

Efficient Backhauling in Cooperative MultiPoint Cellular Networks

TILAK RAJESH LAKSHMANA

Communication Systems Group
Department of Signals and Systems
 CHALMERS UNIVERSITY OF TECHNOLOGY
 Gothenburg, Sweden, 2013

Thesis for the degree of Licentiate of Engineering

Efficient Backhauling in Cooperative MultiPoint Cellular Networks

Tilak Rajesh Lakshmana



CHALMERS

Communication Systems Group
Department of Signals and Systems
Chalmers University of Technology
Gothenburg, Sweden

Lakshmana, Tilak Rajesh
Efficient Backhauling in Cooperative MultiPoint Cellular Networks

Technical Report No. R005/2013
ISSN 1403-266X
Communication Systems Group
Department of Signals and Systems
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
Telephone: + 46 (0)31-772 1000



© 2013 Tilak Rajesh Lakshmana

except where otherwise stated.

No rights reserved.

This thesis has been prepared using LyX.

Front Cover: The cover picture is an abstract illustration of a PHY layer stochastic precoder using particle swarm optimization to find the precoding weights. The swarm of birds flying through the search space try to find their young ones to give them food. The young ones represent the stable or equilibrium solution points in the search space and the food represents the precoding weights.

Printed by Chalmers Reproservice
Gothenburg, Sweden, February 2013.

To *Friends and Family*

Contents

Abstract	1
List of Included Publications	2
Preface	3
List of Abbreviations	5
List of Notations	7
I. Coordinated MultiPoint transmission	9
1. A Brief Introduction	11
1.1. Coordinated MultiPoint	13
1.2. The Motivation and the Problem	14
1.2.1. Precoding, a PHY Layer Design	19
1.2.2. Scheduling, a MAC Layer Design	20
1.2.3. Network Architecture	21
2. Particle Swarm Optimization	23
2.1. The Power of Stochastic Algorithms	25
2.2. Particle Swarm Optimization as a Tool for Precoding	27
2.3. Advantages and Disadvantages	28
3. Organization of this Thesis and Included Papers	29
3.1. Paper A - Partial Joint Processing with Efficient Backhauling in Co-ordinated MultiPoint Networks	29
3.2. Paper B - Partial Joint Processing with Efficient Backhauling using Particle Swarm Optimization	30
3.3. Paper C - Scheduling for Backhaul Load Reduction in CoMP	30
3.4. Paper D - On the Potential of Broadcast CSI for Opportunistic Co-ordinated MultiPoint Transmission	30
3.5. Visions and Future Work	31
3.6. List of Related Publications	31
Bibliography	33

II. Included Papers	37
4. Partial Joint Processing with Efficient Backhauling in Coordinated Multi-Point Networks	41
4.1. Introduction	41
4.2. System Model	43
4.3. Particle Swarm Optimization	45
4.4. Simulation Results	48
4.5. Conclusion	50
Bibliography	52
5. Partial Joint Processing with Efficient Backhauling using Particle Swarm Optimization	55
5.1. Introduction	55
5.1.1. State of the art techniques	58
5.1.2. Contributions	59
5.2. System model	61
5.2.1. Linear beamforming	62
5.2.2. Limitations of the state of the art	62
5.3. Particle swarm optimization for precoding in the PJP framework . . .	65
5.3.1. Objective function	66
5.3.2. Termination criteria	68
5.3.3. Convergence	68
5.3.4. Computation complexity analysis	69
5.4. Analysis of interference using Gershgorin's discs	70
5.5. Simulation results	71
5.5.1. Objective function: Weighted interference minimization . . .	75
5.5.2. Objective function: Sum rate maximization	80
5.5.3. Gershgorin's circles	80
5.6. Conclusions	82
Bibliography	84
6. Scheduling for Backhaul Load Reduction in CoMP	89
6.1. Introduction	89
6.2. System Model	91
6.3. Scheduling	93
6.3.1. Block Diagonalization (BD)	94
6.3.2. Unconstrained Scheduling (US)	95
6.3.3. Constrained Scheduling (CS)	95
6.4. Performance Evaluation	96
6.5. Conclusion	101
Bibliography	103

7. On the Potential of Broadcast CSI for Opportunistic Coordinated Multi-Point Transmission	107
7.1. Introduction	107
7.2. Signal and System Model	109
7.3. Decentralized Opportunistic CoMP	111
7.3.1. Centralized joint processing with Unicast CSI	111
7.3.2. Decentralized Opportunistic CoMP with Broadcast CSI	111
7.3.3. Decentralized Joint Processing with Broadcast CSI exploiting diversity	113
7.4. Performance Evaluation	113
7.5. Conclusion	117
Bibliography	121
III. Appendix	123
A. Max-min SINR Comparison with Weighted Interference Minimization	124
B. Performance of PSO based on Field Measurements Data	127
Bibliography	130

Abstract

The efficient use of the spectrum in cellular systems has given rise to cell-edge user equipments (UEs) being prone to intercell interference. In this regard, coordinated multipoint (CoMP) transmission is a promising technique that aims to improve the UE data rates. In a centralized network architecture, the users need to feed back the channel state information (CSI) to its anchor base station (BS). The CSI is then forwarded to a central coordination node (CCN) for precoder design to jointly mitigate interference. However, feeding back the CSI consumes over-the-air uplink resources as well as backhaul resources. To alleviate this burden, the quantity of CSI being fed back is limited via relative thresholding. That is, the CSI feedback is limited to those BSs whose signal strength fall above a threshold relative to the strongest BS. Moreover, with limited CSI, efficient backhauling of the precoding weights is necessary, as the user data is routed based on the path taken by the precoding weights from the CCN to the corresponding BSs. The focus of this thesis is mainly on a physical (PHY) layer and a medium access control (MAC) layer approach for reducing the backhaul load in a CoMP system, with minimal penalty on the potential CoMP gains. Furthermore, broadcasting the CSI in a decentralized network architecture is considered in order to reduce backhaul latency.

In the PHY layer approach, the precoder design is based on stochastic optimization such as particle swarm optimization (PSO). This method has no constraints on the scheduling of the UEs. The PSO based precoder design was also applied to field measurement data with CSI imperfections due to prediction errors and quantization errors. It was found to perform the best compared to other robust precoders developed in the EU FP7 ARTIST4G consortium. With the MAC layer approach, a simple zero forcing precoder is assumed, which focuses on how to schedule the UEs in such a way that they achieve the backhaul load reduction. Lastly, the decentralized network architecture is explored, where the UEs broadcast the CSI. The BSs coordinate by sharing minimal scheduling information, thereby achieving data rates comparable to the centralized network architecture.

In this thesis, the backhauling is defined to be *efficient* when the total number of CSI coefficients aggregated at the CCN is equal to the total number of precoding weights for a given time-frequency resource, in a centralized architecture with the PHY layer approach. In the MAC layer approach, the total number of precoding weights is less than or equal to the total number of CSI coefficients. In the decentralized network architecture, the CCN does not exist. The BSs can coordinate over a less stringent backhaul, thereby reducing the backhaul load and latency.

Keywords: Coordinated multipoint transmission, centralized, decentralized architecture, particle swarm optimization, precoding, prediction errors, quantization errors, scheduling, stochastic optimization

List of Included Publications

This licentiate thesis is based on the following appended papers. They are:

- [A] T.R. Lakshmana, C. Botella, and T. Svensson, “Partial Joint Processing with Efficient Backhauling in Coordinated MultiPoint Networks,” in *Proc. IEEE Vehicular Technology Conference (VTC)*, Jun. 2012.
- [B] T.R. Lakshmana, C. Botella, and T. Svensson, “Partial Joint Processing with Efficient Backhauling using Particle Swarm Optimization,” *EURASIP Journal of Wireless Communications and Networking.*, vol. 2012, 2012.
- [C] T.R. Lakshmana, J. Li, C. Botella, A. Papadogiannis and T. Svensson, “Scheduling for Backhaul Load Reduction in CoMP,” in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2013.
- [D] T. R. Lakshmana, A. Papadogiannis, J. Li, and T. Svensson, “On the Potential of Broadcast CSI for Opportunistic Coordinated MultiPoint Transmission,” in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2012.

Preface

It gives me immense pleasure to present this licentiate thesis. The thesis has been organized in three parts. In the first part, the topic is introduced, keeping an undergraduate or a master student in mind. Hence, the initial chapters could be verbose for researchers. In the second part of the thesis, the published papers are appended whose layout has been revised. The final part is the Appendix, where additional results that have not been published in any conference or journal are presented. The appendix is a complement to the material covered in this work and one can see it as a section that makes the thesis more complete.

I am still learning to use the creative commons. However, I would like to highlight that “Sita sings the blues” has inspired me to make this work available under



I hope you enjoy reading this thesis as much as I have enjoyed writing it.

Acknowledgments

I would like to thank my supervisor Tommy Svensson for giving me the opportunity to work with some amazing people. It has been great to have our lively discussions in the corridor, tram, metro, plane, train, at Skatteverket and in our offices, at various times, on various topics. Thank you for being there virtually every step of the way. I would like to thank my co-supervisor Carmen Botella for helping me take the initial steps in this field of research. I miss your guidance, kindness and your thoughtful comments. I would also like to thank my current co-supervisor Agisilaos Papadogiannis for guiding me with new ideas. I admire your attention to detail, where you quickly grasp the results and also identify my pitfalls. I thought I knew English reasonably well until I received comments from my supervisors. I thank you all. I would also like to thank Mattias Wahde for triggering my curiosity with his lectures, advice and his book on “Biologically Inspired Optimization Methods”. Thanks to Erik Ström for giving me the opportunity to be here at ComSys, and also for the pleasure of being a teaching assistant for the wireless communication course. This course has taught me how to teach and interact with students. In particular, my understanding on how to supervise and be supervised. I hope, someday I can teach like the way you do it.

Working with my colleague Jingya Li has been a tremendous learning experience in every way. Apart from the awesome discussions, and constant exchange of thoughts, ideas and songs, your innocence has enriched our work environment with a feeling of leisure. I have learnt a lot about Chinese culture and I thoroughly enjoy your food. I thank you for enjoying - or putting up with - my taste in music.

I would like to thank Tommy Svensson for creating an amazing VR project meeting environment where I got to meet and collaborate with researchers from other universities in Sweden. The VR project meetings with Rikke Apelfröjd, Mikael Stenard, Anna Brunström and Annika Klockar would never be the same without your helpful critical comments. In particular, I would like to thank Rikke and Mikael at Uppsala University for the results presented in Appendix B with the field measurement data.

Thanks to Erik Agrell for introducing me to the sport of Fox Hunting (a.k.a. radio orienteering). This sport has helped me understand contours, search spaces, optimization and experience Sweden's beautiful nature.

Thanks to Lars Börjesson for the encyclopedia of Swedish punk, in particular your taste of music has been an enriching experience, and of course thank you for the prompt help. Thanks to Agneta for making my environment at ComSys so research-friendly. Thanks to Madeleine for administrative help, and to Natasha for all the Visa renewals.

My present and former colleagues at S2, you are a lovely bunch. Thanks to Arash, Srikar, Behrooz, Kasra, Mohammad, Yutao, Rajet, Lotfollah, Johnny, Naga, Keerthi, Henk, Nima/s, Ali, Gabo, Rajet, Reza, Fateme, Ashkan, Satya, Misha, Christian, Christopher, Eija, Marie, Abu, Deepthi, Lennart, Astrid, Benji, Ahmed, Lars, Tomas. You all have crossed my path with memories to reflect upon.

Inspiration and helping hands are not enough with a thank you. I have this section and I can at least thank Brita, Prakash, Martin, Kasyab, Emma, Patricia, Manu, Vijay, Anish, Sumana, Suri, Preeth, Shiv, Aithal, Lakshmi, Kempe, Vishwas, the folks at St. Andrews, Totoro and Winnie the Pooh. I would like to thank my parents, my grandmother Rita, my sister Leela, and my lovely niece Trishel, for holding everything together. I love you all.

To be where I am now, I thank you all.

Carpe Diem!

Tilak Rajesh Lakshmana
Gothenburg, February 2013

Thanks to the following bodies: the Swedish Governmental Agency for Innovation Systems (VINNOVA), the Swedish Research Council (VR) and the Seventh Framework Program (EU FP7-ARTIST4G) for supporting my work.

List of Abbreviations

1G	First Generation
2G	Second Generation
3G	Third Generation
3GPP	3rd Generation Partnership Project
4G	Fourth Generation
AMPS	Advanced Mobile Phone Service
BC	BroadCast
CCN	Central Coordination Node
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CoMP	Coordinated MultiPoint
CQI	Channel Quality Indicator
CS	Coordinated Scheduling
CSI	Channel State Information
DPC	Dirty Paper Coding
DPC	Dirty Paper Coding
FBH	Full BackHauling
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
FFB	Full FeedBack
GSM	Global System for Mobile
HARQ	Hybrid Automatic Repeat reQuest

ICIC	InterCell Interference Coordination
LBH	Limited BackHauling
LFB	Limited FeedBack
LTE	Long Term Evolution
MAC	Medium Access Control
MIMO	Multiple Input Multiple Output
MMSE	Minimum Mean Square Error
NMT	Nordic Mobile Telephone
OFDM	Orthogonal Frequency Division Multiplexing
OSI	Open Systems Interconnection
PHY	Physical Layer
PSO	Particle Swarm Optimization
RLC	Radio Link Control
RLP	Robust Linear Precoding
RLP-ACFF	RLP - Automatic Control robust Feed-Forward
RRC	Radio Resource Control
SINR	Signal to Interference plus Noise Ratio
SNR	Signal to Noise Ratio
TACS	Total Access Control/Communications System
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
UE	User Equipment
UNESCO	United Nations Educational, Scientific and Cultural Organization
WCDMA	Wideband Code Division Multiple Access
WIM	Weighted Interference Minimization
ZF	Zero Forcing

List of Notations

The following notation has been used in PART I of the thesis.

x	scalar
\mathbf{x}	vector
\mathbf{X}	Matrix
$\mathbf{X}(i, j)$	The i th row and the j th column
$\widetilde{\mathbf{X}}$	Sparse matrix
\mathbb{C}	A set consisting of complex numbers

Part I.

Coordinated MultiPoint transmission

1. A Brief Introduction

“For millions of years, mankind lived just like the animals. Then something happened which unleashed the power of our imagination. We learned to talk and we learned to listen. Speech has allowed the communication of ideas, enabling human beings to work together to build the impossible. Mankind’s greatest achievements have come about by talking, and its greatest failures by not talking. It doesn’t have to be like this. Our greatest hopes could become reality in the future. With the technology at our disposal, the possibilities are unbounded. All we need to do is make sure we keep talking.”

— Stephen W. Hawking

Communications have been the very essence of human interaction. There is an inherent need to communicate, as individuals have moved farther away exploring nature. As the invention of the wheel made humans explore places far and wide, so did communications evolve with tools such as smoke signals, Morse code, the telephone, etc. In particular, the technology that enables smart phones has created a different world with unbounded possibilities. In this section, a brief evolution of cellular communications is presented from an interference point of view, including how coordinated multipoint transmission comes into play.

In the nineteenth century, telegraph wires were used for long distance communications. However, they suffered from attenuation. It was a time when speaking could be regarded as the only form of wireless communication for human interaction. Maxwell’s equations were merely theoretical until Heinrich Hertz showed how to capture electromagnetic waves. This can be considered the trigger that started wireless communications. Transatlantic telegraph cables provided a good communication link for countries like Sweden and the United States of America. However, with severed cables, wireline communications is easily hindered. In Sweden, Grimeton radio station in Varberg was constructed for making wireless telegraphy. This was the main mode of communication in the early twentieth century. In 2004, Grimeton radio station was tagged as a world heritage site by UNESCO. In its prime days, the radio station occupied an entire building consuming 200 kW as shown in Figure 1.1. It was operating at 17.2 kHz, giving rise to a wavelength of ~ 17.4 kms. This required six antenna towers, placed 380 m apart, to accommodate the radiating antenna elements. A brief history of communications from the nineteenth to the twenty-first century is well captured in [ComSoc]. Henceforth, the focus of this chapter is mainly on cellular wireless communications.



Figure 1.1.: A preelectronic radio transmitter for transatlantic communication at Grimeton, east of Varberg, Sweden.

In the early 1980s, the first generation (1G) of wireless hand-held cellular phones were introduced. In different countries, this took different forms. In the United States, Advanced Mobile Phone Service (AMPS) used frequency modulation to carry voice (analog) service to a mobile phone or a user equipment (UE). A given UE is assigned a specific frequency exclusively available in each cell. This assignment can be called frequency division multiple access (FDMA). In the United Kingdom, Total Access Control/Communications System (TACS) used frequency shift keying with FDMA. In the Nordic countries, 1G took the form of Nordic Mobile Telephone (NMT). Inherent design of this system constituted the UEs to be allocated on different frequencies and to have low cochannel interference from neighbouring cells. This gave rise to the concept of frequency reuse [G05] where the frequencies used in a particular cell are not reused in adjacent cells, but reused only after some distance. This also means that the cellular system is spectrally very inefficient in a given cell, as the entire spectrum available to a network operator is not fully utilized in every cell.

In 1991, the second generation (2G) of cellular systems, Global System for Mobile (GSM) communications brought forth digital cellular networks. Keeping interoperability in mind, GSM triggered a new era for widespread use of mobile communications. The UEs were scheduled in a given time-slot in a time division multiple access (TDMA) fashion. In the United States, cdmaOne was introduced, where the UEs are separated by means of spreading codes, giving rise to code division multiple access (CDMA). Similar to the 1G system, the spectrum in GSM is inefficiently used as every cell is assigned a given frequency, and the UEs in a given cell are separated in time. Typically, a misalignment in time for a voice service gives rise to crosstalk. In CDMA, both the time and frequency resources are used at the same time, and the UEs are separated by codes. With nonorthogonal codes, interference between the UEs within a cell will cause intracell interference, while intercell interference is

due to the codes reused in other cells [G05]. The spreading gain of the code can be used to lower the intercell and intracell interference. One could view this to be inefficient as a limited number of orthogonal codes need to be reused.

In 2001, the third generation (3G) cellular networks based on wideband code division multiple access (WCDMA) was commercially available. All the frequencies are used in every cell but the UEs are separated by codes, similar to CDMA. In WCDMA, Gold codes are used as scrambling codes to reduce the intercell interference, and nonorthogonality of codes leads to intracell interference. In short, the frequency reuse factor is one. In the spring of 2009 in Sweden, the fourth generation (4G) cellular systems employed orthogonal frequency division multiplexing (OFDM) where all the frequency-time resources are utilized in every cell. This means that the neighboring cells also use the same resource, thereby potentially using the spectrum resources efficiently. However, this gives rise to severe intercell interference. In Long Term Evolution (LTE), intercell interference coordination (ICIC) is proposed to overcome this problem. With ICIC, the BSs need to avoid scheduling UEs simultaneously at the cell-edge. However, this poses some restriction in using all the resources all the time, resulting in reduced efficiency.

To mitigate intercell interference, [KFV06] introduced coordination of BSs for spectrally efficient systems where the clustered BSs sent useful signals to the UEs. This is popularly known as network MIMO. In the 3rd Generation Partnership Project (3GPP) context, this is more widely known as coordinated multipoint. Currently, this is a working item as part of the LTE-Advanced [TR36819].

In the following sections, CoMP is introduced as a means to handle intercell interference as defined in 3GPP. Then, the focus transitions towards why backhauling aspects in CoMP cellular networks are important and how one can reduce the backhaul load.

1.1. Coordinated MultiPoint

Coordinated multipoint (CoMP) is a promising technique that aims to improve the UE data rates, especially at the cell-edge via coordinating BSs. The coordination of BSs can be used to mitigate intercell interference [GHH+10, MF11].

CoMP is broadly classified as downlink CoMP or uplink CoMP. In this regard, the downlink or the forward link is referred to as CoMP transmission, while the uplink or the reverse link is referred to as CoMP reception. CoMP reception is relatively simple compared to CoMP transmission [IDM+11], as there are no transceiver modifications required in the UE while the BSs can estimate the channel in the uplink. Hence, there is no need for the UEs to feed back the channel state information (CSI). In the case of a typical CoMP transmission, the UEs need to feed back the CSI to their serving BS(s). An entity in the network collates the CSI of the UEs and jointly calculates the precoding weights to mitigate interference. The CSI feedback applies to a frequency division duplex (FDD) system where the uplink and the downlink are on different carrier frequencies, while in the case of a time division duplex (TDD)

system, the uplink and downlink use the same frequency. The reciprocity nature of the channel can be used here and the feed back of CSI can be avoided, as long as the channel remains stationary from the time of reception until transmission. It should be noted that the reciprocity of the channel demands a well calibrated transmitter and receiver chain.

Some of the CoMP categories based on 3GPP [TR36819] are discussed here. CoMP transmission for a given time-frequency resource is categorized based on the availability of UE data at various BSs. When the UE data is available at more than one BS to jointly serve a given UE, it is referred to as joint processing. If the UE data is only available at one BS so that the scheduling of the UE is coordinated with other BSs, it is referred to as coordinated scheduling, or coordinated beamforming when the beamforming decisions are coordinated with other BSs. Alternately, a hybrid category uses joint processing and coordinated scheduling/beamforming. Joint processing can be further categorized as joint transmission, dynamic point selection (muting), and a hybrid of joint transmission and dynamic point selection. In joint transmission, the UEs are served from multiple BSs, either coherently or non-coherently to improve their data rates. In the case of dynamic point selection, the serving BSs dynamically changes over a given time-frequency resource. A very stringent requirement on synchronization is needed for those UEs that coherently receive their data. However, non-coherent joint transmission from different BSs can be viewed as the same signal taking different multipaths to reach the UE. The Table 1.1 is inspired from [RS10] and it summarizes the downlink CoMP categories.

Table 1.1.: Downlink CoMP categories

Joint Processing		Coordinated Beamforming/Scheduling
Joint Transmission	Dynamic Cell Selection	Beamforming/Scheduling
Data at each BS	Data at each BS	Data at serving BS only
Transmission from multiple BSs simultaneously	Transmission from one BS at a time	Transmission from one BS, but scheduling decisions are coordinated

In this thesis, the focus is on joint transmission combined with dynamic point selection for CoMP transmission in an FDD system.

1.2. The Motivation and the Problem

Recall that spectrally efficient systems are limited by interference. In this regard, consider a homogenous network as shown in Figure 1.2. The coverage area of each BS is approximated by the hexagonal cell where the cell-edge is marked with dashed

lines. The darker shaded hexagonal structure in the middle is defined as the cluster area where the BSs are allowed to cooperatively serve the UEs in this area. Modern cellular systems are spectrally efficient, as the same frequency-time resource is used in a given cluster area. This gives rise to intracell interference. If one were to visualize Figure 1.2 being replicated around itself, then the interference from the other clusters could be seen as intercluster interference. The UEs at the cluster edge are prone to intercluster interference that can potentially degrade the system performance. To overcome this problem, the clusters also need to be coordinated. However, full coordination is practically impossible. In [ZCA+09], limited intercluster coordination is performed for the disjoint clusters, and in [LBS12], frequency reuse schemes are proposed to mitigate the intercluster interference. The main focus of this thesis is on the UEs at the cell-edge in an isolated cluster, as illustrated in Figure 1.2. Hence, the intercluster interference is not considered in this study.

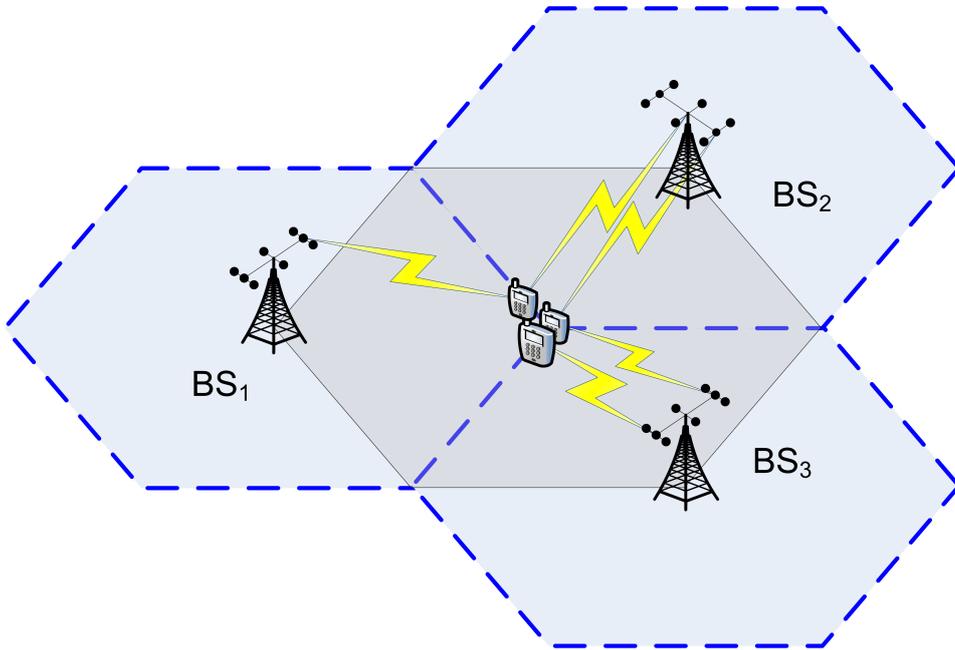


Figure 1.2.: A homogenous hexagonal network layout.

To mitigate interference in a centralized network architecture, the UEs need to feed back the CSI of those BSs from which they would like to be served. The CSI is usually fed back to their anchor BS. The anchor BS forwards this information to the central coordination node (CCN) where the precoding is performed to mitigate interference. A centralized joint processing network architecture is illustrated in Figure 1.3. The backhaul constitutes all the connections and network entities used to interconnect the BSs. In Figure 1.3, this would constitute the CCN and the connections between the BSs and the CCN. The focus of the thesis is mainly on the backhaul traffic, consisting of the CSI and the precoding weights.

In a practical scenario, the user data constitutes a major portion of the backhaul

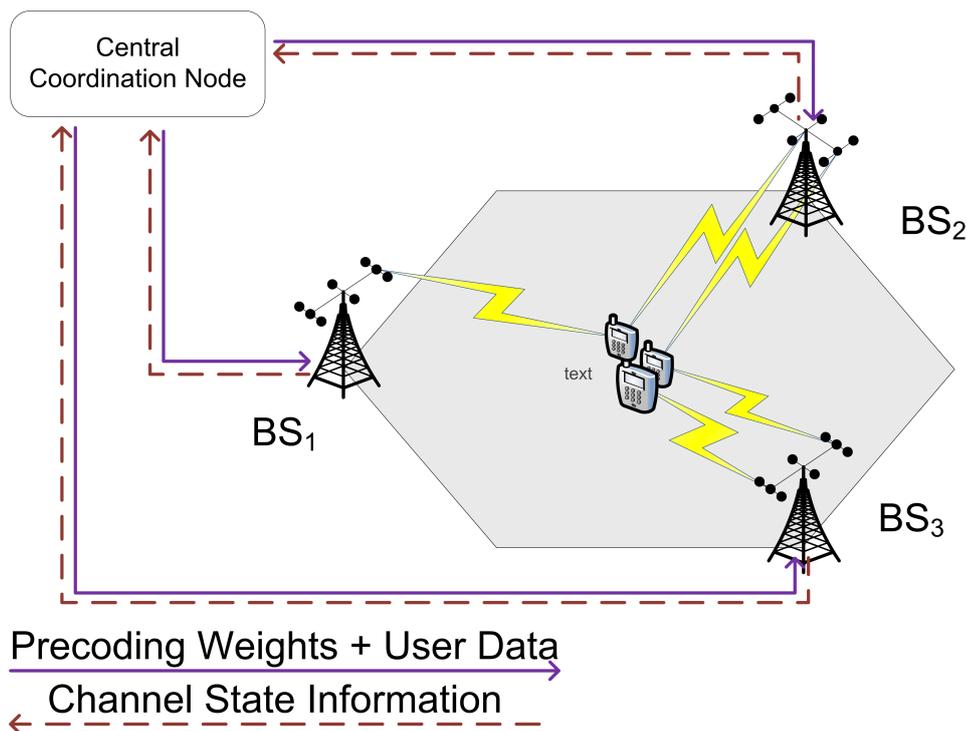


Figure 1.3.: A centralized joint processing network architecture.

traffic. However, in a centralized network architecture, the user data could be assumed to be routed based on the path taken by the precoding weights. Thus, the focus is more on the control signaling part of the backhaul traffic. As mentioned earlier, to coordinate all the BSs in the network would be impractical, and hence, clusters of BSs are formed [PGH08]. A predefined set of BSs forming a cluster that does not change with time is referred to as static clustering [ZCA+09]. Likewise, dynamic clusters of BSs can be formed depending on the channel conditions [PGH08]. Moreover, depending on where the clustering decisions are performed, it can be classified as network centric or UE centric clustering. Various combinations of the clustering can be performed. In this thesis, a *dynamic UE centric clustering* is performed, where the UE dynamically chooses the set of BSs from which it would like to be served [BSX+10]. To alleviate the problems of the CSI feedback overhead within a cluster area, absolute thresholding and relative thresholding can be considered [PBG+08, BSX+10]. In the case of absolute thresholding, the UEs are instructed to feed back the CSI of links that are above a certain value, while in the case of relative thresholding, the UEs are instructed to feed back links that fall within a window relative to the best link.

There is an inherent dimensionality limitation in estimating the channel from different BSs [CRP10]. Clustering of BSs could alleviate this dimensionality problem of coordinating the pilots for obtaining the CSI from different BSs. The UEs estimating the channel is one aspect of obtaining the CSI. The inherent delays due to

the control loop emphasizes the other important aspect of predicting the channel. Typically, a UE would employ a Kalman filter to estimate and predict the channel well in advance. The prediction horizon defines the duration of time for which the channel is predicted. A short prediction horizon will indirectly limit the UE velocity and it imposes a fast backhauling network with very low latency, in the order of milliseconds. The predicted CSI is quantized and fed back to the anchor BS. Quantization by itself gives rise to quantization errors and the process of feeding back the CSI also occupies the uplink resources. These practical aspects are considered in the precoder design and the results are presented in Appendix B.

To present the main problem being addressed in this thesis, it is assumed that the CSI obtained at the CCN is error free. Based on relative thresholding, consider the following channel matrix aggregated at the CCN as shown in Table 1.2, where UE_1 feeds back the CSI of BS_1 and BS_2 while CSI of BS_3 is not fed back as it falls outside the relative threshold window. Likewise, other UEs also feed back the CSI that falls above the threshold. Modeling of CSI that is not available at the CCN as zeros may not be the best way to go about it. However, intuitively it makes sense to treat them as zeros. These zeros denote the feedback reduction obtained with relative thresholding. The question that one would like to ask is, if an equivalent backhaul reduction can be obtained in terms of the precoding weights as shown in Table 1.3. That is, can the number of CSI feedback coefficients fed back by a given UE, have an equivalent number of precoding weights in the backhaul? More importantly, this is a desired property for the precoding matrix. The main reason for this is that the user data is routed based on the precoding weights designed at the CCN, in the case of a centralized architecture, as the user data is several orders of magnitude greater than the control information (precoding weights). This desired property will alleviate the burden on the backhaul, and the need for the user data to be present at all the cooperating BSs is reduced. The network architecture is illustrated in Figure 1.4. In the following subsections, a brief explanation of how this can be solved is presented.

Table 1.2.: Aggregated Channel Matrix at the CCN

$\widetilde{\mathbf{H}}$	BS_1	BS_2	BS_3
UE_1	h_{11}	h_{12}	0
UE_2	0	h_{22}	h_{23}
UE_3	0	0	h_{33}

The notion of backhaul savings is partly inspired from [PBG+11] for the open systems interconnection (OSI) layered approach. Alternate network architecture is also considered for backhaul savings. To understand the subsequent sections better, a brief idea of the OSI model is presented. The OSI model is depicted in Figure 1.5. The layered structure of the communication software makes it easier to realize complex systems. Every layer of the OSI model performs a dedicated task. This provides an opportunity to design and test the layers in parallel. The

Table 1.3.: Desired precoding matrix based on $\widetilde{\mathbf{H}}$

$\widetilde{\mathbf{W}}$	UE_1	UE_2	UE_3
BS_1	w_{11}	0	0
BS_2	w_{21}	w_{22}	0
BS_3	0	w_{32}	w_{33}

lowest layer is called the physical (PHY) layer or Layer1. It is mostly concerned with channel coding and modulation. The second layer is the data link layer. In the protocol stack of the UE, this corresponds to the radio link control (RLC) and medium access control (MAC). The RLC performs the segmentation of the data packets obtained from Layer3 which is the radio resource control (RRC), and reassembly of data packets obtained from the PHY layer. The MAC layer performs the scheduling as to when the PHY layer should transmit a given data block. In this thesis, the focus is mostly on the control plane aspects related to the PHY layer and MAC layer. More details about the functions of various protocol stack layers can be found in [TS25301].

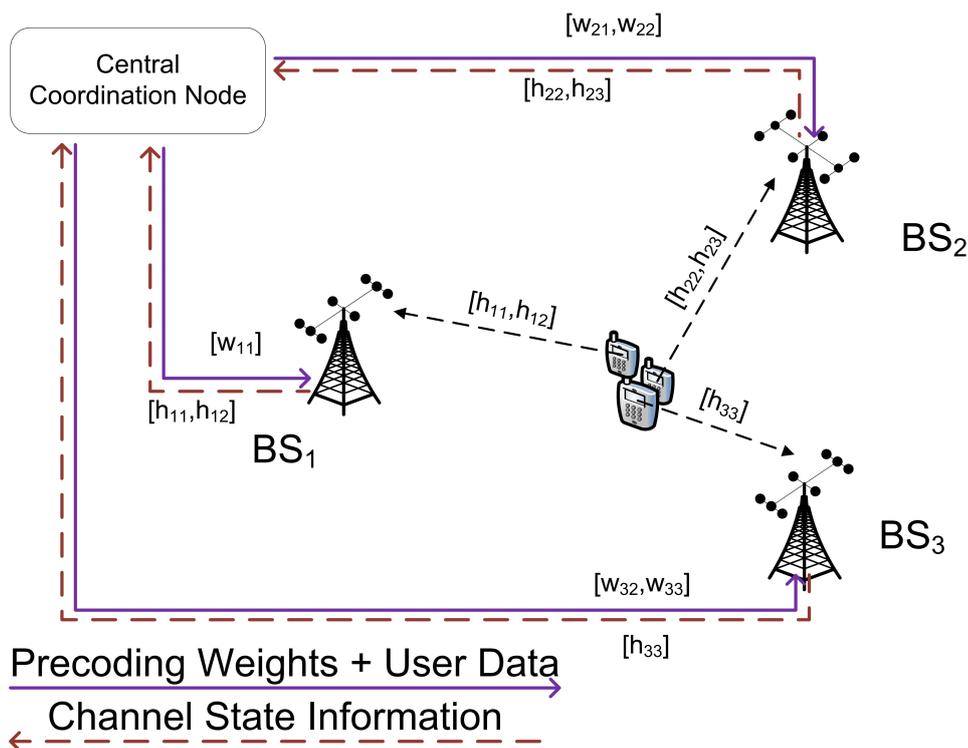


Figure 1.4.: An illustration of the equal number of CSI coefficients and the precoding weights. The uneven distribution of the CSI coefficients and the precoding weights is also captured in the backhauling links. Moreover, the user data is taking the path of the precoding weights.

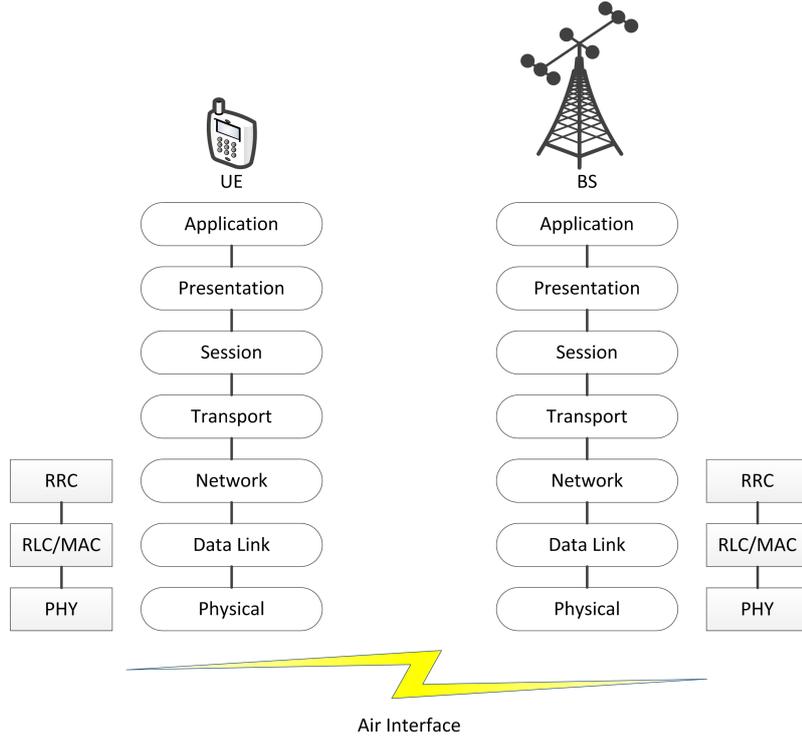


Figure 1.5.: An illustrative mapping of the OSI model that maps to the protocol stack of the UE and BS highlighted in rectangular blocks.

1.2.1. Precoding, a PHY Layer Design

To understand precoding, consider the following simplified communication system as illustrated in Figure 1.6 (a) where typically the receiver performs channel inversion to compensate the effects of the channel via equalization. The complexity of the receiver can be reduced if this is performed at the transmitter as shown in Figure 1.6 (b). However, the receiver needs to feed back the CSI to the transmitter, as shown by the dashed lines. Finally, in Figure 1.6 (c), the precoding weights are multiplied just before the symbols are upconverted, amplified and transmitted. In the case of CoMP systems, the transmitter is distributed at multiple geographically separated BSs, and in a centralized architecture, the precoder design resides in the CCN.

Now consider an aggregated channel matrix formulated at the CCN as shown in Table 1.2. Given a dummy scheduler (at the MAC layer) in a network where any given set of UEs need to be served. The question now is how to remove interference and achieve backhaul savings equivalent to the feedback savings. One can approach the problem with a PHY layer design for precoding. With a dummy scheduler, the complexity is pushed to the precoder for interference mitigation and also to achieve backhaul savings with limited CSI feedback. In this regard, a stochastic precoder is proposed in this thesis to achieve this goal. In chapter 2, stochastic optimization is discussed in general and the particular case of using particle swarm optimization as

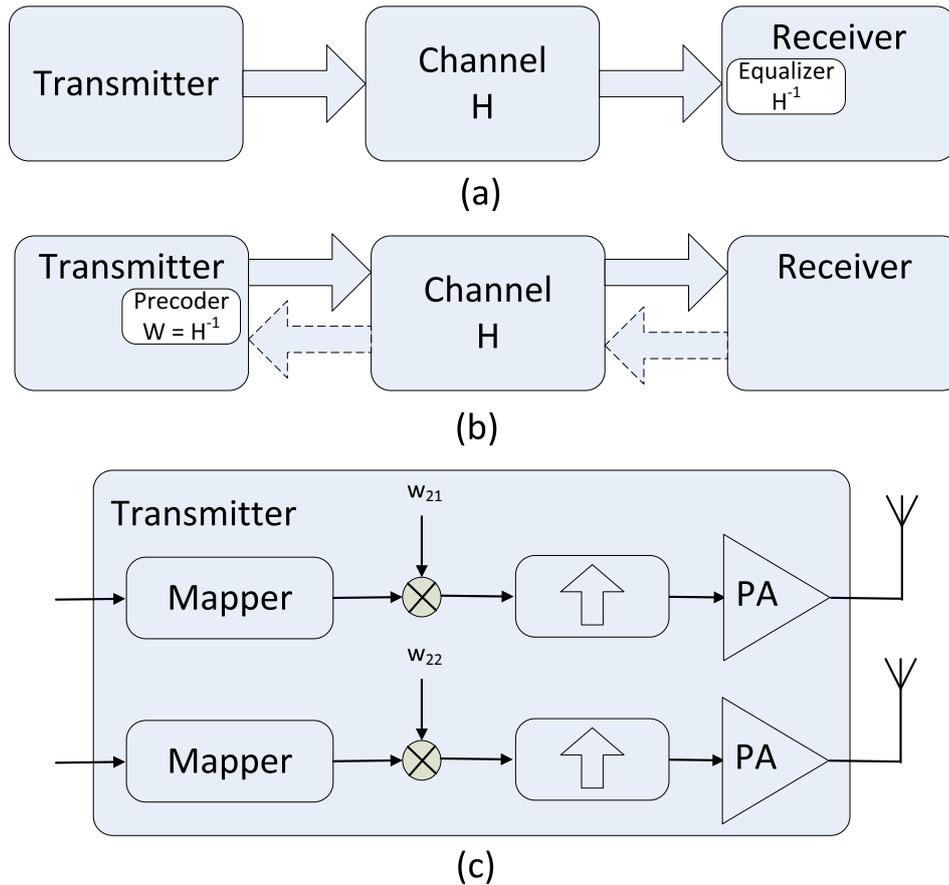


Figure 1.6.: A simplified illustration of precoding. In a centralized CoMP architecture, the precoder design resides in the CCN, as the transmitters are distributed.

a tool for designing the precoder is emphasized.

In this precoding approach, the backhaul usage is defined to be efficient when the total number of CSI coefficients fed back by the UEs is equal to the total number of precoding weights available at the cooperating BSs. This is a desired property of the precoder as mentioned in the above section. Moreover, a UE feeding back the CSI has spent the uplink resources in notifying the preferred BSs, and it is imperative that the spent resources are made worthwhile for that UE.

1.2.2. Scheduling, a MAC Layer Design

Alternately, for a given frequency-time resource, the goal of interference mitigation and backhaul savings comparable to the limited CSI feedback can be achieved with a MAC layer design. In this regard, a simple precoder such as zero forcing (ZF) is considered. The simplicity of this linear precoding approach is very much preferred from an implementation point of view. This means that the complexity needs

to be handled by the scheduler, residing at the CCN. In [SSB+09], it was shown that a simple user selecting procedure combined with a suboptimal ZF beamformer performs close to the upper bound obtained with joint multicell dirty paper coding (DPC). This is achieved asymptotically with the number of UEs per cell under equal per-cell power constraint.

In this MAC layer approach, the backhaul usage is defined to be efficient when the total number of precoding weights is less than or equal to the total number of CSI coefficients. This is primarily due to the scheduling constraint where a given set of UEs that feed back the CSI coefficients is not guaranteed to be served.

1.2.3. Network Architecture

A decentralized architecture is a centralized architecture without the CCN. The removal of the CCN relaxes the stringent constraints on the backhaul latency. However, the CSI needs to be present at the coordinating BSs. Hence, the UEs need to broadcast the CSI coefficients. This novel approach was proposed in [PHG08] and it is illustrated in Figure 1.7. This approach could be very useful in designing a network whose backhaul is limited. In [TPK11], a decentralized beamforming solution under limited backhauling is proposed, where the CSI is locally available at the BSs. However the BSs exchange some cross-interference parameters with adjacent neighbors. Alternatively, sharing the scheduling information between coordinating BSs can yield a user data rate that is comparable to the centralized architecture. Any form of sharing between BSs requires some backhaul support. Hence, such an architecture could be viewed as though every BS has the CCN, which could be treated as a logical entity.

In this decentralized network architecture with broadcasting of CSI coefficients, the reduction in the number of hops in the backhaul, complimented with sharing the scheduling information instead of CSI can be regarded as an improvement in the reduction in the backhaul latency. However, this approach could be treated as a step towards efficient usage of the backhaul.

In the next chapter, the precoder design based on particle swarm optimization is presented.

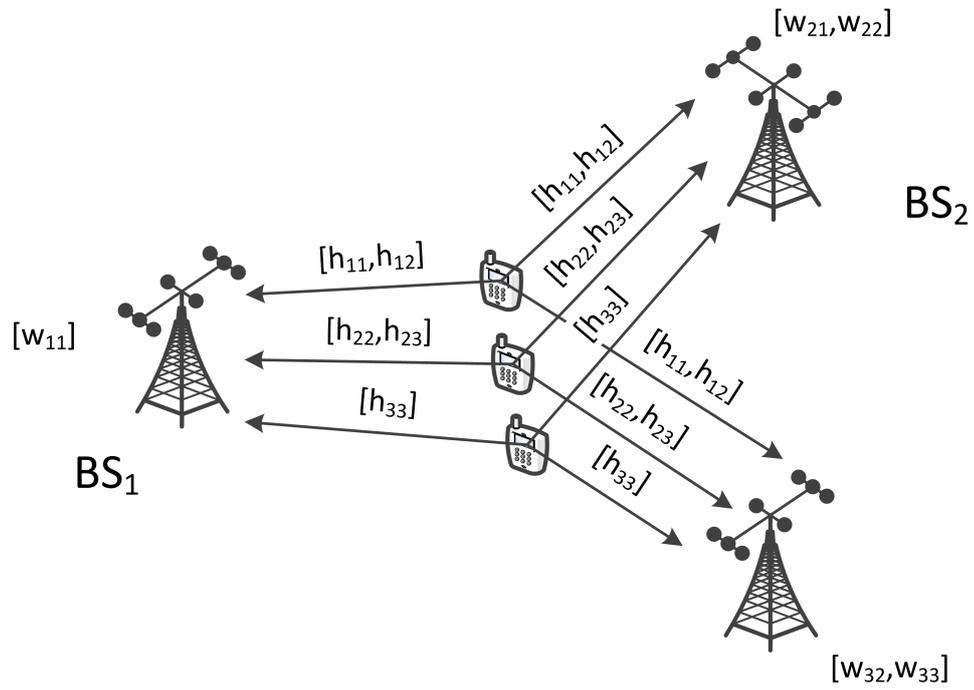


Figure 1.7.: An illustration of the UEs broadcasting the CSI over the air in a decentralized network architecture.

2. Particle Swarm Optimization

“I wandered lonely as a Cloud
That floats on high o’er Vales and Hills,
When all at once I saw a crowd
A host of dancing Daffodils;
Along the Lake, beneath the trees,
Ten thousand dancing in the breeze.”

— William Wordsworth

In this chapter, a brief introduction to stochastic optimization algorithms inspired from nature is presented. The question of how one can extract the power of stochastic algorithms, is also touched upon. The design of particle swarm optimization (PSO) as a tool for precoding is presented in particular. Nature provides a lot of inspiration to gain insights into the working forces around us. An interesting part is how evolution has brought forth optimization as one of its core elements. Evolutionary algorithms are stochastic algorithms whose driving force is optimization. There are various evolutionary algorithms, in particular the movement of ants, the swarming of birds and genetic algorithms are briefly presented in this chapter.

The movement of ants has given rise to the ant colony optimization, where the ants are involved in laying a scent trail along the path to find food. Most ants wander in search for food, however the ones that find food return to their colony’s nest with it, at the same time marking the trail. Other ants tend to follow this trail. In effect, the path to the food and the nest are optimized towards the best path. The fascinating thing about ants is that there is apparently no leader and the goal that they achieve is much bigger than what can be achieved as an individual. For example, ants making an ant-bridge to get other ants to the food source is indeed a remarkable feat. The foraging of ants is shown in Figure 2.1, where the ants are taking the scented path. This is mainly due to the cooperation between ants, helping in the survival of the species.

A flock of birds or a shoal of fish or a swarm of bees tend to move together as a group. The fish tend to avoid the shark by moving in a group, thus making it harder for the predator to catch its prey. The birds move together looking for food, as more eyes can increase the chances of finding food. Scientists simulating the movement of the flock of birds discovered that the birds were performing optimization [KE95]. In Figure 2.2, the flock of birds can be seen flying together. This helps in reducing the drag and the effort needed for flying.



Figure 2.1.: Foraging of ants at Randers rainforest, Denmark.

Genetic algorithms are evolutionary algorithms derived from the workings of the genetic material such as the chromosomes. These undergo biological changes that are favorable to sustain their species. These changes are evolutionary in nature when one looks at certain characteristics for many generations. The features that are not useful are degenerated, for example the appendix in humans. Hence, the individual of a given species tends to introduce changes and those changes that are useful are carried forward to the next generation. The variables in a given optimization problem need to be translated into a string of digits called chromosomes. Interested readers are requested to refer to [W08] for more details on the algorithms discussed above.

On a philosophical note, the notion of evolutionary algorithms can be applied to the human society where diversity in the population working together can achieve the global optimum compared to a society that is confined within itself.

Stochastic algorithms are used in designing hardware. For example, PSO is used for designing chipsets for lowering the heat dissipation or the run length of wires in a given circuitry. It is also used for designing antennas with a desired side-lobe level or the antenna element positions in a nonuniform array [JR07]. A comprehensive analysis of the publications on the applications of PSO is presented in [P08]. PSO has been proposed to be used in some parts of a communication system. Limiting ourselves to the scope of this thesis, PSO has been proposed to find the optimal precoding vector that maximizes the throughput in a MU-MIMO system [FWL09]. It is also used for optimizing the scheduling in the downlink for a MU-MIMO system [HLY+09]. Apart from [P08], PSO was also recently proposed in a MIMO-OFDM receiver for the initialization of channel estimation [KHT+11].

One can argue that stochastic optimization may give a solution to a given problem, however the solution is not easily tractable and may not guarantee a global optimum, unless special triggers are introduced within the algorithm. In particular, for PSO to achieve global optimum, triggers such as random PSO or multistart PSO need to be considered [E05]. In the field of communication engineering, technology such



Figure 2.2.: Birds flying together to minimize the drag. This picture is taken by Peter M. Prehn, a Flickr user and it is used here under  license.

as convex optimization is preferred over stochastic optimization. This is due to the simple fact that closed form solutions are preferred with the emphasis on linear algebra as it is generally easier to implement them on digital signal processors. In the following section, a counter argument is presented as to why stochastic optimization should be considered.

2.1. The Power of Stochastic Algorithms

Given a problem, typically one needs to reformulate or relax the problem so that the problem becomes convex. Then, the optimization variables need to be well formulated for the convex optimization machinery to take over [CVX12]. Reformulating problems to be convex is an art of its own, as one needs to be careful with the approximations that make the problem convex. One may end up solving a different problem than the actual problem at hand, as they are not equivalent. However, looking at the brighter side, with convex optimization, solving the dual problem gives access to the bound for the original primal problem. While in the case of biologically inspired stochastic optimization algorithms, all one needs to do is map a given problem to a stochastic optimization tool such as PSO. The problem solved by PSO will be the actual problem solved. However, the solution obtained may not be optimal or it could be difficult to prove that it is optimal. Nevertheless, one can safely say that the PSO provides a useful solution.

There are various advantages and disadvantages of these tools. The most important aspect of a stochastic algorithm is that it provides a solution to any given

problem. The key aspect that one can learn from using a stochastic algorithm is that there is room for improving the existing linear techniques as shown in Part II and the Appendix of this thesis. To illustrate this notion, consider the cumulative distribution function (CDF) of the user data rate as shown in Figure 2.3. In this illustration, a simple linear precoding technique such as the ZF or minimum mean square error (MMSE), a stochastic approach such as the PSO and an “unknown” upper bound are presented under limited feedback (LFB) of CSI and limited backhauling (LBH) of the precoding weights. The rightmost curve denotes the nonlinear dirty paper coding (DPC) with full feedback (FFB) and full backhauling (FBH). Recall the aggregated channel matrix considered in Table 1.2 and the desired precoding matrix in Table 1.3. The zeros in the aggregated channel matrix are due to the LFB and the zeros in the desired precoding matrix are due to the LBH. When these matrices are full, i.e., without any zeros, the aggregated channel matrix and the precoding matrix are regarded as FFB and FBH, respectively. For the given problem at hand, where there is LFB at the CCN and LBH, there is no reasonable upper bound, i.e., the upper bound is mostly obtained under relaxed conditions. For example, an upper bound is obtained when there is no interference, thereby making the upper bound very loose. However, the PSO provides a solution to the given problem, wherein the linear ZF model could be adapted to perform better or at least as close as the stochastic algorithm. From an engineering point of view, by convention, linear techniques are very much preferred for implementation. Hence, the power of stochastic algorithms should be embraced as they act as inputs to enhance the design of linear techniques.

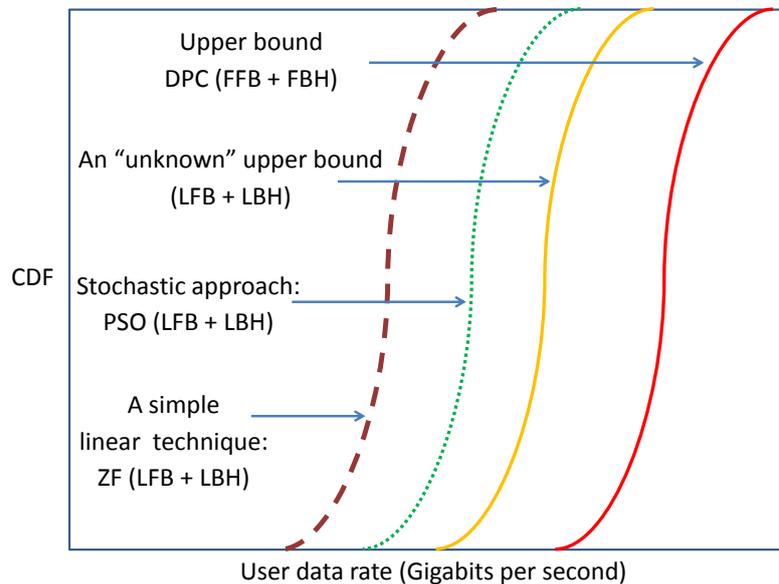


Figure 2.3.: An illustration of the use of stochastic algorithm in the context of a known linear technique and an “unknown” upper bound.

2.2. Particle Swarm Optimization as a Tool for Precoding

In the following section, PSO is presented as a tool for precoder design. In this regard, the basic PSO is introduced. Recall that with limited CSI feedback, the aggregated channel matrix at the CCN could be as shown in Table 1.2. The aim is to use PSO to invert this matrix, $\widetilde{\mathbf{H}}$ where $h_{mk} \in \mathbb{C}$ is the CSI of the channel from the k th BS to the m th UE. The total number of coefficients in the $\widetilde{\mathbf{H}}$ based on Table 1.2 is 5 complex coefficients. For efficient backhauling, the precoding matrix $\widetilde{\mathbf{W}}$ should be as shown in Table 1.3, where the number of coefficients in $\widetilde{\mathbf{W}}$ is 5 complex coefficients. Therefore, the total number of nonzero coefficients in the precoding matrix is $2 \cdot 5 = 10$ due to the real and imaginary component of each weight. Let the total number of particles or birds in the swarm be q . Let the matrix $\mathbf{X}(i, j)$ denote the i th particle carrying the j th real valued precoding weight and $\mathbf{X}(i, j + 1)$ denote the i th particle carrying the j th imaginary part of the precoding weight, i.e., the mapping of the i particle is

$$\mathbf{X}(i, j) \leftarrow \Re\{\widetilde{\mathbf{W}}(k, m)\} \quad (2.1)$$

$$\mathbf{X}(i, j + 1) \leftarrow \Im\{\widetilde{\mathbf{W}}(k, m)\}. \quad (2.2)$$

Initially the non-zero weights of $\widetilde{\mathbf{W}}$ are treated as place holders and the zeros are presented to map the backhaul savings equivalent to the feedback savings. The size of the matrix \mathbf{X} is $q \times 10$. The position and the velocity of the particles are stochastically initialized as

$$\mathbf{X}(i, j) = x_{\min} + r \cdot (x_{\max} - x_{\min}) \quad (2.3)$$

$$\mathbf{V}(i, j) = \frac{1}{\Delta t} \left(-\frac{(x_{\max} - x_{\min})}{2} + s \cdot (x_{\max} - x_{\min}) \right) \quad (2.4)$$

where r and s are drawn from the uniform distribution in the interval $[0, 1]$. The maximum starting position of a particle is intuitively chosen as $x_{\max} = \frac{1}{\max|\widetilde{\mathbf{H}}(i, j)|}$ and $x_{\min} = -x_{\max}$. The intuition is that it aids the swarm to converge. Now the particles are allowed to fly through the search space by evaluating the objective function such as maximizing sum rate or weighted interference minimization. For each iteration, each particle will keep track of its own best value achieved so far by saving the corresponding precoding weights as $\mathbf{X}^{\text{pb}}(i, :)$. There will be one particle in the swarm denoted as \mathbf{x}^{sb} that has the best value achieved. After every iteration, each particle needs to update its velocity, and its corresponding position can be calculated based on the velocity. This is performed based on the belief that each particle has on the other. The velocity update is calculated as

$$\mathbf{V}(i, j) \leftarrow w \cdot \mathbf{V}(i, j) + c_1 \cdot p \cdot \left(\frac{\mathbf{X}^{\text{pb}}(i, j) - \mathbf{X}(i, j)}{\Delta t} \right) + c_2 \cdot q \cdot \frac{\mathbf{x}^{\text{sb}}(j) - \mathbf{X}(i, j)}{\Delta t}, \quad (2.5)$$

where w is the inertia weight acting as a bias, as w is initially a value greater than one, giving the swarm a chance to explore the search space. As the particles begin to swarm, the w is damped to a value less than one. The c_1 is the cognitive component that weighs its own self belief, while c_2 is the social component as to how much a given particle can believe in the swarm's best particle. These beliefs are stochastic in nature, governed by p and q that are drawn from the uniform distribution in the interval $[0, 1]$.

Finally, when the PSO algorithm has converged, the \mathbf{x}^{sb} is demapped to obtain the precoding matrix $\bar{\mathbf{W}}$. Hence, a backhaul reduction can be obtained that is equivalent to the feedback reduction, in terms of the number of coefficients. In the next section, the advantages and disadvantages of using PSO are discussed.

2.3. Advantages and Disadvantages

The PSO has various advantages compared to other stochastic algorithms such as ant colony optimization or genetic algorithm. The primary reason is that there is very little bookkeeping that needs to be performed in the algorithm on comparative scales. In the context of precoding with limited CSI feedback, where any given set of users needs to be served, the PSO can dynamically adjust to the number of precoding weights that need to be optimized. More importantly, PSO guarantees a solution for the given problem.

Based on the analysis of the PSO algorithm for precoding, the complexity is $\mathcal{O}(cM^2KN_T)$, where M is the number of single antenna UEs and K is the number of BSs with N_T antennas. Furthermore, c is the number of iterations required for the PSO algorithm to converge. In contrast, the complexity of ZF is $\mathcal{O}(M^2KN_T)$. Hence, the PSO could take a longer time to reach a solution, due to the iterative nature of the algorithm. The other disadvantage foreseen is the performance of the PSO when the size of the problem is increased. Understanding how the PSO scales is an on-going work. However, in the context of precoder design with PSO, the algorithm is executed at the CCN whose computational power can be harnessed with better implementation of PSO. Moreover, the clustering algorithms for grouping BSs and UEs aim to reduce the feedback and backhaul overhead, resulting in a problem size that is manageable. In this thesis, the basic PSO whose parameters are not optimized, is used, i.e., a basic PSO from the field of swarm intelligence is used as it is. There is scope for improving the basic PSO parameters. Also, there is scope for the global PSO such as the multistart PSO, where the global solution is guaranteed [E05].

3. Organization of this Thesis and Included Papers

“What day is it,?” asked Pooh.

“It’s today,” squeaked Piglet.

“My favorite day,” said Pooh,

— A.A. Milne

The focus of this thesis is to find solutions to efficiently use the backhaul resources in a CoMP system. The papers presented in this work can be categorized based on the following: (a) backhaul reduction based on a PHY layer approach, (b) backhaul reduction based on a MAC layer approach, and finally (c) backhaul reduction based on a decentralized network architecture. The papers can be summarized as depicted in Table 3.1.

Table 3.1.: A high level view of the contributions in this thesis

CoMP	Backhaul reduction approaches	Contributions in
Network	Decentralized architecture	Paper D
MAC	Scheduling	Paper C
PHY	PSO based Precoding	Paper A/B

3.1. Paper A - Partial Joint Processing with Efficient Backhauling in Coordinated MultiPoint Networks

In this PHY layer approach for backhaul reduction, the state-of-the-art physical layer block diagonalization solution is compared with PSO. The comparison can only be performed when the block diagonalization of the aggregated channel matrix is feasible. The infeasibility is brought about due to the limited CSI feedback achieved by relative thresholding.

3.2. Paper B - Partial Joint Processing with Efficient Backhauling using Particle Swarm Optimization

This is an extension of the above conference paper, where the PSO is analyzed in greater detail for backhaul load reduction. The PSO is guaranteed to work without any scheduling constraint. The objective function of sum rate maximization is shown to be biased towards users with good signal to interference plus noise ratio (SINR) compared to the low SINR users. A new metric called *weighted interference minimization* is proposed, where the objective function is to minimize interference and improve the weak SINR users. In Appendix A, the weighted interference minimization is compared with the maximization of the minimum SINR user, and the benefits of considering weighted interference minimization is highlighted.

The PSO was used with real field measurement data where the statistical uncertainty of CSI or imperfect CSI were considered. Various algorithms from different partners of the ARTIST4G consortium [ARTIST4G] were considered. It was found that the PSO outperformed all the other algorithms in the scenarios considered. Some of the interesting results are provided in Appendix B, which complement the work performed on PSO.

3.3. Paper C - Scheduling for Backhaul Load Reduction in CoMP

In this MAC layer approach, scheduling is explored for backhaul load reduction, where a subset of UEs and BSs combinations are considered. Constrained and unconstrained scheduling are proposed. In particular the constrained scheduling approach outperforms the state of the art block diagonal approach, in terms of the average sum rate per backhaul use.

3.4. Paper D - On the Potential of Broadcast CSI for Opportunistic Coordinated MultiPoint Transmission

A decentralized network architecture is considered for backhaul load reduction. In this setting, the CSI coefficients broadcasted by the UE undergo a certain probability of error as they are received at different BSs, hence, giving rise to precoding loss and scheduling loss. It is shown that with a minimal exchange of scheduling information, the decentralized architecture can achieve the rates comparable to the centralized approach that makes use of the central coordination node, thereby reducing the stringent latency constraints in the backhaul.

In this work, the phase information of the CSI alone is considered to undergo errors while the amplitude or channel quality indicator (CQI) can be assumed to

be error free. This assumption is reasonable and goes well with the LTE standards, where only 4 bits are used for feeding back the CQI. Moreover, one could protect the CQI with robust channel codes, whose overhead is not significant.

3.5. Visions and Future Work

Broadcasting of CSI coefficients in a decentralized architecture is attractive from the backhaul latency point of view. The CSI feedback can be envisioned to be of two levels. Firstly, the quantity of the feedback should be focused upon, i.e., how much CSI should a given UE feed back, and secondly, what should be the quality of the CSI that becomes available at the BSs. Should the BSs that decode the broadcasted CSI be clustered? Reducing the quantity of feedback in any architecture will cause limited CSI feedback and PSO will be an attractive solution to mitigate interference. However, one needs to find a trade off with distributing the complexity of mitigating interference to the PHY layer, MAC layer and the network architecture as well. Interference mitigation should be considered at various layers in the protocol stack. As an analogy to the OSI model of the protocol stack in cellular communications, segmentation and reassembly of packets is performed at various layers.

As part of the future work, the bounds for limited feedback and the equivalent backhaul reduction beg for some attention. This can provide inputs as to whether, the complexity should be handled by the precoder or the scheduler or if they should work together. This is still an open question.

The recent trends in the widespread use of communications has driven the cellular network operators to consider dense deployment of BSs to improve capacity [LFJ+13]. In dense networks, interference could potentially destroy the gains. Hence, suitable interference mitigation techniques should be considered. Dynamic clustering could play an important role to decide which BSs should serve a given UE. Feeding back the CSI could be an overhead, possibly hybrid automatic repeat request (HARQ) with dynamic cell selection could be considered to reduce the backhaul dependency.

3.6. List of Related Publications

- B. Makki, **T.R. Lakshmana**, T. Eriksson and T. Svensson, “On Resource Allocation in Coordinated Multi-point Networks”, to be submitted to *IEEE Trans. on Commun.*, Feb. 2013.
- C. Botella, L. Brunel, C. Ciochina, L. Cottatellucci, V. D’Amico, P. Kerret, D. Gesbert, J. Giese, N. Gresset, J. Guillet, H. Halbauer, X. Jiang, H. Khanfir, **T.R. Lakshmana**, B. Melis, D. Sabella, S. Saur, T. Svensson, P. Tortelier, W. Zirwas, “ARTIST4G, D1.3 Innovative scheduling and cross-layer design techniques for interference avoidance”, EU FP7 INFSO-ICT-247223, Mar. 2011.

[Online] Available: <https://ict-artist4g.eu/projet/deliverables> [Accessed: Feb. 2013].

- V. D’Amico, H. Halbauer, D. Aronsson, C. Botella, S. Brueck, C. Ciochina, T. Eriksson, R. Fritzsche, D. Gesbert, J. Giese, N. Gresset, **T.R. Lakshmana**, B. Makki, B. Melis, R.O. Abildgaard, M.L. Pablo, H. Phan, T. Dinh, S. Saur, M. Sternad, T. Svensson, R. Zakhour, W. Zirwas, “ARTIST4G, D1.2 Innovative advanced signal processing algorithms for interference avoidance”, EU FP7 INFISO-ICT-247223, Dec. 2010. [Online] Available: <https://ict-artist4g.eu/projet/deliverables> [Accessed: Feb. 2013].
- **T.R. Lakshmana**, C. Botella, T. Svensson, X. Xu, J. Li, X. Chen, “Partial joint processing for frequency selective channels”. in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2010.
- A.T. Toyserkani, **T.R. Lakshmana**, E.G. Ström, A. Svensson, “A Low-Complexity Semi-Analytical Approximation to the Block Error Rate in Nakagami-m Block Fading Channels”, in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2010.

Bibliography

- [ARTIST4G] ARTIST4G: Advanced Radio Interface Technologies for 4G Systems. [Online] Available: <https://ict-artist4g.eu/> [Accessed: Jan. 2013].
- [ARTD12] ARTIST4G D1.2, “Innovative advanced signal processing algorithms for interference avoidance”, ARTIST4G technical deliverable, Dec. 2010. [Online] Available: <https://ict-artist4g.eu/projet/deliverables>, [Accessed: Dec. 2012].
- [BSX+10] C. Botella, T. Svensson, X. Xu, H. Zhang, “On the performance of joint processing schemes over the cluster area,” in *Proc. IEEE Veh. Technol. Conf.*, May 2010.
- [ComSoc] IEEE Communications Society, A Brief History of Communications, 2nd Edition, 2012.
- [CRP10] G. Caire, S.A. Ramprasad, H.C. Papadopoulos, “Rethinking network MIMO: Cost of CSIT, performance analysis, and architecture comparisons,” *Information Theory and Applications Workshop (ITA)*, Jan. 31-Feb. 5, 2010.
- [CVX12] CVX Research, Inc. CVX: Matlab software for disciplined convex programming, version 2.0 beta. <http://cvxr.com/cvx>, Sep. 2012.
- [E05] A. Engelbrecht, Fundamentals of computational swarm intelligence, John Wiley & Sons, 2005.
- [FWL09] S. Fang, G. Wu, S-Q. Li, “Optimal multiuser MIMO linear precoding with LMMSE receiver”, *EURASIP J. on Wireless Commun. and Netw.*, Article ID 197682, 2009.
- [G05] A. Goldsmith, Wireless Communications, Cambridge University Press New York, NY, USA, 2005, ISBN:0521837162.
- [GHH+10] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, “Multi-Cell MIMO Cooperative Networks: A New Look at Interference,” *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.
- [HLY+09] Y. Hei, X. Li, K. Yi, H. Yang, “Novel scheduling strategy for down-link multiuser MIMO system: particle swarm optimization”, *Science in China F*, 52(12), pp. 2279–2289, 2009.

- [IDM+11] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H. Mayer, L. Thiele, V. Jungnickel, “Coordinated multipoint: Concepts, performance, and field trial results,” *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102-111, Feb. 2011.
- [JR07] N. Jin and Y. Rahmat-Samii, “Advances in particle swarm optimization for antenna designs: real-number, binary, single objective and multiobjective implementations,” *IEEE Trans. on Antennas Propag.*, vol. 55, no. 3, pp. 556– 567, 2007.
- [KFV06] MK Karakayali, GJ Foschini, RA Valenzuela, “Network coordination for spectrally efficient communications in cellular systems”, *IEEE Wireless Commun. Mag.*, vol 13, no. 4, pp. 56-61, Aug. 2006.
- [KE95] J. Kennedy, R.C. Eberhart, “Particle swarm optimization,” in *Proc. IEEE Int. Conf. on Neural Netw.*, pp. 1942–1948, 1995.
- [KHT+11] C. Knievel, P.A. Hoeher, A. Tyrrell, G. Auer, “Particle swarm enhanced graphbased channel estimation for MIMO-OFDM”, in *proc. IEEE Veh. Technol. Conf. Fall*, May 2011.
- [LBS12] J. Li, C. Botella, and T. Svensson, “Resource allocation for clustered network MIMO OFDMA systems,” *EURASIP J. Wireless Comm. and Netw.*, vol. 2012, 2012.
- [LFJ+13] S. Landström, A. Furuskär, K. Johansson, L. Falconetti and Fredric Kronstedt, “Heterogeneous Networks (HetNets) – an approach to increasing cellular capacity and coverage”, [Online] Available: <http://www.ericsson.com> [Accessed: Jan. 2013].
- [MF11] P. Marsch, and G. Fettweis (eds), *Coordinated Multi-Point in Mobile Communications: From Theory to Practice*. Cambridge University Press. 2011.
- [P08] R. Poli, “Analysis of the Publications on the Applications of Particle Swarm Optimisation”, *J. of Artificial Evol. and Appl.*, Vol. 2008, No. 3, Jan. 2008.
- [PBG+08] A. Papadogiannis, H.J. Bang, D. Gesbert, and E. Hardouin, “Down-link Overhead Reduction for Multi-Cell Cooperative Processing enabled Wireless Networks,” in *Proc. IEEE Intl. Symposium on Pers., Indoor and Mobile Radio Commun. (PIMRC)*, Sep. 2008.
- [PBG+11] A. Papadogiannis, H.J. Bang, D. Gesbert, E. Hardouin, “Efficient selective feedback design for multicell cooperative networks,” *IEEE Trans. on Vehicular Tech.*, vol. 60, no. 1, pp. 196-205, Jan. 2011.

- [PGH08] A. Papadogiannis, D. Gesbert, and E. Hardouin, “A dynamic clustering approach in wireless networks with multi-cell cooperative processing,” in *Proc. IEEE Intl. Conf. on Commun. (ICC)*, May 2008.
- [PHG08] A. Papadogiannis, E. Hardouin, and D. Gesbert, “A Framework for Decentralising Multi-Cell Cooperative Processing on the Downlink,” in *Proc. IEEE GLOBECOM Workshops*, Dec. 2008.
- [RS10] Rohde and Schwarz, “LTE-Advanced Technology Introduction,” White Paper, 2010.
- [SSB+09] O. Somekh, O. Simeone, Y. Bar-Ness, A. Haimovich, and S. Shamai, “Cooperative multicell zero-forcing beamforming in cellular downlink channels,” in *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3206–3219, Jul. 2009.
- [TPK11] A. Tölli, H. Pennanen, and P. Komulainen, “Decentralized Minimum Power Multi-cell Beamforming with Limited Backhaul Signalling,” in *IEEE Trans. on Wireless Comm.*, vol. 10, no. 2, pp. 570 - 580, Feb. 2011.
- [TR36819] 3GPP TR 36.819-b10, “3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Coordinated multi-point operation for LTE physical layer aspects (Release 11) ,” Dec. 2010.
- [TS25301] 3GPP TS 25.301-b00, “3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Radio Interface Protocol Architecture (Release 11) ,” Sep. 2012.
- [W08] M. Wadhe, *Biologically Inspired Optimization Methods: An Introduction*, WIT press, Southampton, Boston, 2008, ISBN: 978-1-84564-148-1.
- [ZCA+09] J. Zhang, R. Chen, J. Andrews, A. Ghosh, R. Heath, Networked MIMO with clustered linear precoding,” in *IEEE Trans. Wirel. Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.

Part II.
Included Papers

Paper A

Partial Joint Processing with Efficient Backhauling in Coordinated MultiPoint Networks

Tilak Rajesh Lakshmana, Carmen Botella and Tommy Svensson

Published in
IEEE Vehicular Technology Conference (VTC), May 2012

©2012 IEEE

The layout has been revised.

4. Partial Joint Processing with Efficient Backhauling in Coordinated MultiPoint Networks

Abstract

Joint processing between base stations is a promising technique to improve the quality of service to users at the cell edge, but this technique poses tremendous requirements on the backhaul signaling capabilities. Partial joint processing is a technique aimed to reduce feedback load, in one approach the users feed back the channel state information of the best links based on a channel gain threshold mechanism. However, it has been shown in the literature that the reduction in the feedback load is not reflected in an equivalent backhaul reduction, unless additional scheduling or precoding techniques are applied. The reason is that reduced feedback from users yields sparse channel state information at the Central Coordination Node. Under these conditions, existing linear precoding techniques fail to remove the interference and reduce backhaul, simultaneously, unless constraints are imposed on scheduling. In this paper, a partial joint processing scheme with efficient backhauling is proposed, based on a stochastic optimization algorithm called particle swarm optimization. The use of particle swarm optimization in the design of the precoder promises efficient backhauling with improved sum rate.

Keywords—Joint Processing, Zero Forcing, Backhaul load reduction, Particle Swarm Optimization, Stochastic Optimization

4.1. Introduction

Future cellular communication systems tend to have a frequency reuse factor of one, causing intercell interference and reducing user experience close to the cell-edge. Joint Processing (JP) between Base Stations (BSs) is one of the techniques that falls in the framework of Coordinated MultiPoint (CoMP) transmission [1]. In downlink JP, the user receives its data from multiple coordinating BSs.

In a typical Centralized Joint Processing (CJP) approach, the cluster of BSs jointly coordinates and transmits the data to the intended user, without causing interference to other users. This poses users to feed back the Channel State Information (CSI) of all the BSs in the cluster to their serving BS. Then, the CSI needs

to be forwarded over the backhaul towards the Central Coordination Node (CCN) to precancel the interference via BeamForming (BF) and power allocation. This non-casual availability of CSI at the CCN for interference avoidance can be treated as casual for a stationary channel, but needs regular updates for non-stationary channels. Nevertheless, the need for the CSI being available at the CCN and for sending the precoding weights and user data from the CCN to the corresponding BSs puts tremendous requirements on backhauling. To alleviate them, clusters of BSs are usually arranged. The clustering techniques can be divided into network-centric or user-centric, depending on where the clustering decision is carried out. To this end, Partial Joint Processing (PJP) has been proposed to reduce the CSI feedback load [2].

PJP can be viewed as a general framework for feedback and backhaul reduction. In the particular approach considered in this paper, a CCN or the serving BS might instruct the User Equipments (UEs) to report the CSI of the links in the cluster of BSs whose channel gain fall within an active set threshold or window, relative to their best link (usually the serving BS). With this PJP scheme, feedback load is reduced, as CSI of only a subset of BSs is fed back per UE. This subset is also referred to as an active set. Note that this user-centric clustering technique results in the formation of overlapping active set of BSs for each user, and the user should preferably only receive its data from the set of BSs included in its active set. The CSI being available at the CCN are marked as active links and those that are not available (or not reported) are marked inactive. The CCN forms an aggregate channel matrix based on these active and inactive links for interference avoidance. As a result, the aggregated channel matrix is now sparse, due to the reduced CSI feedback giving rise to inactive links, which are modeled as zeros.

In JP, linear BF techniques such as Zero Forcing (ZF) can be used for interference avoidance, as long as the aggregated channel matrix is well conditioned for inversion. It has been shown in the literature that the reduction in the CSI feedback load is not necessarily reflected in an equivalent BF backhaul reduction, unless additional scheduling or precoding techniques are applied [3], or unless the aggregated channel matrix is diagonal or block-diagonal. In other words, when calculating the ZF BF based on the sparse aggregated channel matrix, one inactive link may be mapped into a non-zero BF weight for that link. This causes unnecessary backhauling, since the UE has reported that link as inactive and that BS is then outside the active set of that UE, i.e., the resources at this BS can be used to serve other UEs. A brute force approach would be that the CCN might resort to nulling the BF weights where the links are inactive, but this might lead to inefficient power allocation and increased interference, resulting in reduced sum rate of the system in the cluster area.

To the best of our knowledge, this problem has only been addressed in [3], where two solutions are proposed, one based on scheduling (MAC layer approach) and the other based on ZF precoding (PHY layer approach). In the scheduling solution of [3], a suitable set of UEs are selected for transmission such that the sparse aggregated channel matrix available at the CCN is block-diagonal and hence, invertible. These

suitable sets are formed by arranging disjoint active sets in each time slot, i.e., each BS involved in serving a UE belongs to only one active set. The drawback of this approach is that, in each time slot, a given set of disjoint BS active sets is selected for transmission; if the UEs prefer services from the same set of BSs they need to wait for their turn to be served in a TDMA fashion. Fairness is guaranteed but at the cost of UEs needing to wait for a long time. In case of the ZF precoding solution in [3], no constraints are assumed on scheduling. To reduce the backhaul load, the zeros in the sparse aggregated channel matrix are mapped to the aggregated BF matrix. The interference is reduced by formulating this as a constrained optimization problem. The proposed solution needs a well constructed aggregated channel matrix and hence, it is heavily dependent on scheduling. On the other hand, there is no linear technique existing in the literature that can invert the aggregated channel matrix with zeros (inactive links) and preserve these zeros in the transposed version of the inverse, when the aggregated channel matrix is not diagonal or block-diagonal.

In this paper, Particle Swarm Optimization (PSO), a stochastic optimization method, is proposed as a tool to design a BF that can achieve a backhaul reduction in terms of zero BF weights equivalent to the CSI feedback reduction. Compared to the ZF precoding solution in [3], this technique works on sparse aggregated channel matrices, without any constraint on scheduling. PSO has already been shown to obtain the optimal multiuser MIMO linear precoding vector, where the objective function of the PSO was to maximize the system capacity [4]. Whereas, in our paper, the PSO is used in a multicell scenario performing PJP CoMP with perfect CSI with the main objective of minimizing interference.

The paper is organized as follows, in section 4.2 the system model is described. The section 4.3 introduces PSO and discusses how the BF weights are treated as particles. Simulation results are presented in section 4.4, and section 4.5 concludes the paper. The notation used in this paper is summarized in the footnote below.

4.2. System Model

Consider a cluster of K single antenna BSs involved in the downlink transmission to M single antenna UEs located at the cluster center, as shown in Figure 4.1. Any form of intercluster interference affecting the demodulation of signals at the UE is assumed to be negligible and is thus neglected. With a frequency reuse factor of one in this layout, the transmission to a UE will cause interference to other UEs.

Notation: Boldface upper-case letters denote matrices, \mathbf{X} , boldface lower-case letters denote vectors, \mathbf{x} , and italics denote scalars, x . The $\mathbb{C}^{m \times n}$ is a complex valued matrix of size $m \times n$. The $(\cdot)^H$ is the conjugate transpose of a matrix. The $\|\cdot\|_F$ is the Frobenius norm, $\text{OffDiag}(\mathbf{X})$ is an operation on the matrix \mathbf{X} that sets the elements in the main diagonal to zero. $\mathbf{X}(i, j)$ is the (i, j) th element of matrix \mathbf{X} . $\text{vec}(\mathbf{X})$ is the vector of stacked columns of matrix \mathbf{X} and \otimes denotes the Kronecker product. $\Re\{\mathbf{X}(i, j)\}$ and $\Im\{\mathbf{X}(i, j)\}$ are the real part and the imaginary parts of $\mathbf{X}(i, j)$.

Assuming CJP, the discrete time signal received at M UEs, $\mathbf{y} \in \mathbb{C}^{M \times 1}$ is

$$\mathbf{y} = \mathbf{H}\mathbf{W}\mathbf{x} + \mathbf{n}, \quad (4.1)$$

where $\mathbf{H} \in \mathbb{C}^{M \times K}$ is the aggregated channel matrix of the form $[\mathbf{h}_1^T \mathbf{h}_2^T \dots \mathbf{h}_M^T]^T$, $\mathbf{h}_m \in \mathbb{C}^{1 \times K}$ is the channel from all the BSs in the cluster to the m th UE, $\mathbf{W} \in \mathbb{C}^{K \times M}$ is the aggregated BF matrix of the form $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_M]$, $\mathbf{w}_m \in \mathbb{C}^{K \times 1}$ is the BF for the m th UE, $\mathbf{x} \in \mathbb{C}^{M \times 1}$ is the transmitted symbols to the M UEs, and \mathbf{n} is the spatially and temporally white receiver noise with variance σ^2 , and it is uncorrelated with the transmitted symbols.

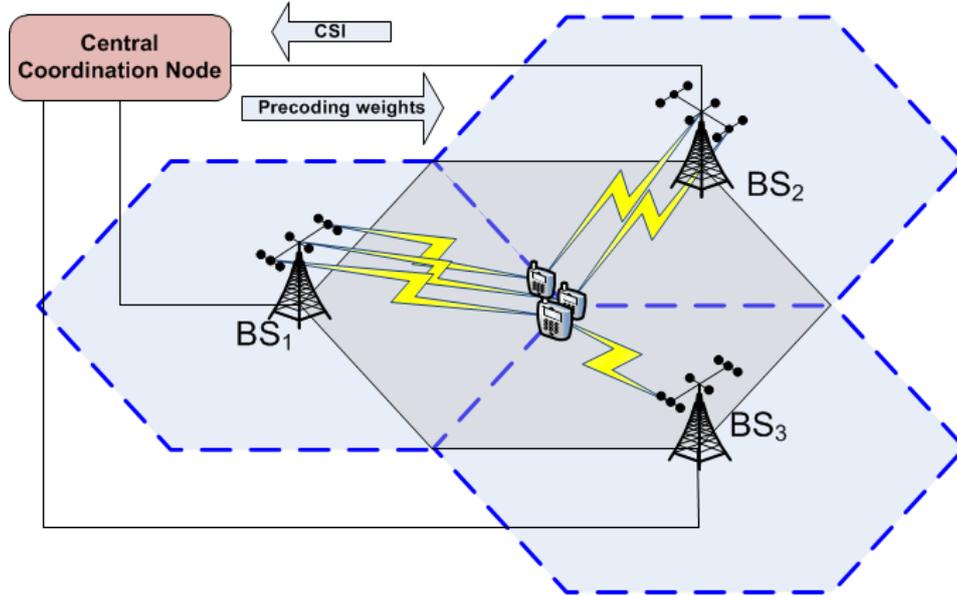


Figure 4.1.: The cluster layout, the hexagon in the middle denotes the cluster area under consideration.

When CJP is used, the CCN has a full channel matrix \mathbf{H} . In the literature, a ZF BF matrix \mathbf{W} is obtained by taking the right inverse of \mathbf{H} ,

$$\mathbf{W} = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1}. \quad (4.2)$$

CJP can be seen as a particular case of PJP when the threshold is high, such that all links are active for a given UE. For convenience, at the CCN, the active and inactive links can be represented as an active set, a binary matrix of size $[M \times K]$, whose (m, k) th element represents the (m, k) th link between the m th user and the k th BS. These elements take the value '1' and '0' representing links whose CSI is available (active) and not available (inactive), respectively [2]. Few links are active in some scenarios, e.g., small values of the active set threshold result in a sparse aggregated channel matrix $\tilde{\mathbf{H}}$ at the CCN. If $\tilde{\mathbf{H}}$ is invertible, the $\tilde{\mathbf{W}}$ thus formed may not have zeros at places where needed, i.e., a UE will receive its data from BSs outside its

active set, corresponding to its inactive BSs. For example, say UE_1 reported CSI for BS_1 , BS_3 and not for BS_2 i.e., BS_2 falls outside the active set of UE_1 , hence, $\widetilde{\mathbf{H}}(1, 2) = 0$. The CCN having the CSI of all the UEs in that cluster tries to invert the aggregated channel matrix to obtain the BF weights. These weights are only needed at BS_1 and BS_3 for UE_1 , but they might show up at BS_2 for UE_1 , due to the behavior of the pseudo-inverse involved with ZF, i.e., $\widetilde{\mathbf{W}}(2, 1) \neq 0$. This is highly undesirable as it results in extra and unnecessary backhaul load on the cluster. It should be pointed out that BS_2 being inactive is not involved in JP for serving UE_1 but BS_2 can be involved in serving other UEs. A BS that is not involved in serving any UE need not be considered in this setup at all. Due to the overlapping clusters formed with PJP, the subset of BSs reported by the UEs differ for a given frequency/time resource. Hence, a BS serving only one UE at the cluster center, should be included in the precoding design as this UE is sharing the same frequency/time resource and the interference thus generated needs to be accounted.

To realize the gains of the active set based PJP scheme, the problem at the CCN under a ZF assumption, is two fold: firstly, invert a sparse matrix and secondly, obtain null BF weights in the correct places. Hence, BF is an important challenge to be realized, especially without the need to have any special constraints on scheduling.

Every BS involved in JP has a maximum per-BS power constraint of P_{\max} . The precoding matrix $\overline{\mathbf{W}}$ is realized such that at least one of the BSs can transmit at maximum power as defined in

$$\overline{\mathbf{W}} = \left(\sqrt{P_{\max} / \left(\max_{k=1, \dots, K} \|\widetilde{\mathbf{W}}^k\|_F^2 \right)} \right) \cdot \widetilde{\mathbf{W}}, \quad (4.3)$$

where $\widetilde{\mathbf{W}}^k$ stands for the BF weights of the k th BS towards the M users.

The Signal to Interference plus Noise Ratio (SINR) for the m th UE is given as

$$\text{SINR}_m = \frac{\|\mathbf{h}_m \overline{\mathbf{w}}_m\|^2}{\sum_{j=1, j \neq m}^M \|\mathbf{h}_m \overline{\mathbf{w}}_j\|^2 + \sigma^2}, \quad (4.4)$$

and the sum rate per cell at the cluster center is given as

$$R_{\text{tot}} = \frac{1}{K} \sum_{m=1}^M \log_2 (1 + \text{SINR}_m) [\text{bps}/\text{Hz}/\text{cell}]. \quad (4.5)$$

4.3. Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a stochastic optimization algorithm inspired from the movement of a flock of birds, a shoal of fish, etc [5]. The birds are modeled as particles traveling in the search space to find food or the feasible solution of a given objective function. Their social behavior is modeled as a swarm. In [5], the algorithm simulating the social behavior was simplified and was observed to be performing optimization. PSO being metaheuristic does not guarantee a global optimum, but

when implemented as a stochastic optimization, randomness is injected into the algorithm to move away from the local solution when searching for a global optimum. In this paper, a basic PSO capable of finding an equilibrium solution with *inertia weight* and without the *craziness* operator [5], is presented. As per [6], a basic PSO does not satisfy the convergence condition for global search. PSO is chosen over classical optimization methods, as the overlapping clusters (dynamic changes in the aggregated channel matrix in every frequency/time resource) formed with PJP make the linear ZF BF technique presented in [3] difficult to realize.

The ZF precoding solution proposed in [3] for single antenna systems is simplified into a classic linear algebra problem $\mathbf{A}\mathbf{x} = \mathbf{b}$, where \mathbf{A} is formed by block diagonalizing the aggregated channel matrix; \mathbf{x} and \mathbf{b} are the vectorized BF and identity matrices, respectively. The zeros in the \mathbf{x} eliminate the columns of \mathbf{A} and the solution reduces to a classic right inverse as in (Equation 4.2). This can be written compactly as $(\mathbf{I}_M \otimes \widetilde{\mathbf{H}}) \cdot \text{vec}(\widetilde{\mathbf{W}}_e) = \text{vec}(\mathbf{I}_M)$, where the inactive links or zeros in $\widetilde{\mathbf{W}}_e$ have eliminated the columns of the Kronecker product. This elimination can give rise to an overdetermined system and the right inverse does not exist. An overdetermined system is encountered $\sim 83\%$ of the time, when this approach is applied for PJP with active set threshold of 10 dB in our evaluation setup, see section 4.4. Hence, we propose PSO to overcome the limitations in the state of the art ZF BF solution in [3], and we later show that PSO performs better.

The BF weight matrix $\widetilde{\mathbf{W}}$ is stochastically initialized, and zeros are inserted according to the active set matrix. The non-zero BF weights are mapped to the i th particle as $\mathbf{X}(i, j) \leftarrow \Re\{\widetilde{\mathbf{W}}(m, n)\}$ and $\mathbf{X}(i, j + 1) \leftarrow \Im\{\widetilde{\mathbf{W}}(m, n)\}$. The search space of the particles is initially limited to $[x_{\min}, x_{\max}]$, where $x_{\max} = 1/\max\{|\widetilde{\mathbf{H}}(i, j)|\}$. This value is chosen as the starting limit of the particles in the search space for faster convergence. PSO stochastically changes these limits in every iteration. The aim of the PSO is to optimize the position of the particles of the swarm based on the objective function. The resulting best particle is chosen as the best BF weights. In this work, two different cases are considered:

$$\text{case a) } \arg \min_{\widetilde{\mathbf{w}}_a} \left\{ \|\widetilde{\mathbf{H}}\widetilde{\mathbf{W}}_a - \mathbf{I}\|_F \right\} \quad (4.6)$$

$$\text{case b) } \arg \min_{\widetilde{\mathbf{w}}_b} \left\{ \|\text{OffDiag}(\widetilde{\mathbf{H}}\widetilde{\mathbf{W}}_b)\|_F \right\} \quad (4.7)$$

Both objective functions for the PSO are subject to $\forall i, j : \widetilde{\mathbf{H}}(i, j) = 0$ maps to $\widetilde{\mathbf{W}}_x(j, i) = 0$, where x represents case a) or case b). In case a), the identity matrix is subtracted from the product of the sparse aggregated channel matrix and the aggregated BF matrix. The identity matrix tries to ensure that all the users are fairly served, i.e., it aims for equal receive power to all the users. This is based on the ZF philosophy that $\mathbf{H}\mathbf{W} = \mathbf{I}$. The interference is minimized considering fairness between users. In case b), only the off diagonal elements of the product of the sparse aggregated channel matrix and the aggregated BF matrix is considered. The non-zero off diagonal elements represent the presence of interference in the system

involved in JP. The aim of this objective function is to minimize the interference alone.

To evaluate the objective function, the positions of the particles are demapped to form the complex BF weight, i.e., $\widetilde{\mathbf{W}}(m, n) \leftarrow \{\mathbf{X}(i, j)\} + i \cdot \{\mathbf{X}(i, j + 1)\}$, where ‘i’ is the imaginary unit. The core of the PSO algorithm is described in (Equation 4.8)-(Equation 4.11). These equations are evaluated in every iteration, i.e., $\forall i = 1, \dots, N; j = 1, \dots, n$, where the time step length is $\Delta t = 1$, N is the number of particles and n is the number of variables. (Equation 4.8) updates the velocity of the particle, where the term involving c_1 is called the *cognitive* component (weighs the self confidence of that i th particle) and the term involving c_2 is called the *social* component (weighs the reliability of other particles for that i th particle). $\mathbf{X}^{pb}(i, :) \leftarrow \mathbf{X}(i, :)$ is the best position attained by the i th particle itself carrying the best BF weights it could find, \mathbf{x}^{sb} is the best position attained by any particle carrying the best BF weights in the entire swarm. p and q are uniform random numbers in $[0, 1]$. (Equation 4.9) restricts the maximum velocity of the particle to v_{\max} , such that the particles do not diverge and (Equation 4.10) updates the position of the particle. An *inertia weight*, w , is used to bias the current velocity based on its previous value in (Equation 4.8), such that when the inertia weight is initially 1.4, being greater than 1, the particles are biased to explore the search space. When w decays to 0.4, due to a constant decay factor β in (Equation 4.11), the cognitive/social components are given more attention [7].

$$\begin{aligned} \mathbf{V}(i, j) \leftarrow & w \cdot \mathbf{V}(i, j) + c_1 \cdot p \cdot (\mathbf{X}^{pb}(i, j) - \mathbf{X}(i, j)) / \Delta t \\ & + c_2 \cdot q \cdot (\mathbf{x}^{sb}(j) - \mathbf{X}(i, j)) / \Delta t \end{aligned} \quad (4.8)$$

$$|\mathbf{V}(i, j)| < v_{\max} \quad (4.9)$$

$$\mathbf{X}(i, j) \leftarrow \mathbf{X}(i, j) + \mathbf{V}(i, j) \cdot \Delta t \quad (4.10)$$

$$w \leftarrow w \cdot \beta \quad (4.11)$$

The *cognitive* factor, c_1 , and the *social* factor, c_2 , are equal to 2 as highlighted in [5, 6]. These references also indicate that the choice of the nonlinear decreasing in the inertia weight with an initial value of 1.4 ensures that the particles cover a large search space and then the particles focus on refining the solutions. The choice of the decay factor as 0.99 indicates slow decaying in the inertia weight. The number of particles is also fixed throughout the simulation. The choice of the number of particles, $N = 30$, was found to be a reasonable compromise with the computational complexity and the ability of the PSO to find an equilibrium solution, see [6] and the references therein. The values of the PSO parameters used in this initial work are the typical values and are summarized in Table 4.1. The influence of the PSO parameters chosen for the design of the precoding matrix needs to be investigated further as part of our future work, and also with the channel data from field measurements. PSO is preferred over other genetic algorithms, as it has the least number of parameters to configure and promises better computational efficiency.

4.4. Simulation Results

Consider the scenario with $K = M = 3$ single antenna BSs/UEs involved in JP. Each BS covers a hexagonal cell of radius, $R = 0.5$ kms. The UEs are placed at the cluster center along an ellipse with semi-major and semi-minor axis of length $\frac{R}{16}$ and $\frac{h/2}{16}$, respectively, where h is the height of the hexagon, as illustrated in Figure 4.1. The pathloss model in [8] is used,

$$\gamma_{\text{PL}}(\text{dB}) = 128.1 + 37.6\log_{10}R.$$

The channel is realized as follows, $\mathbf{H} = \Gamma\sqrt{G \cdot \gamma_{\text{PL}} \cdot \gamma_{\text{SF}}}$, where $\Gamma \sim \mathcal{CN}(0, 1)$ are i.i.d complex Gaussian fading coefficients, $\gamma_{\text{SF}} \sim \mathcal{N}(0, 8 \text{ dB})$ is the shadow fading component, and G is the transmit antenna gain of 9 dB. The system Signal to Noise Ratio (SNR) or the reference value of one UE located at the cell-edge, is fixed at 15 dB, giving rise to a maximum BS transmit power of 0.0603 W. It has been shown that it is difficult to estimate channels with pilot overhead for PJP with active set threshold greater than 15 dB at the cell-edge [9]. Hence, in this setup, a PJP threshold of 10 dB is considered.

The PSO with two different objective functions with case a) and b) can be compared with case e) based on [3]. The case c) is a typical ZF BF as in (Equation 4.2) obtained from limited feedback where backhaul reduction is achieved with explicit nulling of the BF coefficients, i.e., zeros or nulls are placed in the BF matrix where needed. Similarly, case d) achieves backhaul reduction with explicit nulling but this is a genie aided case where full feedback is allowed i.e., complete CSI is available at the CCN. CJP has the full feedback and full backhauling, without any reduction in feedback or backhaul.

Table 4.1.: Particle Swarm Parameters

Parameters	Values
Number of Particles, N	30
$x_{\max} = -x_{\min}$	$1/\max\{ \tilde{\mathbf{H}}(i,j) \}$
Max. velocity, v_{\max}	$(x_{\max} - x_{\min})/\Delta t$
Cognitive factor, c_1	2
Social factor, c_2	2
Inertia Weight, w	1.4 \rightarrow 0.4
Constant decay factor, β	0.99

The convergence of the PSO algorithm is shown in Figure 4.2 for various aggregated channel matrices at CCN. The PSO in case b) converges the fastest within 100 iterations. Once the BFs are obtained from various algorithms, a simple power allocation per BS as in (Equation 4.3) is performed, where there is at least one BS transmitting at maximum power. PSO being an iterative procedure, the per-BS power constraint can be applied in every iteration or after convergence. It should be

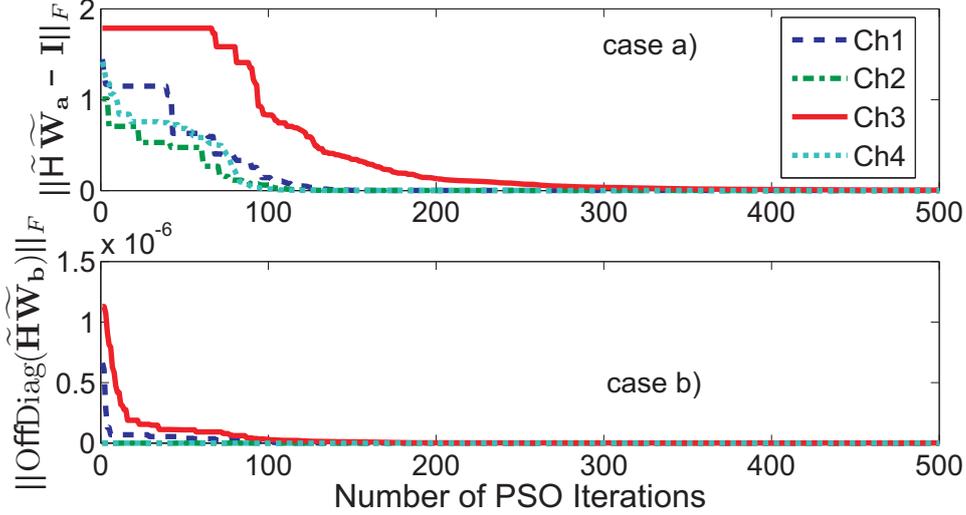


Figure 4.2.: Convergence behavior of PSO for 4 different channel realizations

noted that in case a), as defined in (Equation 4.6), the power constraint is applied after the PSO algorithm has converged and in case b), as defined in (Equation 4.7), the power constraint is applied after evaluating the objective function in every iteration. Other combinations were not considered, as the convergence was poor when the power constraint was applied in every iteration in case a). Applying the power constraint after convergence in case b) had more residual interference with slower convergence. Hence, the best combinations were considered. Figure 4.3 shows the Cumulative Distribution Function (CDF) of the BS transmit power. The values in the legend show the mean value of a given CDF. It can be observed that the PSO in case b) uses the BS power constraint, P_{\max} more effectively compared to others, as it is the bottom-most curve and it uses 2.85% relatively more power than case e), in average. All CDFs exhibit maximum power for 33.3% of the time, this is due to (Equation 4.3), as at least one of the 3 BSs is transmitting at maximum power.

The residual interference power in the system was calculated based on $\|\text{OffDiag}(\mathbf{H}\widetilde{\mathbf{W}}_x)\|_F^2$, where \mathbf{H} is the actual channel and x can be any one of the cases. It was observed that case a) trying to fairly serve all the users with equal power leaves 0.69% relatively more interference in the system when compared to case e). While, case b) reduces the interference by 3.94% relatively compared to case e). The CDF of the residual interference is not shown here, due to lack of space.

The CDF of the sum rate is shown in Figure 4.4. Case a) and b) perform the best compared to the other cases, with a PJP-10 dB threshold on CSI feedback. Case b) has a relative improvement in the average sum rate by 2.45% compared to case e), state of the art [3]. It should be noted that in this simulation setup, for fair comparison, only those cases are considered where the active set does not produce an overdetermined system, so that the state of the art solution in case e) works. This also means that PSO can be applied in all the scenarios without any

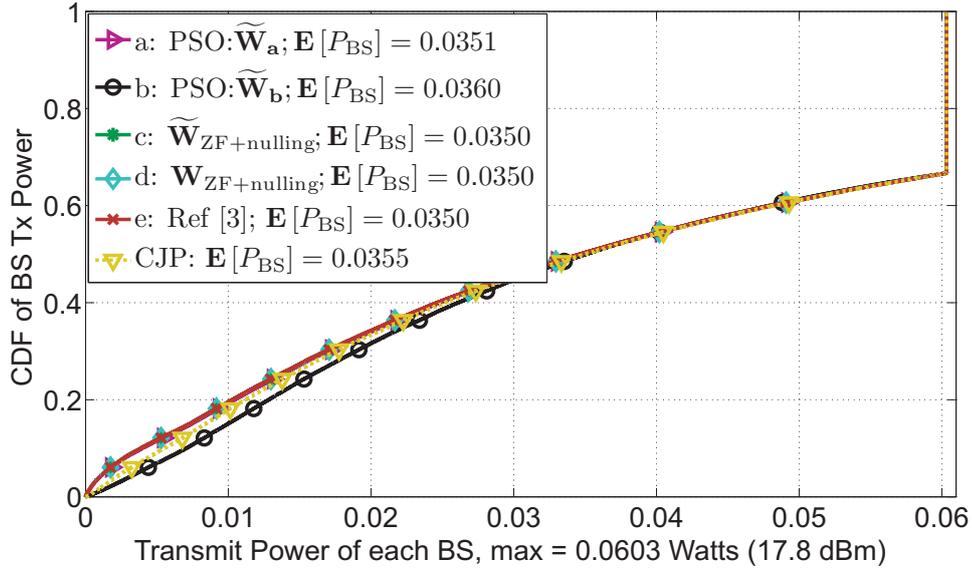


Figure 4.3.: CDF of the BS power (P_{BS}) transmitted with various precoding algorithms. The k th BS power is calculated as $\|\overline{\mathbf{W}}^k\|_F^2$.

restriction regarding scheduling or the need to have a well conditioned aggregated channel matrix. Hence, PSO not only achieves the reduction in backhaul, there is also a greater gain compared to the state of the art, as it always finds a solution. The case b) with PSO performs 2.98% relatively better than the genie aided full feedback case d) with reduced backhaul. The complexity analysis of the PSO algorithm with the constraint of backhaul load being equivalent to feedback load is treated as part of our future work.

4.5. Conclusion

Efficient backhauling techniques are needed to realize the gains of joint processing with reduced feedback. Existing techniques of achieving efficient backhauling in this context have constraints on scheduling or need full channel state information to be fed back. The particle swarm optimization used in this paper, is able to perform without any such constraints.

When the state of the art technique can find a solution, the average sum rate of the stochastic optimization algorithm performs 2.45% better than the state of the art solution, without any restriction on scheduling, such that the backhauling load is equivalent to that of the feedback load. The proposed algorithm is also capable of performing even when the state of the art solution fails due to an overdetermined system. The best algorithm proposed in this paper converges in merely 100 iterations, with effective usage of the base station power and lowering interference. With parallel computing and with more improved flavors of swarm algorithms, this complexity is feasible. Robustness of the proposed particle swarm algorithm with

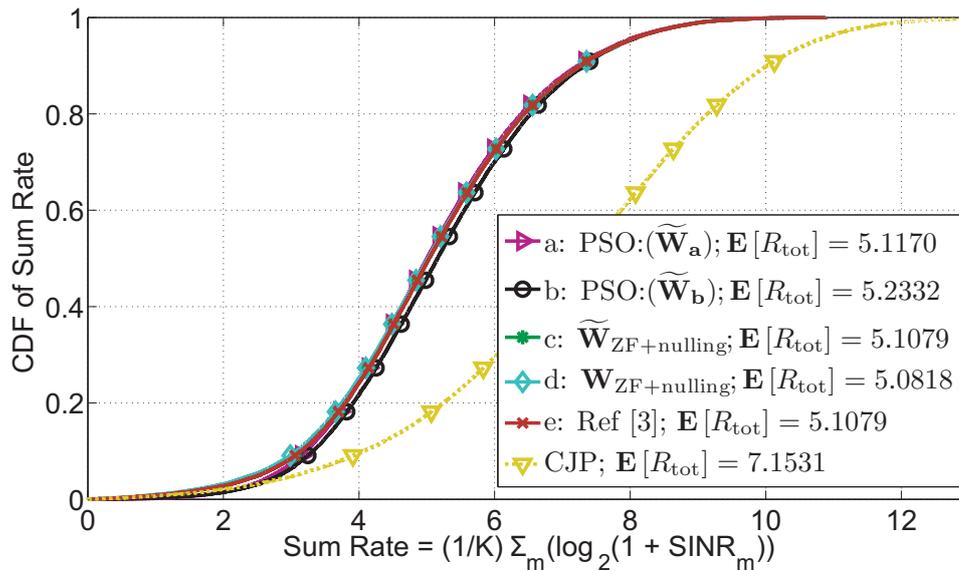


Figure 4.4.: Sum rate per cell, R_{tot} , of each algorithm and $\mathbf{E}[R_{\text{tot}}]$ is the mean of R_{tot} . CJP can be viewed as the upper bound for ZF.

optimized parameters and imperfect channel state information will be studied as part of our future work.

Bibliography

- [1] 3GPP TR 36.814-900, “3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Further Advancements for E-UTRA Physical Layer Aspects (Release 9),” Mar. 2010.
- [2] C. Botella, T. Svensson, X. Xu, H. Zhang, “On the performance of joint processing schemes over the cluster area,” in *Proc. IEEE Vehicular Tech. Conf.*, May 2010.
- [3] A. Papadogiannis, H.J. Bang, D. Gesbert, E. Hardouin, “Efficient selective feedback design for multicell cooperative networks,” *IEEE Trans. on Vehicular Tech.*, vol. 60, no. 1, pp. 196-205, Jan. 2011.
- [4] Fang Shu, Wu Gang, Li Shao-Qian, “Optimal multiuser MIMO linear precoding with LMMSE receiver,” *EURASIP Journal on Wireless Comm and Networking*, vol. 2009, Article ID 197682, Mar. 2009.
- [5] J. Kennedy, R.C. Eberhart, “Particle swarm optimization,” in *Proc. IEEE Int. Conf. on Neural Networks*, pp. 1942–1948, 1995.
- [6] AP Engelbrecht. *Fundamentals of computational swarm intelligence*, (John Wiley & Sons, 2005), pp. 171-172
- [7] Y.H. Shi, R.C. Eberhart, “Parameter Selection in Particle swarm optimization,” *The 7th Annual Conf. on Evolutionary Programming*, San Diego, USA, 1998.
- [8] 3GPP TR 36.942-a20, “3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios (Release 10),” Jan. 2011.
- [9] ARTIST4G D1.2, “Innovative advanced signal processing algorithms for interference avoidance”, *ARTIST4G technical deliverable*, 2010. [Online] Available: <https://ict-artist4g.eu/projet/deliverables>

Paper B

Partial Joint Processing with Efficient Backhauling using Particle Swarm Optimization

Tilak Rajesh Lakshmana, Carmen Botella and Tommy Svensson

Published in
EURASIP J. Wireless Commun. and Netw., vol. 2012, 2012

Notation error: In section 5.3, the scalar quantity Q should be q .
©2012 Springer

The layout has been revised.

5. Partial Joint Processing with Efficient Backhauling using Particle Swarm Optimization

Abstract

In cellular communication systems with frequency reuse factor of one, user terminals at the cell-edge are prone to intercell interference. Joint processing is one of the coordinated multipoint transmission techniques proposed to mitigate this interference. In the case of centralized joint processing, the channel state information fed back by the users need to be available at the central coordination node for precoding. The precoding weights (with the user data) need to be available at the corresponding base stations to serve the user terminals. These increase the backhaul traffic. In this paper, partial joint processing is considered as a general framework that allows reducing the amount of required feedback. However, it is difficult to achieve a corresponding reduction on the backhaul related to the precoding weights, when a linear zero forcing beamforming technique is used. In this work, particle swarm optimization is proposed as a tool to design the precoding weights under feedback and backhaul constraints related to partial joint processing. The precoder obtained with the objective of *weighted interference minimization* allows some multiuser interference in the system, and it is shown to improve the sum rate by 66% compared to a conventional zero forcing approach, for those users experiencing low signal to interference plus noise ratio.

Keywords—Coordinated MultiPoint, Joint Processing, Particle Swarm Optimization, Precoding, Stochastic Optimization

5.1. Introduction

Future cellular communication systems tend to be spectrally efficient with a frequency reuse factor of one. The aggressive reuse of frequency resources causes interference between cells, especially at the cell-edge. Therefore, the user experience is affected and the performance of such systems is interference limited. To overcome this problem, Coordinated MultiPoint (CoMP) transmission/reception is proposed [1]. Joint Processing (JP) is one of the techniques that falls into the framework of CoMP transmission. In the downlink, JP involves the coordination of Base Stations

(BSs) such that the interfering signals are treated as useful signals when transmitting to a User Terminal (UT). Note that this technique was previously referred to as network coordination [2].

For JP, UTs need to feed back the Channel State Information (CSI) of their BS-UT links. In Centralized Joint Processing (CJP), the CSI is collected at a node in the network called Central Coordination Node (CCN), to form an aggregated channel matrix [3, 4]. The CCN can be treated as a logical node that can be implemented at a BS. Based on this aggregated channel matrix, the CCN obtains the precoding weights, consisting of the beamforming weights after power allocation. These precoding weights need to be available along with the user data at the corresponding BSs to control interference via JP. In this work, the backhaul traffic mainly comprises of transporting the CSI coefficients from the cooperating BSs to the CCN, the precoding weights from the CCN to the cooperating BSs and the user data. We restrict the definition of the *backhaul load* as transporting the precoded weights from the CCN to the cooperating BSs. The *feedback load* is the traffic due to the CSI forwarding from UTs to the BSs. These definitions are illustrated in Figure 5.1. Along with the user data, this traffic poses tremendous requirements on the network backhaul [4, 5, 6]. It also imposes delay constraints due to non-stationary channels, but the delay constraints are beyond the scope of this work.

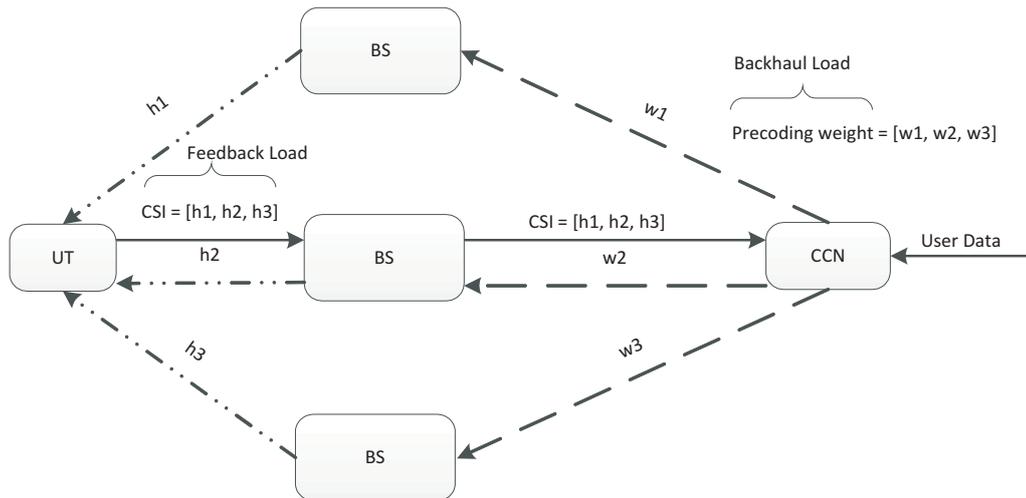


Figure 5.1.: This figure illustrates the feedback load comprising of the CSI coefficients from the UT to the BS. The backhaul load consists of the precoding weights from the CCN to the BS. The equivalence can be seen in the number of CSI coefficients, h_1, h_2, h_3 and the number of precoding weights w_1, w_2, w_3 for a given UT. The user data is assumed to be routed based on the non-zero precoding weights at the CCN.

One of the approaches to alleviate the complexity requirements in JP is to arrange the BSs in clusters [3]. The BSs involved in JP within a cluster control the intracluster interference, while the BSs belonging to neighboring clusters give rise to

intercluster interference. In a static clustering approach the cooperating set of BSs does not change with time, but this can create unfairness for UTs on the cluster edge. Hence, dynamic clustering helps in maintaining fairness among UTs. An example of dynamic clustering could be a family of clusters operating in round robin fashion where each cell takes its turn to be at the cluster boundary. Clustering techniques can also be divided into user-centric or network-centric depending on where the clustering decision is taking into account the UT determined channel conditions. Since CJP implies full cooperation, it requires extensive feedback and backhaul resources in the cooperative cluster. In order to bring JP close to realistic scenarios, one can further reduce the complexity for a given cluster through suboptimal approaches.

Several such approaches have been considered in the literature to reduce the requirements of CJP, such as limited feedback [2, 5] and limited backhauling [5, 6, 3, 10]. Partial Joint Processing (PJP) is a general framework aiming to reduce the complexity requirements of CJP, basically the feedback and backhaul load. In the particular PJP approach considered in this paper, a CCN or the serving BS instructs the UTs to report the CSI of the links in the cluster of BSs whose channel gain fall within an active set threshold or window, relative to their best link (usually the serving BS) [7]. This is summarized in Algorithm 5.1. Note that a similar approach is used in [8]. PJP can be regarded as a user-centric clustering when it is implemented over a static cluster, since overlapping *subclusters* or *active sets* of BSs are dynamically formed. Note that CJP is a particular case of PJP when the threshold tends to infinity.

Algorithm 5.1 Active set thresholding for limited feedback based on [7]

```
1: Choose: threshold = 10 dB
2: for each UT do
3:   Measure the channel gain from all BSs
4:   bestLink = max{channel strength from all BSs}
5:   if (bestLink - otherLink) ≤ threshold then
6:     UT feed backs the CSI of otherLink
7:     CCN marks this link as active
8:   else
9:     Feedback load reduction:
10:    UT does not feed back the otherLink
11:    CCN marks this link as inactive
12:   end if
13:   UT feeds back the bestLink
14:   CCN marks this link as active
15: end for
```

In PJP, the CSI of the links reported by the UTs to the CCN are marked as *active* links and those not reported are marked as *inactive*. Based on these, the CCN forms an aggregated channel matrix for interference control, where the coefficients of the inactive links are set to zero. In this paper, the CCN identifies the BSs that fall

outside the threshold window for a given UT based on the links for which the UT has not reported CSI. It is assumed that the obtained CSI is error free. Protocol aspects of this communication need to be addressed in more detail in a real system implementation. As a result, the aggregated channel matrix is now sparse. Linear techniques such as Zero Forcing (ZF) can invert the aggregated channel matrix to remove interference, but these techniques fail to invert a sparse aggregated channel matrix and at the same time reduce the backhaul load, such that only the BSs in the active set of a UT receive the precoding weights [9].

The question thus arises, in the PJP framework, can the gains achieved with CSI feedback load reduction translate to an *equivalent* backhaul load reduction, in the sense that the number of CSI coefficients constituting the feedback load (assuming a single tap channel for simplicity) is the same as that of the precoding weights in the backhaul (Figure 5.1 illustrates this notion). Particle Swarm Optimization (PSO) is proposed in this paper as a tool to obtain a solution that fits this requirement, since it can find the precoding weights without actually inverting the sparse aggregated channel matrix.

5.1.1. State of the art techniques

Precoding design for clustered scenarios under JP is a recent problem. In [11], a large network is divided into a number of disjoint clusters of BSs. Linear precoding is carried out within these clusters to suppress intracluster interference as well as intercluster interference. In the case of overlapping clusters, Soft Interference Nulling (SIN) precoding technique is proposed in [12]. For SIN, the complete CSI is available at all BSs and the user data is made available only to the BSs in the coordination cluster. Hence, the BSs can jointly encode the message for transmission. Moreover, in [12], multiple spatial streams are allowed up to the total number of transmit antennas in the coordination clusters. As the exhaustive search for the best clustering combination has a very high complexity, two simple clustering algorithms are proposed in [12]. They are: a) Nearest bases clustering and b) Nearest interferers clustering. The SIN iterative precoder optimization algorithm does not remove the interference completely, but performs better than or equal to any linear interference-free precoding scheme [12, Proposition 1]. SIN precoding relaxes the restriction to have zero interference, due to that SIN precoding works even when the number of transmit antennas is less than the total number of receive antennas within a coordination cluster. It should be noted that SIN achieves backhaul reduction in terms of the precoded weights and user data being available at BS where needed, but it does not provide feedback load reduction.

For JP, as long as the aggregated channel matrix at the CCN is well conditioned for inversion, linear ZF beamforming (BF) techniques can be used for interference control. It has been shown in [9] that when using techniques that achieve CSI feedback reduction, such as active set thresholding in PJP, this reduction does not translate to an equivalent backhaul load reduction with the linear ZF BF. When calculating the ZF BF based on the sparse aggregated channel matrix, a link that

has been defined as an inactive link may be mapped with a non-zero BF weight for that link. This causes unnecessary backhauling, since the UT has reported that link as inactive and that BS is already outside the active set of that UT. Instead, the BS could use this resource to serve another UT. An intuitive approach could be that the CCN resorts to nulling the BF weights where the links are expected to be inactive. This is a suboptimal solution. In [13], a partial ZF precoding design is proposed based on [14] to remove the interference in a PJP scenario. This solution performs better than the linear ZF BF with a weight nulling assumption, and works even for a sparse aggregated channel matrix at the CCN, but it does not achieve an equivalent backhaul load reduction. On the other hand, there is no linear technique in the literature that can invert the sparse aggregated channel matrix and preserve the zeros in the transposed version of the inverse, when the aggregated channel matrix is not diagonal or block-diagonal.

To the best of our knowledge, the problem of backhaul load reduction equivalent to feedback load reduction has only been addressed in [9], where two solutions are proposed. One based on scheduling (Medium Access Control - MAC layer approach) and a second one based on a ZF precoding PHY layer approach. The limitations of this approach are discussed in section 2.2.

5.1.2. Contributions

The active set thresholding technique in PJP (limited feedback of CSI) is used to achieve the feedback load reduction and these gains need to be preserved with an equivalent backhaul load reduction (limited backhauling of precoding weights). To achieve this, a stochastic optimization algorithm such as PSO is proposed for precoding design. PSO has been shown to obtain the optimal linear precoding vector, aimed to maximize the system capacity in a MultiUser-Multiple Input Multiple Output (MU-MIMO) system [15]. The main distinguishing factor of our paper compared to [15] is that the PSO is used for designing the precoder under a multicell setting with PJP. PSO has also been proposed as a tool for a scheduling strategy in a MU-MIMO system [16]. Recently, a multiobjective PSO has been proposed for accurate initialization of the channel estimates in a MIMO-OFDM iterative receiver [17]. Drawing inspiration from [4, 16, 17], and combining the state of the art PSO implementation with expendable parallel computing power at the CCN, a PSO based precoder should be feasible for the scenario under consideration.

In this paper, two objectives are studied using PSO. They are:

1. *Weighted interference minimization*: Minimize the interference for the UTs and improve the UT experiencing the minimum Signal to Interference plus Noise Ratio (SINR).
2. Sum rate maximization.

In addition, to fairly compare the linear ZF-based precoder and the proposed PSO-based precoder, the use of perturbation theory and Gershgorin's discs is introduced.

These discs can be used to obtain a quick graphical snapshot of the intracluster interference remaining in the system. The sum rate bounds under a constrained backhaul and imperfect channel knowledge are important [18] and it is part of our future work.

The paper is organized as follows. The system model and the limitations in the state of the art linear solutions are discussed in section 5.2. The PSO as a tool for precoder design is presented in section 5.3. In this section, the objective function, the termination criteria, the convergence of PSO and the complexity in terms of the big \mathcal{O} notation are analyzed. An interesting connection is made between the signal to interference ratio (SIR) and Gershgorin's discs in section 5.4. The simulation results are presented in section 5.5 and the conclusions are drawn in section 5.6.

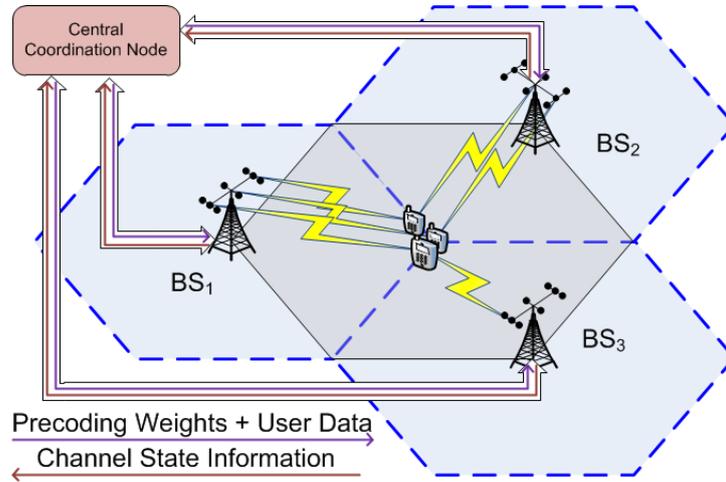


Figure 5.2.: The cluster layout, the hexagon in the middle denotes the cluster area under consideration where the UTs are located at the cluster center.

Notation: The boldface upper-case letters, boldface lower-case letters and italics such as \mathbf{X} , \mathbf{x} and x denote matrices, vectors and scalars, respectively. The $\mathbb{C}^{m \times n}$ is a complex valued matrix of size $m \times n$. The $(\cdot)^H$ is the conjugate transpose of a matrix. $\|\cdot\|_F$ is the Frobenius Norm, $\text{diag}(\mathbf{A})$ and $\text{OffDiag}(\mathbf{A})$ are the diagonal and off-diagonal elements of the matrix \mathbf{A} . Block diagonalizing the matrices \mathbf{A} and \mathbf{B} is denoted as $\text{blockdiag}(\mathbf{A}, \mathbf{B})$. The i th row and the j th column of a matrix \mathbf{A} is represented as $\mathbf{A}(i, j)$. To access all the elements of the i th row of a matrix \mathbf{A} is $\mathbf{A}(i, :)$ and for the j th column is $\mathbf{A}(:, j)$. $\text{vec}(\mathbf{A})$ is the vector of stacked columns of matrix \mathbf{A} . $\Re\{\mathbf{A}(i, j)\}$ and $\Im\{\mathbf{A}(i, j)\}$ are the real part and the imaginary parts of $\mathbf{A}(i, j)$, respectively. \mathbf{H} and $\tilde{\mathbf{H}}$ denote the aggregated channel matrix at the CCN due to full CSI feedback and the sparse aggregated channel matrix at the CCN due to limited CSI feedback, respectively. \mathbf{W} and $\tilde{\mathbf{W}}$ denote the BF matrix and sparse BF matrix, respectively. The BF matrix with power allocation forms the precoding matrix, $\bar{\mathbf{W}}$.

5.2. System model

Consider the downlink of a static cluster of K BSs with N_T antennas serving M single antenna UTs [7]. In this model, the intracluster interference caused due to the transmission to the UTs located at the cluster center is considered, as shown in Figure 5.2. For simplicity, the intercluster interference is assumed to be negligible. Assuming CJP between BSs in the cluster, the discrete time signal received at M UTs, $\mathbf{y} \in \mathbb{C}^{M \times 1}$ is

$$\mathbf{y} = \mathbf{H}\bar{\mathbf{W}}\mathbf{x} + \mathbf{n}. \quad (5.1)$$

The aggregated channel matrix available in the CCN is $\mathbf{H} \in \mathbb{C}^{M \times KN_T}$, and it is of the form $\mathbf{H} = [\mathbf{h}_1^T \mathbf{h}_2^T \dots \mathbf{h}_M^T]^T$, where $\mathbf{h}_m \in \mathbb{C}^{1 \times KN_T}$ is the channel from all the BSs to the m th UT in the cluster. The precoding matrix $\bar{\mathbf{W}}$ is obtained from the aggregated BF matrix $\mathbf{W} \in \mathbb{C}^{KN_T \times M}$ after power allocation. The BF matrix is of the form $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_M]$, $\mathbf{w}_m \in \mathbb{C}^{KN_T \times 1}$ is the BF for the m th UT. The transmitted symbols to the M UTs are $\mathbf{x} \in \mathbb{C}^{M \times 1}$. The receiver noise \mathbf{n} at the UTs is spatially and temporally white with variance σ^2 , and it is uncorrelated with the transmitted symbols.

In the case of ZF BF, the constraint $K \cdot N_T \geq M$ needs to be satisfied to maintain orthogonality between UTs [19]. The matrix \mathbf{W} is then obtained by taking the Moore-Penrose pseudoinverse of \mathbf{H} as

$$\mathbf{W} = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1}. \quad (5.2)$$

Each BS is constrained to a maximum transmit power, P_{\max} . The suboptimal power allocation based on [20] is performed for ZF under per-BS power constraints, where at least one of the BSs is transmitting at maximum power, and it is defined as

$$\bar{\mathbf{W}} = \sqrt{\frac{P_{\max}}{\left(\max_{k=1, \dots, K} \|\mathbf{W}(k_{N_T}, :)\|_F^2\right)}} \cdot \mathbf{W}, \quad (5.3)$$

where k_{N_T} selects the rows of the BF matrix \mathbf{W} of the k th BS with its N_T antennas towards the M UTs. The SINR at the m th UT is given as

$$\text{SINR}_m = \frac{\|\mathbf{h}_m \bar{\mathbf{w}}_m\|^2}{\sum_{\substack{j=1 \\ j \neq m}}^M \|\mathbf{h}_m \bar{\mathbf{w}}_j\|^2 + \sigma^2}, \quad (5.4)$$

and the sum rate per cell in terms of bits per second per Hertz per cell (bps/Hz/cell) is given as

$$R_{\text{tot}} = \frac{1}{K} \sum_{m=1}^M \log_2 (1 + \text{SINR}_m). \quad (5.5)$$

5.2.1. Linear beamforming

As stated in section 5.1, the link for which the CSI is reported to the CCN is marked as an active link and the unreported CSI is marked as an inactive link. These active and inactive links can be represented with a binary matrix of size $M \times K$. The (m, k) th element in this matrix corresponds to the link between the m th UT and the k th BS. An active link is represented with ‘1’ and an inactive link is represented with ‘0’.

In (Equation 5.2), the linear ZF BF completely removes the interference by inverting the aggregated channel matrix \mathbf{H} . With small active set thresholds, there are few active links, forming a sparse aggregated channel matrix $\widetilde{\mathbf{H}}$ at the CCN. If the sparse aggregated channel matrix $\widetilde{\mathbf{H}}$ is invertible, then the BF matrix $\widetilde{\mathbf{W}}$ thus formed may not have zeros at places where needed. If each BS were to have N_T antennas each, then the pseudoinverse could generate BF weights for some of the N_T antennas and not for the BS as a whole. Moreover, a UT might receive its data from a BS outside of the active set of a given UT. The effects of ZF are highly undesirable as it results in extra and unnecessary backhaul load on the cluster, as well as unnecessary transmissions on these links. The ZF solution over a sparse aggregated channel matrix without any scheduling constraint cannot achieve an equivalent reduction in backhaul load.

In this paper, the following ZF scenarios are considered, where the ZF is performed using the pseudoinverse as in (Equation 5.2) on the aggregated channel matrix at the CCN. The main focus is on the ZF with Limited FeedBack (LFB) and Limited Backhauling (LBH), where the gains of feedback load reduction need to be preserved in the backhaul load reduction. This is denoted as ZF:LFB+LBH. The LFB is achieved based on the active set thresholding technique. The LBH with ZF is achieved with an intuitive approach of nulling of the BF coefficients based on the inactive links in the binary matrix. When the UT is allowed to feed back all the CSI (Full FeedBack, FFB) and allowing Full Backhauling (FBH), it is represented as ZF:FFB+FBH. This scenario is considered to show the upper bound of the ZF technique, as in the case of CJP. The scenario ZF with FFB and LBH is considered to have a similar configuration as that of the SIN precoding technique [12]. This is denoted as ZF:FFB+LBH. Finally, the scenario ZF with LFB and FBH is considered, similar to that considered in [9], where the ZF is allowed to have the precoded weights at BSs where it is not desired and allowing FBH. This is represented as ZF:LFB+FBH. It should be noted that this approach does achieve some backhaul reduction, but not necessarily equivalent to the feedback load.

5.2.2. Limitations of the state of the art

The following subsections capture the limitations with the state of the art solutions.

The invertibility of the aggregated channel matrix:

To maintain the orthogonality between the UTs, as highlighted earlier, the condition $K \cdot N_T \geq M$ needs to be satisfied. Due to this, the number of columns of the matrix \mathbf{H} is always greater or equal to the number of rows, and the only way to invert the aggregated channel matrix is by using the right inverse as shown in (Equation 5.2). The invertibility of the linear ZF BF is limited by the ability to invert $(\mathbf{H}\mathbf{H}^H)^{-1}$ or in other words, the rank of $\mathbf{H}\mathbf{H}^H$ should be equal to the number of UTs, whose channels are linearly independent.

In the PJP framework, the active set threshold can be increased such that the UTs can feed back the CSI of any additional BSs that fall within this window, thereby increasing the chances of inverting the aggregated channel matrix as proposed in [13]. The worst case could be that the UTs would need to feed back the complete CSI from all the BSs like in the case of CJP. The CCN can now invert the aggregated channel matrix to obtain the BF weights, but at the expense of increasing the feedback load.

Required nulls in beamformer:

As stated before, to the best of our knowledge, to overcome the invertibility of the aggregated channel matrix and the required nulls in the BF, the MAC layer and the PHY layer approaches are proposed in [9]. These approaches are analyzed for the remaining part of this section.

In the scheduling MAC layer approach, BS subgroups are formed such that the transmission to the UTs in each time slot is disjoint, where each BS is transmitting in only one subgroup. These disjoint sets give rise to a sparse aggregated channel matrix at the CCN, which presents a block-diagonal form. Note that the scheduling approach can be mapped to a disjoint clustering solution. This approach solves the problem of equivalent backhaul load reduction, as the inverse of a block diagonal matrix is block diagonal itself, thereby retaining the zeros or nulls in the BF weights where needed. In a given time slot, if the collocated UTs prefer services from the same set of BSs, then the MAC layer approach can only serve the UTs in a time division multiplexing fashion, as disjoint BS sets need to be selected for transmission. To guarantee fairness, such UTs will have to wait for a long time to be served.

The proposed PHY layer ZF precoding solution does not require any specific constraints on scheduling [9], and it allows the formation of overlapping clusters. The interference is reduced by formulating a constrained least squares optimization problem, whose solution is showed to be a pseudoinverse [9]. The closed-form solution to find the non-zero BF weights as obtained in [9, eq. (29)] is

$$\mathbf{w}_{el} = \tilde{\mathbf{H}}_{el}^H \left(\tilde{\mathbf{H}}_{el} \tilde{\mathbf{H}}_{el}^H \right)^{-1} \text{vec}(\mathbf{I}_K), \quad (5.6)$$

where $\tilde{\mathbf{H}}_{el}$ is obtained after processing the sparse aggregated channel matrix $\tilde{\mathbf{H}}$ in the CCN, after eliminating the columns corresponding to the zeros from $\text{vec}(\tilde{\mathbf{W}})$. These zeros correspond to the nulls expected in the BF, \mathbf{I}_K is the identity matrix

of size $K \times K$, where $K = 3$. \mathbf{w}_{el} contains the vectorized non-zero BF weights that need to be remapped to form the final BF matrix $\widetilde{\mathbf{W}}$.

To illustrate the limitations in the PHY layer approach in [9], consider single antenna BSs serving single antenna UTs with the aggregated channel matrix at the CCN as shown in Table 5.1. The first step is to build a block diagonal matrix as $\widetilde{\mathbf{H}}_{\mathbf{d}} = \text{blockdiag}(\widetilde{\mathbf{H}}, \widetilde{\mathbf{H}}, \widetilde{\mathbf{H}})$, and then to eliminate the columns of $\widetilde{\mathbf{H}}_{\mathbf{d}}$ corresponding to the predetermined zeros in the vectorized BF matrix, $\text{vec}(\widetilde{\mathbf{W}})$. In this example, columns 3, 4, 7 and 8 should be eliminated from the matrix $\widetilde{\mathbf{H}}_{\mathbf{d}}$ to obtain the matrix $\widetilde{\mathbf{H}}_{el}$ of size 9×5 . Due to this, the rows 3 and 7 in $\widetilde{\mathbf{H}}_{el}$ become zeros and the (Equation 5.6) is badly conditioned for right inverse. Hence, the PHY layer algorithm *should be modified to eliminate all rows that contain only zeros in $\widetilde{\mathbf{H}}_{el}$ before evaluating the right inverse*. Proceeding with this modification for the solution in (Equation 5.6), the matrix $\widetilde{\mathbf{H}}_{el}$ is now of size 7×5 , but it still has a problem of having more rows (equations) than columns (variables). There is no solution to this overdetermined system and the right inverse as shown in the closed-form in (Equation 5.6) is not feasible. There could only be solutions when the rows are linearly dependent, i.e., two or more UTs see the same channel, which is a highly unlikely scenario. More examples can be found, where the closed-form solution as per (Equation 5.6) breaks down.

Table 5.1.: An example of a sparse aggregated channel matrix giving rise to an overdetermined system when PHY layer precoding is applied.

$\widetilde{\mathbf{H}}$	BS_1	BS_2	BS_3
UT_1	h_{11}	h_{12}	0
UT_2	0	h_{22}	h_{23}
UT_3	0	0	h_{33}

The PHY layer solution does not comment on the fact that there is no PHY layer solution without scheduling the UTs, as the invertible part of the pseudoinverse in $(\widetilde{\mathbf{H}}_{el} \widetilde{\mathbf{H}}_{el}^H)^{-1}$ may not be feasible. In short, the PHY layer solution needs some scheduling constraints to obtain the BF weights.

Due to the limitations in this closed-form solution in (Equation 5.6), a proper comparison of the proposed PSO with this PHY layer solution is not possible. Hence, the PHY layer solution of [9] is not considered in the simulations. However, an interested reader can refer to [21] where the comparison is performed when [9] is feasible. In the subsequent section, PSO is presented as a tool for precoder design for backhaul load reduction equivalent to the feedback load reduction in the PJP framework.

5.3. Particle swarm optimization for precoding in the PJP framework

The PSO was inspired from the movement of a swarm, such as a shoal of fish, a flock of birds, etc, to find food or to escape from enemies, by splitting up into groups. There is no apparent leader of the swarm other than the social interactions between the bird like objects (or *boids*). The coherent movement of these boids is modeled based on their social interactions with their neighbors. The algorithm simulating these social aspects was simplified in [22] and it was found to perform optimization. In this paper, a basic PSO algorithm [23] with inertia weight and velocity restriction is implemented and it is capable of finding a stable solution based on a given objective function.

Classical optimization methods are especially preferred when the optimization problem is known to be convex but this is not the case here. Numerical methods such as Newton's method are not feasible as the objective function is non-differentiable. Other classical techniques could fail but PSO would always find an equilibrium/stable solution. PSO was chosen over other evolutionary algorithms, as it requires very few parameters to configure, it is easier to understand with computationally lesser bookkeeping and it fits well for reducing the backhaul load. In [23], PSO is viewed as a paradigm within the field of swarm intelligence and the performance measures of basic PSO are highlighted. This reference also provides detailed differences between PSO and other evolutionary algorithms.

In this paper, each bird in a swarm carries the real and imaginary parts of the non-zero elements of the BF matrix, i.e., the i th member of the swarm is the i th *particle* that carries all the ($n = 2 \cdot K \cdot N_T \cdot M$) BF coefficients. The '2' is due to PSO treating the real and the imaginary part of the complex BF coefficients as another dimension to the search space. Hence, the particle having the best n values needs to be found for a given objective function. For example, an infinite threshold would yield $n = 2 \cdot K \cdot N_T \cdot M$ non-zero CSI coefficients in the aggregated channel matrix of size $[M \times K \cdot N_T]$. With an active set threshold of 0 dB then only the best link (or reference link) would be fed back by each UT yielding $n = 2 \cdot 1 \cdot N_T \cdot M$. The real and the imaginary parts of the non-zero BF matrix, $\widetilde{\mathbf{W}}$, are mapped to a particle. This mapping, during initialization, is only for illustrating how the BF is translated to a particle. These steps can be omitted in the actual implementation. The position, $\mathbf{X}(i, j)$, and the velocity, $\mathbf{V}(i, j)$, of the i th particle with the j th BF coefficient are stochastically initialized as $\mathbf{X}(i, j) = x_{\min} + r \cdot (x_{\max} - x_{\min})$ and $\mathbf{V}(i, j) = \frac{1}{\Delta t} \left(-\frac{(x_{\max} - x_{\min})}{2} + s \cdot (x_{\max} - x_{\min}) \right)$, respectively. Here r and s are random numbers picked from a uniform distribution in the interval $[0, 1]$, and x_{\max} is the maximum value that a BF coefficient is initialized with. This does not mean that the position of the particle will not exceed this value, i.e., the particles in the PSO can actually go beyond these limits. The same holds for the velocity of the particle, but it is restricted by a maximum velocity, v_{\max} , so that the particle does not diverge. Δt is the time step length. The total number of particles is Q . Recall

that each particle is indexed using the variable i , where each particle is carrying n BF coefficients. These coefficients are indexed using the variable j .

A given objective function is evaluated for every particle i carrying the BF coefficients, and it is demapped to form the BF matrix as $\widetilde{\mathbf{W}}(l, m) \leftarrow \{\mathbf{X}(i, j)\} + i \cdot \{\mathbf{X}(i, j + 1)\}$, $l \in \{1, \dots, KN_T\}$, $m \in \{1, \dots, M\}$. The i th particle keeps a record of its best BF as $\mathbf{X}^{pb}(i, :)$, and the best BF achieved by any of the particles in the swarm is stored as \mathbf{x}^{sb} . The equations governing the update of the velocity and the position of a particle are:

$$\mathbf{V}(i, j) \leftarrow w \cdot \mathbf{V}(i, j) + c_1 \cdot p \cdot \left(\frac{\mathbf{X}^{pb}(i, j) - \mathbf{X}(i, j)}{\Delta t} \right) + c_2 \cdot q \cdot \frac{\mathbf{x}^{sb}(j) - \mathbf{X}(i, j)}{\Delta t}, \quad (5.7)$$

$$\mathbf{X}(i, j) \leftarrow \mathbf{X}(i, j) + \mathbf{V}(i, j) \cdot \Delta t. \quad (5.8)$$

The variables p and q are random numbers drawn from a uniform distribution in the interval $[0, 1]$. The terms involving c_1 and c_2 are called the *cognitive* component and the *social* component, respectively. The cognitive component tells how much a given particle should rely on itself or believe in its previous memory, while the social component tells how much a given particle should rely on its neighbors. The cognitive and social constant factors, c_1 and c_2 , are equal to 2, as highlighted in [22]. An *inertia weight*, w , is used to bias the current velocity based on its previous value, such that when the inertia weight is initially being greater than 1 the particles are biased to explore the search space. When the inertia weight decays to a value less than 1, the cognitive and social components are given more attention [24]. The decaying of the inertia weight is governed by a constant decay factor β , such that $w \leftarrow w \cdot \beta$.

The pseudocode of PSO described above is summarized in Algorithm 5.2.

5.3.1. Objective function

The particle with the best BF coefficients is demapped to obtain the BF matrix, $\widetilde{\mathbf{W}}$. The maximum transmit power at each BS is constrained to P_{\max} and power allocation based on [20] is applied as per (Equation 5.3). This is referred to as *power adjustment* on the BF matrix, forming a precoding matrix, $\overline{\mathbf{W}}$. There are two ways in which this can be applied, either in every iteration of the PSO (in short, *PwrAdj*) or after obtaining the best particle from the PSO (in short, *NoPwrAdj*). Making sure that at least one BS is transmitting at maximum power in every iteration consumes more computational resources, but on the contrary, if this is done after running the PSO algorithm, then this normalization skews or disfigures the best precoding weights. Both cases of power normalization are considered in the objective functions below. It should be noted that for the *NoPwrAdj* case, the objective function is evaluated without any restriction on the BS transmit power. This means that it is possible to exceed the BS power constraint when evaluating the objective function. Nevertheless, the final precoding weights after applying (Equation 5.3) satisfy the

BS power constraint. The flexibility of choosing an objective function gives another degree of freedom for the PSO-based precoder.

Algorithm 5.2 Pseudocode for obtaining the BF via PSO. Steps 3 to 5 are only mentioned for illustration and can be avoided prior to initialization.

```

1: Initialization:
2: Determine the number of non-zero coefficients  $n$  needed in the BF matrix,  $\widetilde{\mathbf{W}}$ 
3: Map the BF to the particle:
4:  $\mathbf{X}(i, j) \leftarrow \Re\{\widetilde{\mathbf{W}}(l, m)\}, l \in \{1, \dots, KN_T\}, m \in \{1, \dots, M\}$ 
5:  $\mathbf{X}(i, j+1) \leftarrow \Im\{\widetilde{\mathbf{W}}(l, m)\}$ 
6: Stochastically initialize particles with BF coefficients:
7:  $x_{\max} = 1/\max|\widetilde{\mathbf{H}}(i, j)|$ 
8:  $x_{\min} = -x_{\max}$ 
9: Position:  $\mathbf{X}(i, j) = x_{\min} + r \cdot (x_{\max} - x_{\min})$ 
10: Velocity:  $\mathbf{V}(i, j) = \frac{1}{\Delta t} \left( -\frac{(x_{\max} - x_{\min})}{2} + s \cdot (x_{\max} - x_{\min}) \right)$ 
11: while Termination Criterion do
12:   for the  $i$ th particle in the swarm do
13:     Demap the variables in a particle to form the BF matrix
14:      $\widetilde{\mathbf{W}}(l, m) \leftarrow \{\mathbf{X}(i, j)\} + i \cdot \{\mathbf{X}(i, j+1)\}$ 
15:     Evaluate the objective function  $f(\mathbf{X}(i, :))$ 
16:     Store:
17:     if  $f(\mathbf{X}(i, :)) < f^{pb}(\mathbf{X}(i, :))$  then
18:       Particles' Best:  $\mathbf{X}^{pb}(i, :) \leftarrow \mathbf{X}(i, :)$ 
19:     end if
20:     if  $f(\mathbf{X}(i, :)) < f^{sb}(\mathbf{X}(i, :))$  then
21:       Swarm's Best:  $\mathbf{x}^{sb} \leftarrow \mathbf{X}(i, :)$ 
22:        $\widetilde{\mathbf{W}}^{sb}(l, m) \leftarrow \{\mathbf{x}^{sb}(j)\} + i \cdot \{\mathbf{x}^{sb}(j+1)\}$ 
23:     end if
24:   end for
25:   for Each particle in the swarm with BF coefficients do
26:     Update:
27:     Velocity:  $\mathbf{V}(i, j) \leftarrow w \cdot \mathbf{V}(i, j) + c_1 \cdot p \cdot \left( \frac{\mathbf{X}^{pb}(i, j) - \mathbf{X}(i, j)}{\Delta t} \right) + c_2 \cdot q \cdot \frac{\mathbf{x}^{sb}(j) - \mathbf{X}(i, j)}{\Delta t}$ 
28:     Restrict velocity:  $|\mathbf{V}(i, j)| < v_{\max}$ 
29:     Position:  $\mathbf{X}(i, j) \leftarrow \mathbf{X}(i, j) + \mathbf{V}(i, j) \cdot \Delta t$ 
30:   end for
31:    $w \leftarrow w \cdot \beta$ 
32: end while
33: return BF Weight Matrix,  $\widetilde{\mathbf{W}}^{sb}$ 

```

In this paper, two different objective functions are considered for the PSO to optimize.

Weighted interference minimization:

Based on our experience, choosing a single direct objective function of minimizing only the interference skews the PSO algorithm to prefer only the good SINR UTs

and to leave out the weak SINR UTs. This gives rise to power savings at the BS, thereby lowering the sum rate of the UTs. One can choose to maximize the weak SINR UTs but then the total interference is not taken into account. Hence, the objective function should not only minimize the interference but also improve the SINR of the weakest UT (*minSINRuser*). We call this objective function *weighted interference minimization*, where the interference is minimized with the weight of the SINR of the weakest UT in each iteration. Note that the weakest SINR UT can change in every iteration. Thus, a multiobjective function evaluated for the i th particle in every iteration is defined as

$$f(\mathbf{X}(i, :)) := \frac{\|\text{OffDiag}(\mathbf{HW})\|_F}{\text{minSINRuser}}. \quad (5.9)$$

The goal of every particle is to minimize this multiobjective function iteratively. Finally, the swarm's best particle will contain the best BF that has managed to minimize (Equation 5.9).

Sum rate maximization:

The PSO presented in Algorithm 5.2 involves minimization of the objective function. Hence, to maximize the sum rate, the objective function is written as $f(\mathbf{X}(i, :)) := -R_{\text{tot}}$. This means that prior to evaluating the objective function, the sum rate per cell as in (Equation 5.5) needs to be calculated for every iteration.

5.3.2. Termination criteria

In [23], various stopping conditions are discussed. A few of them are listed here for completeness. The algorithms can be terminated when at least one of the following conditions is triggered:

1. Maximum number of iterations has been exceeded.
2. A solution fulfilling a target value is found.
3. No improvement is observed over a number of iterations.
4. Normalized swarm radius is close to zero.

In practice, any one of the above mentioned criteria can be used for termination. In this paper, the third criterion is used for termination.

5.3.3. Convergence

With a basic PSO, the notion of convergence means that the swarm has moved towards an equilibrium state [23]. The *lemma 14.2* in [23] shows that the basic PSO does not satisfy the convergence condition for global search. In our paper, a basic PSO with basic variations such as velocity restriction and inertia weight has

been used. Proving the optimality conditions of the PSO is not easy, but what can be said is that a stable solution can be achieved. Hence, suitable variations of the PSO algorithm need to be considered in future work, such as *Random Particle PSO* or *Multistart PSO*, since they satisfy the convergence condition for global search and can be considered for global optimization [23].

5.3.4. Computation complexity analysis

The Big \mathcal{O} notation is used to determine the complexity of implementing PSO as a function of the number of UTs M , based on the pseudocode presented in Algorithm 5.2. The computational complexity of PSO is

$$\mathcal{O}(\text{Block1} + c(\text{Block2} + \text{Block3})), \quad (5.10)$$

where c refers to the number of iterations in the while loop, which depends on the convergence of the algorithm. In this paper, it was observed that the algorithm converges within 100 iterations with no further improvements for the case of LFB and LBH.

- *Block1*: The initialization of particles carrying the BF coefficients from step 6 to 10, has a computational complexity of $\mathcal{O}(Qn)$, where Q is the number of particles which is a constant throughout the simulation and n is the number of BF coefficients.
- *Block2*: From steps 12 to 24, the complexity is $\mathcal{O}(Q \cdot \text{complexity of the objective function})$. Demapping from the i th particle to the BF matrix consumes $\mathcal{O}(n)$, which is independent of the objective function. But now we shall represent the dimension of the BF matrix \mathbf{W} in terms of n and M as $\left[\frac{n}{2M} \times M\right]$.

Objective function: Weighted interference minimization

- The complexity of \mathbf{HW} is $\mathcal{O}(Mn)$, the Frobenius norm constitutes $\mathcal{O}(M^2)$ and the SINR of the m th user is $\mathcal{O}(2\frac{n}{2M}M)$. To find the minimum SINR user, the SINR for all the M users is calculated as $\mathcal{O}(Mn)$. Therefore, the complexity of weighted interference minimization is $\mathcal{O}(Mn) + \mathcal{O}(M^2) + \mathcal{O}(Mn)$ and can be simplified to $\mathcal{O}(Mn)$.

Objective function: Sum rate maximization

- The calculation of SINR and consequently the sum rate per cell yields $\mathcal{O}(Mn)$.

Therefore, considering the worst case objective function, the complexity of Block2 is $\mathcal{O}(QMn)$.

- *Block3*: From steps 25 to 30, the time and space complexity can only grow with the number of BF coefficients. Hence, the computational complexity is $\mathcal{O}(Qn)$.

Finally, the overall complexity of the PSO is $\mathcal{O}(Qn + c(QMn + Qn))$ and can be simplified to $\mathcal{O}(cMn)$, ignoring the constants and lower order terms.

In this paper, we consider M single antenna UTs and K BSs with N_T antennas each. As shown in Algorithm 5.2, the number of BF coefficients carried by a particle is $n = 2 \cdot M \cdot K \cdot N_T$. Therefore, the complexity of the PSO is $\mathcal{O}(cM^2KN_T)$. Assuming that orthogonality is maintained in the system such that the number of UTs is $M = K \cdot N_T$, we have $\mathcal{O}(cM^3)$. The complexity of ZF BF is merely that of the pseudoinverse which is of the order $\mathcal{O}(M^2KN_T)$ and can be simplified to $\mathcal{O}(M^3)$ under orthogonality constraint. Comparison between PSO and ZF in terms of execution time may not be fair as only a basic PSO with basic variations is being implemented in MATLAB and ZF is bound to perform better. But, it should be noted that the PSO always provides an equilibrium solution while the ZF might not. Hence, it is difficult to perform a completely fair comparison.

5.4. Analysis of interference using Gershgorin's discs

In the case of ZF BF, the feasibility of the solution is determined by the ability to invert $(\mathbf{H}\mathbf{H}^H)^{-1}$. In [25] and [26], it is shown that any approach to improve the channel inversion must aim to reduce the effects of the largest eigen value. Another metric that has been used in the framework of ZF in a single-cell setup is the Frobenius norm of the channel \mathbf{H} , since it is proportionally related to the link level performance as shown in [25]. Their proposed network coordinated BF algorithm combines both metrics such that the mean of the largest eigen value of $(\mathbf{H}\mathbf{H}^H)^{-1}$ should be small and the mean of the Frobenius norm of \mathbf{H} should be large, so that SINR of the UT is large and the bit error rate is improved, respectively.

In the case of PSO, analyzing the properties of the obtained precoder via $(\mathbf{H}\mathbf{H}^H)^{-1}$ is not meaningful. To evaluate the performance of a PSO-based precoder, $\mathbf{H}\overline{\mathbf{W}}$ is analyzed here. But, $\|\mathbf{H}\overline{\mathbf{W}}\|_F$ does not give an insight into the properties of the precoder, as the off-diagonal elements are the residual intracluster interference remaining in the system. Interference is completely removed when the off-diagonal elements of $\mathbf{H}\overline{\mathbf{W}}$ are zeros. However, note that complete removal of interference is not maximizing the sum rate and therefore suboptimal in that sense.

In the framework of perturbation theory [27], these off-diagonal elements can be seen as a perturbation over the diagonal elements of $\mathbf{H}\overline{\mathbf{W}}$. In this context, Gershgorin circle theorem [27] can be used to analyze the behavior of different precoding techniques. Gershgorin circle theorem says that for a given square matrix \mathbf{A} , the elements in the main diagonal give an estimate of the eigen values on the complex plane. For a given element in the diagonal, the sum of the absolute values of the corresponding row is the length of the radius of the Gershgorin disc around this estimated eigen value. The circumference of this disc is called the Gershgorin's circle. The Gershgorin's circle theorem tells that all the eigen values of the matrix \mathbf{A} lie within the union of these discs. This theorem was mainly used to describe

how well the elements in the diagonal of a matrix approximate their eigen values. Hence, Gershgorin's discs can be used here to fairly visualize how the intracluster interference is removed with the PSO-based precoder and the linear ZF BF, as shown in subsection 5.5.3.

Applying Gershgorin's circle theorem, the matrix $\mathbf{A} = \mathbf{H}\bar{\mathbf{W}}$ can be perturbed as $\mathbf{H}\bar{\mathbf{W}} = \mathbf{D} + \mathbf{F}$, where $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_M)$ and \mathbf{F} has zero as its diagonal entries while the off-diagonal elements are the perturbation. The elements in the diagonal of \mathbf{D} form the useful signal strength for the UTs, while the off-diagonal elements of the matrix \mathbf{F} are the multiuser interference in the system. The i th Gershgorin disc, D_i , is computed based on

$$D_i = \left\{ z \in \mathbb{C} : |z - d_i| \leq \sum_{j=1}^M |\mathbf{F}(i, j)| \right\}, \quad (5.11)$$

where the right hand side of the inequality is the radius of the i th disc.

Table 5.2.: Simulation parameters

System Parameters	Values
Number of BSs\UTs	3\6
Number of antennas at BS\UT	3\1
Shadow fading, γ_{SF}	$\mathcal{N}(0, 8 \text{ dB})$
Pathloss model, γ_{PL} (d in Kms)	$128.1 + 37.6 \cdot \log_{10}(d)$
Rayleigh fast fading, Γ	$\mathcal{CN}(0, 1)$
BS antenna gain, G	9 dBi
Correlation between antennas at BS, ρ	0.5
Number of channel realizations	10^4
Max. BS Tx power with cell-edge SNR=15 dB	0.0603 W (17.8 dBm)
Noise bandwidth	1 MHz
Noise figure	0 dB
Active set threshold for LFB	10 dB

5.5. Simulation results

Consider the cluster layout in Figure 5.2, where $K = 3$ BSs with $N_{\text{T}} = 3$ antennas each, are serving $M = 6$ single antenna UTs. The UTs are uniformly dropped at the cluster center, along an ellipse with semi-major and semi-minor axis of length $\frac{R}{16}$ and $\frac{h/2}{16}$, respectively. $R = 500$ m is the radius of the cell and h is the height of the hexagon of the cluster area. The correlation between the antennas at the BS is $\rho = 0.5$. The pathloss, γ_{PL} , is modeled based on 3GPP pathloss model [28], with shadow fading, γ_{SF} , of $\mathcal{N}(0, 8 \text{ dB})$ and a Rayleigh fast fading component, Γ , which is

simulated as a circularly symmetric complex Gaussian random variable as $\mathcal{CN}(0, 1)$. The channel between the k th BS and the m th UT is calculated as

$$\mathbf{h} = \Gamma \cdot \mathbf{C}^{\frac{1}{2}} \cdot \sqrt{G \cdot \gamma_{\text{SF}} \cdot \gamma_{\text{PL}}}, \quad (5.12)$$

where G is the gain of the antenna at the BS and \mathbf{C} is the correlation matrix of size $N_{\text{T}} \times N_{\text{T}}$. The simulation parameters are summarized in Table 5.2. Note that in order to compare PSO and ZF, only the cases providing invertibility of the aggregated channel matrix are considered. As the focus is at the cluster center along an ellipse as defined earlier, the ZF approach fails to invert only 0.22% of the time. Nevertheless, the probability for failure to invert increases as the UTs move closer to a BS as shown in [13] with a realistic WINNER II channel model (scenario B1, urban micro-cell, non-line of sight). But, PSO would still be able to find a solution when ZF fails. The parameters governing PSO are summarized in Table 5.3.

Various configurations of the ZF and PSO precoders are considered for comparison. They are summarized in Table 5.4. A simple power allocation is performed as per (Equation 5.3). In case of ZF, the power constraint is always applied after the pseudoinverse, and for the scenarios involving PSO, the power constraint is applied in every iteration or after convergence, (refer to subsection 5.3.1 for a more detailed explanation). To obtain a better equilibrium, the solution obtained from PSO with FFB and LBH is fed to one of the particles in the PSO with FFB and FBH during the stochastic initialization stage, for PwrAdj and NoPwrAdj cases, respectively. The scenario PSO:FFB+LBH is considered to simulate the same environment as that of SIN precoding and to compare the same with the corresponding ZF scenario. For the cases of LFB or LBH, an active set threshold of 10 dB was pessimistically considered. This value was decided based on a recent study [29], in which it was found that it is difficult to jointly estimate the channels for a UT with an active set threshold greater than 15 dB at the cell-edge. Also in [29, Figure 4.19], the 10 dB threshold defines a cooperation area that is more focused on the cluster center, while a 15 dB threshold considers the cluster center and cell-edges as well.

Table 5.3.: PSO parameters

Parameters	Values
Number of particles, Q	30
Number of variables, n	Number of \Re & \Im BF coeffs.
$x_{\max} = -x_{\min}$	$1/\max\{ \tilde{\mathbf{H}}(m,l) \}$
Time step length, Δt	1
Max. velocity, v_{\max}	$(x_{\max} - x_{\min})/\Delta t$
Cognitive factor, c_1	2
Social factor, c_2	2
Inertia weight, w	1.4 \rightarrow 0.4
Constant decay factor, β	0.99
Max. number of iterations	500

Table 5.4.: Various precoding configurations (Figure legends): Limited Feedback (LFB) and Limited BackHauling (LBH) refer to an active set threshold of 10dB while Full FeedBack (FFB) and Full BackHauling (FBH) refer to an active set threshold of ∞ .

Nos.	Precoder	Feedback	Backhaul	Power Constraint
1	PSO:FFB+FBH+PwrAdj	Full	Full	Every iteration
2	PSO:FFB+FBH+NoPwrAdj	Full	Full	After convergence
3	PSO:LFB+LBH+PwrAdj	Limited	Limited	Every iteration
4	PSO:FFB+LBH+PwrAdj	Full	Limited	Every iteration
5	PSO:FFB+LBH+NoPwrAdj	Full	Limited	After convergence
6	ZF:LFB+LBH	Limited	Limited	After ZF
7	ZF:FFB+LBH	Full	Limited	After ZF
8	ZF:FFB+FBH	Full	Full	After ZF

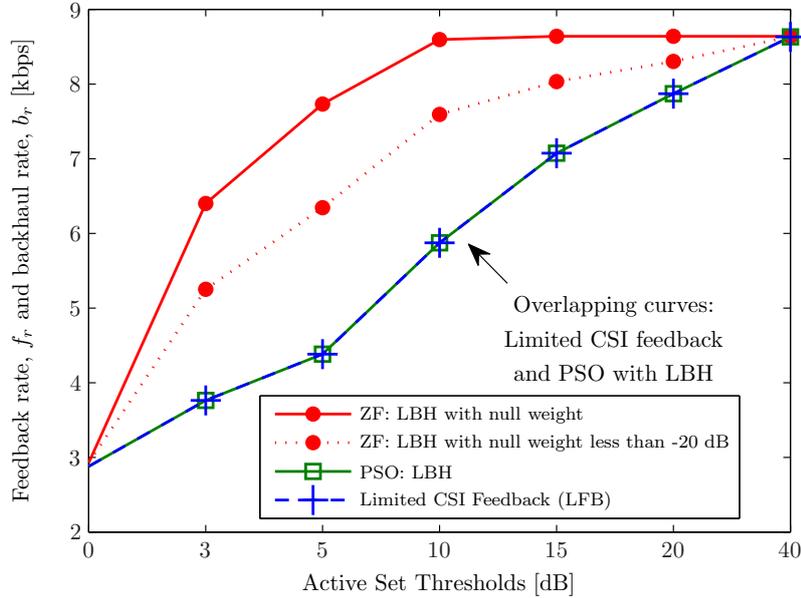


Figure 5.3.: Rate at which the average number of CSI feedback coefficients and the average number of precoding weights in the backhaul that needs to be transmitted for various active set thresholds for all the 6 UTs is shown in the above figure. User data is routed at the CCN based on the non-zero precoding weights which also translates to the reduction in the user data paths in the backhaul.

Figure 5.3 shows the rate at which the average number of CSI coefficients are fed back (LFB) and the rate at which the average number of precoding weights are backhauled (LBH) for various active set thresholds. In particular, for an active set threshold of 10 dB, the feedback rate, f_r due to the CSI coefficients of all the $M = 6$

UTs is 587.3 kbps, which is calculated as,

$$f_r = \frac{(\text{Average number of coefficients}) \cdot (\text{Number of bits})}{\text{Scheduling interval}}, \quad (5.13)$$

assuming that every complex coefficient takes 16 bits for quantization and a scheduling interval of 1 ms (LTE). Likewise, following a similar approach, the backhauling rate, b_r , due to precoding weights with PSO is 587.3 kbps. Hence, the feedback load reduction is equivalent to the backhaul load reduction. In case of a ZF approach, precoders that show zeros for nulls in the beamformer have a higher rate of 859.4 kbps, thereby increasing the backhaul load. Relaxing the null constraint for the ZF approach by treating a threshold of less than 20 dB as a null in the BF, still yields a higher backhaul rate of 759.3 kbps. The reduction in the backhaul load in terms of precoding weights also translates to the reduction in the user data distribution in the backhaul, as the user data can be selectively routed to a given BS based on the non-zero precoding weights. It should be reiterated here that the ZF approach could have nulled the weights when the BSs required them and thereby reducing the sum rate. This is not accounted in this figure. Hence, for a given active set threshold, PSO achieves the exact bound for the backhaul load being equivalent to the feedback load. Note that in a wideband system, the CSI would be estimated and fed back based on the pilot positions. The estimated CSI would be interpolated for a group of subcarriers, as they are smaller than the coherence bandwidth of the channel and thus this group of subcarriers would experience flat fading. The precoding weights obtained at the CCN would be based on the estimated CSI and could be applied over this group of subcarriers. Hence, every CSI coefficient fed back by the UT still would map to a corresponding precoding weight. However, with a ZF approach, the user data being routed at the CCN, as shown in Figure 5.1, would cause a substantial and unnecessary increase in the backhaul. This could be avoided with the proposed PSO. It should be noted that in Figure 5.3 the backhaul rate, b_r , (due to precoding weights) does not include the user data rate. The user data rate would be several orders of magnitude larger than the feedback rate, and could be proportionally reduced with selective routing as described above.

5.5.1. Objective function: Weighted interference minimization

PSO with LFB and LBH performs better than ZF under comparable configurations. The Cumulative Distribution Function (CDF) of the sum rate per cell is shown in Figure 5.4. The PSO with LFB and LBH performs better than the ZF with LFB and LBH by 66.53% on average. PSO with LFB and LBH also performs better than ZF with FFB and LBH by 43.73% on average. The ZF with FFB and FBH performs better than PSO with FFB and FBH with PwrAdj in every iteration, but PSO with NoPwrAdj shows the best average sum rate compared to the other scenarios considered. This is primarily due to the fact that PSO with NoPwrAdj effectively uses the BS peak power constraint. The ZF with LFB and FBH (without backhaul load reduction) performs better than the ZF with FFB and LBH (with backhaul load reduction). This is similar to the results observed in [9], since the signals received by the UT from BSs outside the active set are seen as desired signals and thus help the UTs to accumulate more energy, but it leads to unnecessary backhauling. Due to this minor gain and the undesired additional backhauling, this scenario, ZF with LFB and FBH, is not considered in the following plots.

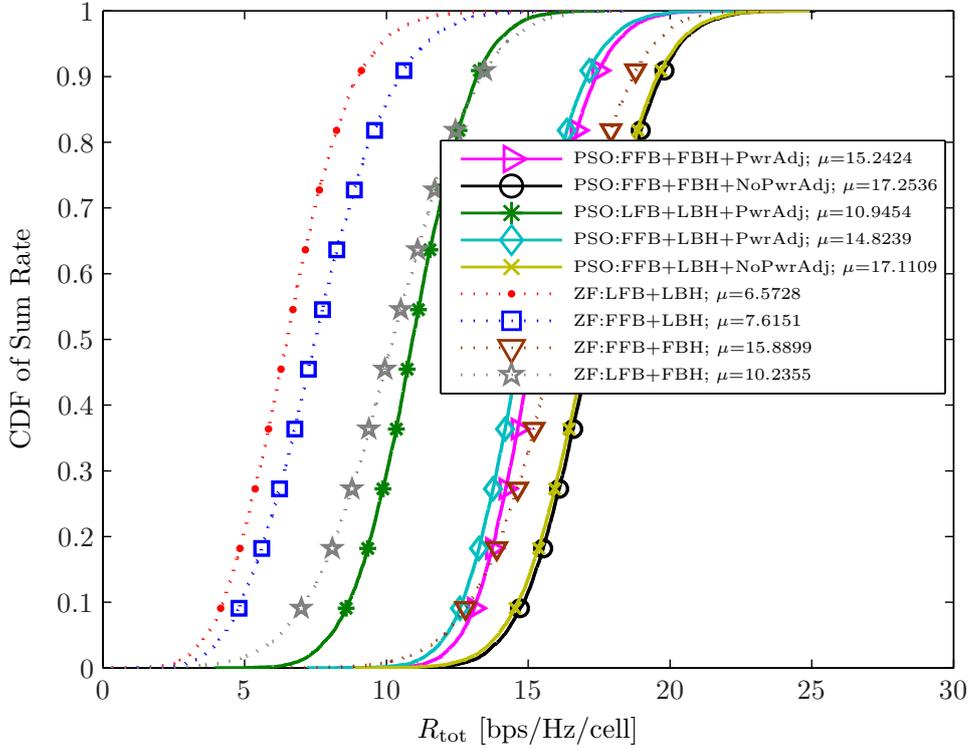


Figure 5.4.: CDF of the sum rate. PSO with objective function: weighted interference minimization. The value in the legend denotes the mean value μ in a given CDF.

Alternately, PSO with the objective of maximizing the minimum SINR of the UT was simulated. In the case of LFB and LBH, a 2.1% relative increase in the average

sum rate per cell was observed when compared to weighted interference minimization but at the cost of 7.7% relative increase in BSs power consumption and 45% relative increase in interference. As expected interference is greatly affected, hence, weighted interference minimization is preferred.

PSO utilizes the available transmit power constraint of P_{\max} per BS more effectively, and at the same time, it improves the weakest SINR UT. The CDF of the SINR of any UT for various precoding algorithms in any channel realization is shown in Figure 5.5 with reasonable improvement in the SINR of the weakest UT (the lower part of the CDF). There is an improvement of 2.97% in the average SINR of PSO compared to ZF, under the same conditions of LFB and LBH. We define the SINR difference as $[\Delta \text{SINR}]_{\text{dB}} = [\text{SINR}_m]_{\text{dB}} - [\text{SINR}_{m'}]_{\text{dB}}$, where UT m , $m \neq m'$, experiences the best SINR while UT m' experiences the worst SINR in a *given* channel realization. The CDF of this SINR difference is shown in Figure 5.6. With this objective function, where the worst SINR UT is taken care explicitly, the PSO has a much lower variance compared to the ZF approach. As expected, the ZF approach with FFB and FBH has all the UTs with equal SINR, hence the difference is zero. It is interesting to note that PSO with FFB and LBH with PwrAdj and NoPwrAdj are nearly 15 dB apart in the SINR difference between the best and the worst UT. This is because in the case of NoPwrAdj, applying (Equation 5.3) disfigures the BF weights obtained from the PSO after convergence.

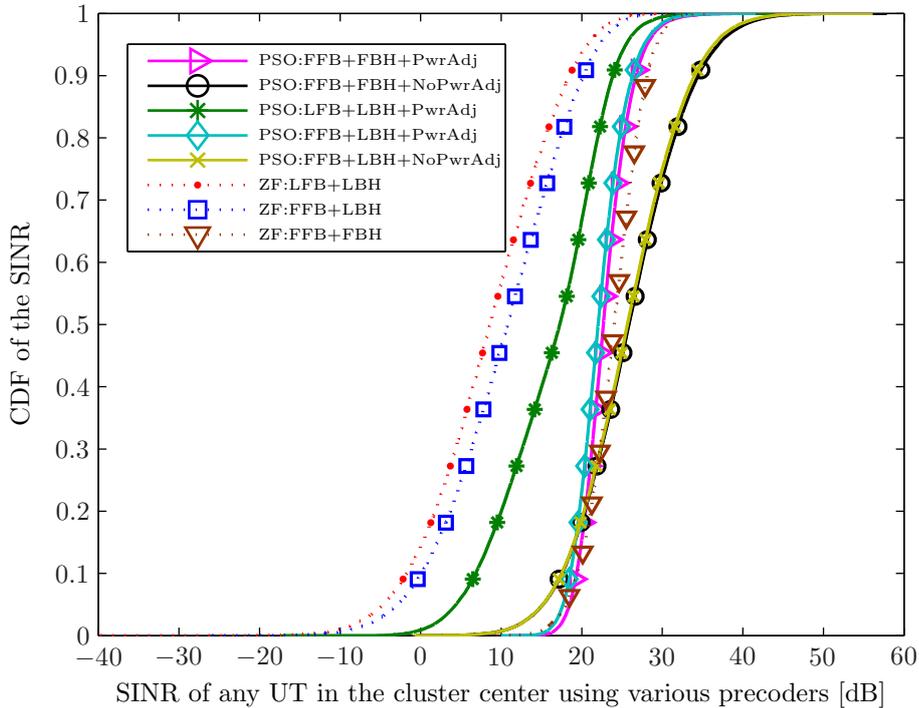


Figure 5.5.: CDF of the SINR of any UT for various precoders in any channel realization. PSO with objective function: weighted interference minimization.

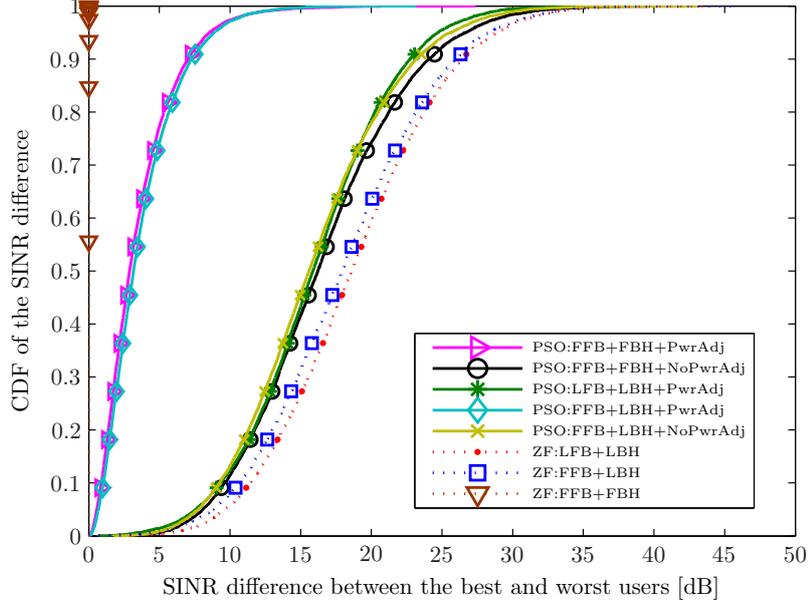


Figure 5.6.: CDF of the SINR difference, $[\Delta SINR]_{dB} = [SINR_m]_{dB} - [SINR_m']_{dB}$ between the best and the worst UTs in a *given* channel realization. PSO with objective function: weighted interference minimization.

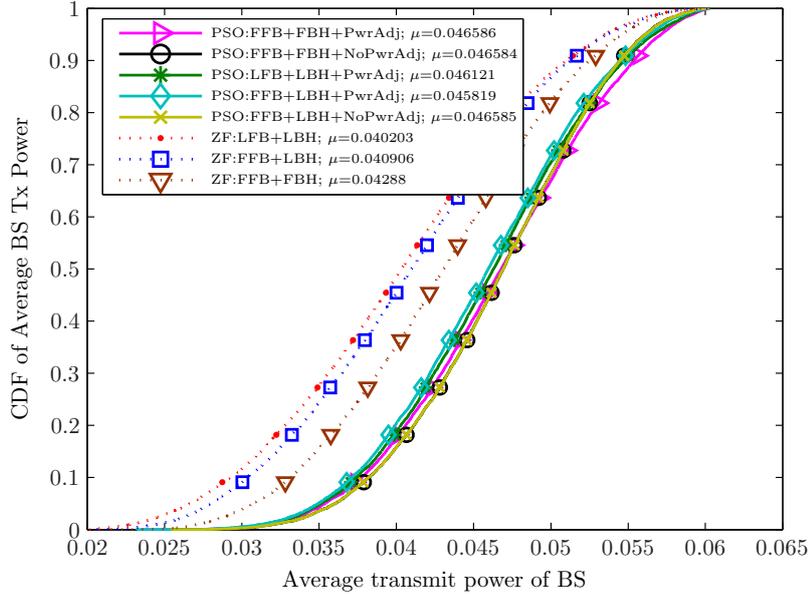


Figure 5.7.: CDFs of the average BS Transmit Power, $\max = 0.0603 \text{ W} (17.8 \text{ dBm})$, with cell-edge SNR of 15 dB. PSO with objective function: weighted interference minimization. The value in the legend denotes the mean value μ in a given CDF.

The CDF of the average transmitted power at the BS is shown in Figure 5.7. The maximum BS transmit power is 0.0603 W, corresponding to a cell-edge SNR of

15 dB. In fact, the PSO keeps the BS power amplifiers *on* at a higher power, most of the time, which is a desired property for amplifier efficiency. The power consumption is reduced beforehand with the limited feedback and limited backhauling. PSO uses BS transmit power more effectively when there is a null constraint on the BF (LBH).

It can be seen in Figure 5.8 that the PSO allows interference compared to the ZF scenarios. The sum rate is improved even when some interference is remaining in the system. This is similar to the SIN technique in [12], but the SIN technique requires full CSI at the CCN. It is also observed that ZF with FFB and FBH completely removes the interference, but this scheme does not use the available BS transmit power effectively as shown in Figure 5.7. In the case of PSO, as observed in Figure 5.8, if LBH is preferred then LFB should also be preferred, as PSO with LFB and FFB under PwrAdj shows the same residual interference in the system.

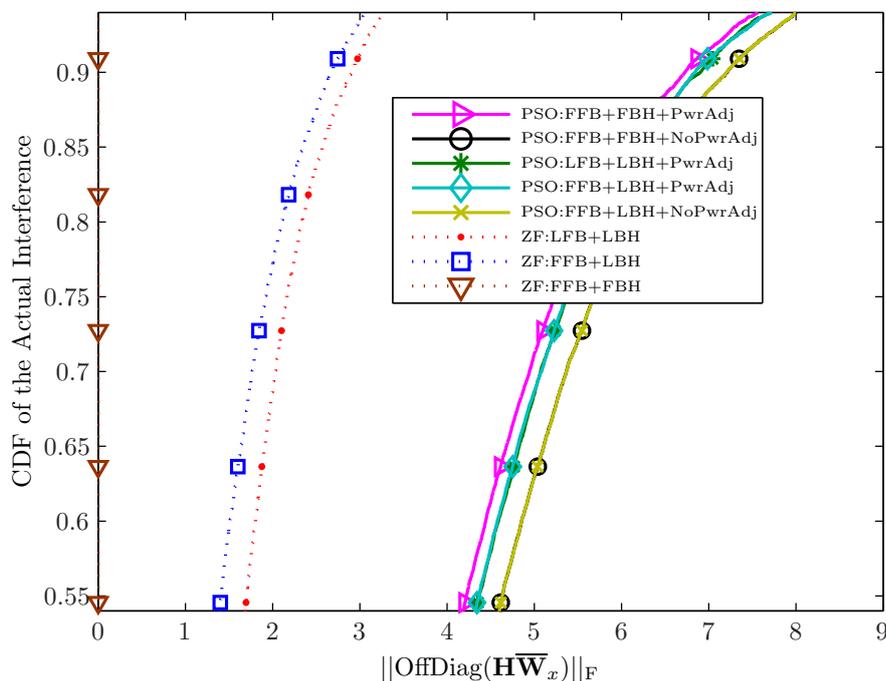


Figure 5.8.: The plot shows the cropped CDF of the actual interference due to the various precoders. ZF with FFB and FBH completely removes the interference, hence the yellow curve is on the y-axis. PSO with objective function: weighted interference minimization.

The convergence of the PSO algorithm when evaluating the objective function of weighted interference minimization for four randomly chosen aggregated channel matrix realizations is shown in Figure 5.9 for various precoder configurations. This objective function converges in less than 100 iterations when PSO is applied with LFB and LBH. It can be observed that the number of iterations to find a stable solution is comparatively fast when the number of BF coefficients is small. The convergence pattern of FBH is not examined here. This is because the PSO with

FFB and FBH has one of the particles fed with the corresponding solution of PSO with FFB and LBH, during the stochastic initialization phase. This was done to show that the PSO implemented in this paper only finds a stable equilibrium solution and not the global optimum, as increasing the dimensionality of the problem makes it harder for the PSO, i.e., PSO with unconstrained backhauling, FBH, yielded a slightly poor solution compared to the PSO with constrained backhauling (LBH), when the objective function was sum rate maximization. To unify our PSO proposal, both objective functions followed the same procedure. This is one of the main reasons why the convergence curves of PSO with FBH remain relatively flat.

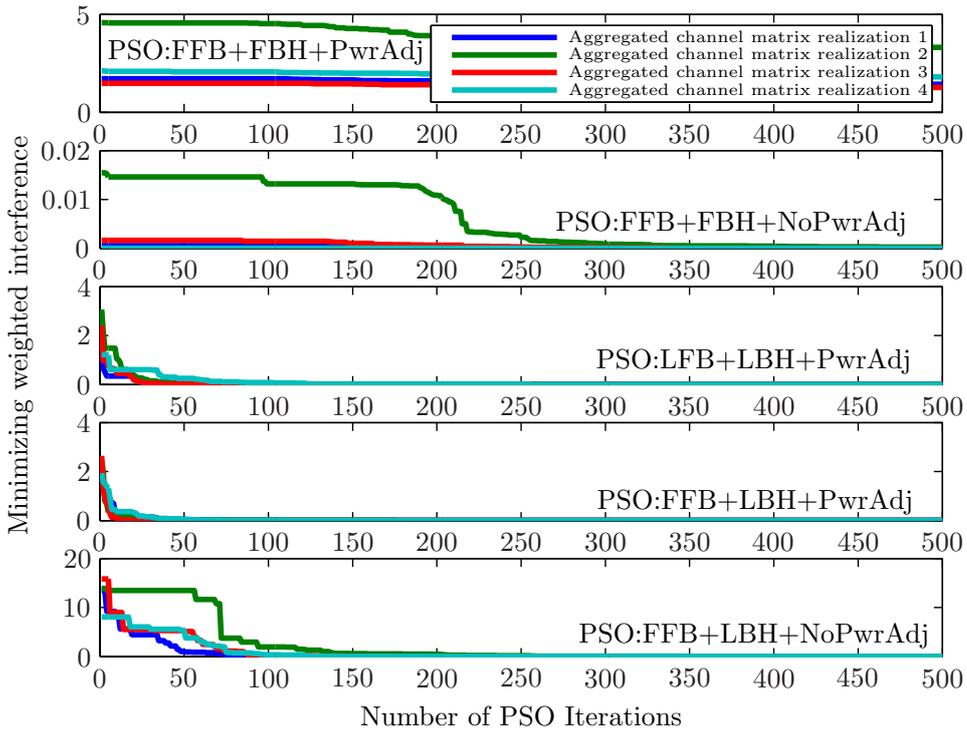


Figure 5.9.: Convergence of PSO for various configurations. PSO with objective function: weighted interference minization.

Based on the analysis in subsection 5.3.4 and on the prior experience, the number of BF coefficients carried by a particle decreases with the sparsity of the aggregated channel matrix. With LBH, the PSO converges faster than the case when there is FBH. Reference to Figure 5.9 could be unfair, due to the reason cited earlier that the solution of PSO with LBH is fed to one of the particles in the case of FBH. If this is not performed, then the faster convergence of the PSO is observed (not shown here).

5.5.2. Objective function: Sum rate maximization

When the objective of the PSO is to maximize the sum rate, the maximization is indirectly related to the particles in the PSO carrying the BF weights via the logarithm. This objective is very sensitive to the power adjustment performed after the PSO algorithm has converged. The CDF of the SINR of any UT is shown in Figure 5.10. It can be observed that the ambition of improving only the sum rate of the system penalizes the weak SINR UTs.

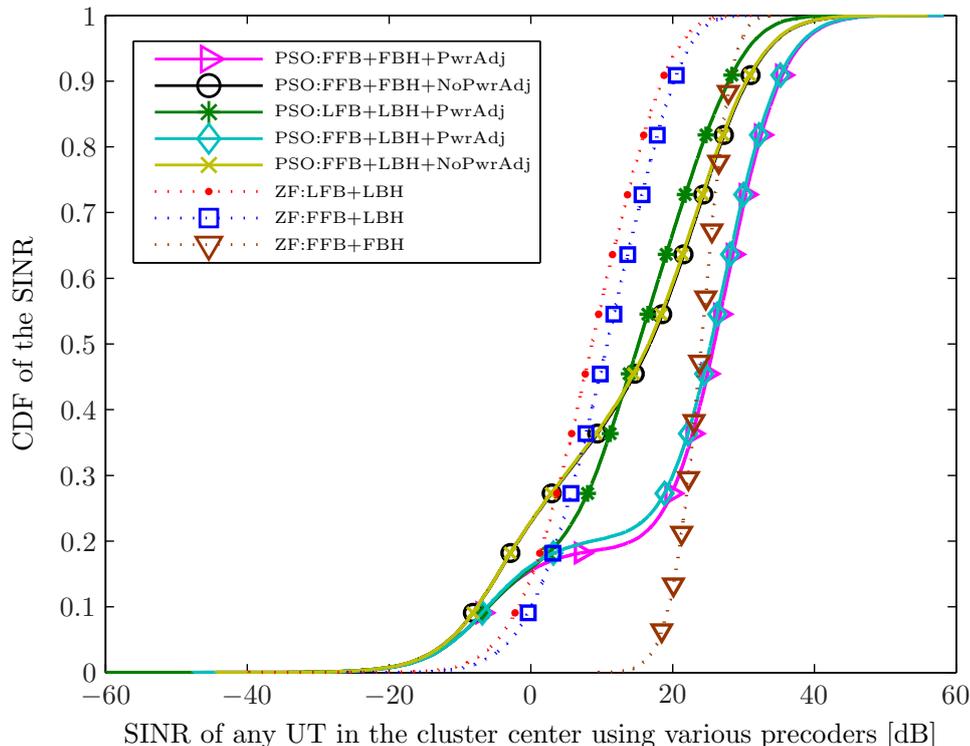


Figure 5.10.: CDF of the SINR of any UT for various precoders in any channel realization. PSO with objective function: sum rate maximization.

5.5.3. Gershgorin's circles

In the complex plane, Figure 5.11 shows the circumference of the Gershgorin's discs for various precoders, with the objective of the PSO being weighted interference minimization. This figure is plotted for a given reference SINR value of the PSO with FFB and FBH. The receiver noise is assumed to be uniform across all the UTs, and SINR is plotted instead of SIR. The green '+' refers to the elements in the diagonal of the matrix $\mathbf{D} = \text{diag}(\mathbf{H}\overline{\mathbf{W}})$, representing the Gershgorin's estimate of the eigen value. The absolute sum of the off diagonal elements forms the radius of the blue Gershgorin's circles for that eigen value and it is plotted with the green '+' as its center. The blue bigger circles show the multiuser interference remaining in the system for a given precoder. The actual eigen values are plotted in red squares

as ‘□’. It can be seen that the PSO gives more freedom for the eigen values to move around in the complex plane, thereby increasing the power transmitted to the UTs. ZF with FFB and FBH completely removes the interference and hence the blue multiuser interference circles are not visible. The ZF approach aims to serve all the UTs equally and hence their actual eigen values are closer together.

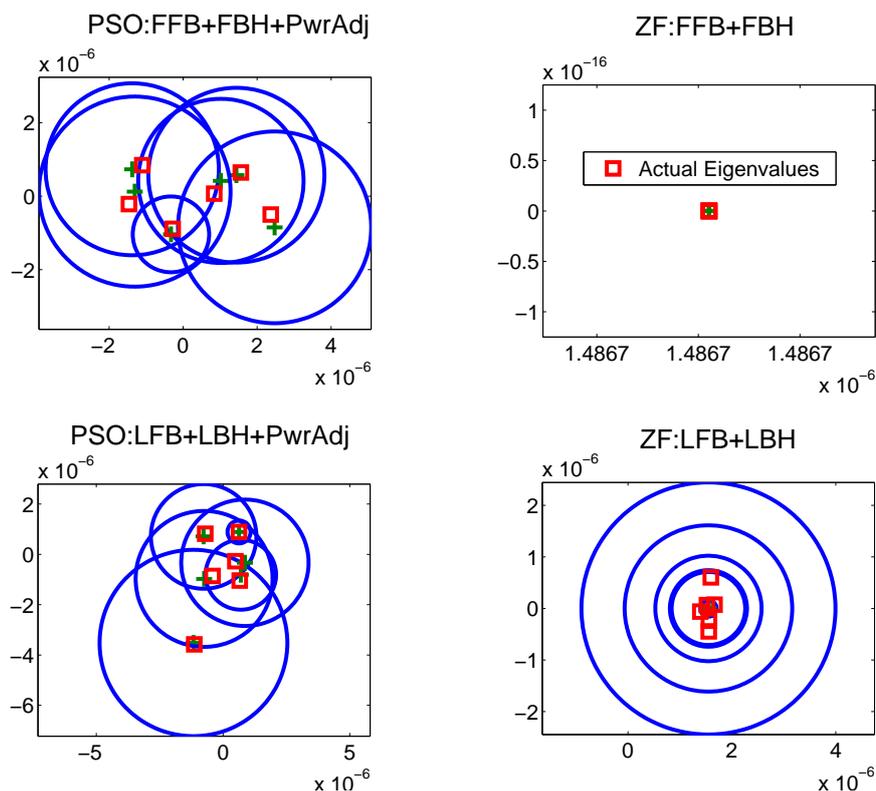


Figure 5.11.: Gershgorin discs of $\overline{\mathbf{H}\mathbf{W}}$. The green “+” is the Gershgorin’s estimate of the eigen values. Bigger blue circles denote the multiuser interference remaining in the system. This plot is mapped to the SINR of PSO with FFB and FBH with PwrAdj equal to 26 dB. PSO with objective function: weighted interference minimization.

It is interesting to note that for the PSO with LFB and LBH, the actual eigen values map closely to the estimated Gershgorin’s eigen values, unlike the ZF with LFB and LBH. From an interference point of view, having concentric circles helps containing the interference within the largest circle. ZF with LFB and LBH shows this attribute.

5.6. Conclusions

In this work, a particle swarm stochastic optimization algorithm has been proposed in a partial joint processing framework to design the precoding weights for efficient backhauling, achieving a backhaul reduction proportional to the reduction in the channel state information feedback. In this context, two objective functions have been considered, a weighted interference minimization and a sum rate maximization. In the proposed weighted interference minimization, the SINR of the weakest user terminal is iteratively improved, in addition to the interference minimization. With the limited feedback and limited backhaul constraints, and the weighted interference minimization as the objective function, the average sum rate per cell of the user terminals is improved by 66.53% with respect to a zero forcing precoder. The particle swarm based precoder allows some multiuser interference to remain in the system, still improving the sum rate, and it uses the BS transmit power more effectively.

With recent developments in swarm intelligence, the complexity and the feasibility can be improved to achieve a faster and a more robust particle swarm algorithm. There is potential for improving the particle swarm optimization algorithm with capabilities to perform global search, such as random particle swarm optimization, which should improve the already promising results presented in this paper.

List of abbreviations

BF	Beamformer (-ing)
bps	bits per second
BS(s)	Base Station(s)
CCN	Central Coordination Node
CDF	Cumulative Distribution Function
CJP	Centralized Joint Processing
CoMP	Coordinate MultiPoint (transmission)
CSI	Channel State Information
FBH	Full BackHauling
FFB	Full FeedBack
JP	Joint Processing
LBH	Limited BackHauling
LFB	Limited FeedBack

MAC	Medium Access Control
MUMIMO	MultiUser Multiple Input Multiple Output
NoPwrAdj	No Power Adjustment
PJP	Partial Joint Processing
PHY	Physical
PwrAdj	Power Adjustment
PSO	Particle Swarm Optimization
SIR	Signal to Interference Ratio
SINR	Signal to Interference plus Noise Ratio
SNR	Signal to Noise Ratio
SIN	Soft Interference Nulling
UT(s)	User Terminal(s)
ZF	Zero Forcing

Competing Interests

The authors declare that they have no competing interests.

Acknowledgements

The authors would like to thank the reviewers for their critical comments that greatly improved the paper. This work has been partly supported by the Swedish Research Council, within the project 621-2009-4555 Dynamic Multipoint Wireless Transmission. This work is also been supported by The Swedish Agency for Innovation Systems (VINNOVA) and by the EU FP7 project INFOS-ICT-247223 ARTIST4G. C. Botella's work has been supported by the Spanish MEC Grants CONSOLIDER-INGENIO 2010 CSD2008-00010 "COMONSENS" and COSIMA TEC2010-19545-C04-01. The authors would also like to acknowledge the members of the VR meetings, in particular Jingya Li and Agisilaos Papadogiannis for the valuable discussions. Thanks to Bhavishya Goel for the fruitful discussions on complexity. The computations were performed on C³SE computing resources.

Bibliography

- [1] 3GPP TR 36.814-900, 3rd Generation Partnership Project; Technical specification group radio access network; Further advancements for E-UTRA physical layer aspects (Release 9), (2010)
- [2] MK Karakayali, GJ Foschini, RA Valenzuela, Network coordination for spectrally efficient communications in cellular systems, *IEEE Wireless Communications*, **13**(4), pp. 56-61, (2006)
- [3] D Gesbert, S Hanly, H Huang, SS Shitz, O Simeone, Wei Yu, Multi-cell MIMO cooperative networks: a new look at interference, *IEEE Journal on Selected Areas in Communications*, **28**(9), pp. 1380-1408, (2010)
- [4] A Papadogiannis, E Hardouin, D Gesbert, Decentralising multi-cell cooperative processing: a novel robust framework, *EURASIP Journal on Wireless Communications and Networking*, Article ID 890685, (2009)
- [5] P Marsch and G Fettweis, A framework for optimizing the downlink of distributed antenna systems under a constrained backhaul, in *Proc. European Wireless Conference* (2007)
- [6] P Marsch and G Fettweis, On multicell cooperative transmission in backhaul-constrained cellular systems, *Ann. Telecommun.*, **63**, p.253, (2008)
- [7] C Botella, T Svensson, X Xu, H Zhang, On the performance of joint processing schemes over the cluster area, in *Proc. IEEE Vehicular Technology Conference*, (2010)
- [8] A Papadogiannis, H Bang, D Gesbert, E Hardouin, Downlink overhead reduction for multicell cooperative processing enabled wireless networks, *IEEE Personal, Indoor and Mobile Radio Communications* (2008)
- [9] A Papadogiannis, HJ Bang, D Gesbert, E Hardouin, Efficient selective feedback design for multicell cooperative networks, *IEEE Transactions on Vehicular Technology*, **60**(1), pp. 196-205, (2011)
- [10] F Boccardi, H Huang, A Alexiou, Network MIMO with reduced backhaul requirements by MAC coordination, in *Proc. IEEE Asilomar Conference on Signals, Systems and Computers*, (2008)

- [11] J Zhang, R Chen, J G Andrews, A Ghosh, and RW Heath, Networked MIMO with clustered linear precoding, *IEEE Transactions on Wireless Communications*, **8**(4), pp. 1910-1921, (2009)
- [12] CTK Ng, H Huang, Linear precoding in cooperative MIMO cellular networks with limited coordination clusters, *IEEE Journal on Selected Areas in Communications*, **28**(9), pp. 1446-1454, (2010)
- [13] TR Lakshmana, C Botella, T Svensson, X Xu, J Li, X Chen, Partial joint processing for frequency selective channels, *IEEE 72nd Vehicular Technology Conference Fall (VTC 2010-Fall)*, (2010)
- [14] X Wei, T Weber, A Kuhne, A Klein, Joint transmission with imperfect partial channel state information, in *Proc. IEEE Vehicular Technology Conference*, (2009)
- [15] S Fang, G Wu, S-Q Li, Optimal multiuser MIMO linear precoding with LMMSE receiver, *EURASIP Journal on Wireless Communications and Networking*, Article ID 197682, (2009)
- [16] Y Hei, X Li, K Yi, H Yang, Novel scheduling strategy for downlink multiuser MIMO system: particle swarm optimization, *Science in China F*, **52**(12), pp. 2279–2289, (2009)
- [17] C Knievel, PA Hoeher, A Tyrrell, G Auer, Particle swarm enhanced graph-based channel estimation for MIMO-OFDM, *IEEE 73rd Vehicular Technology Conference Fall (VTC 2011-Fall)*, (2011)
- [18] P Marsch and G Fettweis, On downlink network MIMO under a constrained backhaul and imperfect channel knowledge, *Global Telecommunications Conference* (2009)
- [19] Q Spencer, A Swindlehurst, M Haardt, Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels, *IEEE Transactions on Signal Processing*, **52**(2), pp. 461-471, (2004)
- [20] H Zhang and H Dai, Cochannel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks, *EURASIP Journal on Wireless Communications and Networking*, **2**, pp. 222-235, (2004)
- [21] TR Lakshmana, C Botella, T Svensson, “Partial Joint Processing with Efficient Back-hauling in Coordinated MultiPoint Networks”, in *Proc IEEE Vehicular Technology Conference*, pp. 1–5 (2012)
- [22] J Kennedy, RC Eberhart, Particle swarm optimization, in *Proc. IEEE International Conference on Neural Networks*, pp. 1942–1948, (1995)

-
- [23] AP Engelbrecht. *Fundamentals of computational swarm intelligence*, (John Wiley & Sons, 2005), pp. 171-172
- [24] YH Shi, RC Eberhart, Parameter selection in particle swarm optimization, *The 7th Annual Conference on Evolutionary Programming*, San Diego, USA, (1998)
- [25] C-B Chae, S-H Kim, RW Heath, Network coordinated beamforming for cell-boundary users: Linear and Nonlinear Approaches, *IEEE Journal of Selected Topics in Signal Processing*, **3**(6), pp. 1094-1105, (2009)
- [26] CB Peel, BM Hochwald, AL Swindlehurst, A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization, *IEEE Transactions on Communications*, **53**(1), pp. 195-202, (2005)
- [27] RA Horn and CR Johnson, *Matrix Analysis*, (Cambridge University Press, 1985), pp. 344-347
- [28] 3GPP TR 36.942-a20, 3rd Generation Partnership Project; Technical specification group radio access network; Evolved universal terrestrial radio access; Radio frequency system scenarios (Release 10), (2011)
- [29] ARTIST4G D1.2, Innovative advanced signal processing algorithms for interference avoidance, *ARTIST4G technical deliverable*, <https://ict-artist4g.eu/projet/work-packages/wp1/documents/d1.2/d1.2.pdf>, pp. 84. Accessed 17 June 2010

Paper C

Scheduling for Backhaul Load Reduction in CoMP

Tilak Rajesh Lakshmana, Jingya Li, Carmen Botella, Agisilaos Papadogiannis, and
Tommy Svensson

Published in
IEEE Wireless Communications and Networking Conference (WCNC), Apr. 2013

©2012 IEEE

The layout has been revised.

6. Scheduling for Backhaul Load Reduction in CoMP

Abstract

Coordinated multi-point (CoMP) transmission has received a lot of attention, as a way to improve the system throughput in an interference limited cellular system. For joint processing in CoMP, the user equipments (UEs) need to feed back the channel state information (CSI), typically to their serving base stations (BSs). The BS forwards the CSI to a central coordination node (CCN) for precoding. These precoding weights need to be forwarded from the CCN to the corresponding BSs to serve the UEs. In this work, a feedback load reduction technique is employed via partial joint processing to alleviate the CSI feedback overhead. Similarly, to achieve backhaul load reduction due to the precoding weights, scheduling approaches are proposed. The state of the art block diagonalization solution is compared with our proposed constrained and unconstrained scheduling. Our main contribution is the method of choosing the best subset of the BSs and UEs at the CCN that yields the best sum rate under the constraint of efficient backhaul use. In particular, with constrained scheduling, the choice of a smaller subset proportionally reduces the backhaul load. Simulation results based on a frequency selective WINNER II channel model, show that our proposed constrained scheduling outperforms the block diagonalization approach in terms of the average sum rate per backhaul use.

Keywords—Backhaul Load Reduction, Scheduling, CoMP, Partial Joint Processing, Zero Forcing

6.1. Introduction

In future cellular communication systems, coordinated multi-point (CoMP) transmission is a promising technique proposed to improve the throughput of the user equipments (UEs) at the cell edge, being limited by interference [1]-[2]. To realize these gains, the UEs need to feed back the channel state information (CSI) typically to their serving base station (BS). The CSI is forwarded to the central coordination node (CCN) to form the aggregated channel matrix that is used to create the precoding matrix to jointly mitigate interference. To reduce the overhead of feeding back the CSI, clusters of BSs are formed [2]. In particular, partial joint processing (PJP) was proposed in [3] for feedback load reduction, in which dynamically

overlapping clusters of BSs are formed.

PJP can be seen as a framework that attempts to categorize the trade off between how much load can be reduced for a given perceivable loss in the system performance. In this regard, the CSI feedback load reduction can be achieved by limiting the quantity of feedback by the UEs. To this end, a relative thresholding is proposed in [3], where the UE only feeds back the CSI for a set of BSs links that fall within a threshold relative to its strongest BS. The CSI of the BSs that fall outside this threshold are modeled as *zeros* in the aggregated channel matrix. Likewise, in this context, the signaling in the backhaul (BSs-CCN) is primarily due to the distribution of the precoding weights from the CCN to the cooperating BSs. Limiting the feedback causes the aggregated channel matrix to be sparse and poses problems in the case of precoders such as zero forcing (ZF). Under these circumstances, achieving an efficient use of the backhaul is difficult. Hence, the structure in the aggregated channel matrix formed at the CCN needs to be exploited, such that the zeros are correspondingly preserved in the precoder matrix for reducing the backhaul load. To achieve this, backhaul load reduction can be carried out at the medium access control (MAC) layer or physical (PHY) layer, as proposed in [4].

The MAC layer approach is a scheduling based scheme, where disjoint BS sub-groups are formed, such that the aggregated channel matrix is block diagonal. The main benefit of this approach is that the inverse of a block diagonal matrix is still block diagonal and that the zeros are preserved. The PHY layer approach is a ZF precoding approach, where the aggregated channel matrix is repeated such that a block diagonal structure is created, and the precoding matrix is created with zeros where needed. The limitations of the PHY layer approach are discussed in [5].

In this paper, we propose a constrained scheduling (CS) and an unconstrained scheduling (US) for backhaul load reduction. In the CS approach, an exhaustive search is carried out to find the best subset of the aggregated channel matrix, such that zeros are avoided in the aggregated channel matrix. This approach directly aims at reducing the backhaul load as the zeros are disallowed. The US approach is similar to the CS approach except that the zeros are allowed to be present in the aggregated channel matrix, and the backhaul load reduction is achieved by explicit nulling of the precoding weights corresponding to the zeros in the aggregated channel matrix. We compare our techniques with the MAC layer scheduling based block diagonalization (BD) technique proposed in [4]. The BD approach achieves the backhaul load reduction by forming a block diagonal structure of the aggregated channel matrix. All the above techniques are evaluated with the PJP based CSI feedback load reduction as proposed in [3]. To summarize our contribution, our proposed CS and US algorithms (i) reduce the backhaul load, (ii) significantly increase the performance, as a larger feasible subset is considered compared to the BD approach, which poses a stricter constraint of being block diagonal, and (iii) the

Part of this work has been performed in the framework of the FP7 project ICT-317669 METIS, which is partly funded by the European Union. This work is also supported by the Swedish Research Council VR under the project 621-2009-4555 Dynamic Multipoint Wireless Transmission. The computations were performed on C³SE.

best subset of BSs and UEs are clustered.

The paper is organized as follows, in section 6.2 the system model is introduced with the focus on how the feedback and backhaul load reduction are achieved in a frequency selective channel. Discussions on the scheduling strategies for backhaul load reduction are presented in section 6.3. The performance of these scheduling strategies are discussed in section 6.4 and finally the main results are concluded in section 6.5. The notation used in this paper is summarized in the footnote.

6.2. System Model

Consider the cluster layout as shown in Figure 6.1, where $K = |\mathcal{K}|$ single antenna BSs need to serve $M = |\mathcal{M}|$ single antenna UEs. $\mathbf{h}_m = [h_{m,1}, h_{m,2}, \dots, h_{m,K}]$ is the CSI of the links from the K BSs to the m th UE. In this work, we study block-fading channels where the CSI available at the CCN is considered to be error free, i.e., the quantization loss and the backhaul delays are assumed to be negligible [6]. In a wideband system, consisting of a number of subcarriers, each UE feeds back the CSI for a given frequency resource. The CSI being fed back can be applied to a group of subcarriers. The CSI feedback process is performed under the PJP framework proposed in [3], using a relative active set thresholding as summarized in [5, Algo.1]. The CSI feedback from the m th UE based on the channel from K BSs can be represented as

$$\tilde{\mathbf{h}}_m = \mathbf{h}_m \odot \mathbf{t}_{x,m}, \quad (6.1)$$

where $\mathbf{t}_{x,m}(k) \in \{0, 1\}, \forall k = 1, \dots, K$. The operation is independently performed over a collection of subcarriers (frequency adaptive thresholding) [7], where all the M UEs report the CSI for all the subcarriers. When the m th UE feeds back the CSI of k th BS, it is denoted “1” while a “0” denotes that the CSI was not fed back. The feedback load reduction can be seen as a masking operation by a binary threshold vector $\mathbf{t}_{x,m}$ via element wise multiplication with the CSI measured by the m th UE. The subscript x denotes the threshold value in dB. When $x = 0$ dB, it represents a scenario where only the strongest BS is serving the UE, while the threshold of $x = \infty$ dB represents that all the links are fed back. If $K = 3$ then the m th UE will feed back in one of the following ways: $\mathbf{t}_{x,m} = \{\{1, 0, 0\}, \{0, 1, 0\}, \{0, 0, 1\}, \{1, 1, 0\}, \{0, 1, 1\}, \{1, 0, 1\}, \text{ and } \{1, 1, 1\}\}$. However, with relative thresholding [3], $\mathbf{t}_{x,m} = \{0, 0, 0\}$ will never occur, as it enables the UE to feedback at least its strongest BS.

For backhaul load reduction, consider a subset of the cluster formed at the CCN, with $N = |\mathcal{N}|$ UEs and $L = |\mathcal{L}|$ BSs, where $\mathcal{N} \subseteq \mathcal{M}$ and $\mathcal{L} \subseteq \mathcal{K}$. The maximum

Boldface upper-case letters represent matrices, \mathbf{X} , boldface lower-case letters represent vectors, \mathbf{x} , and italics represent scalars, x . The $\mathbb{C}^{m \times n}$ is a complex valued matrix of size $m \times n$. The $(\cdot)^T$ and $(\cdot)^H$ is the transpose and conjugate transpose, respectively. $\mathbf{E}_x \{\cdot\}$ is the expectation with respect to x . The $\|\cdot\|_F$ is the Frobenius norm. $\mathbf{X}(i, j)$ is the (i, j) th element of matrix \mathbf{X} and $\mathbf{x}(i)$ is the i th element of the vector \mathbf{x} . The i th row of a matrix \mathbf{X} is $\mathbf{X}(i, :)$ and the j th column of a matrix \mathbf{X} is $\mathbf{X}(:, j)$. The sets are indicated in calligraphic letters and $|\mathcal{X}|$ denotes the cardinality of the set \mathcal{X} . The operator \odot is the element wise multiplication.

number of BSs that can be chosen is $L_{\max} = K$ and the maximum number of UEs that can be scheduled in a given subcarrier group/resource is $N_{\max} = L_{\max}$. The discrete time signal received at the N selected UEs, $\mathbf{y} \in \mathbb{C}^{N \times 1}$ is

$$\mathbf{y} = \mathbf{H}\widetilde{\mathbf{W}}\mathbf{x} + \mathbf{n}, \quad (6.2)$$

where $\mathbf{H} \in \mathbb{C}^{N \times L}$ is the channel matrix for the subset of the cluster. $\widetilde{\mathbf{W}} \in \mathbb{C}^{L \times N}$ is the precoding matrix and \mathbf{n} is the receiver noise at the UEs, which are spatially and temporally white with variance σ^2 .

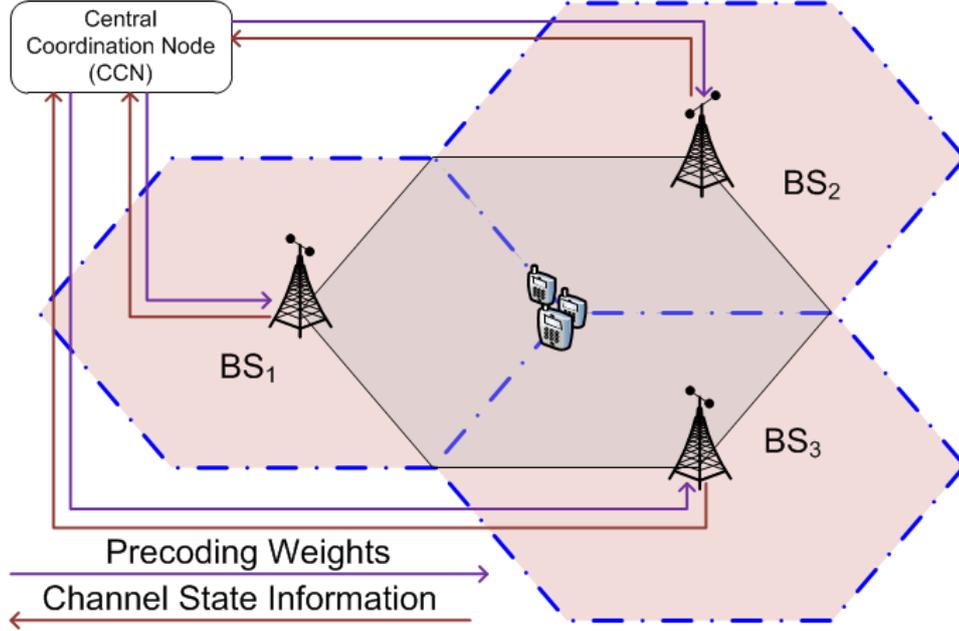


Figure 6.1.: The cluster layout

A linear ZF precoding is considered in this work. The precoding matrix is firstly calculated as the Moore-Penrose pseudoinverse of the aggregated channel matrix $\widetilde{\mathbf{H}}$

$$\widetilde{\mathbf{W}} = \widetilde{\mathbf{H}}^H (\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^H)^{-1}, \quad (6.3)$$

where $\widetilde{\mathbf{H}} = \mathbf{H} \odot \mathbf{T}_x$ and $\mathbf{T}_x = [\mathbf{t}_{x,1}^T, \mathbf{t}_{x,2}^T, \dots, \mathbf{t}_{x,N}^T]^T$. Then, the columns of $\widetilde{\mathbf{W}}$ are normalized to have a unit norm [4]. Finally, based on equal user rate power allocation [8], the precoding matrix can be obtained as

$$\overline{\widetilde{\mathbf{W}}} = \sqrt{\frac{P_{\max}}{\left(\max_{l=1, \dots, L} \|\widetilde{\mathbf{W}}(l, :)\|_F^2\right)}} \cdot \widetilde{\mathbf{W}}, \quad (6.4)$$

where P_{\max} is the maximum power at which a BS can transmit on a given resource, i.e., we are not considering optimal power allocation over the parallel resources. The signal to interference plus noise ratio (SINR) for the n th UE is given as

$$\text{SINR}_n = \frac{\|\mathbf{h}_n \widetilde{\mathbf{W}}(:, n)\|^2}{\sum_{j \in \mathcal{N}, j \neq n} \|\mathbf{h}_n \widetilde{\mathbf{W}}(:, j)\|^2 + \sigma^2}. \quad (6.5)$$

The sum rate in bps/Hz for scheduling the N different UEs on the same frequency/time resource is

$$R_{\text{tot}} = \sum_{n \in \mathcal{N}} \log_2(1 + \text{SINR}_n). \quad (6.6)$$

Due to feedback load reduction, the channel matrix $\widetilde{\mathbf{H}}$ might have zero elements depending on the threshold x dB. Hence, a sparse channel matrix is used to obtain the precoding matrix, $\widetilde{\mathbf{W}}$. The zeros in $\widetilde{\mathbf{H}}$ pose problems for the ZF precoder for backhaul load reduction. For example, if the n th UE does not feed back the CSI for the l th BS, then $\widetilde{\mathbf{H}}(n, l) = 0$. Applying the pseudoinverse in (Equation 6.3), the sparse aggregated channel matrix $\widetilde{\mathbf{H}}$ of size $N \times L$, will create $\widetilde{\mathbf{W}}$ of size $L \times N$, however, it could lead to $\widetilde{\mathbf{W}}(l, n) \neq 0$. This will lead to unnecessary backhaul, given that the UE has not fed back the CSI while the ZF solution still tries to serve the n th UE from the l th BS. Also, consider the situation where the n th UE has fed back the CSI for the $(l + 1)$ th BS such that $\widetilde{\mathbf{H}}(n, l + 1) \neq 0$, however (Equation 6.3) might lead to a situation where $\widetilde{\mathbf{W}}(l + 1, n) = 0$. This is poor backhauling as the uplink resources are already being spent for the n th UE to feed back the CSI.

Hence, suitable scheduling strategies need to be developed in achieving an efficient use of the backhaul. These are discussed in the subsequent section.

6.3. Scheduling

Let a set of $\mathcal{N} \subseteq \mathcal{M}$ UEs be chosen to be served from a set of $\mathcal{L} \subseteq \mathcal{K}$ BSs. To maintain orthogonality with a linear ZF precoder, the number of UEs chosen are $N = |\mathcal{N}|$ and the BSs chosen are $L = |\mathcal{L}|$, such that $N \leq L$. The particular choice of the set of UEs and BSs are driven by the combination that maximizes the sum rate as

$$\{\mathcal{L}^*, \mathcal{N}^*\} = \arg \max_{\{\mathcal{L}, \mathcal{N}: |\mathcal{N}| \leq |\mathcal{L}|\}} \sum_{n \in \mathcal{N}} \log_2(1 + \widehat{\text{SINR}}_n), \quad (6.7)$$

$$\widehat{\text{SINR}}_n = \frac{\sum_{l \in \mathcal{L}} \widetilde{\mathbf{H}}(n, l) \mathbf{W}(l, n)}{\sum_{\substack{j \neq n \\ j \in \mathcal{L}}} \sum_{l \in \mathcal{L}} \widetilde{\mathbf{H}}(n, l) \mathbf{W}(l, j) + \sigma^2}, \quad (6.8)$$

where $\widetilde{\mathbf{H}}$ is the channel sub-matrix of size $N \times L$ related to the set $\{\mathcal{N}, \mathcal{L}\}$ of UEs and BSs. Applying (Equation 6.3) and (Equation 6.4) to this $\widetilde{\mathbf{H}}$ results in

the precoding matrix \mathbf{W} . In the following subsections, we evaluate the scheduling strategies considered in this work.

6.3.1. Block Diagonalization (BD)

In [4], a MAC-layer approach is proposed where disjoint subgroups of BSs are formed to preserve the block diagonal structure of the aggregated channel matrix. An important property of a block diagonal structure is that it is preserved even under matrix inversion. This property is key to backhaul load reduction which conserves an equivalent feedback load reduction for the scheduled UEs belonging to the set \mathcal{N} , i.e., if $\tilde{\mathbf{H}}(n, l) = 0$ then $\tilde{\mathbf{W}}(l, n) = 0$. However, it should be noted that this BD approach based on [4] always requires N_{\max} UEs to be scheduled. Therefore, with feedback load reduction, M UEs feeding back the CSI results in $N_{\max} = L_{\max} = K$ UEs being scheduled. The choice of N_{\max} corresponds to an exhaustive search for the best combination of UEs that maximizes the sum rate given that the aggregated channel matrix is block diagonal. The BD approach can be summarized as choosing the combination of the UEs as in Algorithm 6.1. The block diagonal channel matrix $\tilde{\mathbf{H}}_{\text{BD}}$ is extracted based on \mathbf{T}_x and correspondingly $\tilde{\mathbf{W}}_{\text{BD}}$ is obtained from (Equation 6.3).

Due to the BD structure, the positions of zeros in $\tilde{\mathbf{H}}_{\text{BD}}$ and $\tilde{\mathbf{W}}_{\text{BD}}$ are identical, and the aggregated channel matrix needs to be a square matrix such that $N_{\max} = L_{\max}$. This gives rise to some ill-effects. Consider $N < N_{\max}$ and $L < L_{\max}$ then $N < L$ is not considered which could potentially produce a better sum rate translating to a better system performance. This is due to the stringent constraint of the BD approach that $N_{\max} = L_{\max} = K$, where the feasible set is considerably reduced. Also, the feedback can be significantly reduced at the cluster center, via small relative thresholds [7].

Algorithm 6.1 BD approach: Note that the BD algorithm [4] always considers $L = L_{\max} = K$.

```

1:  $M$  UEs feed back the CSI as defined in section 6.2
2: for every  $\mathcal{N}$  from  $\mathcal{M}$  such that  $N_{\max} = |\mathcal{N}| = L_{\max}$  do
3:   Form  $\mathbf{T}_x(n, l) \in \{0, 1\}, \forall n = 1, \dots, N_{\max}; \forall l = 1, \dots, L_{\max}$ 
4:   if permuted  $\mathbf{T}_x$  is block diagonal then
5:     Found  $\mathbf{T}_x$  to have a block diagonal structure, evaluate (Equation 6.7)
6:     Save  $\mathbf{T}_x$  based on the best  $\{\mathcal{L}^*, \mathcal{N}^*\}$  achieved so far
7:   else
8:     Failed to find a block diagonal structure
9:   end if
10: end for
11: return Schedule the subset formed with  $\{\mathcal{L}^*, \mathcal{N}^*\}$  using  $\mathbf{T}_x$ 

```

For example, a threshold of 5 dB creates a sparse aggregated channel matrix which is difficult to block diagonalize. One of the ways to overcome this limitation of the BD approach is to increase the threshold towards infinity. However, this increases

the feedback load. On the contrary, a full aggregated channel matrix with few zeros also renders the BD approach difficult to realize. As a generalization, the BD approach proposed in [4] can be extended to consider the cases when $N = L < K$. However, in this work, we confine our study to the original algorithm proposed in [4].

6.3.2. Unconstrained Scheduling (US)

With feedback load reduction, the aggregated channel matrix is sparse depending on the threshold. The choice of a feasible subset of BSs and UEs, $\{\mathcal{L}^*, \mathcal{N}^*\}$, that produces the best sum rate is summarized in Algorithm 6.2.

Algorithm 6.2 US approach

```

1:  $M$  UEs feed back the CSI as defined in section 6.2
2: Assign  $L = K$ 
3: while  $L \geq 1$  do
4:   for  $\mathcal{L} : \mathcal{L} \subseteq \mathcal{K}; |\mathcal{L}| = L$  do
5:     for every  $\mathcal{N}$  from  $\mathcal{M}$  such that  $N = |\mathcal{N}| \leq L$  do
6:       Form  $\mathbf{T}_x(n, l) \in \{0, 1\}, \forall n = 1, \dots, N; \forall l = 1, \dots, L$ 
7:       Evaluate (Equation 6.7)
8:       Save  $\mathbf{T}_x$  based on the best  $\{\mathcal{L}^*, \mathcal{N}^*\}$  achieved so far
9:     end for
10:  end for
11:   $L = L - 1$ 
12: end while
13: return Schedule the subset formed with  $\{\mathcal{L}^*, \mathcal{N}^*\}$  using  $\mathbf{T}_x$ 

```

The scheduled BSs and UEs in matrix form can be written as $\mathbf{T}_x(n, l) \in \{0, 1\}, \forall n = 1, \dots, N$ and $\forall l = 1, \dots, L$. Hence, the sparse aggregated channel matrix can be written as $\widetilde{\mathbf{H}}_{\text{US}} = \mathbf{H} \odot \mathbf{T}_x$. Compared to the BD approach, the US approach has a flexibility in considering $N \leq L$, and $\widetilde{\mathbf{W}}$ is obtained from $\widetilde{\mathbf{H}}_{\text{US}}$ by applying (Equation 6.3) and (Equation 6.4). This is followed by explicit nulling of the precoded weights as $\widetilde{\mathbf{W}}_{\text{US}} = \widetilde{\mathbf{W}} \odot (\mathbf{T}_x)^T$, to achieve backhaul load reduction based on the nulls due to feedback load reduction. However, explicit nulling gives rise to multi-user interference to remain in the system. It should be noted that the explicit nulling is automatically taken care of in the BD approach in subsection 6.3.1. Explicit nulling seems like an intuitive approach but the ZF precoder has its own limitations when there are zeros in the aggregated channel matrix (see section 6.2).

6.3.3. Constrained Scheduling (CS)

The CS approach is similar to the US approach with an important constraint that the aggregated channel matrix $\widetilde{\mathbf{H}}_{\text{CS}}$ is full due to the proper selection of UEs and BSs, i.e., $\widetilde{\mathbf{H}}_{\text{CS}} = \mathbf{H} \odot \mathbf{T}_x$, where $\widetilde{\mathbf{H}}_{\text{CS}} \in \mathbb{C}^{N \times L}$ and $\widetilde{\mathbf{H}}_{\text{CS}}(i, j) \neq 0, \forall i, j$ as $\mathbf{T}_x(n, l) \in \{1\}, \forall n = 1, \dots, N$ and $\forall l = 1, \dots, L$. This simplifies the ZF in (Equation 6.3). The

CS approach is summarized in Algorithm 6.3. The main advantage of this approach is that the backhaul load reduction is automatically achieved by this constrained scheduling approach as smaller subset of a matrix, $\widetilde{\mathbf{H}}_{\text{CS}}$, is formed from \mathbf{H} . Also, multi-user interference is removed from the system.

Algorithm 6.3 CS approach

```

1:  $M$  UEs feed back the CSI as defined in section 6.2
2: Assign  $L = K$ 
3: while  $L \geq 1$  do
4:   for  $\mathcal{L} : \mathcal{L} \subseteq \mathcal{K}; |\mathcal{L}| = L$  do
5:     for every  $\mathcal{N}$  from  $\mathcal{M}$  such that  $N = |\mathcal{N}| \leq L$  do
6:       Form  $\mathbf{T}_x(n, l) \in \{1\}, \forall n = 1, \dots, N; \forall l = 1, \dots, L$ 
7:       Evaluate (Equation 6.7)
8:       Save  $\mathbf{T}_x$  based on the best  $\{\mathcal{L}^*, \mathcal{N}^*\}$  achieved so far
9:     end for
10:  end for
11:   $L = L - 1$ 
12: end while
13: return Schedule the subset formed with  $\{\mathcal{L}^*, \mathcal{N}^*\}$  using  $\mathbf{T}_x$ 

```

Illustrative Example

To illustrate the above algorithms with an example, consider $\mathbf{T}_x = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

This subset is feasible with the BD approach, while the CS approach requires the zeros to be removed. Hence, a feasible subset \mathbf{T}_x after removing the zeros can be any of these $\left\{ \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, [1 \ 1], [1] \right\} \Rightarrow \mathcal{S}_{\text{CS}} = \{2 \times 2, 1 \times 2, 1 \times 1\}$ while $\{3 \times 3, 2 \times 3, 1 \times 3\} \notin \mathcal{S}_{\text{CS}}$ for this particular case of \mathbf{T}_x . As for the US approach, all possible combinations are feasible. From our proposed algorithms, what clearly falls out is that they offer an inherent seamless mode switching capability between CoMP and single cell 1×1 scenario. When N users are selected to be served from L BSs, they are expressed as $N \times L$. Table 6.1 summarizes the possible combinations of the various user scheduling strategies described above. In all the scheduling strategies, it should be noted that the UEs that are not currently being served can be expected to be served in another resource, thereby achieving user fairness.

6.4. Performance Evaluation

Consider the cluster center where M single antenna UEs moving at 3 kmph are dropped as shown in Figure 6.1. The radius of the cell is $R = 500$ m. These UEs are uniformly dropped in an ellipsoid in \mathbb{R}^2 , whose center is the cluster center. The major and minor axis of the ellipsoid are $(2\Delta x, 2\Delta y)$ where $0 \leq \Delta x \leq \frac{R}{16}$,

Table 6.1.: Summary of the scheduling approaches with $K = 3$

	BD	US	CS
Feasible Set [†] , $\mathcal{S}_{\text{algo}}$	$\{3 \times 3\}$	$\{3 \times 3, 2 \times 3, 2 \times 2, 1 \times 3, 1 \times 2, 1 \times 1\}$	$\{3 \times 3, 2 \times 3, 2 \times 2, 1 \times 3, 1 \times 2, 1 \times 1\}$
Search	Exhaustive	Exhaustive	Exhaustive
Cardinality	$ \mathcal{S}_{\text{BD}} < \mathcal{S}_{\text{CS}} $	$ \mathcal{S}_{\text{US}} $	$ \mathcal{S}_{\text{CS}} \leq \mathcal{S}_{\text{US}} $
Zeros [‡]	Allowed	Allowed	Not Allowed
Interference	Removed	Partially	Removed

[†] The subscript “algo” refers to BD or US or CS.

[‡] The zeros in the aggregated channel matrix, $\widetilde{\mathbf{H}}$, formed at the CCN.

$0 \leq \Delta y \leq \frac{h/2}{16}$ and h is the height of the hexagon or cluster. $K = 3$ single antenna BSs are positioned as shown in Figure 6.1. A realistic WINNER II channel model [9] corresponding to scenario B1: urban micro-cell, non-line of sight with pathloss and shadow fading is considered with 500 independent channel realizations at 2 GHz center frequency. The signal to noise ratio at the cell-edge (reference value for one user located at the cell-edge) is fixed at 15 dB. For the B1 scenario, the channel provided by the WINNER II model is converted to the frequency domain with a 256-fast Fourier transform, where 32 consecutive subcarriers correspond to one resource for simplicity. The feedback load reduction is performed for one such resource, \mathbf{T}_x , where x takes the values 0, 5 and 40 dB. The results presented are averaged over the Monte Carlo simulations over all the resources.

Figure 6.2 shows the average sum rate of the various scheduling algorithms considered in this work. As expected for lower thresholds, the unconstrained scheduler, US, performs better than the constrained scheduler, CS. However, this is achieved at the expense of the backhaul. This is due to the smaller sets of \mathcal{L} and \mathcal{N} being formed with CS unlike US. For CS and US, the sum rate increases with the increase in the feedback threshold. However in the case of BD, for lower number of UEs, the 0 dB threshold outperforms the 5 dB. This is related to Figure 6.3 where the original BD algorithm is unable to find a block diagonal structure. The BD 0 dB case has a better chance of finding an identity matrix that results in block diagonal structure than the BD 5 dB, as the 0 dB thresholding maps to the UEs feeding back at least the strongest BSs, while the 5 dB allows the UE to feed back more BSs, thereby making it hard to find a block diagonal structure. However, the situation improves when the number of UEs increase and the scheduler is able to find a block diagonal structure. Theoretically, an ∞ dB threshold is the case when CS, US and the generalized BD scheduling algorithm converge to the same solution.

The BD algorithm performs reasonably well in terms of sum rate, however, it is not feasible when the number of UEs are small with lower threshold. Figure 6.3 captures this in terms of the probability of failure to find a block diagonal subset of UEs, P_f , such that $N_{\text{max}} = L_{\text{max}} = K = 3$, for a given threshold for feedback

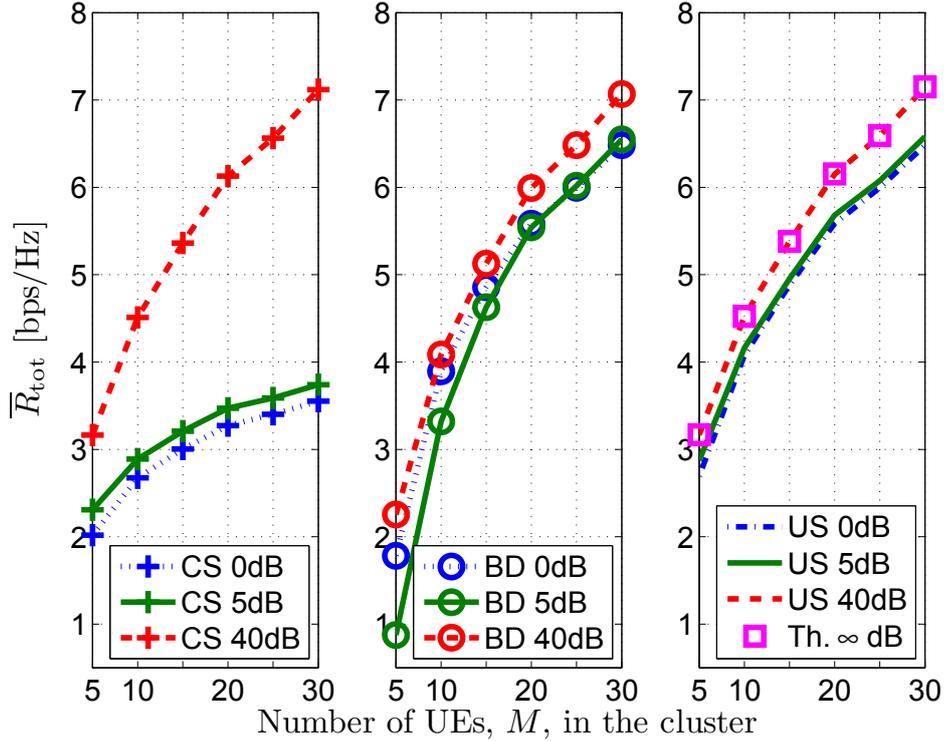


Figure 6.2.: Average sum rate versus the increase in the number of UEs.

load reduction. This failure maps to Algorithm 6.1, step 8. P_f goes to zero when the number of UEs exceeds 25 for all the thresholds considered in this work. With small number of UEs, the failure is due to the relative thresholding. The ratio of the number of unsuccessful attempts to the total number of attempts to find a block diagonal structure when performing the exhaustive search is 77.8%, 94.9%, and 0.6% for 0 dB, 5 dB, and 40 dB, respectively. These values do not change with the increase in the number of UEs. Let us consider the number of UEs to be 30. This translates to the total number of attempts being $\binom{30}{3} = \frac{30!}{3!27!} = 4060$. With a feedback load reduction threshold of 5 dB, corresponding to the cluster center [10, Fig. 4.19], the BD approach cannot be evaluated for potentially 94.9% of the time. With a threshold of 40 dB, the failure to find a block diagonal structure is as low as 0.6%, this is due to the fact that a bigger threshold allows the UE to feed back the CSI from more BSs, causing $\tilde{\mathbf{H}}$ to be a full matrix more often than not. Hence, the BD procedure can easily be applied to a 40 dB threshold. However, it should be noted that this failure can be avoided if the BD approach in [4] is generalized, such that the subsets $\{2 \times 2, 1 \times 1\}$ are also included. This is treated as part of our future work.

We define the average feedback load reduction, \bar{f}_{LR} as the average of the number of zeros in a sparse aggregated channel matrix $\tilde{\mathbf{H}} \in \mathbb{C}^{M \times K}$ i.e., the cardinality of set $\mathcal{S}_{\text{FB}} = \{\tilde{\mathbf{H}}(i, j) = 0, \forall i, j \in \mathbb{N}^+, i \leq M, j \leq K\}$. The average feedback load

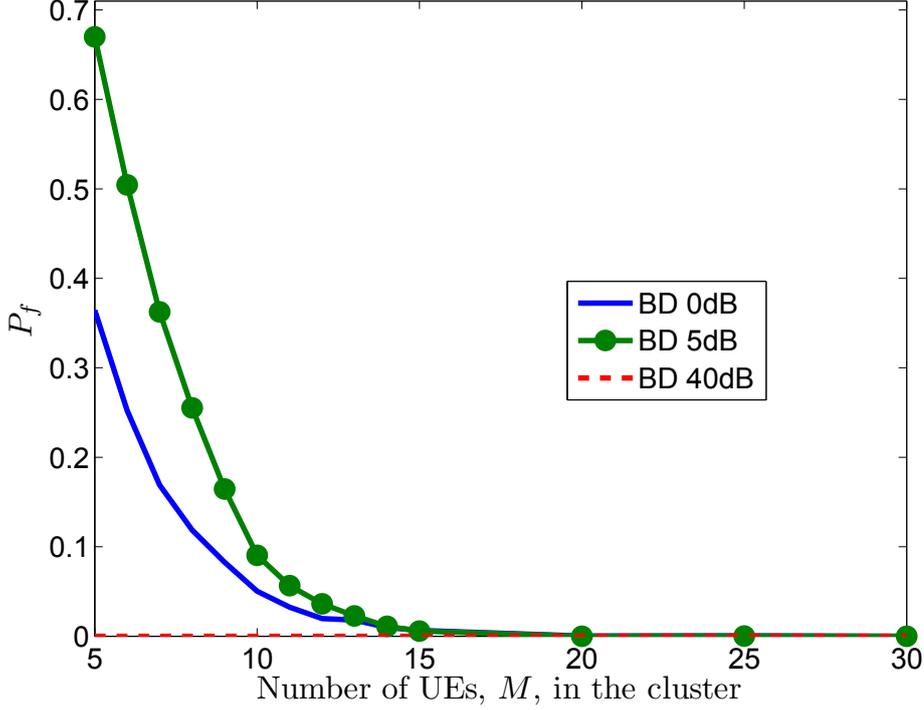


Figure 6.3.: Probability of failure to find a block diagonal subset, P_f

reduction is calculated as

$$\bar{f}_{\text{LR}} = \mathbf{E}_{\tilde{\mathbf{H}}} \{ |\mathcal{S}_{\text{FB}}| \}. \quad (6.9)$$

Figure 6.4 shows the \bar{f}_{LR} due to various thresholds that were applied to all the scheduling algorithms considered in this work. As more number of UEs feed back the CSI, the same needs to be available at the CCN. The savings in the feedback load is linear whose slope decreases with increasing threshold. As expected, the feedback load reduction with threshold of 40 dB and ∞ dB has poor savings.

Now we discuss the impact on the backhaul due to the precoding weights. We define the normalized average backhaul load reduction as the relative difference in the total number of UEs and BSs to the cardinality of the set, \mathcal{S}_{BH} consisting of non-zeros in the precoded matrix, $\widetilde{\mathbf{W}}_{\text{algo}} \in \mathbb{C}^{L \times N}$, i.e.,

$\mathcal{S}_{\text{BH}} = \left\{ \widetilde{\mathbf{W}}_{\text{algo}}(j, i) \neq 0, \forall i, j \in \mathbb{N}^+, i \leq N, j \leq L \right\}$. The normalized average backhaul load reduction is calculated as

$$\bar{b}_{\text{LR}} = \frac{(L_{\text{max}} N_{\text{max}}) - \mathbf{E}_{\tilde{\mathbf{H}}} \{ |\mathcal{S}_{\text{BH}}| \}}{L_{\text{max}} N_{\text{max}}}, \quad (6.10)$$

where $L_{\text{max}} = K = 3$ and $N_{\text{max}} = L_{\text{max}}$ as the maximum number of UEs served is limited by the maximum number of BSs selected. This is captured in Figure 6.5 for the various scheduling algorithms considered in this work. The CS approach has

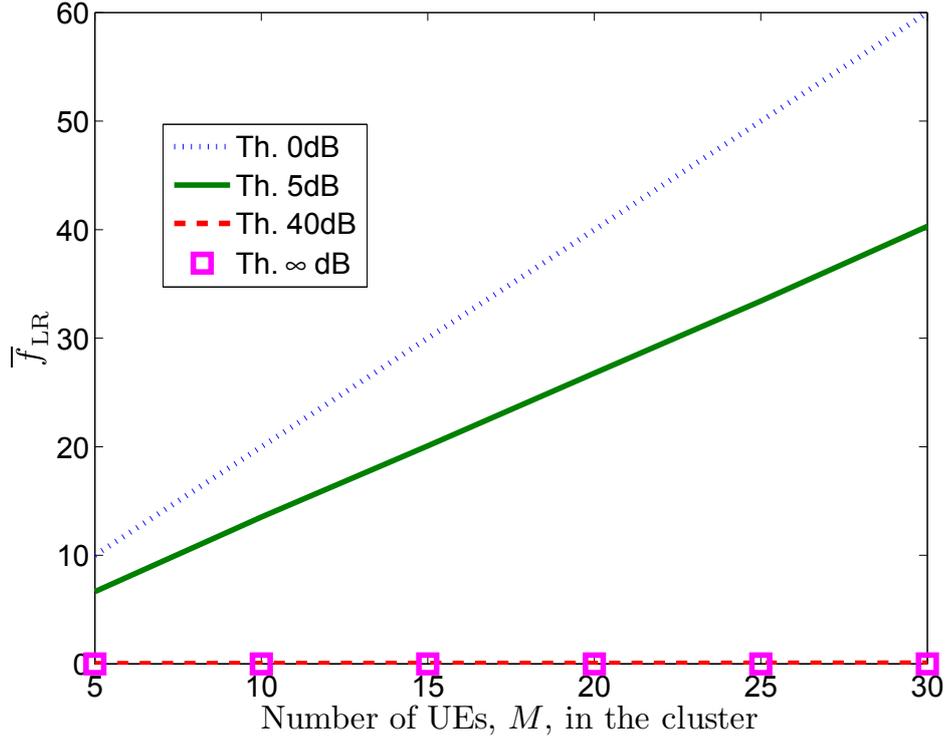


Figure 6.4.: Average feedback load reduction, \bar{f}_{LR} , achieved via PJP

nearly 90% backhaul savings with the smallest feedback load reduction threshold, and the savings diminish as the threshold increases. As the number of UEs grows, it is interesting to note that with CS 40 dB and US 40 dB, the savings are nearly similar, with both undergoing an exponential decay. An ∞ dB threshold also shows this decay, resulting in savings in the backhaul. This is due to the fact that the scheduler is capable of finding a smaller set of BSs and UEs that can achieve a better sum rate. With smaller thresholds, the CS and US both tend to have higher savings in the backhaul. There is no backhaul savings when the feedback threshold is 40 dB in the case of BD, as the aggregated channel matrix is full. The BD 5 dB has better savings in the backhaul compared to BD 0 dB when M is small. This is due to the failure to find a block diagonal structure that results in the savings in the backhaul as observed in Figure 6.3.

The metric average sum rate per backhaul use is considered, as the user data will be routed at the CCN based on the non-zero precoding weight. This will dominate the backhaul compared to the CSI feedback [5, Fig. 1]. Hence, the average sum rate per backhaul use is calculated as

$$\bar{R}_{\text{tot}}^{\text{per backhaul use}} = \frac{\bar{R}_{\text{tot}}}{(1 - \bar{b}_{LR}) L_{\text{max}} N_{\text{max}}} = \frac{\bar{R}_{\text{tot}}}{\mathbf{E}_{\tilde{\mathbf{H}}} \{|\mathcal{S}_{\text{BH}}|\}}, \quad (6.11)$$

and Figure 6.6 captures this metric. It can be observed that our proposed CS

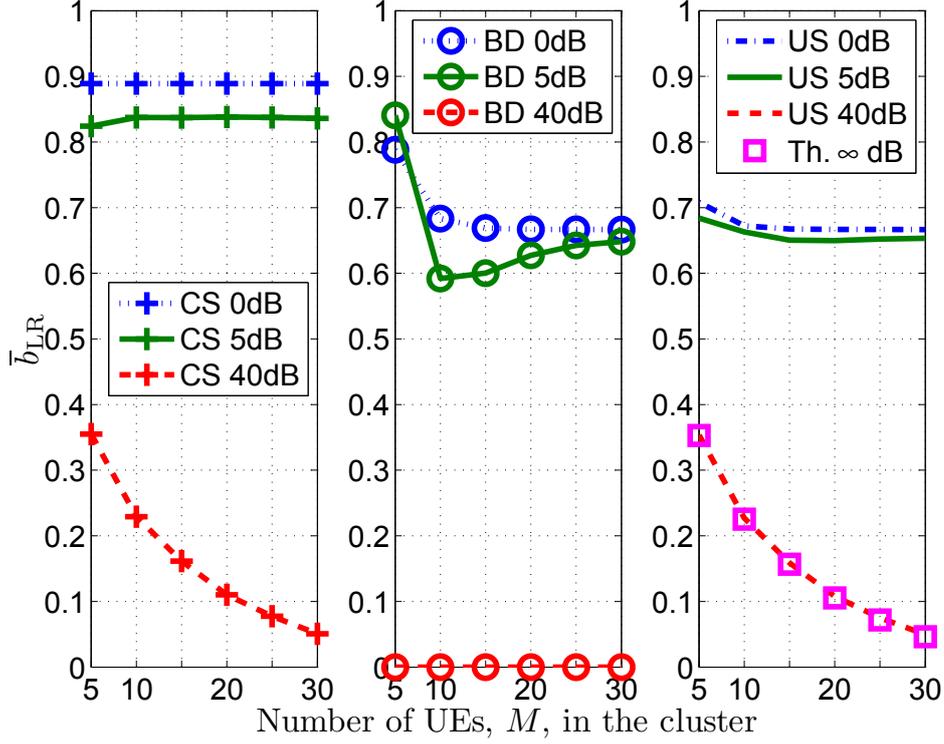


Figure 6.5.: Average normalized backhaul load reduction, \bar{b}_{LR}

algorithm performs the best compared to all the other algorithms, providing the best sum rate per backhaul use.

The limitation of the proposed approaches is that they need to perform an exhaustive search to find the best possible set of BSs and UEs that gives the best sum rate. However, a greedy based user selection can be easily implemented based on the proposed algorithm in order to reduce the complexity [11]. It should be noted that the algorithms presented in this paper are independent of the scheduling criteria (7).

6.5. Conclusion

In this work, we explore scheduling techniques that can efficiently use the backhaul for distributing the precoding weights (from CCN to corresponding BSs) under feedback load reduction achieved via partial joint processing for coordinated multipoint transmission. We proposed the constrained and unconstrained scheduling schemes, comparing them to the state of the art MAC layer block diagonalization technique for backhaul load reduction. The constrained scheduling achieves the best tradeoff in terms of the sum rate per backhaul use. The block diagonalization technique performs well in terms of the sum rate when the number of users is large, however, they fail to find a block diagonal structure when the number of users are small.

As part of our future work, combining the constrained scheduling and the block diagonalization technique can harness the gains of both these approaches and over-

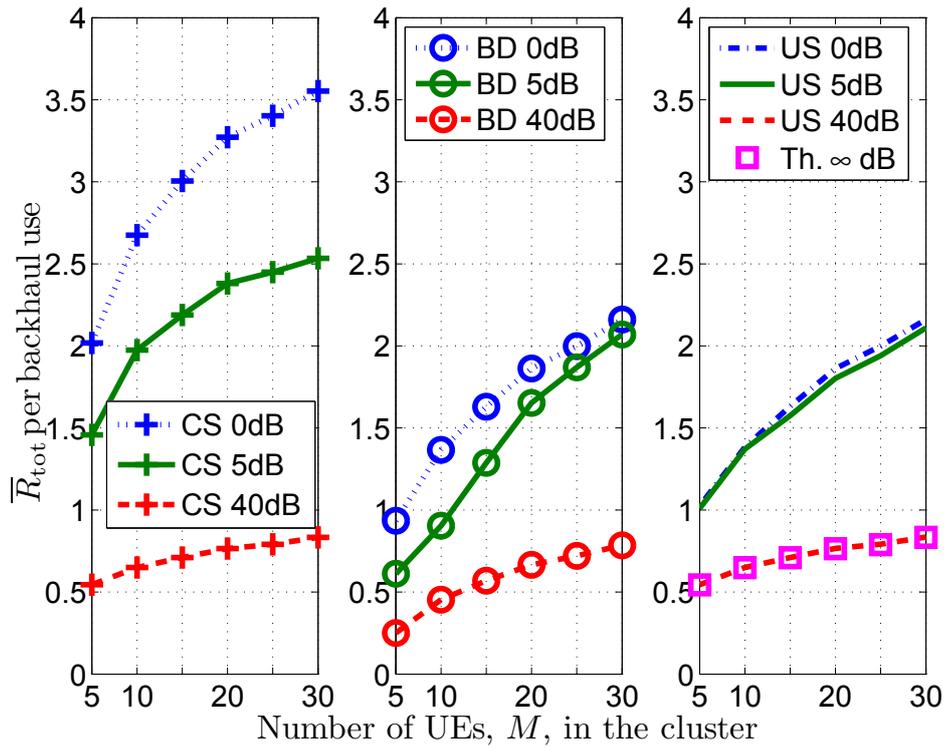


Figure 6.6.: The average sum rate per backhaul use

come their limitations simultaneously. This combined technique can achieve a better tradeoff between the sum rate and backhaul use. Also, generalizing the block diagonalization technique, such that $N = L \leq K$ can improve these preliminary results.

Bibliography

- [1] P. Marsch and G. Fettweis, *Coordinated Multi-Point in Mobile Communications: from theory to practice*, Cambridge University Press, 2011.
- [2] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and Wei Yu, "Multi-Cell MIMO Cooperative Networks: A New Look at Interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.
- [3] C. Botella, T. Svensson, X. Xu, and H. Zhang, "On the performance of joint processing schemes over the cluster area," in *Proc. IEEE Veh. Technol. Conf.*, pp. 1-5, May 2010.
- [4] A. Papadogiannis, H.J. Bang, D. Gesbert, and E. Hardouin, "Efficient Selective Feedback Design for Multicell Cooperative Networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, pp. 196-205, Jan. 2011.
- [5] T.R. Lakshmana, C. Botella, and T. Svensson, "Partial Joint Processing with Efficient Backhauling using Particle Swarm Optimization," *EURASIP J. Wireless Commun. and Netw.*, vol. 2012, 2012.
- [6] B. Makki and T. Eriksson, "On Hybrid ARQ and Quantized CSI Feedback Schemes in Quasi-Static Fading Channels," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 986-997, Apr. 2012.
- [7] T.R. Lakshmana, C. Botella, T. Svensson, X. Xu, J. Li and X. Chen, "Partial Joint Processing for Frequency Selective Channels," in *Proc. IEEE VTC Fall*, Sept. 2010.
- [8] H. Zhang and H. Dai, "Cochannel Interference Mitigation and Cooperative Processing in Downlink Multicell Multiuser MIMO Networks," *EURASIP J. Wireless Commun. and Netw.*, vol. 2004, no. 2, pp. 222-235, 2004.
- [9] P. Kyösti, J. Meinilä, et al., "D1.1.2 WINNER II channel models: Part I channel models," IST-4-027756 WINNER II, September 2007.
- [10] ARTIST4G D1.2, Innovative advanced signal processing algorithms for interference avoidance, *ARTIST4G technical deliverable*, <https://ict-artist4g.eu/projet/work-packages/wp1/documents/d1.2/d1.2.pdf>, pp. 84. Accessed 17 Aug. 2012.

- [11] J. Li, T. Svensson, C. Botella, T. Eriksson, X. Xu, and X. Chen, “Joint Scheduling and Power Control in Coordinated Multi-Point Clusters”, in *Proc. IEEE VTC Fall*, Sept 2011.

Paper D

On the Potential of Broadcast CSI for Opportunistic Coordinated Multi-Point Transmission

Tilak Rajesh Lakshmana, Agisilaos Papadogiannis, Jingya Li, and Tommy Svensson

Published in
*IEEE International Symposium on Personal, Indoor and Mobile Radio
Communications (PIMRC)*, Sep. 2012

©2012 IEEE

The layout has been revised.

7. On the Potential of Broadcast CSI for Opportunistic Coordinated Multi-Point Transmission

Abstract

Coordinated Multi-Point transmission is a promising technique to improve the performance of the users at the cell-edge. To achieve this, in case of a centralized approach, users need to unicast the quantized channel state information (CSI), typically to the anchor base station (BS), and then each BS forwards this information to a central coordination node for precoding and scheduling. In the case of a decentralized approach, users broadcast the quantized CSI such that the coordinating BSs could simultaneously receive the CSI. The advantage of a decentralized approach is that it does not require a central coordination node, thereby not imposing stringent latency constraints on the backhaul. The CSI transmission over the erroneous feedback channel in the uplink gives rise to precoding loss and scheduling loss. In the decentralized framework, the feedback errors could result in BSs receiving a different version of the CSI. In this work, we propose a decentralized opportunistic scheduling approach, which only requires a minimal sharing of scheduling information between BSs. The results show that the sum rate achieved with the proposed method is comparable to that of the centralized approach even when there is a high bit error probability introduced by the feedback channel. We also show that when the bit error probabilities in the feedback channel are less than 10^{-4} , the decentralized approach achieves the sum rate of the centralized approach.

Keywords—Broadcast CSI, CoMP, Decentralized Architecture, Scheduling

7.1. Introduction

In cellular systems, Coordinated Multi-Point (CoMP) transmission is a promising technique to improve the user experience, especially at the cell-edge as the user throughput is limited primarily due to intercell interference [1]-[3]. To harvest these gains in a frequency division duplex system, the User Equipments (UEs) need to feedback the Channel State Information (CSI) to their anchor Base Stations (BSs).

We define anchor BS for a specific UE as the BS that provides the best average channel gain.

This information is then forwarded to a Central Coordination Node (CCN) where user scheduling and data transmission are designed. This approach is called *Centralized Joint Processing* (CJP) with unicast CSI [2, 4]. Figure 7.1 shows the centralized architecture. The main drawback of the centralized architecture is the backhaul latency introduced due to the forwarding of CSI and precoding weights, to and from the CCN.

To avoid the stringent latency constraints in the backhaul, a novel feedback approach is proposed in [4, 5] where the CSI is broadcasted by the UE to the coordinated BSs. We define this approach as *Decentralized Joint Processing* (DJP) with broadcast CSI [4]. Figure 7.2 captures the decentralized architecture (without any information exchange between BSs). In [6], it is shown that the CSI distribution over the air (without backhaul) outperforms the CSI distribution over the backhaul in terms of the user rate even with backhaul latency of 4 ms. One of the main benefits of a DJP approach is that it does not need the backhaul for CSI exchange. In a real system, the performance of Joint Processing (JP) is severely affected by errors in the CSI feedback due to quantization and delays in the backhaul. With CJP, the CSI needs to undergo two hops to reach the CCN via its anchor BS, for every UE involved in JP. Similarly, the precoding weights need to traverse the path of CSI in the backhaul, taking one hop from the CCN to the corresponding BS. With DJP, the backhaul is not used, and hence, it reduces the total latency related to the CSI and the precoding weights per scheduled UE by two hops.

The CSI that needs to be fed back always suffers from quantization loss [7]. The errors in the CSI due to the feedback channel gives rise to precoding loss and scheduling loss. Quantization, precoding and scheduling play an important role in harnessing the gains of CoMP. Building on the ideas based on [5, 6] as described above; in this work, we propose an opportunistic scheduling (UEs that result in the best sum rate are served) and sharing minimal scheduling information between BSs for a decentralized architecture where the UEs broadcast the CSI. This is shown in Figure 7.2. We show that minimal exchange of scheduling information between the coordinating BSs following DJP with a broadcast CSI approach can realize the gains of CJP with unicast CSI. These gains are valid for a range of bit error probabilities experienced by the collaborating BSs. Unlike [8], where the CSI delay is considered based on the feedback rate, and each BS broadcasts the selected user index and the CSI to other BSs, under equal power allocation. To position our work in comparison to [8], we consider the quantization loss in the CSI feedback and a BS shares only the scheduled user index to the BSs involved in JP under optimal power allocation. The iterative broadcasting of CSI from each BS to other BSs are avoided in our proposal.

The paper is organized as follows: section 7.2 discusses the signal and system model. The proposed decentralized opportunistic CoMP and various network architectures are presented in section 7.3. In section 7.4, the potential gains and open issues of the network architectures are discussed. Finally, the main conclusions of this work are summarized in section 7.5.

Here unicasting refers to the transmission of the CSI from a UE to an anchor (single) BS.

The following notation is used in this paper: boldface upper-case letters denote matrices, \mathbf{X} , boldface lower-case letters denote vectors, \mathbf{x} , and italics denote scalars, x . The absolute value of the elements in a vector \mathbf{x} is denoted as $|\mathbf{x}|$. The $\mathbb{C}^{m \times n}$ is a complex valued matrix of size $m \times n$. The $(\cdot)^T$ and $(\cdot)^H$ is the transpose and the conjugate transpose, respectively. $\mathbf{E}_x \{\cdot\}$ is the expectation with respect to x . The $\|\mathbf{x}\|_2$ is the 2-norm of \mathbf{x} . $\mathbf{X}(i, j)$ is the (i, j) th element of matrix \mathbf{X} . The i th row of a matrix \mathbf{X} is $\mathbf{X}(i, :)$. The sets are indicated in calligraphic letters and $|\mathcal{X}|$ denotes the cardinality of the set \mathcal{X} . The $\langle \mathbf{x}, \mathbf{y} \rangle$ represents the inner product between \mathbf{x} and \mathbf{y} . The operator \otimes is the modulo-2 addition.

7.2. Signal and System Model

Consider K single antenna BSs that need to serve $M = |\mathcal{M}|$ single antenna UEs, where \mathcal{M} is the set of all the active UEs requiring service. In this regard, two different architectures are considered. They are centralized and decentralized architectures. \mathcal{U} is the set of UEs selected for JP such that $\mathcal{U} \subseteq \mathcal{M}$ and $|\mathcal{U}| \leq K$, so that orthogonality can be maintained under a linear precoding assumption [9]. These UEs need to feed back the quantized CSI. For simplicity, the channel norm, g_m , is assumed not to be corrupted by errors and it is available at the BS from the m th UE as

$$g_m = \|\mathbf{h}_m\|_2, \quad (7.1)$$

where $\mathbf{h}_m = [h_{m,1}, h_{m,2}, \dots, h_{m,K}]$ is the CSI of the links from the K BSs to the m th UE. In other words, g_m is well protected with suitable channel coding, and being a scalar the overhead of feeding back g_m can be considered negligible compared to the CSI. The discrete time signal received at $|\mathcal{U}|$ UEs, $\mathbf{y} \in \mathbb{C}^{|\mathcal{U}| \times 1}$ is

$$\mathbf{y} = \mathbf{H}\overline{\mathbf{W}}\mathbf{x} + \mathbf{n}. \quad (7.2)$$

In a centralized approach, the channel matrix serving $|\mathcal{U}|$ UEs is $\mathbf{H} \in \mathbb{C}^{|\mathcal{U}| \times K}$, $\overline{\mathbf{W}} \in \mathbb{C}^{K \times |\mathcal{U}|}$ is the precoding matrix and \mathbf{n} is the receiver noise at the UEs, which are spatially and temporally white with variance σ^2 . Random vector quantization [10]-[12] is used to quantize the direction of the CSI after normalizing it with the channel norm such that the generated codebook vectors are on a unit sphere and is represented as $\tilde{\mathbf{h}}_m$ for the normalized CSI from the m th UE. This approach simplifies the codebook, $\mathbf{B} \in \mathbb{C}^{2^N \times K}$, required at the UEs and the BSs, where N is the number of bits required to represent the quantized CSI. Random vector quantization mainly aligns the channel vector to that of the codebook and can be summarized as follows:

$$\tilde{\mathbf{h}}_m = \mathbf{h}_m / g_m \quad (7.3)$$

$$b' = \arg \max_b |\langle \mathbf{B}(b, :), \tilde{\mathbf{h}}_m \rangle|, \quad (7.4)$$

where $\mathbf{B}(b, :) \in \mathbb{C}^{1 \times K}$ such that the elements of $\mathbf{B}(b, :)$ are iid circularly symmetric complex Gaussian distributed as $\mathcal{CN}(0, 1)$, b' is the codebook index which can be represented as a vector, \mathbf{v} , of length N bits and is fed back by the m th UE. These bits are independently flipped with a probability depending on the bit error probability, P_e , of the feedback channel. An error is declared on the n th bit of \mathbf{v} as

$$e = \begin{cases} 1, & \text{if } P_e > r, \\ 0, & \text{otherwise no error} \end{cases} \quad (7.5)$$

$$\hat{\mathbf{v}}(n) = \mathbf{v}(n) \otimes e, \quad (7.6)$$

where r is a random number chosen from a uniform distribution in the interval $[0, 1]$. The BSs perform the reverse processing in extracting the CSI of the m th UE as $\hat{\mathbf{h}}_m$. The errors in feedback channel affect the CSI feedback vector \mathbf{h}_m to have a different value from what was transmitted. This is due to the decoding of the codebook vector based on an incorrect codebook index. If the decoded codebook index is \hat{b}' , then the decoded CSI for the m th UE can be written as

$$\hat{\mathbf{h}}_m = g_m \mathbf{B}(\hat{b}', :) \quad (7.7)$$

When the feedback channel is error free, the quantization error is the difference between $\hat{\mathbf{h}}_m$ and \mathbf{h}_m . It should be noted that in this work, we do not aim to optimize quantization.

As the main focus of this study is on the network architecture, a linear zero forcing beamformer (BF) is considered in this work. The BF is calculated as

$$\mathbf{W} = \hat{\mathbf{H}}^H (\hat{\mathbf{H}} \hat{\mathbf{H}}^H)^{-1}, \quad (7.8)$$

which is the Moore-Penrose pseudoinverse and $\hat{\mathbf{H}} \in \mathbb{C}^{|\mathcal{U}| \times K}$ is the estimated CSI available for beamforming. The optimization problem that jointly optimizes the user scheduling and power allocation is formulated as

$$\text{maximize } \left\{ \sum_{m \in \mathcal{U}} \log_2 \left(1 + \frac{p_m}{\sigma^2} \right) \right\} \quad (7.9)$$

subject to

$$\begin{aligned} |\mathbf{W}(k, :)|^2 \mathbf{p} &\preceq P_{\max} \mathbf{1}_{|\mathcal{U}| \times 1} \\ \mathbf{p} &\succeq \mathbf{0}_{|\mathcal{U}| \times 1} \end{aligned}$$

where $\mathbf{p} = [p_1, \dots, p_{|\mathcal{U}|}]^T \in \mathbb{R}^{|\mathcal{U}| \times 1}$ is the power transmitted to the selected $|\mathcal{U}|$ UEs. Note that for each fixed user set, the optimization problem is a convex problem and the optimal solution can be obtained in a water filling fashion [13]. Finally, the power allocated to each UE is absorbed into the BF giving a precoding vector to the m th UE as

$$\bar{\mathbf{W}}(:, m) = \mathbf{W}(:, m) \sqrt{p_m} \quad (7.10)$$

The Signal to Interference plus Noise Ratio (SINR) for the m th UE is given as

$$\text{SINR}_m = \frac{\|\mathbf{h}_m \overline{\mathbf{W}}(:, m)\|^2}{\sum_{\substack{j=1, j \neq m \\ j, m \in \mathcal{U}}}^{|\mathcal{U}|} \|\mathbf{h}_m \overline{\mathbf{W}}(:, j)\|^2 + \sigma^2}, \quad (7.11)$$

The average sum rate per cell in bps/Hz/cell for scheduling $|\mathcal{U}|$ different UEs on the same frequency/time resource is

$$\overline{R}_{tot} = \frac{1}{K} \mathbb{E}_{\mathbf{h}} \left\{ \sum_{m \in \mathcal{U}} \log_2(1 + \text{SINR}_m) \right\}. \quad (7.12)$$

Scheduling

There are two types of scheduling approaches considered in this paper, i.e., random scheduling and opportunistic scheduling. Random scheduling involves arbitrarily choosing \mathcal{U} UEs for JP while the opportunistic scheduling picks \mathcal{U} UEs based on the combination of UEs that produces the best sum rate, i.e.,

$$\mathcal{U}^* = \arg \max_{\mathcal{U}} \sum_{m \in \mathcal{U}} \log_2(1 + \widehat{\text{SINR}}_m) \quad (7.13)$$

$$\widehat{\text{SINR}}_m = \frac{\|\hat{\mathbf{h}}_m \overline{\mathbf{W}}(:, m)\|^2}{\sum_{\substack{j=1, j \neq m \\ j, m \in \mathcal{U}}}^{|\mathcal{U}|} \|\hat{\mathbf{h}}_m \overline{\mathbf{W}}(:, j)\|^2 + \sigma^2}. \quad (7.14)$$

7.3. Decentralized Opportunistic CoMP

In this section, different network architectures are discussed. The potential of using a decentralized opportunistic CoMP transmission is investigated.

7.3.1. Centralized joint processing with Unicast CSI

Each UE feed back or unicasts the CSI to its anchor BS. The UEs quantize the CSI and feed it back to their anchor BS with a bit error probability of P_{e1} . The BSs decode the CSI based on the codebook mapping and then forward this CSI to the CCN. In this setup, the backhaul is assumed to be error free. The CCN performs precoding and sends back the precoding weights to the BSs. This is illustrated in Figure 7.1.

7.3.2. Decentralized Opportunistic CoMP with Broadcast CSI

A decentralized approach aims to avoid the stringent latency constraints on the backhaul unlike the centralized approach. In the decentralized architecture, the UEs

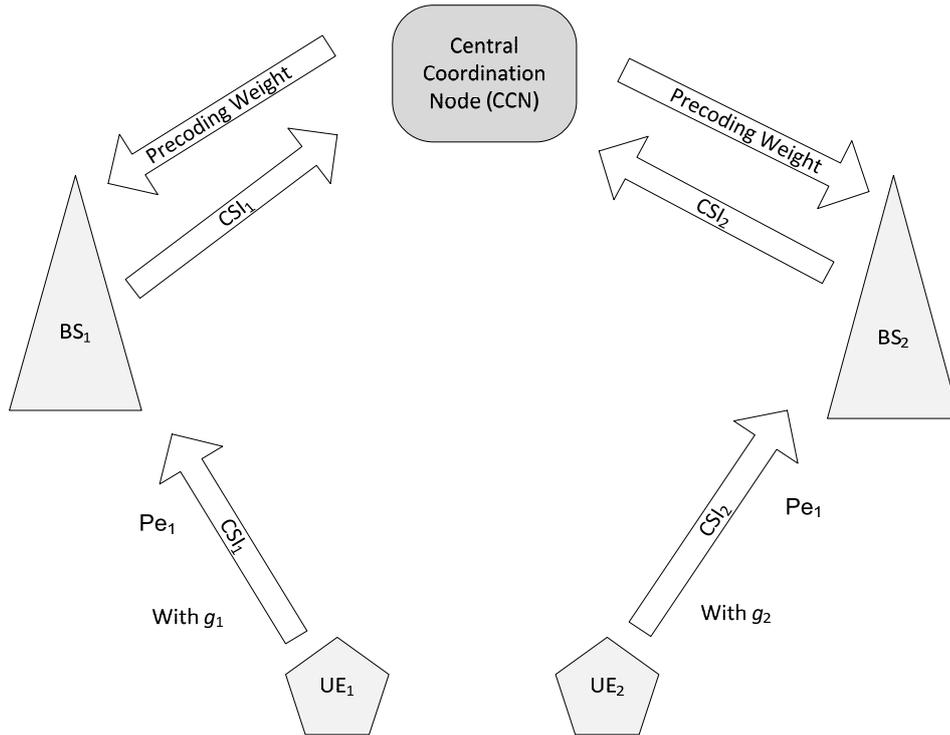


Figure 7.1.: Joint processing Architecture: Centralized. Here g_1 and g_2 are assumed not to be corrupted by errors.

broadcast the CSI such that the BSs receive a version of the CSI that undergoes a different probability of error. For simplicity, considering two BSs, the anchor BS to the UE undergoes P_{e1} and the other BS experiences P_{e2} . This is as shown in Figure 7.2. Each BS could potentially have a different version of the CSI estimated by the UE. This implies that each BS will potentially generate different precoding vectors depending on the UEs being scheduled. In our proposed method, the BS_1 losslessly shares the scheduling information with BS_2 , such that joint transmission is made possible. This implies an extra hop but the amount of this information required to be shared is negligible compared to sharing the CSI. This is illustrated in Figure 7.2. It should be noted that in the CJP approach, the scheduling information is also needed to be passed on from the CCN to the BSs. The scheduling information is merely an index consisting of the UEs being scheduled. BS_1 may decide this based on the scheduling algorithm, e.g., opportunistic scheduling. Then, BS_2 selects the same UEs as those selected by BS_1 . The index values are integers, thus lossless compression can be applied when sharing this information between BSs. It should be noted that the best UEs selected by BS_1 via opportunistic scheduling may not be the best UEs to be served by BS_2 .

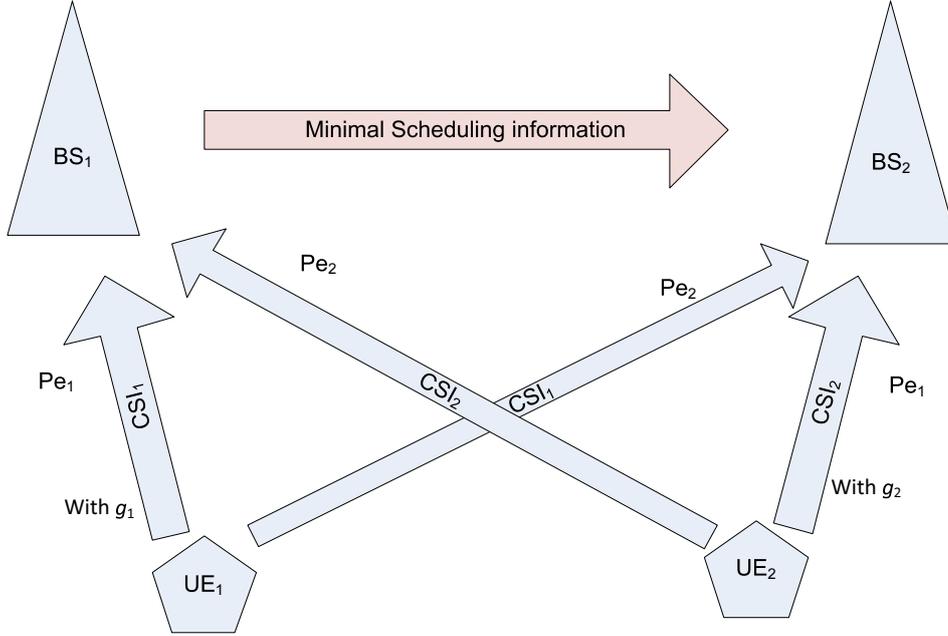


Figure 7.2.: Joint processing Architecture: Decentralized with minimal exchange of scheduling information. Here g_1 and g_2 are assumed not to be corrupted by errors.

7.3.3. Decentralized Joint Processing with Broadcast CSI exploiting diversity

A potential alternative architecture would be a hybrid architecture where the system is decentralized as shown in Figure 7.2 and the CCN is introduced. The broadcasted CSI undergoing different feedback errors are received by different BSs and forwarded to the CCN where the CSI can be coherently combined to exploit diversity. This architecture could be useful when there is high uncertainty in the CSI obtained at the BSs. But, these potential diversity gains could diminish due to the latency involved in the two hops required for the CSI to be available at the CCN. This hybrid architecture combines the CJP and DJP approaches but this causes additional increase in the backhaul traffic as different variations of the same CSI reaches the CCN. Hence, this hybrid architecture is more of an overhead and this architecture does not motivate further study. However, if the backhaul is unconstrained and there is a need for a better quality of the CSI then this hybrid architecture could be considered.

7.4. Performance Evaluation

Consider $K = 2$ single antenna BSs located at the center of two hexagonal cells, as shown in Figure 7.3. The distance separating the BSs is the height of the hexagon,

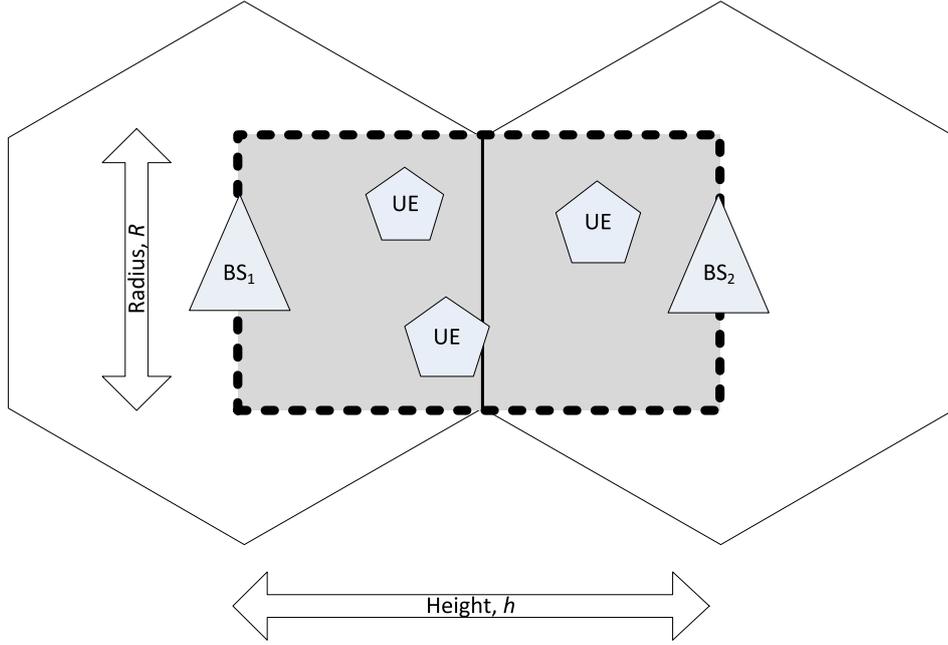


Figure 7.3.: Serving area where the M UEs are dropped.

h . The length of any side of the hexagon is the same as the cell radius, $R = 1$ km. Hence, the rectangular area of concern is h by R , where M UEs are dropped. This is illustrated as the shaded region in Figure 7.3. In every instance \mathcal{U} UEs are scheduled, where $|\mathcal{U}| = 2$. The scheduling is based on the CSI fed back from the M UEs, based on which random or opportunistic scheduling is performed. Initially, the codebook is generated and shared between the UEs and the BSs. The number of bits (or size), $N = 16$, required for the feedback is chosen for a given set of K BSs. This was chosen based on the simulations and [5]. In a real system, the CSI feedback data would be encapsulated as a packet and suitably protected. In these simulations, the size of the packet can be considered to be N and the packet is never discarded even if they contain errors. It is assumed that the CSI feedback is not protected for any error correction. Instead, the UEs are still served based on the erroneous CSI feedback. Also, as the errors are introduced at the bit level, the results are presented in terms of the bit error probabilities instead of block errors. This gives an intuitive feel for the potential benefits with the decentralized architecture when there are errors in the CSI feedback.

The maximum power, P_{\max} , at which the BS can transmit is determined based on cell-edge signal to noise ratio of 15 dB. Water filling based power allocation as formulated in (Equation 7.9) is implemented using CVX [14]. A Rayleigh fading component, Γ , is simulated as a circularly symmetric complex Gaussian random variable as $\mathcal{CN}(0, 1)$. The channel between the k th BS and the m th UE is calculated as

$$\mathbf{h} = \Gamma \sqrt{\gamma_{\text{SF}} \cdot \gamma_{\text{PL}}}, \quad (7.15)$$

where shadow fading is $\gamma_{\text{SF}} \sim \mathcal{CN}(0, 8 \text{ dB})$ and γ_{PL} is based on the 3GPP pathloss model [15] as

$$\gamma_{\text{PL}}(\text{dB}) = 128.1 + 37.6 \log_{10} R. \quad (7.16)$$

In Figure 7.6, the average sum rate as determined by (Equation 7.12) is evaluated for various centralized and decentralized architectures with equal error probabilities. In the legend of Figure 7.6, 1) C: $P_e = 0$; Opport. represents that it is a centralized architecture without any errors in the feedback and opportunistic scheduling is applied. Similarly, 7) involves random scheduling and 5) is the single cell system which does not perform any precoding but only considers the unicast of the channel strength and is affected by interference due to the transmission to the other UE. It should be noted that quantization is not considered for the curves that appear flat in Figure 7.6. This is to provide an upper bound for the corresponding scenarios undergoing bit errors. Scenarios 2) and 8) are the centralized approaches with opportunistic and random scheduling. Their counterparts in the decentralized approach are 3) and 9), where the scheduling information is shared by BS_1 to BS_2 . It can be observed that the centralized and the decentralized curves nearly overlap. The decentralized approach performs marginally below the centralized approach due to the broadcasted CSI undergoing different errors. More importantly, the UEs selected at BS_1 with opportunistic approach might not be appropriate to be opportunistically scheduled at BS_2 . This loss is more at high bit error probabilities. Hence, diversity could be exploited to overcome this loss, as explained in subsection 7.3.3. Finally, scenario 4) shows the typical DJP approach without sharing any scheduling information. In this case, the UEs are scheduled by each BS running the same opportunistic scheduler. For a given bit error rate of 0.01, the DJP (scenario 4) has a loss of 2.94 bps/Hz/cell compared to the CJP approach (scenario 2) while the DJP with sharing the scheduling information (scenario 3) only loses out by 0.56 bps/Hz/cell. Also, when the feedback bit errors are less than 10^{-4} , even sharing the scheduling information can be avoided. The single cell system with perfect CSI is shown in scenario 5) and those with errors are shown in scenario 6). A random scheduler being a dummy approach performs the worst.

Figure 7.5 captures the average sum rate for different bit errors experienced at different BSs, for scenarios 3) and 4), where BSs share and not share the scheduling information, respectively. It captures the scenarios for values of P_{e1} and P_{e2} . It can be observed that if the feedback channel has low bit error probabilities then one can even avoid sharing the scheduling information between BSs. This implies that the performance of the decentralized approach in terms of sum rate would be comparable to the centralized approach. Figure 7.6 shows the effect of the number of UEs on the average sum rate, for a given bit error probability $P_{e1} = P_{e2} = 0.001$. The increase in the average sum rate with the increase in the number of users is due to the multiuser diversity gains, as opportunistic scheduling comes into play. The average sum rate with the decentralized approach without cooperation of BSs is much smaller compared to CJP. While the decentralized approach with sharing the

scheduling information, i.e., 3) D : P_{e1}, P_{e2} ; Opport., $BS_1 \rightarrow BS_2$, always catches up CJP, i.e., 2) C: P_e ; Opport. It should be noted that if errors are considered in the backhaul links, then the gains with CJP would reduce. Hence, the decentralized approach is an attractive alternative. It is interesting to note that the single cell system with no feedback errors performs better than the CJP with feedback errors. This performance loss can be attributed to the precoding loss due to the ZF approach while given a large set of UEs, the single cell system favors those UEs close to the BS.

Table 7.1 summarizes the differences with the architectures considered in this work. The CSI is unicast in case of the centralized approach while it is broadcast in the case of the decentralized approach. The CCN is required in the centralized while this logical entity can be omitted with the decentralized approach. A high capacity backhauling link is needed with the centralized architecture while none is required for signaling the CSI or precoding weights in case of the decentralized architecture. However, the decentralized architecture where minimal scheduling information needs to be shared only requires a low capacity link. The backhaul latency in case of the centralized architecture is two hops, as every UE needs to feedback the CSI to its anchor BS and the same needs to be forwarded to the CCN. These hops are avoided in case of a decentralized architecture. However, sharing scheduling information requires one hop per JP UEs. Scheduling the UEs is decided at the CCN in case of the centralized architecture while in the case of the decentralized architecture sharing the scheduling information can be treated as semi-decentralized due to the extra hop required for sharing. Some performance loss in terms of the average sum rate can be expected with completely decentralized scheduling, as observed in scenario 4) where the BSs do not share the scheduling information. The quantization loss when feeding back the CSI can be expected in the architectures discussed here. However, the centralized architecture has an additional quantization loss due to the BF weights that need to be transported from the CCN to the corresponding BSs.

Table 7.1.: Comparison of centralized and decentralized architectures, with minimal scheduling information exchange

Parameters	Centralized	Decent. w/ Sched.†	Decent.‡
CSI	Unicast	Broadcast	Broadcast
CCN	Req.	Not req.	Not req.
Backhaul (BH)	High	Low	None
BH Latency	2 hops/UE	1 hop/JP UEs	None
Scheduling	Centralized	Semi-decent.	Decent.
Quantization Loss	CSI & BF*	CSI only	CSI only

† Corresponds to scenario 3) D: P_{e1}, P_{e2} ; Opport., $BS_1 \rightarrow BS_2$

‡ Corresponds to scenario 4) D: P_{e1}, P_{e2} ; Opport. (same user by chance)

* The BF weights with quantization loss is not studied here.

Sharing the scheduling information between BSs can happen harmoniously within

the cooperating BSs, i.e., without the need of a master-slave relationship between the cooperating BSs. As the UEs about to be scheduled can be losslessly exchanged, and each BS is at liberty to choose what the other BS has planned to schedule. Alternatively, the BSs receiving the scheduling information can choose the best UEs that should be scheduled given this new information.

7.5. Conclusion

Scheduling is an important function that should be exercised to harness the gains in CoMP. In this work, we proposed a decentralized approach with broadcasting channel state information with opportunistic scheduling and sharing minimal scheduling information between cooperating base stations. The proposed approach yields a sum rate comparable to the centralized joint processing. A purely decentralized approach yields poorer performance when the bit errors in the feedback channel occur with high probability. The main advantage of using a decentralized approach is that the stringent latency constraints on the backhaul imposed by the centralized approach can be circumvented. The decentralized approach should be preferred when the bit error probability in the feedback channel are less than 10^{-4} , otherwise the precoding loss would diminish these gains. Without sharing the scheduling information, the decentralized approach still performs well when the bit error probability is low but the scheduling loss kicks in when the BSs do not cooperate and the bit errors are high. However, when both BSs apply the same opportunistic algorithm, the performance is reasonable under low bit error conditions.

The results in this paper were obtained using single-antenna nodes, and a relevant question is how they generalize for multi-antenna BSs and UEs. The minimal scheduling exchange could potentially be avoided if a given BS can predict what other BSs are about to schedule, based on the location information of the UEs. Then there would be no need to exchange the scheduling information between BSs. Also, a joint opportunistic scheduling could further improve the sum rate. These research items are considered as part of our future work.

Acknowledgment

This work has been supported by the Swedish Agency for Innovation Systems (VINNOVA), within the P36604-1 MAGIC project and the Swedish Research Council VR under the project 621-2009-4555 Dynamic Multipoint Wireless Transmission. The computations were performed on C³SE computing resources. The authors would like to thank the members of the VR project meetings from Uppsala University and Karlstad University for their valuable inputs.

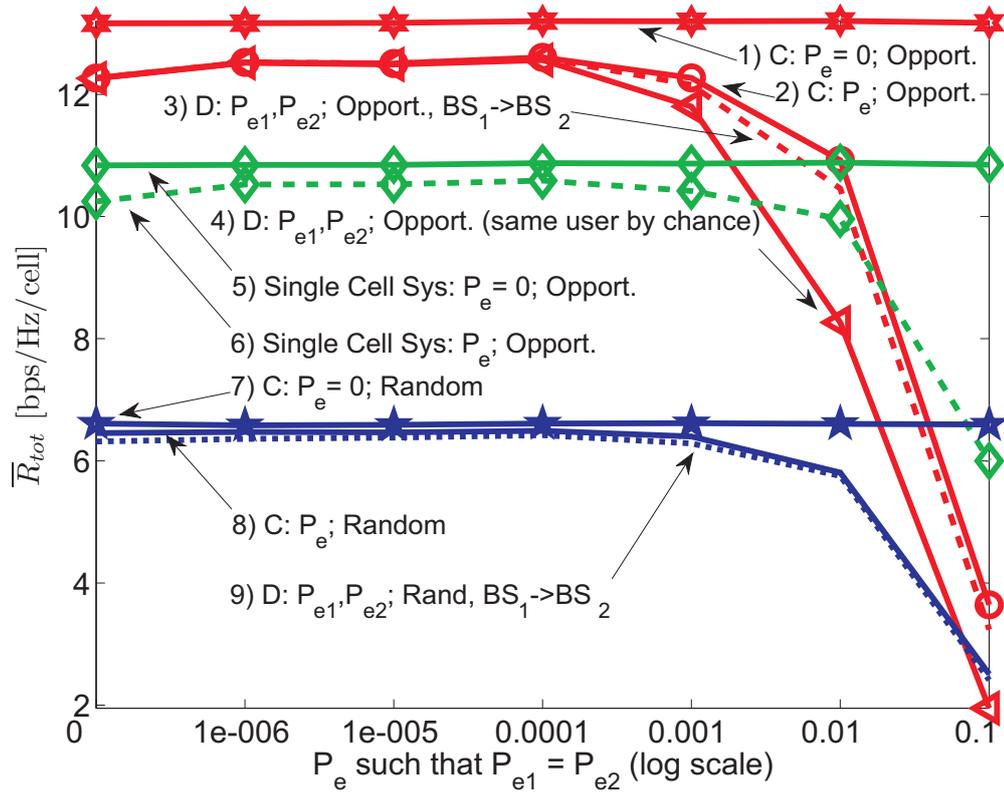


Figure 7.4.: Average sum rate with $P_{e1} = P_{e2}$ simulating a cell-edge scenario for CSI feedback at both BSs, where $M = 10$, and $N = 16$.

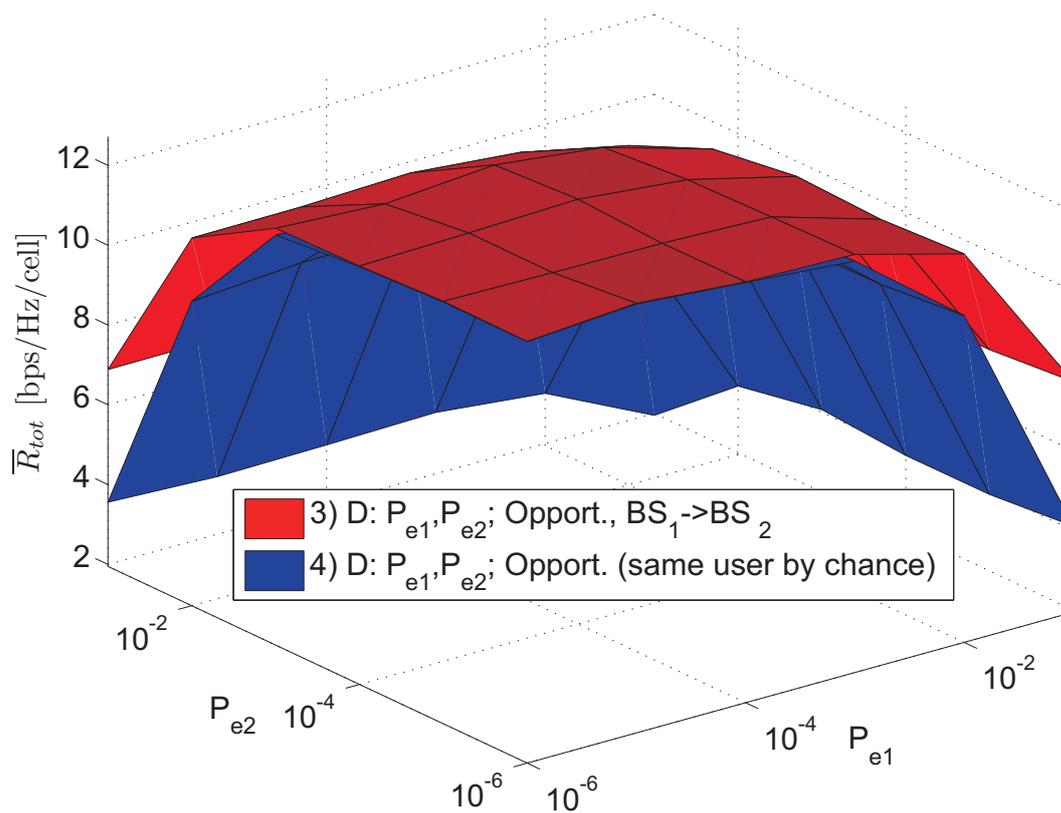


Figure 7.5.: Decentralized opportunistic scheduling where the BS is sharing the scheduling information, and the decentralized scheduling where the BSs do not share any information. The variation in the average sum rate for different error probabilities at both BSs, where $M = 10$, and $N = 16$.

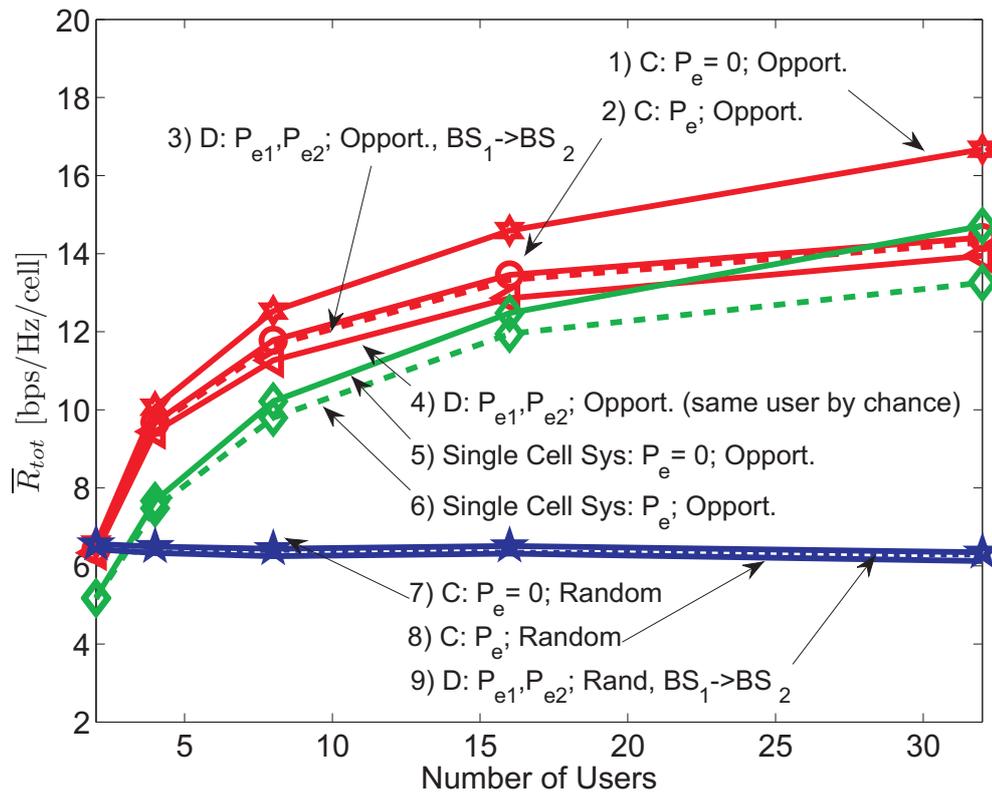


Figure 7.6.: Average sum rate increases logarithmically with the number of UEs with opportunistic scheduling undergoing a bit error probability, $P_{e1} = P_{e2} = 0.001$.

Bibliography

- [1] L. Daewon, S. Hanbyul, B. Clerckx, E. Hardouin, D. Mazzaresse, S. Nagata, K. Sayana, “Coordinated multipoint transmission and reception in LTE-advanced: deployment scenarios and operational challenges,” *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 148–155, Feb. 2012.
- [2] M. Karakayali, G. Foschini, and R. Valenzuela, “Network Coordination for Spectrally Efficient Communications in Cellular Systems,” *IEEE Wireless Commun.*, vol. 13, no. 4, pp. 56–61, 2006.
- [3] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and Wei Yu, “Multi-Cell MIMO Cooperative Networks: A New Look at Interference,” *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.
- [4] A. Papadogiannis, E. Hardouin, and D. Gesbert, “A Framework for Decentralising Multi-Cell Cooperative Processing on the Downlink,” in *Proc. IEEE GLOBECOM Workshops*, Dec. 2008.
- [5] A. Papadogiannis, E. Hardouin, and D. Gesbert, “Decentralising Multicell Cooperative Processing: A Novel Robust Framework,” *EURASIP J. Wireless Commun. and Netw.*, vol. 2009, Article ID 890685, 10 pages, 2009.
- [6] R. Fritzsche and G. Fettweis, “CSI Distribution for Joint Processing in Cooperative Cellular Networks,” in *Proc. IEEE VTC Fall*, Sept. 2011.
- [7] B. Makki and T. Eriksson, “On Hybrid ARQ and Quantized CSI Feedback Schemes in Quasi-Static Fading Channels,” *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 986–997, Apr. 2012.
- [8] S. Han, Q. Zhang, C. Yang, “Distributed coordinated multi-point downlink transmission with over-the-air communication,” *IEEE CHINACOM Conf.*, pp.1-5, Aug. 2010
- [9] Q. Spencer, A. Swindlehurst, and M. Haardt, “Zero-forcing Methods for Downlink Spatial Multiplexing in Multiuser MIMO Channels,” *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.
- [10] C. Au-Yeung and D. J. Love, “On the Performance of Random Vector Quantization Limited Feedback Beamforming in a MISO System,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 458–462, Feb. 2007.

-
- [11] D. J. Love, R. W. Heath Jr., V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An Overview of Limited Feedback in Wireless Communication Systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, 2008.
- [12] N. Jindal, "MIMO Broadcast Channels With Finite-Rate Feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, 2006.
- [13] A. Armada, M. Sanchez-Fernandez, and R. Corvaja, "Waterfilling Schemes for Zero-Forcing Coordinated Base Station Transmission," in *Proc. IEEE Glob. Telecom. Conf.*, Hawaii, USA, 2009.
- [14] M. Grant and S. Boyd, CVX: Matlab Software for Disciplined Convex Programming, version 1.21, <http://cvxr.com/cvx/>, Apr. 2011.
- [15] 3GPP TR 36.942-a20, "3rd Generation Partnership Project, Technical Specification Group Radio Access Network, Evolved Universal Terrestrial Radio Access, Radio Frequency System Scenarios (Release 10)," Jan. 2011.

Part III.
Appendix

A. Max-min SINR Comparison with Weighted Interference Minimization

In this section, PSO is evaluated under another objective function of maximizing the minimum SINR (max-min SINR). This is compared with the weighted interference minimization (WIM) as proposed in Paper B. Recall the objective function of WIM for the i th particle of the PSO can be expressed as

$$f(\mathbf{X}(i, :)) := \frac{\|\text{OffDiag}(\mathbf{H}\mathbf{W})\|_F}{\text{minSINR}_{user}}, \quad (\text{A.1})$$

where \mathbf{H} is the aggregated channel matrix that is available at the CCN, \mathbf{W} is the precoding matrix obtained with PSO and minSINR_{user} is the user with the least SINR. The minimizing objective function for max-min SINR for the i th particle of the PSO can be formulated as

$$f(\mathbf{X}(i, :)) := -\text{minSINR}_{user}. \quad (\text{A.2})$$

In particular, the case of limited CSI feedback being equivalent to the limited backhaul is considered. Figure A.1 shows that max-min SINR has a 2.1% relative increase in average sum rate compared to WIM. However, this is achieved at the expense of 7.7% relative increase in average BSs power consumption as shown in Figure A.2, and also, at the expense of nearly doubling the interference, i.e., a 45% relative increase in actual interference is observed. This is captured in Figure A.3.

To summarize, the WIM objective proposed in Paper B performs nearly as good as the max-min SINR in terms of the sum-rate achieved. However, there are greater benefits with using WIM in terms of the relatively lesser power consumption and lowered interference remaining in the system.

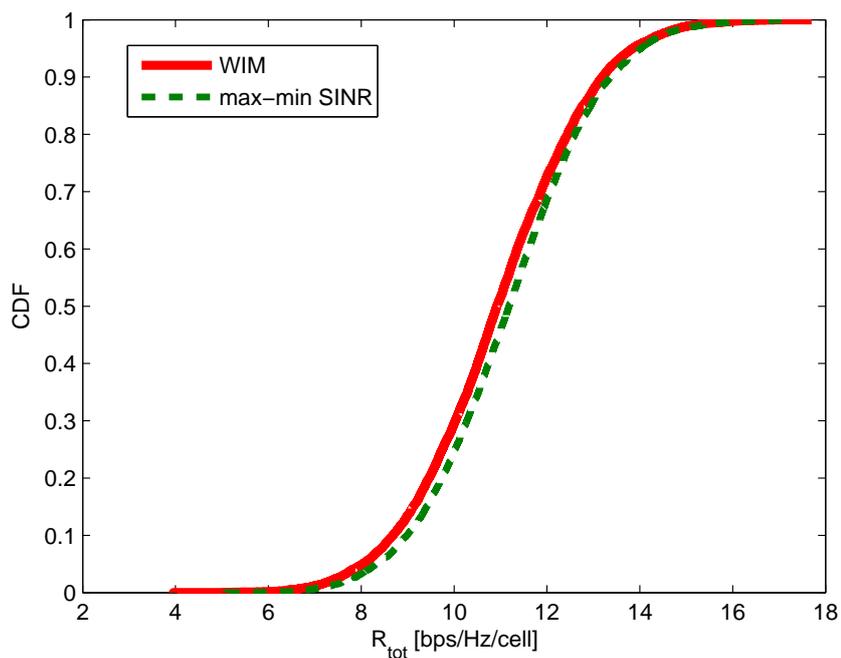


Figure A.1.: CDF of the Sum Rate.

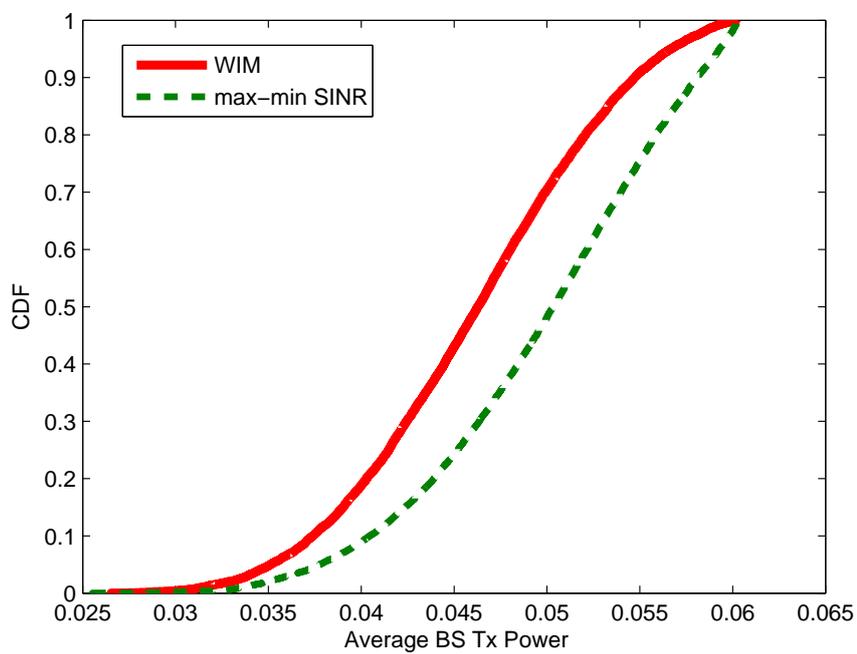


Figure A.2.: CDF of the average BS transmit power.

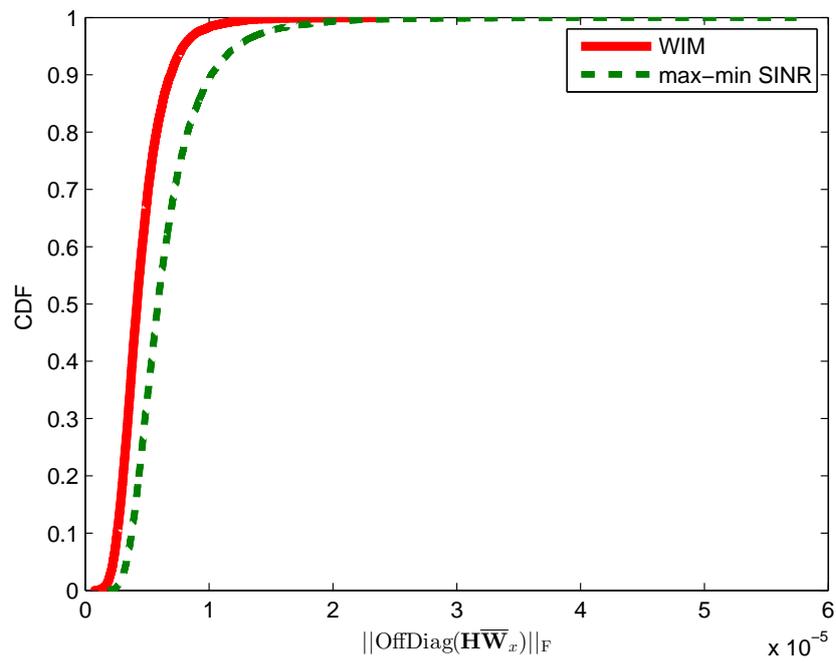


Figure A.3.: CDF of the Actual Interference remaining in the system.

B. Performance of PSO based on Field Measurements Data

In this section, the results from a collaborative evaluation of the precoder design with Uppsala University based on field measurements are presented. The performance of the PSO with perfect and imperfect CSI is also covered.

The channel measurement data was collected by Ericsson Research at Kista a suburb in Stockholm, Sweden. The specific details of these LTE based channel measurements can be found in [ARTD12, Sec. 4.4.1] and the references therein. The results presented here are confined to the most interesting cases compiled from [ARTD14]. In this regard, the robust linear precoder (RLP) based on automatic control robust feed-forward (RLP-ACFF) filter is considered [ASA12]. The RLP-ACFF iteratively minimizes a given objective function and is comparable to PSO with the iterative nature of the algorithm. When UEs are moving fast, obtaining reliable CSI at the transmitter can be very difficult. Hence, as a fallback mechanism, simpler techniques such as coordinated scheduling (CS) and FDMA can be adopted. In the CS, the UE with the strongest instantaneous channel gets the whole bandwidth such that one eNodeB is transmitting while the other is not. Whereas in the case of FDMA, the bandwidth is equally divided between the UEs.

The setup consists of two single antenna eNodeB with two single antenna UEs at a distance of 500m. The UEs are moving at 5 kmph. The evaluation consists of three different scenarios. They are (i) UEs are at the cell-edge, (ii) UEs are at a distance 125 m from their serving eNodeB or the master eNodeB, and (iii) one UE is located at a distance of 125 m from the eNodeB while the other is at the cell-edge. A Wiener filter is used for estimating the channel 8 ms in the future from outdated CSI. The predicted channel is quantized with 5 bits for the phase of the channel and is fed back. The imperfections in CSI are primarily due to prediction errors and quantization errors. A cell-edge signal to noise ratio (SNR) of 15 dB is chosen. A per-eNodeB power constraint is applied in every iteration of the PSO.

Numerical results based on the above three scenarios show that the PSO performs the best. This is captured in Figure B.1-Figure B.3. More importantly, PSO is robust against imperfect CSI. The errors in the channel are dominated by channel prediction errors in the low SNR region while the quantization errors dominate at high SNR region.

This is an important result. The prediction horizon of 8 ms is reasonable considering the feedback loop and the delays especially in the backhaul. PSO is robust against prediction errors based on the Wiener filter, although a Kalman filter would do a better job. One may argue that PSO is computing intensive. However, the

precoder is computed at the CCN in a centralized network architecture or at the eNodeB in a decentralized network architecture. The properties of the PSO are very desirable for precoder design.

A RLP based MMSE solution was also evaluated. However, RLP-MMSE performed poorly and therefore these results are not shown here. More detailed results are presented in the ARTIST4G project in deliverable D1.4 [ARTD14].

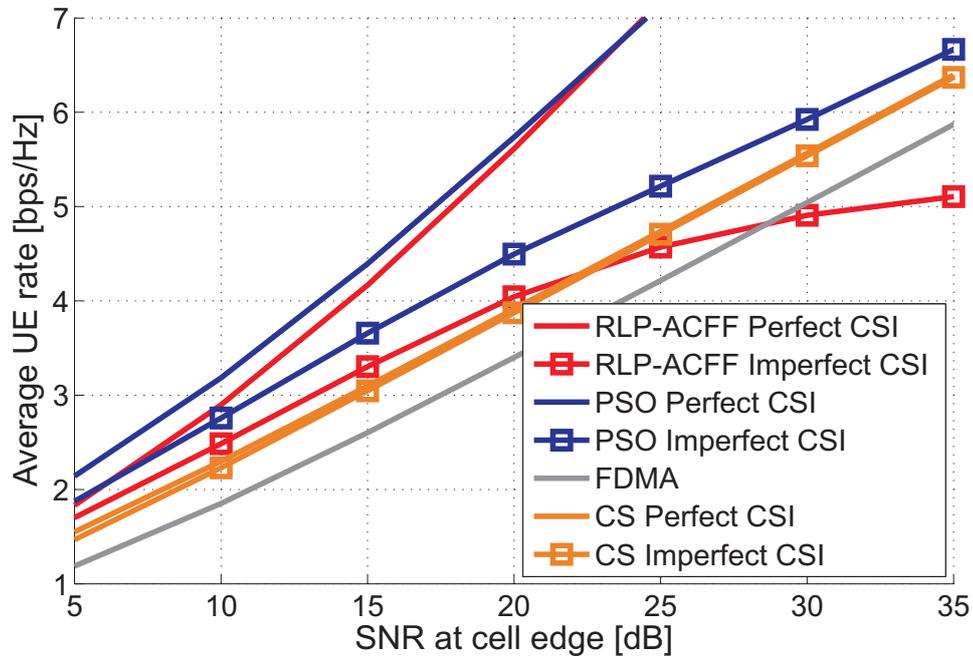


Figure B.1.: Scenario 1: Both users are in their respective cells, close to the cell border. (Reformatted and used with permission from R. Apelfröjd, Uppsala University)

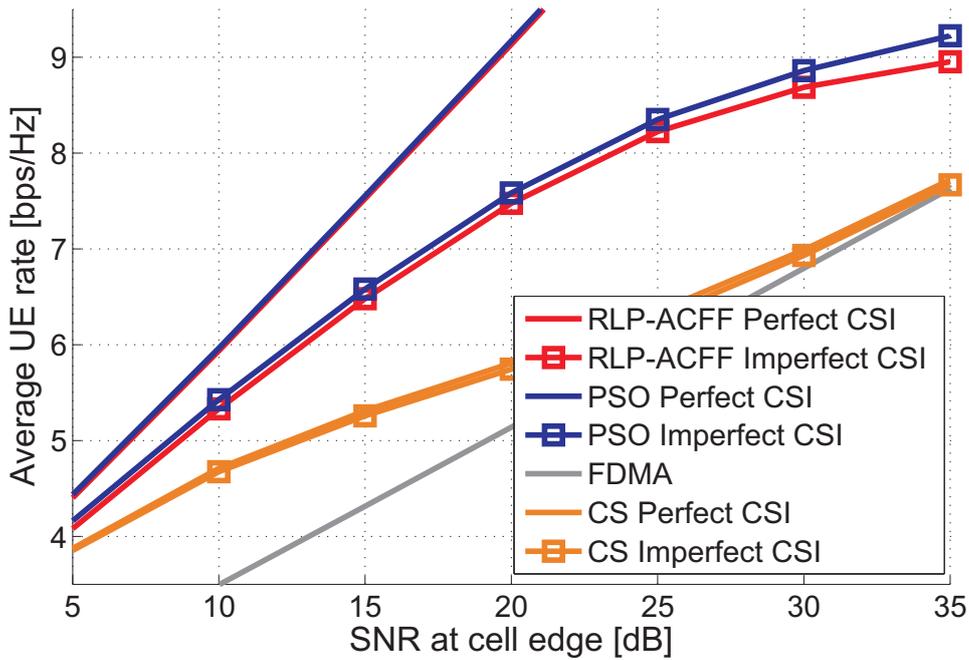


Figure B.2.: Scenario 2: Both users are at a distance of 125 m from their Master eNodeB. (Reformatted and used with permission from R. Apelfröjd, Uppsala University)

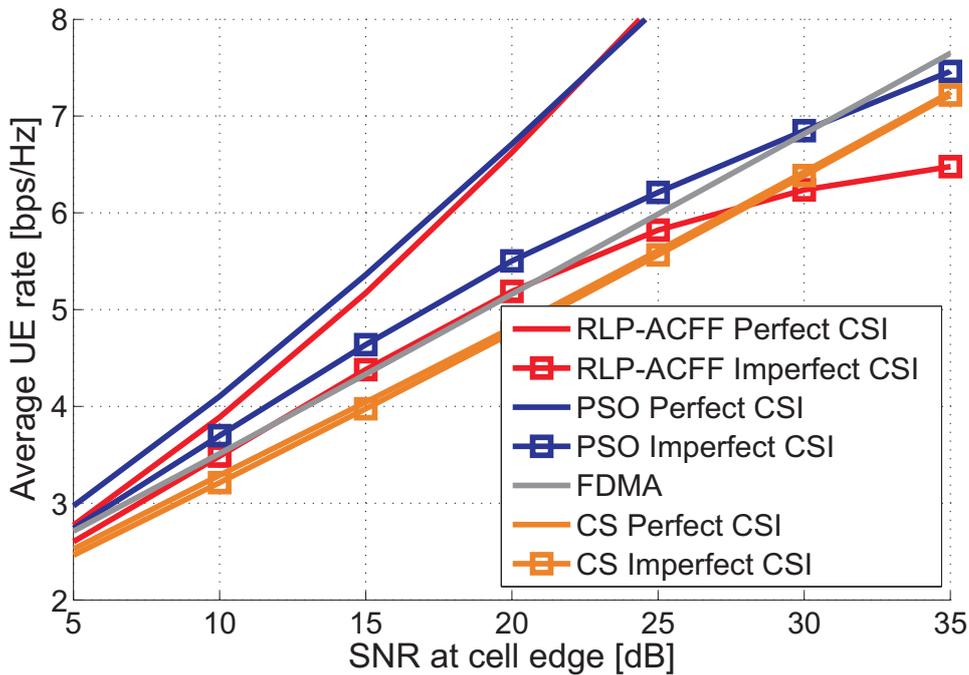


Figure B.3.: Scenario 3: One user located at the cell edge while one user is 125 m from its Master eNodeB. (Reformatted and used with permission from R. Apelfröjd, Uppsala University)

Bibliography

- [ARTD12] ARTIST4G D1.2, “Innovative advanced signal processing algorithms for interference avoidance”, ARTIST4G technical deliverable, Dec. 2010. [Online] Available: <https://ict-artist4g.eu/projet/deliverables>, [Accessed: Dec. 2012].
- [ARTD14] ARTIST4G D1.4, “Interference Avoidance Techniques and System Design”, ARTIST4G technical deliverable, Jul. 2012. [Online] Available: <https://ict-artist4g.eu/projet/deliverables>, [Accessed: Dec. 2012].
- [ASA12] R. Apelfröjd, M. Sternad and D. Aronsson, “Measurement-based evaluation of robust linear precoding for downlink CoMP”, in *proc. IEEE Int. Conf. on Commun. (ICC)*, 2012.