequency resources to those BSs experiencing higher α_k (see (7.8)).

7.4.3 Mapping the solution into real resources

The optimal resource provisioning has been found as a continuous distribution of the total available spectrum. However, practical systems dispose of an integer number of frequency resources to be allocated. So we should convert the optimal distribution into a discrete number of frequency resources per-BS satisfying the constraint in (7.9): $\sum_{k \in \mathcal{K}} x_k \leq N$, and then map it into real frequency resources. The round down (i.e. use floor function) would always satisfy the constraint. But any other rounding that satisfies the constraint is a valid one. Afterwards, mapping the number of frequency resources per-BS into real frequency resources is straightforward.

7.5 Graph-Based Resource Provisioning

In this section, we derive a long-term resource provisioning scheme when reuse of frequency resources among BSs that have non-overlapping coverage areas is permitted. Similar criterion as in Section 7.4 is adopted, but now we will impose orthogonality only among the sets of frequency resources assigned to BSs in an interference situation. To do so, we use information given by the *interference graph* (see Section 7.3) to formulate the problem and then exploit graph coloring to solve it.

7.5.1 Problem formulation and resolution

Based on the interference graph, we impose that neighbor BSs (those connected through an edge in the interference graph) cannot reuse frequency resources. This is done through the inclusion of proper constraints into the optimization problem. We define an orthogonality constraint for each maximal clique (see Definition 7.6 in Appendix 7.A) of the interference graph, where each constraint includes the resources of all the vertices that conform each maximal clique.

As an example, Fig. 7.3 shows the interference graph in a one-tier network composed of $|\mathcal{K}|{=}7$ BSs. In this case, the set of orthogonality constraints derived from the interference graph in Fig. 7.3, which will be represented by $\mathcal{C}$ in what follows, is given by:

$$\mathcal{C} = \left\{ \begin{array}{ll} x_1 + x_2 + x_4 \leq N, & x_2 + x_3 + x_4 \leq N, \\ x_3 + x_5 \leq N, & x_5 + x_6 + x_7 \leq N \end{array} \right\}. \tag{7.20}$$

This set of orthogonality constraints defines a necessary condition for an implementable solution to exist (i.e. any valid assignation of frequency resources that can be satisfied without conflicts in the interference graph will satisfy the constraint set). Moreover, as we will see in Section 7.5.2, said constraint set defines a necessary and sufficient condition for an implementable solution to
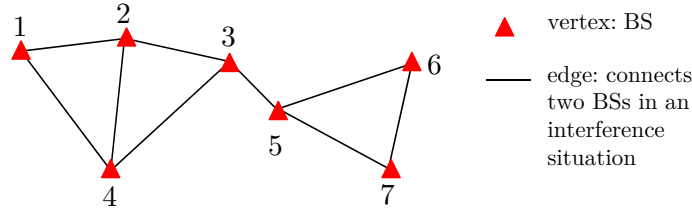
Figure 7.3: Interference graph for a network of $|\mathcal{K}|{=}7$ BSs.

exist and to be found if the interference graph has certain properties. Otherwise, the interference graph can be slightly modified by adding some edges to meet the properties and hence ensuring the sufficient condition.

According to the constraint set, we define the optimization problem by following the minimization of a function of the RU ($\rho_k{=}\frac{\alpha_k}{x_k}$) subject to the constraints extracted from the interference graph:

$$(\text{P}_{7,4}): \quad \underset{\{x_k\}}{\text{minimize}} \quad f\left(\left\{\frac{\alpha_k}{x_k}\right\}\right) \tag{7.21}$$
$$\text{subject to} \quad \mathcal{C},$$

where the objective function $f\left(.\right)$ depends on the optimization criterion:

$$f\left(\left\{\frac{\alpha_k}{x_k}\right\}\right) = \begin{cases} \max_k \left(\frac{\alpha_k}{x_k}\right) & \text{if max RU} \\ \sum_{k\in\mathcal{K}} \mu_k \frac{\alpha_k}{x_k} & \text{if weighted sum RU} \end{cases}. \tag{7.22}$$

Both max RU criterion and weighted sum of RU criterion have an objective function $f\left(\left\{\frac{\alpha_k}{x_k}\right\}\right)$ that is convex w.r.t. $\{x_k\}$. Then, as the constraint set $\mathcal{C}$ is linear (and hence convex) on the optimization variables, problem $(\text{P}_{7,4})$ in (7.21) is convex w.r.t. $\{x_k\}$. Although a closed-form solution for resource provisioning cannot be obtained, problem $(\text{P}_{7,4})$ in (7.21) can be solved in polynomial time using convex optimization tools as, for instance, interior point methods [92]. Let us denote by $\{x_k^*\}$ to the optimal solution to problem $(\text{P}_{7,4})$ in (7.21).

7.5.2   Mapping the solution into real resources

Once the optimal continuous distribution of spectrum to be assigned to each BS is obtained, $\{x_k^\star\}$, we should map it into specific frequency resources. First, rounding the optimal distribution into an integer number of frequency resources per BS is required. Any rounding within the constraint set $\mathcal{C}$ in (7.21) is a valid one. After the rounding, we should determine which are the specific frequency resources to be used by each BS, i.e. we should map the number of frequency resources per BS into real frequency resources. Multiple mappings may exist, but the key point is to guarantee that there exists at least one implementable mapping. To obtain the mapping, we define an extended graph and then use graph coloring to color it (where each color represents a resource).
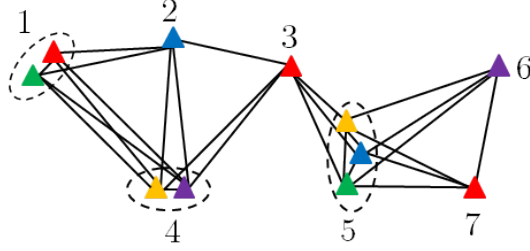
Figure 7.4: Extended graph of the interference graph in Fig. 7.3. $|\mathcal{K}|=7$ BSs, $N=5$ resources, $x_1^\star=2$, $x_2^\star=1$, $x_3^\star=1$, $x_4^\star=2$, $x_5^\star=3$, $x_6^\star=1$, $x_7^\star=1$.

The *extended graph* is built as follows:

- for each k-th BS, as many vertices as the number of frequency resources obtained from discretizing the optimal solution to problem ($P_{7,4}$) in (7.21) (i.e. $x_k^\star$) are included, and
- all the vertices corresponding to the k-th BS have edges to all vertices corresponding to all other BSs that were connected in the interference graph to the k-th BS.

Note that the extended graph is formed through the replication of vertices in the interference graph according to the discretized solution of problem ($P_{7,4}$) in (7.21). An example of the extended graph corresponding to the interference graph in Fig. 7.3 is shown in Fig. 7.4 assuming that $N=5$ frequency resources were available for optimization of problem ($P_{7,4}$) in (7.21) and that the obtained discretized solution was: $x_1^\star=2$, $x_2^\star=1$, $x_3^\star=1$, $x_4^\star=2$, $x_5^\star=3$, $x_6^\star=1$, $x_7^\star=1$. Then, we apply graph coloring over the extended graph.

Graph coloring is an NP-hard problem for arbitrary graphs. In addition, most of the coloring algorithms were proposed to color a graph under the assumption that there are enough colors to color the graph. When the number of colors is limited to N, *proper N-coloring* (see Definition 7.9 in Appendix 7.A) may not be possible. This is the reason why proper N-coloring cannot be ensured in recent resource provisioning schemes provided in the literature (either in the long-term [32] or in the short-term [170]): they use a graph that may not have a proper N-coloring due to the randomness in the traffic loads and in the network deployment, see [32, 170].

However, a well-known fact about graph coloring is that the chromatic number of a graph $\mathcal{X}(\mathcal{G})$ (see Definition 7.10 in Appendix 7.A) is lower bounded by the clique number of the graph $\omega(\mathcal{G})$ (see Definition 7.7 in Appendix 7.A): i.e. $\mathcal{X}(\mathcal{G}) \geq \omega(\mathcal{G})$, being the bound tight for perfect graphs [174, 175] (see Definition 7.12 in Appendix 7.A). In our case, the clique number of the extended graph $\mathcal{G}_E$ is exactly N: $\omega(\mathcal{G}_E)=N$, as it was imposed in the constraint set (see (7.20)). Therefore, as perfection is preserved through replication of vertices [175], if the interference graph is a perfect graph then the extended graph will be a perfect graph with $\mathcal{X}(\mathcal{G}_E)=\omega(\mathcal{G}_E)=N$ such that the extended graph will have a proper N-coloring. So, for perfect graphs, the proposed constraint set is a necessary and sufficient condition for an implementable solution to exist.

Note that the interference graph in Fig. 7.3 and its extended graph in Fig. 7.4, as well as the interference graph used for simulations in Fig. 7.6, are perfect graphs.

Lemma 7.1. *After determining the number of frequency resources required per BS by discretizing the optimal solution to problem* ($P_{7,4}$) *in* (7.21) *with N available frequency resources, proper N-coloring of the extended graph is possible (i.e. an implementable mapping into real frequency resources exists) if the interference graph is a perfect graph.*

Proof. The clique number of the extended graph is $\omega(\mathcal{G}_E){=}N$, as it is imposed by the constraints to problem ($\mathrm{P}_{7,4}$) in (7.21) where N frequency resources are available (see (7.20)). Under the assumption that the interference graph is a perfect graph (and, by replication, also the extended graph is), then: $\mathcal{X}(\mathcal{G}_E){=}N$. So proper N-coloring of the extended graph is possible.   ∎

Perfect graphs can be recognized in polynomial time [100]. There are many classes of perfect graphs, see [99] where up to 120 classes are described. Moreover, any graph can be transformed into a chordal graph (a class of perfect graphs, see Definition 7.14 in Appendix 7.A) by adding few edges such that the graph possesses no cycles of length ≥ 4 [176]. Therefore, in case the interference graph is not a perfect graph, we can transform it into a perfect graph, define the constraint set accordingly, and hence ensure the sufficient condition.

Finally, optimal coloring of perfect graphs can be obtained with Greedy algorithms in polynomial time [177, Sect. 9]. Therefore, an implementable solution does not only exist but can also be found for any conditions of the BS loads. For example, we can color the extended graph with $\mathcal{X}(\mathcal{G}_E){=}N$ colors by applying any of the well-known low-complexity algorithms in [101], [178], or [179].

7.6   Simulation Results

The scenario consists of a deployment of $|\mathcal{K}|{=}6$ outdoor SeNBs (that act as BSs), which are randomly placed within a circular area of 100 m radius with a minimum distance of 40 m among them. $|\mathcal{I}_k|$ users are randomly placed around each k-th BS in a concentric 40 m radius circle, with a minimum distance between users and BSs of 10 m. All BSs operate on the same carrier frequency at 2 GHz with 10 MHz bandwidth, where there are $N{=}50$ RBs (i.e. frequency resources) available. Path loss and shadowing models follow specifications in [63] for multi-cell pico scenario. Downlink transmission is evaluated. The antenna pattern is omnidirectional and the transmit power is 24 dBm at BS. Noise spectral density is -174 dBm/Hz.

As the objective of the following simulations is to show the benefits of the proposed long-term resource provisioning schemes when the load varies per BS, the number of users associated to each BS ($|\mathcal{I}_k|$) is set to $20, 16, 6, 24, 8, 15$ users, respectively. A deployment example is shown in Fig. 7.5. DL transmission is considered for simulations.

The non-full buffer FTP3 traffic model [61] is used for traffic generation, where packets for the same user arrive according to a Poisson process with arrival rate λ (in packets/s) and the packet length is fixed and equal to $L = 0.5$ Mbits (i.e. $L_k{=}L$, $\forall k$). Then, the packet arrival rate for the k-th BS is proportional to the number of users, i.e. $\lambda_k{=}\lambda|\mathcal{I}_k|$ in (7.2).

Results are averaged over 1000 random deployments of the users but, for simplicity, the BSs' positions are kept fixed. We assume a threshold distance equal to 80 m. Hence, the interference graph associated to the deployment in Fig. 7.5 is the one depicted in Fig. 7.6, from which the set of orthogonality constraints for the graph-based reuse schemes are:

$$\mathcal{C}{=}\left\{ \begin{array}{ll} x_1 + x_2 + x_6 \leq 50, & x_2 + x_4 \leq 50, \\ x_1 + x_5 \leq 50, \; x_4 + x_5 \leq 50, \; x_3 + x_4 \leq 50 \end{array} \right\}. \tag{7.23}$$

1000 frames of 10 ms are simulated, being the simulation time equal to 10 s. On each frame, the active users per BS (i.e. users with packets in the queues) are uniformly distributed among
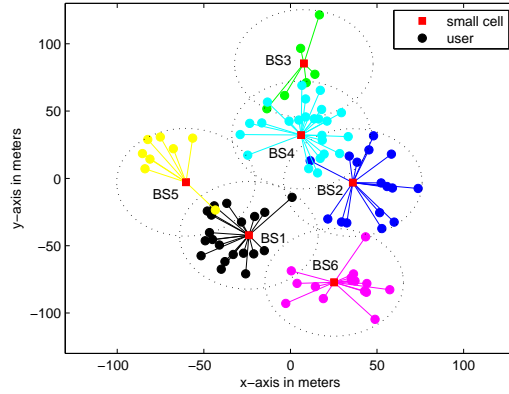
Figure 7.5: Example of network deployment used for simulations with $|\mathcal{K}|=6$ BSs. The number of users at each BS is: $20, 16, 6, 24, 8, 15$, respectively.
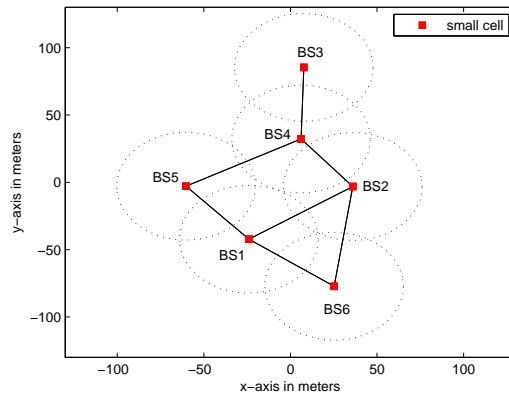

Figure 7.6: Interference graph associated to the network deployment in Fig. 7.5. The threshold distance is 80 m.

the available RBs. The number of available RBs per BS is determined in the long-term based on the different schemes.

The following techniques are evaluated:

- **orthogonal uniform**: the 50 RBs are orthogonally distributed among the 6 BSs in an almost uniform manner.

- **orthogonal maxRU**: the 50 RBs are orthogonally distributed among the 6 BSs according to the proposed resource provisioning in Section 7.4 to minimize the max RU (see $(\text{P}_{7,1})$ in (7.9)). The distribution follows the closed-form solution in (7.13).

- **orthogonal sumRU**: the 50 RBs are orthogonally distributed among the 6 BSs according to the proposed resource provisioning in Section 7.4 to minimize the sum of RU (see $(\text{P}_{7,3})$ in (7.14)). $\mu_k=1$ is used, $\forall k$. The distribution follows the closed-form solution in (7.19).

- **full reuse**: the 50 RBs can be employed by all the 6 BSs whenever they have packets to transmit. Interference is generated if two neighboring BSs transmit on the same resource.

- **frequency reuse 1/3**: frequency reuse scheme with reuse factor equal to 1/3, i.e. the 50 RBs are split into three disjoint sets (or subbands) and the neighboring BSs transmit on different subbands.

- **graph-based reuse maxRU**: the 50 RBs are distributed among the 6 BSs according to the proposed resource provisioning in Section 7.5 to minimize the max RU subject to the constraints imposed by the interference graph in Fig. 7.6 (see $(P_{7,4})$ in (7.21)).

- **graph-based reuse sumRU**: the 50 RBs are distributed among the 6 BSs according to the proposed resource provisioning in Section 7.5 to minimize the sum of RU subject to the constraints imposed by the interference graph in Fig. 7.6 (see $(P_{7,4})$ in (7.21)). $\mu_k=1$ is used, $\forall k$.

In all the proposed schemes based on RU optimization, i.e. 'orthogonal maxRU/sumRU' and 'graph-based reuse maxRU/sumRU', the average spectral efficiency is assumed to be equal for all BSs ($C_k=C, \forall k$) as all BSs (that correspond to SeNBs) dispose of the same power, users are uniformly distributed within the same area, and a round-robin scheduler is adopted. Then, as the packet arrival rate ($\lambda_k=\lambda|\mathcal{I}_k|$) and the packet length ($L_k=L$) are fixed, the ratio among the average traffic load and the average spectral efficiency is given by: $\alpha_k=|\mathcal{I}_k|\lambda L/C$ (see (7.2)).

Fig. 7.7.(a) shows the maximum RU versus λ (in packets/s). The maximum RU corresponds to the average over deployments of the maximum RU among the 6 BSs. It can be observed that 'graph-based reuse maxRU' reduces significantly the max RU as compared to all the other schemes, except for very low traffic loads ($\lambda=0.5$) where full reuse is better. 'orthogonal maxRU' reduces the max RU as compared to 'orthogonal uniform' and, as the traffic load (i.e. λ) increases, it also improves 'full reuse' and 'frequency reuse-3'. This is because i) in 'full reuse' there is an increased level of interference that cannot be controlled and ii) in 'frequency reuse-3' the resources are equally distributed, which make these schemes being saturated (in terms of occupancy) earlier than 'orthogonal maxRU' as λ increases. Therefore, resources are fairly allocated with the proposed schemes based on RU optimization because they allow distributing the resources by taking into account the per-BS traffic loads and spectral efficiencies such that all BSs have a similar RU. To conclude, the scheme that provides the lowest max RU and can work in a large range of traffic loads without being saturated is the proposed 'graph-based reuse maxRU', as it allows an efficient reuse and distribution of resources where strong interference conditions are avoided.

Fig. 7.7.(b) shows the sum of RU versus λ (in packets/s). The sum RU corresponds to the average over deployments of the sum of RU of the 6 BSs. Similar conclusions as in Fig. 7.7.(a) can be extracted, but now in terms of sum RU. The scheme that provides the lowest sum RU and can work in a large range of traffic loads without being saturated is the proposed 'graph-based reuse sumRU'.

Fig. 7.8 displays the total traffic served by each scheme and the total offered traffic (in Mbits/s) versus λ (in packets/s). All the orthogonal-based schemes ('orthogonal uniform' and 'orthogonal maxRU/sumRU') get saturated in terms of served traffic as λ increases because the number of resources per BS is low and all BSs are highly occupied for high values of λ. The remaining schemes have already not reached the maximum occupancy at all the BSs (although maybe in some BSs) with $\lambda=3$ packets/s. It can be observed that 'graph-based reuse maxRU/sumRU' allow serving all data traffic for low values of λ but, in addition, it allow serving the maximum quantity of the traffic as λ increases.

Fig. 7.9 depicts the mean packet delay (in s) versus λ (in packets/s). The packet delay is computed from the moment the packet arrives at the queue to the moment in which the transmission of the whole packet is completed. During the simulation time, only packets whose transmission is completed are used to compute the mean. It can be observed that the mean
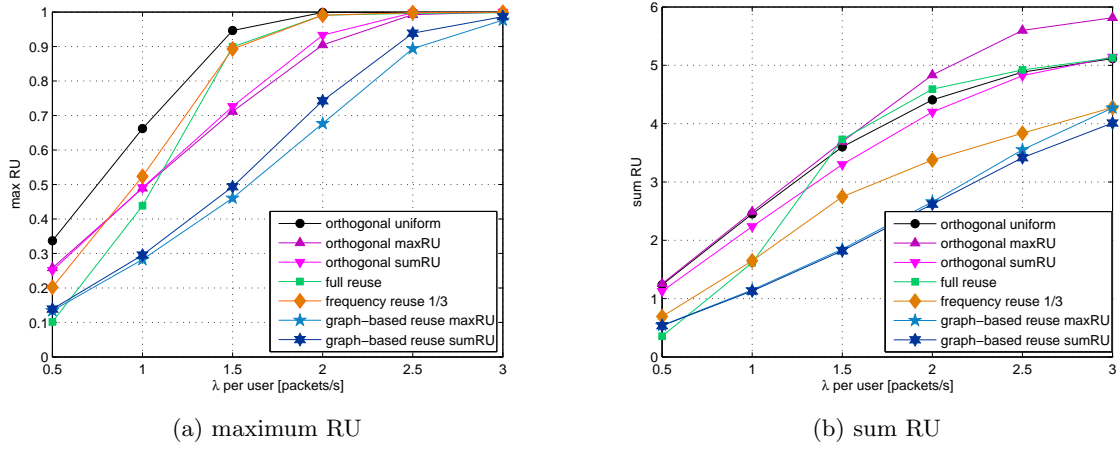
(a) maximum RU
(b) sum RU

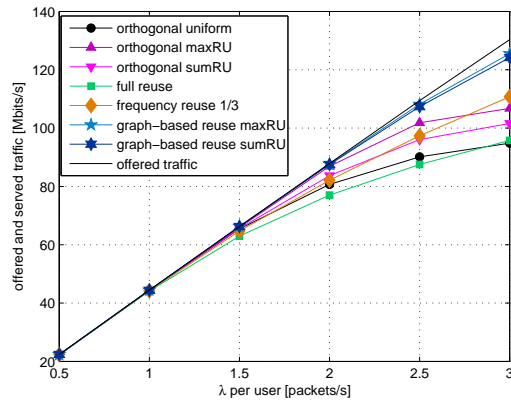Figure 7.7: RU-function vs. packet arrival rate per user (λ, in packets/s).

Figure 7.8: Served traffic and offered traffic (in Mbits/s) vs. packet arrival rate per user (λ, in packets/s).
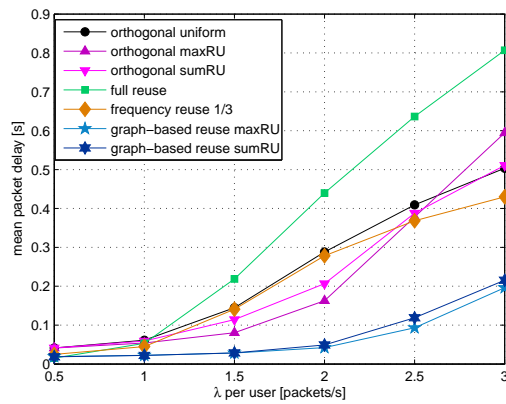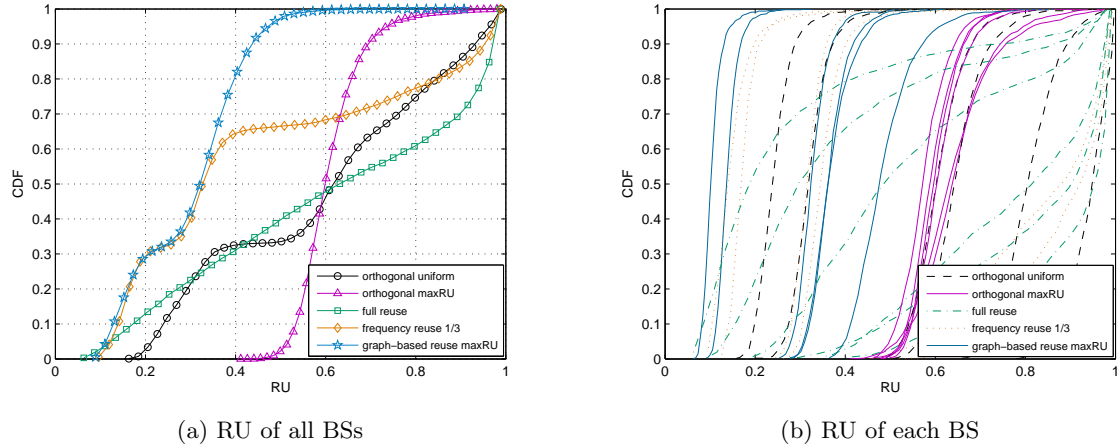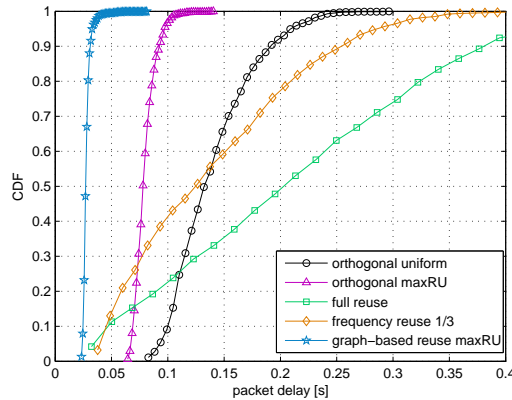

Figure 7.9: Mean packet delay (in s) vs. packet arrival rate per user (λ, in packets/s).

(a) RU of all BSs

(b) RU of each BS

Figure 7.10: CDF of RU for λ=1.5 packets/s.

Figure 7.11: CDF of packet delay (in s) for λ=1.5 packets/s.

packet delays are reduced with the proposed schemes for medium values of λ, and that 'graph-based reuse maxRU' scheme provides a significantly lower value of the mean packet delay for all the simulated packet arrival rates λ.

Therefore, although we have focused on the minimization of the occupancies (or RU) of all the BSs, it can be concluded from Fig. 7.8 and Fig. 7.9 that optimization of the RU is effectively translated into an increase of the served traffic and a reduction of the mean packet delay in a multi-cell scenario.

Fig. 7.10.(a) shows the cumulative distribution function (CDF) of the RU of all BSs in the network for λ=1.5 packets/s. It can be observed that the dispersion of RU values with 'orthogonal uniform', 'full reuse', and 'frequency reuse 1/3', is very large. On the contrary, the proposed 'orthogonal maxRU' and 'graph-based reuse maxRU' allow having the RU of all BSs in a reduced range of values, i.e. the variability in the occupancy of the BSs is much low thanks to the efficient distribution of radio resources.

Fig. 7.10.(b) displays the CDF of the RU of each of the 6 BSs separately for λ=1.5 packets/s. Again, the dispersion of the RU values can be observed to be different with each scheme, and to

be more concentrated with 'orthogonal maxRU' and 'graph-based reuse maxRU' schemes. Also, it is important to note that with 'orthogonal maxRU' all BSs have a similar RU, because equal RU for all BSs was the optimal solution to problem ($P_{7,1}$) in (7.9). In contrast, in 'graph-based reuse maxRU', the RU values of the different BSs are not as concentrated as in 'orthogonal maxRU' because the optimal solution for the problem ($P_{7,4}$) in (7.21) does not imply all BSs having the same RU. It can also be observed that 'graph-based reuse maxRU' is the scheme that has a lowest max RU.

Fig. 7.11 shows the CDF of the packet delay for λ=1.5 packets/s. We can observe that the packet delays are significantly reduced with the proposed schemes, and specially with 'graph-based reuse maxRU'. Also, the dispersion of the packet delay values is much low in the proposed schemes than in the static resource allocation approaches.

7.7 Conclusions

This chapter proposes long-term frequency resource provisioning schemes for dense OFDMA-based networks. We derive procedures to allocate frequency resources among different TDD BSs by following the optimization of the resource utilization factors, or occupancies, of the BSs (both in DL and UL transmissions). The mean traffic load per BS in DL and UL transmit directions as well as the average effective DL/UL capacity per BS are taken into account in the problem formulation. The optimization is performed in the long-term, hence being independent of the users to be served at a specific time instant but dynamic enough to follow significant variations of the average DL/UL traffic load per BS. When orthogonal resource usage among BSs is assumed, the optimal distribution of resources to be used per BS is obtained in closed-form. In case of allowing to reuse resources among BSs, we present a procedure in which the optimization problem is set according to the interference graph, then it is optimally solved using convex optimization tools and discretized to get the number of resources per-BS, and finally graph coloring is applied to map the obtained solution into real resources.

Simulation results show that the maximum RU and the sum of RU are significantly reduced with the proposed graph-based reuse scheme, as it allows an efficient reuse of resources while avoiding strong interference conditions. In addition, it is shown that the optimization of the RU of all BSs translates into an improvement of the total amount of served traffic and an effective reduction of the mean packet delay for all the simulated traffic load conditions.

Appendices

7.A   Preliminaries for Graph Theory

Graph theory is a wide area of discrete mathematics, for which a complete reference can be found in [102]. This section formalizes a few concepts about graph theory, and specially about graph coloring.

Some basic definitions are presented next.

Definition 7.1. *An undirected graph $\mathcal{G}$ is a mathematical structure consisting of an ordered pair $\mathcal{G}=(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the finite set of elements called vertices and $\mathcal{E}$ is a finite unordered pairs of vertices called edges.*

Definition 7.2. *Two vertices $u, v \in \mathcal{V}$ are called adjacent if $\{u, v\} \in \mathcal{E}$. In other words, two vertices are adjacent if there is a line, the edge, connecting them.*

Definition 7.3 (Vertex degree)**.** *The degree of a vertex $v \in \mathcal{V}$, denoted by $\deg(v)$, is the number of edges incident to vertex v.*

Definition 7.4 (Maximum vertex degree)**.** *The degree of a graph $\mathcal{G}$, denoted by $\Delta(\mathcal{G})$, is the maximum degree of a vertex in $\mathcal{G}$, i.e. $\Delta(\mathcal{G})=\max_v \deg(v)$.*

Definition 7.5 (Clique)**.** *A clique of a graph $\mathcal{G}$ is a complete subgraph of $\mathcal{G}$, i.e. a subset of $\mathcal{V}$ such that for every two vertices there exists an edge connecting the two.*

Definition 7.6 (Maximal Clique)**.** *A maximal clique of a graph $\mathcal{G}$ is a complete subgraph of $\mathcal{G}$ that cannot be extended by including one more adjacent vertex, meaning it is not a subset of a larger complete subgraph of $\mathcal{G}$.*

Definition 7.7 (Clique number)**.** *The clique number of a graph $\mathcal{G}$, denoted by $\omega(\mathcal{G})$, is the cardinality of the largest maximal clique of the graph.*

The degree ($\Delta(\mathcal{G})$) and the clique number ($\omega(\mathcal{G})$) of a graph are two important measures. The degree can be easily computed from the connections of the vertices. Algorithms to find all the maximal cliques of a graph are available in the literature (e.g. in [180]), from which the clique number can be determined.

Now, concepts about graph coloring (in the sense of vertex coloring) are presented[27].

---

[27]Graph coloring can refer either to edge coloring or to vertex coloring. Within the Ph.D. dissertation, graph coloring is used to refer to vertex coloring.

Vertex coloring is an assignment of labels or colors to each vertex of a graph such that no edge connects two identically colored vertices. The most common type of vertex coloring seeks to minimize the number of colors for a given graph. Such a coloring is known as a minimum vertex coloring, and the minimum number of colors which with the vertices of a graph may be colored is called the chromatic number, denoted by $\mathcal{X}(\mathcal{G})$, as introduced in the following.

Definition 7.8 (Proper Coloring). *A proper vertex coloring (or proper coloring) of a graph is an assignment of colors to the vertices of the graph such that no adjacent vertices have the same color.*

Definition 7.9 (Proper N-coloring). *A proper N-coloring of a graph is an assignment of colors to the vertices of the graph such that no adjacent vertices have the same color when N colors are available.*

Definition 7.10 (Chromatic number). *The chromatic number of a graph $\mathcal{G}$, denoted by $\mathcal{X}(\mathcal{G})$, is the minimum number of different colors required for a proper coloring of the graph.*

Therefore, every graph has a proper N-coloring for $N \geq \mathcal{X}(\mathcal{G})$.

Graph coloring is an NP-hard problem for arbitrary graphs. In addition, most of the vertex coloring algorithms were proposed to color a graph under the assumption that there are enough colors to color the graph. When the number of colors is limited to N, *proper N-coloring* may not be possible (i.e. it might happen that $N < \mathcal{X}(\mathcal{G})$).

However, a well-known fact about graph coloring is that the chromatic number $\mathcal{X}(\mathcal{G})$ is lower and upper bounded as stated in the following theorem.

Theorem 7.1 (Bounds on the chromatic number). *Every graph can be properly colored with one more color than the maximum vertex degree and no less than the clique number, i.e.:*

$$\omega(\mathcal{G}) \leq \mathcal{X}(\mathcal{G}) \leq \Delta(\mathcal{G}) + 1 \tag{7.24}$$

The upper bound in (7.24) comes from the Greedy coloring algorithm [177], which shows that every graph can be colored with one more color than the maximum vertex degree. The upper bound in (7.24) is in fact improved by the Brook's theorem, which states that $\mathcal{X}(\mathcal{G}) \leq \Delta(\mathcal{G})$ unless $\mathcal{G}$ is a complete graph or an odd cycle.

The lower bound in (7.24) arises directly from the fact that if a graph has a maximum clique of size $\omega(\mathcal{G})$ then at least $\omega(\mathcal{G})$ colors are required to color the vertices of that graph. The lower bound in (7.24) is tight for perfect graphs [174, 175] (which are also known as Berge graphs), for which we pay a special attention.

Definition 7.11. *An induced subgraph of a graph $\mathcal{G}$ is another graph, formed from a subset of the vertices of $\mathcal{G}$ and all of the edges connecting pairs of vertices in that subset.*

Definition 7.12 ([175] - Perfect graph). *A perfect graph is a graph in which the chromatic number of every induced subgraph, including the graph itself, equals the size of the largest maximal clique of that subgraph.*

Perfect graphs have many important properties. In all perfect graphs, the graph coloring problem and the maximum clique problem can all be solved in polynomial time [177]. Another important property is that perfection is preserved through the replication of vertices in the

graph [175] (i.e. for a perfect graph, we can replicate any vertex of the graph and the resulting graph is also a perfect graph).

For many years the complexity of recognizing perfect graphs remained open. Finally, subsequent to the proof of the strong perfect graph theorem (presented in what follows), a polynomial time algorithm was discovered to recognize perfect graphs [100].

Definition 7.13. *The complement or inverse of a graph $\mathcal{G}$ is a graph $\mathcal{H}$ on the same vertices such that two distinct vertices of $\mathcal{H}$ are adjacent if and only if they are not adjacent in $\mathcal{G}$.*

Theorem 7.2 ([174] - Strong perfect graph theorem)**.** *A graph $\mathcal{G}$ is a perfect graph if and only if neither $\mathcal{G}$ nor its complement have an odd-length induced cycle of length 5 or more.*

There are many classes of perfect graphs, including many important families of graphs. Up to 120 classes are described in [99]. Some important classes are: bipartite graphs, chordal graphs (which include interval graphs as subclasses), line graphs, and comparability graphs.

Moreover, any graph can be transformed into a chordal graph (and, hence, a perfect graph) by adding some edges such that the graph possesses no cycles of length 4 or more [176, 181].

Definition 7.14 (Chordal graph)**.** *A chordal graph is a graph in which every cycle of length four and greater has a cycle chord, i.e. an edge.*

Chapter 8

Conclusions

This Ph.D. dissertation has focused on interference management for dense MIMO cellular networks. In particular, we have investigated:

- advanced signaling schemes (subsuming improper Gaussian signaling (IGS) and proper Gaussian signaling (PGS)) in Chapter 2,
- transmit coordination strategies for inter-cell interference management in Chapters 3-4-5, including transmit coordination and transmit cooperation as well as the coexistence of linear and widely linear transceivers, and
- traffic-aware and interference-aware resource management techniques for flexible duplexing in Chapters 6-7, comprising short-term and long-term optimizations.

Advanced signaling schemes

The first part of this work (i.e. Chapter 2) analyzes advanced signaling schemes for interference-limited wireless scenarios by focusing on different statistical characterizations of the signals. In this regard, majorization theory has been exploited to formally quantify the strict superiority of IGS over PGS for the simple MIMO point-to-point channel with interference (P2P-I). Strictly positive bounds have been derived for the achievable rate and the mean square error (MSE) improvement when using IGS as compared to PGS. It has been observed that the achievable rate and the MSE gains with IGS are larger as the interference level increases.

Based on these results, the potential benefits of IGS have been applied to different MIMO interference-limited scenarios through an efficient design of widely linear transceivers that does generate IGS. First, a practical IGS-based scheme is derived for the MIMO Z-interference channel (Z-IC), which does not require knowledge of the interference channel but allows trading on system fairness and sum-rate performance. In this case, it is again observed that the larger the interference level is the larger the gain is obtained with IGS-based schemes.

Second, the benefits of IGS are applied to the downlink (DL) transmission of co-channel heterogeneous cellular networks (HCNs) where macrocells and smallcells coexist and multiple MIMO Z-IC (with high levels of interference) appear. In this scenario, 3GPP-compliant simulations show that user throughputs are significantly improved with the proposed IGS-based scheme as compared to conventional time-sharing solutions, PGS-based schemes, and eICIC techniques, due to the fact that the proposed scheme can cope with a full reuse of the frequency bands and

time slots while providing enough flexibility to combat the predominant cross-tier interference generated from a macrocell to the users served by smallcells in co-channel HCNs. The proposed IGS-based scheme does not only avoid the estimation of the interference channels but also avoids optimization of the muting ratio and addressing the on/off switching of the macrocells (needed, e.g., for eICIC and time-sharing). It is therefore highly recommended for systems where a single strong interference appears such that the MIMO Z-IC can be identified.

Future research lines include:

- finding tighter bounds for the achievable rate and MSE improvement in the MIMO P2P-I when using IGS as compared to PGS. Positive bounds have been derived so far, but they have been shown to be loose and larger differences between IGS and PGS usage are observed through simulations (see Fig. 2.4). In this sense, majorization theory for strongly Schur-convex and strongly Schur-concave functions [130] should be followed as well, but the strong-convexity and strong-concavity properties of the achievable rate and MSE functions should be deeply characterized.

- the derivation of the optimal precoding design for weighted sum-rate maximization in the MIMO Z-IC when IGS is adopted. In this work, a practical and heuristic design has been proposed for the MIMO Z-IC (which does already outperform PGS-based solutions thanks to the use of IGS), but the optimal scheme remains unknown. The optimal scheme for the SISO Z-IC with IGS has been derived in [122]. Extension to MIMO systems could be investigated.

- analyzing the application of IGS to more complex interference-limited scenarios, like the multi-cell multi-user MIMO scenario where multiple potential interfering BSs might appear. Also, application to other interference-limited scenarios is highly relevant as well. There are already some works available in the literature that show the superiority of IGS for underlay cognitive radio systems (see [114–117]) and full-duplex systems (see [118,119]).

Transmit coordination strategies

The second part of the Ph.D. dissertation aims at deriving transmit coordination procedures to manage interference in dense networks when full reuse of time and frequency resources among BSs densely deployed in the network is adopted. To do so, a wide deployment of MIMO systems has been leveraged to develop advanced multi-antenna signal processing techniques.

In Chapter 3, decentralized and coordinated precoding designs have been derived to manage interference in the DL of a multi-cell multi-user MIMO system. The proposed procedures exploit channel reciprocity in TDD systems so as to estimate the interference-cost matrix (which allows managing inter-cell interference) based on the covariance matrix of the received interference-plus-noise signal in a properly configured uplink (UL) transmission. Thus, the interference-cost matrix is obtained over-the-air without requiring estimation of the interfering channel matrices. Thanks to that, the impact of imperfect estimation of the interfering channel matrices and the stringent backhaul requirements for transmit coordination in dense networks are alleviated, only at the cost of a synchronized DL-UL protocol that has been defined. In this sense, decentralized designs have been developed assuming either perfect or imperfect channel knowledge of the direct channel matrices.

The monotonic convergence of the proposed decentralized and iterative procedure is demonstrated when all parameters are perfectly acquired and simultaneous per-BS optimizations are

performed. This proof is highly relevant for practical implementation issues, as when only one iteration (or a few) of the algorithm can be performed a performance improvement with parallel and independent per-BS optimizations is guaranteed.

Methods to implement the proposed technique by following 3GPP-based mechanisms have been investigated. It has been recommended to use a coordinated transmission of the already defined sounding reference signals (SRS) in LTE so as to estimate the interference-cost matrices. Finally, robustness under imperfect acquisition of the interference-cost matrices (either due to the use of non-orthogonal pilot signals or due to non-ideal TDD channel reciprocity) and imperfect estimation of the direct channel matrices has been shown through simulations.

Next, in Chapter 4, the concepts presented in Chapter 3 have been extrapolated to manage interference in the DL transmission of a multi-cell multi-user MIMO system where transmit cooperation among a small number of BSs is allowed. We have derived procedures for the joint design of BSs clustering and transmit precoding under a user-centric approach in which a cluster of BSs is defined per user (i.e. UE-centric clustering). The problem has been properly formulated so as to allow obtaining small cluster sizes by including a penalizing term in the objective function that allows taking out of the cluster those BSs using low power. Differently from Chapter 3, for decentralized implementation, not only the interference-cost matrices are acquired but also the coupling in the transmit power among BSs belonging to different clusters has been decoupled.

It is important to recall that the proposed procedure could also be used to optimize precoding design for CoMP-JT under a BS-centric clustering scheme. In this sense, UE-centric clustering and BS-centric clustering schemes have been compared through simulations. Results have shown the suitable applicability of UE-centric clustering schemes for dense and irregular deployments of smallcells, since a significant improvement is obtained in the 5%-tile per-user rate. Also, an effective reduction of the cluster size is achieved with the proposed design.

To finalize the study, the complementarities of CoMP-JT and the use of multiple antennas at BSs have been analyzed. Simulation results in 3GPP-compliant dense smallcell network scenarios show that 4 smallcells are enough for transmit cooperation in most of the cases, otherwise increasing the cluster size provides low performance benefits. Whether the larger gains are obtained by increasing the cluster size for CoMP-JT or by using additional transmit antennas at BSs depends on the scenario (symmetric or non-symmetric) and on the channel conditions (high/medium/low antenna correlation). However, it is worth recalling that the potential complementarities of using CoMP-JT and adding multiple antennas per BS are observed in all cases.

Then, in Chapter 5, the procedures derived in Chapter 3 have been extended to the case of an heterogeneous (linear and widely linear) MIMO transceiver deployment in multi-cell scenarios. The heterogeneous scenario arises naturally due to the use of different signaling schemes: IGS and PGS (which have been previously analyzed in Chapter 2). In this sense, coordinated precoding designs are proposed for multi-cell MIMO systems, which do jointly exploit the benefits of transmit coordination and the use of IGS. Both centralized and decentralized implementations have been investigated.

Simulation results show that the gains from the heterogeneity of linear and widely linear transceivers in terms of sum-rate are significant in two different non-exclusive conditions: when the interference level increases (either due to a large number of interfering BSs or owing to the presence of a strong interfering BS) or when the number of transmit/receive antennas is low. If the interference level is not significant or dimensions are already provided by adding multiple

antennas at transmitters and receivers, the proposed coordinated transceiver scheme leads to a PGS-based solution.

Let us recall that weighted sum-rate maximization in the MIMO interference broadcast channel (which models a multi-cell multi-user MIMO system) is a non-convex (and NP-hard) problem. In this sense, all the approaches for interference management presented in this part of the doctoral thesis are not guaranteed to reach the global maximum weighted sum-rate solution, but rather a local optimum is achieved. Accordingly, the weighted sum-rate maximization with linear transceivers in MIMO interference broadcast channels is still an open problem.

Future work involves:

- the application of the proposed schemes to FDD systems, whereby transformations in the frequency domain should be derived. Frequency-domain transformations have been studied in [182, 183], and should be redefined to allow estimation of the interference-cost matrix through an UL transmission that is carried out in a different band.

- investigating larger levels of decentralization for UE-centric clustering and precoding design in cooperative-based transmissions (i.e. CoMP-JT). In Chapter 4, the decentralization is achieved at the cluster-level (i.e. for a small set of BSs) assuming knowledge (at the BS master of the cluster) of the channel matrices from the BSs in the cluster towards the user. However, it would be interesting to perform a decentralization at the BS-level when only local channel matrix knowledge is available (i.e. from the BS towards the user). This way, processing among cooperative BSs that belong to the same cluster could be decoupled. Such level of decentralization corresponds to the worst-case scenario for CoMP-JT, in which only data sharing among BSs is assumed. Its investigation has already been started in [184] for given cluster sizes. Extensions of the strategies presented in this work could be used to find joint optimal clustering and precoding design for the fully decentralized CoMP-JT case.

- including per-BS backhaul rate constraints into the problem formulation for joint clustering and precoding design in cooperative-based transmissions (i.e. CoMP-JT). The resulting optimization problem is very difficult to deal with. It has been addressed in [82] whereby, after applying approximations of the norm function, the per-BS backhaul rate constraints were approximated by a fixed value that was iteratively updated. Therefore, gap for further improvement seems to be available.

Traffic-aware duplexing techniques

The last part of this thesis focuses on interference-awareness and traffic-awareness radio resource allocation, where not only interference but also traffic load and traffic asymmetry conditions in dense networks are taken into account for resource management. Two different optimization time frames are followed in two separated chapters. For short-term management, the deployment of MIMO systems is capitalized to develop advanced coordinated strategies. For long-term management, graph coloring is exploited to distribute resources among densely deployed BSs.

In Chapter 6, short-term optimization procedures have been derived for flexible duplexing in MIMO TDD systems when a full reuse of the spectrum is assumed and a short-length single-direction frame structure is adopted. In this sense, the joint design of transmit precoding, user scheduling and transmit direction selection is investigated. The concepts for interference management presented in Chapter 3 have been used, but the whole problem formulation has

been modified to include DL/UL traffic asymmetry conditions among BSs and to address the new types of interference that appear under flexible duplexing. To solve the general problem, decomposition methods (including stochastic approximation recursions and the interference-cost concept) have been applied. The derived approach provides high flexibility to match the average DL/UL traffic asymmetries at each BS while managing interference at the short-term, but assumes knowledge of all channel matrices.

Simulation results show gains in DL and UL average rates for different traffic asymmetries, network densities and user densities as compared to existing schemes for dynamic TDD. But the gains are larger for systems with a low number of antennas because they have less flexibility for interference management through the coordination of transmit/receive spatial filters and hence larger gains can be obtained by jointly optimizing the transmit direction.

Interesting future work comprises:

- the investigation of the multi-user access (i.e. multiple users being simultaneously served at every frame either in DL or in UL), whereby user grouping strategies should be optimized together with the precoding design and the transmit direction selection. To do so, the constraint in (6.4) should be modified (e.g. by setting a maximum number of simultaneous users that the BS can serve) and the system model should be reformulated in order to include intra-cell interference.

- the application of the proposed approaches to OFDMA-based networks, in which users from the same BS are assigned to orthogonal resources at each frame (and each frame can be either for DL or for UL). In this case, a new constraint should be included in the problem formulation in (6.13), in a way such that a single transmit direction (either DL or UL) is chosen for all frequency resources at every BS on each frame.

- the inclusion of per-user DL/UL minimum average rate constraints in (6.9) to allow fulfilling quality of service requirements of the users. In this sense, one should investigate how these constraints translate into instantaneous per-user DL/UL rate constraints to be included in the decomposed per-frame problem in (6.13).

In Chapter 7, graph coloring has been exploited to derive long-term resource provisioning schemes for OFDMA-based networks. We have focused on long-term resource management with the objective of distributing frequency resources among BSs densely deployed in the network according to their average traffic load and average traffic asymmetry conditions while avoiding strong interference conditions. So, the distribution is independent of the specific users to be served at a specific time instant. In this sense, a novel approach based on optimization of the resource utilization (RU), or occupancies, of the BSs is proposed. The proposed approach is based on solving an RU-based optimization problem and then uses graph coloring to map the obtained result into real frequency resources. Differently from previous works, the proposed approach ensures that the application of graph coloring will provide an implementable solution for any conditions of the traffic loads.

In addition to improving load distribution and enhancing RU-based network metrics, simulation results have shown that the optimization of the RU factors of all BSs is translated into an improvement of the total amount of served traffic and an effective reduction of the mean packet delay, since resources are fairly distributed among BSs according to their traffic loads and spectral efficiencies. The proposed approaches are useful for dense, irregular, and interfered wireless networks, in which neither the conventional long-term resource provisioning schemes nor the full frequency reuse (without short-term coordination) are shown to be efficient.

Potential research lines include:

- the use of power control in frequency-domain (i.e. per frequency resource, which can be related to resource blocks, subbands, or component carriers) for the RU-based long-term resource provisioning schemes. Note that in Chapter 7 we have assumed an spectral power mask per frequency resource, such that the average spectral efficiencies in DL and UL do not depend on the number of resources assigned per BS. However, under power control, the average spectral efficiencies in DL and UL would depend on the number of actual resources given to each BS as well as on the power distribution. To allow it, the RU factors presented in (7.1)-(7.2) should be modified in order to include the DL/UL transmit powers as optimization variables within the DL/UL average spectral efficiencies, and the adequate problem should be formulated and solved.

- analyzing graph-based resource provisioning schemes (as the ones proposed in Section 7.5) where the spatial reuse factor is further improved and, at the same time, inter-cell interference is controlled. This could be done, for instance, by using different power levels per frequency resource and defining different interference-graphs according to the established power levels. This way, a BS could reuse a frequency resource employed by a neighbor BS provided that a low power level is used on that resource.

- investigating the decentralized implementation of the graph-based resource provisioning schemes proposed in Section 7.5. Both the RU-based optimization problem in (7.21) and the vertex-coloring algorithm are performed in a centralized fashion. But, as we move towards dense networks, the decentralized implementation of resource allocation and interference management problems (or, at least, hybrid solutions between centralized and decentralized) is receiving much attention. Therefore, it would be interesting to find decentralized solutions to both: the RU-based optimization problem and the graph coloring. Decentralized implementation of graph coloring algorithms can be found, e.g., in [185].

Bibliography

[1] P. Bhat *et al.*, "LTE-Advanced: an operator perspective," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 104–114, Feb. 2012.

[2] 3GPP Long Term Evolution (LTE). [Online]. Available: www.3gpp.org/.

[3] F. Boccardi *et al.*, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[4] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[5] I. Hwang, B. Song, and S. S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 20–27, Jun. 2013.

[6] N. Bhushan *et al.*, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82–89, Feb. 2014.

[7] R. Zhang *et al.*, "LTE-unlicensed: the future of spectrum aggregation for cellular networks," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 150–159, Jun. 2016.

[8] T. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: it will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[9] E. G. Larsson *et al.*, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[10] G. C. Alexandropoulos *et al.*, "Advanced coordinated beamforming for the downlink of future LTE cellular networks," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 54–60, Jul. 2016.

[11] V. Jungnickel *et al.*, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, May 2014.

[12] V. Fernandez *et al.*, "Improving dense network performance through centralized scheduling and interference coordination," *IEEE Trans. Vehicular Technology*, Sep. 2016.

[13] A. Khandekar *et al.*, "LTE-Advanced: heterogeneous networks," in *IEEE European Wireless Conference*, Apr. 2010.

[14] T. Taleb and A. Kunz, "Machine type communications in 3GPP networks: potential, challenges and solutions," *IEEE Commun. Magazine*, vol. 50, no. 3, pp. 178–184, 2012.

[15] H. Shariatmadari *et al.*, "Machine-type communications: current status and future perspectives toward 5G systems," *IEEE Commun. Magazine*, vol. 53, no. 9, pp. 10–17, Sep. 2015.

[16] A. Damnjanovic *et al.*, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 10–21, Jun. 2011.

[17] 3rd Generation Partnership Project (3GPP). [Online]. Available: www.3gpp.org/.

[18] 3GPP Universal Mobile Telecommunications System (UMTS). [Online]. Available: www.3gpp.org/.

[19] L. Wan, M. Zhou, and R. Wen, "Evolving LTE with flexible duplex," *IEEE Global Commun. Conf. Workshops*, pp. 49–54, Dec. 2013.

[20] M. Feng *et al.*, "Cooperative small cell networks: high capacity for hotspots with interference mitigation," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 108–116, Dec. 2014.

[21] X. Ge *et al.*, "5G ultra-dense cellular networks," *IEEE Wireless Commun.*, vol. 23, no. 1, pp. 72–76, Feb. 2014.

[22] 3GPP, *Feasibility study for further advancements for E-UTRA (LTE-Advanced)*, TR 36.912, Release 11, v11.0.0, Sep. 2012.

[23] ——, *Evaluation assumptions for small cell enhancements*, R1-130856, Huawei, HiSilicon.

[24] Accelleran, "The essential importance of LTE TDD for small cell deployments," Jul. 2013.

[25] Z. Shen *et al.*, "Dynamic uplink-downlink configuration and interference management in TD-LTE," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 51–59, Nov. 2012.

[26] IEEE 802.16 Task Group m. [Online]. Available: www.ieee802.org/16/tgm/.

[27] Qualcomm, "5G waveform and multiple access techniques," Nov. 2015.

[28] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireless Communications*. Cambridge (UK), Cambridge University Press, 2003.

[29] 3GPP, *Study on network-assisted interference cancellation and suppression (NAICS) for LTE*, TR 36.866, Release 12, v12.0.0, Mar. 2014.

[30] T. D. Novlan *et al.*, "Analytical evaluation of fractional frequency reuse for OFDMA cellular networks," *IEEE Trans. on Wireless Commun.*, vol. 10, no. 12, pp. 4294–4305, Dec. 2011.

[31] N. Saquib, E. Hossain, and D. I. Kim, "Fractional frequency reuse for interference management in LTE-Advanced HetNets," *IEEE Trans. on Wireless Commun.*, pp. 113–122, Apr. 2013.

[32] R. Y. Chang *et al.*, "A graph approach to dynamic fractional frequency reuse (FFR) in multi-cell OFDMA networks," *IEEE Int. Conf. on Commun.*, pp. 1–6, Jun. 2009.

[33] Q. H. Spencer *et al.*, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[34] E. G. Larsson and E. A. Jorswieck, "Competition versus collaboration on the MISO interference channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1059–1069, Sep. 2008.

[35] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[36] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, no. 4, pp. 774–780, Aug. 1972.

[37] A. Goldsmith *et al.*, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[38] J. Lee *et al.*, "Coordinated multipoint transmission and reception in LTE-Advanced systems," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 44–50, Nov. 2012.

[39] M. Sadek, A. Tarighat, and A. H. Sayed, "A leakage-based precoding scheme for downlink mulit-user MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1711–1721, May 2007.

[40] E. A. Jorswieck, E. G. Larsson, and D. Danev, "Complete characterization of the pareto boundary for the MISO interference channel," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5292–5296, Oct. 2008.

[41] R. Zhang and S. Cui, "Cooperative interference management with MISO beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5450–5458, Oct. 2010.

[42] R. Mochaourab and E. A. Jorswieck, "Optimal beamforming in interference networks with perfect local channel information," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1128–1141, Mar. 2011.

[43] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3309–3322, Jun. 2011.

[44] D. Gesbert *et al.*, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[45] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 56–61, Aug. 2006.

[46] A. S. Hamza *et al.*, "A survey on inter-cell interference coordination techniques in OFDMA-based cellular networks," *IEEE Commun. Surveys & Tut.*, vol. 15, no. 4, pp. 1642–1670, Aug. 2013.

[47] R. Zhang, "Cooperative multi-cell block diagonalization with per-base-station power constraints," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1435–1445, Dec. 2010.

[48] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[49] P. Viswanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[50] P. Marsch and G. P. Fettweis, *Coordinated multi-point in mobile communications: from theory to practice.*   Cambridge University Press, 2011.

[51] R. Irmer *et al.*, "Coordinated multipoint: concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 28, no. 9, pp. 1380–1408, Feb. 2011.

[52] S. A. Jafar, "Interference alignment: a new look at signal dimensions in a communication network," *Foundations and Trends in Commun. and Inf. Theory*, vol. 7, no. 1, pp. 1–134, Jun. 2010.

[53] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[54] A. Motahari *et al.*, "Real interference alignment: exploiting the potential of single antenna systems," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4799–4810, Aug. 2014.

[55] M. Torrellas *et al.*, "The DoF of the 3-user (p,p+1) MIMO interference channel," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3842–3853, Nov. 2014.

[56] B. Soret *et al.*, "Multicell cooperation for LTE-Advanced heterogeneous network scenarios," *IEEE Wireless Commun.*, vol. 20, no. 1, pp. 27–34, Feb. 2013.

[57] D. López-Pérez, X. Chu, and I. Guven, "On the expanded region of picocells in heterogeneous networks," *IEEE J. Sel. Topics in Signal Process.*, vol. 6, no. 3, pp. 281–294, Jun. 2012.

[58] B. Soret *et al.*, "Dynamic enhanced intercell interference coordination for realistic networks," *IEEE Trans. Vehicular Technology*, vol. 65, no. 7, pp. 5551–5562, Jul. 2016.

[59] J. Zhang *et al.*, "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, p. 19101921, Apr. 2009.

[60] M. Hong *et al.*, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 226–240, Feb. 2013.

[61] 3GPP, *Further advancements for E-UTRA physical layer aspects*, TR 36.814, Release 9, v9.0.0, Mar. 2010.

[62] ——, *Coordinated multi-point operation for LTE physical layer aspects*, TR 36.819, Release 11, v11.2.0, Sep. 2013.

[63] ——, *Further enhancements to LTE time division duplex (TDD) for downlink-uplink (DL-UL) interference management and traffic adaptation*, TR 36.828, Release 11, v11.0.0, Jun. 2012.

[64] A. Lozano, J. G. Andrews, and R. W. Heath, "Spectral efficiency limits in pilot-assisted cooperative communications," *IEEE Int. Symp. on Inf. Theory Proceedings*, pp. 1132–1136, Jul. 2012.

[65] P. J. Schreier and L. L. Scharf, *Statistical signal processing of complex-valued data: the theory of er and noncircular signals.* Cambridge (UK), Cambridge University Press, 2010.

[66] Y. C. Yoon and H. Leib, "Maximizing SNR in improper complex noise and applications to CDMA," *IEEE Commun. Lett.*, vol. 1, no. 1, pp. 5–8, Jan. 1997.

[67] W. Gerstacker, R. Schober, and A. Lampe, "Receivers with widely linear processing for frequency-selective channels," *IEEE Trans. Commun.*, vol. 51, no. 9, pp. 1512–1523, Sep. 2003.

[68] Z. K. M. Ho and E. A. Jorswieck, "Improper Gaussian signaling on the two-user SISO interference channel," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3194–3203, Sep. 2012.

[69] Y. Zeng *et al.*, "Transmit optimization with improper Gaussian signaling for interference channels," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2899–2913, Jun. 2013.

[70] 3GPP, *Hardware calibration requirement for dual layer beamforming*, R1-092359, Huawei, Jul. 2009.

[71] ——, *Performance comparison between Tx diversity and single stream precoding*, R1-091368, Nokia Siemens Networks, Nokia, Mar. 2009.

[72] F. Kaltenberger *et al.*, "Relative channel reciprocity calibration in MIMO/TDD systems," in *IEEE Future Network and Mobile Summit*, Jun. 2010.

[73] 3GPP, *Channel reciprocity modeling and performance evaluation*, R1-094622, Alcatel-Lucent Shanghai Bell, Alcatel-Lucent, Nov. 2009.

[74] F. Huang *et al.*, "Antenna mismatch and calibration problem in coordinated multi-point transmission system," *IET Commun.*, vol. 6, no. 3, pp. 289–299, Feb. 2012.

[75] X. Zhang, D. P. Palomar, and B. Ottersten, "Statistically robust design of linear MIMO transceivers," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3678–3689, Aug. 2008.

[76] A. Pascual-Iserte *et al.*, "A robust maximin approach for MIMO communications with partial channel state information based on convex optimization," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 364–360, 2006.

[77] A. M. Hamza and J. W. Mark, "A timing synchronization scheme in coordinated base-stations cooperative communications," *IEEE Wireless Commun. Signal Process.*, pp. 1–6, Oct. 2012.

[78] Y. Zhang and J. Zhang, "Multiple CFOs compensation and BER analysis for cooperative communication systems," *IEEE Wireless Commun. Netw. Conf.*, pp. 1–6, Apr. 2009.

[79] Y. Mostofi and D. C. Cox, "Mathematical analysis of the impact of timing synchronization errors on the performance of an OFDM system," *IEEE Trans. Commun.*, vol. 54, no. 2, pp. 226–230, Feb. 2006.

[80] L. Zhao *et al.*, "An enhanced signal-timing-offset compensation algorithm for coordinated multipoint-to-multiuser systems," *IEEE Commun. Letters*, vol. 18, no. 6, pp. 983–986, Jun. 2014.

[81] 3GPP, *Coordinated multi-point operation for LTE with non-ideal backhaul*, TR 36.874, Release 12, v12.0.0, Dec. 2013.

[82] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access: Special section on recent advances in C-RAN*, vol. 2, pp. 1326–1339, Oct. 2014.

[83] M. Peng *et al.*, "Heterogeneous cloud radio access networks: a new perspective for enhancing spectral and energy efficiencies," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 126–135, Dec. 2014.

[84] ——, "System architecture and key technologies for 5G heterogeneous cloud radio access networks," *IEEE Netw.*, vol. 29, no. 2, pp. 6–14, Mar. 2015.

[85] K. M. Abadir and J. R. Magnus, *Matrix algebra.* Cambridge (UK), Cambridge University Press, 2005.

[86] R. Bhatia, *Matrix analysis.* Springer, New York, 1997.

[87] G. H. Golub and C. F. V. Loan, *Matrix computations.* The Johns Hopkins University Press, 1996.

[88] A. Hjorungnes, *Complex-calued matrix derivatives.* Cambridge (UK), Cambridge University Press, 2011.

[89] K. B. Petersen and M. S. Pedersen, *The matrix cookbook.* Technical University of Denmark, 2012.

[90] A. Hjorungnes, D. Gesbert, and D. P. Palomar, "Unified theory of complex-valued matrix differentiation," *IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, pp. 345–348, Apr. 2007.

[91] D. P. Bertsekas, *Convex optimization theory.* Athena Scientific, 2009.

[92] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge, 2004.

[93] M. Chiang *et al.*, "Layering as optimization decomposition: a mathematical theory of network architectures," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 225–312, Jan. 2007.

[94] D. P. Palomar and M. Chiang, "Alternative distributed algorithms for network utility maximization: framework and applications," *IEEE Trans. Automatic Control*, vol. 52, no. 12, pp. 2254–2269, Dec. 2007.

[95] G. Scutari *et al.*, "Decomposition by partial linearization: parallel optimization of multi-agent systems," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 641–656, Feb. 2014.

[96] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: theory of majorization and its applications.* Springer Series in Statistics, 2009.

[97] E. A. Jorswieck and H. Boche, "Majorization and matrix monotone functions in wireless communications," *Foundations and Trends in Communun. and Inf. Theory*, vol. 3, no. 6, pp. 553–701, Jul. 2007.

[98] D. P. Palomar and Y. Jiang, "MIMO transceiver design via majorization theory," *Foundations and Trends in Commun. and Inf. Theory*, vol. 3, no. 4-5, pp. 331–551, 2007.

[99] S. Hougardy, "Classes of perfect graphs," *Discrete Mathematics*, vol. 306, pp. 2529–2571, 2006.

[100] M. Chudnovsky *et al.*, "Recognizing Berge graphs," *Combinatorica*, vol. 25, no. 2, pp. 143–186, Mar. 2005.

[101] D. Brelaz, "New methods to color the vertices of a graph," *J. Commun. ACM*, vol. 22, no. 4, pp. 251–256, Apr. 1979.

[102] D. B. West, *Introduction to Graph Theory, 2nd ed.* Englewood Cliffs, NJ: Prentice-Hall, 2000.

[103] MATLAB Inc., Mathworks. [Online]. Available: www.mathworks.com/products/matlab/.

[104] 3GPP, *Small cell enhancements for E-UTRA and E-UTRAN - physical layer aspects*, TR 36.872, Release 12, v12.1.0, Dec. 2013.

[105] I. Telatar, "Capacity of multi-antenna Gaussian channels," in *AT&T Technical Memorandum*, Jun. 1995.

[106] V. R. Cadambe, S. A. Jafar, and C. Wang, "Interference alignment with asymmetric complex signaling - Settling the Høst–Madsen-Nosratinia conjecture," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4552–4565, Sep. 2010.

[107] G. Tauböck, "Complex-valued random vectors and channels: entropy, divergence, and capacity," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2729–2744, May 2012.

[108] C. Hellings, M. Joham, and W. Utschick, "QoS feasibility in MIMO broadcast channels with widely linear transceivers," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1134–1137, Nov. 2013.

[109] C. Hellings and W. Utschick, "Iterative algorithms for transceiver design in MIMO broadcast channels with improper signaling," *Proc. 10th Int. ITG Conf. Systems, Commun., Coding*, Feb. 2015.

[110] ——, "Performance gains due to improper signals in MIMO broadcast channels with widely linear transceivers," *IEEE Int. Conf. Acoustics, Speech and Signal Process.*, pp. 4379–4383, May 2013.

[111] C. Lameiro and I. Santamaria, "Degrees-of-freedom for the 4-user SISO interference channel with improper signaling," *IEEE Int. Conf. Commun.*, pp. 3053–3057, Jun. 2013.

[112] M. Torrellas, A. Agustin, and J. Vidal, "DoF-delay trade-off for the K-user MIMO interference channel with delayed CSIT," *eprint arXiv:1504.05498*, Jun. 2015.

[113] Y. Zeng *et al.*, "Optimized transmission with improper Gaussian signaling in the K-user MISO interference channel," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6303–6313, Dec. 2013.

[114] C. Lameiro, I. Santamaria, and J. P. Schreier, "Analysis of maximally improper signaling schemes for underlay cognitive radio networks," *IEEE Int. Conf. Commun.*, pp. 1398–1403, Jun. 2015.

[115] ——, "Benefits of improper signaling for underlay cognitive radio," *IEEE Wireless Commun. Letters*, pp. 22–25, Feb. 2015.

[116] O. Amin, W. Abediseid, and M.-S. Alouini, "Underlay cognitive radio systems with improper Gaussian signaling: outage performance analysis," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4875–4887, Jul. 2016.

[117] ——, "Outage performance of cognitive radio systems with Improper Gaussian signaling," *IEEE Int. Symp. Inf. Theory*, pp. 1851–1855, Jun. 2015.

[118] C. Kim *et al.*, "Asymmetric complex signaling for full-duplex decode-and-forward relay channels," *Proceedings of the Int. Conf. on ICT Convergence*, pp. 28–29, Oct. 2012.

[119] M. Gaafar *et al.*, "Underlay spectrum sharing techniques with in-band full-duplex Systems using improper Gaussian signaling," *eprint arXiv:1601.00256*, Jan 2016.

[120] D. Palomar, J. Cioffi, and M. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

[121] M. Costa, "On the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. 31, no. 5, pp. 607–615, Sep. 1985.

[122] E. Kurniawan and S. Sun, "Improper Gaussian signaling scheme for the Z-interference channel," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3912 – 3923, Jul. 2015.

[123] C. Lameiro, I. Santamaria, and J. P. Schreier, "Rate region boundary of the Z-interference channel with improper signaling," *eprint arXiv:1605.04201*, May 2016.

[124] R. Merris, "An improvement of the Fischer inequality," *Journal of Research of the Notional Bureau of Standards - B. Mathematical Sciences*, vol. 75B, no. 1-2, Jun. 1971.

[125] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1261–1282, Apr. 2005.

[126] S. Barbarossa, *Multiantenna wireless communication systems.* Mobile Commun. series, 2005.

[127] Q. Shi *et al.*, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[128] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[129] C. Hellings and W. Utschick, "Block-skew-circulant matrices in complex-valued signal processing," *IEEE Trans. Signal Process.*, vol. 63, no. 8, pp. 2093–2107, Jan. 2015.

[130] K. Nikodem, T. Rajba, and S. Wasowicz, *Functions generating strongly Schur-convex sums.* Springer International Series of Numerical Mathematics, v. 161, pp. 175-182, 2012.

[131] F. Hiai, "Log-majorizations and norm inequalities for exponential operators," *Banach Center Publications*, vol. 38, no. 1, pp. 119–181, 1997.

[132] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 57–53, Feb. 2008.

[133] C. Shi *et al.*, "Distributed interference pricing for the MIMO interference channel," *IEEE Int. Conf. Commun.*, pp. 1–5, Jun. 2009.

[134] G. Scutari *et al.*, "Distributed dynamic pricing for MIMO interfering multiuser systems: a unified approach," *IEEE Int. Conf. Network Games, Control and Optimization*, Oct. 2011.

[135] S. Shi, M. Schubert, and H. Boche, "Downlink MMSE transceiver optimization for multiuser MIMO systems: duality and sum-MSE minimization," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5436–5446, Nov. 2007.

[136] S. Christensen *et al.*, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[137] H. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[138] M. Biguesh and A. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[139] D. P. Bertsekas, *Nonlinear programming.* 2nd ed. Belmong, MA: Athena Scientific, 1999.

[140] P. Tseng, "Convergence of a block coordinate descent method for non-differentiable minimization," *J. Optimization Theory and Applications*, vol. 103, no. 9, pp. 475–494, 2001.

[141] 3GPP, *Enhanced performance requirement for LTE user equipment (UE)*, TR 36.829, Release 11, v11.1.0, Dec. 2012.

[142] A. Bourdoux, B. Come, and N. Khaled, "Non-reciprocal transceivers in OFDM/SDMA systems: impact and mitigation," in *IEEE Proc. Radio and Wireless Conference*, Aug. 2003, pp. 183–186.

[143] 3GPP, *Coordinated sounding for CoMP BF (CoMP BF-CoS) exploiting TDD channel reciprocity*, R1-134291, DAC-UPC, IAESI, Oct. 2013.

[144] G. Scutari, D. P. Palomar, and S. Barbarossa, "The MIMO iterative waterfilling algorithm," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1917–53, May. 2009.

[145] E. Dahlman, S. Parkvall, and J. Sköld, *4G: LTE/LTE-Advanced for mobile broadband.* Elsevier Ltd, Academic Press, 2010.

[146] D. G. Luenberger and Y. Ye, *Linear and nonlinear programming, Third edition.* Springer, 2008.

[147] D. Zhu, M. Lei, and A. Goldsmith, "Coordinated resource allocation in centralized radio access networks with dynamic downlink/uplink reconfiguration," *IEEE Global Commun. Conf.*, pp. 3625–3630, Dec. 2013.

[148] D. Zhu and M. Lei, "Cluster-based dynamic DL/UL reconfiguration method in centralized RAN TDD with dense deployment of remote radio units," *IEEE 77th Vehicular Technology Conf. (VTC Spring)*, pp. 1–5, Jun. 2013.

[149] A. A. Dowhuszko *et al.*, "A decentralized cooperative uplink/downlink adaptation scheme for TDD small cell networks," *IEEE 24th Int. Symp. Personal Indoor and Mobile Radio Commun.*, pp. 1682–1687, Sep. 2013.

[150] M. S. ElBamby *et al.*, "Dynamic uplink-downlink optimization in TDD-based small cell networks," *Int. Symp. in Wireless Commun. Systems*, pp. 939–944, Aug. 2014.

[151] V. Venkatasubramanian *et al.*, "On the performance gain of flexible UL/DL TDD with centralized and decentralized resource allocation in dense 5G deployments," *IEEE 25th Int. Symp. on Personal, Indoor and Mobile Radio Commun.*, pp. 1840–1845, Sep. 2014.

[152] *Performance analysis of DL-UL interference management and traffic adaptation in multi-cell pico-pico deployment scenario*, 3GPP R1-121529, Intel Corporation, Mar. 2012.

[153] A. M. Fouladgar *et al.*, "Joint interference alignment and bi-directional scheduling for MIMO two-way multi-link networks," *IEEE Int. Conf. on Commun.*, pp. 4126–4131, Jun. 2015.

[154] P. Mogensen *et al.*, "Centimeter-wave concept for 5G ultra-dense small cells," *IEEE 79th Vehicular Technology Conf. (VTC Spring)*, pp. 1–6, May 2014.

[155] Qualcomm, "The 5G unified air interface," Nov. 2015.

[156] F. Kelly, "Charging and rate control for elastic traffic," *European Trans. Telecommun.*, pp. 33–37, 1997.

[157] J. Lee and N. Jindal, "Dirty paper coding vs linear precoding for MIMO broadcast channels," in *Asilomar Conf. Signals, Systems and Computers*, Oct. 2006, pp. 779–783.

[158] X. Wang, G. B. Giannakis, and A. G. Marques, "A unified approach to QoS-guaranteed scheduling for channel-adaptive wireless networks," *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2410–2431, Dec. 2007.

[159] H. Kushner and G. Yin, *Stochastic approximation algorithms and applications.* 2nd ed. Berlin, Germany, Springer-Verlag, 2003.

[160] H. Kushner and P. A. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions," *IEEE Trans. on Wireless Commun.*, vol. 3, no. 4, pp. 1250–1259, Jul. 2004.

[161] R. Agrawal and V. Subramanian, "Optimality of certain channel-aware scheduling policies," *Proc. 40th Allerton Conf. Commun., Control and Computing*, pp. 636–647, Oct. 2002.

[162] A. Agustin *et al.*, "Decentralized weighted sum rate maximization in MIMO-OFDMA femtocell networks," *IEEE Global Commun. Conf.*, Dec. 2011.

[163] S.-J. Kim and G. B. Giannakis, "Optimal resource allocation for MIMO ad hoc cognitive radio networks," *IEEE Trans. on Inf. Theory*, pp. 3117–3131, May. 2011.

[164] J. Huang, R. A. Berry, and M. L. Honig, "Distributed interference compensation for wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 5, pp. 1074–1084, May 2006.

[165] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.

[166] C. Shi, R. A. Berry, and M. L. Honig, "Monotonic convergence of distributed interference pricing in wireless networks," in *ISIT Conf.* IEEE, Seoul, Korea, Jun. 2009.

[167] W. Yu, "Multiuser water-filling in the presence of crosstalk," *Proc. Inf. Theory and App. Workshop*, Jan. 2007.

[168] T. M. Cover and J. A. Thomas, *Elements of information theory.* Wiley, 2006.

[169] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: A comprehensive survey," *IEEE Commun. Surveys Tutorials*, vol. 3, no. 2, pp. 10–31, Apr. 2000.

[170] S. Sadr and S. R. Adve, "Partially-distributed resource allocation in small-cell networks," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 12, pp. 6851–6862, Dec. 2014.

[171] A. Kumar, D. Manjunath, and J. Kuri, *Communication Networking: an analytical approach.* Morgan Kaufmann Publishers, Elsevier, 2004.

[172] T. Bonald, "Flow-level performance analysis of some opportunistic scheduling algorithms," *European Trans. on Telecommunications*, vol. 16, no. 1, pp. 65–75, Jan. 2005.

[173] H. Klessig, A. Fehske, and G. Fettweis, "Admission control in interference-coupled wireless data networks: a queuing theory-based network model," *Fifth Int. Workshop on Indoor and Outdoor Small Cells*, pp. 151–158, May. 2014.

[174] M. Chudnovsky *et al.*, "The strong perfect graph theorem," *Annals of Mathematics*, vol. 164, no. 1, pp. 51–229, 2006.

[175] L. Lovász, "Normal hypergraphs and the perfect graph conjecture," *Discrete Mathematics*, vol. 2, no. 3, pp. 253–267, Jun. 1972.

[176] A. Parra and P. Scheffler, "Characterizations and algorithmic applications of chordal graph embeddings," *Discrete Applied Mathematics*, vol. 79, no. 1–3, pp. 171–188, Nov. 1997.

[177] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric algorithms and combinatorial optimization.* Springer Verlag, 1993.

[178] D. J. A. Welsh and M. B. Powell, "An upper bound for the chromatic number of a graph and its application to timetabling problems," *The Computer Journal*, vol. 10, pp. 85–86, 1967.

[179] O. Petelin and R. Adve, "Distributed resource allocation in femtocell networks," *13th Canadian Workshop on Inf. Theory*, pp. 102–107, 2013.

[180] C. Bron and J. Kerbosch, "Finding all cliques of an undirected graph [H]," *J. Commun. ACM*, vol. 16, no. 9, pp. 575–577, Sep. 1973.

[181] P. Heggernes *et al.*, "Interval completion with few edges," *ACM Symp. Theory of Computing*, pp. 374–381, 2007.

[182] B. Friedlander and A. J. Weiss, "Direction finding using noise covariance modeling," *IEEE Trans. Signal Process.*, vol. 43, no. 7, pp. 1557–1567, Jul. 1995.

[183] J. M. Goldberg and J. R. Fonollosa, "Downlink beamforming for spatially distributed sources in cellular mobile communications," *Elsevier Journal Signal Process.*, vol. 65, no. 2, pp. 181–197, Mar. 1998.

[184] J. Kaleva *et al.*, "Decentralized coherent coordinated multi-point transmission for weighted sum rate maximization," *IEEE Global Commun. Conf.*, Dec. 2015.

[185] P. Fraigniaud *et al.*, "Distributed computing with advice: information sensitivity of graph coloring," *Distributed Computing*, vol. 21, no. 6, pp. 395–403, Dec. 2009.