

Bassoy, Selcuk (2020) *Self-organised multi-objective network clustering for coordinated communications in future wireless networks.* PhD thesis.

https://theses.gla.ac.uk/81360/

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk

Self-Organised Multi-Objective Network Clustering for Coordinated Communications in Future Wireless Networks

Selcuk Bassoy

Submitted in fulfillment of the requirements for the
Degree of Doctor of Philosophy

School of Engineering
College of Science and Engineering
University of Glasgow



May 2020

Abstract

The fifth generation (5G) cellular system is being developed with a vision of 1000 times more capacity than the fourth generation (4G) systems to cope with ever increasing mobile data traffic. Interference mitigation plays an important role in improving the much needed overall capacity especially in highly interference-limited dense deployment scenarios envisioned for 5G. Coordinated multi-point (CoMP) is identified as a promising interference mitigation technique where multiple base stations (BS) can cooperate for joint transmission/reception by exchanging user/control data and perform joint signal processing to mitigate inter-cell interference and even exploit it as a useful signal. CoMP is already a key feature of long term evolution-advanced (LTE-A) and envisioned as an essential function for 5G. However, CoMP cannot be realized for the whole network due to its computational complexity, synchronization requirement between coordinating BSs and high backhaul capacity requirement. BSs need to be clustered into smaller groups and CoMP can be activated within these smaller clusters.

This PhD thesis aims to investigate optimum dynamic CoMP clustering solutions in 5G and beyond wireless networks with massive small cell (SC) deployment. Truly self-organised CoMP clustering algorithms are investigated, aiming to improve much needed spectral efficiency and other network objectives especially load balancing in future wireless networks. Low complexity, scalable, stable and efficient CoMP clustering algorithms are designed to jointly optimize spectral efficiency, load balancing and limited backhaul availability.

Firstly, we provide a self organizing, load aware, user-centric CoMP clustering algorithm in a control and data plane separation architecture (CDSA) proposed for 5G to maximize spectral efficiency and improve load balancing. We introduce a novel re-clustering algorithm for user equipment (UE) served by highly loaded cells and show that unsatisfied UEs due to high load can be significantly reduced with minimal impact on spectral efficiency. Clustering with load balancing algorithm exploits the capacity gain from increase in cluster size and also the traffic shift from highly loaded cells to lightly loaded neighbours.

Secondly, we develop a novel, low complexity, stable, network-centric cluster-

ing model to jointly optimize load balancing and spectral efficiency objectives and tackle the complexity and scalability issues of user-centric clustering. We show that our clustering model provide high spectral efficiency in low-load scenario and better load distribution in high-load scenario resulting in lower number of unsatisfied users while keeping spectral efficiency at comparably high levels. Unsatisfied UEs due to high load are reduced by 68.5% with our algorithm when compared to greedy clustering model. In this context, the unique contribution of this work that it is the first attempt to fill the gap in literature for multi-objective, network-centric CoMP clustering, jointly optimizing load balancing and spectral efficiency.

Thirdly, we design a novel multi-objective CoMP clustering algorithm to include backhaul-load awareness and tackle one of the biggest challenges for the realization of CoMP in future networks i.e. the demand for high backhaul bandwidth and very low latency. We fill the gap in literature as the first attempt to design a clustering algorithm to jointly optimize backhaul/radio access load and spectral efficiency and analyze the trade-off between them. We employ 2 novel coalitional game theoretic clustering methods, 1-a novel merge/split/transfer coalitional game theoretic clustering algorithm to form backhaul and load aware BS clusters where spectral efficiency is still kept at high level, 2-a novel user transfer game model to move users between clusters to improve load balancing further. Stability and complexity analysis is provided and simulation results are presented to show the performance of the proposed method under different backhaul availability scenarios. We show that average system throughout is increased by 49.9% with our backhaul-load aware model in high load scenario when compared to a greedy model.

Finally, we provide an operator's perspective on deployment of CoMP. Firstly, we present the main motivation and benefits of CoMP from an operator's viewpoint. Next, we present operational requirements for CoMP implementation and discuss practical considerations and challenges of such deployment. Possible solutions for these experienced challenges are reviewed. We then present initial results from a UL CoMP trial and discuss changes in key network performance indicators (KPI) during the trial. Additionally, we propose further improvements to the trialed CoMP scheme for better potential gains and give our perspective on how CoMP will fit into the future wireless networks.

Dedication

To:

My brother ***Halim Bassoy*** *&*
my sister ***Zehra Basmaci****,*

Two great souls, always being an inspiration for me with their courage and their dealings with difficult situations especially in the past few years.

University of Glasgow

College of Science & Engineering

Statement of Originality

Name: Selcuk Bassoy

Registration Number: 2287363

I certify that the thesis presented here for examination for a PhD degree of the University of Glasgow is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it) and that the thesis has not been edited by a third party beyond what is permitted by the University's PGR Code of Practice.

The copyright of this thesis rests with the author. No quotation from it is permitted without full acknowledgment.

I declare that the thesis does not include work forming part of a thesis presented successfully for another degree.

I declare that this thesis has been produced in accordance with the University of Glasgow's Code of Good Practice in Research.

I acknowledge that if any issues are raised regarding good research practice based on review of the thesis, the examination may be postponed pending the outcome of any investigation of the issues.

Signature: .....

Date: .....

List of Publications

Journals

1. **Selcuk Bassoy**, Hasan Farooq, Muhammad A. Imran, and Ali Imran. "Coordinated multi-point clustering schemes: A survey." IEEE Communications Surveys & Tutorials 19, no. 2 (2017): 743-764.

2. **Selcuk Bassoy**, Mona Jaber, Muhammad Ali Imran, and Pei Xiao. "Load aware self-organising user-centric dynamic CoMP clustering for 5G networks." IEEE Access 4 (2016): 2895-2906.

3. **Selcuk Bassoy**, Muhammad Ali Imran, Shufan Yang, and Rahim Tafazolli. "A Load-Aware Clustering Model for Coordinated Transmission in Future Wireless Networks." IEEE Access 7 (2019): 92693-92708.

4. **Selcuk Bassoy**, Mona Jaber, Muhammad Ali Imran. "Radio & Backhaul Load Aware Multi-Objective Clustering in Multi-cell MIMO Cooperative Networks". To be submitted to "IEEE Transactions on Wireless Communications"

Book Chapters

1. **Selcuk Bassoy**, Mohamed Aziz, Muhammad A. Imran. Chapter: Coordinated Multi-point for Future Networks: Field Trial Results. Book: Access, Fronthaul and Backhaul Networks for 5G and Beyond. Institution of Engineering and Technology (IET), 2017.

Invited papers and presentations

1. **Selcuk Bassoy** and Muhammad Ali Imran, “Determining how to ensure QoS and QoE during the transition to LTE: Self Organising Networks,” presented at the IQPC Mobile Network Performance Management Conference, London, 02-Jun-2014.

2. **Selcuk Bassoy** and Muhammad Ali Imran, “How to reduce interference on 4.5G/5G Networks: Coordinated Multi-point (CoMP),” presented at

the IQPC Mobile Network Performance Management Conference, London, 04-Jun-2015.

3. Syed Ali Raza Zaidi, Muhammad Ali Imran, Mounir Ghogho, and **Selcuk Bassoy**, "Big Data Empowered Self Organisation For Split Plane 5G Cellular Networks," presented at the Policy Making in the Big Data Era: Opportunities and Challenges, University of Cambridge, 16-Jun-2015.

4. **Selcuk Bassoy** and Muhammad Ali Imran, "Self Organised Coordinated Multipoint Clustering for 5G Wireless Networks" presented at the Self-Organising Networks Conference, London, 18-Oct-2016.

Contents

Abstract i

Dedication iii

Statement of Originality v

List of Publications vii

List of Tables xiii

List of Figures xv

List of Acronyms xix

Mathematical Notations xxi

Acknowledgements xxv

1 Introduction 1

1.1 Motivation 2
1.2 Objectives 5
1.3 Research Contributions 6
1.4 Thesis Outline 7

2 Background and State-of-the-Art 9

2.1 Introduction 9
2.2 Related Work 10
2.3 CoMP - Essential Background 12
2.3.1 Enabling Technologies for CoMP 17
2.4 SON as an enabler for CoMP Clustering 18
2.5 Clustering Taxonomy based on Self Organization 20
2.5.1 Static Clustering 21
2.5.2 Semi-Dynamic Clustering 23

2.5.3 Dynamic Clustering 24
2.6 Dynamic Clustering Taxonomy Based on Objective Function . . . 32
2.6.1 Spectral Efficiency 34
2.6.2 Backhaul Optimisation & Caching at RAN for JT-CoMP . 35
2.6.3 Energy Efficiency 38
2.6.4 Load Balancing 40
2.6.5 Multi-Objective Clustering 42
2.7 Conclusions and Outlook 43

3 Load Aware Dynamic CoMP Clustering 47
3.1 Introduction 47
3.2 Related Work 48
3.3 System Model 49
3.4 User-centric Clustering Algorithm 52
3.5 Clustering with Load Balancing 53
3.5.1 Cell Load and Unsatisfied Users Metric 53
3.5.2 Clustering Algorithm with Load Balancing 55
3.6 Numerical Results 56
3.6.1 Dense Deployment with High Load Scenario 59
3.6.2 Dense/Medium/Sparse Deployment 60
3.6.3 Dense Deployment with High/Medium/Light load 61
3.7 Conclusion 64

4 Load Aware Network-Centric Clustering for CoMP 67
4.1 Introduction 67
4.2 Related Work and Problem Statement 68
4.3 System Model 69
4.3.1 Network Model 69
4.3.2 CoMP Performance and Overhead Metrics 72
4.4 Dynamic Network-centric Clustering Problem as a Coalition Game 75
4.4.1 Coalition Formation Game Concepts 76
4.4.2 Utility Function 77
4.4.3 Merge/Split Operation 80
4.4.4 Algorithm Complexity 82
4.4.5 Partition Stability 83
4.5 Simulation Results 86
4.5.1 Homogeneous Network (HN) Scenario 88
4.5.2 Random Network (RN) Scenario 93
4.6 Conclusion 96

5 Multi-Objective CoMP Clustering **99**
5.1 Introduction . . . 99
5.2 Related Work . . . 100
5.3 System Model . . . 101
5.4 CoMP Performance Metrics . . . 104
5.4.1 RAN and Backhaul Load . . . 104
5.4.2 Cell Throughput . . . 105
5.4.3 Unsatisfied Users . . . 106
5.4.4 Pilot Overhead . . . 106
5.5 Coalition Game for Multi-Objective Clustering . . . 107
5.5.1 Coalitional Game Model for SC Clustering Sub-game . . . 107
5.5.2 Coalitional Game Model for User Transfers Sub-game . . . 110
5.5.3 Algorithm Stability . . . 114
5.5.4 Algorithm Complexity . . . 115
5.6 Simulation Results . . . 116
5.7 Conclusion . . . 124

6 CoMP for Future Networks: Field Trial Results **125**
6.1 Introduction . . . 125
6.2 Motivation and Benefits of CoMP - Operator Perspective . . . 125
6.3 Operational Requirements . . . 126
6.4 Uplink CoMP Field Trial for LTE-A . . . 127
6.4.1 Uplink CoMP Introduction . . . 127
6.4.2 UL CoMP in Trial Area . . . 129
6.4.3 Trial Performance Results . . . 130
6.5 Evolution into 5G . . . 134
6.6 Conclusions . . . 137

7 Conclusions and Future Research Directions **139**
7.1 Summary of Contributions . . . 141
7.1.1 Load aware user-centric CoMP clustering . . . 141
7.1.2 Load aware network-centric CoMP clustering . . . 141
7.1.3 Backhaul aware multi-objective CoMP clustering . . . 141
7.1.4 Real Network CoMP Field Trial Results . . . 142
7.2 Future Research Directions . . . 142

List of Tables

2.1 DL CoMP Types, Associated Benefits and Challenges 15
2.2 Summary of CoMP Clustering Approaches based on Self-Organisation 33
2.3 Summary of CoMP Clustering Approaches based on Objective Function 44

3.1 Simulation parameters. 57

4.1 Simulation Parameters 89

5.1 Simulation Parameters 118

6.1 Site Density Details for the 3 Deployment Scenarios 130

List of Figures

1.1 Proposed capacity enhancements for 5G. 2
1.2 Dynamic multi-objective CoMP clustering illustration in CDSA architecture. 3

2.1 Main downlink CoMP types for LTE-A [6]. 12
2.2 Coherent and non-coherent joint processing. 14
2.3 CoMP clustering taxonomy based on self organization. 20
2.4 Dynamic CoMP clustering taxonomy. 25
2.5 CoMP clustering taxonomy based on objective function. 34

3.1 Control-data separation architecture (CDSA). 49
3.2 Simulation network topology illustration, SC and UE locations, hotspot and non-hotspot areas, cell borders following Voronoi tessellation. 58
3.3 Unsatisfied UEs and spectral efficiency changes in dense deployment scenario with high load. 59
3.4 Dense deployment scenario with high load. 60
3.5 Unsatisfied UEs in dense/medium/sparse deployment scenario with high load. 61
3.6 Dense/medium/sparse deployment scenario with high load. 62
3.7 SC load distribution in dense/sparse deployment scenario with high load. 62
3.8 Unsatisified UEs in dense deployment scenario with high/medium/light load. 62
3.9 Dense deployment scenario with high/medium/light load. 63
3.10 SC load distribution in dense deployment with light load. 63
3.11 Unsatisfied UEs for dense deployment with high load for different max. cluster size. 64

4.1 System model for downlink MU JT-CoMP. 70
4.2 Optimum pilot overhead vs CoMP cluster size [86]. 74

4.3 Complexity vs cluster size $|\mathcal{C}_i|$ when $(C^n_{max} = 6)$. 75
4.4 Utility function $v_1(SC_m, \mathcal{C}_i)$ vs SC load $\hat{l}_m$ for different cluster sizes when $\hat{u}_m = 50$ and $C^n_{max} = 6$. 79
4.5 An illustration of $\mathbb{D}_c$ stable partitions. 84
4.6 $\mathbb{D}_c$ stable partitions from merge/split cluster formation game. . . 86
4.7 Average spectral efficiency and cluster size in HN with/without hotspot scenarios. 91
4.8 Snapshot of SE-GR, SE-GA and L-GA clusters in HN with hotspot scenario. 92
4.9 eNodeB load distribution in HN with hotspot. 92
4.10 Unsatisfied UEs in HN with hotspot. 93
4.11 Random network simulation setup. 94
4.12 Average spectral efficiency and cluster size in RN with/without hotspot scenarios. 95
4.13 eNodeB load CDF in RN with hotspot. 95
4.14 Unsatisfied UEs in RN with hotspot. 95
4.15 Payoff changes with merge/split iterations for L-GA algorithm in RN with/without hotspot scenario. 96
4.16 Unsatisfied UEs/spectral efficiency changes with split/merge iterations for L-GA algorithm in RN with hotspot scenario. 97

5.1 Snapshot of SE-GR, L-GA and LBH-GA clusters in HN with hotspot scenario. 119
5.2 Average spectral efficiency and CS in RN without hotspot scenerio when VDSL2 rate= 33%. 120
5.3 Cluster size and spectral efficiency comparision for all backhaul cases. 121
5.4 Unsatisfied UEs and system throughput comparison for all backhaul cases. 122
5.5 LBH-GA game actions vs. unsatisfied UEs for 50% VDSL rate. . 122
5.6 LBH-GA unsatisfied UEs vs spectral efficiency and payoff for 50% VDSL rate. 123
5.7 SC overall load and backhaul load distribution for 50% VDSL rate. 123

6.1 Network topology of the 3 deployment scenarios in the trial. . . . 129
6.2 An illustration of intra-BS Uplink CoMP: Joint processing with IRC 130
6.3 PUSCH SINR with/without UL CoMP in the 3 deployment scenarios 131
6.4 Employed uplink modulation types with/without UL CoMP in the 3 deployment scenarios. 132

6.5 Uplink BLER with/without UL CoMP in the 3 deployment scenarios. 133
6.6 Average UL user throughput in 3 deployment scenarios. 134
6.7 Power limited UEs in 3 deployment scenarios. 134
6.8 Uplink packet loss and drops in dense deployment at 800 MHz. . . 134

List of Acronyms

3GPP	Third Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
AGWN	Additive White Gaussian Noise
BBU	Baseband Processing Unit
BLER	Block Error Rate
BS	Base Station
CAPEX	Capital Expenditure
CB	Coordinated Beamforming
CCU	CoMP Control Unit
CDR	Call Data Record
CDSA	Control and Data Plane Separation
CoMP	Coordinated Multi-point
C-RAN	Centralized Radio Access Network
CS	Coordinated Scheduling
CSI	Channel State Information
D2D	Device to Device
DAS	Distributed Antenna System
DL	Downlink
DPS	Dynamic Point Selection
EPA-A	Extended Pedestrian-A
GBR	Guaranteed Bit Rate
HetNet	Heterogeneous Network
HN	Homogeneous Network
ICIC	Intercell Interference Cancellation
IRC	Interference Rejection Combining
JR	Joint Reception
JT	Joint Transmission
KPI	Key Performance Indicator
LTE-A	Long Term Evolution-Advanced

MBS	Macro Base Station
MD	Macro Diversity
MDT	Minimization of Drive Tests
MEC	Mobile Edge Computing
MIMO	Multiple Input-Multiple Output
MU	Multi-User
NLOS	Non-Line-of-Sight
NOMA	Non-Orthogonal Multiple Access
OPEX	Operational Expenditure
PPP	Poisson Point Process
PRB	Physical Resource Block
QoS	Quality of Service
RAN	Radio Access Network
RN	Random Network
RRU	Remote Radio Unit
RSRP	Reference Signal Received Power
RSSI	Received Signal Strength Indicator
SC	Small Cell
SFR	Soft Frequency Reuse
SINR	Signal-to-Interference-plus-Noise Ratio
SON	Self Organizing Networks
SU	Single User
TP	Transmission Point
UE	User Equipment
UL	Uplink
ZF	Zero Forcing

Mathematical Notations

Unless stated otherwise, following ruleset is used in the thesis for mathematical notation:

- Uppercase, calligraphic letters represent *sets* (e.g $\mathcal{C}$)
- Lowercase, italic letters represent *scalars* (e.g. d_k)
- Uppercase, italic letters represent *constants* (e.g. P_{min})
- Lowercase, bold-face letters represent *vectors* (e.g. $\mathbf{w}_k$)
- Uppercase, bold-face letters represent *matrices* (e.g. $\mathbf{H}$)
- Variables with a hat on top represent *estimates* (e.g. $\hat{z}_{im}$)
- Superscript indices on variables are used to represents users, subscript indices are used to represent SCs (e.g. $\mathcal{C}_i^k$ refers to user-centric cluster for UE_k in SC cluster $\mathcal{C}_i$

The following variables are adapted in this thesis:

b_{PRB}	Bandwidth for user-data in one PRB
B_{PRB}	Total bandwidth for one PRB
B_{tot}	Total system bandwidth
$\mathcal{C}$	Set of all small cells
$\mathcal{C}^k$	Set of SCs in user-centric cluster for UE_k in $\mathcal{C}$
C_{max}	Max. user-centric cluster size limit
C_{max}^h	Max. user-centric cluster size limit for UEs served by highly loaded SCs
C_{max}^n	Soft limit for network-centric cluster size
$\mathcal{C}_i$	Set of SCs in network-centric cluster i
$\mathcal{C}_i^k$	Set of SCs in user-centric cluster for UE_k in $\mathcal{C}_i$
d_{im}^{BH}	Backhaul throughput demand for SC_m in $\mathcal{C}_i$
$\hat{d}_{im}^{RAN}$	Estimated dedicated RAN throughput demand for SC_m in $\mathcal{C}_i$
d_{im}^{RAN}	RAN throughput demand for SC_m in $\mathcal{C}_i$
d_k	GBR requirement for UE_k
$\hat{f}_{im}^{BH}$	Estimated backhaul throughput capacity at SC_m in $\mathcal{C}_i$

g_{km}	Channel coefficient scalar from SC_m to UE_k
$\mathbf{H}$	Channel matrix between SCs in $\mathcal{C}_i^k$ and UEs in $\mathcal{U}_i^k$
$\mathbf{h}_k$	Channel vector for UE_k
L_{min}	Minimum SC load threshold
$\hat{l}_m^{BH}$	Estimated backhaul load at SC_m
$\hat{l}_m^{RAN}$	Estimated RAN load at SC_m
N_0	Noise spectral density
n_k	User-centric cluster size for UE_k
P_Δ	Max. power offset from best serving cell for user-centric clustering
P_Δ^{nei}	Max. power offset from best serving cell for neighbour definition
p_{k1}	Received power at UE_k from best serving SC
p_{km}	Received power at UE_k from SC_m
P_{min}	Min. received power threshold for user-centric clustering
P_{min}^{nei}	Min. received power threshold for neighbour definition
P_{Tx}	Total transmit power for any SC
r_{im}	Total number of PRBs required at SC_m in $\mathcal{C}_i$
R_{tot}	Total number of resource blocks for any SC
$\hat{r}_k$	Estimated dedicated PRB count for UE_k at each SC in $\mathcal{C}_i^k$
r_k	Average number of PRBs allocated for UE_k at each SC in $\mathcal{C}_i^k$
SE_Δ^{max}	Maximum allowed spectral efficiency loss for re-clustering
$SINR_{min}$	Min SINR required for re-clustering
$\hat{t}_{im}$	Estimated dedicated throughput for SC_m in $\mathcal{C}_i$
$\mathcal{U}$	Set of all active UEs
$\mathcal{U}^k$	Set of UEs scheduled in the same resources as UE_k in $\mathcal{C}^k$
$\mathcal{U}_{im}$	Set of UEs attached to SC_m in $\mathcal{C}_i$
$\hat{u}_{im}$	Estimated dedicated user count attached to SC_m in $\mathcal{C}_i$
$\mathcal{U}_{im}^{best}$	Set of UEs attached to SC_m in $\mathcal{C}_i$ where SC_m is the best serving cell
$\mathcal{U}_m$	Set of UEs attached to SC_m in $\mathcal{C}$
$\hat{u}_m$	Estimated dedicated UE count attached to SC_m in $\mathcal{C}$
$\mathcal{U}_m^{best}$	Set of UEs attached to SC_m in $\mathcal{C}$ where SC_m is the best serving cell
$\mathcal{U}_h$	Set of UEs served by highly loaded SCs
$\mathcal{U}_i$	Set of UEs attached to $\mathcal{C}_i$
$\mathcal{U}_i^k$	Set of UEs scheduled in the same resources as UE_k in $\mathcal{C}_i^k$
$v_1(SC_m, \mathcal{C}_i)$	Load aware utility of $SC_m \in \mathcal{C}_i$
$v_2(SC_m, \mathcal{C}_i)$	Spectral efficiency based utility of $SC_m \in \mathcal{C}_i$
$\mathbf{W}$	Precoding matrix at SCs in $\mathcal{C}_i^k$ for UEs in $\mathcal{U}_i^k$
w_{mk}	Precoding scalar on SC_m for UE_k
$\mathbf{w}_k$	Precoding vector for UE_k

y_k	Spectral efficiency at UE_k
$\hat{z}_{im}$	Estimated unsatisfied UE count at SC_m in $\mathcal{C}_i$
$\hat{z}_m$	Estimated unsatisfied UE count at SC_m in $\mathcal{C}$
δ_Δ	Min payoff gain required for user transfer operation
$\delta_{v_{ij}}$	Payoff gain of merge operation for $(\mathcal{C}_i, \mathcal{C}_j)$
$\delta_{v_{ixj}}$	Payoff gain for SC/UE transfer operation for SC_{ix}/UE_{ix} from $\mathcal{C}_i$ to $\mathcal{C}_j$
$\lambda_\mathcal{C}$	SC PPP distribution density
$\lambda_{\mathcal{U}_{high}}$	PPP UE density in high traffic area
$\lambda_{\mathcal{U}_{low}}$	PPP UE density in low traffic area

Acknowledgements

Firstly, I really would like to express my deepest gratitude to my supervisor Prof. Muhammed Imran for this excellent guidance and support in my long PhD journey, keeping me motivated at all times. I would like to also extend my thanks to members of the research team at University of Glasgow, especially to Dr Paulo Henrique Valente Klaine for this continuous support on various matters which made me very welcome to the department as a part time student. My deepest gratitude also goes to Dr Shufan Yang who has given me great support and feedback on my work and on one of our jointly published paper.

I would like to also extend my thanks to my second supervisor Prof. Pei Xiao at 5GIC (University of Surrey) for his support during my first year with fruitful discussions and helping me on finding my way into my research topic. I am also grateful for Dr Mona Jaber for her support and constructive feedback on my work during the first couple of years.

Last but least, I would like express my sincerest gratitude to my wife Bengisu Bassoy and our lovely two children Omer Ediz and Elif for their patience and support during this long journey. I am so thankful to them for sacrificing their share of my time, giving me the space and extra time for my PhD work. Without my wife's support and encouragement from the very first day till completion for the whole 6 years, this thesis wouldn't have been possible.

Chapter 1

Introduction

Future wireless cellular networks will be under tremendous pressure with the increasing data demand as the user behavior changes with popular high bandwidth applications. While smart phones become very popular, high bandwidth hungry applications like video streaming, multimedia file sharing etc. becomes more popular. With 5G, more diverse applications like massive machine type communications, ultra-reliable and low latency communications and enhanced mobile broadband will be widely available. Massive additional capacity is required to handle these wide-range of diverse applications. Mobile data traffic has been growing rapidly and it is expected to grow at an annual growth rate of 46% over the next 5 years i.e. a 7-fold increase is expected by 2022 [53]. Moreover, a 1000 fold increase in mobile data traffic is expected for 5G beyond 2020 [101, 129]. To enable 5G to cope with this tremendous increase in data growth, following three development areas in the emerging wireless landscape are proposed [80, 101, 129].

1. Network Densification - Massive small cell deployment
2. Increased Spectral Efficiency - CoMP, Multiple Input-Multiple Output (MIMO), Enhanced coding techniques
3. Additional Spectrum

Figure 1.1 illustrates the potential capacity gains expected from each of the three key capacity enhancement proposed for 5G [80, 101, 129]. Biggest capacity gains are expected from network densification: a massive deployment of SCs will be required [87, 111] in search for additional capacity. Dense SC deployment in heterogeneous cellular networks (HetNet) will lead to a severely interference limited network depending on the available frequency spectrum. More advanced inter-cell interference mitigation techniques will need to be deployed to combat interference and improve spectral efficiency. Improved spectral efficiency will lead

to much needed capacity enhancement as highlighted above as one of the three key development areas for 5G capacity requirement.

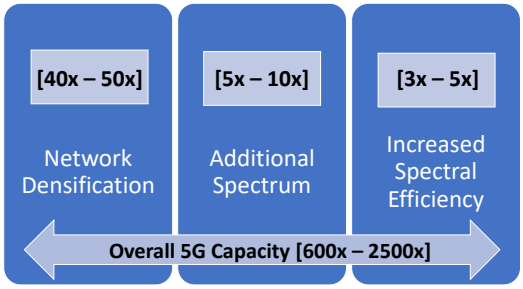


Figure 1.1: Proposed capacity enhancements for 5G.

CoMP or network-MIMO is the emerging technology which is proposed to reduce interference especially at the cell edge and hence improve high data rate footprint especially in dense deployment interference limited scenarios. CoMP has been introduced for LTE-A by the third generation partnership project (3GPP) in Release 11 [6] and it is widely discussed in literature as a key feature for 5G [112, 114, 186]. CoMP is expected to be part of 5G in 3GPP Release 16 which is expected to be released in June-2020 [67]. Massive MIMO is has been already part of 5G in the first release (Release 15) and it is being extended to CoMP with non-coherant joint transmission in Release 16. Introduction of CoMP will especially aim to improve reliability of ultra-reliable, low latency use case for 5G. Further CoMP enhancements are also expected from 3GPP in Release-17 especially on channel state information (CSI) acquisition to tackle blockage issue in mmWave frequencies [67].

1.1 Motivation

CoMP technology makes use of the shared data between coordinating transmission points (TP) and inter-cell interference is mitigated or even exploited as meaningful signal at the receiver. Coordination between all cells in the network is very complex due to precise synchronization requirement within coordinated cells, additional pilot overhead, additional signal processing, complex beamforming design and scheduling among all BSs. It will require high bandwidth backhaul links due to CSI and/or user data exchange between all BSs [82, 93]. To reduce this overhead, smaller size cooperation clusters are required where coordination only takes place within the cluster. Optimum cooperating cluster selection is key for maximizing the benefits of CoMP.

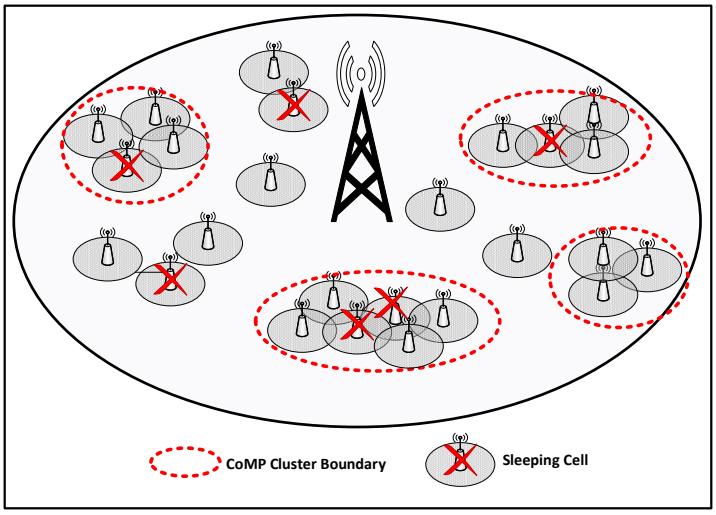


Figure 1.2: Dynamic multi-objective CoMP clustering illustration in CDSA architecture.

A number of challenges need to be critically evaluated for a comprehensive CoMP clustering approach to maximise the benefits of CoMP:

- **Is it efficient to deploy CoMP ?** The first question which need to be answered is, if it is worth deploying CoMP for individual cells in a given network setup. Would the overheads for deploying CoMP be more than the gains it provides ? As illustrated in Figure 1.2, cells closer to each other need to form clusters for cooperation as the CoMP gains would be maximized when there is severe inter-cell interference which can be mitigated. However, isolated cells may need to work without coordination, based on the limited amount of inter-cell interference experienced from other cells. In addition, users close to the cell center may not experience high inter-cell interference, however cell edge users will suffer from high interference hence, it can be more efficient to deploy CoMP for cell edge users only. In [63], authors presented a dynamic clustering scheme and suggested no spectral efficiency gain in employing CoMP in high signal-to-interference-plus-noise ratio (SINR) region due to additional pilot signalling required for CoMP, reducing spectral efficiency more than the expected gains. Users are allocated CoMP clusters or CoMP is not used based on their SINR from the local serving BS. It is shown that CoMP gains are maximized when received power levels from coordinating cells are close to the received power levels of the local serving cell. Hence it can be concluded that CoMP gains vary with network density and CoMP may not need to be deployed for some cells based on their location, user profile and the amount inter-cell interference.

- **How many cells in the cluster ?** Cluster size is another key parameter for optimal CoMP clustering. Too small clusters will fail to provide full achievable gains from CoMP, on the other hand, big cluster size will lead to increased overhead on CSI feedback and backhaul capacity [117]. Increased cluster size will give better weighted sum rate [145] but with the cost of additional signal processing and increased feedback and signalling. Moreover, increased cluster size can lead to energy inefficiency in terms of achieved bits/joule [52]. As illustrated in Figure 1.2 for an example CDSA architecture, some clusters will have 6 cells, others will have 5 or 4 and some others will reduce cluster size by switching off some cells within the cluster for energy efficiency. Hence, there is no ideal fixed cluster size, instead, cluster size needs to be a dynamic parameter in the clustering algorithm which needs to change based on channel conditions, user profile and network density.

- **Which cells to switch off for energy efficiency?** As illustrated in Figure 1.2, some cells can be switched off by forming intelligent CoMP clusters to enhance SINR and make sure minimum SINR is provided while some cells are switched off for energy efficiency. A number of network objectives will need to be considered for BS switch-off:

  - Can the remaining capacity in the cooperating cluster cope with the traffic demand for a given quality of service (QoS)?
  - Is SINR provided by the cooperating cluster without the sleeping cell over the minimum threshold ?
  - Do the cells within the cooperated set have enough backhaul bandwidth to cope with increased traffic when a cell is switched off for energy efficiency ?

- **Multi-objective Clustering:** CoMP clustering need to maximize spectral efficiency improvement as the key objective of CoMP, however other network objectives and limitations need to be taken into account for CoMP clustering for a more realistic approach. Cooperation introduce additional capacity in the network by improving spectral efficiency [82]. Intelligent clustering algorithms can be employed to support load balancing by shifting traffic from highly loaded cells to its neighbouring clusters. Increased cluster size can also uplift capacity in hotspot areas based on network topology. However, backhaul bandwidth requirement will also increase with increased cluster size. So backhaul load need to be considered while increasing cluster

size. On the other hand, BS sleeping can be considered to reduce energy consumption and also backhaul requirement. Overall, alongside with spectral efficiency, other network objectives like load balancing, energy efficiency and backhaul limitations need to be considered for intelligent CoMP clustering.

Given the challenges for CoMP clustering design as discussed above, static clustering based on a fixed topology will fail to give expected gains for future networks as the network topology will be dynamically changing with on/off sleeping cells, user deployed cells with unknown location etc. Moreover, spatio-temporal distribution of users and service demands dynamically changes. To maximize CoMP gains, clustering algorithms need to be able to accurately respond to these dynamically changing network conditions and user profiles. Self-organised, dynamic CoMP clustering algorithms are required to maximize multiple objectives like spectral efficiency, energy efficiency and load balancing while taking limitations into account, such as increased complexity and available bandwidth.

1.2 Objectives

This PhD project aims to investigate dynamic CoMP clustering techniques with multiple objectives to maximize CoMP gains for future wireless networks. There has been a number of studies carried out for CoMP clustering already but there is very few work on multi-objective clustering where multiple network objectives are jointly optimized for CoMP clustering. As stated in above section, energy efficiency, load balancing, backhaul availability and spectral efficiency are directly related with each other and severely related to how CoMP cluster should be structured. CoMP is likely to be deployed in densely populated areas where there is interference limited dense deployment and there is inevitable hotspot areas at certain times of the day which results in significant load imbalance between BSs. This research fills the gap in literature to introduce load balancing as a key objective to CoMP clustering and provides insights on dynamic CoMP clustering solutions to jointly optimize load balancing and other network objectives.

In this context, the objectives of this thesis are:

- To provide an extensive literature review of CoMP clustering solutions and critically review strong and weak points of available solutions, provide gaps in literature and identify future research directions.
- To provide insights of load aware dynamic CoMP clusters where backhaul load, RAN load and spectral efficiency are jointly optimized and trade-off

between the performance of various objectives and overheads/complexity are analyzed.

- To provide a real CoMP implementation case from a major network operator in the UK to assess the deployment challenges against the impact of CoMP on main KPIs and provide essential improvements required for future networks.

- To provide future trends and research directions for CoMP clustering and conclusions from our work.

1.3 Research Contributions

Main contributions of this thesis are summarized as follows:

- A comprehensive literature survey has been conducted on CoMP clustering. Enabling technologies for CoMP like centralized radio access network (C-RAN) and CDSA are analyzed, various types of CoMP techniques are compared against their implementation challenges. State of the art has been surveyed and two novel taxonomies for CoMP clustering are introduced based on self organization and clustering objectives. This extensive survey has been successfully published in IEEE Communications Surveys and Tutorials journal [30].

- A novel load aware, self-organised user-centric CoMP clustering algorithm is developed where load balancing and spectral efficiency objectives have been jointly optimized for CoMP clustering. This work is the first attempt to fill the gap in literature introducing load balancing as one of the key objectives for CoMP clustering. Numerical results promises significant improvement on load balancing by dynamically changing the cluster size and re-clustering to move load from congested cells. This work is successfully published in IEEE Access journal [32].

- User-centric clustering solutions provide an upper bound for CoMP gain, but introduce high complexity for realistic CoMP implementation, especially when the network size is large. To reduce this limitation, a novel load aware, network-centric clustering solution is provided where user-centric clustering algorithm can be implemented within smaller network-centric clusters. Network clustering problem is formulated as a coalitional game where a novel utility function is developed to jointly optimize load balancing

and spectral efficiency. Stability and complexity aspects of the algorithm is extensively studied. This work is successfully published in IEEE Access journal [31].

- A further network-centric clustering model is developed to include backhaul load as an additional network objective alongside with RAN load and spectral efficiency. As backhaul bandwidth is a key dependency for CoMP implementation, backhaul load balancing is an important objective for any CoMP implementation. Two unique coalitional games are designed, first one for forming and dynamically updating the BS clusters and the second one is for moving the users between the BS clusters to further reduce load on congested BSs and improve backhaul/RAN load balancing and user satisfaction. This work is being submitted for publication in "IEEE Transactions on Wireless Communications".

- A real network CoMP deployment for a major UK operator is analyzed, various deployment challenges and performance impact is assessed. Additional CoMP enhancements are identified for improved performance in 5G and beyond future networks. This work is published as a chapter in Book: Access, Fronthaul and Backhaul Networks for 5G and Beyond. Institution of Engineering and Technology (IET), 2017 [81].

1.4 Thesis Outline

The rest of the thesis is organised as follows:

Chapter 2 presents the state of the art on CoMP clustering. We first present the need for CoMP and the clustering challenge in future wireless networks and briefly evaluate different types of CoMP implementation. We then introduce self organization concept as an important framework for dynamic CoMP clustering. Next, we present two novel taxonomies on existing CoMP clustering solutions, based on self organization and aimed objective function. Strengths and weaknesses of the available solutions in the literature are critically discussed. This extensive literature review has been published in IEEE Communications Surveys and Tutorials Journal [30].

Chapter 3 presents our first study on a self-organizing user-centric CoMP clustering algorithm in a CDSA scenario proposed for 5G. We aim to maximize spectral efficiency for a given maximum cluster size, and we further improve this clustering algorithm to distribute load from highly loaded cells to other lightly loaded neighbours for multi-user (MU), joint transmission (JT) CoMP case. We

introduce a novel re-clustering algorithm for UEs served by highly loaded cells and show that unsatisfied UEs due to high load can be significantly reduced with minimal impact on spectral efficiency. This work has been published in IEEE Access journal [32].

In **Chapter 4**, we extend our work on Chapter 3 to a novel load aware network-centric clustering model. We develop a load aware clustering model by employing a merge/split concept from coalitional game theory. A load aware utility function is introduced to maximize both spectral efficiency and load balancing objectives. We show that proposed load aware clustering model dynamically adapts into the network load conditions providing high spectral efficiency in light load conditions and results in better load distribution with significantly less unsatisfied users in over-load conditions while keeping spectral efficiency at comparable levels when compared to a greedy clustering model. Simulation results show that proposed solution can significantly reduce the number of unsatisfied users due to over-load conditions when compared to greedy clustering algorithm. Furthermore, we analyze the stability of the proposed solution and prove that it converges to a stable partition in both homogeneous network (HN) and random network (RN) with and without hotspot scenarios. Additionally, we show the convergence of our algorithm into the unique clustering solution with the best payoff possible when such solution exists.

In **Chapter 5**, we further extend our work from Chapter 4 and develop a multi-objective, load aware dynamic CoMP clustering model to optimize backhaul load in addition to spectral efficiency and RAN load. We formulate our load aware model as two coalitional sub-games for SC and UE clustering respectively. Merge/split/transfer actions for each sub-game are defined, complexity and stability analysis are provided. Extensive simulation results show that our model dynamically allocates clusters to avoid backhaul limited sites and achieves significantly better load balancing with reduced unsatisfied users and increased throughput in high load scenario. Our backhaul aware model provides an additional 21.9% average throughput when compared to the same model without backhaul awareness in the case when all BSs are backhaul limited.

In **Chapter 6**, we present a commercial network CoMP trial results, and discuss operational challenges and further improvements to currently available solutions. This work is published as a book chapter in: "Access, Fronthaul and Backhaul Networks for 5G and Beyond. Institution of Engineering and Technology (IET), 2017" [81].

Chapter 7 provides a summary of major findings presented, draws conclusions and identifies future research directions.

Chapter 2

Background and State-of-the-Art

In this chapter, we provide a comprehensive survey on the state-of-the-art on one of the key challenges of CoMP implementation: CoMP clustering. As a starting point, we present a brief essential background about CoMP, enabling network architectures and the clustering challenge. We then introduce self organizing networks (SON) as an important concept for effective dynamic CoMP clustering to maximize CoMP gains. Next, we present two novel taxonomies on existing CoMP clustering solutions, based on self organization and aimed objective function. Strengths and weaknesses of the available clustering solutions in the literature are critically discussed. We then conclude the chapter with a summary of lessons learnt, future open research areas and how our work presented in the following chapters have fulfilled some of the gaps in literature. Most of this work has been published in IEEE Communications Surveys and Tutorials journal [30].

2.1 Introduction

Optimal CoMP clustering is one of the key challenges for CoMP implementation for future wireless networks. Selecting the right group of BSs for cooperation for a given network/user profile is key to maximize potential CoMP gains. Trade-off between the overhead and interference cancellation benefits needs to be taken into account for optimum cluster size design. There are multiple objectives for CoMP clustering and the right balance between the various efficiency/overhead indicators is a challenge. For example, maximizing spectral efficiency with CoMP clustering can degrade energy efficiency and backhaul limitations may prevent such cluster design. Hence, a comprehensive clustering approach should be considered to achieve the right balance between multiple objectives of future networks such as energy efficiency, load balancing, spectral efficiency and backhaul availability.

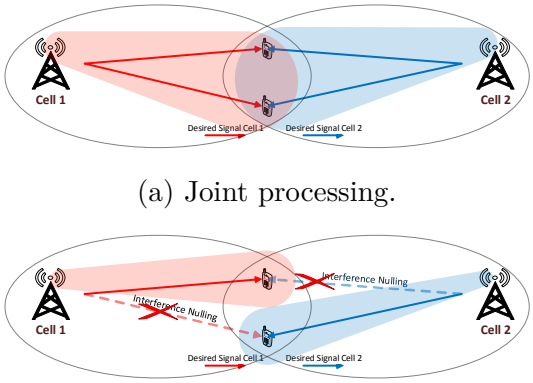
Main scope of this chapter is to provide an extensive survey of CoMP clustering techniques in the literature over the last decade. We provide a novel taxonomy on CoMP clustering techniques, critically discuss the strengths and weaknesses of the available solutions in the literature. The rest of the chapter is structured as follows:

In Section 2.2, we review the relevant work on CoMP clustering and show our novel contribution with this survey. In Section 2.3, we provide an essential background about CoMP, main types of CoMP implementation, associated challenges and the enabling network architectures. In Section 2.4, we introduce a key framework for CoMP clustering challenge and present self organizing networks (SON) as an important platform to implement effective dynamic CoMP clustering algorithms. In Section 2.5, a novel self-organization based taxonomy on CoMP clustering in the literature is introduced. Various CoMP clustering approaches are discussed and critical review is provided based on self organization, complexity, scalability and practical use. In Section 2.6, a further taxonomy is introduced based on the aimed objective function of CoMP clustering. An extensive survey of existing clustering approaches based on different objective functions like spectral efficiency, energy efficiency, load balancing and backhaul optimization are presented and criticized in detail. In Section 2.7, we conclude the chapter with a summary of lessons learnt, open research areas for CoMP clustering and the role of our presented work in this thesis on filling some of the gaps in literature.

2.2 Related Work

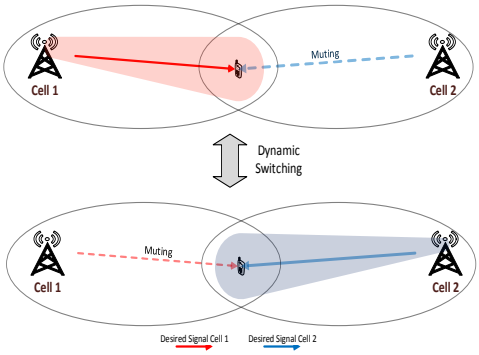
A number of works have already been conducted for CoMP in general [82,87,131] and more specifically for LTE-A implementation in [93, 152]. Deployment scenarios and brief clustering reviews are presented in these works, however there is no study in literature that extensively surveys clustering challenge for CoMP. In [87], CoMP clustering is reviewed briefly and a subset of static overlapping clusters are presented, however this work lacks a comprehensive survey on all clustering models in literature, especially missing the advanced clustering techniques i.e. dynamic and/or multi-objective based clustering. CoMP concept and trial results are presented in [82] with a dynamic clustering algorithm trialed in a test network, however the paper again lacks a review of other available clustering models. Authors in [93] discuss CoMP implementation challenges and various deployment scenarios for LTE-A, however clustering challange is not exploited in the paper. Backhaul capacity and latency requirement for different CoMP schemes are investigated in [36]. A user-centric CoMP clustering approach is

studied to investigate available backhaul capacity/latency impact on CoMP clustering. However the paper lacks on an extensive review of other available CoMP clustering algorithms which can be employed to dynamically adapt to available backhaul capacity. Beylerian et al. presents a service-aware resource allocation for non-coherent JT-CoMP in C-RAN architecture in [34] where a static and a user-centric clustering approach is presented. Same authors propose a further resource allocation solution combining non-orthogonal multiple access (NOMA) scheme with CoMP in [33] to exploit power and space domain multiplexing and further improve capacity. A static clustering approach of a fixed cluster size of two is employed in this work, however both studies does not intend to cover all clustering solutions available, especially missing the dynamic clustering algorithms which can reduce high complexity on user-centric clustering solution in large clusters of cells. Rao et al. presents a survey on energy efficient resource management for cooperative networks in [144] however energy efficient cooperative clustering challenge is not reviewed extensively. A comprehensive book is published about CoMP [131], two example clustering techniques, one for static, one for dynamic clustering is presented however it again fails to present an extensive review for CoMP clustering. Coalitional game theory is introduced in [76] as an important analytical tool to form CoMP clusters. An example clustering algorithm is also presented for UE clustering in the uplink, maximizing the sum-rate capacity. Nonetheless, the book fails to provide a review of all CoMP clustering approaches available. An extensive survey is provided on CDSA for future networks in [120], however this survey lacks a review on CoMP within the CDSA architecture. Mustafa et al. provides a survey on device to device (D2D) CoMP within the CDSA architecture in [122] and discuss CoMP clustering briefly with one dynamic clustering example. Both papers [120, 122] lack a wider review of all CoMP clustering solutions available in literature. In [22], an extensive review for SON is provided, however CoMP clustering is not discussed in relation to SON framework. To the best of our knowledge, there is no comprehensive survey in the literature about CoMP clustering. This chapter provides an extensive survey on the existing CoMP clustering approaches in literature. Two novel taxonomies on CoMP clustering based on aimed objective and self organization are presented. Strengths and weaknesses of available solutions are critically reviewed and future research directions are identified.



(a) Joint processing.

(b) Coordinated beamforming.



(c) Dynamic point selection.

Figure 2.1: Main downlink CoMP types for LTE-A [6].

2.3 CoMP - Essential Background

In this section, we provide an essential background of CoMP before moving to the main scope of this chapter, i.e. CoMP clustering.

Network coordination deals with inter-cell interference, reducing the interference especially at the cell edge, resulting in much needed additional capacity and increased UE throughput. By making use of the shared data between coordinating transmission points (CSI/scheduling/user data etc.), inter-cell interference can be mitigated or even exploited as meaningful signal at the receiver. In [93], authors show that more CoMP gains are achievable for cell edge users in scenarios where more interference is experienced. Similarly, more CoMP gains are presented for HetNet scenario where pico cells experience severe interference from macro base stations (MBS). Transmission points (TP) are different antenna ports of MIMO enabled cells which may or may not be located at the same place.

Coordination between TPs can be at different levels ranging from simple CS schemes to more complex precise coherent joint transmission/reception (JT/JR) CoMP [131]. 3GPP study mainly focuses on 3 CoMP schemes on downlink (DL) for LTE-A [6] based on backhaul bandwidth requirements and scheduling/precoding complexity. An illustration of downlink CoMP types is given in Figure 2.1.

1. Joint Transmission (JT):
   This type of coordination improves the signal quality and throughput for the user by simultaneous data transmission from multiple TPs in a time/frequency resource. Data transmission can be possible to a single user or multiple users from all or a subset of the TPs in the coordinating set. User data need to be available on all TPs serving to the same user, hence this type of coor-

dination require high backhaul bandwidth to share user data between the TPs. Joint transmission can be implemented coherently or non-coherently. In non-coherent transmission, user data is transmitted from multiple TPs with independent local precoding at each TP in the coordinating set. This is a simpler scheme which does not require CSI exchange and hence the absence of CSI delay problem. Multi-user non-coherent joint processing is depicted in Figure 2.2a. Coherant transmission is a more advanced scheme to further mitigate inter-cell interference where precoding is done centrally and global CSI knowledge is required. This scheme also require precise synchronization and CSI exchange between all TPs with minimal latency. In a typical multi-user coherant CoMP system, assuming T as the number of TPs in coordination and R as the number of users jointly served, a $R \times T$ virtual MIMO system is formed. Received signal for each UE_k in R can be expressed as:

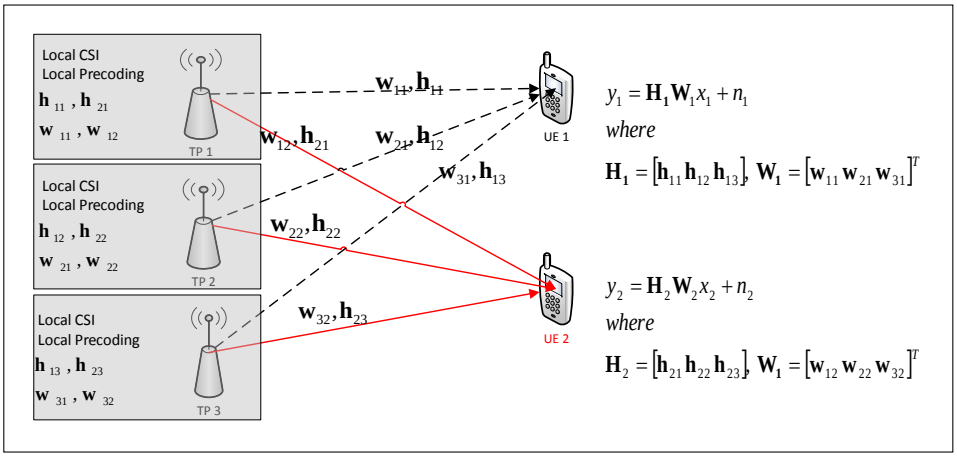
$$\mathbf{y} = \mathbf{HWx} + \mathbf{n}, \mathbf{H} \in \mathbb{C}^{R \times T}, \mathbf{W} \in \mathbb{C}^{T \times R} \tag{2.1}$$

where channel matrix can be expressed as $\mathbf{H} = \begin{bmatrix}\mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_R\end{bmatrix}^T$ and channel vector at UE_k is expressed as $\mathbf{h}_k = \begin{bmatrix}h_{k1} h_{k2} \dots h_{kT}\end{bmatrix}$. Similarly, precoding matrix can be expressed as $\mathbf{W} = \begin{bmatrix}\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_R\end{bmatrix}$ and beamforming vector for UE_k is expressed as $\mathbf{w}_k = \begin{bmatrix}w_{1k} w_{2k} \dots w_{Tk}\end{bmatrix}^T$. Received signal at UE_k can be expressed as:

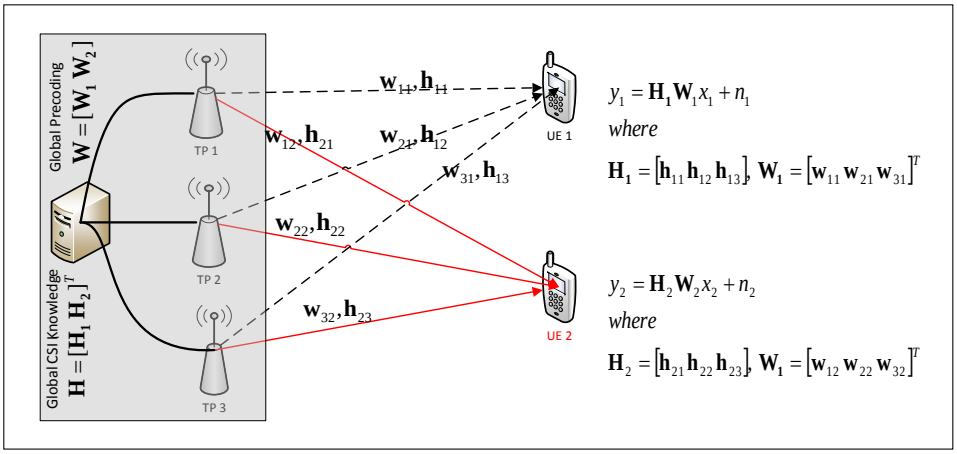
$$y_k = \mathbf{h}_k \mathbf{w}_k x_k + \sum_{i \in R/k} \mathbf{h}_k \mathbf{w}_i x_i + n_k \tag{2.2}$$

where first term represents the desired signal received at UE_k from all TPs in coordination and second term represents the interference received at UE_k, followed by n_k which represents the additive white Gaussian noise (AGWN) at UE_k. In the ideal scenerio of perfect CSI knowledge with minimal latency, second term representing inter-cell interference can be minimized to neglegable levels. An illustration of coherant transmission in the case of 3 TPs and 2 UEs is given in Figure 2.2b.

2. Dynamic Point Selection (DPS):
   User is served by only one TP in this CoMP type, however the serving TP dynamically changes at each subframe (i.e. 1ms) to the best preferred signal, exploiting the fast fading variations in the wireless channel. Similar to JT, user data needs to be available at each TP in the CoMP set, hence this scheme also require high backhaul bandwidth. Non-serving TPs in the



(a) Multi-user non-coherent joint processing.



(b) Multi-user coherent joint processing.

Figure 2.2: Coherent and non-coherent joint processing.

CoMP set can remain silent to improve SINR for the user [93]. Figure 2.1c depicts the DPS scheme between two cells where time/frequency resource at one cell is muted while the other cell is transmitting to the user.

3. Coordinated Scheduling/Beamforming (CS/CB):
   CSI and scheduling information is shared between the TPs in the CoMP set where scheduling and beamforming decisions are made centrally to reduce interference between the TPs. User data is only available at one TP and it is transmitted from the same TP while reducing interference to other TPs in the CoMP set. This scheme is a lighter version of JT where user data does not need to be shared between the TPs, hence backhaul bandwidth requirement is reduced. CB CoMP between two cells is depicted in Figure 2.1b.

A brief summary of each DL CoMP type, its associated challenges and benefits are given in Table 2.1.

Table 2.1: DL CoMP Types, Associated Benefits and Challenges

CoMP Type	Method	Challenges	Benefits
JT	User data/CSI is shared UE is served by multiple TPs	High backhaul bandwidth requirement	Desired signal from multiple TPs
DPS	User data is shared UE is served by one TP only at any given moment	High backhaul bandwidth requirement	Fast fading changes are exploited
CS/CB	Only CSI is shared UE is served by one TP only	Lower backhaul bandwidth requirement	Interference is reduced/eliminated

There are two main uplink (UL) CoMP transmission categories identified by 3GPP in [6]

1. Coordinated Scheduling/Beamforming (CS/CB):
   User scheduling and precoding design is done by coordination between the TPs however user data is only received by one TP.

2. Joint Reception:
   User data is received by multiple TPs jointly. Similar to downlink JT, uplink joint reception offers higher gains but with the cost of increased complexity and higher backhaul bandwidth requirement.

CoMP is one of the key features, standardized for LTE-A to uplift the network performance. 3GPP initiated a study item on LTE-Advanced in March 2008 and the requirements for radio interface enhancements are published in [1]. To satisfy these requirements, 3GPP published the physical layer enhancements in [2] where CoMP has been identified as one of the key features. A further feasibility study for CoMP in LTE-A is undertaken by 3GPP in Release 11 [6], where physical layer aspects of CoMP is studied. Simulation results from various sources are presented in this study where it is shown that CoMP can offer a significant performance improvement especially at the cell edge for different network deployment scenarios [6]. An overview of 3GPP Release11 CoMP techniques for LTE-A is presented in [159] where similar spectral efficiency gains are observed. CoMP is further enhanced in the following releases. In Release 12, impact of non-ideal backhaul with 5ms and 50ms backhaul delay is assessed for inter-BS CoMP [7]. Further enhanced CoMP scenarios are assessed in Release 14 showing simulation results for non-coherent JT-CoMP and also CS/CB CoMP with full dimension (FD)

MIMO [11]. More recently, self-organised dynamic CoMP clustering scenarios are presented in Release 15 [12] where backhaul latency and spatio-temporal traffic changes are taken into account to dynamically change CoMP clusters.

CoMP is also an essential part of 5G, delivering both capacity and ultra-reliable connectivity. 5G test-bed results from Qualcomm show how CoMP helps to exploit spacial multiplexing to increase capacity with MU JT-CoMP and also spatial diversity with same data being transmitted from each TP in the case of single user (SU) CoMP, where spatial diversity is exploited for ultra reliable connectivity in 5G [141]. CoMP is a promising inter-cell interference mitigation technology to provide better performance but it also have deployment challenges like backhaul capacity and latency requirements, BS synchronization issues and computational complexity. In [47], JT-CoMP performance is studied in dense SC deployment scenario where high inter-cell interference is expected without CoMP. A number of user-centric clustering algorithms are presented and backhaul reliability and capacity limitations are assessed. JT-CoMP provides a significant uplift in SINR and hence improve spectral efficiency but on the other hand, synchronization requirement between all SCs with JT-CoMP are highlighted as one of the major drawbacks especially in the presence of user-centric clustering solution. Influence of backhaul limitations, clock synchronization and imperfect CSI on CoMP performance is further evaluated in [156] and field test results are presented showing the impact of these limitations on achieved spectral efficiency. New emerging technologies like mobile edge computing (MEC) is proposed to reduce high backhaul capacity requirement for CoMP in [48]. Caching data on the MEC servers at the SCs will eliminate the need of popular data being transmitted from core network over the backhaul, reducing the backhaul capacity and latency requirements needed for CoMP.

Alongside with orthogonal multiple access (OMA) scheme employed in LTE-A and 5G, CoMP is also a key feature for non-orthogonal multiple access (NOMA) schemes envisioned for beyond 5G wireless networks. CoMP performance in heterogeneous ultra-dense networks is compared for both OMA and NOMA in [179] and it is shown that CoMP can significantly increase network performance for both OMA and NOMA schemes. In [160], authors propose downlink CB-CoMP for NOMA and OMA schemes and conclude that NOMA with CB-CoMP outperforms NOMA without CB-CoMP in medium to high SNR regime. Similarly, NOMA with CB-CoMP outperforms OMA with CB-CoMP in the same medium to high SNR regime and underperforms in low SNR regime due to the fact that NOMA with CB-CoMP does not have sufficient power for each user in low SNR. A similar outcome is presented for JT-CoMP in [107] where authors show that

NOMA with JT-CoMP significantly outperforms NOMA without coordination in a typical HetNet scenario. NOMA wit JT-CoMP in a two tier HetNet is studied in [19], where simulation results present a significantly better spectral efficiency for NOMA with JT-CoMP scheme when compared to OMA with JT-CoMP. In summary, CoMP is a promising feature for both NOMA and OMA schemes and performance of CoMP is higher for NOMA especially in high SNR range as further inter-cell interference is mitigated in the presence of non-orthogonal resource allocation.

2.3.1 Enabling Technologies for CoMP

The requirement for network densification for future cellular networks has initiated research on a number of new network architectures to optimize increased energy consumption, signalling and complex mobility management etc. These recently emerging radio access network (RAN) architectures will also help to overcome the challenges for CoMP (i.e. backhaul limitation, complex precoding, signalling etc.), enabling CoMP to be one of the main features of future wireless networks.

- **Control/Data Plane Separation Architecture (CDSA)** Motivated by proposed dense HetNet deployment and energy efficiency concerns, a control and data plane separation architecture (CDSA) is proposed for MBSs to provide coverage layer and handle most of the control signalling and SC layer under the MBS to provide the required data services [120,122]. CDSA is one of key enablers of CoMP implementation where MBSs can be enhanced to function as CoMP control unit (CCU) with strong backhaul links to the SCs within its coverage area. CCU functionality on the MBS can handle central precoding design, baseband processing and can make intelligent clustering decisions centrally within the SC layer, taking various efficiency metrics into account i.e. energy efficiency, load balancing, spectral efficiency etc. With all SCs connected to the associated MBS, there is no need for high bandwidth backhaul between the SCs in CDSA.

- **Cloud Radio Access Networks (C-RAN)** Another architecture envisioned for network densification is C-RAN where baseband processing unit (BBU) is decoupled from remote radio unit (RRU). A pool BBU is proposed in the cloud where there is high bandwidth front-haul between the cloud and RRUs [51, 119, 161]. Baseband resource sharing can be maximized and CoMP can easily be realized in this architecture [132]. Cloud can be enhanced to handle CCU function and make intelligent clustering decisions

for the connected RRUs. A BBU+RRU based CoMP example has been studied in [167] for LTE-A giving promising spectral efficiency gains as expected. The downside of C-RAN is the requirement for high bandwidth fronthaul [21]. Larger CoMP cluster size in C-RAN can be feasible with ideal fronthaul [55] due to centralized BBUs handling main CoMP functions. Concept of self organizing cloud cells is proposed in [24] where SCs within the coverage area of a MBS are connected to the MBS. MBS then handles the decision making on which SCs to be allocated for user data service to improve blocking probability, energy consumption and handover probability. This setup can also be easily extended to enable CoMP and enhance MBS to handle CoMP-CCU functionality. More recently CoMP in C-RAN is proposed as a key feature to mitigate inter-cell interference in densely deployed SC networks in [132] and a pilot reuse scheme is introduced to avoid high pilot overhead for CSI measurement.

2.4 SON as an enabler for CoMP Clustering

As discussed earlier, CoMP can only be realized within small cluster of cells due to its complexity which generally increases with the number of coordinating cells. Optimum cooperating cluster selection is key for maximizing the benefits of CoMP. Static clustering solutions will fail to respond to dynamic changes in the network and hence dynamic clustering solutions are required to adapt to changing spatio-temporal changes in the network and user profiles. Self-organised CoMP clustering algorithms are needed to form optimum clustering by reading various network data and making clustering decisions based on the changing conditions, maximizing the objectives like spectral efficiency, energy efficiency, backhaul availability and load balancing while keeping the fairness between the users.

In this section, we propose SON as the key enabler for dynamic CoMP clustering and give brief introduction on SON:

SON is an emerging concept in wireless cellular networks to automate some of the operational tasks in closed loop to overcome the challenges of a complex multi-layer network [22]. Network conditions are monitored dynamically by exploiting Big Data from various sources and intelligent algorithms are employed to effectively manage the network based on the changing local conditions. SON can also be utilized for predictive algorithms where network and user profile changes are predicted by employing machine learning techniques and optimization algorithms can proactively adapt to the changing conditions [91]. Dynamic CoMP clustering

can also be deployed within the SON platform as an enhancement to other SON modules which utilizes the Big Data for making proactive CoMP clustering decisions. A recent 3GPP technical report focuses on SON based CoMP clustering use cases to dynamically form clusters for spatio-temporal traffic changes and also based on backhaul availability [12].

SON algorithms can be designed as a distributed or centralized function depending on the requirements of the tasks, especially time and scalability limitations. Given the increasing complexity of the wireless cellular networks, SON will have a strong, enhanced presence in future networks. Future networks will need to deploy effective SON algorithms to improve capacity and QoS and reduce capital expenditure (CAPEX) and operational expenditure (OPEX) by reducing labour costs. SON has been an important part of 3GPP LTE/LTE-Advanced standardization starting from Release 8 and continued to evolve in ongoing releases [5, 8, 10, 13] and it is also considered as a key function for 5G by 3GPP where [16, 17] covers the technical specification for SON concepts, use cases for 5G in Release 16.

SON is mainly categorized in three folds:

1. Self Configuration:
   This group of SON modules aim to manage new entities integrated in the network. A considerable amount of operational expenditure (OPEX) and capital expenditure (CAPEX) is spent for new site configuration during network rollout and it will increase significantly with proposed massive deployment of SCs. Self configuration algorithms aim to automate new site configuration, initial automated neighbour relations and software updates [14].

2. Self Optimization:
   This group of SON modules aim to optimize ongoing services in the network. Self optimization algorithms will monitor network performance data and derive optimization changes in the network in open and/or closed loop, aiming to reduce OPEX costs and also improve network spectral efficiency, energy efficiency, network capacity and overall QoS. Dynamic CoMP clustering can be incorporated to self-optimization module set to derive closed-loop dynamic clustering decisions based on network data already available in the SON platform. Self optimization is an important part of LTE/LTE-A standardization [13] and there are already commercialized algorithms deployed in the current LTE networks. Self optimization tasks can be mainly grouped in three folds [22].

(a) Load balancing

(b) Coverage and Capacity Improvement

(c) Interference Control

3. Self Healing:
   This group of algorithms aim to detect faults in network elements, analyze the fault by gathering relevant information, diagnose and clear the fault. For time consuming fault restoration, self healing also aims to perform compensation actions on neighbour cells until the faulty cell is restored. 3GPP has standardized self healing for LTE/LTE-Advanced as an important feature of SON platform [15].

2.5 Clustering Taxonomy based on Self Organization

In this section, CoMP clustering algorithms in literature are critically discussed based on self organization. Three main clustering types are identified:

1. Static Clustering
2. Semi-Dynamic Clustering
3. Dynamic Clustering

A summary of clustering taxonomy based on self organization is given in Figure 2.3.

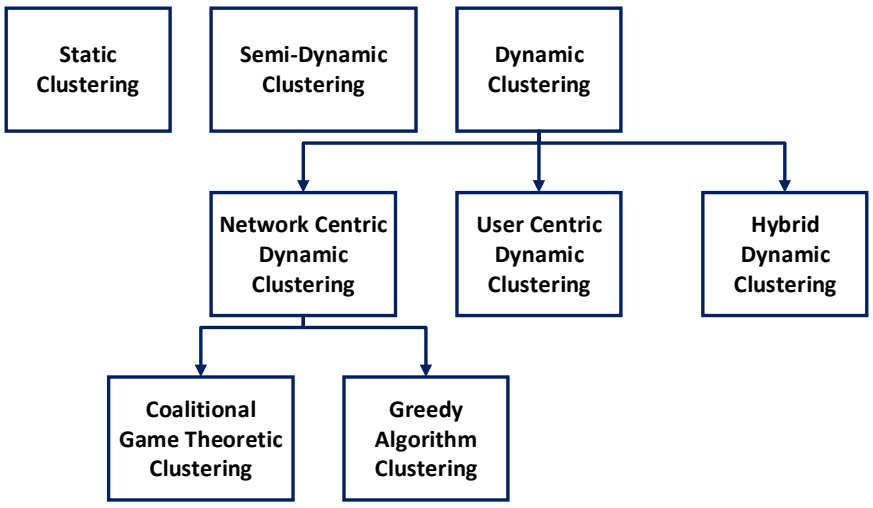


Figure 2.3: CoMP clustering taxonomy based on self organization.

Static clustering method is less complex with less signalling overhead but this method is not responsive to changes in the network nodes or user locations, hence the performance gains are limited. Semi-dynamic clustering is an enhanced version of static clustering where a number of static clusters are formed and employed dynamically to improve the potential gains. Complexity increases with additional signalling but performance is also improved when compared to static clustering. However, this method still lacks on truly responding to the dynamic changes in the network. Dynamic clustering methods are developed to respond to network and user mobility changes, i.e. new sites, sleeping cells, load changes etc. This scheme comes with increased complexity on scheduling and beamforming design but it gives the best results, reducing inter-cluster interference by moving the clusters dynamically. Dynamic clustering can be classified in three main categories within itself based on the approach. In network-centric clustering approach, all users in the same cluster use the same set of cells, however in user-centric clustering, users can be assigned their own clusters which comes with additional complexity. Hybrid approach combines both approaches which can be a good balance of complexity vs. performance.

In the subsequent subsections, we present an extensive literature review for each category and criticize available techniques based on complexity, scalability and potential spectral efficiency gains.

2.5.1 Static Clustering

CoMP clusters are formed in a static way, mostly based on topology and do not change according to changes in the network. This method offers a less complex solution which can be a good candidate to deploy in the initial phase of deployment. Static clustering within cells in the co-located site is the most basic and practical option which does not require data exchange between the sites, hence not reliant on fast backhaul.

The work presented in [20] propose a static clustering scheme, where sectors looking into each other are clustered to improve SINR. Authors assume a hexagonal grid in deployment which is non-realistic in real network deployments. This is usually the downside for most static clustering solutions. In [155], static intra-site and inter-site CoMP clustering is considered with orthogonal frequency reuse where antenna bore-sights are shifted to face into each other for extra CoMP gain. Dead-spots would be created with this new topology where SCs are proposed to fill in the dead-spots. CoMP and HetNet deployment are merged in this solution to identify locations for SC deployment, however an idealistic hexagonal grid is assumed again, which is unrealistic. A disjoint and overlapped static clustering

model is presented in [117] where static clusters are formed to maximize mean SINR or to minimize SINR outage at possible user locations. In the overlapped solution, one cell can be in three clusters where system resources are splitted into each of the three clusters. Presented solution is better than the clustering types based on regular patterns as it can apply to realistic network topology. However the proposed work is not scalable as the complexity of the solution increases with the number of possible user locations. More recently, in [45], static clustering is proposed in RN scenario with Poisson point process (PPP) distributed BS topology to analyze the performance of MU JT-CoMP for a given maximum cluster size. Increasing maximum cluster size and the number of antennas at each BS improves spectral efficiency however the increase is sub-linear and dependent on BS load. This study again eliminates the typical limitation of regular network deployment scenarios applied in static clustering solutions, however static nature of the algorithm lacks in responding to dynamic user/network changes.

A number of drawbacks for CoMP clustering have been investigated in [29]. Authors have investigated an inter-cell interference model in HetNet scenario with pico-cells to offload macro network. Time-domain resource partitioning is considered between the MBS and pico layer within the MBS's coverage area. A static CB-CoMP method is applied with centralized beamforming and scheduling for the cluster of all pico-cells and its connected MBS. CoMP failed to improve the performance further from enhanced inter-cell interference coordination (eICIC) due to the additional overhead required to implement CoMP i.e. mainly the UE-RS signal introduced with CoMP in LTE-Advanced. In [73], time synchronization limitation between coordinated cells is investigated. Authors have shown that time synchronization will need to be taken into account for a network with large inter-site distance (7km studied), however there is minimal inter-symbol interference (ISI) issues for inter-site distance of < 1 km due to cyclic prefix (CP) length.

In summary, static clustering is an attractive approach with its significantly less complexity for initial CoMP deployment for LTE-A networks. Intra-BS CoMP is a promising solution which eliminates the need for high backhaul bandwidth requirement between the BSs. On the other hand, inter-BS static clustering algorithms are mostly based on the assumption of hexagonal grid layout, which is not applicable to real networks. Furthermore, this method will fail to give the much needed spectral efficiency gains and increased system capacity for future 5G networks. Semi-dynamic and/or fully dynamic solutions are required to respond to changing network/user profile conditions and maximize CoMP gains.

2.5.2 Semi-Dynamic Clustering

Semi-dynamic clusters are more advanced than static clusters where several layers of static clusters are designed to avoid inter-cluster interference. More than one static clustering patterns are formed where users are able to select the most suitable cluster. This method also mostly relies on hexagonal grid network topology which is unrealistic in practical networks.

A two layer static clustering, based on regular network topology is proposed in [78] to extend on static clustering. This approach is then extended for several layers for dynamic clustering. It is proposed for users to pick one of the available clusters based on power. While the solution is an improved algorithm compared to static clustering, overlapping nature of the proposed algorithm adds to the scheduling complexity and require increased backhaul bandwidth. A semi-dynamic clustering scheme is introduced in [143] where static clusters are formed based on hexagonal grid topology and multiple shifted cluster patterns are created with different sub-channels allocated for each shifted cluster. A joint, centralized scheduling is developed for this clustering type. In [138], static cluster shift idea from [143] is further enhanced with "full shift" and different frequency bands are allocated on shifted clusters. Static clusters are formed to maximise neighbouring cells in the same cluster for a given hexagonal network layout. Shifted clusters reduce the inter-cluster interference, maximizing the CoMP gain, however solution is based on hexagonal grid topology which is not applicable to real networks. In [158], a semi-dynamic clustering scheme is proposed for downlink time division duplex (TDD) JT-CoMP scenario. Solution is based on large size (nine cells) static clustering and creating different static patterns of sub-clusters in each large static cluster. Dynamically selecting sub-clusters achieves almost as good as large cluster spectral efficiency but with reduced complexity. Inter-cluster interference between the large clusters is an important drawback in the proposed solution. Moreover, the solution is not able to adapt to network changes within the big static clusters i.e. new/sleeping cells etc. More recently, authors in [139] present a semi-dynamic clustering solution where multiple static clustering patterns are rotated in a hexagonal HetNet scenario to provide a less complex alternative to dynamic network clustering. However, provided solution gets more complex as the rotating speed increases. The static nature of the solution lacks on dynamically adapting to user profile/network changes.

In summary, semi-dynamic clusters are an improved version of static clusters with minimal overhead increase, however most solutions are based on idealistic hexagonal grid topology which is not realistic. Furthermore, majority of semi-dynamic algorithms propose orthogonal frequency allocation from each cell to

its assigned static clusters. Based on the utilization of dedicated bandwidth for each static cluster, proposed algorithms can reduce the overall spectral efficiency. Moreover, static nature of clusters is not able to respond fully to the spatio-temporal changes in user profiles and the network elements. Dynamic clustering algorithms is discussed in the next section which is mostly applicable to real network topology and can adopt to changing user profile and network conditions.

2.5.3 Dynamic Clustering

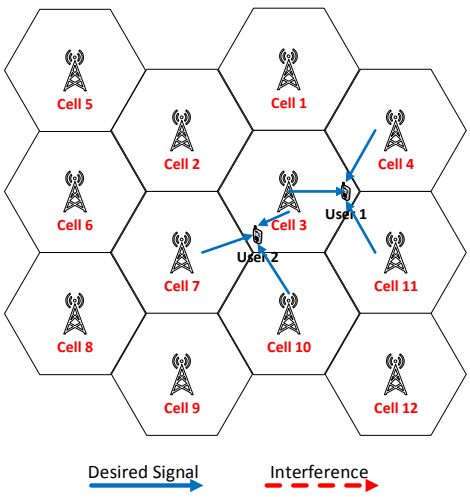
Dynamic CoMP clustering is more complex with increased signalling overhead but it is more responsive to the changes in the network. Inter-cluster interference can be minimized and cluster size for individual users can be optimized dynamically for an optimum balance. Dynamic CoMP clustering can be classified in three groups based on network elements considered for clustering:

1. Network-Centric Clustering
2. User-Centric Clustering
3. Hybrid Clustering

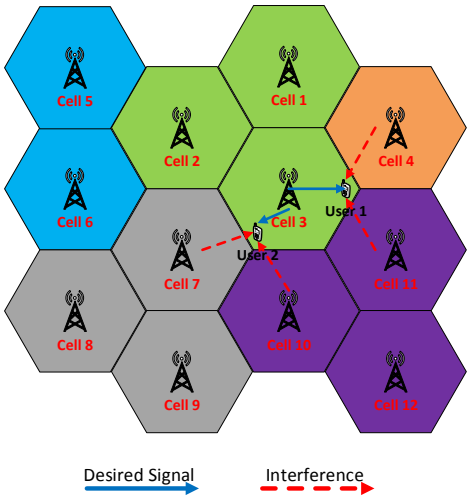
An illustration of the three types of dynamic clustering is given in Figure 2.4. CoMP benefits are illustrated for two sample users for an identical network with different clustering schemes. For example, user-1 is located at the edge of cell-3, receiving strong interference from cell-4 and cell-11. Network-centric clustering is the most limited scenario where user-1 is located at the edge of the cluster. Its cluster consists of cell-3 only and there is interference from cell-4 and cell-11. Hybrid clustering employs larger network-centric clusters, which improves user-1's cluster to cell-3 and cell-4. User-1's SINR is improved in this clustering type but there is still interference from cell-11. The most beneficial clustering scheme is the user-centric one where user-1's cluster consists of all three surrounding cells i.e. cell-3, cell-4 and cell-11. Although user-centric clustering seem to be most beneficial one, it comes with additional scheduling/precoding complexity and increased backhaul requirement. The three types of dynamic clustering are reviewed in detail in the subsequent subsections.

Network-Centric Clustering

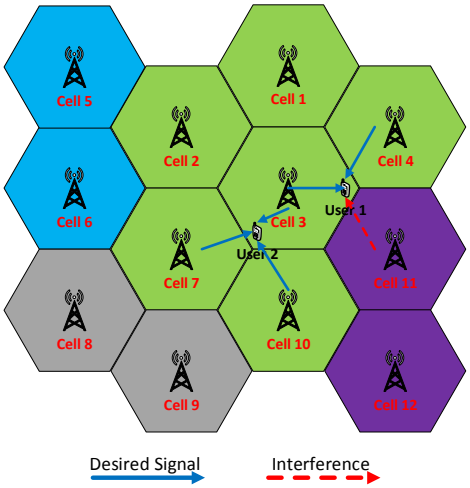
In network-centric clustering approach, cells are clustered in groups where all users within the serving area of a certain BS cluster are served by a sub-group of BSs in the cluster. A simple illustration of network-centric clustering is given



(a) User-centric clustering.



(b) Network-centric clustering.



(c) Hybrid clustering.

Figure 2.4: Dynamic CoMP clustering taxonomy.

in Figure 2.4b. It is less complex when compared to user-centric clustering, especially from scheduling point of view. However cluster edge users suffer from inter-cluster interference. Dynamic network-centric clustering can minimize this effect by moving cluster boundary dynamically.

We classify the methodologies followed to design network-centric clusters and review the solutions in literature as follows:

Greedy Algorithms Greedy algorithms are widely used for CoMP cluster formation in literature. Clusters are formed iteratively, starting from a randomly chosen BS to maximize the main objective, typically spectral efficiency. Best cluster is formed for the randomly chosen BS, maximizing the CoMP gains, however

the clusters formed in later stages of the algorithm suffer from sub-optimal clusters. It is relatively less complex but may not achieve as good results as the other methods, i.e. game theoretic clusters. A greedy uplink clustering algorithm is studied in [134] aiming to maximize spectral efficiency. It is shown that dynamic clustering with cluster size of two cells outperforms static clustering with much larger cluster size. A predefined fixed cluster size is proposed which is not the optimal solution for some clusters. A similar approach is employed in [108] but a dynamic cluster size is proposed. Authors have designed a dynamic clustering solution for UL multi-user distributed antenna system (MU-DAS), where one cell has a number of RRUs placed in the cell's coverage area with fast fiber connection to their cell. BSs are merged based on highest interference created to the other users. However, clustering takes only scheduled users into account at any point in time, hence not taking load into account for cluster formation. Starting the iterations from the highly loaded cells can improve the system throughput as the CoMP gains will be maximized for clusters formed in early stages of the algorithm. Also clustering is proposed to change with each scheduling interval in this solution [108] which increases signalling due to high frequency cluster changes. Both proposed algorithms in [108, 134] offer disjoint clusters where inter-cluster interference is still an important factor reducing spectral efficiency. An overlapping dynamic clustering is proposed in [58] to improve network average sum rate and fairness. A greedy approach is considered starting from a random BS. Authors have shown better results with cluster size of four with overlap size of two when compared to cluster size of eight with no overlap. The solution lacks scalability where large network size can lead to increased complexity. Overlapping clusters will also require more complex scheduling but overlap and cluster size parameters are introduced in the proposed algorithm to control this complexity.

Greedy algorithms provide lower computational complexity however lack on sub-optimal clusters especially for clusters formed at later stages of the algorithm. Shortcomings of greedy algorithm can be improved by employing coalitional game theory for cluster formation with a utility function to maximize system throughput and other key network objectives. Game theory can also provide distributed solutions with reduced signalling overhead as opposed to centralized greedy algorithms, however coalitional game theoretic algorithm's computational complexity is higher than greedy algorithms [70]. Coalitional game theoretic clustering is discussed in detail in the next paragraph.

Game Theoretic Clustering There is an increasing interest in applying coalitional game theory to design self-organised, distributed cooperative clusters. A

utility function is introduced to formulate the cost and CoMP gain trade-off for forming clusters. Proposed utility function can limit the cluster size dynamically based on BS locations and user profiles. Coalitional game theory can provide distributed, stable, converging solutions to maximize CoMP gains. An extensive tutorial on coalitional game theory for wireless communications applications is presented in [148].

In [99], authors proposed a dynamic network-centric clustering method employing a utility function to maximize the second best servers of the cell edge users in the same cluster. Cluster size is fixed to two only which leads to sub-optimal clustering for varying network conditions. Also network clustering formation is based on exhaustive search for collusion, hence not scalable, i.e. complexity increases with network size. Moon et al. have studied a dynamic cluster formation algorithm in [121] which merges cells into clusters based on the improvement on spectral efficiency, with configurable maximum cluster size and the minimum efficiency gain. This algorithm is semi-distributed where SINR measurements are based on pilot signal measurements but still need a CCU for cluster decision-making. It implicitly takes the number of users into account and hence clusters are formed based on cell load. Although a more flexible cluster size is introduced in [121] when compared to [99], algorithm still lacks on scalability as the complexity increases with the number of BSs involved. Walid et al. presented an application of a coalitional formation game for user clustering in the uplink, maximizing the sum-rate capacity with a cost function based on power requirements which is dependant on the distance between the users in [147]. Inspired by [147], authors in [70] developed a coalitional game theoretic clustering method where utility function dictates average cluster size and targets for higher spectral efficiency. It is a distributed algorithm which does not need a central entity and reduces signalling overhead. SINR at the cell edge is significantly improved when compared to a greedy algorithm. On the other hand, solution lacks on scalability where the cluster formation complexity increase with network size. Computational complexity of such algorithms can be reduced by limiting the candidate sites for coalition to neighbour cells only. Utility function for forming coalitional game theoretic clusters play an important role for optimal clusters. Utility function need to include a realistic model for the cost of cluster formation and the relevant CoMP gains. Dynamic cluster size can be self-imposed with accurate implementation of a utility function. Also multi-objective clustering can be implemented by including multiple metrics into the utility function i.e. energy efficiency, load balancing, spectral efficiency and backhaul bandwidth limitations.

Other Dynamic Network-Centric Clustering Algorithms A self organizing dynamic clustering method is presented in [168] where candidate clusters are formed from reported list of cells from users. CCU is proposed to arrange cluster solution by listing the candidate clusters with minimal cost, where the cost function takes into account the cluster size, number of users and reference signal received power (RSRP). This algorithm is a basic one where cost function can be improved to maximize SINR / spectral efficiency for more optimal solutions. It lacks on scalability with increasing complexity of handling high number of candidate clusters as the network size / number of users increase. Time averaged measurements from users is considered where fast fading is eliminated. Weber's algorithm [168] is further enhanced in [28] by replacing the cost function based on received power levels to a utility function with the aim of maximizing the weighted sum rate. Unlike [168], authors in [28] proposed a fast changing cluster design, responding to fast fading channel variations which will lead to increased signalling and possible ping-pong cluster re-selections. To reduce signalling overhead, cluster change frequency can be reduced to a wider time-frame and averaging algorithm can be used for user measurements which can eliminate fast fading variations. Authors in [135] use only macroscopic path loss for SINR estimation at each pixel of the network area and derive a clustering algorithm to maximize average SINR at all pixels. The solution is formulated as a binary optimization problem and solved with a standard solver like CVX [69].This solution eliminates the fast fading changes and can adapt to network changes, however it is not able to adapt clustering based on user profile changes as it optimizes all network area, rather than where users are located. Additionally the proposed solution is not scalable as the solution gets more complex in-line with growing size of the network. This clustering solution is the utilized in [136] where soft frequency reuse (SFR) and CoMP are employed together to improve cell edge user performance. An analytical framework is driven to optimize SFR parameters to maximize the overall cluster capacity, cell edge user throughout and required backhaul. A further dynamic network-centric clustering is proposed in [88] for CoMP-DPS to maximize the 2nd best serving cells based on SINR in the same cluster for each UE. Backhaul load is also taken into account for any cell to be granted into a cluster to improve backhaul load balancing. An online learning based dynamic CoMP clustering model is presented in [89] where clusters are formed based on historic received signal strength measurements from the UEs. Clusters which provide the maximum SINR are selected as the final clusters of the solution. Both works are not scalable to bigger networks as the complexity increases with the number of cells/users in the network. Mobility

data is utilized in some works to form the network clusters. A graph partitioning based dynamic clustering method is proposed in [44] where pairwise handover data and distance between the BSs are utilized to form DL CoMP clusters in a real network scenario. Proposed method provides a low complexity solution which can dynamically adapt to spatio-temporal changes in user/network profiles, and maximize spectral efficiency however other network objectives are not taken into account, like backhaul load, RAN load etc. More recently, Narmanlioglu *et al.* uses handover data to form network clusters in [123] which aims to include cells with higher handovers into the same cluster based on historic mobility data. This work can be utilized for CoMP clustering as well as other SON functionalities which may require network clusters. It provides a practical, low complexity, centralized solution, however the clusters will be slow to adjust into network/user profile changes as the cluster formation is based on mobility data and new mobility data will need to be available before any re-clustering decision is made to respond to network changes.

User-Centric Clustering

Users are allocated their own cluster of cells individually in user-centric clustering approach. Although this method can give better SINR/throughput gains, it is more complex, especially in terms of scheduling where user clusters overlap with each other. To reduce complexity, user-centric clustering can be implemented in small groups of cells rather than the whole network.

In [64], authors have studied macro diversity CoMP with dynamic user-centric clustering, comparing random network and hexagonal network topologies. It is shown that CoMP gives higher capacity results and bigger cluster size are required in random networks due to the random nature of BSs with more potential for inter-cell interference. Authors had no limitation on user-centric clustering which leads to complex scheduling between the BSs. To reduce complexity, user-centric clustering can be limited to groups of cells for easier scheduling, less signalling overhead and data exchange.

A three-tier clustering approach is presented in [175], wherein it has been proposed that cell center users will not use CoMP, users within the same site will use static clustering between intra-BS cells and a user-centric clustering is proposed for intra-site cluster edge users. Fixed cluster size is assumed which can lead to unnecessary complexity or less efficient coordination, depending on user location and SINR conditions. Similar complexity arises in works presented in [64] and [175] where no limitation is proposed on user-centric approach to any group of cells which will lead to higher complexity with a large number of

BSs cooperating at the same time. More recently, another user-centric clustering algorithm is presented in [100] with a scheduling and power control algorithm to improve throughput for the cell edge users. Authors propose to limit user-centric clustering with smaller network-centric clusters but static approach is considered for network-centric clustering. In [154], a user-centric clustering model is proposed for JT-CoMP in CDSA architecture where an optimum value of average received power difference is proposed for forming the user-centric clusters. Only two cells are allowed within the clusters and cell center users are proposed not to have CoMP enabled based on 2nd best server average received power levels. The solution takes radio resource management into account where CoMP users are allocated less resources due to higher resource allocation for these users but it statically allocates only two cells for each user-centric cluster which is sub-optimal for some users experiencing interference from multiple cells. Authors in [98] tackle this problem and provide a dynamic user-centric cluster size where the number of BSs at each user-centric-cluster is driven based on the relative distance of serving BS to the interfering BSs. A threshold for relative power levels is used to identify BSs in each user-centric cluster. User-centric cluster size is further optimized in [178] where authors consider the costs and gain trade-off for CoMP deployment and propose minimum number of cells in each user-centric cluster subject to QoS constraints.

User-centric clusters performance is compared to network-centric models in [18] where user-centric clustering is proposed in C-RAN architecture to eliminate inter-cluster interference in the case when network-centric clustering solutions are deployed. Significant performance gain is observed with user-centric clustering when compared to network-centric clustering however, the solution requires an ideal network with all baseband processing to be centralized and RRH units to be connected back to C-RAN BBU with fiber which is not physically possible or too costly to implement in some scenarios. Performance gain of user-centric clustering against network-centric clusters is further studied in [72], where HetNet CRAN scenario is taken into account and a biasing algorithm is employed for user association to offload MBS to SCs. CoMP is deployed to improve SINR for offloaded users from MBS. It is shown that user-centric clustering clearly provides better SINR, and hence better user data rates, on the other hand, it is worse for energy efficiency due to increased base-band processing complexity.

User-centric clustering approach is an ideal scenario to provide an upper limit of CoMP spectral efficiency gain, however it is not realistic due to increased complexity. Hybrid clustering is discussed next, which limits the user-centric approach to a group of BSs only to reduce complexity.

Hybrid Clustering

Hybrid clustering approach is the combination of network and user-centric approaches where users are allocated their own preferred cells but limited to a bigger group of cells which can be dynamically changing to adapt to changing network conditions. Hybrid clustering is driven from the complexity/throughput gain trade-off where user-centric clustering is used for better throughout but its complexity is kept at manageable levels by introducing network-centric clustering where users are limited to select cells only within the network-centric cluster.

In [182], authors developed a hybrid clustering method where a pre-defined network-centric clustering is used for cell center users and a number of pre-defined overlapping clusters are used for cell edge users to pick the best overlapping cluster to maximise SINR for the cell edge user. Inter-cluster interference on overlapping clusters is eliminated by orthogonal frequency allocation. Presented solution lacks on self organization as the pre-defined clusters are static, i.e. cannot respond changes in the network (e.g. new sites, sleeping cells). Although overlapping cluster patterns improve cluster edge user performance, orthogonal frequency use prevents the optimal use of the bandwidth. A simple downlink user-centric clustering is studied in [170] where users coordinate with two best serving cells according to the received power levels under a bigger static cluster. Proposed static network-centric clusters will suffer from high inter-cluster interference and also fixed user-centric cluster size can lead to unnecessary coordination, waste of resources and also possibly not being able to cancel severe interference from third best server for some users. A self-organised, dynamic network-centric clustering can reduce inter-cluster interference further and also dynamic user-centric cluster size can be employed for better performance. In [110], a hybrid clustering model for downlink SU-COMP is studied. Authors proposed static network-centric clusters and cell edge users are proposed to have user-centric clusters of fixed size of three within each network-centric cluster. Authors also presented a review of SU-COMP scheduling and a SU-COMP joint scheduling algorithm is provided for the proposed clustering scheme. The presented clustering scheme has low complexity, but further work is required to introduce dynamic network-centric clustering for improved cluster design. Fixed cluster size is also another shortcoming of the algorithm which can generate sub-optimal clusters.

In summary, dynamic CoMP clustering is a promising concept which can improve performance over static/semi-static alternatives. However, increased complexity and performance trade-off need to be evaluated for optimal solutions. User-centric clustering provides a theoretical upper bound for maximum performance gain but it requires complex precoding design, scheduling and in-

creased backhaul bandwidth [64, 175]. To reduce complexity, user-centric clustering solutions need to be limited to smaller network-centric clusters. Main approaches in network-centric clustering in literature are greedy algorithms studied in [58,108,134], coalitional game theoretic approaches deployed in [70,99,121] and more recently machine learning aided clustering algorithm based on mobility data [44, 123]. Greedy algorithms provide less complexity but results in sub-optimal clustering especially for the clusters formed at the later stages of the algorithm. Coalitional game theoretical clustering is a promising methodology which provides flexibility on implementing multiple network objectives and provides more optimized clusters when compared the greedy models. Moreover, coalitional game models can be implemented in a distributed way to reduce complexity. The key balance between additional complexity from user-centric clusters and the potential CoMP gains can be achieved by hybrid solutions where user-centric clustering is deployed within network-centric clusters [110,170,182]. However, hybrid solutions in current literature focuses either on dynamic user-centric approach with static network-centric clustering or dynamic network-centric clustering with no focus on user-centric clustering. Our work in Chapter 4 and Chapter 5 fills this gap with a dynamic network-centric clustering model where dynamic user-centric model is deployed within each network-centric cluster. Each of these algorithms target multiple network objectives to optimize, the final example in this thesis in Chapter 5 provides a clustering model to optimize backhaul load, RAN load and spectral efficiency collectively. A summary of CoMP clustering approaches based on self-organization and their shortcomings are provided in Table 2.2.

2.6 Dynamic Clustering Taxonomy Based on Objective Function

In this section, a novel CoMP clustering taxonomy is presented based on the objective function. The main objective of CoMP is to mitigate interference from neighbour cells and improve spectral efficiency in general but a more comprehensive approach is required to include other metrics/limitations for CoMP clustering. Backhaul bandwidth limitations for CoMP implementation and energy efficiency concerns for future wireless networks need to be included in comprehensive CoMP clustering algorithms. Moreover, with exponentially growing mobile data demand, better utilization of system capacity with load balancing will be a key concept which need to be taken into account for CoMP cluster design. Based on our detailed literature survey, main objective functions studied are:

Table 2.2: Summary of CoMP Clustering Approaches based on Self-Organisation

Clustering Type	Proposed Method	Shortcomings	Reference
Static clustering	Hexagonal grid topology, static clustering pattern	Non-realistic hexagonal grid approach, unable respond to network/user profile changes	[20], [155]
Static clustering	Mean/Outage SINR optimised, overlapping clusters	Increased complexity for bigger size network clustering	[117], [45]
Semi-dynamic clustering	Multiple static clustering patterns	Non-realistic hexagonal grid approach, Increased scheduling/beamforming complexity, increased backhaul bandwidth requirement	[78], [143], [138], [158]
Network-centric dynamic clustering	Greedy algorithm	Sub-optimal clusters which are formed in later stages of the algorithm. Fixed cluster size	[134]
Network-centric dynamic clustering	Greedy algorithm	Sub-optimal clusters which are formed in later stages of the algorithm. Cell load not taken into account	[108]
Network-centric dynamic clustering	Greedy algorithm-overlapping clusters	Sub-optimal clusters which are formed in later stages of the algorithm. Lacking scalability	[58]
Network-centric dynamic clustering	Coalitional game theoretic clustering	Exhaustive search, higher complexity, fixed cluster size	[99]
Network-centric dynamic clustering	Coalitional game theoretic clustering	High computational complexity, not scalable	[121], [70]
Network-centric dynamic clustering	Machine learning based	Not agile, single objective only	[44], [89], [123]
User-centric dynamic clustering	User-centric design	Higher complexity for beamforming design and scheduling	[64], [175], [98], [178], [18]
Hybrid dynamic clustering	User-centric clustering within static network-centric clustering	Static network-centric cluster design, not able to respond to dynamic changes in the network/user/service profiles	[170], [110]

1. Spectral Efficiency
2. Backhaul Optimization
3. Energy Efficiency
4. Load Balancing

A summary of CoMP clustering taxonomy based on objective function is given in Figure 2.5. In the following subsections, each objective function is critically discussed and extensive literature review is presented.

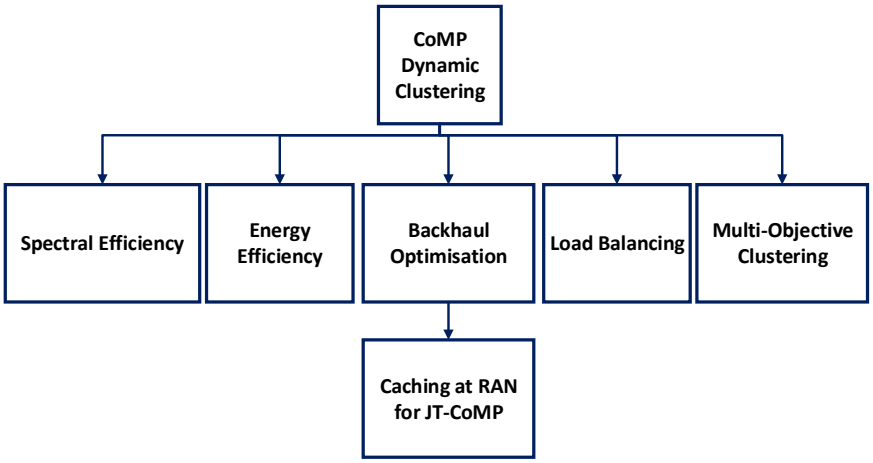


Figure 2.5: CoMP clustering taxonomy based on objective function.

2.6.1 Spectral Efficiency

Main objective of CoMP is to mitigate inter-cell interference within the cooperating cluster. Interference cancellation leads to better SINR and improved spectral efficiency. Cluster formation algorithms are designed to maximize spectral efficiency as a common objective, however other utilities such as backhaul bandwidth optimization, energy efficiency and load balancing have also been studied in the literature. Trade-off between spectral efficiency and other objectives for optimum clustering has been also in the interest of research community [90].

3GPP identified three CoMP deployment scenarios for LTE-Advanced and released a feasibility study, presenting simulation results from over 20 sources showing significant spectral efficiency improvements by deploying CoMP especially at the cell edge [6]. The most basic, intra-site static clustering is studied as scenario-1 and over 20% increase in spectral efficiency is observed at the cell edge with MU-MIMO JT-CoMP case [6]. Inter-site static clustering solutions

are employed in [20], [155] which is not able to respond to the dynamic changes in the network and user/service profiles, hence limiting the CoMP gains. Semi-dynamic clusters are proposed in [78], [143], [138], [158] where multiple static clustering patterns are designed to mitigate inter-cluster interference. This type of approach is more responsive to the dynamic changes of the network and user profile, however it still lacks on providing full spectral efficiency gain. Dynamic network-centric clustering methods can further increase spectral efficiency by dynamically changing CoMP clusters based on the spatio-temporal changes in user profiles and network elements. A game-theoretic, network-centric clustering approach is employed in [70]. Authors in [134] used a greedy clustering algorithm for uplink network-centric clustering to maximise spectral efficiency. More recently, machine-learning aided algorithms based on handover data are also used to form CoMP dynamic network-centric CoMP clusters [44, 123]. User-centric dynamic clustering approaches are studied in literature [64, 104, 175] which provide an upper bound on spectral efficiency gain but with increased complexity. Hybrid solutions reduce this complexity where user-centric clustering is limited only within a network-centric cluster [110,170,182]. Dynamic clustering solutions require more complex precoding design and scheduling, and increased backhaul. Complexity and additional requirements are reduced in semi-static clustering, and further simplified in static clustering with the cost of reduced spectral efficiency gain. An extensive critical review of CoMP clustering solutions based on static/semi-static/dynamic approaches and the trade-off between complexity and the additional spectral efficiency gains are provided in Section 2.5. A summary of different approaches and their shortcomings are presented in Table 2.2.

2.6.2 Backhaul Optimisation & Caching at RAN for JT-CoMP

As discussed in previous sections, one of the key requirement of CoMP is high backhaul bandwidth and low latency. Depending on the type of CoMP, backhaul requirement will vary. JT-CoMP will require more bandwidth due to user data being shared between cooperated cells. Huang *et al.* studied backhaul bandwidth requirements for network MIMO in [151] and concluded that backhaul requirement for CSI and scheduling information exchange is negligible when compared to user data sharing. Hence, JT-CoMP require much larger backhaul bandwidth than CS/CB CoMP. Backhaul requirement is also strongly dependent on user SINR and cluster size. Most users with high SINR will increase air interface capacity which will then increase backhaul demand. Biermann *et al.* have stud-

ied distributed JT-COMP feasibility in terms of high backhaul bandwidth and latency requirements especially in hotspot scenarios where certain backhaul links are more loaded than others [35]. In the proposed algorithm, all CSI is sent from cooperated cells to the serving cell where it is processed for precoding design and sent back from serving cell to the cooperating cells. Hence serving cell backhaul demand is more than other BSs in the cluster. Based on the backhaul load on each BS, a dynamic serving BS re-assignment algorithm is proposed by using "forced handover" to distribute backhaul load evenly. In [180], authors have designed a user-centric clustering strategy to minimize the backhaul data transfer for the JT-CoMP scenario where user data exchange between the BSs will be very high. An optimized number of links is proposed for a given CoMP cluster based on minimum SINR requirement of each user. Heuristic approach is used to reduce the links based on channel strength and "Signal to Leakage" (SLR) ratio (i.e. taking signal power and also the interference caused to other users into account). Authors have further improved this design in [181]. An optimization problem is formulated and approximate results are obtained by convex relaxation. An iterative algorithm is followed to further reduce the number of BSs in each user's cluster. A control unit (CU) is proposed for the semi-distributed solution where each BS is connected and share CSI with CU. Limited backhaul capacity constraint is further studied in [124] where limited backhaul capacity and per-BS power constraints are taken into account for a transmit precoding design to maximize sum rate. Authors in [71] aim to minimize the total transmission power by optimizing user-centric RRH clusters with their precoding and transmission power in a CRAN architecture with limited fronthaul capacity while maintaining user's QoS and fronthaul capacity constraints. In [116], CS-CoMP deployment feasibility under different backhaul infrastructures are analyzed in terms of convergence delay in exchanging scheduling information between base stations over X2. It identifies star topology as the best solution in terms of deployment cost and convergence delay where each SCs is connected to a central unit, possibly at MBS. Same authors further enhance this work in [115] and propose a bandwidth allocation scheme to prioritize X2 traffic over S1 traffic for CS-CoMP and hence reduce convergence time i.e. scheduling information exchange latency when TDM PON is used as backhaul in a 5G network. These works study backhaul limitation in isolation, but other network objectives like radio resource load and spectral efficiency is not taken into account collectively. On the other hand, higher complexity of implementing a user-centric clustering approach for CoMP makes these solutions not scalable for big networks. User-centric clustering can be employed within the a bigger network-centric cluster to limit computational

complexity and backhaul capacity demand.

Caching at the RAN for JT-CoMP

There is an increasing interest in the research community to explore potential benefits of MEC and caching popular multimedia content closer to the user to reduce high backhaul requirement for CoMP [48]. A significant amount of network data usage is due to duplicate downloads of few popular multimedia content from Netflix, Youtube, Facebook etc. Caching the popular content at various points in the network, i.e. RAN, core network or even the user devices can reduce the high backhaul requirement and give opportunity for JT-CoMP deployment, where high backhaul capacity is not available [46]. Furthermore, caching closer to the user can improve overall energy efficiency. A recent study on an operational 4G network shows [142] that 73% of the data volume is cachable and 54% of the cachable data is revisited, so significant gains are possible with caching.

In [105], caching at the BS is proposed and an opportunistic cooperative MIMO is employed without high backhaul requirement. Cells within the same cluster are proposed to cache identical data aiming to be employed for MU JT-CoMP. For users where requested data is available at the cache, JT-CoMP is proposed. If the requested data is not available at the cache, CB-CoMP is proposed where user-data exchange between the BSs is not required but CSI exchange is still employed for joint precoding. Authors presented a JT-CoMP solution in a limited backhaul capacity scenario in [56] where BS data caching is introduced to reduce required backhaul capacity for user data, hence increasing available backhaul capacity for CSI sharing. Improved backhaul availability for CSI sharing improves the accuracy in CSI knowledge at the central node, resulting in better precoding, hence improved interference cancellation. In [174], data caching-aware user-centric clustering model is presented for JT-CoMP to minimize backhaul traffic. In a similiar work in [77], Hou *et al.* present a user-centric clustering model where data catching is employed to reduce backhaul requirement showing that limited backhaul capacity and data caching have a significant impact on CoMP performance. In [49], a further user-centric CoMP clustering model based on cached data at SCs is proposed for optimum user association to reduce backhaul traffic demand and increase network throughput jointly for a given maximum cluster size. Zhou *et al.* propose to categorize data content into popular and and a less popular set and a cache placement strategy is driven for DL CoMP to minimize the outage probability [185]. In [173], mm-wave mesh backhaul links between SCs are proposed where few of the SCs have a wired backhaul to the core-network via MBSs. Each SC stores popular files in their cache and proposed

solution employs JT-CoMP where identical user data is jointly transmitted from each SC serving to the user. In the case when the data is not cached on the SC, this data is transmitted to the SC via the mm-wave backhaul links from another SC where the data is cached. Proposed solution aims to reduce the data traffic on the mm-wave mesh backhaul by optimizing the selection of the SC group to serve to each user and the routing paths between SCs to deliver the missing data to each SC.

Realization of CoMP depends on high backhaul bandwidth availability, hence CoMP clustering algorithms need to take this limitation into account. Caching popular multimedia proves to be one of the ways to reduce backhaul bandwidth requirement for CoMP realization. Cluster size and type of cooperation are other factors that can change the backhaul bandwidth requirement. Furthermore redistribution of backhaul data transfer to less-loaded cells can be deployed for better CoMP gains.

2.6.3 Energy Efficiency

Energy efficiency has recently become an important topic for wireless networks for both economical and environmental reasons [40, 62, 113]. In mobile communications, more than 80% of the power is consumed in RAN, especially BSs [113]. As briefly discussed in the introduction section, network densification, massive MIMO and CoMP are some of the key tools envisioned for 5G to meet ever increasing traffic demand which will severely increase energy consumption and OPEX costs. New architectures like CDSA [120] and C-RAN [51, 119, 161] have been envisioned to enable energy efficiency and reduce OPEX and CAPEX costs in future wireless networks, mainly by providing SC coverage only when it is required. Enabling CoMP will also improve energy efficiency [144]. It has been in the attention of research to design CoMP clusters to maximize energy efficiency and to optimise the trade-off between spectral efficiency and energy efficiency. On one hand, CoMP can reduce cell/UE output power for a given QoS but there is also additional energy consumed for additional signal processing and backhaul requirement.

CoMP clustering can be optimized for energy efficiency by increasing the number of sleeping BSs and/or their sleeping duration. In [74], BS sleeping with CoMP has been studied for energy efficiency with static clustering and assuming one cell is sleeping on each cluster during off-peak hours. A joint sub-carrier and power allocation algorithm is proposed to minimize the power requirements for coordination and compensate for sleeping cell for a given QoS. Cao et al. in [42] has compared the energy efficiency gains between CoMP and wireless relaying by

maximising the number of sleeping cells. Based on the traffic demand, it is shown that, energy efficiency gains are almost constant in lightly loaded traffic conditions where network is mainly coverage limited. In high traffic load, there is almost no energy gains possible, whereas in "energy efficient region", dynamic energy efficiency algorithms can provide bigger energy efficiency. As BS density increases, the "energy efficient region" also increases and the region for larger CoMP gains decreases. In [52], user-centric CoMP clustering for all cells within 3dB window is studied for cell switch-off on lightly loaded cells to improve energy efficiency. It is shown that unnecessary increase in cluster size and imperfect channel knowledge can lead to energy inefficiency. Up to 24% more energy efficiency is observed in perfect CSI when compared to imperfect CSI conditions.

More recently, Zeng *et al.* propose energy efficient CoMP scheduling for green CRAN architecture in [176] where RRHs are powered by both distributed renewable resources and traditional power grid. To reduce non-renewable energy consumption, an algorithm is proposed to optimize the number of RRHs to be activated for a given QoS constraints for each user. In [177], CoMP CB scheme is proposed for improved energy efficiency and show that energy efficiency is sensitive to the number of SC and MBS deployed, QoS constraints and massive MIMO deployment. A BS sleeping mechanism is developed in [92] for energy efficiency where CoMP and antenna tilt have been employed to maximize energy efficiency and maintain the same level of QoS and network performance. The number of sleeping SCs are maximized especially when there is less traffic demand during the quieter times of the day at night. In [96], Leng *et al.* derives an energy efficiency metric to find the optimum BS density to maximize energy efficiency with a coverage probability constraint in 3D mobile network model. Authors also present CoMP and sleeping cell concept as key tools for further energy efficiency.

CoMP helps to reduce inter-cell interference and can provide increased number of sleeping cells for a given QoS, however additional power consumption is introduced with CoMP due to increased backhaul usage. Popular data caching at the BS is a promising concept to reduce backhaul load as discussed in the previous section and it can also be useful to reduce additional power consumption due to increased backhaul demand. In [50], Chiang *et al.* optimizes energy efficiency by employing CoMP with BS sleeping alongside with data caching. First BS clusters are formed to maximize BS sleeping for a given QoS and then shareable caching is deployed to minimize backhaul power consumption.

Besides BS switch-off, deployment costs can be reduced and energy efficiency can be maximized by taking network coordination into account at network planning stage. In [125], a BS planning scheme is proposed to reduce the total number

of required BSs for a given coverage and traffic QoS by deploying CoMP. A SU-MIMO CoMP scheme with user-centric clustering method is employed to choose the optimal BS locations for deployment from a number of candidate BS locations to maximize energy efficiency. A typical example of this work is to reduce the number of BSs required from three to two BSs where some users cannot be served without the third BS if CoMP is not employed.

Deployment of CoMP and realization of future network architectures like CDSA and C-RAN will enable energy efficiency by increasing the number of sleeping cells. However, most studies in the literature are lacking the load conditions in the network but concentrate on coverage requirements only for BS switch-off to maximize energy efficiency. With predicted mobile data growth, network capacity will be under pressure and will require to be managed more intelligently. BS switch-off with CoMP clustering algorithms will need to include data demand and available capacity in the network. Hence, a more comprehensive approach for dynamic CoMP clustering should optimize energy efficiency and load balancing jointly. We discuss CoMP clustering in relation to load balancing in the next subsection.

2.6.4 Load Balancing

Load balancing has always been an important topic for cellular networks due to non-even distribution of user traffic, resulting in some BSs overloaded whereas other BSs not fully utilized. Network planning process takes traffic distribution into account and BS locations are planned accordingly, however unpredictable nature of user activities like traffic accidents, mass events etc. still cause overloaded cells. With ever increasing mobile data traffic globally, it is expected that the number of connected devices will increase up to 100 billion by 2030 [157]. The new use cases with 5G i.e. enhanced mobile broadband, massive machine type communications and ultra-reliable and low latency communications will enable a wide range of new applications be widely available with rapid deployment of 5G [130]. Data intensive applications such as virtual reality and gaming will be more widely used [57]. To meet this increased capacity demand, load balancing becomes even more important in future cellular networks. Various load balancing schemes have already been studied in literature [25, 79, 172] for cellular networks however there is very limited load balancing studies with CoMP deployment.

CoMP is likely to be deployed in dense deployment areas where there is severe inter-cell interference and these areas typically form hotspots at certain times of the day due to special events, accidents etc. Load balancing must be considered as one of the key network objectives for CoMP deployment to resolve load imbalance

in hotspot areas and improve customer satisfaction.

In our work presented in Chapter 3 (also published in [32]), we proposed the first load aware CoMP clustering algorithm in literature where cell-load is taken into account to design user-centric CoMP clusters for MU JT-CoMP scenario. UEs at the cell edge for highly loaded cells are transferred to lightly loaded cells to improve load balancing. Further work followed our study where Liu *et al.* proposed a load aware JT-CoMP scheme in C-RAN architecture and utilize sleeping BSs to improve coverage probability and enhance ergodic link throughput [106]. In [109], a two step load aware CoMP clustering solution is provided for non-coherent macro diversity (MD) CoMP, where a load aware user-centric clustering solution is developed by using game theory in the first step and then an inter-cell resource scheduling algorithm is provided for the given clusters in the second step. A further user-centric clustering algorithm is developed in [43] for MD-CoMP to improve load balancing while keeping spectral efficiency at comparably high levels. Shift factor and edge factor metrics are driven to minimize spectral efficiency reduction and improve fairness for cell edge users. Both works considered MD CoMP only to reduce complexity, however MD CoMP uses additional dedicated resources from each BS within the cluster for the same user, hence spectral efficiency gain has to be much higher than the additional resource demand to be able to make the solution spectral efficient. When coherent MU JT-CoMP is utilized, multiple users can be scheduled within the same resource, and hence additional spectral efficiency/load gain is expected, however full user-centric clusters are too complex for MU JT-CoMP due to additional processing required for central precoding and scheduling. To limit this complexity, smaller network-centric clusters are required to enable MU JT-CoMP within these BS clusters only. In our work presented in Chapter 4, we proposed a novel, low complexity network-centric clustering solution for MU JT-CoMP, utilizing coalitional game theory to optimize load balancing and spectral efficiency jointly. This work is published in [31]. We further enhanced this study to take backhaul load conditions into account for CoMP clustering in Chapter 5 where a two step coalitional game framework is presented to form load aware BS clusters and adjust user clusters to optimize RAN load and backhaul load and spectral efficiency jointly.

In summary, CoMP will introduce spectral efficiency gain and increased throughput especially at the cell edge [82]. Additional capacity from CoMP can be utilized for load balancing through dynamic CoMP clustering based on cell load. In recent years, self-organised load aware CoMP clustering algorithms have been in the interest of research community but more work is required especially to design CoMP clusters not to optimize load balancing only but to include a number of key

objectives to form multi-objective clusters. Multi-objective clusters are discussed in the next subsection.

2.6.5 Multi-Objective Clustering

As seen in aforementioned subsections, dynamic CoMP clustering can aim to improve not only spectral efficiency but also other objectives like energy efficiency and load balancing. Most of the CoMP clustering works in literature deals with one network objective as reviewed in previous subsections and but there is also few works which focus on optimizing two objectives jointly. In [90], authors have compared a number of static clustering options for trade-off between throughput and energy efficiency in sparse, medium and dense deployment scenarios. They have identified transmit power, inter-site distance and SINR service demands as the main inputs for this trade-off. Li et al. proposed a dynamic CoMP clustering algorithm with BS sleeping to maximize energy efficiency while maintaining high achievable rate for all users [102]. Candidate clusters are formed by all possible combinations of groups of cells with predefined cluster size and each BS selects a suitable cluster from the candidate clusters by maximizing achievable rate for its users. Developed algorithm then looks for cell load and moves users from cells with light load onto other clusters to increase the number of sleeping cells and hence better energy efficiency. Proposed clustering algorithm lacks on scalability as number of candidate clusters increase with network size, leading to high computational complexity. Moreover, proposed algorithm fails to look at total system capacity and load balancing aspects, i.e. BS sleeping decisions will need to look at not only coverage limitations but also incorporate the network capacity against demand to make sure there is no BS congestion issues once a neighbour BS is switched off. Hence, energy efficiency and load balancing will need to be jointly optimized for an improved multi-objective CoMP clustering algorithm. Furthermore, backhaul load also needs to be taken into account for BS sleeping decisions to make sure backhaul capacity from the remaining active BS can provide the required backhaul capacity. Again, a number of works in literature deal with backhaul availability problem for CoMP clustering [115, 116, 124], implications of backhaul channel reliability on spectral efficiency of the CoMP clusters [118] and more recently caching user-data at the RAN is employed to reduce backhaul requirement for CoMP [77, 173, 185], however these studies lack on a comprehensive approach to jointly optimize multiple objectives. A multi-objective optimization problem is setup in [137] to optimize overall network capacity, cell edge performance and backhaul requirements jointly in CoMP network with SFR. Inspired from [37], authors define a scalar objective function which combines all objec-

tives into one by taking weighted arithmetic mean of all objectives where weight figures can be configured depending on the priority of each objective. This objective function is then employed for SFR design a fixed clustering scheme is used, lacking on a multi-objective clustering model.

Existing literature focuses on one limiting objective and investigates the trade-off against spectral efficiency gains. However, a more comprehensive CoMP clustering approach need to take all limiting factors in the same algorithm for intelligent clusters which jointly optimize backhaul bandwidth, energy efficiency, load balancing and spectral efficiency. For example, BS switch-off is a widely studied concept in literature as part of CoMP clustering, however only the SINR constraints are taken into account to make sure there is enough coverage for BS switch-off. However, other constraints like RAN capacity, load balancing, backhaul bandwidth availability also need to be considered in a realistic network for BS switch-off decision. Our work presented in Chapter 5 provides a clustering solution to jointly optimize backhaul and RAN load alongside with spectral efficiency, however more research is required for multi-objective CoMP clustering algorithms with above mentioned constraints. A comparison of CoMP clustering algorithms based on aimed objective and their shortcomings are provided in Table 2.3.

2.7 Conclusions and Outlook

This chapter provides an extensive survey on CoMP clustering methods for future cellular networks. We first give the motivation for CoMP for future wireless networks and briefly provide an outline of CoMP implementation challenges and the need for CoMP clustering. We then provide a section to give brief tutorial about different types of cooperation, associated challenges and propose network architectures like CDSA and C-RAN which will enable CoMP implementation. We then introduce SON as a key framework for much needed dynamic CoMP clustering algorithms. The core of the chapter provides an extensive survey on CoMP clustering techniques available in the literature and introduce two novel taxonomies for CoMP clustering algorithms based on self-organization and aimed objective function.

Firstly, we provide a CoMP clustering taxonomy based on self organization, and critically discuss static, semi-static and dynamic CoMP clustering works in literature. Dynamic clustering algorithms are further divided to network-centric and user-centric clustering solutions and different approaches are discussed, their benefits and shortcomings are highlighted.

Table 2.3: Summary of CoMP Clustering Approaches based on Objective Function

Clustering Objective	Proposed Approach	Impact / Shortcomings	Reference
Spectral Efficiency	Dynamic clustering as summarised in Table 2.2	As summarised in Table 2.2	[134], [178], [58], [18], [121], [70], [64], [98], [170], [110]
Backhaul Bandwidth Optimization	Dynamic serving BS reassignment	Re-distribute backhaul load from serving cell to cooperating cells	[35]
Backhaul Bandwidth Optimization	Minimise cluster size based on min. SINR requirement	Reduced backhaul requirement, however spectral efficiency is sacrificed.	[180], [181], [115], [71], [124]
Backhaul Bandwidth Optimization	Caching at the BS. Switch between CB/JT CoMP based on backhaul availability	Reduced backhaul bandwidth requirement by caching popular multimedia at the BS	[105], [56]
Backhaul Bandwidth Optimization	Caching at the BS. Exchange cache data between BS via mm-wave backhaul	Reduced backhaul bandwidth requirement by caching popular multimedia at the BS	[173]
Backhaul Bandwidth Optimization	Caching at the BS. Cache-aware user-centric clustering	Reduced backhaul bandwidth requirement by caching popular multimedia at the BS	[174], [77], [49]
Energy Efficiency	CoMP clustering to maximize BS switch-off	CoMP clustering to switch-off lightly loaded cells for better energy efficiency. Only coverage constraints are considered, network load constraints need to be jointly optimized	[74], [42], [52], [176], [92], [96]
Energy Efficiency	Minimize number of BS deployment by employing CoMP	CoMP clustering to reduce the number of BSs required for deployment for better energy efficiency and cost saving. Only coverage constraints are considered, network load constraints need to be jointly optimized	[125]
Load Balancing	Load aware user-centric clustering	Higher complexity, not scalable user-centric clusters Network-centric clusters are required for lower complexity, scalable solutions	[109], [43], [32]
Multi-Objective Clustering	Energy efficiency without BS switch-off and Spectral Efficiency jointly optimised	Energy efficiency by deploying CoMP without BS switch-off. Comparing different CoMP static clustering schemes for energy efficiency/spectral efficiency trade-off.	[90]
Multi-Objective Clustering	Energy efficiency with BS switch-off and Spectral Efficiency jointly optimized	Energy efficiency by deploying BS Switch-off with CoMP while maximizing spectral efficiency	[102]
Multi-Objective Clustering	Backhaul Bandwidth and Spectral Efficiency jointly optimized	Effect of backhaul channel reliability on spectral efficiency for CB/JT CoMP. Reduced backhaul requirement by caching at the RAN.	[118], [115], [116], [56]

Secondly, we present a novel CoMP clustering taxonomy based on the objective function. CoMP clustering algorithms aiming for spectral efficiency, energy efficiency, backhaul optimization and load balancing are extensively discussed. More focus is given on comprehensive multi-objective clustering, available works in literature are presented, shortcomings are identified in detail.

Based on this extensive survey, key lessons learnt are summarized below:

- CoMP is a key part of future wireless networks to mitigate inter-cell interference in interference-limited densely deployed SC networks envisioned for 5G. However CoMP comes with significant overheads depending on the type of CoMP deployment. This limits the CoMP implementation to small clusters only. Optimum CoMP cluster design is key to be able to maximize CoMP gains.

- Future networks will be more complex with multiple layers and even with cells where operators will not have much control over (e.g. user-deployed cells). Hence, network elements and user profiles change dynamically and truly realistic CoMP clustering models need to be dynamic to be able to respond to spatio-temporal changes in the network and users. SON platform provides a promising framework to design dynamic CoMP clusters.

- Multiple network objectives are taken as primary objective for CoMP clustering models studied in literature, however there is limited work considering multiple objectives and optimize them jointly. A realistic CoMP clustering model will need to take all key objectives into account like backhaul availability, energy efficiency, spectral efficiency and load balancing as key objectives.

- CoMP is likely to be deployed in interference-limited areas where there is dense population and there are inevitable hotspot areas at certain times of the day at specific locations. Some cells will be highly loaded where others will be under-utilized. Load balancing appears to be a key metric to be optimized in CoMP clustering design too. We make the first attempt to fill the gap in literature with CoMP clustering design which optimize load balancing alongside with spectral efficiency. A load aware user-centric CoMP clustering work presented in Chapter 3 where user-centric clusters are formed to maximize spectral efficiency and load balancing objectives jointly. This work is followed by Chapter 4 where a novel network-centric, load aware clustering model is presented. This work provides a scalable solution for the potential complexity and scalability issues with user-centric clustering.

- CoMP requires a high backhaul bandwidth and low latency. This is a key dependency especially with JT-CoMP but also applies to other types of CoMP at a lower scale. A realistic CoMP design will need to take backhaul availability into account to maximize CoMP gains. Backhaul availability is considered for CoMP clustering design in literature, however a holistic approach has been missing where backhaul awareness is taken into account along side with other key metrics. In Chapter 5, we attempt to fill this gap with a clustering design to optimize RAN and backhaul load alongside with spectral efficiency.

- CoMP deployment and intelligent clustering solutions can improve energy efficiency especially with increasing the number of sleeping BSs [92,96,176]. BS sleeping has been employed in most works to improve energy efficiency, however only SINR constraints are taken into account for BS sleeping to make sure there is coverage available for all users. As discussed in Section 2.6.5, other constraints like system capacity and backhaul bandwidth will need to be taken into account for BS sleeping. A future open research area is to study a more realistic clustering approach where both RAN and backhaul load balancing are considered while making decision for BS switch-off with the aim of maximizing energy efficiency.

- CoMP clustering models in literature focus on reactive design where CoMP clusters are adapted into the network/user profiles changes after it happens. Reactive solutions will inevitably cause delay in adapting the clusters into the spatio-temporal changes in the network/user profiles which then will provide sub-optimal performance. We envision proactive CoMP clustering to accommodate much faster response rates required for 5G. Historic mobility data has been utilized in few studies for cluster formation in some recent works in [89, 123], however more intelligent solutions can be studied, Big Data in the context of wireless networks can be utilized to empower machine learning based algorithms to form prediction based CoMP clusters.

Chapter 3

Load Aware Dynamic CoMP Clustering

3.1 Introduction

In this chapter, we introduce a load aware, user-centric clustering model for JT-CoMP as a first attempt in literature to include load balancing as one of the primary objectives in CoMP deployment scenario. We first develop a self-organizing, user-centric CoMP clustering algorithm, maximizing spectral efficiency for a given maximum cluster size. We then further develop this clustering algorithm for load awareness and present a novel re-clustering algorithm in two stages. In stage-1, maximum cluster size is allowed to increase further for highly loaded cells to introduce more capacity in the system. A novel re-clustering algorithm is presented in stage-2 to distribute traffic from highly loaded cells to lightly loaded neighbours for MU JT-CoMP case. We show that unsatisfied UEs due to high load can be significantly reduced with minimal impact on spectral efficiency. Clustering with load balancing algorithm exploits the capacity gains from increase in cluster size and also the traffic shift from highly loaded cells to lightly loaded neighbours.

The rest of the chapter is organised as follows. In Section 3.2, we discuss the load balancing problem and present existing literature. Our system model is presented in Section 3.3. Our dynamic user-centric clustering algorithm is presented in Section 3.4. We further enhance the our user-centric clustering in Section 3.5 and introduce a re-clustering algorithm to take load balancing into account to distribute the load evenly to lightly loaded cells. In Section 3.6, we present results from our simulation and Section 3.7 concludes our work with the outcome and further discussion. Part of this work is published in [32].

3.2 Related Work

Mobile network operators experience an exponential increase in mobile data traffic and it is expected to continue with the wide range of additional services and devices becoming available in 5G. A 1000-fold capacity increase is projected for the next decade for 5G [101]. Given the very high capacity requirement, load balancing becomes even more important in future cellular networks. On the other hand, to cope with high capacity demand, ultra-dense small cell networks are envisioned where inter-cell interference will be severe. To mitigate inter-cell interference, CoMP is identified as a key feature for LTE-A by 3GPP [6] and it is likely to be a key feature for 5G [101]. As discussed in Chapter 2, CoMP and the associated clustering problem has been studied extensively in literature and various network objectives like spectral efficiency, energy efficiency and backhaul availability are optimized for CoMP clustering. However, load balancing in CoMP networks is a new concept and there is only very recent interest in CoMP clustering solutions aiming to optimize load balancing objective. Various load balancing schemes have already been studied in the literature for traditional networks without CoMP [25,79,172]. We make the first attempt in literature to provide a novel load aware, user-centric CoMP clustering model for JT-CoMP in this chapter exploiting the additional capacity from increased cluster size and load shift from highly loaded cells to lightly loaded cells. Our work in this chapter is followed by a similar user-centric model for MD-CoMP in [43] where a dynamic access threshold based on SC load is driven for each SC to prioritize UEs for scheduling. For highly loaded SCs, some UEs are shifted to alternative lightly loaded SCs where possible. Based on average received pilot power, UEs find alternative cells with minimal spectral efficiency degradation and request access to these alternative SCs. SCs receive access requirements and confirm access for UEs based on priorities driven from CoMP access threshold. Liu *et al.* provides a load aware user-centric CoMP clustering solution [109] for non-coherent MD CoMP where game theory is applied to find load aware user-centric clusters. UEs are set as players aiming to maximize their throughput and they are implicitly penalized not to prefer highly loaded cells as UEs will be allocated less resources in highly loaded cells. In both works, MD-CoMP scenario is studied where same user is allocated dedicated resources from all SCs in the cluster and hence MD-CoMP is not spectral efficiency efficient unless proposed SINR increase is more than the additional amount of resource usage. Different to these studies, our work employs MU JT-CoMP where capacity dynamics are different as same resources from each SC within the cluster can be allocated to multiple users and hence

additional capacity is possible by increasing cluster size provided that complexity and other overheads are kept at manageable level. Our model exploits clusters size and dynamically increase clusters size for highly loaded cells to provide additional capacity where required. In [106], authors propose a load aware JT-CoMP scheme in C-RAN architecture, where sleeping BSs are utilized to improve coverage probability and enhance ergodic link throughput. The additional capacity from the void BSs are utilized to improve spectral efficiency and hence reduce load. Solution assumes a number of void BSs and all BSs to be able to coordinate with JT-CoMP, however computational complexity can be severely high for all BSs to cooperate depending on the number of BSs involved in cooperation.

3.3 System Model

We consider CDSA architecture in our model where MBS is used to provide coverage and handle most of the control signalling and SCs under the MBS provide the required data services [120]. We consider MBSs have enhanced functions as CCU with fiber backhaul links to the SCs within its coverage area as illustrated in Figure 3.1. CCU functionality on the MBS will handle central precoding design, baseband processing and make intelligent clustering decisions centrally within the SC layer. With all SCs connected to the associated MBS, there will be no need for high bandwidth backhaul between the SCs. In addition to CDSA model, our presented algorithm can also be implemented in C-RAN architecture [119, 161] where the clustering decision, precoding, scheduling functions can take place at the "cloud" centrally.

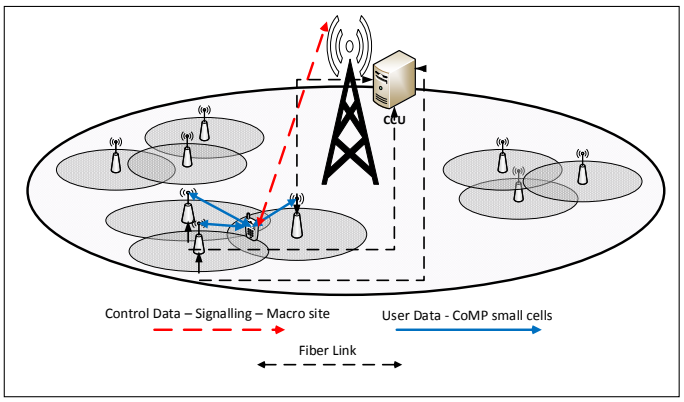


Figure 3.1: Control-data separation architecture (CDSA).

Assume there are $\mathcal{C}$ small cells and $\mathcal{U}$ users in one MBS's coverage area where all SCs are connected to the MBS with fiber via which SCs share its CSI. Global precoding is designed and scheduling is performed at the MBS. MBS acts as CCU

for all SCs in the serving area. Frequency spectrum used for SC layer is different to macro layer hence no interference between MBS and SCs are assumed. Similar two frequency approach is also employed in 3GPP LTE-A HetNet deployment scenario [4]. We propose to use different time frames for pre-coding and clustering. Precoding is calculated in much faster rate in response to the fast fading channel conditions, however clustering decisions are updated in longer time intervals based on averaged receive power levels eliminating fast fading effects [64, 68, 168]. This gives extra resilience on clustering algorithm's imperfect CSI knowledge and reduces additional signalling required for more frequent cluster changes [133].

User-centric clustering is employed in this work, where each UE is assigned its own cluster within the group of SCs connected to the same MBS. MU-JT CoMP is employed where user data is available at all SCs within the cluster. Ideal backhaul and perfect CSI knowledge are assumed. Zero forcing (ZF) precoding is employed where intra-cluster interference is completely canceled. Maximum transmit power P_{Tx} from each SC is assumed to be equal.

Assume UE_k is assigned a cluster of SCs defined as $\mathcal{C}^k$ where $|\mathcal{C}^k| = T$. A group of UEs defined as $\mathcal{U}^k$ including UE_k are scheduled at the same physical resource block (PRB) in this cluster where $|\mathcal{U}^k| = R$. Each UE and SC are assumed to have one TP for simplicity. Received signal for each UE in $\mathcal{U}^k$ can be expressed as:

$$\mathbf{y} = \mathbf{HWx} + \mathbf{n}, \mathbf{H} \in \mathbb{C}^{R \times T}, \mathbf{W} \in \mathbb{C}^{T \times R} \tag{3.1}$$

Channel vector at UE_k is expressed as:

$$\mathbf{h}_k = \begin{bmatrix} h_{k1} h_{k2} \dots h_{kT} \end{bmatrix} \tag{3.2}$$

where $\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_R \end{bmatrix}^T$

Beamforming vector for UE_k is expressed as:

$$\mathbf{w}_k = \begin{bmatrix} w_{1k} w_{2k} \dots w_{Tk} \end{bmatrix}^T \tag{3.3}$$

where $\mathbf{W} = \begin{bmatrix} \mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_R \end{bmatrix}$

Received signal at UE_k can be expressed as:

$$y_k = \mathbf{h}_k^{\mathcal{C}^k} \mathbf{w}_k^{\mathcal{C}^k} x_k + \sum_{i \in \mathcal{U}^k / UE_k} \mathbf{h}_k^{\mathcal{C}^k} \mathbf{w}_i^{\mathcal{C}^k} x_i + \sum_{j \in \mathcal{U} / \mathcal{U}^k} \mathbf{h}_k^{\mathcal{C} / \mathcal{C}^k} \mathbf{w}_j x_j + n_k \tag{3.4}$$

First term in (3.4) represents the desired signal, followed by intra-cluster in-

terference from cells within the cluster $\mathcal{C}^k$ and the third term represents inter-cluster interference from all SCs outside of the cluster. Last term n_k represents the AGWN.

SINR at UE_k can be written as:

$$SINR_k = \frac{\left|\mathbf{h}_k^{\mathcal{C}^k}\mathbf{w}_k^{\mathcal{C}^k}x_k\right|^2}{\left|\sum_{i\in\mathcal{U}^k/UE_k}\mathbf{h}_k^{\mathcal{C}^k}\mathbf{w}_i^{\mathcal{C}^k}x_i\right|^2+\left|\sum_{j\in\mathcal{U}/\mathcal{U}^k}\mathbf{h}_k^{\mathcal{C}/\mathcal{C}^k}\mathbf{w}_j x_j\right|^2+|n_k|^2} \quad (3.5)$$

We assume perfect channel knowledge with ZF precoder and equal transmit power for all PRBs within all SCs. Also equal total transmission power (P_{Tx}) is assumed for all cells. With ideal backhaul and perfect CSI knowledge, any typical ZF precoder results in the cancellation of intra-cluster interference. Similar assumptions are also made in literature for CoMP clustering algorithms in [70, 90, 117, 134]. Consequently, (3.5) can be simplified to:

$$\hat{SINR}_k = \frac{P_{Tx}\sum_{i\in\mathcal{C}^k}|h_{ki}|^2}{P_{Tx}\sum_{j\in\mathcal{C}/\mathcal{C}^k}|h_{kj}|^2+N_0B_{tot}} \quad (3.6)$$

where N_0 is the noise spectral density and B_{tot} is the total system bandwidth.

Ideal backhaul and perfect CSI knowledge is an over-estimation but gives an illustrative bound for our work. Non-ideal backhaul is studied in Chapter 5 where impact of non-ideal backhaul is taken into account as reduction in achieved spectral efficiency when compared to the ideal case as studied in [39].

Channel coefficient h_{ki} is made up of 2 terms, static distance based path loss component with shadow fading and fast fading complex coefficients:

$$h_{ki} = g_{ki} * f_{ki} \quad (3.7)$$

In (3.7), g_{ki} is the distance based path-loss and shadow fading component and f_{ki} is the complex fast fading channel coefficient. As discussed earlier, clustering decisions are proposed to be based on long term received power levels, hence fast fading competent in (3.6) will be averaged out. Consequently, (3.6) can be further simplified to eliminate fast fading component for clustering decisions:

$$\hat{SINR}_k = \frac{P_{Tx}\sum_{i\in\mathcal{C}^k}|g_{ki}|^2}{P_{Tx}\sum_{j\in\mathcal{C}/\mathcal{C}^k}|g_{kj}|^2+N_0B_{tot}} \quad (3.8)$$

3.4 User-centric Clustering Algorithm

We consider user-centric clustering where each UE_k has its cluster of SCs based on the received power levels. Each UE will report its average received reference signal power levels from each of the SCs within its range to its best serving SC. Collected signal levels from each UE will be sent to the CCU located at MBS through the fiber backhaul from the SCs to the MBS. CCU will process this information to assign a cluster of SCs to each UE for cooperation. We propose to limit the complexity of user-centric clustering by keeping the clustering only to the SCs which are connected to the same MBS. This approach can also be considered as a hybrid-clustering approach where all SCs connected to the same MBS form a network-centric cluster and user-centric clustering is employed within the network-centric cluster. Management of inter-cluster interference between the SCs under different MBS is out of scope for this work. Cluster size is designed to dynamically change for each UE_k based on received power levels. Clusters are designed from SCs within closer range of serving SC received power level and a minimum power threshold P_{min} is applied to avoid unnecessary cells in the clusters.

The proposed user-centric clustering algorithm works as follows:

1. For each UE_k, compute the average received power levels from all SCs within the MBS. Received power levels will be averaged out in time, eliminating the fast fading component. Hence g_{km} in (3.9) consists of path loss and shadow fading only, i.e., the received power from any SC_m at UE_k can be expressed as:

$$p_{km} = P_{Tx} * |g_{km}|^2, m \in \mathcal{C} \tag{3.9}$$

2. p_{km} is sorted for each UE_k.

$$p_{k1} = \arg\max_{m} p_{km}, m \in \mathcal{C} \tag{3.10}$$

   p_{k1} is the received power for serving cell for UE_k. Similarly p_{k2} indicates the received power for 2nd best serving cell and so on.

3. Choose cluster $\mathcal{C}^k$ for UE_k from cells with highest received power levels with following conditions:

   (a) Number of cells do not exceed the maximum cluster size defined for the algorithm. This is a tunable input parameter to the algorithm

where complexity against CoMP efficiency trade-off can be balanced.

$$|\mathcal{C}^k| \leq C_{max} \tag{3.11}$$

(b) Received power level should not be lower than a minimum threshold. This ensures no unnecessary cells added to the cluster preventing increased signalling and wasted resources without significant CoMP gains.

$$p_{km} > P_{min} \tag{3.12}$$

(c) For any cell in the cluster, $p_{km}/p_{k1} > P_{\Delta}$. This ensures that cells within a similar received power range to the serving cell are included in the cluster.

3.5 Clustering with Load Balancing

In this section, we discuss how we utilize user-centric clustering algorithm defined in the previous section and propose a novel load balancing algorithm to dynamically change clusters to distribute load evenly in hotspot areas. In the following subsection, we define cell load and unsatisfied users metrics for MU JT-CoMP clustering scenario. Then, the user-centric clustering algorithm with load balancing is detailed in the following subsection.

3.5.1 Cell Load and Unsatisfied Users Metric

In [164], a mathematical framework is developed for cell load for traditional networks and a term called "unsatisfied users" is introduced for UEs where available throughput is below the guaranteed bit rate (GBR) for their service. Based on this work, we derive the cell load and unsatisfied users metrics for MU JT-CoMP scenario. Our proposed CoMP clustering algorithm will aim to minimize the number of unsatisfied users by user-centric cluster formation taking cell load into account .

We assume total number of PRBs at each SC as R_{tot} where each PRB has a bandwidth of B_{PRB}. Based on the Shannon capacity formula, the maximum achievable throughput from one PRB can be estimated as:

$$y_k = B_{PRB} log_2(1 + \hat{SINR}_k) \tag{3.13}$$

We assume constant bit rate d_k is required for each user UE_k, hence the average required PRBs for each user for no CoMP scenario is $r_k = d_k/y_k$. But

in the MU-JT CoMP case, user data for UE_k is also transmitted from the other SCs in the cluster $\mathcal{C}^k$. So UE_k will require resources from each of the SCs in the cluster. Additionally, same resources allocated for UE_k are shared with other UEs $\in \mathcal{U}^k$ scheduled in the same cluster. We assume the number of UEs sharing the same PRB in the same cluster is equal to the the cluster size for UE_k i.e., $|\mathcal{C}^k| = |\mathcal{U}^k| = n_k$. Since same PRB is shared between n_k number of users, we can estimate a dedicated PRB allocation for each user as $\hat{r}_k = r_k/n_k$ in MU JT-CoMP scenario. So the estimated average dedicated PRB required for UE_k from all SCs in the $\mathcal{C}^k$ cluster can be defined as:

$$\hat{r}_k = \frac{d_k}{y_k n_k} \tag{3.14}$$

For example, assume a three cell cluster with three UEs scheduled at the same time on the same one PRB from each of the SCs in the cluster. An estimated average dedicated PRB requirement for each UE from each SC is 1/3, and hence all three UEs PRB requirements in one SC will add up to one PRB.

Let $\mathcal{U}_m$ be the active UE list attached to SC_m. Cell load l_m can be defined as proportion of the number of utilized PRBs to the total PRB count R_{tot} on the SC. Since load can not exceed one, l_m can be expressed as:

$$l_m = \min\left(1, \frac{\sum_{k\in\mathcal{U}_m} \hat{r}_k}{R_{tot}}\right) \tag{3.15}$$

From l_m in (3.15), we can also define an estimated cell load $\hat{l}_m$ which is allowed to go beyond one, and give a measure of how much overloaded the cell is:

$$\hat{l}_m = \frac{\sum_{k\in\mathcal{U}_m} \hat{r}_k}{R_{tot}} \tag{3.16}$$

From (3.16), we can define an unsatisfied users term to indicate the load on cell and use it as target function to minimise unsatisfied users in CoMP clustering. Given that all users are assumed to require constant GBR d_k, the users will be defined as "satisfied" if they obtain their GBR, otherwise unsatisfied. For example, when $\hat{l}_m \leq 1$, all associated users are satisfied in SC_m and when load increases to $\hat{l}_m = 4$, only one fourth of the users are satisfied [164].

To be able to calculate unsatisfied users for each SC, we need to express an estimated dedicated number of UEs associated with each cell in the MU JT-CoMP scenario. As defined above, $\mathcal{U}_m$ represents the active UE list in SC_m, however UEs are associated to multiple SCs in the MU-JT CoMP case. Since each UE is repeated on all cells in its cluster, an estimated dedicated UE count for each SC can be found by adding up all associated UEs with a factor of $1/n_k$ i.e. its

cluster size. Estimated dedicated number of UEs associated with each cell can be expressed as:

$$\hat{u}_m = \sum\nolimits_{k \in \mathcal{U}_m} \frac{1}{n_k} \tag{3.17}$$

Number of unsatisfied users on SC_m can then be defined as:

$$\hat{z}_m = \max\left(0, \hat{u}_m \left(1 - \frac{1}{\hat{l}_m}\right)\right) \tag{3.18}$$

3.5.2 Clustering Algorithm with Load Balancing

User-centric clustering algorithm discussed in Section 3.4 is adjusted to take SC load into account with the aim of balancing the load across the SCs. Clustering for load balancing is designed in 2 stages:

1. **Stage-1: Increase cluster size:**
   Increased cluster size will increase the capacity in a given cluster with MU JT-CoMP at the expense of additional complexity as extensively discussed in Chapter 2. In our proposed algorithm, the capacity/complexity trade-off is managed with assigning different maximum cluster size limit for low and high load scenarios separately. We define L_{min} as an input parameter for the high load threshold and identify the set of users $\mathcal{U}_h$ which are served by highly loaded SCs i.e. any UE_k where $\hat{l}_m > L_{min}$ for any $SC_m \in \mathcal{C}^k$. We set an increased maximum cluster size C^h_{max} for UEs in $\mathcal{U}_h$ and the maximum cluster size set for light load C_{max} is increased in iteration until the high load is cleared or the C^h_{max} is reached. Algorithm flow is illustrated in Algorithm 1.

Algorithm 1 Stage-1: Increase cluster size:

```
while C_max < C^h_max and U_h ≠ ∅ do
  C_max = C_max + 1
  for all UE_k ∈ U_h do
    if UE_k ∈ U_h then
      Re-cluster with incremented C_max
      Update U_h
    end if
  end for
end while
```

2. **Stage-2: Re-cluster to exclude highly loaded SCs:**
   If $\mathcal{U}_h \neq \emptyset$ after cluster size increase in Stage-1, UEs in $\mathcal{U}_h$ will be subject to re-clustering in Stage-2 with the aim of excluding highly loaded SCs from

the user-centric clusters. We define a maximum allowable spectral efficiency loss parameter SE_{Δ}^{max} in the algorithm and start re-clustering UEs at the cell edge where there are alternative clusters available with a lower spectral efficiency loss. Allowed spectral efficiency loss $SE_{\Delta}^{max-iter}$ is incremented in iterations upto SE_{Δ}^{max} to make sure cell edge users are re-clustered first and users closer to the cell center are re-clustered only when there is still highly loaded SCs after cell-edge users are re-clustered. Furthermore, our algorithm looks for a candidate cluster for each UE in $\mathcal{U}_h$ where none of the SC load levels within the candidate clusters are greater than L_{min}. If this is not achievable, then L_{min} is incremented to a higher value to look for candidate clusters where only the highest loaded SCs in the cluster are excluded, but relatively lower loaded SCs are still included in the candidate cluster. This part of the algorithm is illustrated in Algorithm 2.

Algorithm 2 Stage-2: Re-cluster excluding loaded cells:

while $SE_{\Delta}^{max-iter} \leq SE_{\Delta}^{max}$ **do**
  for all $UE_k \in \mathcal{U}_h$ **do**
    while $\arg\max(l_m \text{ for } \forall SC_m \in \mathcal{C}^k) \geq L_{min}$ **do**
      Find $\hat{\mathcal{C}}^k$ where $\hat{l}_m$ for $\forall SC_m \in \hat{\mathcal{C}}^k < L_{min}$
      if $\hat{SINR}_k > SINR_{min}$ and $SE_{\Delta} < SE_{\Delta}^{max-iter}$ **then**
        Recluster UE_k
        Break while-loop and move to next UE
      else
        Increment L_{min}
      end if
    end while
  end for
  Increment $SE_{\Delta}^{max-iter}$
end while

3.6 Numerical Results

In order to evaluate the performance of our novel, load aware, user-centric CoMP clustering method presented in this chapter, SCs within one MBS coverage area is considered where MBS coverage area is assumed to be a circle with radius $r_b = 0.4m$. To simulate the unplanned nature of SC deployment in future cellular networks, SCs are modelled as RN following PPP distribution with density parameter $\lambda_{\mathcal{C}}$. UEs are also randomly distributed following PPP distribution with density$\lambda_{\mathcal{U}_{high}}$ and $\lambda_{\mathcal{U}_{low}}$. MBS coverage area is assumed to have un-even traffic distribution where there is high user density $\lambda_{\mathcal{U}_{high}}$ within the inner circle and low user density $\lambda_{\mathcal{U}_{low}}$ in the outer ring. SCs deployed within the inner circle

Parameter Name	Parameter Value
Simulation Environment	Urban Microcell [83]
Frequency Carrier	5 GHz
Channel Bandwidth	5 MHz
PRB Bandwidth (B_{PRB})	180 kHz
Number of PRBs/SC (R_{tot})	25
Shadow fading std	4 dB [83]
UE Antenna Gain	0 dBi
UE Thermal Noise Density	-174 dBm/Hz
TP Total Transmit Power (P_{Tx})	41dBm [83]
UE Noise Figure	7dB
TP Noise Figure (Cable loss)	5dB
BS antenna gain (boresight)	17dBi
User-centric cluster: Min RX Power (P_{min})	-110dBm
User-centric cluster: Max RX power offset (P_{Δ})	20dB
User-centric cluster: Min Required SINR ($SINR_{min}$)	0
Max Cluster Size (no Overload) (C_{max})	3
Max Cluster Size (Overload) (C^{h}_{max})	6
Guaranteed bit rate for UEs (d_k)	512 kbps
SC Density Dense Deployment ($\lambda_{\mathcal{C}}$)	80SC/km^2
SC Density Medium Deployment ($\lambda_{\mathcal{C}}$)	40SC/km^2
SC Density Sparse Deployment ($\lambda_{\mathcal{C}}$)	20SC/km^2
MS Density Hotspot High Load ($\lambda_{\mathcal{U}_{high}}$)	12000UE/km^2
MS Density Hotspot Medium Load ($\lambda_{\mathcal{U}_{high}}$)	10000UE/km^2
MS Density Hotspot Light Load ($\lambda_{\mathcal{U}_{high}}$)	6000UE/km^2
MS Density Non-Hotspot ($\lambda_{\mathcal{U}_{low}}$)	800UE/km^2
Simulation Area Radius (r_b)	0.4km
Non-Hotspot Area Radius	0.2km
Hotspot Area Radius	0.1km
Minimum Load (L_{min})	80%

Table 3.1: Simulation parameters.

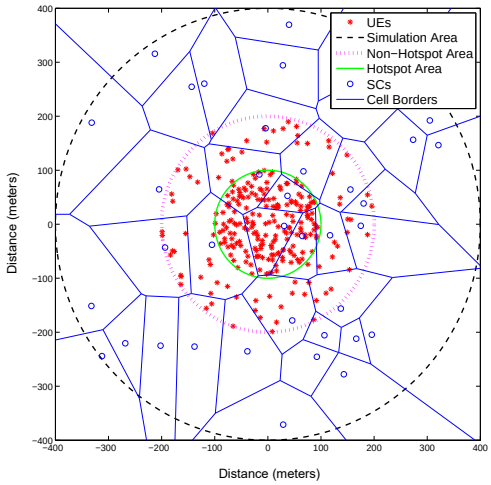


Figure 3.2: Simulation network topology illustration, SC and UE locations, hotspot and non-hotspot areas, cell borders following Voronoi tessellation.

will be highly loaded and the aim is to reduce the load on these SCs by shifting traffic from highly loaded cells to under-utilized SCs by dynamic user-centric CoMP clustering. High user density area radius is assumed to be 0.1m and low density user radius is assumed to be 0.2m. SCs are deployed within a larger area ($r_b = 0.4m$) to avoid border effect and make sure UEs at the border receive interference from within 0.2km outside the UE radius. The simulation setup with network topology is illustrated in Figure 3.2.

Each SC is assumed to have one cell with omnidirectional antenna. The ITU-R microcell urban non-line-of-sight (NLOS) path loss in [83] is employed for SC path loss as given below where d is the distance between SC and the UE in meters and fc is the carrier frequency in GHz.

$$PL = 36.7 \log_{10}(d) + 22.7 + 26 \log_{10}(fc) \tag{3.19}$$

Antenna bore-sight gain is assumed to be 17dBi and TP noise figure is assumed to 5dBm as suggested for ITU-R microcell urban test environment [83]. Frequency carrier is selected as 5GHz to simulate medium bandwidth spectrum range (i.e. $<$ 6GHz) proposed for 5G [95].

MU JT-CoMP with coherent combining is employed, however proposed algorithm can be easily adapted to other coordination methods, SU-JT-CoMP or CS/CB CoMP. The rest of the simulation parameters are provided in Table 3.1. We run Monte Carlo simulations for a number of scenarios to simulate dense/medium/sparse deployment with high/medium/light load and each scenario has been run for one hundred snapshots. Average of all one hundred

snapshots are presented for each scenario.

3.6.1 Dense Deployment with High Load Scenario

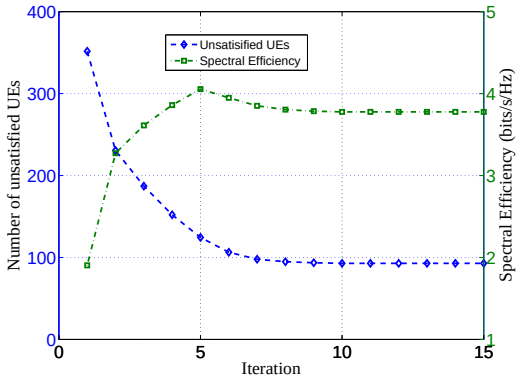
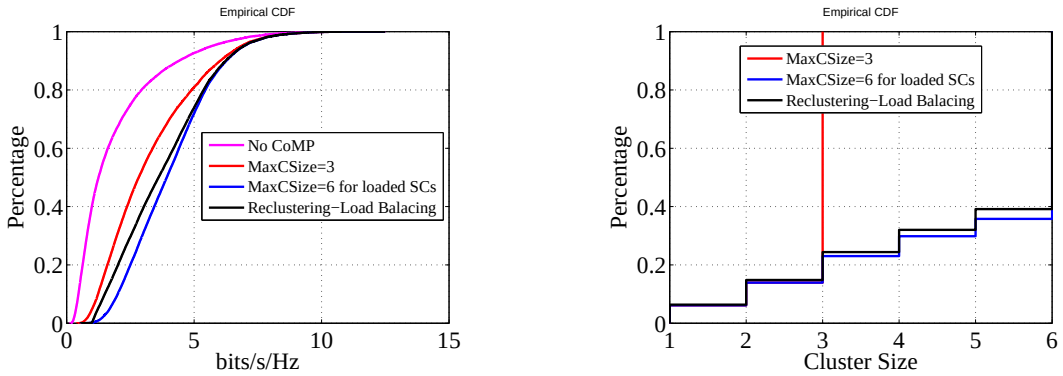


Figure 3.3: Unsatisfied UEs and spectral efficiency changes in dense deployment scenario with high load.

Figure 3.3 depicts the changes in number of unsatisfied users and average spectral efficiency in iterations for dense network deployment with high UE load case. First iteration shows the unsatisfied users when CoMP is not employed. Our presented user-centric CoMP clustering is employed in the next iteration with max cluster size of three without taking SC load into account. This reduces the number of unsatisfied users by 34.4% due to the additional capacity introduced with MU JT-CoMP. Load balancing algorithm is used at the next stage where only the UEs attached to highly loaded cells are allowed to increase cluster size beyond the original value of three. The additional three iterations in Figure 3.3 gives the reduction in the number of unsatisfied UEs by increasing maximum cluster size to 4,5 and 6 respectively. Unsatisfied UEs are reduced by an additional 30.2% at this stage. Spectral efficiency continues to increase as CoMP cluster size increases, and no cells are excluded from clusters until iteration 6. Once cluster size is increased to the maximum $C_{max}^{h} = 6$ for loaded cells, then Stage-2 of the load balancing algorithm starts to further reduce the unsatisfied users based on re-clustering UE clusters for UEs which are served by any SC_m where $\hat{l}_m > L_{min} = 80\%$. UEs are re-clustered only if the spectral efficiency loss is below a certain spectral efficiency loss threshold $SE_{\Delta}^{max-iter}$ at each iteration. This threshold is increased at each iteration until either all unsatisfied UEs are cleared, or the max allowed limit for spectral efficiency loss threshold SE_{Δ}^{max} is reached. This ensures that UEs located at the cell edge of the loaded SCs are re-clustered to other neighbour SCs first and gradually more UEs are re-clustered until cell load is reduced to $< L_{min} = 80\%$. Spectral efficiency loss steps $SE_{\Delta}^{max-iter}$ are set to 1, and max spectral efficiency

loss threshold SE_{Δ}^{max} is set to 5 in this simulation. An additional 9% of the unsatisfied UEs are reduced in Stage-2 in dense deployment case.(i.e. iterations 6-15 in Figure 3.3). Overall, number of unsatisfied UEs are reduced by 73.6%. Re-clustering in Stage-2 of the algorithm comes with the cost of reduced spectral efficiency as some of the UEs served by loaded SCs area handed over to the non-best serving cells. 6.84% reduction in spectral efficiency is observed in the dense deployment with high load case. In return for spectral efficiency loss, more users have been allocated their GBRs, resulting in the reduction of unsatisfied UEs by 9%. Spectral efficiency distribution at different stages of the algorithm is shown in Figure 3.4a.



(a) Spectral efficiency distribution.

(b) Cluster size distribution.

Figure 3.4: Dense deployment scenario with high load.

Figure 3.4b depicts the cluster size distribution at 3 different iteration points, i.e., iteration 2, 5 and 15 capturing the cluster size distribution when maximum cluster size C_{max} is set to 3, 6 and at the end of the re-clustering iterations. 86.1% of the UEs had 3 cells in their cluster when the initial clustering algorithm was deployed, however when the cluster size is increased to 6 for load balancing, UEs with maximum cluster size of 6 is reduced to 60.9%. This is due to clustering algorithm not allowing cluster size increase if it is not required for load balancing. Figure 3.7a shows the SC load distribution at the same iteration points where load distribution is clearly improved when cluster size is increased to 6 for loaded cells and it is further improved after re-clustering.

3.6.2 Dense/Medium/Sparse Deployment

Simulations are run for dense/medium/sparse deployment scenarios with high UE load i.e. $\lambda_{\mathcal{C}}$=80,40 and 20 SC/km^2 with $\lambda_{\mathcal{U}_{high}}$=12000 UE/km^2 respectively to compare the effectiveness of the algorithm. Figure 3.5 shows the unsatisfied UEs reduced by 73.6%, 64.8% and 56.6% for dense, medium and sparse deploy-

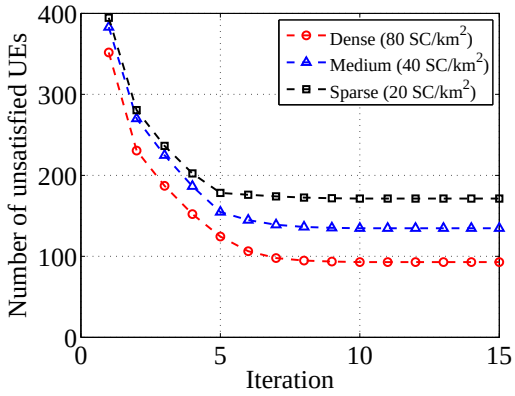
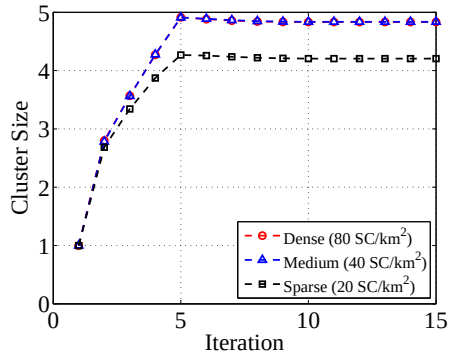


Figure 3.5: Unsatisfied UEs in dense/medium/sparse deployment scenario with high load.

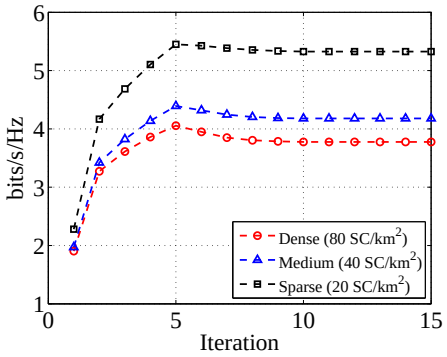
ment respectively. Results clearly show that presented algorithm is more effective in the dense deployment scenario. As presented in Figure 3.6a, sparse deployment results in significantly lower cluster size, due to lack of available SCs with overlapping coverage, hence limiting the re-clustering options for load balancing. Spectral efficiency changes are compared in Figure 3.6b showing negligible spectral efficiency loss in re-clustering phase (from iteration 5 on-wards) in sparse deployment due to minimal amount of re-clustering activity. Higher spectral efficiency is achieved in sparse deployment due to limited inter-cell interference with lack of too many neighbour SCs. Figure 3.7b shows the SC load distribution at sparse deployment scenario where re-clustering is not effective. However in dense deployment scenario, SC load distribution shows a clear improvement due to re-clustering in Figure 3.7a. Overall, our algorithm is significantly more effective in dense deployment scenario which is the likely deployment case in 5G and beyond wireless networks to cope with high capacity demand. Dense deployment scenario provides alternative clustering options to shift load effectively which has been utilized in our novel algorithm.

3.6.3 Dense Deployment with High/Medium/Light load

We have also evaluated the proposed scheme in dense deployment scenario with different UE load conditions. Figure 3.8 shows the the change in unsatisfied UEs for dense deployment in high/medium/light load scenarios. In the light load scenario, unsatisfied UEs have almost completely disappeared at iteration 6, limiting the spectral efficiency loss allowed for re-clustering ($SE_{\Delta}^{max-iter}$) to 1 only. On the other hand, Figure 3.9a shows that average cluster size is significantly lower in the light load scenario. The algorithm is only applied to the UEs served by loaded cells which is a lower portion of the total UEs for light load case. Lower

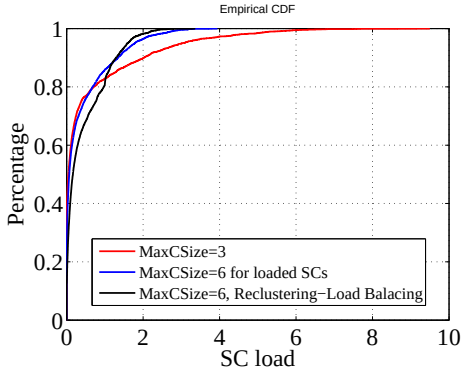


(a) Mean cluster size.

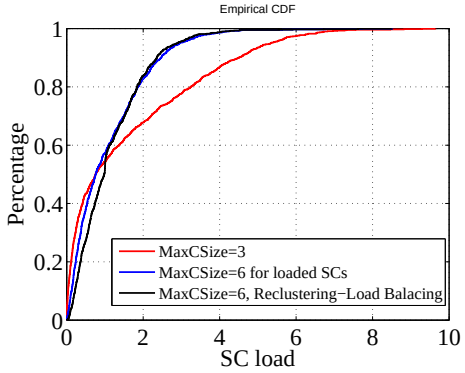


(b) Mean spectral efficiency.

Figure 3.6: Dense/medium/sparse deployment scenario with high load.



(a) SC load distribution in dense deployment with high load.



(b) SC load distribution in sparse deployment with high load.

Figure 3.7: SC load distribution in dense/sparse deployment scenario with high load.

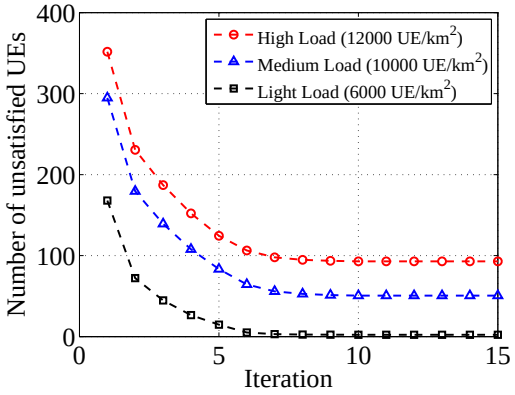
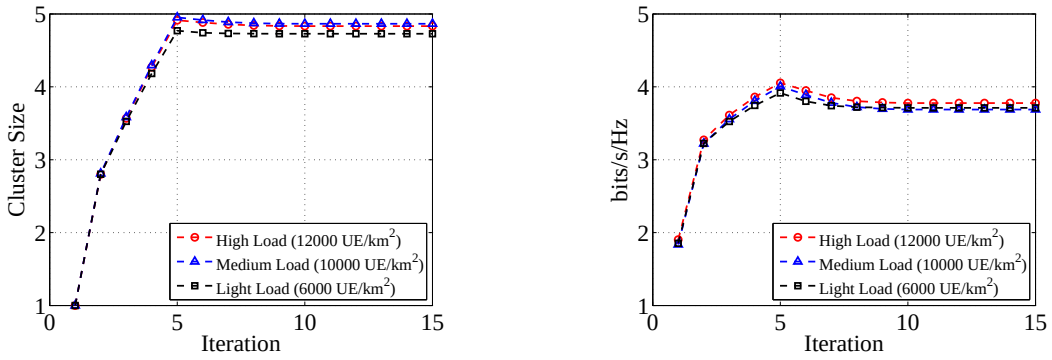


Figure 3.8: Unsatisified UEs in dense deployment scenario with high/medium/light load.

cluster size in light load case has direct effect on spectral efficiency, as light load scenario has the lowest spectral efficiency in Figure 3.9b due to lower cluster size. Furthermore, spectral efficiency loss due to re-clustering is minimum in the

light load scenario as shown in Figure 3.9b from iteration 5 to 12. Figure 3.10 shows the SC load distribution in dense deployment with light load scenario where all SCs are successfully moved to light load range (load <1) after re-clustering. Overall, our algorithm responds to the UE load successfully where cluster size increase (stage-1) and re-clustering (stage-2) algorithms made minimal changes in light load scenario to avoid additional complexity with increased cluster size and spectral efficiency loss in re-clustering



(a) Mean cluster size.

(b) Mean spectral efficiency.

Figure 3.9: Dense deployment scenario with high/medium/light load.

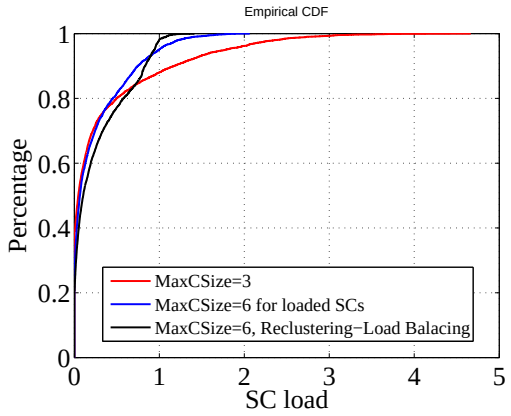


Figure 3.10: SC load distribution in dense deployment with light load.

Finally, Figure 3.11 shows the total number of unsatisfied UEs for different allowed maximum cluster size C^h_{max} in the dense deployment scenario with high load. It can be seen that as the cluster size increases, the impact of the increased cluster size on reducing the unsatisfied UE metric reduces. Based on the density of the deployment, max cluster size need to be optimised carefully for maximising the load balancing gains in return for increased complexity due to high cluster size.

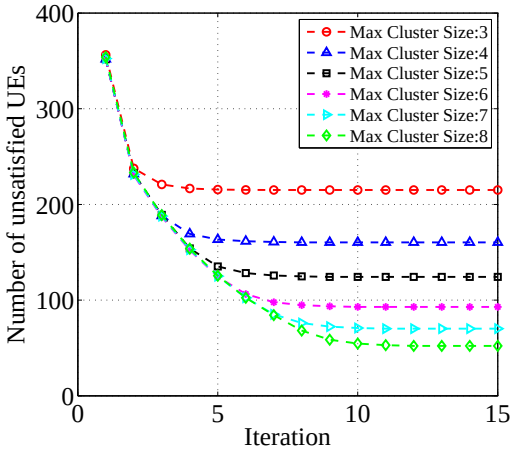


Figure 3.11: Unsatisfied UEs for dense deployment with high load for different max. cluster size.

3.7 Conclusion

In this chapter, we presented a novel, load aware, user-centric CoMP clustering model. It has been shown that additional capacity with MU JT-CoMP can be utilized further by increasing the cluster size when required. Increased complexity with cluster size can be minimized with our proposed algorithm where cluster size is increased only when additional capacity is required. We show that re-clustering UEs to exclude loaded SCs distributes the load to lightly loaded cells, decreasing the number of unsatisfied UEs significantly. Spectral efficiency loss due to re-clustering is minimized by re-clustering UEs at the cell edge first and move closer to the cell center only after the cell edge users are moved and there is still high load. It is shown that presented algorithm is most effective in dense deployment scenario which is the likely case for CoMP deployment in future 5G networks. Furthermore, algorithm is also tested for different load scenarios and we show that cluster size increase and re-clustering are kept to minimum in light load scenarios. So the re-clustering actions are minimized based on load conditions dynamically. Moreover, the effect of maximum allowed cluster size C^h_{max} is also investigated and concluded that maximum allowed cluster size needs to be tuned carefully based on SC density, as a larger cluster size has minimal impact in relatively lower SC density.

Complexity of employing user-centric clustering is limited to the coverage area of the MBS in the proposed algorithm. Depending on the SC density within the MBS coverage area, additional complexity will arise when a high number of SCs will need to coordinate at the same time. To reduce this complexity, CoMP can

be deployed in smaller network-centric clusters and user-centric clusters can be implemented within the smaller network-centric clusters. In the following chapter, we study a novel load aware network-centric CoMP clustering solution to jointly optimize spectral efficiency and load balancing objectives.

Chapter 4

Load Aware Network-Centric Clustering for CoMP

4.1 Introduction

In this chapter, we present a novel, load aware, network-centric clustering algorithm to eliminate the complexity issues arise with user-centric clustering models. Our goal is to find the clustering structure $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_s\}$ which best satisfies the network objectives in terms of CoMP performance and its overhead costs. Finding the best cluster formation by exhaustive search of every possible cluster combination is too complex especially when the network size is larger. As discussed in Chapter 2, most of the existing solutions lack on scalability due to exponential increase in processing complexity as the network size increase. Applications of coalitional game theory in wireless networks is an emerging concept especially with CoMP and network coordination in general [148] to reduce this complexity. It provides a flexible analytical framework to provide distributed, low overhead, less complex solutions. We utilize coalition game theory and setup a merge/split coalition formation game to model the dynamic clustering problem for MU-JT CoMP in downlink and optimize spectral efficiency and load balancing objectives jointly. A load aware utility is designed to formulate the trade-off between cluster size/complexity and spectral efficiency vs. load balancing. A dynamic cluster size adaptation is formed where maximum cluster size is dynamically increased in high load conditions to improve spectral efficiency and reduce load. We show that our proposed merge/split cluster formation framework provides a low complexity solution and always converges to a stable partition in both HN and RN scenarios with different load conditions. Moreover, our load aware clustering model achieve high spectral efficiency in light load scenario and better load distribution in high-load scenario resulting in significantly lower number of

unsatisfied users while keeping spectral efficiency at comparably high levels. We analyze the trade-off between additional complexity of bigger cluster size and the improvement in spectral efficiency and load balancing in both HN and RN scenarios. Simulation results are compared to an improved version of greedy clustering model presented in [134]. We show that our solution outperforms the greedy solution and it provides a low complexity, scalable and stable clustering solution. In this context, the unique contribution of this paper is that we introduce load balancing as one of the key objectives for network-centric clustering for the first time in literature and develop a novel, low complexity and stable network-centric clustering model as a first attempt to fill the gap in literature for load aware network-centric CoMP clustering, jointly optimizing load balancing and spectral efficiency. Part of this work is published in [31].

The rest of the chapter is organised as follows. In Section 4.2, we briefly discuss the related work in literature and then, we present our system model for MU JT-CoMP and discuss key performance metrics and overheads for CoMP in Section 4.3. In Section 4.4, we first present coalition formation game concepts. Next, we introduce our spectral efficiency based and load aware utility functions employed in our coalitional game. We then present merge/split game operation in detail and discuss its complexity and stability. In Section 4.5, we present simulation results for HN and RN with and without hotspot scenarios. Finally, conclusions are drawn in Section 4.6.

4.2 Related Work and Problem Statement

Capacity is one of the biggest challenges for future networks with ever growing data demand globally. Mobile data traffic is expected to grow at an annual growth rate of 46% over the next 5 years and a 7-fold increase is expected by 2022 [53]. Clearly, load balancing is a key network objective to utilize all resources effectively to cope with this capacity demand and not to waste much needed resources. Load balancing has been studied in the literature for traditional networks without coordination for a long time [25, 79, 172] but there has been only limited load balancing studies available for CoMP networks. Our work presented in Chapter 3 is the first work in literature offering a load aware, user-centric clustering solution for MU JT-CoMP scenario. Following on from our work, a load aware user-centric clustering algorithm is proposed in [106] where sleeping cells are utilized to improve ergodic link throughput. More recently, Cao *et al.* provides a further load aware, user-centric clustering algorithm for MD-CoMP with an interactive approach where users request access from SCs and SCs schedule users based on

load aware threshold function. Another load aware, user-centric clustering model is presented in [109] for MD-CoMP where game theory is utilized to form load aware user-centric clusters. All of these works in literature provide user-centric solutions to optimize load balancing, however user-centric clusters will increase complexity exponentially for CoMP implementation especially with coherent JT-CoMP where there will be higher challenges such as synchronization requirement for all SCs and increased scheduling and precoding complexity etc. To reduce this complexity to manageable levels for a realistic CoMP deployment, smaller network-centric clusters need to be formed and user-centric clustering solutions should be deployed in limited small network-centric clusters only. As discussed in Chapter 2, network-centric clustering problem is studied extensively in literature aiming to optimize spectral efficiency, backhaul availability and energy efficiency, but there is no work yet to optimize load balancing. Our work presented in this chapter attempts to fill this gap, providing a load aware, network-centric clustering model for JT-COMP optimizing load balancing and spectral efficiency objectives jointly for the first time in literature. We deploy a user-centric clustering model within the network-centric clusters to present a novel, low complexity, hybrid clustering model.

4.3 System Model

4.3.1 Network Model

We consider a similar setup to our work in Chapter 3 where we assume a HetNet scenario with one MBS and a set of SCs ($\mathcal{C}$) are deployed within the coverage area of the MBS. A set of users ($\mathcal{U}$) are also distributed in the same area. The SCs are connected to the MBS with fast fiber backhaul links where all SCs share their CSI with MBS. Similar to the approach taken by 3GPP scenario in [4], a designated frequency spectrum is assumed at each layer, hence no interference is expected between MBS and the SC layer.

MU-JT CoMP is employed at SC layer where user data is made available in all SCs within the same network-centric cluster. Network-centric clustering and associated precoding/scheduling is performed at CCU located at the MBS. We propose that re-clustering activities do not aim to exploit the fast fading changes (i.e. in miliseconds) but it will respond to spatio-temporal changes in user/demand profile and the network. Hence, we propose re-clustering activity at a slower rate i.e. in seconds/minutes where fast fading changes are averaged out within this time window. This provides extra resilience in clustering decisions

to issues like imperfect CSI knowledge and also reduce the additional signaling required for faster re-clustering [133]. Precoding within the cluster takes place at much faster rate (i.e. in milliseconds) where fast fading changes are exploited. We assume ideal backhaul and perfect CSI knowledge where intra-cluster interference is reduced to negligible levels with a typical precoder like ZF precoder. Similar assumptions are made in other clustering works such as [90, 134] and in our work in Chapter 3.

Assume that the SC layer is partitioned into smaller clusters of SCs $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_s\}$ and users are assigned to each SC cluster forming user clusters $\mathcal{U} = \{\mathcal{U}_1, \ldots, \mathcal{U}_s\}$ i.e. user group $\mathcal{U}_i$ is assigned to SC cluster $\mathcal{C}_i$. Suppose any user $UE_k \in \mathcal{U}_i$ is assigned a network-centric cluster $\mathcal{C}_i$ and a user-centric cluster of $\mathcal{C}_i^k$ where $|\mathcal{C}_i^k| = T$ and $\mathcal{C}_i^k \subseteq \mathcal{C}_i$. Let $\mathcal{U}_i^k$ be the group of UEs including UE_k which are scheduled at the same PRB at each SC in $\mathcal{C}_i^k$ where $|\mathcal{U}_i^k| = R$. We assume one antenna for each SC and UE for simplicity. A $T \times R$ virtual MIMO system is formed with SCs in $\mathcal{C}_i^k$ and UEs in $\mathcal{U}_i^k$. An illustration of the system model is shown in Figure 4.1.

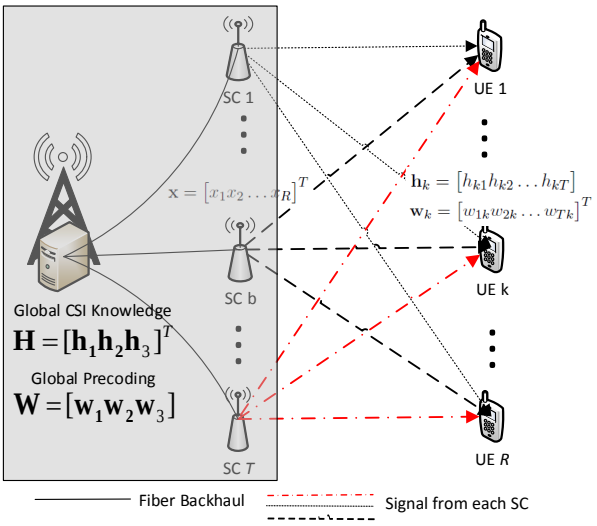


Figure 4.1: System model for downlink MU JT-CoMP.

For each UE in $\mathcal{U}_i^k$, received signal can be expressed as:

$$\mathbf{y} = \mathbf{HWx} + \mathbf{n}, \mathbf{H} \in \mathbb{C}^{R \times T}, \mathbf{W} \in \mathbb{C}^{T \times R} \tag{4.1}$$

where channel matrix can be expressed as: $\mathbf{H} = \left[\mathbf{h}_1 \mathbf{h}_2 \ldots \mathbf{h}_R\right]^T$ and channel vector at UE_k is: $\mathbf{h}_k = \left[h_{k1} h_{k2} \ldots h_{kT}\right]$. Similarly, precoding matrix can be expressed as: $\mathbf{W} = \left[\mathbf{w}_1 \mathbf{w}_2 \ldots \mathbf{w}_R\right]$ and beamforming vector for UE_k is $\mathbf{w}_k = \left[w_{1k} w_{2k} \ldots w_{Tk}\right]^T$.

Received signal at UE_k is:

$$y_k = \mathbf{h}_k^{\mathcal{C}_i^k} \mathbf{w}_k^{\mathcal{C}_i^k} x_k + \sum_{i \in \mathcal{U}_i^k/k} \mathbf{h}_k^{\mathcal{C}_i^k} \mathbf{w}_i^{\mathcal{C}_i^k} x_i + \sum_{j \in \mathcal{U}/\mathcal{U}_i^k} \mathbf{h}_k^{\mathcal{C}/\mathcal{C}_i^k} \mathbf{w}_j x_j + n_k \quad (4.2)$$

First term in (4.2) is the desired signal, where the second term is the intra-cluster interference from SCs within the cluster $\mathcal{C}_i^k$ followed by inter-cluster interference from SCs outside of the cluster $\mathcal{C}_i^k$. The last term n_k is the AGWN at UE_k.

SINR at UE_k can be expressed as:

$$SINR_k = \frac{\left|\mathbf{h}_k^{\mathcal{C}_i^k} \mathbf{w}_k^{\mathcal{C}_i^k} x_k\right|^2}{\sum_{i \in \mathcal{U}_i^k/k} \left|\mathbf{h}_k^{\mathcal{C}_i^k} \mathbf{w}_i^{\mathcal{C}_i^k} x_i\right|^2 + \sum_{j \in \mathcal{U}/\mathcal{U}_i^k} \left|\mathbf{h}_k^{\mathcal{C}/\mathcal{C}_i^k} \mathbf{w}_j x_j\right|^2 + |n_k|^2} \quad (4.3)$$

Intra-cluster interference term $\sum_{i \in \mathcal{U}_i^k/k} \left|\mathbf{h}_k^{\mathcal{C}_i^k} \mathbf{w}_i^{\mathcal{C}_i^k} x_i\right|^2$ in (4.3) becomes negligible when a typical precoder like ZF precoder is employed at the CCU with perfect channel knowledge. We assume equal transmit power on each PRB and also equal total transmit power for each SC. Similar equal transmit power assumption is made in other CoMP clustering works in literature [70,90]. Average SINR term is employed for clustering algorithm as discussed in the previous section. The complex fast fading channel coefficient of the path loss is averaged out in average SINR term and hence, $SINR_k$ can be simplified as:

$$\hat{SINR}_k = \frac{P_{Tx} \sum_{i \in \mathcal{C}_i^k} |g_{ki}|^2}{P_{Tx} \sum_{j \in \mathcal{C}/\mathcal{C}_i^k} |g_{kj}|^2 + N_0 B_{tot}} \quad (4.4)$$

where N_0 is the noise spectral density, B_{tot} is the total system bandwidth and g_{ki} is the distance based path-loss and shadow fading component.

Any user UE_k is first assigned a network-centric cluster $\mathcal{C}_i$ and a user-centric cluster $\mathcal{C}_i^k$ is formed for UE_k from SCs within $\mathcal{C}_i$ based on average received signal level. Inspired from our previous work in Chapter 3, two simple conditions are followed to form user-centric cluster $\mathcal{C}_i^k$ from $\mathcal{C}_i$:

1. Average received power level at UE_k from SC_j in $\mathcal{C}_i^k$ (p_{kj}) should be greater than a minimum threshold i.e. $p_{kj} > P_{min}$. This eliminates any SCs which do not provide the required level of coverage to UE_k.

2. The difference in average received power from the best serving SC_m (p_{km}) to SC_j (p_{kj}) within $\mathcal{C}_i^k$ should not be greater than a threshold i.e. $p_{kj}/p_{km} >$

P_Δ. This ensures only SCs with similar received power levels are in the cluster to maximize interference cancellation from CoMP and prevent unnecessary addition of SCs in $\mathcal{C}_i^k$.

User-centric clusters $\mathcal{C}_i^k$ always have best serving SC and other SCs in the cluster based on above two rules. In this study, we design a network-centric clustering model to jointly optimize load balancing and spectral efficiency and employ this user-centric clustering model within each network-centric cluster. Adjusting user-centric clusters $\mathcal{C}_i^k$ for load balancing presented in Chapter 3 is not considered in this work.

4.3.2 CoMP Performance and Overhead Metrics

The key performance metric for CoMP is the Spectral efficiency improvement achieved by interference mitigation. Spectral efficiency improvement leads to less radio resources utilised, and hence lower cell load. More SCs within the same cluster $\mathcal{C}_i$ will provide additional interference cancellation and better spectral efficiency, but on the other hand, increasing the cluster size will increase the CoMP overheads. Additional pilot channels are required for CSI estimation as cluster size increase, hence reducing the resources available for user data. Moreover, precoding computation gets more complex and additional backhaul bandwidth is required as the cluster size increase. In this section, we formulate CoMP performance and overhead metrics to deploy in the our dynamic clustering problem.

Cell Load

Cell load can be interpreted as one of the key metrics to quantify CoMP gain and cost trade-off. As CoMP cluster size increases, interference from more cells are mitigated, and hence spectral efficiency is improved further which then reduces the cell load. On the other hand, with increased cluster size, more pilot resources are required for channel estimation which will reduce available PRB bandwidth for user data. This will then derive the load higher due to reduced PRB bandwidth.

As discussed in Chapter 3, cell load can be defined as the ratio of required PRBs for all users associated to the cell against the total available PRBs. We first define the average required PRBs for each UE_k at each cell. In no CoMP scenario, assuming constant GBR requirement d_k for UE_k, average PRB requirement for UE_k can be expressed as $r_k = d_k/(y_k B_{PRB})$ where $y_k = log_2(1 + SI\hat{N}R_k)$ and B_{PRB} is the total bandwidth for user data in a single PRB. In MU JT-CoMP, UE_k requires resources from all SCs within its user-centric cluster $\mathcal{C}_i^k$, and PRB resource for UE_k is shared between all users in $\mathcal{U}_i^k$ which are scheduled within

the same cluster. We assume $|\mathcal{C}_i^k| = |\mathcal{U}_i^k| = n_k$ and define an estimated dedicated PRBs for UE_k at each SC within $\mathcal{C}_i^k$ as $\hat{r}_k = r_k/n_k$).

Assume that SC_m is in coalition $\mathcal{C}_i$ and $\mathcal{U}_{im}$ is the associated active UEs in SC_m where $\mathcal{U}_{im} \subseteq \mathcal{U}_i$ i.e. SC_m is not connected to all users in $\mathcal{U}_i$ as user-centric clusters of some users may not include SC_m. Let R_{tot} be the total number of PRBs for each SC, assuming all SCs have same total bandwidth. Cell load on SC_m in cluster $\mathcal{C}_i$ can be expressed as:

$$\hat{l}_{im} = \frac{\sum_{k \in \mathcal{U}_{im}} \hat{r}_k}{R_{tot}} \tag{4.5}$$

Unsatisfied Users

Similar to the unsatisfied users definition we derived in Chapter 3, we define an unsatisfied users term with network-centric clustering notation. In MU JT-CoMP scenario, users are connected to more than one SC, hence associated connected user count for each SC_m ($\mathcal{U}_{im}$) will need to be adjusted for MU JT-CoMP scenario to avoid double-counting. We define an estimated dedicated user count for each SC by distributing the number of users to each SC within its user-centric cluster. Assume UE_k has user-centric cluster of $\mathcal{C}_i^k$ with $|\mathcal{C}_i^k| = n_k$. We define the estimated dedicated user count at SC_m in cluster $\mathcal{C}_i$ as $\hat{u}_{im} = \sum_{k \in \mathcal{U}_{im}} 1/n_k$.

Unsatisfied users for each SC_m in $\mathcal{C}_i$ can then be expressed as:

$$\hat{z}_{im} = \max\left(0, \hat{u}_{im}\left(1 - \frac{1}{\hat{l}_{im}}\right)\right) \tag{4.6}$$

Additional Pilot Overhead:

One of the challenges for CoMP is the requirement for additional pilot channels for CSI estimation in downlink as the number of TPs in coordination increases [86]. Using the optimum pilot overhead estimation for multi-antenna channels in [86]:

$$\begin{aligned}\alpha = &\sqrt{(1+SNR)\frac{\dot{C}(SNR)}{C(SNR)}2n_T f_D} \\ &- \left((1+SNR)\frac{\ddot{C}(SNR)}{\dot{C}(SNR)} + 2 + \frac{1}{2SNR}\int_{-1}^{+1}\frac{d\xi}{\tilde{S}_H(\xi)}\right) n_T f_D \\ &+ O(f_D^{3/2})\end{aligned} \tag{4.7}$$

where

$C(SNR) = \mathbb{E}[\log_2(1 + SNR|H|^2)]$
$\dot{C}(SNR) = \frac{1}{SNR}\left(\log_2 e - \frac{C(SNR)}{SNR}\right)$
$\ddot{C}(SNR) = \frac{1}{SNR^2}\left[\log_2 e + \dot{C}(SNR - 2\frac{C(SNR)}{SNR}\right]$

$\tilde{S}_H(\xi)$ is the doppler spectrum of the wireless channel.
f_D is the normalised doppler frequency
n_T is the number of transmit antennas
and α is percentage pilot overhead bandwidth requirement.

Figure 4.2 shows the optimum overhead required for three typical wireless channels widely used by 3GPP [3] for Clarke-Jakes spectrum with SNR=10dB. To estimate the pilot training overhead for any cluster $\mathcal{C}_i$, we adapted the pilot requirement from (4.7) for extended pedestrian-a (EPA-A) case where: $f_D = 0.000357$ and the term $\int_{-1}^{+1} \frac{d\xi}{\tilde{S}_H(\xi)}$ simplifies to $\pi^2/2$ for Clarke-Jakes spectrum. We assume SNR=10 for training overhead estimation and one antenna for each SC, hence $n_T = |\mathcal{C}_i|$.

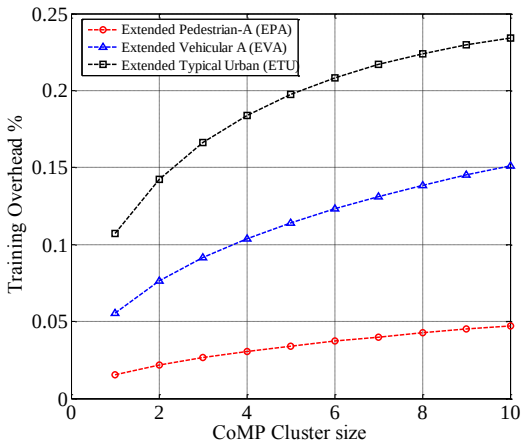


Figure 4.2: Optimum pilot overhead vs CoMP cluster size [86].

Pilot overhead increases with cluster size $|\mathcal{C}_i|$, and hence the actual bandwidth of a PRB for user data is reduced. For example, for EPA-A wireless channel with above assumptions, pilot overhead for each PRB will be 2.18% when cluster size is 2 and it increases to 3.05% and 3.71% for cluster size 4 and 6 respectively as depicted in Figure 4.2. Consequently, the available bandwith for user data for each PRB will reduce to 97.82%, 96.95% and 96.29% compared to total PRB bandwidth for cluster size 2,4 and 6 respectively. This will then be reflected on the overall available capacity/load of all SCs within cluster $\mathcal{C}_i$. Adjusted PRB bandwidth available for user data can be expressed as:

$$b_{PRB} = B_{PRB}(1 - \alpha) \tag{4.8}$$

Other Challenges

There are other challenges of CoMP implementation such as precoding, scheduling complexity and required backhaul bandwidth which increase as cluster size $|\mathcal{C}_i|$ increases. To account for these additional costs, we define complexity factor $c(|\mathcal{C}_i|)$. A soft maximum cluster size limit is imposed within the complexity factor where the cost of CoMP is sharply increased beyond a maximum cluster size limit $|\mathcal{C}_i| > C^n_{max}$. $|\mathcal{C}_i|$ can still increase beyond C^n_{max} in extreme conditions where the associated spectral efficiency/load gain is higher than the increased cost. For any cluster $\mathcal{C}_i$, complexity function is estimated as a sigmoidal function as follows:

$$c(|\mathcal{C}_i|) = \frac{1}{1 + e^{-(|\mathcal{C}_i| - C^n_{max})}} \tag{4.9}$$

C^n_{max} is designed to be an input parameter for the algorithm where it can be adjusted based on signal processing capacity and backhaul availability of the network. Figure 4.3 depicts the complexity factor used in our simulations when soft maximum cluster size is set to $C^n_{max} = 6$. A similar sigmoidal function is employed in [70] to introduce a soft limit to cluster size and penalize cluster size above a certain limit.

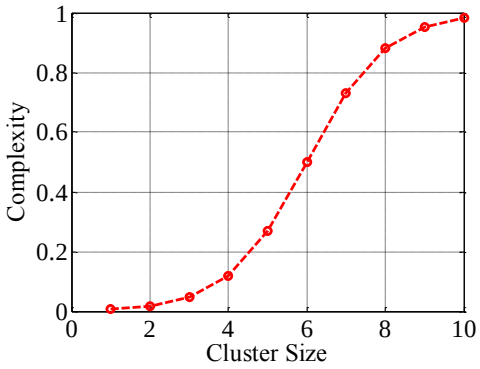


Figure 4.3: Complexity vs cluster size $|\mathcal{C}_i|$ when ($C^n_{max} = 6$).

4.4 Dynamic Network-centric Clustering Problem as a Coalition Game

The core aim of this chapter is to design the optimum network clustering model $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_s\}$ for any given list of SCs in $\mathcal{C}$ to optimize spectral efficiency and load balancing jointly. As briefly discussed in Section 4.1, we utilize coalitional game theory to design a novel, scalable, low complexity framework for a

load aware CoMP clustering structure to maximize load balancing and spectral efficiency objectives. In this section, we first define the concepts for our coalition formation game model based on two simple transformation rules: merge and split. We then define two novel utility functions to employ in our coalition formation game. We discuss load aware utility in detail and trade-off between performance improvement in load balancing/spectral efficiency against the increased system complexity. We then present our novel network-centric clustering algorithm as a merge/split coalition game and discuss its complexity and stability properties.

4.4.1 Coalition Formation Game Concepts

Let $\mathcal{C} = \{SC_1, SC_2, ..., SC_n\}$ be the players of the game, i.e. each player representing an SC in our scenario. Grand coalition is defined as the unique group of all cells in the game, i.e. $\mathcal{C}$ itself. Any cluster of SCs within the grand coalition is defined as a coalition $\mathcal{C}_i = \{SC_{i1}, SC_{i2}...SC_{iz}\}$. A collection is defined as a group of coalitions $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_s\}$ and a collection is called a partition if all coalitions within $\mathcal{C}$ are disjoint coalitions i.e. $\forall i \neq j, \mathcal{C}_i \cap \mathcal{C}_j = \varnothing$ and all players (SCs) are included in one of the coalitions i.e. $\cup_{i=1}^{s} \mathcal{C}_i = \mathcal{C}$.

The utility (payoff) of a coalition $\mathcal{C}_i$ within the partition $\mathcal{C}$ is defined as $v(\mathcal{C}_i, \mathcal{C})$ and the overall coalition game is uniquely defined by $(\mathcal{C}, v)$ pair. Utility of any coalition should reflect the overall CoMP gain including both the benefit and the cost for cooperation. Utility function for CoMP clusters in our scenario takes spectral efficiency and cell load distribution into account as benefits and include a cost factor to account for increased computational complexity, pilot overhead and backhaul requirement with increased cluster size. The cost factor in the utility prevents a super-additive game, i.e. the cost increases with cluster size and hence it is mostly impossible to get all SCs to cooperate in a single cluster.

Characteristic form of a coalition game is defined such that the utility of any coalition $v(\mathcal{C}_i)$ does not depend on how the rest of the partition $(\mathcal{C} \backslash \mathcal{C}_i)$ is structured i.e. $\forall i \; v(\mathcal{C}_i, \mathcal{C}) = v(\mathcal{C}_i)$. In our scenario, since we propose clustering changes in longer time intervals (seconds, minutes) where fast fading changes are averaged out as expressed in (4.4), the amount of interference created from the cells outside of the cluster are the same regardless of their clustering structure. Hence our scenario can be modelled as a coalition game in characteristic form. We make use of this property to reduce complexity of our algorithm as detailed in Section 4.4.4.

To compare the preference between two collections $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_z\}$ and $\mathcal{H} = \{\mathcal{H}_1, \mathcal{H}_2, ..., \mathcal{H}_b\}$ of the same subset of players $\mathcal{C}_{sub}$ where $\mathcal{C}_{sub} \subseteq \mathcal{C}$, we define a comparison relation $\triangleright$, where $\mathcal{G} \triangleright \mathcal{H}$ means that coalitions in $\mathcal{G}$ is preferred to

the coalitions in $\mathcal{H}$. Various comparison orders are discussed in [27] but two orders are of notable importance for coalitional games for cooperative wireless networks [148]. First one is the utilitarian order which compares the utility of the overall collection. The players in $\mathcal{C}_{sub}$ prefer to move to collection $\mathcal{G}$ from collection $\mathcal{H}$ i.e. $\mathcal{G} \triangleright \mathcal{H}$ if $\sum_{i=1}^{z} v(\mathcal{G}_i) > \sum_{i=1}^{b} v(\mathcal{H}_i)$, in other words, the total utility of all coalitions within collection $\mathcal{G}$ is greater than the one in collection $\mathcal{H}$, irrespective of individual player utilities. The second important order is known as pareto order which compares the individual player utilities to make sure none of the players are worse off due to new collection formation and at least one player is better off. For a given subset of players $\mathcal{C}_{sub}$, the utility of player SC_m in collection $\mathcal{G}$ is denoted as $v(SC_m, \mathcal{G})$; then $\mathcal{G} \triangleright \mathcal{H}$ if $\forall i \in \mathcal{C}_{sub},\ v(SC_i, \mathcal{G}) \geq v(SC_i, \mathcal{H})$.

It is highly appealing to employ utilitarian order in our coalition game to maximize the overall system utility. The aim of our proposed coalition formation game is to maximize the total utility regardless of the utility for any individual SC. In other words, if the utility gain of a group of SCs is higher than the utility loss of the remaining SCs, then the corresponding clustering change shall be performed. In a typical hotspot scenario, cluster changes aim to reduce load for SCs with very high load (players with better payoff) but this will inevitably cause increased traffic in other SC where load is not as high (players with worse payoff). This clustering change is preferred in utilitarian order if the overall utility is increased however this is not allowed in pareto order as some players are worse off regardless of the overall utility.

To form coalitions and dynamically adapt the coalitions based on user profile/network changes, two simple transformation rules are followed:

- **Merge:** Players (SCs) in any two or more coalitions $\{\mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_z\}$ prefer to merge into one coalition $\mathcal{F} = \cup_{i=1}^{z} \mathcal{G}_i$ i.e. $\cup_{i=1}^{s} \mathcal{G}_i \triangleright \{\mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_z\}$, if $v(\mathcal{F}) > (\sum_{i=1}^{s} v(\mathcal{G}_i))$ following the utiliterian order.

- **Split:** Players (SCs) prefer to split from any coalition $\mathcal{C}_i$ into smaller coalitions $\{\mathcal{C}_{i1}, \mathcal{C}_{i2}, ..., \mathcal{C}_{iy}\}$ where $\mathcal{C}_i = \cup_{j=1}^{y} \mathcal{C}_{ij}$ i.e. $\{\mathcal{C}_{i1}, \mathcal{C}_{i2}, ..., \mathcal{C}_{iy}\} \triangleright \mathcal{C}_i$ if $(\sum_{j=1}^{y} v(\mathcal{C}_{ij}) > v(\mathcal{C}_i)$ following utiliterian order.

4.4.2 Utility Function

Utility function $v(SC_m, \mathcal{C}_i)$ is defined to calculate payoff for any SC_m (player) in coalition $\mathcal{C}_i$ and payoff for any coalition $v(\mathcal{C}_i)$ is simply the total payoff of all SCs within the coalition i.e. $v(\mathcal{C}_i) = \sum_{SC_j \in \mathcal{C}_i} v(SC_j, \mathcal{C}_i)$. Utility function should reflect both the proposed performance improvement and the associated overhead costs of any coalition formation. Firstly, we define a load aware utility function aiming

to jointly maximize spectral efficiency and load balancing objectives. The goal is to distribute SC load evenly and relieve congestion in hotspot scenarios while keeping spectral efficiency at high levels and also provide high spectral efficiency in non-hotspot scenarios when load balancing is not required. Secondly, we define an spectral efficiency based utility intending to maximize spectral efficiency only for comparison to our load aware utility.

1. **Load aware utility:** For any SC_m in coalition $\mathcal{C}_i$, load-based utility function is defined as follows:

$$v_1(SC_m, \mathcal{C}_i) = \begin{cases} \frac{-(\hat{l}_{im})}{1-c(|\mathcal{C}_i|)}\hat{u}_{im} & \hat{l}_{im} < 1 \\ \frac{-(\hat{l}_{im})^3}{1-c(|\mathcal{C}_i|)}\hat{u}_{im} & \hat{l}_{im} \geq 1 \end{cases} \tag{4.10}$$

The main aim of the load aware utility is to jointly optimize load balancing and spectral efficiency by reducing SC load $\hat{l}_{im}$ which then implicitly enforces for better spectral efficiency. When spectral efficiency is improved, less radio resources are used for any given demand, and hence load is reduced. Payoff for each SC_m e.g. $v_1(SC_m, \mathcal{C}_i)$ is reduced as SC_m load $\hat{l}_{im}$ increases. Once the cell is congested (i.e. $\hat{l}_{im} \geq 1$), any load increase is penalized more than the case when $\hat{l}_{im} < 1$. This is achieved by increasing the impact of load with the term $(\hat{l}_m)^3$ in the utility function in (4.10) when $\hat{l}_{im} \geq 1$. In other terms, additional payoff incentive is introduced for reducing the load in high load range, when compared to light load, i.e. enabling load distribution from congested SCs to lightly loaded SCs. In the high load range, distribution of load is given higher priority and hence clustering decisions in this range will prioritize load balancing improvement despite other clustering solutions may be available with better overall spectral efficiency. In light load range, $v_1(SC_m, \mathcal{C}_i)$ will provide similar results to spectral efficiency based utility as SC load reduction implicitly enforces higher spectral efficiency. Load aware utility $v_1(SC_m, \mathcal{C}_i)$ is also directly proportional with estimated dedicated user count $\hat{u}_{im}$ i.e. highly loaded cells with more active users are given more incentive to reduce load and achieve better payoff. This promotes fairness in the system and aims to reduce the total number of unsatisfied users $\hat{z}_{im}$ at each SC. Term $c(|\mathcal{C}_i|)$ in $v_1(SC_m, \mathcal{C}_i)$ represents the complexity factor as the cluster size increases. Complexity function $c(|\mathcal{C}_i|)$ enforces low cluster size $|\mathcal{C}_i|$, by introducing high payoff penalty as the cluster size increases. Cluster size is only increased when the payoff incentive from reducing the load is higher than the payoff penalty introduced with $c(|\mathcal{C}_i|)$.

Figure 4.4 illustrates the utility function $v_1(SC_m, \mathcal{C}_i)$ against SC load $\hat{l}_{im}$ for different cluster sizes $|\mathcal{C}_i|$ for $\hat{u}_{im}$=50 and $c(|\mathcal{C}_i|) = \frac{1}{1+e^{-(|\mathcal{C}_i|-C^n_{max})}}$ when $C^n_{max} = 6$. It can be seen that payoff only gradually increases as the load decrease in light load range, whereas there is sharper payoff increase in high load, in other terms, the load aware utility provides additional payoff incentive to reduce load in high load range. On the other hand, increasing cluster size is penalized with complexity factor $c(|\mathcal{C}_i|)$ where a sharp payoff penalty is observed especially moving from $|\mathcal{C}_i| = 5$ to $|\mathcal{C}_i| = 6$ in this example. A higher cluster size is expected in high load when compared the light load as the payoff incentive for reducing the load is higher in high load range as introduced in $c(|\mathcal{C}_i|)$. A dynamic trade-off between cluster size/complexity and spectral efficiency/load is formed with this utility where maximum cluster size limit is dynamically adjusted based on load situation in the network. The cost/gain factors and the trade-off between system complexity and load balancing/spectral efficiency performance in $v_1(SC_m, \mathcal{C}_i)$ provides a sample which can be adjusted based on specific radio network operator priorities. For example, in a highly customer-centric network, performance can be favored more than complexity in hotpots and to minimize the number of unsatisfied users due to congestion, term $\hat{l}^3_{im}$ when $\hat{l}_{im} \geq 1$ can be adjusted to give more payoff incentive for reducing load in high load range. Similarly, $c(|\mathcal{C}_i|)$ can be adjusted to increase maximum allowed cluster size in high/light load ranges.

Our simulation results in Section 4.5 show the proposed dynamic cluster size adaptation depending the load situation, i.e. increasing cluster size dynamically when there is high load and hence improve spectral efficiency/load balancing performance in both HN and RN scenarios.

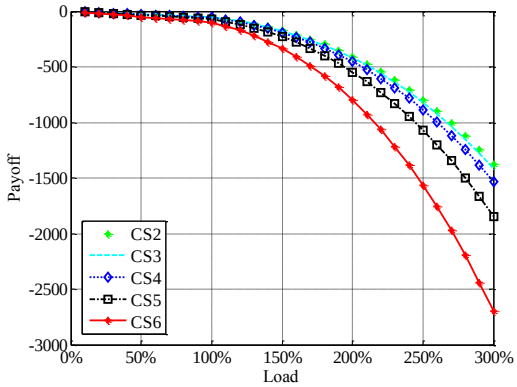


Figure 4.4: Utility function $v_1(SC_m, \mathcal{C}_i)$ vs SC load $\hat{l}_m$ for different cluster sizes when $\hat{u}_m = 50$ and $C^n_{max} = 6$.

2. **Spectral efficiency based utility:** We define a second utility function to maximize spectral efficiency, without considering load conditions. This utility is employed in our game model and in a greedy algorithm to compare with our load aware utility.

   Spectral efficiency based utility function is introduced as follows:

$$v_2(SC_m, \mathcal{C}_i) = \sum\nolimits_{k \in \mathcal{U}_{im}^{best}} y_k(1 - c(|\mathcal{C}_i|)) \tag{4.11}$$

   where:
   $\mathcal{U}_{im}^{best}$ is the list of users where SC_m is the best serving cell based on average received signal power, i.e. a subset of the associated users $\mathcal{U}_{im}$ at the SC_m.
   y_k is the spectral efficiency achieved at UE_k i.e. $y_k = log_2(1 + S\hat{I}\hat{N}R_k)$.

   The spectral efficiency experienced at each user is added up to get the total utility at SC_m and complexity factor $c(|\mathcal{C}_i|)$ is embedded to impose a soft cluster size limit similar to the one in load aware utility in (4.10).

Similar to the two utility functions presented above, other utilities can be designed to optimize different network objectives like spectral efficiency, load balancing, energy efficiency and backhaul availability etc. Furthermore, a combination of different network objectives can be embedded within the same utility function to jointly optimize multiple network objectives. Our novel clustering model based on merge/split coalitional game sets a flexible framework to employ various utility functions aiming for different network objectives.

4.4.3 Merge/Split Operation

Let $\{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_s\}$ be a partition of $\mathcal{C}$, i.e. the current status of the network. We propose to start merge operation with $\mathcal{C}_i$ which has got the maximum absolute payoff value. In both utility functions defined in (4.10) and (4.11), high absolute payoff value refers to coalitions with high number of active users and hence high load. Coalition $\mathcal{C}_i$ looks for neighbour coalitions $\mathcal{C}_j$ for any possible merge operation. We define neighbour coalition concept to avoid exhaustive search for merge operation and reduce complexity, i.e. merge operation will not be tried for every other coalition in the system but only towards the neighbour coalitions. Neighbour definitions are performed by utilizing the average received reference signal level measurements received from the users. Firstly a simple neighbour relations list is performed at SC level. For any UE_k in the serving area of SC_m, the average received signal from all other SC_j where $p_{kj} > P_{min}^{nei}$ are compared. A neighbour rank count is incremented for $\{SC_m, SC_j\}$ pair if $p_{kj}/p_{km} > P_{\Delta}^{nei}$.

Once each SC has a rank based neighbour list, then neighbours at cluster level are calculated in a similar way i.e. the neighbour rank is incremented for $\{\mathcal{C}_m, \mathcal{C}_j\}$ coalition pair when $SC_m \subseteq \mathcal{C}_m$, $SC_j \subseteq \mathcal{C}_j$, $p_{kj}/p_{km} > P_{\Delta}^{nei}$ and $p_{kj} > P_{min}^{nei}$.

Algorithm 3 Merge Operation

For any given network clustering state $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_s\}$, $\forall \mathcal{C}_i \in \mathcal{C}$, set $\mathcal{C}_i$.clustered=0
Merge-ongoing=1
while Merge-ongoing **do**
  Merge-ongoing=0
  Sort $\forall \mathcal{C}_i \in \mathcal{C}$ based on $v(\mathcal{C}_i)$ in descending order
  for all $\mathcal{C}_i$ where $\mathcal{C}_i$.clustered=0 **do**
    Update $\mathcal{C}_i$.nei
    for all $\mathcal{C}_j$ in $\mathcal{C}_i.nei$ where $\mathcal{C}_j$.clustered=0 **do**
      Update payoff gain for possible merge$(\mathcal{C}_i, \mathcal{C}_j)$ i.e. $\delta_{v_{ij}} = v(\mathcal{C}_i \cup \mathcal{C}_j) - \{v(\mathcal{C}_i) + v(\mathcal{C}_j)\}$
    end for
    Find $\mathcal{C}_m \in \mathcal{C}_i.nei$ where $\delta_{v_{im}} = \underset{\mathcal{C}_j \in \mathcal{C}_i.nei}{\arg\max}(\delta_{v_{ij}})$ and $\delta_{v_{im}} > 0$
    while $\mathcal{C}_m$ exist **do**
      Merge$(\mathcal{C}_i, \mathcal{C}_m)$
      $\mathcal{C}_m$.clustered=1
      Update $\mathcal{C}_i$.nei
      for all $\mathcal{C}_j$ in $\mathcal{C}_i.nei$ where $\mathcal{C}_j.clustered$=0 **do**
        Update payoff gain for possible merge $(\mathcal{C}_i, \mathcal{C}_j)$ i.e. $\delta_{v_{ij}} = v(\mathcal{C}_i \cup \mathcal{C}_j) - \{v(\mathcal{C}_i) + v(\mathcal{C}_j)\}$
      end for
      Find $\mathcal{C}_m \in \mathcal{C}_i.nei$ where $\delta_{v_{im}} = \max_{\mathcal{C}_j \in \mathcal{C}_i.nei}(\delta_{v_{ij}})$ and $\delta_{v_{im}} > 0$
    end while
    $\mathcal{C}_i$.clustered=1
    if Any merge operation with $\mathcal{C}_i$ **then**
      Break for-loop and continue with while-loop
      Merge-ongoing=1
    end if
  end for
end while

The possibility of a merge operation is checked for all neighbour coalitions of coalition $\mathcal{C}_i$ and merge is performed with $\mathcal{C}_j$ if $v(\mathcal{C}_i \cup \mathcal{C}_j) > v(\mathcal{C}_i) + v(\mathcal{C}_j)$ based on utilitarian order as described in Section 4.4.1. Once a merge operation is successful, then neighbour lists are updated for the new merged coalition $(\mathcal{C}_i \cup \mathcal{C}_j)$ and further possible merges are searched in a similar fashion until there is no more neighbours left for a possible merge operation. Same process is repeated for the rest of the coalitions in partition $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_s\}$ in absolute payoff value order as illustrated in Algorithm 3 until there is no more merges possible. A new par-

tition $\mathcal{H}$ if formed at the end of the merge operation. Partition $\mathcal{H}$ is then subject to split operation where every coalition $\mathcal{H}_i$ is checked for all possible split options and it is split only when the total payoff of the split coalitions are better than the bigger coalition following utilitarian order i.e. $(\sum_{j=1}^{y} v(\mathcal{H}_{ij}) > v(\mathcal{H}_i)$. Split operation is successively iterated for the rest of the coalitions in partition $\mathcal{H}$ until no more split is possible as outlined in Algorithm 4. Merge and split operations are then performed iteratively until there is no more merge and split possible and the algorithm terminates. Termination of the algorithm is always guaranteed as all merge and split operations aim for the same objective i.e. increase the overall system utility $v(\mathcal{C})$. There is always a finite number of merge/split operations possible for increasing $v(\mathcal{C})$ and the algorithm will terminate when there is no room to increase $v(\mathcal{C})$ by merge/split operations.

Algorithm 4 Split Operation

For any given network clustering state $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_s\}$, $\forall \mathcal{C}_i \in \mathcal{C}$, set $\mathcal{C}_i$.splitpossible=1
Split-ongoing=1
while Split-ongoing **do**
  Split-ongoing=0
  for all $\mathcal{C}_i$ where ($\mathcal{C}_i$.split-possible=1 and $|\mathcal{C}_i| > 1$) **do**
    Update $\mathcal{C}_i$.Split-options
    $\mathcal{C}_i$.split-possible=0
    for all $\mathcal{C}_i$.Split-options **do**
      if Any split option is possible i.e. $(\sum_{j=1}^{y} v(\mathcal{C}_{ij}) > v(\mathcal{C}_i)$ **then**
        Split($\mathcal{C}_i$ to $\{\mathcal{C}_{i1}, \mathcal{C}_{i2}, ..., \mathcal{C}_{iy}\}$
        Split-ongoing=1
        $\forall \mathcal{C}_{ij}$, set $\mathcal{C}_{ij}$.split-possible=1
        Break for-loop and continue with next $\mathcal{C}_i$
      end if
    end for
  end for
end while

4.4.4 Algorithm Complexity

Exhaustive search for any potential merge operation can increase complexity of the algorithm exponentially as the network size increase. Unlike exhaustive search proposed in previous network-centric clustering solutions like [70, 134], we define neighbour cluster concept as described in section 4.4.3 and propose merge operation only with neighbour clusters which reduces merge operation complexity and makes the algorithm scalable for larger networks.

Split operation can be a complex task as the number of possible splits increase exponentially with cluster size. To reduce this complexity, we utilize the characteristic form property of our coalition game where the possibility of any split operation does not depend on how the rest of the SCs are clustered. Once we check a coalition for a possible split and if there is no possible split operation, then even when the rest of the network is re-clustered, marked coalitions will not be checked again for split in the following iterations.

Furthermore, in our coalitional game model, we have a soft maximum cluster size limit C^n_{max} embedded in both utility functions to avoid increased signal processing and backhaul bandwidth required for CoMP. This limitation reduces the complexity on the split operation, i.e. less number of possible split options are available due to limited cluster size. Additionally, the split operation stops searching for other split options once a split option with better utility is found and hence the split operation does not have to go through all split options in most cases.

In summary, we define a low complexity merge and split operation in our novel game-theoretic clustering algorithm: We limit the merge operations to only neighbour clusters which improves scalability of the solution and reduces complexity. Additionally, a soft maximum cluster size limit is embedded in both utility functions which reduces complexity on split operation preventing high number of potential splits. Furthermore, we make use of the characteristic form property of our coalitional game model and reduce split complexity further. The stability of the algorithm and convergence to the best outcome is discussed in the next section.

4.4.5 Partition Stability

We utilize a novel concept of defection function $\mathbb{D}$ introduced in [26] to analyze stability of our merge/split coalition game. Defection function $\mathbb{D}(\mathcal{C})$ of a partition $\mathcal{C}$ associates partition $\mathcal{C}$ with a set of collections. Partition $\mathcal{C}$ is defined as $\mathbb{D}$-stable if none of the players have any incentive to leave the partition to form collections allowed by $\mathbb{D}$.

The most robust stability is defined as $\mathbb{D}_c$ stable if it is the unique partition where the utility is maximum, i.e. there is no intention for any players to deviate into any other partition [26]. A partition $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_s\}$ is $\mathbb{D}_c$ stable only if below 2 conditions are satisfied [26]:

1. $\forall \mathcal{C}_i \in \mathcal{C}$, any disjoint coalitions $\mathcal{C}_{ia}$ and $\mathcal{C}_{ib}$ in $\mathcal{C}_i$ where $\mathcal{C}_{ia} \cup \mathcal{C}_{ib} \subset \mathcal{C}_i$, then $v(\mathcal{C}_{ia} \cup \mathcal{C}_{ib}) \geq v(\mathcal{C}_{ia}) + v(\mathcal{C}_{ib})$ i.e. any a sub-group of players in any coalition

do not have any additional payoff incentive to leave the coalition.

2. For any arbitrary coalition $\mathcal{A} \in \mathcal{C}$ where $\mathcal{A} \notin \mathcal{C}_i$ and all players in coalition $\mathcal{A}$ may not belong to the same coalition in $\mathcal{C}$, then: $\sum_{i=1}^{s} v(\mathcal{C}_i \cap \mathcal{A}) \geq v(\mathcal{A})$

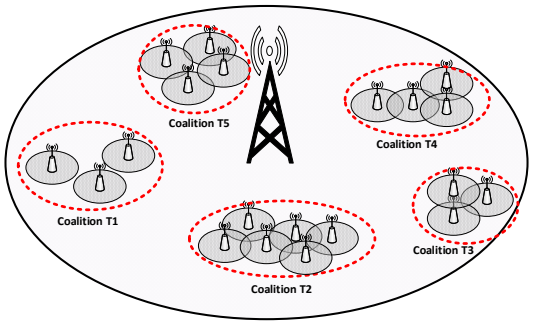


Figure 4.5: An illustration of $\mathbb{D}_c$ stable partitions.

$\mathbb{D}_c$ is the most desired form of stability as it is the unique partition with maximum utility, however partitions formed from merge/split game does not always guarantee $\mathbb{D}_c$ partitions. Our merge/split coalition game results in the $\mathbb{D}_c$ stable partition depending on the network and user profiles. In a typical SC deployment scenario, basic coverage is provided by the MBS, and SCs are deployed in hotspot areas only. There are hotspot areas within the MBS coverage area where most users and SCs are concentrated as illustrated in Figure 4.5. We show that both conditions of $\mathbb{D}_c$ stability are guaranteed in this deployment scenario as follows:

Condition 1 for a $\mathbb{D}_c$-stable partition $\mathcal{C}$ states that for each coalition $\mathcal{C}_i \in \mathcal{C}$, any 2 disjoint sub-coalitions $\mathcal{C}_{ia}, \mathcal{C}_{ib} \in \mathcal{C}_i$ will not have additional payoff to form separate coalitions. In our model, there is dense deployment of SCs and high concentration of users in small hotspot areas where inter-cell interference is high due to dense deployment in the absence of CoMP. Therefore, there is high payoff incentive to form $\mathcal{C}_i$ to include all SCs within the same hotspot area in both utility functions in (4.10) and (4.11) as severe inter-cell interference is mitigated, improving spectral efficiency (y_k in (4.11)) and hence reducing the cell load $\hat{l}_m$ in (4.10). The cost ($c(|\mathcal{C}_i|)$) of forming this coalition is kept low when $|\mathcal{C}_i| < C^n_{max}$ and it increases exponentially when the coalition size increase beyond C^n_{max}. Therefore, our coalition game forms the coalitions to include all SCs within the same hotspot when the number of SCs within the same hotspot do not exceed C^n_{max}.

Let $\mathcal{C}_i$ be the coalition including all SCs within any hotspot location where $|\mathcal{C}_i| < C^n_{max}$, and assume $\mathcal{C}_{ia}$ and $\mathcal{C}_{ib}$ are 2 disjoint sub-coalitions of coalition

$\mathcal{C}_i$. Individual SCs in $\mathcal{C}_i$ will not have better payoff for leaving $\mathcal{C}_i$ to form a smaller sub-cluster $\mathcal{C}_{ia}$ i.e. $\forall SC_m \in \mathcal{C}_i,\ v(SC_m, \mathcal{C}_i) > v(SC_m, \mathcal{C}_{ia}))$. SCs will have better payoff in bigger clusters due to improved inter-cell interference mitigation, provided that the size of the bigger cluster does not exceed C^n_{max} , i.e. for any 2 disjoint sub-clusters, $\forall SC_m \in \mathcal{C}_{ia},\ v(SC_m, (\mathcal{C}_{ia} \cup \mathcal{C}_{ib})) > v(SC_m, \mathcal{C}_{ia})$, and thus $v(\mathcal{C}_{ia} \cup \mathcal{C}_{ib}) > v(\mathcal{C}_{ia}) + v(\mathcal{C}_{ib})$. Condition 1 for $\mathbb{D}_c$ stability is satisfied when $|\mathcal{C}_i| < C^n_{max}$.

For a $\mathbb{D}_c$-stable partition $\mathcal{C}$, condition 2 states that any players from different coalitions $\mathcal{C}_i$ and $\mathcal{C}_j$ have no additional payoff to form another coalition $\mathcal{H}_i$ where $\mathcal{H}_i \notin \mathcal{C}$. Let $\mathcal{C}_i$ and $\mathcal{C}_j$ be the coalitions of SCs in two separate hotspots and $|\mathcal{C}_i| < C^n_{max}$ and $|\mathcal{C}_j| < C^n_{max}$ so condition 1 of a $\mathbb{D}_c$ stability is satisfied i.e. all SCs within the same hotspot are in the same coalition with no incentive to leave and form smaller coalitions. There is no incentive for any $SC_i \in \mathcal{C}_i$ and $SC_j \in \mathcal{C}_j$ to form another coalition if the distance between the two are $> d_0$ where no user UE_k have any incentive to have both SC_i and SC_j in the same user-centric cluster $\mathcal{C}^k_i$. This is guaranteed when for any UE_k, if the average received signal power from SC_i from distance d_{ki} is above the minimum threshold for clustering i.e. $p_{ki}(d_{ki}) > P_{min}$, then $p_{kj}(d_{kj})$ should be below P_{min}. Similarly, if $p_{kj}(d_{kj}) > P_{min}$, then $p_{ki}(d_{ki}) < P_{min}$. Assuming distance based path loss only for simplification, to satisfy this condition for any UE_k, the worst case scenario is considered where UE_k is located in the middle of SC_i and SC_j with equal distance. If the distance between SC_i and SC_j is $> d_0$ where both $p_{ki}(d_0/2) < P_{min}$ and $p_{kj}(d_0/2) < P_{min}$ then it is guaranteed that SC_i and SC_j can never be in the same user-centric cluster $\mathcal{C}^k_i$ i.e. average received power will not be above P_{min} for both SCs for any arbitrary user, hence there is no incentive for SC_i to leave $\mathcal{C}_i$ to form a new coalition with SC_j from coalition $\mathcal{C}_j$, i.e. condition 2 is satisfied.

In summary, our merge/split formation game results in a $\mathbb{D}_c$ stable partition in our typical deployment scenario where merge/split operation results in forming coalitions including all SCs within the local hotspot areas if the number of SCs within the same hotspot is not higher than the maximum cluster size limit and the distance between the hotspot areas are $> d_0$. Figure 4.6 shows the clustering results from our simulations for our merge/split cluster formation game in a typical SC deployment scenario in hotspots. Unique $\mathbb{D}_c$ stable partition is achieved in this deployment scenario where all local SCs within the same hotspot are included in the same cluster.

In the case when the SC deployment is not so dense, or low power SCs are used with almost isolated coverage areas, i.e. there is very limited inter-cell interference, proposed solution will not form clusters around all cells within hotspot

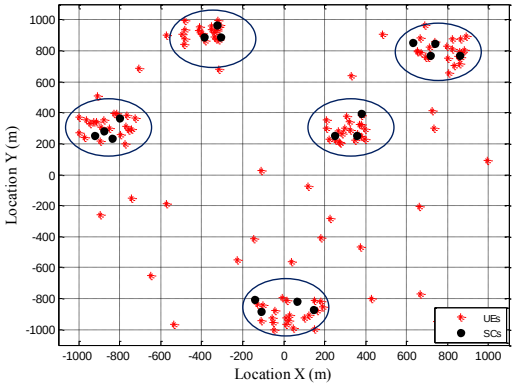


Figure 4.6: $\mathbb{D}_c$ stable partitions from merge/split cluster formation game.

as there will not be enough payoff incentive to justify cluster formation. This intuitively implies that expected CoMP gain would be minimal in this scenario, and there will not be a unique $\mathbb{D}_c$ stable partition around all SCs within the same hotspot area. In other scenarios, where there is no specific hotspot deployment, or the number of cells within hotspot area exceed C^n_{max}, unique $\mathbb{D}_c$ stable partition will not exist. In these cases, a more relaxed defection function $\mathbb{D}_{hp}$ is defined in [26] where for any partition $\mathcal{C}$, players are allowed leave to form another partition only by means of possible merge and splits. As our coalition game only follows merge/split operations and always terminates as there is only finite number of merge/split possible which can increase the overall system utility, all partitions resulting from our merge/split coalition game are always $\mathbb{D}_{hp}$ stable. $\mathbb{D}_{hp}$ stability does not have to be unique and other partitions with better utility may exist. To improve the merge/split game clustering outcome when $\mathbb{D}_c$ stability is not possible, we propose to start merge operations from the coalition with the maximum absolute payoff value aiming to achieve better utility for the loaded cells and maximize the resulting partition utility.

In summary, $\mathbb{D}_c$ stability provides the most desired unique partition with maximum utility and this is achievable in our merge/split game in certain network conditions which is most likely to be the deployment scenario for future networks. In the case when $\mathbb{D}_c$ stability is not possible, all partitions from our merge/split game are $\mathbb{D}_{hp}$ stable.

4.5 Simulation Results

To evaluate the performance of the proposed load aware, game-theoretic clustering framework, simulations are run for both HN and RN scenarios with various hotspot schemes. We excluded a comparison to the ideal partition with the best

utility based on exhaustive search as it is too complex and almost impossible to calculate as network size increase. For example, the number of possible partitions for a 1O SCs is 115.975. To compare our load aware clustering model performance based on load-based utility in (4.10), we employed spectral efficiency based utility in (4.11) as well in our framework and additionally we compared simulation results with an improved version of the greedy algorithm proposed in [134]. We adapted our novel spectral efficiency based utility function (4.11) in the greedy algorithm and lifted the maximum cluster size limit in [134] as the cluster size is self-limited with cost function $c(|\mathcal{C}_i|)$ within the utility function in (4.11). We also reduced complexity of the algorithm in [134] and utilized our neighbour coalition concept where only neighbour coalitions are considered for possible clustering as described in Section 4.4.4. Greedy algorithm starts with a random SC and forms clusters with neighbour SCs starting from the SC with maximum joint payoff. Unlike merge/split game clusters, greedy clusters lack on additional split functionality and also the randomness of the starting SC can provide under-optimized clustering solution depending on which SC the algorithm starts with. Algorithm 5 shows a summary of the enhanced version of the greedy algorithm presented in [134].

Algorithm 5 Greedy Clustering

Initiate Clusters i.e. $\forall \mathcal{C}_i \in \mathcal{C}$, $\mathcal{C}_i = \{SC_i\}$ and $\mathcal{C}_i$.clustered=0
for all $\mathcal{C}_i$ where $\mathcal{C}_i$.clustered=0 **do**
  Update $\mathcal{C}_i$.nei
  for all $\mathcal{C}_j$ in $\mathcal{C}_i.nei$ where $\mathcal{C}_j$.clustered=0 **do**
    Update payoff gain for possible merge$(\mathcal{C}_i, \mathcal{C}_j)$ i.e. $\delta_{v_{ij}} = v(\mathcal{C}_i \cup \mathcal{C}_j) - \{v(\mathcal{C}_i) + v(\mathcal{C}_j)\}$
  end for
  Find $\mathcal{C}_m \in \mathcal{C}_i.nei$ where $\delta_{v_{im}} = \underset{\mathcal{C}_j \in \mathcal{C}_i.nei}{\arg\max}(\delta_{v_{ij}})$ and $\delta_{v_{im}} > 0$
  while $\mathcal{C}_m$ exist **do**
    Merge$(\mathcal{C}_i, \mathcal{C}_m)$
    $\mathcal{C}_m$.clustered=1
    Update $\mathcal{C}_i$.nei
    for all $\mathcal{C}_j$ in $\mathcal{C}_i.nei$ where $\mathcal{C}_j.clustered$=0 **do**
      Update payoff gain for possible merge$(\mathcal{C}_i, \mathcal{C}_j)$ i.e. $\delta_{v_{ij}} = v(\mathcal{C}_i \cup \mathcal{C}_j) - \{v(\mathcal{C}_i) + v(\mathcal{C}_j)\}$
    end for
    Find $\mathcal{C}_m \in \mathcal{C}_i.nei$ where $\delta_{v_{im}} = \underset{\mathcal{C}_j \in \mathcal{C}_i.nei}{\arg\max}(\delta_{v_{ij}})$ and $\delta_{v_{im}} > 0$
  end while
  $\mathcal{C}_i$.clustered=1
end for

Following abbreviation is used in the rest of this section:

SE-GR: Greedy clustering with spectral efficiency based utility.
SE-GA: Game-theoretic clustering with spectral efficiency based utility.
L-GA: Game-theoretic clustering with load-based utility.

A network of SCs within one MBS is considered for our simulations as described in Section 4.3. Each SC is assumed to have one cell with omni directional antenna for simplicity. ITU-R microcell urban NLOS path loss model in [83] is adapted in our simulation as given below where d is is the distance between UE and SC in meters and fc is the carrier frequency in GHz.

$$PL = 36.7 \log_{10}(d) + 22.7 + 26 \log_{10}(fc) \tag{4.12}$$

For user-centric clustering algorithm, minimum received power threshold P_{min} is set to -110dBm and P_{Δ} is set to 20dB to consider all candidate SCs into the user-centric clusters. More relaxed settings can be used to utilize SCs in user-centric clusters where received signal level is more closer to the best serving cell for the right balance between increased cluster size and associated CoMP complexity. Additionally, for the neighbor definitions utilized in merge algorithm, minimum received power threshold P_{min}^{nei} is set to -110dBm similar to P_{min} and maximum received power offset from the serving cell for neighbor definition P_{Δ}^{nei} is set to $-6dB$. A conservative approach is taken to pick only SCs/clusters within 6dB margin as neighbors and hence only these are considered for possible merge operation in the simulations. More relaxed settings can be selected for further merge options but with the expense of increased algorithm complexity as discussed in Section 4.4.4. Rest of the simulation parameters are summarized in Table 4.1.

We ran our simulation 100 times for each deployment scenario described below and present the results from the average of 100 snapshots.

4.5.1 Homogeneous Network (HN) Scenario

Firstly, we evaluate the performance for a HN deployment scenario where 25 SCs are located within the simulation area of 0.5kmx0.5km with 100m inter-site distance. 2 scenarios are evaluated in HN deployment:

- HN without hotspot: 250 UEs are distributed within the simulation area following uniform random distribution. A fixed GBR of 512kbps is assigned to each UE.

- HN with hotspot: 2 hotspot areas are assumed within the simulation area,

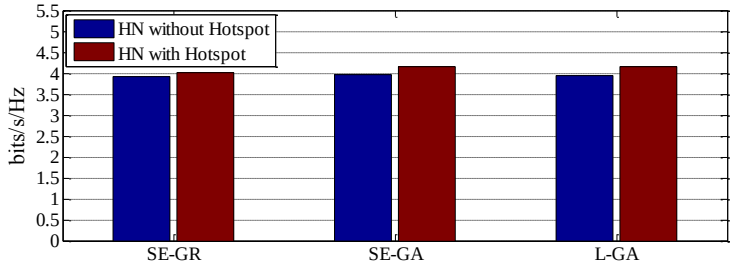
Table 4.1: Simulation Parameters

Parameter Name	Parameter Value
Simulation Enviroment	Urban Microcell [83]
Frequency Carrier	5 GHz
Channel Bandwidth	5 MHz
PRB Bandwidth (B_{PRB})	180 kHz
Number of PRBs/SC (R_{tot}	25
Shadow fading std	4 dB [83]
UE Antenna Gain	0 dBi
UE Thermal Noise Density	-174 dBm/Hz
TP Total Transmit Power (P_{Tx})	41dBm [83]
UE Noise Figure	7dB
TP Noise Figure (inc cable loss)	5dB
SC antenna gain (boresight)	17dBi
User-centric cluster: Min RX Power (P_{min})	-110dBm
User-centric cluster: Max RX power offset (P_{Δ})	20dB
Min RX power for neighbour Def. (P_{min}^{nei})	-110dBm
Max RX power offset for neighbour Def. (P_{Δ}^{nei})	-6dB
GBR for UEs in the hotspot in RN Scenario	512 kbps
GBR for UEs outside the Hotspot in RN Scenario	256 kbps
GBR for UEs in HN Scenario	512 kbps
SC Density for RN ($\lambda_{\mathcal{C}}$)	80SC/km^2
UE Density within Hotspot in RN Scenario ($\lambda_{\mathcal{U}_{high}}$)	4000UE/km^2
UE Density outside Hotspot in RN Scenario ($\lambda_{\mathcal{U}_{low}}$)	200UE/km^2
RN Simulation Area Radius	0.5km
RN SC deployment Area Radius	0.4km
RN Hotspot Area Radius	0.1km

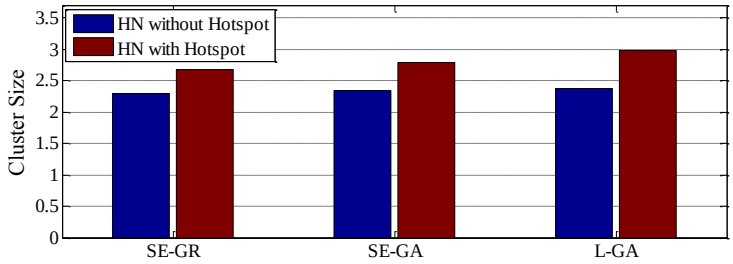
each hotspot is 100mx100m with 125 UEs distributed in each hotspot area following uniform random distribution. 250 additional UEs are uniformly distributed in the whole test area including the hotspot areas. All UEs have a fixed GBR requirement of 512 kbps.

Figure 4.7a depicts the average spectral efficiency and in HN with/without hotspot scenario for each of the clustering algorithms i.e. SE-GR, SE-GA and L-GA respectively. Without the hotspots, we observe similar spectral efficiency improvement on SE-GR and SE-GA algorithms. This is an expected outcome as possible merge-split iterations in the coalitional game model does not play an important role when compared to greedy algorithm in forming clusters due to even distribution of SCs and users. We also observe that L-GA algorithm also achieve similar spectral efficiency when compared to SE-GR/SE-GA, even though the employed utility function does not directly aim to maximize spectral efficiency. Load aware utility in L-GA aim to reduce load and improve load distribution which improves the spectral efficiency indirectly. Average cluster size for each of the algorithm in HN without hotspot scenario is also similar as depicted in Figure 4.7b. CS is controlled by the same cost function $c(|\mathcal{C}_i|)$ in both employed utility functions in (4.11) and (4.10), and hence similar average CS is expected for all 3 schemes in HN without hotspot scenario. It can be concluded that our novel load aware clustering model (L-GA) perform as good as spectral efficiency based algorithm in maximizing spectral efficiency when there is no over-load conditions.

A more realistic network scenario is when the users are not uniformly distributed and there are hotspots at certain locations. Clustering is more challenging in this scenario where any clustering combination without load awareness can potentially reduce achievable performance. Figure 4.8 shows a snapshot of clusters formed from SE-GR, SE-GA and L-GA clusters respectively in HN with hotspot scenario. Due to random selection of SCs for clustering, greedy algorithm (SE-GR) fails to get SCs within the same hotspot in the same cluster in this snapshot. However, SE-GA cluster starts the clustering process from the SC/cluster with maximum absolute payoff value and hence SC/clusters with higher load are given the priority on forming the clusters. SE-GA clusters manage to form clusters including the nearest SCs to the hotspots which then improves the spectral efficiency for majority of the UEs. L-GA clusters form larger clusters around the hotspots when compared to SE-GA and SE-GR. This is due to employed load aware utility (4.10) providing more payoff incentive for reducing load in high load conditions overcoming the cost of increasing cluster size $c(|\mathcal{C}_i|)$, resulting in bigger cluster size, improved inter-cell interference mitigation, better spectral efficiency and hence reduced load with the expense of increased processing complexity and



(a) Average spectral efficiency in HN with/without hotspot.



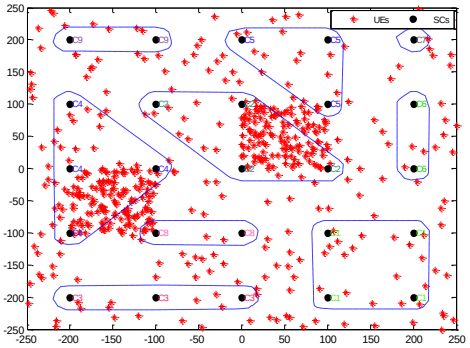
(b) Average cluster size in HN with/without hotspot.

Figure 4.7: Average spectral efficiency and cluster size in HN with/without hotspot scenarios.

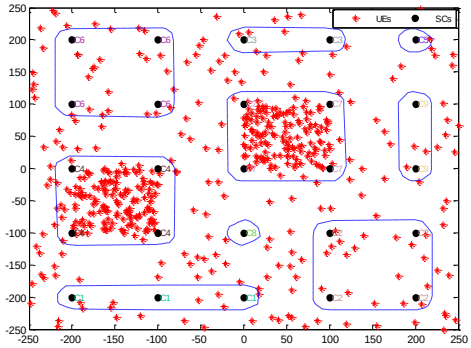
backhaul requirement.

SC load distribution in HN with hotspot scenario is depicted in Figure 4.9 where it is visible that highly loaded SCs are significantly reduced in L-GA clusters when compared to SE-GR/SE-GA clusters resulting in better load distribution. Consequently, a significant reduction in unsatisfied UEs is achieved in L-GA clusters when compared to SE-GR/SE-GA clusters as shown in Figure 4.10. Total number of unsatisfied users is reduced by 34.7% in L-GA when compared to SE-GR. A total of 12.95% of the UEs are unsatisfied in L-GA clusters whereas 19.85% and 18.18% of the UEs are unsatisfied in SE-GR and SE-GA clusters respectively. As depicted in Figure 4.7b, average cluster size is increased in SE-GA and L-GA models by 4.1% and 11.6% respectively when compared to SE-GR algorithm. Load aware utility function (4.10) in L-GA provides additional payoff incentive for reducing load at highly loaded cells which can overcome the cost of increased cluster size resulting in higher cluster size in hotspot scenario and hence the L-GA model responds to hotspots much better than SE-GR and SE-GA.

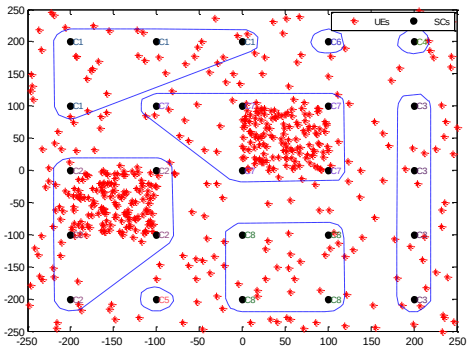
Our spectral efficiency based game theoretic clustering algorithm SE-GA also outperforms the greedy clustering SE-GR in hotspot scenario, where a marginal improvement in spectral efficiency is observed, resulting from the fact that SE-GR algorithm starts from any random cell for clustering, resulting in non-optimum clustering solutions especially in hotspot scenario. Moreover, SE-GR algorithm



(a) SE-GR clusters.



(b) SE-GA clusters.



(c) L-GA clusters.

Figure 4.8: Snapshot of SE-GR, SE-GA and L-GA clusters in HN with hotspot scenario.

lacks on iterative improvements introduced in merge/split game when compared to SE-GA and L-GA algorithms.

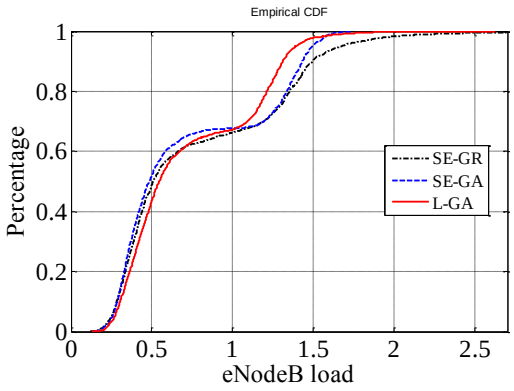


Figure 4.9: eNodeB load distribution in HN with hotspot.

In summary, we show that our novel L-GA clusters result in significantly less number of unsatisfied UEs by distributing load more evenly while keeping spectral efficiency at comparable levels in hotspot scenario. In non-hotspot scenario,

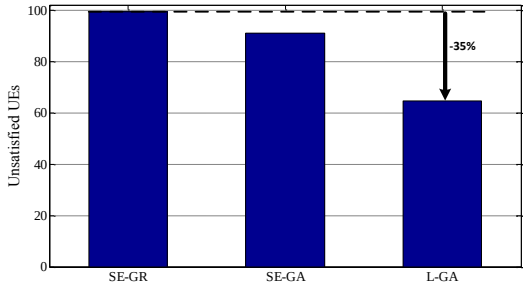


Figure 4.10: Unsatisfied UEs in HN with hotspot.

L-GA clustering performs as good as spectral efficiency based approaches (SE-GR/SE-GA). Overall, L-GA model performs well in all scenarios with/without hotspots providing a multi-objective clustering model which jointly optimizes cell load and spectral efficiency. It is also shown that L-GA provides an interesting dynamic cluster size metric, where average cluster size is increased in hotspot conditions and it is reduced to lower levels when hotspot disappears in the network, providing a control on the additional complexity associated with increased cluster size. Moreover, we also show that our spectral efficiency based game theoretic clustering model (SE-GA) clusters result in better spectral efficiency than greedy algorithm (SE-GR) in HN with hotspot scenario, due to cluster formation priority given to cells in hotspots first, and also the iterative process of merge/split algorithm outperforming greedy cluster formation.

4.5.2 Random Network (RN) Scenario

We evaluate our novel clustering solution further for a RN topology where SCs are randomly distributed within a circle of 0.4km radius following PPP distribution with density parameter $\lambda_{\mathcal{C}}$. All SCs within the circle are assumed to be connected to one MBS as described in Section 4.3. UEs are also randomly distributed following PPP distribution. To simulate the hotspot scenario, higher user density is assumed within an inner circle with 0.1km radius with user density $\lambda_{\mathcal{U}_{high}}$ and a low user density of $\lambda_{\mathcal{U}_{low}}$ is simulated in the outer ring. Outer ring is assumed to go beyond SC deployment radius to make sure users are distributed in the whole coverage area of all SCs. In the non-hotspot scenario, both inner and outer circle user density has been set to the same lower density. A snapshot of the simulated network topology with hotspot is illustrated in Figure 4.11. Simulations are run for 100 times for each RN scenario and various SC/user distribution is generated at each snapshot following PPP distribution with same SC/user density.

Figure 4.12a shows the average achieved spectral efficiency for all clustering types for hotspot and non-hotspot scenarios. Similar to HN scenario, average

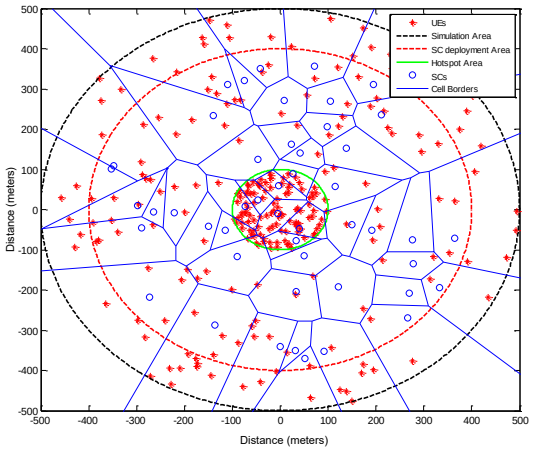


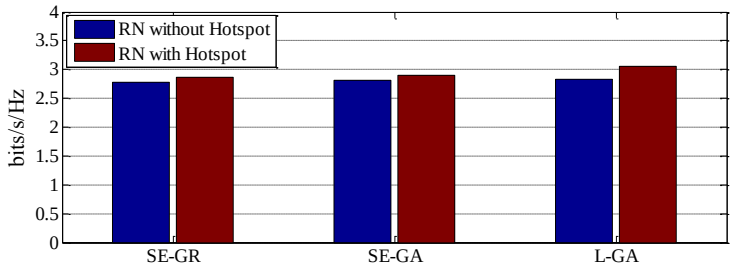
Figure 4.11: Random network simulation setup.

spectral efficiency is comparable on all 3 clustering types in evenly distributed traffic scenario where there is no hotspots, i.e. L-GA clusters perform as good as spectral efficiency based clusters when there is no overload. In hotspot scenario, spectral efficiency based coalitional game model (SE-GA) achieves a 1.57% better spectral efficiency than the greedy model (SE-GR), similar to HN results.

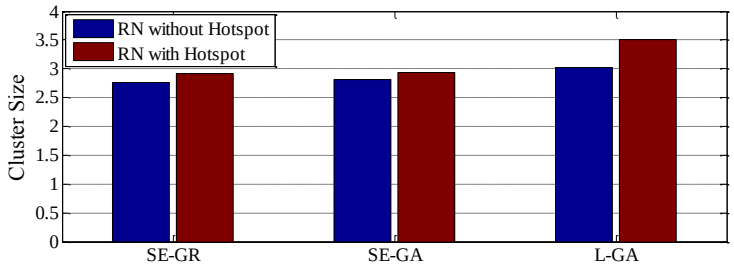
Our novel L-GA model achieves significant improvement in load balancing while keeping spectral efficiency at high levels in hotspot scenario, resulting in reduced number of unsatisfied users. Average achieved spectral efficiency is increased in L-GA model by 6.73% when compared to SE-GR as depicted in Figure 4.12a. Figure 4.13 shows the load distribution of the SCs where L-GA clustering achieves significantly better load distribution with reduced amount of SCs in high load range when compared to SE-GR and SE-GA clusters. Figure 4.14 shows the average total number of unsatisfied UEs for each clustering algorithm in RN with hotspot. L-GA algorithm is significantly more effective in distributing the load and reducing the number of unsatisfied users, resulting in 68.50% less unsatisfied users when compared to SE-GR clusters. 3.63% of the UEs are unsatisfied in L-GA when compared to 11.54% and 12.14% in SE-GR and SE-GA respectively.

Figure 4.12b depicts the average cluster size achieved for each clustering algorithm with and without hotspot scenario. Similar to HN results, average cluster size is increased in L-GA clusters significantly more than SE-GR and SE-GA clusters in hotspot scenario when compared to non-hotspot scenario due to additional payoff incentive in L-GA utility function to reduce load in hotspots. L-GA clusters manage to increase the cluster size in a self-organised way when there is high capacity requirement in hotspot scenario. Cluster size is dynamically reduced when the hotspot disappears and load is evenly distributed.

We further analyze merge/split iterations in RN scenario with and without



(a) Average spectral efficiency in RN with/without hotspot.



(b) Average cluster size in RN with/without hotspot.

Figure 4.12: Average spectral efficiency and cluster size in RN with/without hotspot scenarios.

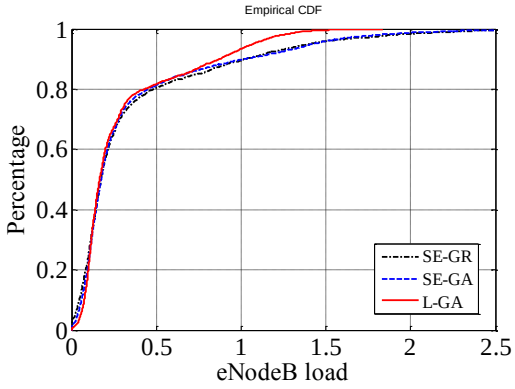


Figure 4.13: eNodeB load CDF in RN with hotspot.

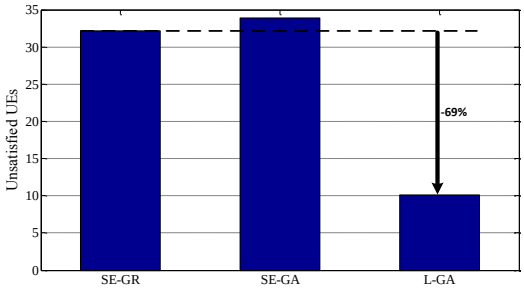


Figure 4.14: Unsatisfied UEs in RN with hotspot.

the hotspots. In Figure 4.15, total payoff of all SCs is shown for each merge/split operation until the final cluster is formed for L-GA clusters in RN scenario with/without hotspot. At each merge/split operation, utilitarian order is followed where merge/split operation is only allowed if the total system payoff is increased. In hotspot scenario, payoff is sharply increased in the first few merge/split operation where highly loaded cells are clustered, resulting in lower SC load and a higher payoff. Payoff increase is more gradual in non-hotpot scenario where merge/split operation gives an average payoff for each cell as they are almost equally loaded. Figure 4.16 depicts the changes in unsatisfied UEs and spectral efficiency for each merge/split iteration for L-GA algorithm in RN with hotspot scenario where our load-based utility function manages to improve both load and spectral efficiency at the same time for each merge/split operation for majority of merge/split operations. Marginal reduction in spectral efficiency is observed in later iterations for forming clusters to distribute load more evenly and therefore reduce unsatisfied UEs further with the expense of marginal spectral efficiency reduction.

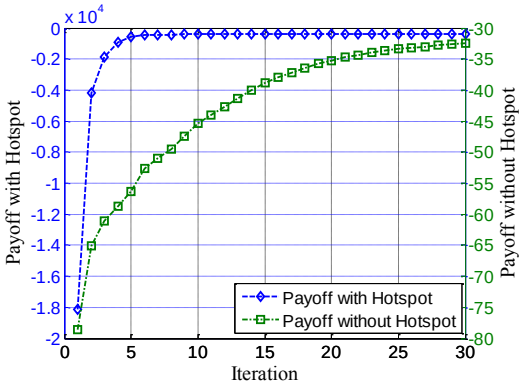


Figure 4.15: Payoff changes with merge/split iterations for L-GA algorithm in RN with/without hotspot scenario.

4.6 Conclusion

In this chapter, we present a novel, load aware network-centric clustering solution based on merge/split coalition game for CoMP deployment in future networks. We introduce merge/split coalitional game concepts and provide analysis on its stability and complexity. We show that our novel algorithm provides the unique partition with maximum utility when it is available, i.e. in the expected SC deployment scenario where SCs are deployed in local hotspot areas. In the case when this is not achievable, a more relaxed stability is always guaranteed where

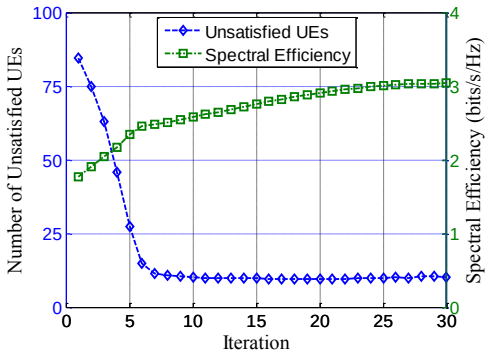


Figure 4.16: Unsatisfied UEs/spectral efficiency changes with split/merge iterations for L-GA algorithm in RN with hotspot scenario.

proposed algorithm converges to a final partition with no more merge/splits possible. Proposed solution is employed with two utility functions: Spectral efficiency based utility is designed to maximize spectral efficiency and load aware utility aims to jointly optimize spectral efficiency and load balancing objectives. It is shown that our spectral efficiency based clustering outperforms greedy algorithm providing better spectral efficiency in scenarios where users are unevenly distributed. Furthermore, we show that our load aware clustering model (L-GA) achieve significantly better load distribution while keeping spectral efficiency at high levels. Unsatisfied UEs are reduced by 68.5% in RN scenario with hotspots in L-GA algorithm when compared to greedy clustering model. Moreover, L-GA model provides a self-organised cluster size metric where CS is increased in hotspot scenarios to reduce high load with the expense of higher processing complexity and backhaul requirement, and it is reduced back down when hotspot disappears. In summary, our novel load-based game theoretical clustering algorithm (L-GA) is shown to be low-complexity, stable clustering solution combining both spectral efficiency and load balancing objectives into the same utility function which can dynamically adapt to both hotspot and non-hotspot scenarios. In the following chapter, we enhance this work further to include backhaul load awareness to CoMP clustering as an additional key objective. We will show that backhaul awareness in CoMP clustering model can increase system throughput significantly and reduce the number of unsatisfied users. We further improve our clustering model with a 2 stage coalitional game model where clusters of SCs are optimized in one game and the groups of users associated to each SC cluster is optimized in the second game to form a multi-objective dynamic clustering model to jointly optimize RAN load, backhaul load and spectral efficiency objectives.

Chapter 5

Multi-Objective CoMP Clustering

5.1 Introduction

In this chapter, we further enhance our work presented in Chapter 4 to include backhaul awareness into CoMP clustering. Realization of CoMP heavily depends on backhaul availability especially for JT-CoMP due to high backhaul requirement to make user-data available on all BSs in the CoMP set. In the case when some SCs in the system are backhaul limited, CoMP clustering algorithm need to take this into account for optimum CoMP cluster design. In this context, we develop a multi-objective, load aware, dynamic clustering model for MU-JT CoMP to jointly optimize spectral efficiency, RAN load and backhaul load. We further improve the RAN-load aware coalitional game model from Chapter 4 and formulate the improved model as two coalitional sub-games for SC and UE clustering respectively. Merge/split/transfer actions for each sub-game are defined, complexity and stability analysis are provided. Extensive simulation results show that our model successfully promotes the SCs with higher backhaul availability in CoMP clusters and SCs with limited backhaul are also included in CoMP sets when the additional spectral efficiency improvement is high. We show that our RAN and backhaul-load aware model provides comparably good spectral efficiency in light load when compared to a greedy model, and significantly better load balancing with reduced unsatisfied users and increased throughput in high load scenario. On average 49% increase in overall system throughout is observed in our simulations when compared to greedy model. Part of this work is submitted for publication, currently under review.

The rest of the chapter is organised as follows. Related work in literature is discussed in Section 5.2. In Section 5.3, we introduce the system model. Key

CoMP performance metrics are defined in Section 5.4. In Section 5.5, we describe our clustering model as SC clustering and UE transfer sub-games and discuss its stability and complexity. Simulation results with insights are presented in Section 5.6 and finally we summarize the findings and conclude the chapter in Section 5.7.

5.2 Related Work

Backhaul capacity and latency are one of the biggest challenges for the realization of CoMP in future networks [36, 85]. Backhaul network limitations and imperfect CSI issues create a significant impact on achievable spectral efficiency with CoMP [156]. There has been a number of CoMP clustering studies in the literature [35, 71, 115, 124, 136, 180, 181] which takes backhaul availability as the main objective as detailed in Chapter 2, Section 2.6.2. For example, required backhaul is taken as one of the key objectives in [136] where SFR and CoMP are employed together to improve cell edge user performance. Limited fronthaul availability is studied in [71] for C-RAN architecture where user-centric clusters of RRHs are optimized to minimize the total transmission power while maintaining user's QoS. More recently, limited backhaul capacity and per-BS power constraints are taken into account to optimize user-centric clusters and design transmit precoding for maximizing the sum rate in [124]. In both studies in [71,124], user-centric clustering models are presented but higher precoding/scheduling/synchronization complexity of user-centric clusters are not resolved. Most works in literature deal with backhaul limitation as a single objective, lacking on a multi-objective clustering solution to optimise all key objectives such as spectral efficiency, load balancing etc.

On the other hand, new emerging technologies like MEC and popular data caching at the BS is a promising concept discussed in literature to reduce backhaul requirement for CoMP [48]. Caching data on the MEC servers at the BSs will eliminate the need for popular data being transmitted from core network over the backhaul, reducing the capacity and latency requirements for backhaul required for CoMP during high load traffic. A number of studies utilize data caching at the BS to reduce backhaul requirement for CoMP in [49, 56, 77, 105, 174]. Similar to above, in these works, backhaul limitation is studied in isolation for CoMP clustering, without considering other network metrics i.e. spectral efficiency, radio access load etc.

RAN load is another key dependency which needs to be taken into account for CoMP clustering. CoMP is likely to be deployed in interference-limited, highly

dense deployment scenarios where hotspot areas will form at certain times. CoMP clusters need to dynamically adjust in order to balance the load and shift traffic from highly loaded BSs to lightly-loaded BSs. In Chapter 3, we proposed a user-centric clustering algorithm where RAN load is taken into consideration for load aware, user-centric clusters for the first time in literature. Following our work, other few load aware, user-centric clusters are studied [43, 106, 109]. However, user-centric clusters are not scalable for large networks due to increased complexity. To avoid the complexity of user-centric clusters, in Chapter 4, we proposed a novel, low-complexity, merge-split coalitional game model to form RAN-load aware network-centric clusters.

Backhaul limitation, radio access load and spectral efficiency objectives have been studied for CoMP clustering but each objective is studied in isolation. There is no CoMP clustering solution in literature to jointly optimize and analyze the trade-off between spectral efficiency and backhaul/RAN load. In this chapter, we further improve our clustering model presented in Chapter 4 and propose a novel RAN and backhaul-load aware multi-objective clustering method to optimize overall RAN/backhaul load and spectral efficiency jointly for SCs in HetNet deployment scenario. We design two coalitional sub-games, 1-an SC clustering sub-game to form RAN/backhaul-load aware SC clusters by merge/split/transfer actions, 2-a novel user transfer sub-game to move users between SC clusters to improve load balancing further. Stability and complexity analysis are provided and extensive simulation results for multiple scenarios are presented to show the performance of the proposed method under different backhaul availability conditions. Results are benchmarked against an improved version of our previous work on RAN-load aware clustering model presented in Chapter 4 and a greedy algorithm in [31].

5.3 System Model

We consider a similar HetNet scenario we assumed in Chapter 4, where a set of SCs ($\mathcal{C}$) are distributed within the coverage area of one MBS. We assume designated frequency spectrum for SC and MBS layer hence no interference is expected between the layers. Network of $\mathcal{C} = \{SC_1, \ldots, SC_n\}$ SCs are grouped into clusters $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_s\}$ for CoMP operation. Each user is assigned an SC cluster $\mathcal{C}_i$, forming clusters of users $\mathcal{U} = \{\mathcal{U}_1, \ldots, \mathcal{U}_s\}$ i.e. user group $\mathcal{U}_i$ is assigned to SC cluster $\mathcal{C}_i$. Assume UE_k is assigned the SC cluster $\mathcal{C}_i$ where $\mathcal{C}_i = \{SC_{i1}, SC_{i2}, \ldots, SC_{iz}\}$. Based on average received reference signal level, UE_k is assigned a user-centric cluster $\mathcal{C}_i^k = \{SC_{i1}^k, SC_{i2}^k, \ldots, SC_{it}^k\}$ within $\mathcal{C}_i$ where

$\mathcal{C}_i^k \subseteq \mathcal{C}_i$. Assuming the best serving cell within $\mathcal{C}_i$ for UE_k is SC_{im}, $\mathcal{C}_i^k$ includes all $SC_{ij} \in \mathcal{C}_i$ where $p_{kj}/p_{km} > P_\Delta$ and $p_{kj} > P_{min}$ where p_{km} and p_{kj} are the average signal power values received at UE_k from $SC_m \subseteq \mathcal{C}_i$ and $SC_j \subseteq \mathcal{C}_i$ respectively.

To evaluate the impact of limited backhaul, we assume two backhaul technologies connecting SCs to MBS: Fiber and VDSL2. Latency and capacity for fiber is considered to be ideal where DL capacity > 10Gbps and latency < 1msec, however DL capacity for VDSL2 is assumed to be limited to 100Mbps and average latency is considered as 3msec [85]. Both backhaul technologies are considered to be robust so outage probability is considered as zero. For SCs where VDSL2 is deployed, backhaul capacity limitation is considered alongside with radio capacity to derive overall cell load values. Backhaul throughput is derived from radio throughput and an additional overhead of 30% is added to account for additional control plane traffic [23,84]. Higher latency in VDSL2 cause CSI imperfection for CoMP, resulting in reduced spectral efficiency, and hence lower throughput. Impact of various latency values in throughput is analyzed in [39] for DL JT-CoMP where an average 15% throughput loss is observed for 3msec latency. Based on [39], we consider 15% loss in spectral efficiency when compared to fiber (very low latency) for UE_k when $\mathcal{C}_i^k$ contains at least one SC with VDSL2 backhaul link to the MBS. The connection between MBS and the core network is assumed to be ideal i.e. fiber, hence no capacity and latency constraints are considered from MBS to the core network.

We assume MU-JT CoMP where multiple users within the same cluster are scheduled to the same PRB and receive user-data from each SC within the cluster, i.e. user data for UE_k is made available at each SC within $\mathcal{C}_i^k$.

In the ideal backhaul scenerio, assume a group of UEs ($\mathcal{U}_i^k$) including UE_k is assigned a user-centric cluster $\mathcal{C}_i^k$ and scheduled in the same PRB at each SC in $\mathcal{C}_i^k$. Assuming one antenna for both UE and SCs for simplicity, a virtual MIMO system is formed with $|\mathcal{C}_i^k| = T$ transmitters and and $|\mathcal{U}_i^k| = R$ receivers.

For each UE in $\mathcal{U}_i^k$, received signal can be expressed as:

$$\mathbf{y} = \mathbf{HWx} + \mathbf{n}, \mathbf{H} \in \mathbb{C}^{R \times T}, \mathbf{W} \in \mathbb{C}^{T \times R} \tag{5.1}$$

where channel matrix $\mathbf{H} = \Big[\mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_R\Big]^T$

and channel vector at UE_k is:

$$\mathbf{h}_k = \Big[h_{k1} h_{k2} \dots h_{kT}\Big] \tag{5.2}$$

Precoding matrix $\mathbf{W} = \Big[\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_R\Big]$

and beamforming vector for UE_k is:

$$\mathbf{w}_k = \left[w_{1k} w_{2k} \dots w_{Tk}\right]^T \tag{5.3}$$

Received signal y_k at UE_k can be expressed as:

$$y_k = \mathbf{h}_k^{\mathcal{C}_i^k} \mathbf{w}_k^{\mathcal{C}_i^k} x_k + \sum_{i \in \mathcal{U}_i^k / k} \mathbf{h}_k^{\mathcal{C}_i^k} \mathbf{w}_i^{\mathcal{C}_i^k} x_i + \sum_{j \in \mathcal{U} / \mathcal{U}_i^k} \mathbf{h}_k^{\mathcal{C} / \mathcal{C}_i^k} \mathbf{w}_j x_j + n_k \tag{5.4}$$

In (5.4), the first term represents the desired signal from each of the SCs within $\mathcal{C}_i^k$, second term represents the interference from within the cluster $\mathcal{C}_i^k$, followed by interference from outside of $\mathcal{C}_i^k$ and the final term is AGWN.

SINR at UE_k is:

$$SINR_k = \frac{\left|\mathbf{h}_k^{\mathcal{C}_i^k} \mathbf{w}_k^{\mathcal{C}_i^k} x_k\right|^2}{\sum_{i \in \mathcal{U}_i^k / k} \left|\mathbf{h}_k^{\mathcal{C}_i^k} \mathbf{w}_i^{\mathcal{C}_i^k} x_i\right|^2 + \sum_{j \in \mathcal{U} / \mathcal{U}_i^k} \left|\mathbf{h}_k^{\mathcal{C} / \mathcal{C}_i^k} \mathbf{w}_j x_j\right|^2 + |n_k|^2} \tag{5.5}$$

Assuming fiber backhaul for each SC within $\mathcal{C}_i^k$ to the MBS, intra-cluster interference would be negligible with highly accurate CSI knowledge and very low latency at the MBS. In the case when $\mathcal{C}_i^k$ contains any SCs with VDSL2 backhaul, higher backhaul latency (3msec) causes imperfect CSI and intra-cluster interference does not get canceled completely causing degradation in SINR. This degradation is taken into account as 15% spectral efficiency degradation based on the findings in [39].

Let the total transmit power P_{Tx} for each SC be the same and the power for each PRB be equal, then (5.5) can be further simplified to:

$$\hat{SINR}_k = \frac{P_{Tx} \sum_{i \in \mathcal{C}_i^k} |h_{ki}|^2}{P_{Tx} \sum_{j \in \mathcal{C} / \mathcal{C}_i^k} |h_{kj}|^2 + N_0 B_{tot}} \tag{5.6}$$

where N_0 is the noise spectral density, B_{tot} is the total system bandwidth and channel coefficient h_{ki} is made up of 2 terms, static distance based path loss component with shadow fading and fast fading complex coefficients $h_{ki} = g_{ki} f_{ki}$.

We propose that clustering decisions are made based on average SINR to respond to spatio-temporal changes in the network and user profiles (in seconds. minutes), but not responding to fast fading changes (in milliseconds). This provides additional resilience for incorrect clustering decisions due to imperfect CSI knowledge and prevents additional signalling overhead required for too frequent

re-clustering decisions [133]. For average SINR, term h_{ki} in (5.6) can be simplified to the distance based path-loss and shadow fading component only i.e. $\hat{h_{ki}} = g_{ki}$ where fast fading component f_{ki} is averaged out over time.

5.4 CoMP Performance Metrics

In this section, we define main CoMP performance metrics and utilize these in our coalitional game model. We recall some of the metrics defined from Chapter 4 and add on backhaul element to them in MU JT-CoMP scenario. Assume UE_k is assigned a network-centric cluster $\mathcal{C}_i$ and user-centric cluster $\mathcal{C}_i^k$ where $\mathcal{C}_i^k \subseteq \mathcal{C}_i$ and let d_k be the GBR requirement for UE_k. The required PRB for UE_k in no CoMP scenario would be $r_k = d_k/(y_k B_{PRB})$ where B_{PRB} is the user-data bandwidth in a single PRB and $y_k = log_2(1 + SI\hat{N}R_k)$. In MU-JT CoMP, a number of UEs ($\mathcal{U}_i^k$) are scheduled on the same PRB at each cell in $\mathcal{C}_i^k$ so we define an estimated dedicated PRB count for UE_k at each SC in $\mathcal{C}_i^k$ as $\hat{r}_k = r_k/n_k$, assuming $|\mathcal{C}_i^k| = |\mathcal{U}_i^k| = n_k$ [31].

5.4.1 RAN and Backhaul Load

The main aim for CoMP is to improve spectral efficiency and hence provide the required throughput with less radio resources i.e. reduce RAN load for the cell. For MU JT-CoMP, increasing CoMP cluster size will improve spectral efficiency further as there is more inter-cell interference cancellation within a bigger CoMP cluster. However, as CoMP cluster size increase, additional pilot channels are required for CSI estimation which then reduce the bandwidth for user data at each PRB. As the available bandwidth for user data reduce, RAN load for the cell will increase. So RAN load is one of the key metrics to measure CoMP performance where it implicitly reflect on spectral efficiency improvement and also the CoMP pilot overhead. We utilize RAN load metric for SC_m driven in Chapter 4 for MU JT-CoMP scenario as follows:

$$l_{im}^{RAN} = \frac{\sum_{k \in \mathcal{U}_{im}} \hat{r}_k}{R_{tot}} \tag{5.7}$$

where $\mathcal{U}_{im}$ is the associated active UEs in SC_m and R_{tot} is the total number of PRBs for each SC, assuming all SCs have same total bandwidth.

A more realistic load figure should also consider backhaul load alongside with RAN load. In a network where some SCs are connected to the MBS with non-ideal backhaul solutions with limited bandwidth and latency, limiting factor for the overall load could be the backhaul load rather than RAN load depending

radio frequency bandwidth, SINR and backhaul type. In MU JT-CoMP scenario, backhaul load increases as the cluster size increase due to user data for all users within $\mathcal{U}_i^k$ will need to be available at all SCs within $\mathcal{C}_i^k$. Moreover, additional latency due to non-ideal backhaul will introduce delay in CSI estimation for precoding and hence reduce spectral efficiency gain and increase RAN load. In summary, along side with RAN load, backhaul load is another key metric which needs to be considered in CoMP clustering.

To define backhaul load $\hat{l}_{im}^{BH}$, firstly, we define RAN throughput demand on SC_m in $\mathcal{C}_i$ as $d_{im}^{RAN} = \sum_{k \in \mathcal{U}_{im}} d_k$. Backhaul throughput demand d_{im}^{BH} is then computed with an average overhead factor of 1.3 to account for additional traffic on backhaul for X2 user/control plane and transport and security overheads [23, 84] i.e. $d_{im}^{BH} = d_{im}^{RAN} \times 1.3$.

Once d_{im}^{BH} is known, $\hat{l}_{im}^{BH}$ can then be defined as:

$$\hat{l}_{im}^{BH} = \frac{d_{im}^{BH}}{\hat{f}_{im}^{BH}} \tag{5.8}$$

where $\hat{f}_{im}^{BH}$ is the backhaul capacity. As discussed in Section 5.3, we consider two backhaul technologies: 1-) Fiber with $>$ 10Gbps capacity and negligible latency and 2-) VDSL2 with 100Mpbs capacity and 3ms latency. When backhaul gets congested i.e. $d_{im}^{BH} > \hat{f}_{im}^{BH}$, then the effective capacity $\hat{f}_{im}^{BH}$ goes down further due to re-transmissions [84]. In the case of VDSL2 link congestion, we consider 10% re-transmission rate i.e. assume the effective capacity of the VDSL link as $\hat{f}_{im}^{BH} = 90Mbps$ in-line with the same assumptions made in [84].

A more realistic SC load definition need to consider both backhaul load and RAN load and pick the highest of the two as the overall load, i.e. the overall load can be defined as $\hat{l}_{im} = \max(\hat{l}_{im}^{RAN}, \hat{l}_{im}^{BH})$.

5.4.2 Cell Throughput

In MU JT-CoMP, user data for UE_k is transmitted from all of the SCs in $\mathcal{C}_i^k$ and hence total RAN throughput demand on SC_m in $\mathcal{C}_i$ i.e. d_{im}^{RAN} does not reflect the "dedicated throughput" value for SC_m as same RAN throughput demand for user UE_k is accounted for in all SCs in $\mathcal{C}_i^k$. An estimated dedicated RAN throughput demand for SC_m in $\mathcal{C}_i$ is defined as $\hat{d}_{im}^{RAN} = \sum_{k \in \mathcal{U}_{im}} d_k / n_k$ where $|\mathcal{C}_i^k| = n_k$.

Based on estimated dedicated RAN throughput demand $\hat{d}_{im}^{RAN}$ for SC_m, estimated dedicated cell throughput $\hat{t}_{im}$ for each SC_m in $\mathcal{C}_i$ can then be defined

as:

$$\hat{t}_{im} = \begin{cases} \hat{d}_{im}^{RAN} & \hat{l}_{im} < 1 \\ \frac{\hat{d}_{im}^{RAN}}{\hat{l}_{im}} & \hat{l}_{im} \geq 1 \end{cases} \tag{5.9}$$

5.4.3 Unsatisfied Users

To quantify the impact of high load on users, we utilize the unsatisfied users metric we presented in Chapter 4 for MU JT-CoMP scenario. Similar metrics are also used in [32, 164].

Unsatisfied users are defined as follows:

$$\hat{z}_{im} = \max\left(0, \hat{u}_{im}\left(1 - \frac{1}{\hat{l}_{im}}\right)\right) \tag{5.10}$$

where $\hat{u}_{im} = \sum_{k \in \mathcal{U}_{im}} 1/n_k$ is defined as the estimated dedicated user count at SC_m. Estimated dedicated user count at each cell is driven from the total users connected at each cell $\mathcal{U}_{im}$ to account for the fact that users are connected to multiple SCs in MU JT-CoMP.

5.4.4 Pilot Overhead

To account for the additional pilot channel overhead, we utilize the pilot overhead estimation for multi-antenna channels in [86] as follows:

$$\begin{aligned} \alpha = & \sqrt{(1+SNR)\frac{\dot{C}(SNR)}{C(SNR)} 2 n_T f_D} \\ & - \left((1+SNR)\frac{\ddot{C}(SNR)}{\dot{C}(SNR)} + 2 + \frac{1}{2SNR}\int_{-1}^{+1} \frac{d\xi}{\tilde{S}_H(\xi)}\right) n_T f_D \\ & + O(f_D^{3/2}) \end{aligned} \tag{5.11}$$

where

$C(SNR) = \mathbb{E}[\log_2(1 + SNR|H|^2)]$
$\dot{C}(SNR) = \frac{1}{SNR}\left(\log_2 e - \frac{C(SNR)}{SNR}\right)$
$\ddot{C}(SNR) = \frac{1}{SNR^2}\left[\log_2 e + \dot{C}(SNR - 2\frac{C(SNR)}{SNR}\right]$
$\tilde{S}_H(\xi)$ is the doppler spectrum of the wireless channel.
f_D is the normalised doppler frequency
n_T is the number of transmit antennas
and α is percentage pilot overhead bandwidth requirement.

Similar to assumptions made in Chapter 4, we assume EPA-A wireless channel from 3GPP [3] for Clarke-Jakes spectrum where $f_D = 0.000357$ and the term $\int_{-1}^{+1} \frac{d\xi}{\tilde{S}_H(\xi)}$ simplifies to $\pi^2/2$. We assume one antenna per SC i.e. $n_T = |\mathcal{C}_i|$ and

SNR=10dB for pilot overhead estimation. Further details about the required pilot overhead bandwidth for various wireless channels from 3GPP [3] against cluster size can be found in Chapter 4 Section 4.3.2.

As cluster size $\mathcal{C}_i$ increases, pilot overhead increases and hence the bandwidth for user data is reduced on each PRB. PRB bandwidth available for user data can be defined as $b_{PRB} = B_{PRB}(1 - \alpha)$.

5.5 Coalition Game for Multi-Objective Clustering

Applications of coalitional game theory have recently become popular in cooperative wireless networks for self-organizing techniques to form CoMP clusters [75,148]. A merge/split coalition formation game is employed in forming user clusters for UL TDMA cooperative network scenario in [147]. Similar merge/split game is utilized in forming BS clusters in DL CoMP for CRAN scenario in [183]. A transfer game is employed alongside with a collage admission game for UL user association problem in HetNet scenario in [149]. In our previous work in Chapter 4, we presented a merge/split game model to form load aware clusters where both spectral efficiency and RAN load are jointly optimized [31]. In this chapter, we formulate two coalitional sub-games to jointly optimize overall load (backhaul and RAN) and spectral efficiency. First, we extend our coalitional game model from Chapter 4 to combine merge/split and transfer games to into a single SC clustering sub-game to form clusters of SCs. Secondly, we drive an additional user transfer sub-game for user groups to transfer users between SC clusters to further distribute the load between the SC clusters. In this section, we formulate and discuss the properties for each sub-game and analyze the overall stability and complexity of the proposed solution.

5.5.1 Coalitional Game Model for SC Clustering Sub-game

In this subsection, we formulate the SC clustering sub-game where SC clusters are formed and dynamically updated based on spatio-temporal changes in the network and/or user profiles. Let $\mathcal{C} = \{SC_1, \ldots, SC_n\}$ be the players of our coalition game i.e. small cells in the network and assume that they are grouped into clusters $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_s\}$. A coalition is defined as the groups of players in the same cluster i.e. $\mathcal{C}_i = \{SC_{i1}, SC_{i2}...SC_{iz}\}$ and a collection is defined as the set of coalitions $\mathcal{H} = \{\mathcal{H}_1, \mathcal{H}_2, ..., \mathcal{H}_b\}$. A collection is called a partition when $\forall i \neq j, \mathcal{H}_i \cap \mathcal{H}_j = \varnothing$ and $\cup_{i=1}^{b} \mathcal{H}_i = \mathcal{C}$. The payoff for coalition $\mathcal{C}_i$ in

partition $\mathcal{C}$ is defined by the utility function $v(\mathcal{C}_i, \mathcal{C})$ and overall SC clustering sub-game is defined by (C, v) pair. The utility function reflects the overall gain for cooperation including multiple objectives of CoMP deployment (e.g. like spectral efficiency and backhaul/RAN load balancing) and also the various cost factors of cooperation (e.g. like additional pilot requirement, signal processing complexity). An accurate utility function is key for better CoMP clustering to maximize the benefits expected from CoMP.

We utilize two utility functions we presented in Chapter 4. Firstly, we employ a load aware utility function which aims to shift load from highly loaded cells to lightly loaded cells and implicitly improve spectral efficiency as follows:

$$v_1(SC_m, \mathcal{C}_i) = \begin{cases} \frac{-(\hat{l}_{im})}{1-c(|\mathcal{C}_i|)}\hat{u}_{im} & \hat{l}_{im} < 1 \\ \frac{-(\hat{l}_{im})^3}{1-c(|\mathcal{C}_i|)}\hat{u}_{im} & \hat{l}_{im} \geq 1 \end{cases} \tag{5.12}$$

where $c(|\mathcal{C}_i|) = \frac{1}{1+e^{-(|\mathcal{C}_i|-C^n_{max})}}$. $c(|\mathcal{C}_i|)$ is defined as the complexity function which represents the additional overhead for CoMP such as precoding processing complexity, synchronization issues and additional backhaul capacity. As the additional overheads for CoMP increase when cluster size increase, the complexity function is designed to introduce a soft limit to maximum cluster size C^n_{max} based on the requirements of the network for the right trade-off between additional spectral efficiency/load gain and CoMP overheads.

An spectral efficiency based utility is also employed in our work within a greedy algorithm for comparison. This utility does not consider cell load but aims to maximize spectral efficiency only. Spectral efficiency based utility function is defined as follows:

$$v_2(SC_m, \mathcal{C}_i) = \sum\nolimits_{k \in \mathcal{U}^{best}_{im}} y_k(1 - c(|\mathcal{C}_i|)) \tag{5.13}$$

where $\mathcal{U}^{best}_{im}$ is the list of users where SC_m is the best serving cell based on average received signal power within $\mathcal{C}_i$, i.e. a subset of the associated users $\mathcal{U}_{im}$ at the SC_m and y_k is the spectral efficiency achieved at UE_k i.e. $y_k = log_2(1 + S\hat{I}NR_k)$.

To compare the utility of two different collections $\mathcal{H} = \{\mathcal{H}_1, \mathcal{H}_2, ..., \mathcal{H}_b\}$ and $G = \{G_1, G_2, ..., G_z\}$ of the same subset of players, we employ the utilitarian comparison order where collection $\mathcal{H}$ is preferable to collection G if the overall utility of the collection is higher, even if the individual players may be worse off, i.e. $\mathcal{H} \triangleright G$ if $\sum_{i=1}^{b} v(\mathcal{H}_i) > \sum_{i=1}^{z} v(\mathcal{G}_i)$ [31, 148].

SC coalitions are formed and adapted into changing network/user profile conditions by 3 different clustering actions:

- **Merge:** Players (SCs) in any two or more coalitions $\{\mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_z\}$ prefer to merge into one coalition $\mathcal{F} = \cup_{i=1}^{z} \mathcal{G}_i$ i.e. $\cup_{i=1}^{z} \mathcal{G}_i \triangleright \{\mathcal{G}_1, \mathcal{G}_2, ..., \mathcal{G}_z\}$, if $v(\mathcal{F}) > (\sum_{i=1}^{z} v(\mathcal{G}_i))$ following the utilitarian order.

- **Split:** Players (SCs) prefer to split from any coalition $\mathcal{C}_i$ into smaller coalitions $\{\mathcal{C}_{i1}, \mathcal{C}_{i2}, ..., \mathcal{C}_{iy}\}$ where $\mathcal{C}_i = \cup_{j=1}^{y} \mathcal{C}_{ij}$ i.e. $\{\mathcal{C}_{i1}, \mathcal{C}_{i2}, ..., \mathcal{C}_{iy}\} \triangleright \mathcal{C}_i$ if $(\sum_{j=1}^{y} v(\mathcal{C}_{ij}) > v(\mathcal{C}_i)$ following utilitarian order.

- **Transfer:** Any player in $\mathcal{C}_i$, i.e. $SC_{ix} \subseteq \mathcal{C}_i$ prefer to transfer from coalition $\mathcal{C}_i$ to $\mathcal{C}_j$ i.e. $\{\mathcal{C}_i \backslash SC_{ix}, \mathcal{C}_j \cup SC_{ix}\} \triangleright \{\mathcal{C}_i, \mathcal{C}_j\}$ if $v(\{\mathcal{C}_i \backslash SC_{ix}) + v(\mathcal{C}_j \cup SC_{ix}) > \{v(\mathcal{C}_i) + v(\mathcal{C}_j)\}$.

Assume $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_s\}$ be any partition of $\mathcal{C}$, i.e. the current network clustering structure. We propose to start with split operation, followed by merge operation and then a transfer operation afterwards. Split/merge/transfer operations are repeated until there is no more re-clustering action possible to improve overall utility.

Split operation checks possible split options for $\forall \mathcal{C}_i$ in $\mathcal{C}$, and implements the split operation when it finds a suitable split option based on utilitarian order i.e. $(\sum_{j=1}^{y} v(\mathcal{C}_{ij}) > v(\mathcal{C}_i)$. Split operation is repeated iteratively until there is no further split is possible as detailed in Algorithm 6. A new partition $\mathcal{H}$ is formed after the split operation. $\mathcal{H}$ is then subject to merge operation as detailed in Algorithm 7. Merge operation starts with coalition $\mathcal{H}_i$ with the maximum absolute payoff value and looks for merge options to its neighbour coalitions. We avoid exhaustive search of possible merge with every other coalition in the network which reduces the algorithm complexity significantly. Merge operation is implemented for $(\mathcal{H}_i, \mathcal{H}_j)$ coalition pair where $\mathcal{H}_j$ is the neighbour coalition for $\mathcal{H}_i$ with maximum additional payoff in the case of a possible merge operation. We adapt the neighbour coalition concept from Chapter 4 where we define them based on the reported average received signal power from the users. For any user UE_k within the serving area of $SC_m \subseteq \mathcal{H}_m$, a neighbour rank value is incremented for $\{\mathcal{H}_m, \mathcal{H}_j\}$ pair if $p_{kj}/p_{km} > P_{\Delta}^{nei}$ and $p_{kj} > P_{min}^{nei}$ where p_{km} and p_{kj} are the average signal power values received from UE_k for $SC_m \subseteq \mathcal{H}_m$ and $SC_j \subseteq \mathcal{H}_j$ respectively. Merge operation continues for $\forall \mathcal{H}_i$ in $\mathcal{H}$ and is repeated for the whole partition until no other merge is possible. Once the merge operation is finished, transfer operation starts with the new partition $\mathcal{G}$ formed after merge operation. For $\forall \mathcal{G}_i \in \mathcal{G}$, each $SC_{ix} \in \mathcal{G}_i$ are checked for a possible transfer to one of the neighbour coalition $\mathcal{G}_j$ i.e. $T(SC_{ix}, \mathcal{G}_i, \mathcal{G}_j)$. Within each coalition $\mathcal{G}_i$, all possible transfer operations are ranked and transfer operation $T(SC_{ix}, \mathcal{G}_i, \mathcal{G}_j)$ is implemented for the one with the maximum additional payoff. Transfer operation

continues for all $\forall \mathcal{G}_i \in \mathcal{G}$ and is repeated for the newly formed partition until there is no further transfer possible with additional payoff, as detailed in Algorithm 8.

Split, merge and transfer operations are then repeated until there is no further SC coalition actions possible.

Algorithm 6 Split Operation

For any given network clustering state $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_s\}$, $\forall \mathcal{C}_i \in \mathcal{C}$, set $\mathcal{C}_i$.splitpossible=1
Split-ongoing=1
while Split-ongoing **do**
  Split-ongoing=0
  for all $\mathcal{C}_i$ where ($\mathcal{C}_i$.split-possible=1 and $|\mathcal{C}_i| > 1$) **do**
    Update $\mathcal{C}_i$.Split-options
    $\mathcal{C}_i$.split-possible=0
    for all $\mathcal{C}_i$.Split-Options **do**
      if Any split option is possible i.e. $(\sum_{j=1}^{y} v(\mathcal{C}_{ij}) > v(\mathcal{C}_i)$ **then**
        Split($\mathcal{C}_i$ to $\{\mathcal{C}_{i1}, \mathcal{C}_{i2}, ..., \mathcal{C}_{iy}\}$
        Split-ongoing=1
        $\forall \mathcal{C}_{ij}$, set $\mathcal{C}_{ij}$.split-possible=1
        Break for-loop and continue with next $\mathcal{C}_i$
      end if
    end for
  end for
end while

5.5.2 Coalitional Game Model for User Transfers Sub-game

Assume $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_s\}$ be the SC partition of $\mathcal{C}$ resulting from the SC clustering sub-game $(\mathcal{C}, v)$. List of users $\mathcal{U} = \{UE_1, \ldots, UE_q\}$ can be expressed as coalitions of users assigned to each SC cluster i.e. $\mathcal{U} = \{\mathcal{U}_1, \ldots, \mathcal{U}_s\}$ where users in $\mathcal{U}_i$ are assigned to SC coalition $\mathcal{C}_i$. We formulate a user transfer sub-game $(\mathcal{U}, v)$ to transfer users between the user coalitions to further distribute the load between the SC clusters.

Transfer operation introduced in SC clustering sub-game in Section 5.5.1 is deployed for the user transfer sub-game i.e. any user $UE_{ix} \subseteq \mathcal{U}_i$ prefer to transfer from coalition $\mathcal{U}_i$ to $\mathcal{U}_j$ i.e. $\{\mathcal{U}_i \backslash UE_{ix}, \mathcal{U}_j \cup UE_{ix}\} \triangleright \{\mathcal{U}_i, \mathcal{U}_j\}$ if $v(\{\mathcal{U}_i \backslash UE_{ix}) + v(\mathcal{U}_j \cup UE_{ix}) > \{v(U_i) + v(U_j)\}$ following utilitarian order. We utilize the load aware utility in (5.12) for user transfer sub-game and transfer users to re-assign to another cluster if the overall utility is improved with this transfer operation.

Neighbour concept introduced in the SC clustering sub-game is employed in user transfer sub-game too at user level, so that each user only looks for the

Algorithm 7 Merge Operation

For any given network clustering state $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_s\}$, $\forall \mathcal{C}_i \in \mathcal{C}$, set $\mathcal{C}_i$.clustered=0
Merge-ongoing=1
while Merge-ongoing **do**
  Merge-ongoing=0
  Sort $\forall \mathcal{C}_i \in \mathcal{C}$ based on $v(\mathcal{C}_i)$ in descending order
  for all $\mathcal{C}_i$ where $\mathcal{C}_i$.clustered=0 **do**
    Update $\mathcal{C}_i$.nei
    for all $\mathcal{C}_j$ in $\mathcal{C}_i$.nei where $\mathcal{C}_j$.clustered=0 **do**
      Update payoff gain for possible merge$(\mathcal{C}_i, \mathcal{C}_j)$ i.e. $\delta_{v_{ij}} = v(\mathcal{C}_i \cup \mathcal{C}_j) - \{v(\mathcal{C}_i) + v(\mathcal{C}_j)\}$
    end for
    Find $\mathcal{C}_m \in \mathcal{C}_i$.nei where $\delta_{v_{im}} = \max_{\mathcal{C}_j \in \mathcal{C}_i.nei}(\delta_{v_{ij}})$ and $\delta_{v_{im}} > 0$
    while $\mathcal{C}_m$ exist **do**
      Merge$(\mathcal{C}_i, \mathcal{C}_m)$
      $\mathcal{C}_m$.clustered=1
      Update $\mathcal{C}_i$.nei
      for all $\mathcal{C}_j$ in $\mathcal{C}_i$.nei where $\mathcal{C}_j$.clustered=0 **do**
        Update payoff gain for possible merge$(\mathcal{C}_i, \mathcal{C}_j)$ i.e. $\delta_{v_{ij}} = v(\mathcal{C}_i \cup \mathcal{C}_j) - \{v(\mathcal{C}_i) + v(\mathcal{C}_j)\}$
      end for
      Find $\mathcal{C}_m \in \mathcal{C}_i$.nei where $\delta_{v_{im}} = \max_{\mathcal{C}_j \in \mathcal{C}_i.nei}(\delta_{v_{ij}})$ and $\delta_{v_{im}} > 0$
    end while
    $\mathcal{C}_i$.clustered=1
    if Any merge operation with $\mathcal{C}_i$ **then**
      Break for-loop and continue with while-loop
      Merge-ongoing=1
    end if
  end for
end while

Algorithm 8 Transfer Operation

For any given network clustering state $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_s\}$
Transfer-ongoing=1
while Transfer-ongoing **do**
  Transfer-ongoing=0
  for all $\mathcal{C}_i \in \mathcal{C}$ **do**
    Update $\mathcal{C}_i$.nei
    for all $SC_{ix} \subset \mathcal{C}_i$ **do**
      for all $\mathcal{C}_j$ in $\mathcal{C}_i$.nei **do**
        Update payoff gain for possible Transfer$(SC_{ix}, \mathcal{C}_i, \mathcal{C}_j)$ i.e. $\delta_{v_{ixj}} = \{v(\mathcal{C}_i \backslash SC_{ix}) + v(\mathcal{C}_j \cup SC_{ix})\} - \{v(\mathcal{C}_i) + v(\mathcal{C}_j)\}$
      end for
    end for
    Find $(SC_{ix}, \mathcal{C}_i, \mathcal{C}_j)$ where $\delta_{v_{ixj}} = \max_{\substack{\mathcal{C}_j \in \mathcal{C}_i.nei \\ SC_{ix} \in \mathcal{C}_i}} (\delta_{v_{xij}})$ and $\delta_{v_{xij}} > 0$
    if $(SC_{ix}, \mathcal{C}_i, \mathcal{C}_j)$ exist **then**
      Transfer$(SC_{ix}, \mathcal{C}_i, \mathcal{C}_j)$
      Transfer-ongoing=1
    end if
  end for
end while

neighbour coalitions instead of all coalitions for a possible transfer. A list of SC clusters are kept as neighbours for UE_k based on received average reference signal level. For any user UE_k within the serving area of $SC_m \subseteq \mathcal{C}_m$, $\mathcal{C}_j$ is included in the neighbour list if $p_{kj}/p_{km} > P_{\Delta}^{nei}$ and $p_{kj} > P_{min}^{nei}$ where p_{km} and p_{kj} are the average signal power values received at UE_k from $SC_m \subseteq \mathcal{C}_m$ and $SC_j \subseteq \mathcal{C}_j$ respectively.

For each user coalition $\mathcal{U}_i \in \mathcal{U}$, users are checked for possible user transfer operation to all of its neighbour coalitions. The best transfer option with maximum additional payoff is implemented for UE_{ix} from $\mathcal{U}_i$ to $\mathcal{U}_j$ and user coalitions are updated. All other user coalitions are then checked for any possible user transfer and single user from each coalition with maximum payoff gain is transferred in a similar way. User transfers are limited to the ones with certain additional payoff δ_{Δ} which is introduced as an input parameter in the algorithm for the right balance between the number of user transfers and additional overall system payoff. User transfer operation is repeated for all user coalitions until no further user transfer is possible, as detailed in Algorithm 9. At the end of user transfer sub-game, a new user partition $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_s\}$ is formed where user coalition $\mathcal{B}_j$ is the associated users in SC cluster $\mathcal{C}_j$.

After forming the new user partition $\mathcal{B}$, SC clustering sub-game is re-deployed for further merge/split/transfer operations. In the case of any SC merge/split/transfer operation, user partition $\mathcal{B}$ is adjusted accordingly. For SC merge operation,

$\mathcal{C}_f = \cup_{i=1}^{s} \mathcal{C}_i$, the associated user coalitions are also merged $\mathcal{B}_f = \cup_{i=1}^{s} \mathcal{B}_i$. In the case of a SC cluster split operation of $\mathcal{C}_i$ into smaller coalitions $\{\mathcal{C}_{i1}, \mathcal{C}_{i2}, ..., \mathcal{C}_{iy}\}$, then associated user coalition $\mathcal{B}_i$ is also splitted to $\{\mathcal{B}_{i1}, \mathcal{B}_{i2}, ..., \mathcal{B}_{iy}\}$ based on each user's best serving SC within the cluster (not necessarily the best serving SC in the network as the user may have been transferred to non-best serving SC coalition during user transfer sub-game). For example, assume $SC_{ix} \in \mathcal{C}_i$ is the best serving SC within $\mathcal{C}_i$ for $UE_k \in \mathcal{B}_i$, then in the case when $\mathcal{C}_i$ splits and SC_{ix} falls in the new coalition $\mathcal{C}_{ix}$, then user coalition $\mathcal{B}_i$ is splitted similarly where $UE_k \in \mathcal{B}_{ix}$. Similarly for transfer operation of $SC_{ix} \subseteq \mathcal{C}_i$ transferring from coalition $\mathcal{C}_i$ to $\mathcal{C}_j$, users in $\mathcal{C}_i$ where SC_{ix} is the best serving SC within $\mathcal{C}_i$ are transferred from $\mathcal{B}_i$ to $\mathcal{B}_j$.

Both SC clustering and user-transfer sub-games are repeated until there is no further SC cluster or user cluster changes. As the utility for both sub-games are the same, each SC/UE coalition change improves the overall utility and converges to a final SC/user partition. We discuss stability of the algorithm and its complexity in the next subsection.

Algorithm 9 User Transfer Operation

For any given network clustering state $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_s\}$ and corresponding user coalitions $\mathcal{U} = \{\mathcal{U}_1, \mathcal{U}_2, ..., \mathcal{U}_s\}$
UserTransfer-ongoing=1
while UserTransfer-ongoing **do**
  UserTransfer-ongoing=0
  for all $\mathcal{U}_i \in \mathcal{U}$ **do**
    for all $UE_{ix} \subset \mathcal{U}_i$ **do**
      for all $\mathcal{U}_j$ in UE_{ix}.nei where $i \neq j$ **do**
        Update payoff gain for possible Transfer($UE_{ix}, \mathcal{U}_i, \mathcal{U}_j$) i.e. $\delta_{v_{xij}} = \{v(\mathcal{U}_i \backslash UE_{ix}) + v(\mathcal{U}_j \cup UE_{ix})\} - \{v(\mathcal{U}_i) + v(\mathcal{U}_j)\}$
      end for
    end for
    Find ($UE_{ix}, \mathcal{U}_i, \mathcal{U}_j$) where $\delta_{v_{xij}} = \max_{\substack{\mathcal{U}_j \in UE_{ix}.nei \\ UE_{ix} \in \mathcal{U}_i}} (\delta_{v_{xij}})$ and $\delta_{v_{xij}} > \delta_\Delta$
    if ($UE_{ix}, \mathcal{U}_i, \mathcal{U}_j$) exist **then**
      Transfer($UE_{ix}, \mathcal{U}_i, \mathcal{U}_j$)
      UserTransfer-ongoing=1
    end if
  end for
end while

5.5.3 Algorithm Stability

In this subsection, we prove that both SC clustering and user transfers sub-games always converge to a final partition and analyze the overall game stability.

Assume that the current state of the SC partition is $\mathcal{C}^1 = \{\mathcal{C}_1^1, \mathcal{C}_2^1, ..., \mathcal{C}_s^1\}$. In SC clustering sub-game, partition $\mathcal{C}^1$ is subject to merge-split-transfer operations which will transfer the network partition to $\mathcal{C}^n$ following a sequence of partitions.

$$\mathcal{C}^1 \rightarrow \mathcal{C}^2 \rightarrow, ..., \rightarrow \mathcal{C}^n \tag{5.14}$$

Any merge/split/transfer operation between coalitions $\mathcal{C}_i$ and $\mathcal{C}_j$ will increase the overall utility of the involved SCs/coalitions following utilitarian preference order i.e. $v(\text{Merge/Split/Transfer}(\mathcal{C}_i^1, \mathcal{C}_j^1)) > (v(\mathcal{C}_i^1) + \mathcal{U}(\mathcal{C}_j^1))$. As detailed in Section 5.3, we assume that clustering decisions are made in longer time intervals (seconds, minutes), fast fading component of the signal is averaged out for clustering decision and hence the interference created from any $SC \in (\mathcal{C}_i^1 \cup \mathcal{C}_j^1)$ to the rest of the network is the same regardless of any merge/split/transfer changes within $(\mathcal{C}_i^1 \cup \mathcal{C}_j^1)$. Hence, $v(\mathcal{C}^1 \backslash (\mathcal{C}_i^1 \cup \mathcal{C}_j^1))$ is unchanged when there is any merge/split/transfer operation between coalitions $\mathcal{C}_i^1$ and $\mathcal{C}_j^1$.

As $v(\text{Merge/Split/Transfer}(\mathcal{C}_i^1, \mathcal{C}_j^1)) > (v(\mathcal{C}_i^1) + v(\mathcal{C}_j^1))$, and there is no change for the rest of the network as a result of this operation, then the overall system utility always increases with every partition in sequence (5.14), i.e.

$$v(\mathcal{C}^n) > v(\mathcal{C}^{n-1}) ... v(\mathcal{C}^2) > v(\mathcal{C}^1) \tag{5.15}$$

where $\mathcal{C}^i \neq \mathcal{C}^j, i \neq j$.

As the overall system utility is always increased with every partition in the sequence i.e. same partition will never be visited again and there are finite number of partitions limited by the Bell number, then the sequence in (5.14) is guaranteed to converge to a final SC partition.

For a given fixed SC partition $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, ..., \mathcal{C}_s\}$, associated user coalitions $\mathcal{U}^1 = \{\mathcal{U}_1^1, \ldots, \mathcal{U}_s^1\}$ will be subject to user transfers which will transform the user coalitions into $\mathcal{U}^t$ and the overall system utility of SC partition $\mathcal{C}$ will increase with every user partition change as per the definition of user transfer rule following utilitarian order i.e.

$$v(\mathcal{U}^t) > v(\mathcal{U}^{t-1}) ... > v(\mathcal{U}^2) > v(\mathcal{U}^1) \tag{5.16}$$

where $\mathcal{U}^i \neq \mathcal{U}^j, i \neq j$.

Similar to SC partition convergence, as there is finite number of user partitions

limited by the Bell number, and user partitions will always evolve to a better utility, then user partition sequence is guaranteed to converge to a final partition.

When both sub-games are employed jointly, the overall system utility is always increased with every SC/user partition change, and hence same SC and user partition will never be re-visited. There will be a finite number of possible SC/user partitions and therefore the overall SC/user partition will always converge to a final SC/user partition.

5.5.4 Algorithm Complexity

Exhaustive search for any SC merge/transfer or user transfer is a highly complex task where the number of possibilities increase exponentially as the network size increase. In our algorithm, we introduce neighbour SC/coalition concept and limit merge/transfer operations for SC and users only to the neighbour coalitions which reduces complexity significantly and makes the algorithm scalable for larger networks. Thresholds set for coalition neighbours can be adjusted for a more relaxed/tighter neighbour definition and increase/reduce the number of clustering actions for the right balance between additional CoMP gain and complexity.

To further reduce complexity, user transfers are limited to only the ones with higher additional payoff, i.e. marginal payoff increase does not trigger user transfers and hence SC clustering sub-game is also not triggered unnecessarily. This is a controlled input parameter in the algorithm to balance a trade-off between higher payoff and increased complexity. A similar additional payoff input threshold can also be introduced for SC clustering sub-game for merge/split/transfer operations to further reduce the complexity but with the expense of lower overall system utility for the final partition.

Split operation can be complex when cluster size is high, as the number of split options increase exponentially with CS. Due to CoMP overheads, CS needs to be kept low, and we incorporated a soft maximum CS (C^n_{max}) into our utility function to limit CS and hence reduce complexity for split operation. Furthermore, for a given SC coalition, once a split option with additional payoff is found, this is implemented without checking the rest of the possibilities to reduce complexity. Moreover, as detailed in Section 5.5.3, any split operation for $\mathcal{C}_i$ does not depend on how the rest of the coalitions are structured, so if any coalition $\mathcal{C}_i$ is checked for split options and there is no split possibility, then it is not checked again in further iterations unless there is further changes in the same coalition $\mathcal{C}_i$ with merge/transfer operation or user transfers.

In summary, we propose a low complexity SC and user clustering model which can be configurable to find the right balance between higher complexity and ad-

ditional system payoff (i.e. better spectral efficiency/load balancing). We utilize the configurable neighbour concept to avoid exhaustive search for merge/transfer operation and enable the solution to deploy in large scale networks. We further reduce complexity by reducing user transfers to only those with significant additional payoff. SC merge/split/transfer operations can be also reduced to limit these operations to only high additional payoff which is configurable to network requirements. Furthermore, we also limit the maximum CS which then reduce the complexity on split operation.

5.6 Simulation Results

In this section, we present the simulation results to evaluate the performance of our RAN and backhaul-load aware clustering model. Firstly, we run our simulation for HN scenario with hotspots and illustrate the clusters formed by each algorithm. We then run extensive simulations for RN scenario with and without hotspots. As described in Section 5.3, we assume a heterogeneous network composed of one MBS overlaid with SCs, where each SC is a single cell with an omni-directional antenna. We compared our RAN/backhaul-load aware model with an improved version of our previous RAN-load aware clustering work in Chapter 4 (also published in [31]). We applied our novel two-stage coalitional game model and employed the same load aware utility (5.12) but only considered RAN load in the utility, rather than a more comprehensive RAN-backhaul overall load. Furthermore, our RAN/backhaul-load aware clustering model is also compared with an improved version of a spectral efficiency based greedy model presented in [134]. Our spectral efficiency based utility function (5.13) is employed in the greedy algorithm where maximum cluster size limitation is lifted with the introduction of implicit soft cluster size limit via the cost function in the utility. Additionally, neighbour concept introduced for our RAN/backhaul-load aware model is also employed in the greedy algorithm. Improved version of the spectral efficiency based greedy algorithm is presented in detail in Chapter 4.

In the rest of the Chapter, following abbreviation is used for the presented clustering models:

- **SE-GR**: Greedy model employing spectral efficiency based utility (5.13).
- **L-GA**: RAN-load aware game theoretic model with load based utility (5.12) considering RAN load only
- **LBH-GA**: RAN and backhaul-load aware game theoretic model with load based utility (5.12) considering combined RAN and backhaul load.

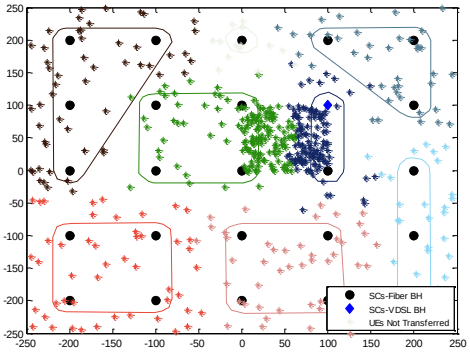
Firstly, we run simulations in HN deployment with hotspot scenario to illustrate the clusters formed by each algorithm. In HN scenario, 25 SCs are deployed in 500mx500m simulation area with 100m inter-site distance. 300 UEs are distributed in the whole simulation area, following uniform random distribution. A further 200 UEs are also uniformly distributed within a 100mx100m area to simulate a hotspot scenario. One of the SCs serving to the hotspot area is assumed to have VDSL2 backhaul and the rest of the SCs are assumed to have fiber backhaul connection to the MBS. Each UE is assumed to have a fixed GBR requirement of 2048kbps. Pathloss model is adapted from ITU-R microcell urban NLOS model [83] as follows: $PL = 36.7\log_{10}(d) + 22.7 + 26\log_{10}(fc)$ where d is distance in meters and fc is the carrier frequency in GHz. Rest of the simulation parameters are summarized in Table 5.1.

Figure 5.1a depicts the clusters formed by SE-GR algorithm in HN deployment scenario with hotpot. As SE-GR clustering starts from random SC, it fails to achieve a cluster around the loaded cells. As shown in Figure 5.1b, L-GA algorithm forms the cluster around the hotspot area as the algorithm utility takes cell load into account, and gives priority to loaded SCs for clustering. Furthermore, L-GA cluster size is increased around the hotpot, giving better spectral efficiency and hence reduced load. Figure 5.1c shows clusters formed by LBH-GA model where a cluster is formed around the hotspot, but the only one VDSL site is excluded from this cluster as backhaul capacity limitation introduce a higher backhaul load than RAN-load reducing the utility gain for forming a cluster. LBH-GA cluster size around the hotspot is still increased similar to L-GA cluster but site with VDSL is excluded, and an additional nearby site is included to increase spectral efficiency and reduce load.

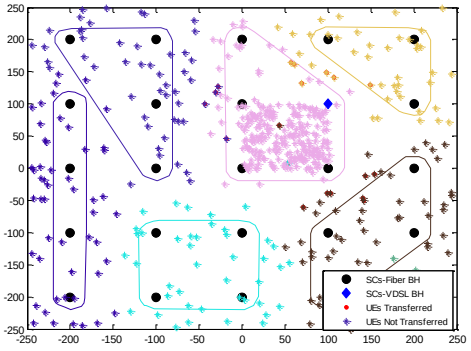
We performed extensive simulations in a more realistic RN scenario with and without hotspots to compare our novel LBH-GA model against SE-GR and L-GA. In our simulation setup, we deployed SCs randomly following Poisson point process (PPP) distribution with density parameter $\lambda_{\mathcal{C}}$ within a circle of 0.4m radius. UEs are also randomly distributed following PPP distribution. In RN with hotspot scenario, we simulated a hotspot area in an inner circle with 0.1m radius where a higher density $\lambda_{\mathcal{U}_{high}}$ of UEs are deployed and a lower density $\lambda_{\mathcal{U}_{low}}$ of UEs are deployed in the outer ring where the radius is set to 0.5m. UE deployment area is set to a bigger radius than SC deployment area to make sure that UEs are distributed to the whole coverage area of the SCs. GBR for UEs within the hotspot are set to 2048kbps whereas GBR for UEs outside of the hotspot ring are set to 256kbps. For RN without hotspot scenario, UE density is set to $\lambda_{\mathcal{U}_{low}}$ for both inner and outer ring areas and GBR is set to 256kbps for all

Table 5.1: Simulation Parameters

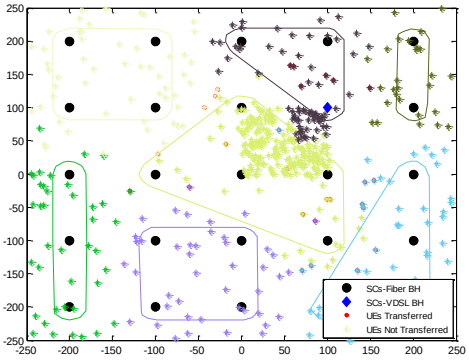
Parameter Name	Parameter Value
Simulation Enviroment	Urban Microcell [83]
Frequency Carrier	5 GHz
Channel Bandwidth	20 MHz
PRB Bandwidth (B_{PRB})	180 kHz
Number of PRBs/SC (R_{tot})	100
Shadow fading std	4 dB [83]
UE Antenna Gain	0 dBi
UE Thermal Noise Density	-174 dBm/Hz
TP Total Transmit Power (P_{Tx})	41dBm [83]
UE Noise Figure	7dB
TP Noise Figure (inc cable loss)	5dB
SC antenna gain (boresight)	17dBi
User-centric cluster: Min RX Power (P_{min})	-110dBm
User-centric cluster: Max RX power offset (P_{Δ})	20dB
Min RX power for neighbour Def. (P_{min}^{nei})	-110dBm
Max RX power offset for neighbour Def. (P_{Δ}^{nei})	-20dB
Min payoff gain for user transfer operation (δ_{Δ})	10
RN Simulation Area Radius	0.5km
RN SC deployment Area Radius	0.4km
RN Hotspot Area Radius	0.1km
SC Density for RN ($\lambda_{\mathcal{C}}$)	80SC/km^2
UE Density within hotspot in RN Scenario ($\lambda_{\mathcal{U}_{high}}$)	6000UE/km^2
UE Density outside hotspot in RN Scenario ($\lambda_{\mathcal{U}_{low}}$)	200UE/km^2
UE Density in RN Scenario without Hotspot ($\lambda_{\mathcal{U}_{low}}$)	200UE/km^2
GBR for UEs in the hotspot in RN Scenario	2048 kbps
GBR for UEs outside the hotspot in RN Scenario	256 kbps
GBR for UEs in RN Scenario without hotspot	256 kbps
GBR for UEs in HN Scenario	2048 kbps



(a) SE-GR Clusters in HN with hotspot.



(b) L-GA Clusters in HN with hotspot.

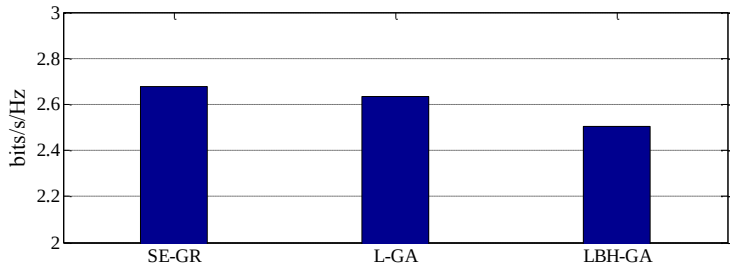


(c) LBH-GA Clusters in HN with hotspots.

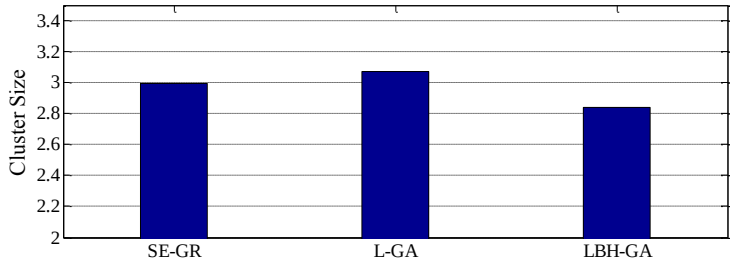
Figure 5.1: Snapshot of SE-GR, L-GA and LBH-GA clusters in HN with hotspot scenario.

UEs.

We first analyze the results in RN without hotspot scenario. We ran our simulation for 100 snapshots where 33% of the SCs are assumed to have VDSL2 backhaul connectivity to the MBS and the remaining ones have fiber. Figure 5.2a and 5.2b depicts the achieved spectral efficiency and CS respectively for all 3 algorithms. We observe that L-GA performs similar to SE-GR when there is no hotspot with marginal difference in achieved spectral efficiency and CS. LBH-GA achieves a slightly lower CS value when compared to L-GA due to backhaul limitations on some sites been taken into account. As observed in HN clustering scenario, LBH-GA tends to exclude sites with VDSL2 connection. For SCs with VDSL2, RAN load is the limiting factor in low cluster size, and as the cluster size increase, backhaul load becomes the limiting factor in our simulation setup with 20 Mhz channel bandwidth. Unlike RAN load, any cluster size increase for the VDSL2 site will always increase backhaul load as there will be additional users



(a) SE in RN without hotspot scenerio.

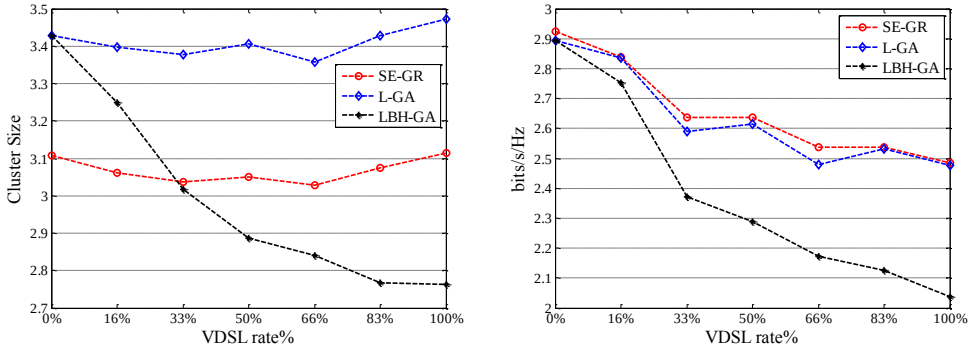


(b) CS in RN without hotspot scenerio.

Figure 5.2: Average spectral efficiency and CS in RN without hotspot scenerio when VDSL2 rate= 33%.

being scheduled within VDSL2 SC and the user-data for the additional users will be added to the backhaul load regardless of the spectral efficiency improvement. Once backhaul load is higher than RAN load, any CS increase will not be allowed as it will increase overall load for VDSL2 sites which introduce extra cost in the utility function i.e. reduction in payoff for the VDSL2 site. When backhaul load is the limiting factor, VDSL2 site only enters into a CoMP set when the additional payoff for other SCs with fiber are greater than the payoff loss for the VDSL site. In other words, when backhaul load is taken into account i.e. for LBH-GA model, it is harder to get backhaul limited SCs within CoMP clusters. Overall, without hotpots, L-GA achieves similar results to SE-GR and LBH-GA achieves marginally less CS due to not promoting CoMP on sites with VDSL2.

We run further simulations in RN with hotspot scenario for different rates of fiber connection availability in the network. Seven different fiber/VDSL2 availability rates are considered and 100 snapshots of simulations are run for each scenario. Figure 5.3a shows the average CS for each VDSL2 rate in hotspot scenario where L-GA CS is consistently higher than SE-GR. This is in-line with HN simulations and the clustering snapshot shown in Figure 5.1 where L-GA cluster size is increased when there is high load to improve spectral efficiency and reduce the load. LBH-GA results in the same CS with L-GA when VDSL2 rate= 0% and average CS is reduced as the VDSL2 rate increases. LBH-GA tends to form

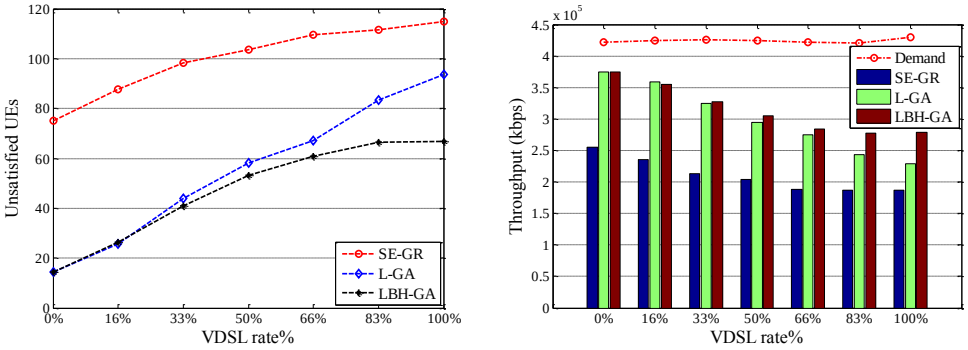


(a) Cluster size comparision. (b) Spectral efficiency comparison.

Figure 5.3: Cluster size and spectral efficiency comparision for all backhaul cases.

clusters without the SCs with VDSL2 for the same reasons we discussed in RN without hotspot scenario. As shown in Figure 5.3b, a similar trend is observed in average spectral efficiency, following average achieved CS as expected. Intuitively, increased CS helps in eliminating further inter-cell interference and hence improve spectral efficiency. Figure 5.4a depicts the unsatisfied UE count for each of the algorithms at different VDSL2 rate scenarios. L-GA model reduces the unsatisfied users by 80.6% when compared to SE-GR model when there is no SC with VDSL connection. As VDSL2 rate increase, unsatisfied users increase in all models as expected, however LBH-GA model achieves the lowest mount of unsatisfied users with 41.7% and 18.4% less unsatisfied users when compared to SE-GR and L-GA respectively in the case when all SCs are connected with VDSL2. LBH-GA model achieves a better load balanced network with less unsatisfied users while CS are kept low and hence low computational complexity for CoMP deployment. Similar to unsatisfied UEs, system throughput is also significantly improved in LBH-GA model when compared to SE-GR model as depicted in Figure 5.4b. An average of 49.9% increase in overall system throughput is observed with LBH-GA when compared to SE-GR across all backhaul scenarios. As the VDSL rate increase, overall throughput gets worse for all models as expected, however LBH-GA throughput gets better when compared to L-GA as LBH-GA model clustering takes backhaul availability into account where SCs with VDSL2 are not preferred in clusters of highly loaded cells. LBH-GA achieves 21.9% higher overall throughput when compared to L-GA in the case when all SCs have VDSL2 backhaul.

We further look at an example scenario of 50% VDSL2 rate in RN with hotspot and analyze the details of each sub-game actions (i.e. SC merge/split/transfer actions and UE transfers) and the changes in spectral efficiency, unsatisfied UEs and game payoff during the iterations. Figure 5.5 shows the changes in average



(a) Unsatisfied UEs comparison. (b) System throughput comparison.

Figure 5.4: Unsatisfied UEs and system throughput comparison for all backhaul cases.

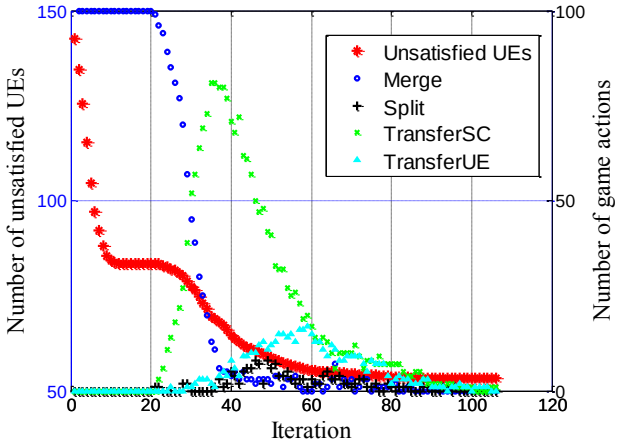
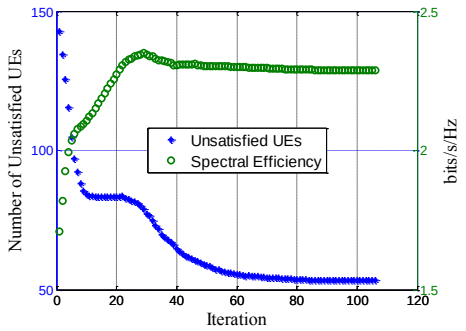


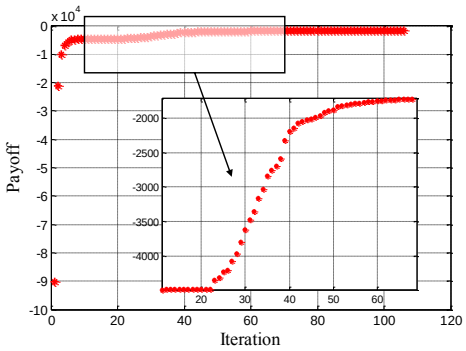
Figure 5.5: LBH-GA game actions vs. unsatisfied UEs for 50% VDSL rate.

number of unsatisfied UEs and the number of each game action at each iteration for the 100 snapshots run in this scenario. At the start of the game, average number of unsatisfied UEs are sharply reduced as the initial clusters are formed and spectral efficiency is improved with merge operations. Later iterations of merge operations only gives marginal improvements and other game actions start increasing. SC transfer actions are significantly high at the next stage where unsatisfied users are further reduced significantly. It can be noted that number of split operations are relatively low when compared to other SC game actions. UE transfer actions are also in relatively high numbers and can be controlled with δ_Δ parameter to allow only most significant UE transfer actions as discussed in Section 5.5.4. Figure 5.6a shows the average changes in spectral efficiency and the number of unsatisfied UEs at each iteration. Spectral efficiency is increased sharply with the initial merge actions and but reduced marginally in the later

actions due to priority on load balancing actions being not necessarily the best action for increasing spectral efficiency. Number of unsatisfied UEs continues to reduce at each game action. Overall system payoff is depicted in Figure 5.6b where a similar pattern to number of unsatisfied UEs is observed where a sharp improvement is observed in the initial merge actions and it continues to improve in smaller intervals during following game actions.

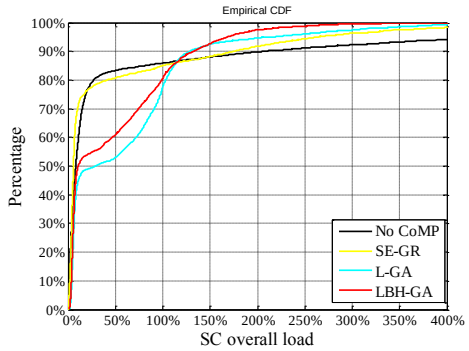


(a) LBH-GA unsatisfied UEs vs spectral efficiency.

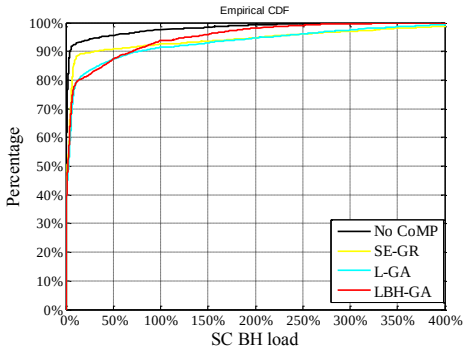


(b) LBH-GA payoff.

Figure 5.6: LBH-GA unsatisfied UEs vs spectral efficiency and payoff for 50% VDSL rate.



(a) SC overall load distribution.



(b) SC backhaul load distribution.

Figure 5.7: SC overall load and backhaul load distribution for 50% VDSL rate.

The resulting overall load distribution of all SCs in all 3 algorithms are shown in Figure 5.7a for the 50% VDSL2 rate case. LBH-GA model clearly achieves better load distribution where highly loaded SCs are reduced and traffic is moved to lightly loaded SCs increasing their load within light load range. Figure 5.7b shows the backhaul load distribution for all SCs, and it is clear that backhaul load increases sharply when CoMP is enabled as user data needs to be available in multiple SCs in our JT-CoMP scenario. A significantly better backhaul load

distribution with low number of SCs with high load is achieved with LBH-GA which results in significantly better system throughput i.e. increased system capacity.

5.7 Conclusion

We have presented a novel low-complexity, multi-objective clustering model in MU JT-CoMP scenario where spectral efficiency, RAN-load and backhaul-load are optimized collectively. An SC merge/split/transfer sub-game and a UE transfer sub-game are designed. Game properties, complexity and stability analysis are presented. It is shown that our novel LBH-GA algorithm is a low complexity model which is scalable and always converges to a final cluster. Simulation results are compared to a RAN-load aware model (L-GA) and an spectral efficiency based greedy (SE-GR) algorithm from our previous work in Chapter 4 to show the impact of backhaul awareness. We show that LBH-GA successfully forms clusters dynamically around the hotspots and excludes backhaul limited SCs when possible to improve the spectral efficiency and reduce overall load. In hotspot scenario where throughput demand is higher than the overall capacity, average system throughout is increased by 49.9% with LBH-GA when compared to SE-GR model. Average throughput is also increased by 21.9% when compared to L-GA model in the case when all SCs are backhaul limited (VDSL2). LBH-GA model is also effective in scenarios without hotpots, dynamically adjusting CS based on backhaul availability and load conditions. Our presented model provides a low complexity, stable framework where it can be enhanced further with improved utility functions to include additional network objectives and provide the right balance between CoMP overhead costs and various objectives based on network requirements.

Chapter 6

CoMP for Future Networks: Field Trial Results

6.1 Introduction

In this chapter, we aim to present an operator's perspective on deployment of CoMP. Firstly, we present the main motivation and benefits of CoMP from an operator's viewpoint. Next, we present the operational requirements for CoMP implementation and discuss practical considerations and challenges of such deployment. Possible solutions for these experienced challenges are reviewed. We then present initial results from a UL CoMP trial and discuss changes in main KPIs during the trial. Additionally, we propose further improvements to the trialled CoMP scheme for better potential gains. Moreover, we give our perspective on how CoMP will fit into the future 5G networks and finally conclude the chapter with a summary of lessons learnt. Most of this work is published as a chapter in [81].

6.2 Motivation and Benefits of CoMP - Operator Perspective

Mobile network operators have been looking for various solutions to improve network capacity as briefly discussed in Chapter 2. Driven from the capacity improvement objective, LTE networks have been rapidly deployed, degree of sectorisation is increased in densely populated areas, micro,pico and wifi cells are deployed to offload the MBS network. Furthermore, multiple antenna solutions (i.e. MIMO) is introduced especially with the LTE network deployment. Moreover, additional spectrum is utilized where possible and solutions for increasing

spectral efficiency on existing spectrum has been explored extensively. Operators globally have been re-purposing their existing frequency spectrum, usually used for 2G/3G technologies, to LTE/4G in the drive to achieve increased efficiency, grow capacity and offer improved user experience.

Further to all above, CoMP has also been in the interest of operators to further improve spectral efficiency in densely deployed LTE-A networks. CoMP can eliminate interference from cells within the coordinating cluster and even can exploit this signal as useful signal. In interference-limited, densely deployed networks, CoMP has got the potential to improve spectral efficiency and increase user throughout especially at the cell edge where there is significantly more interference. Given that some of the CoMP types do not require additional infrastructure and relatively low cost (such as intra-site uplink CoMP), it is in the interest of operators to deploy CoMP for LTE-A to improve much needed capacity. More complex CoMP deployment scenarios introduce challenges such as high backhaul bandwidth requirement and very low latency, UE capability, clustering challenge, complex precoding and precise synchronization requirement etc. These challenges and possible solutions are briefly discussed in the next section.

6.3 Operational Requirements

As discussed in Chapter 2, CoMP can generally be categorized into three main types of coordination schemes; CS/CB, JP, and DPS. Each of these techniques impose varying delay and bandwidth requirements on the backhaul which in turn depend on backhaul technology employed (fibre, ADSL, copper etc.). Studies conducted by the 3GPP and other literature [6, 127, 128] show that highest gain in spectral efficiency, and thus consequent capacity requires virtually zero latency and unlimited capacity backhaul. This practically requires C-RAN architecture using fibre connectivity to connect RRHs to central BBU. In practice, there will be constraints that will limit achievable gain in distributed RAN architecture [126].

Operators with fibre optics assets face relatively easier route to deployment of CoMP than those with no or limited fibre connectivity. Although varies by market conditions, acquiring fibre optic connectivity is likely to attract certain OPEX, CAPEX and internal organization resources. Justifying such investment and upgrade programmes solely on the basis of CoMP benefits is likely to be a difficult business decision.

CoMP algorithms and associated functionality are implementation dependent and requires certain processing power. Even simplest deployment on intra-site CoMP requires additional uplifting of the physical baseband resources. In some

cases, where there is no baseband capacity headroom left, additional site visits are required to upgrade site physical processing capacity as part of CoMP deployment. This incurs certain costs and potentially a degree of service disruption if outage is required to carry out physical upgrade activities. Operational and cost impact will scale up, proportionate to network size and the number of sites to be upgraded. Service providers will need to factor that in their upgrade plans. Unnecessary multiple site visits and potential service disruption resulting from upgrade activity could be prevented by including CoMP requirements at the early stage of the design process to ensure site physical resources are optimally dimensioned.

Propagation environment and carried traffic load will influence the achievable gain in spectral efficiency everything being equal. Selection of cells that make up the CoMP set is very important factor for inter-site deployment. A planning process is required to identify cooperating cells and group them. Such a process will take into account coverage overlap, traffic load, cell type, availability and type of backhaul. After completing an initial planning processing and identifying the cells that make up the CoMP set, it is important to ensure CoMP set is kept optimum by having a feedback loop to cater for changes in the RF environment resulting from various reasons such as new sites, new building development, amongst others. Managing this process manually is likely to be tedious and inaccurate and therefore automation is very important capability to enable large scale deployment of CoMP.

6.4 Uplink CoMP Field Trial for LTE-A

In this section, we briefly introduce UL-CoMP for LTE-A in general and provide intra-site UL-CoMP trial results from an operational LTE-A network for 3 deployment scenarios in 2 trial areas. Improvements in various KPIs are reviewed and limitations of the trialled CoMP scheme are critically discussed. Further potential enhancements for future networks are highlighted to maximize CoMP gain.

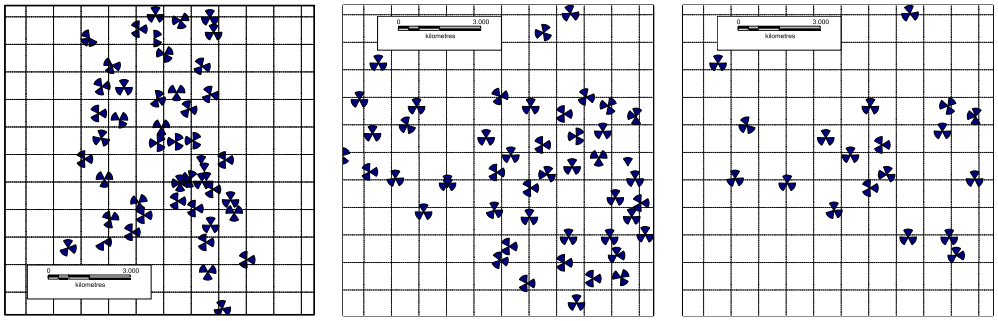
6.4.1 Uplink CoMP Introduction

Uplink CoMP makes use of the UE signal at different BSs to improve SINR and hence cell throughput especially at the cell edge where severe inter-cell interference is usually experienced. Unstandardized applications of exploiting uplink signal from different receiver ports may already be available in conventional net-

works [94] but recent standardized efforts by 3GPP introduced UL-CoMP for LTE-A networks [6]. Two types of UL CoMP categories are identified by 3GPP, namely joint reception(JR) and coordinated scheduling/beamforming (CS/CB). In JR, UE signal is received by multiple points and processed jointly to eliminate interference. Signal processing techniques like interference rejection combining (IRC) are employed to mitigate interference and improve SINR. A tutorial on IRC in LTE networks can be found in [97]. In CS/CB scheme, UE data is intended for one point only and interference is minimised within the coordinated set by central scheduling and beamforming.

UL CoMP is a relatively less complex option for operators to deploy initially, as it is transparent to the UE, and also channel estimation is done by uplink reference signals, i.e. there is no feedback signalling overhead unlike downlink CoMP. Current LTE-A networks are under more pressure for DL capacity than UL as data demand for DL is typically eight times more than UL. However, in special scenarios such as football matches, music festivals, UL data demand increases upto half of the DL data demand. This is evidently explained by consumption of application and services that generate significant data in UL direction like users taking photos/video and sharing with friends and family on social networks like Instagram and Facebook. This, combined with relatively low complexity and operational overhead makes deployment of UL CoMP in such scenarios very attractive proposition to enhance capacity and improve user experience.

As we have demonstrated in this thesis, one of the important factors for maximizing CoMP gains is to decide which and how many cells to coordinate for finding the right balance between complexity/overhead costs and CoMP gain. On the other hand, these clusters will need to dynamically change in response to the spatio-temporal changes in user profile/demand distribution and network elements. Although, fully dynamic CoMP clustering solutions are studied extensively in literature, such solutions are not yet available for current LTE-A networks to our knowledge. Furthermore, available backhaul is mostly limited, preventing such dynamic design for non-co-located BSs in the CoMP cluster. Therefore, realistic CoMP cluster solution for current LTE-A networks for initial deployment is the intra-site CoMP where coordination only takes place within the same site with joint baseband processing unit. Since all cells are geographical collocated and belong to same site, backhaul delay can be assumed to be virtually zero for intra-site deployment.



(a) Dense deployment. (b) Medium density deployment. (c) Sparse deployment.

Figure 6.1: Network topology of the 3 deployment scenarios in the trial.

6.4.2 UL CoMP in Trial Area

Intra-site UL JR-CoMP was trialled in 2 different major cities in the UK where LTE-A BSs are deployed at macro layer with typically 3 cells at each BS. Trial area-A covers a middle-size city in the United Kingdom (UK) with dense deployment. One frequency layer at 800 MHz is deployed in this area. Trial area-B covers outskirts of another city in the UK where LTE-A carrier at 800 MHz is deployed at medium site density in this area. An additional layer at 1800 MHz is also deployed in trial area-B as a capacity layer in local areas where it is required for additional capacity. Hence deployment at 1800 MHz is at lower density than 800 MHz. In the rest of the chapter, we refer to trial area-A as "dense deployment at 800 MHz", and trial area-B as "medium density deployment at 800 MHz" and "sparse deployment at 1800 MHz". Figure 6.1 depicts the site layout and Table 6.1 shows the approximate areas covered and average inter-site distance for each deployment scenario. We discuss UL CoMP gains achieved separately for each of these 3 deployment scenarios. CoMP sets are formed from co-located cells on the same frequency band of the same site, where each cell consists of 2 RX antennas. Co-located cells share the same baseband processing unit and hence there is no backhaul requirement for data/signalling exchange between the cells. Maximum of 2 cells are allowed (i.e. 4 RX antennas in total) to coordinate at any time where IRC receiver is employed to extract the main signal and eliminate inter-cell interference. For each UE, serving cell and the strongest neighbour cell from the CoMP set form the coordination set. CoMP is enabled on the uplink data channel only i.e. PUSCH, it is not deployed on the control channels i.e. physical random access channel (PRACH) or physical uplink control channel (PUCCH). Uplink CoMP scheme in the trial area is illustrated for 2 cells within the same BS in Figure 6.2.

Table 6.1: Site Density Details for the 3 Deployment Scenarios

Deployment Scenario	**Approx. Area** (km^2)	**Aver. Inter-Site Dist** (km)
Dense Deployment	63.6	0.733
Medium Density Deployment	112.4	0.937
Sparse Deployment	112.4	1.509

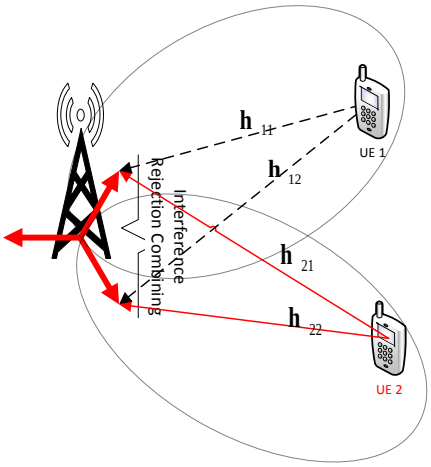


Figure 6.2: An illustration of intra-BS Uplink CoMP: Joint processing with IRC

6.4.3 Trial Performance Results

Intra-site UL CoMP is enabled in the 3 LTE-A network layers at 2 trial areas as described above for 2 weeks and various KPIs are benchmarked against the identified benchmarking time window of 2 weeks before and after the trial. Changes in the KPIs are also presented for a wider time window. Figure 6.3 depicts the average SINR change in logarithmic scale on PUSCH based on network counters for the 3 deployment scenarios in the trial. As expected, SINR value is higher on dense deployment in general and it reduces in medium density deployment and further reduced in sparse deployment. Average SINR is improved with UL-CoMP in all 3 deployment scenarios. Largest increase in average SINR is observed in medium density deployment scenario with 8.57% (+0.67 dB) SINR improvement, followed by 4.55% (+0.32 dB) and 3.56% (+0.34 dB) increase in sparse and dense deployment scenarios respectively.

Average SINR gain due to UL CoMP is also reflected in the employed modulation type on UL. Figure 6.4 shows the average percentage usage of the 2 modulation types employed in uplink i.e. QPSK and 16QAM. Intuitively, higher SINR in dense deployment scenario is reflected in wider use of higher modulation type

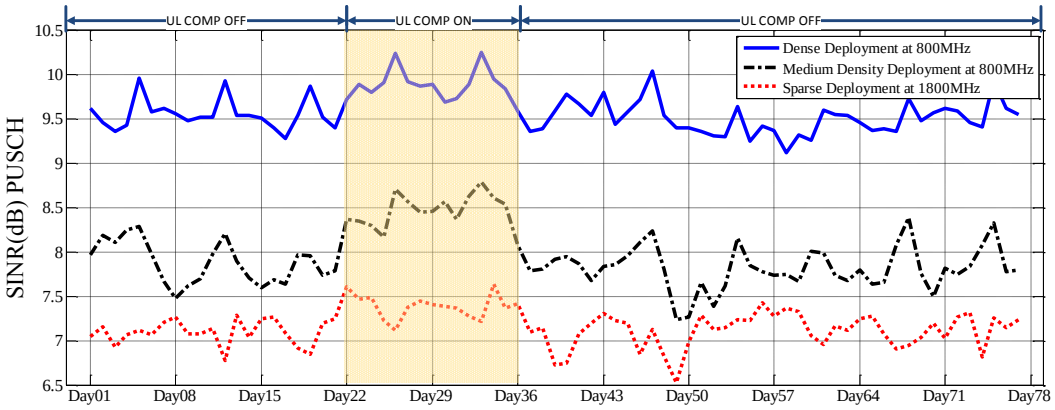


Figure 6.3: PUSCH SINR with/without UL CoMP in the 3 deployment scenarios

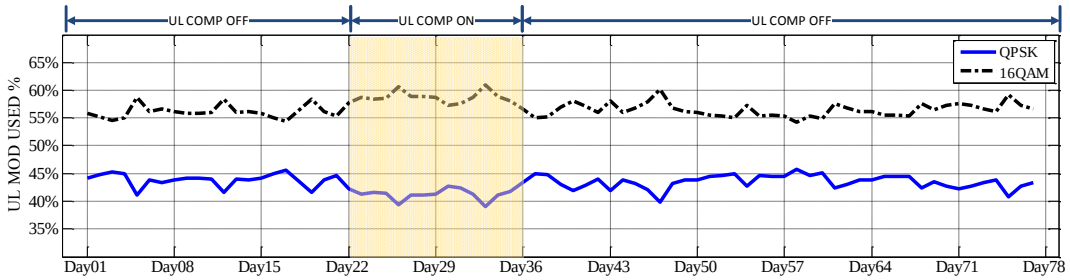
when compared to lighter deployment scenarios. The average percentage usage of 16QAM against QPSK is increased by 7.46%, 4.68% and 4.00% in medium density, sparse and dense deployment scenarios respectively following a similar gain pattern from average SINR increase. UL CoMP can be more effective in dense deployment scenarios, where inter-cell interference is expected to be higher, however intra-site CoMP is employed in this trial where interference from cells located at different BS locations are not mitigated.

Figure 6.5 depicts the average achieved block error rate (BLER) for each modulation type on UL during the trial. BLER improvement is observed in all modulation types for all deployment scenarios apart from inconclusive results in BLER for QPSK modulation for sparse deployment scenario. Overall, an increase is observed on the average ratio of higher modulation usage and also average BLER for each modulation is improved by 7.69%, 9.51% and 4.28% for dense, medium density and sparse deployment scenarios respectively.

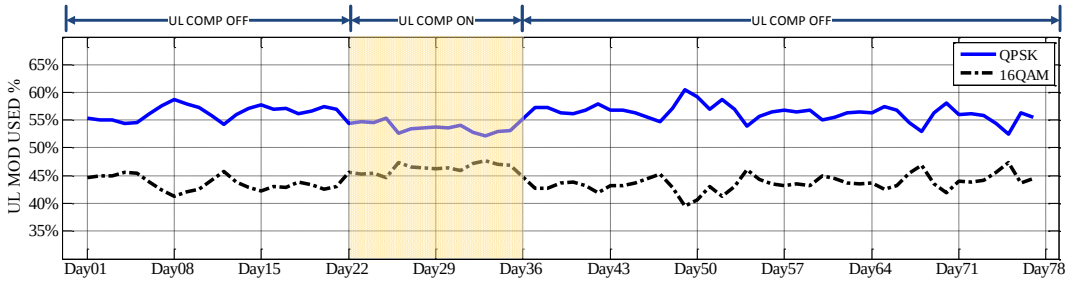
Improvement in the employment of higher order modulation type is also reflected in average UL user throughput. Figure 6.6 depicts the increase in average UL user throughput in all 3 scenarios. Average improvement observed during the trial are 17.86%, 8.68% and 7.42% in medium density, sparse and dense deployment scenarios respectively.

CoMP is more effective at cell edge, as there is more inter-cell interference expected and UEs use full power to overcome this interference. Figure 6.7 depicts the percentage of UEs which are power limited, where a slight reduction can be seen during the trial. An average of 1.92%, 1.85% and 1.17% reduction is observed in the amount of power limited UEs for dense, medium and sparse deployment scenarios respectively during the trial.

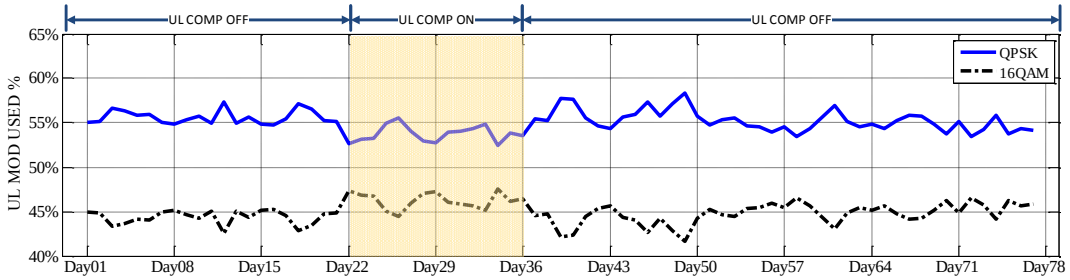
From customer experience point of view, number of drops and overall UL packet loss ratio is also monitored during the trial. A clear reduction in UL Packet



(a) Employed uplink modulation type in dense deployment at 800 MHz.



(b) Employed uplink modulation type in medium density deployment at 800 MHz.

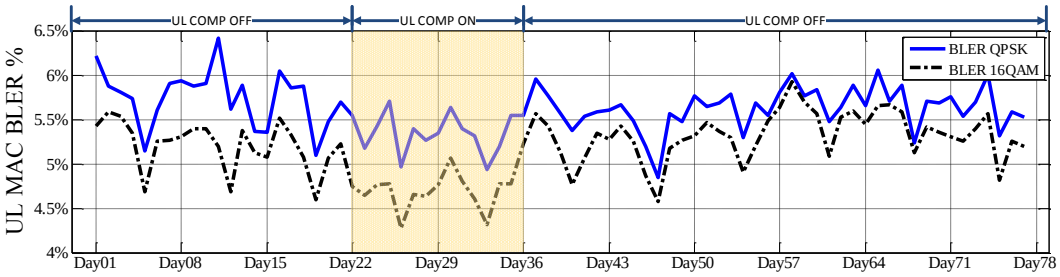


(c) Employed uplink modulation type in sparse deployment at 1800 MHz.

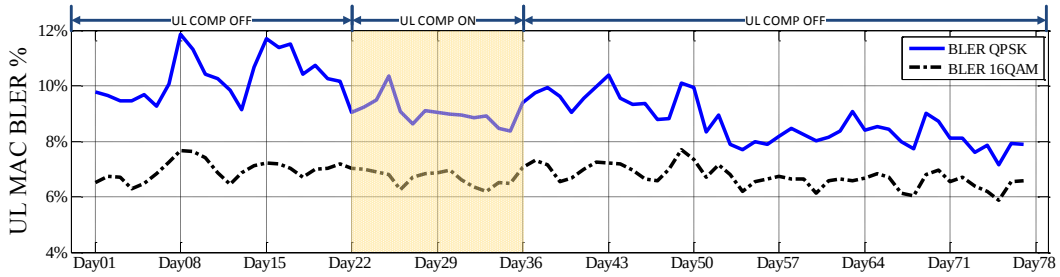
Figure 6.4: Employed uplink modulation types with/without UL CoMP in the 3 deployment scenarios.

loss rate is observed in all 3 scenarios, where the average improvement has been 8.20%, 10.48% and 14.68% for dense, medium and sparse deployment scenarios respectively. Number of drops is also reduced in dense and medium deployment by 1.75% and 7.19% respectively and the results for sparse deployment was not conclusive. Figure 6.8 shows the radio drops and UL packet loss rate for dense deployment case where an improvement on both metrics can be observed clearly during the trial.

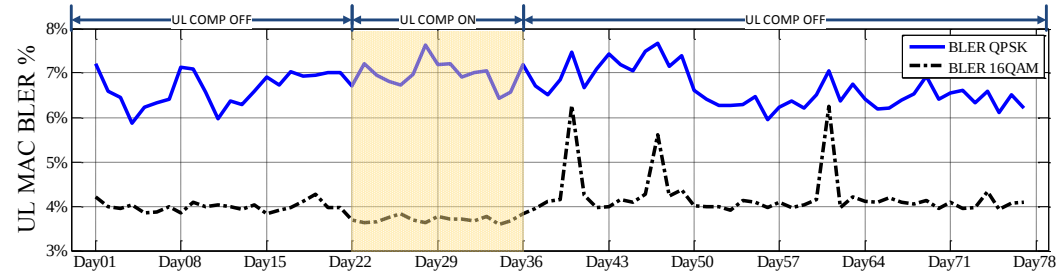
In summary, a significant performance gain is achieved by enabling intra-site UL CoMP in 3 different macro network deployment scenarios. A notable increase in average SINR on PUSCH is observed which is also reflected on the usage of higher order modulation in UL. Consequently, average UL user throughput is also increased and overall session drops are reduced. However, a number of factors have limited further CoMP gains achievable. Firstly, current network



(a) Uplink BLER in dense deployment at 800 MHz.



(b) Uplink BLER in medium density deployment at 800 MHz.



(c) Uplink BLER in sparse deployment at 1800 MHz.

Figure 6.5: Uplink BLER with/without UL CoMP in the 3 deployment scenarios.

load on UL is quite low and inter-cell interference is mostly avoided by inter-cell interference cancellation (ICIC) schemes available in the network so it is expected that the observed gain from the trial will increase when UL network load increases. Additionally, trial results are based on daily average network counter values so higher CoMP gains are expected at the busy hour of the day, where UL load is usually higher than daily average. Secondly, CoMP sets are formed from intra-site cells only during the trial, i.e. interference due to cells from other BSs are not mitigated. Inter-BS interference is increased in dense deployment scenarios where CoMP gains can be maximized when inter-BS CoMP is enabled. Furthermore, instead of static clustering schemes, intelligent dynamic clustering algorithms can be deployed to design CoMP sets to dynamically adapt to spatio-temporal network/user profile changes and increase CoMP gain further.

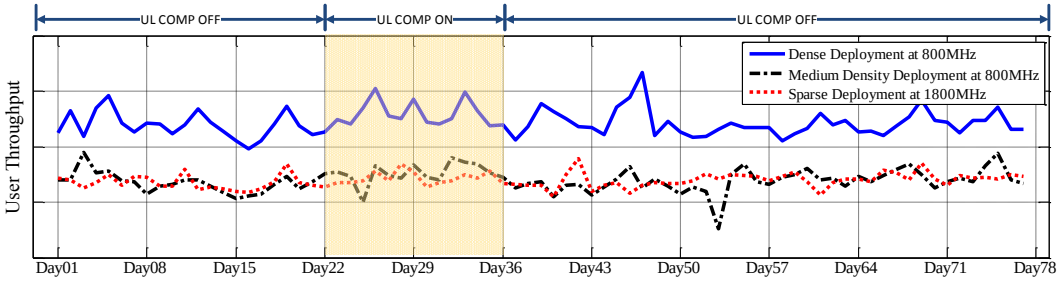


Figure 6.6: Average UL user throughput in 3 deployment scenarios.

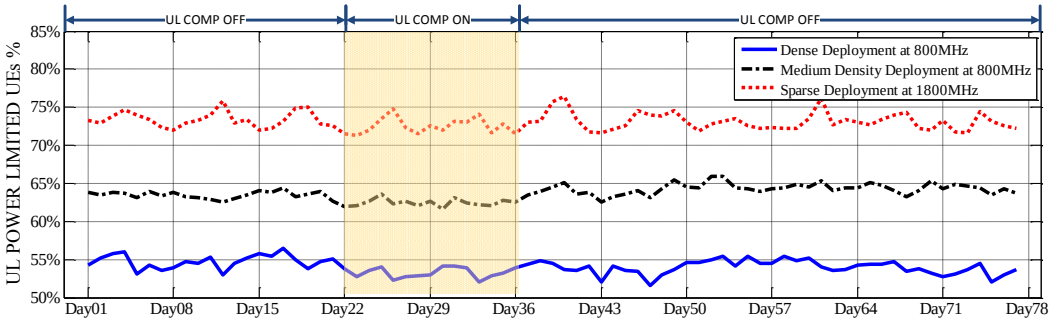


Figure 6.7: Power limited UEs in 3 deployment scenarios.

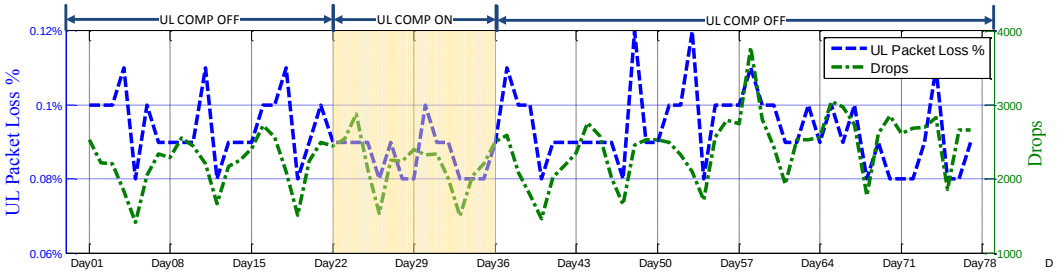


Figure 6.8: Uplink packet loss and drops in dense deployment at 800 MHz.

6.5 Evolution into 5G

In addition to the relatively simpler UL CoMP deployment presented in this chapter, we present the potential deployment of CoMP in future 5G networks where additional KPI improvements can be made. We discuss how CoMP can provide solutions for some of the key challenges of 5G like spectral efficiency, energy efficiency, backhaul bandwidth challenge and load balancing.

Cellular networks need to increase their capacity extensively to be able to meet ever increasing mobile data demand. Given the spectrum shortage to meet with this demand, 10 times more spectral efficiency is envisioned for 5G [165](i.e. from $2-3$b/s/Hz on LTE to 20b/s/Hz for 5G [87]). Network densification i.e. ultra dense small cell deployment, enhanced use of massive MIMO and millimeter-

wave bandwidth communications are proposed as key development areas to meet the challenging capacity targets of 5G [153, 171]. A much larger small cell density is envisioned for 5G when compared to existing 4G deployments, upto 10^3 SCs/km^2 [66, 132, 184]. Additionally, with the introduction of millimeter-wave band for 5G, the range of the cells will reduce to below 100m due to the significantly higher path loss in millimeter-wave band. A much larger deployment density will be required in millimeter-wave band deployment to be able to provide the required 5G applications. For the right design of ultra dense networks, the challenge of spatio-temporal fluctuations in the traffic demand, inter-cell interference, backhaul limitations, increased power consumption and energy efficiency concerns need to be handled carefully [184]. To mitigate severe inter-cell interference in such high density deployment, CoMP is envisioned as a key technology for 5G and expected to be part of 3GPP release 16 [67]. Emerging new architectures like CRAN enables the realization of CoMP in ultra-dense deployment scenario where baseband signal processing can be done centrally at a BBU pool [103, 132]. In millimeter-wave transmission, blockage is a key problem due to high path loss in NLOS scenario. CoMP is also a promising technology to reduce blockage with more then one BS serving to the user, reducing the user outage rate [112]. Ultra dense small cell deployment will need a different frequency allocation approach for network coordination. Different frequency allocation for small cell and macro layer will not give the sufficient spectral efficiency gains. New JT-CoMP schemes are required to achieve better spectral efficiency gains in multi-layer complex 5G network architectures [87]. JT-CoMP will play a key role on improving spectral efficiency in such dense small cell deployment scenario underlayed to macro network, where there will be extensive inter-cell interference presence due to overlapping coverage. An integrated approach will be required to adapt massive MIMO and JT-CoMP where required for maximal efficiency gains with minimal effort [87].

As demonstrated in Chapter 2, to further improve CoMP gains, CoMP clusters need to be designed efficiently and should be dynamically changing to adapt to changing network and user profiles [30]. Dynamic CoMP clustering algorithms for future networks are extensively studied in literature where cells within each cluster and the cluster size dynamically change, responding to dynamic user demand and network profile changes to maximize spectral efficiency [44, 88, 89]. Efficient deployment of CoMP will be able to reduce high deployment costs with less number of BSs required to provide the same QoS with improved spectral efficiency with CoMP [125].

One of the other challenges for future 5G networks is load balancing. In-

evitably, traffic load is usually not evenly distributed, traffic hotspots are formed in some areas at certain times. Intelligent load balancing algorithms are required to distribute load from congested cells to its relatively unloaded neighbours. Deployment of CoMP, especially JT-CoMP MU-MIMO can improve capacity of the congested BS, and intelligent algorithms can be deployed to design load aware CoMP clusters. In Chapter 3, we presented a load aware user-centric CoMP clustering algorithm and shown a significant reduction in unsatisfied users due to overload conditions and Chapter 4 presents a RAN-load aware network-centric clustering model to optimize CoMP clusters and achieve better load balancing.

Realization of JT-CoMP in 5G heavily depends on the requirement of high bandwidth, low latency backhaul as user data and CSI exchange are required between the coordinating cells. However, networks have a range of backhaul solutions and not all are ideal for JT-CoMP. Improved JT-CoMP schemes will need to take limited backhaul bandwidth availability into account for better gain. A number of backhaul aware clustering models are studied in literature [71, 115, 124,136,181] as presented in Chapter 2 and also we provide a backhaul-load aware multi-objective clustering solution in Chapter 5 to optimize RAN/backhaul load and spectral efficiency jointly. On the other hand, caching popular multimedia at the RAN is an increasingly popular concept to reduce user data sharing between the BSs and hence reducing the high backhaul bandwidth requirement for JT-CoMP [49, 56, 77, 105, 174].

Alongside with much needed spectral efficiency, energy efficiency is another key challenge for 5G networks for environmental reasons and to reduce energy costs [40, 62]. Enabling CoMP will also improve energy efficiency [144] as CoMP can reduce UE/BS power requirements, however it requires additional energy for additional signal processing and backhaul. Trade-off between energy efficiency and throughput gain need to be carefully designed for future CoMP deployment in 5G [90]. CoMP deployment can also improve energy efficiency by maximizing the number of sleeping cells when the user demand is low [92, 96]. Small cells are switched on, only when additional capacity is required in the network due to increasing demand.

In summary, CoMP will have an important role in 5G networks to mitigate inter-cell interference in densely deployed, multi-layer complex networks. An integrated approach to adapt massive MIMO and JT-CoMP is envisioned to maximize spectral efficiency. Additionally, CoMP is key for some of the other challenges for future 5G networks like energy efficiency, load balancing and high deployment costs. CoMP is envisioned to increase capacity for the congested cells, and load aware CoMP clusters can support load balancing further. Intel-

ligent CoMP clusters can dynamically adjust the active SCs based on demand maximizing the number of sleeping cells when not required. Wider availability of high bandwidth backhaul such as fiber is expected in future which will enable larger scale deployment of CoMP in 5G.

6.6 Conclusions

In this chapter, we first present the motivation for CoMP deployment from an operator perspective. Next, we analyze the practical limitations for CoMP deployment from an operator perspective and provided possible solutions. We then present performance results for an intra-site UL CoMP trial for 3 different deployment scenarios for a commercial LTE-A operator in the UK. We briefly introduce the CoMP scheme deployed in the trial and discuss improvement on various KPIs. Trial was conducted for a basic UL-CoMP scheme where only intra-site cells were allowed in the same CoMP set and only 2 cells could cooperate for the same UE at the same time. An average of 5.56% SINR improvement is observed on PUSCH. This improvement is reflected on the employed modulation types, where percentage usage of higher order 16QAM modulation on UL is increased by 5.38% on average. One of the main objectives of UL CoMP deployment is to improve user throughput especially at the cell edge. An average of 11.32% increase in user throughput is observed during this trial. Ratio of power limited UEs is reduced by 1.65% which shows the improvement observed especially at the cell edge where UE power is maximized and limited on extra power. Overall customer experience is improved with reduced radio related drops by 2.68% and UL packet loss rate is reduced by 11.12% on average. We further discussed the limitations of the UL CoMP scheme trialled and presented potential improvements for better CoMP gains for future networks. Finally, we presented the evolution of CoMP into 5G from an operator perspective. The need for CoMP to support some of the 5G challenging network objectives like energy efficiency, load balancing, spectral efficiency are presented.

Chapter 7

Conclusions and Future Research Directions

As mobile data traffic increases at a rapid rate, 1000 fold increase is envisioned for 5G beyond 2020 [129, 169]. 5G is expected to provide upto 10 Gb/s user experience and connect 100 billion devices globally [169]. Three major use cases have been identified for 5G to provide a diverse range of applications: Enhanced mobile broadband (eMBB) to support high bandwidth applications such as high definition video streaming, augmented reality/virtual reality applications, massive machine type communication (mMTC) to support massive deployments of IoT devices and ultra-reliable-low latency communications (URLLC) to support latency sensitive applications such as tactile internet. To be able to support applications with such diverse requirements and meet stringent 5G capacity demand, three key technologies emerge for 5G RF interface [153, 171]:

- Mm-wave communications
  Mobile networks heavily utilize spectrum below 6 GHz already. Although this band can provide wide area coverage with less deployment cost, this spectrum is heavily congested and wont be able to provide required capacity. Available spectrum above 24GHz (mm-waveband) is envisioned for 5G to be able to provide ultra-high bandwidth, very low latency applications [95]. Due to higher path loss characteristics of mm-waveband, cell range is reduced to below 100m and NLOS cause blockage for users. CoMP is proposed as a key solution to provide increased diversity and reduce user outage probability [112].

- Massive MIMO (M-MIMO)
  Especially with the utilisation of mm-wave bandwidth, high number of antenna arrays can be deployed at the BS for spatial multiplexing and/or

diversity gain. Multiple users can be served at the same time from the same physical resource improving the spectral efficiency and hence overall system throughput [153]. M-MIMO is a key technology proposed for 5G to improve spectral efficiency [67]. CoMP utilities the same M-MIMO technology where multiple antenna arrays located at different BSs are coordinated for multi-user transmission. CoMP can be seen as an enhancement to M-MIMO and it is already considered by 3GPP as part of 5G enhancements in Release 16 [67].

- Ultra-dense networks
  To meet high capacity demands, an ultra dense small cell deployment is envisioned for 5G up to 10^3 SCs/km^2 [184]. Such dense deployment will create severe inter-cell interference and hence advanced interference mitigation technologies are required to reduce interference and improve spectral efficiency. CoMP is a key technology to mitigate inter-cell interference and utilize as desired signal depending on the type of CoMP implementation.

As summarised above, CoMP is a key technology to support all three RF technologies proposed for 5G. However CoMP can not be realized in the whole network due to additional overheads i.e. additional backhaul requirement, complex precoding, synchronization requirements etc. CoMP will be realized in small clusters of cells to limit the additional overheads. To maximize CoMP gains, CoMP clusters need to be carefully formed. Moreover, these clusters will need to be able to adapt to spatio-temporal changes in the network and user profiles.

In this thesis, we provide an extensive survey of the currently available CoMP clustering solutions, provide two novel taxonomies based on self-organization and aimed objective. We critically review the strong and weak points of each solution. Part of this literature review is published in IEEE Communications Surveys & Tutorials journal [30]. Individual network objectives for CoMP clustering like spectral efficiency, energy efficiency and backhaul availability are well covered in literature where there are wide range of solutions studied however, there is only minimal work on CoMP clustering covering multiple objectives with a holistic network view. Moreover, load balancing is a key objective especially in densely populated areas where CoMP is likely to get deployed but there is very limited work on load balancing optimization as part of CoMP clustering. Our research fills the gap in literature providing a multi-objective CoMP clustering solution where backhaul availability, load balancing and spectral efficiency are jointly optimized.

7.1 Summary of Contributions

7.1.1 Load aware user-centric CoMP clustering

We first introduce a user-centric CoMP clustering model in Chapter 3 for CDSA architecture where user clusters are optimized for load balancing. We adjust CoMP cluster size for individual users to dynamically respond to congestion and create additional capacity by increasing cluster size when required in high load conditions. Additionally, we move users from highly loaded cells to lightly loaded cells to improve load balancing while keeping certain QoS. This has been the first attempt in literature to introduce load aware user-centric CoMP clusters. Part of this work is successfully published in IEEE Access journal [32]

7.1.2 Load aware network-centric CoMP clustering

User-centric clusters provide an upper bound performance for CoMP, however it comes with additional scheduling/precoding complexity and the solution lacks on scalability. Our second work in Chapter 4 provides a solution to this complexity and present a load aware, network-centric clustering solution to optimize load balancing and spectral efficiency jointly. A hybrid solution is proposed where user-centric clusters are employed within a wider network-centric cluster. We show that our algorithm is a low complexity, stable solution where it provides the best available clustering solution when there is one available. Part of this work is successfully published in IEEE Access journal [31].

7.1.3 Backhaul aware multi-objective CoMP clustering

One of the key dependencies for realization of CoMP is backhaul availability. We further extend our work from load aware network-centric clustering in Chapter 4 and provide a backhaul-load aware multi-objective clustering solution in Chapter 5 to optimize RAN load, backhaul load and spectral efficiency jointly. Two novel coalitional game theoretic clustering solutions are presented, one for optimizing the BS clusters and second one to transfer users between BS coalitions to optimize RAN and backhaul load jointly while keeping spectral efficiency at comparably high levels. We show that backhaul aware model provides an additional 21.9% average throughput when compared to same model without backhaul awareness in the case when all BSs are backhaul limited. Complexity and stability of the proposed solution is analyzed in detail and complexity against performance gain trade-off is adjusted with input parameters to adapt to any real network needs

with different backhaul availability and performance requirements. Part of this work is submitted for publication, currently under review.

7.1.4 Real Network CoMP Field Trial Results

We present CoMP field trial results in dense, medium and sparse deployment scenarios for a major network operator in the UK and analyze the performance impact in detail. We then discuss operational challenges for the real network implementation and provide further improvements we envision for future CoMP deployments for 5G. Part of this work is published as chapter in Book: Access, Fronthaul and Backhaul Networks for 5G and Beyond, Institution of Engineering and Technology (IET) [81].

7.2 Future Research Directions

Our research presented in this thesis filled in some of the gaps in literature to form a realistic multi-objective CoMP clustering solution as discussed above. However, there are still open questions to compliment our current work. We summarize these open research points as below:

- **Multi-objective clustering to include energy efficiency** Energy efficiency is one the prime objectives for future wireless network for both environmental and financial reasons as discussed in Chapter 2. Energy efficiency has been studied in literature as a primary objective for CoMP clustering but no other objectives are taken into account. Furthermore, most of the energy efficiency based studies focus on BS sleeping and maximize the number of sleeping cells while making sure SINR target is met for all active UEs. However capacity constraints on existing active BSs also need to be taken into account to make sure there is enough system capacity available before any BS is set to be switched off for energy efficiency. Furthermore, backhaul availability also need to be taken into account to make sure backhaul load on active BSs does not exceed required levels when any BS is set to sleep mode. In this context, our research could be exploited further to include energy efficiency objective alongside with backhaul/RAN load and spectral efficiency to make it a more comprehensive solution.

- **Multi-objective CoMP clustering for HetNet** Our research is conducted on one layer of small cells, however 5G and beyond future wireless networks will have multiple frequency and technology layers (4G,5G and

beyond) where each layer will have its own CoMP sets. The key metrics like load balancing, backhaul availability, energy efficiency are inter-related between all layers/technologies deployed in the same area. Hence an open research direction is to investigate on HetNet scenario as a whole while defining CoMP sets at each layer. For example, one layer can be selected as primary coverage layer and load balancing/energy efficiency metrics can be applied to make sure all users have sufficient coverage within the primary layer. Knowing that there is a primary layer satisfying coverage requirements, another "capacity layer" can be switched off based on capacity constraints only, rather than any coverage (SINR) concerns. Primary coverage layer and capacity layers can be dynamically adjusted based on network deployment settings to maximize energy efficiency while providing required coverage and capacity by a combination of layers/technologies available in the area. A holistic approach to include all layers in HetNet could provide additional benefits when the interaction between the layers are taken into account.

- **Proactive CoMP clustering** The state-of-the-art research on dynamic CoMP clustering in general have a reactive line of action i.e., CoMP clustering are designed/optimized with respect to current network conditions. For example load balancing targeted CoMP clustering will kick in when congestion is observed or diagnosed. Certain time is required to observe the current conditions, find optimum clustering with respect to the objective function and then trigger the appropriate clustering action. One future research direction is to look into proactive or predictive approach to CoMP clustering paradigm such that spatio-temporal future network state in terms of channel variation, mobility behavior and capacity requirements can be predicted beforehand. This is possible by inferring network-level intelligence from the massive amount of control, signalling, and contextual data known as Big Data as proposed in [80]. By leveraging a dexterous combination of advanced techniques of machine learning, statistics and optimization, Big Data can be tapped to enable and empower CoMP clustering algorithms to achieve true performance gains of CoMP. Endowed with predictive capabilities, CoMP clustering algorithms can track, learn and dynamically build user mobility and demand profiles as well as channel characteristics models to predict future user locations coupled with service requests and channel state information. This can lead to timely efficient CoMP clustering as well as can help to alleviate high backhaul requirements. Another advantage of exploiting Big Data in CoMP clustering is that, it can represent the global

state of the network which enables the global optimal CoMP clustering with respect to the defined objective functions such as energy efficiency, spectral efficiency or load balancing as opposed to relying only on the local information that may lead to only locally optimal CoMP clustering solutions. Proactive CoMP clustering paradigm could be investigated in 3 folds:

- **MDT based proactive CoMP clustering** Big Data can play crucial role in proactive selection of BSs for cluster formation. One of the vital sources of Big Data in mobile communications are minimization of drive tests (MDT) reports consisting of reference signal received power (RSRP) and other channel quality related metrics reported by the users to their serving BS [9,146]. The averaged RSRP values of the BSs, as reported by the UEs, can be compared to a threshold to determine which of these BSs should cooperate. Based on current MDT reports, future channel conditions can be predicted through conventional time-series forecasting methods. Therefore, instead of waiting for actual MDT reports, the predicted RSRP measurements can be fed to the cluster optimization algorithm that proactively adapts the cell clustering in CoMP perspective.
- **CoMP with Big Data Aided Mobility Prediction** Big Data aided mobility prediction can play important role in proactive CoMP clustering decisions. Mobility prediction utilizes person's mobility history, i.e. a series of locations and corresponding dwell times to predict this person's next location, as well as his/her dwell time in that location [54, 61, 65, 150, 162, 163, 166]. In this way, CoMP clustering algorithms can plan in advance the clustering decisions thereby meeting the strict latency requirements of 5G networks. Big Data as identified in [80] also contains handover reports which contain Cell IDs and corresponding timestamps whenever user is handed over to new cell. Several techniques such as mobility pattern matching using mobility database, periodicity and multi-class classification and bio-inspired approaches as presented in [54, 65, 166] can be used to predict user mobility behavior. Received signal strength indicator (RSSI) available in MDT reports can also be utilized to predict future location as has been done in [150, 162]. The identified future location of the users along with the corresponding time stamps can be fed to the CoMP dynamic clustering algorithms (both user-centric as well as network-centric) for computing optimal clusters. Mobility behavior of the users directly affects the

CoMP clustering decisions as CSI has small validity period for high speed users and clustering decisions needs to be performed frequently leading to high computational overheads. One solution can be that low speed or static users can be served by CoMP cluster BSs, however, high speed mobile users continue to be served by single BS. By utilizing RSSI and the cell sizes information embedded in Big Data and predicted future user locations, CoMP clustering algorithms can be executed beforehand leading to significant reduction in latency and bandwidth requirements.

– **CoMP with Big Data Aided User Profiling** Call Data Records (CDRs) are one of the key elements of the Big Data that can be harnessed from a cellular network. CDRs reflect mobile user's behavior and give out clues on how the users utilize the network resources. CDRs contain information about the voice calls and data usage pattern and are important markers of temporal-spatial capacity requirements across the deployed network [41, 59, 60]. CDRs can be utilized to profile the network usage behavior of the mobile users which in turn can be utilized for user-centric or behavior-centric CoMP clustering. By applying machine learning and statistical tools on CDRs, we can determine the capacity requirements of the users at different time periods and can utilize this profile information to cluster the CoMP enabled BSs to satisfy the expected QoS requirements of the users. Social media feeds are another element within Big Data that give helpful insights about the interaction of the users and expected temporal-spatial demand of network resources across the network. Among many online social networks, Twitter is one of the popular ways users share information and experience socially on the web. Twitter data can be mined through application program interface (Twitter APIs) wherein each time-stamped tweet contains number of useful information like location, number of re-tweets, number of favorites, message itself and hashtags. Twitter data can be utilized to estimate traffic demand as number of tweets is highly correlated with the number of people in confined places [38]. It can also be utilized to assess network's QoE from subscriber's perspective [140]. The social media feeds together with the CDRs can be taped to accurately model the user behavior and can be utilized to optimize user-centric CoMP clustering algorithms.

In a nutshell, Big Data driven predictive analytics predicting the future spatio-temporal state of the network accurately and using this

knowledge for proactive CoMP clustering is an open research field to understand the overheads and the additional gains when compared to existing reactive solutions.

Bibliography

[1] 3GPP. Requirements for further advancements for evolved universal terrestrial radio access (E-UTRA) (LTE-Advanced). TR 36.913 V9.0.0, Dec. 2009.

[2] 3GPP. Evolved universal terrestrial radio access (E-UTRA); Further advancements for E-UTRA physical layer aspects. TR 36.814, Mar. 2010.

[3] 3GPP. Evolved universal terrestrial radio access (E-UTRA); Base station (BS) radio transmission and reception. TS 36.104, Jun. 2011.

[4] 3GPP. Evolved universal terrestrial radio access (E-UTRA); Mobility enhancements in heterogeneous networks. TR 36.839, Jun. 2011.

[5] 3GPP. Evolved universal terrestrial radio access network (E-UTRAN); self-configuring and self-optimizing network (SON) use cases and solutions. TR 36.902 R9 v9.3.1, Mar. 2011.

[6] 3GPP. Coordinated multi-point operation for LTE physical layer aspects. TR 36.819 R11 v11.2.0, Sept 2013.

[7] 3GPP. Coordinated multi-point operation for LTE with non-ideal backhaul. TR 36.874 R12 v12.0.0, Dec. 2013.

[8] 3GPP. Study on LTE - high rate packet data in 3GPP2 (HRPD), Inter radio-access-technology (RAT) self-configuring and self-optimizin network (SON) use cases and solutions. TR 37.813 v12.1.0, Jun. 2013.

[9] 3GPP. Universal terrestrial radio access (UTRA) and evolved universal terrestrial radio access (E-UTRA); Radio measurement collection for minimization of drive tests (MDT); overall description; stage 2. TS 37.320 R12 v12.2.0, Sept. 2014.

[10] 3GPP. Study on next generation self-optimizing network (SON) for UTRAN and E-UTRAN. TR 37.822 v12.1.0, Mar. 2015.

[11] 3GPP. Study on further enhancements to Coordinated Multi-Point (CoMP) operation for LTE. TR 36.741 R14 v14.0.0, Mar. 2017.

[12] 3GPP. Study on SON for eCoMP. TR 36.742 R15 v2.0.0, Jun. 2017.

[13] 3GPP. Self-organizing networks (SON) policy network resource model (NRM) integration reference point (IRP); requirements. TS 28.627 v15.0.0, Jan. 2018.

[14] 3GPP. Telecommunication management; self-configuration of network elements; concepts and requirements. TS 32.501 r15 v15.0.0, Jun. 2018.

[15] 3GPP. Telecommunication management; self-organizing networks (SON); self-healing concepts and requirements. TS 32.541 R15 v15.0.0, Oct. 2018.

[16] 3GPP. Self-organizing networks (SON) for 5G networks-release 16. TS 28.313 v0.2.0, Nov. 2019.

[17] 3GPP. Telecommunication management;study on the self-organizing networks (SON) for 5G networks (release 16). TR 28.861 v16.0.0, Dec. 2019.

[18] Mst Rubina Aktar, Md Farhad Hossain, and Md Al-Hasan. Dynamic clustering approach for interference cancellation in downlink C-RAN. In *International conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, pages 1–4. IEEE, 2018.

[19] Md Shipon Ali, Ekram Hossain, Arafat Al-Dweik, and Dong In Kim. Downlink power allocation for CoMP-NOMA in multi-cell networks. *IEEE Transactions on Communications*, 66(9):3982–3998, 2018.

[20] Saley Saidou Ali and Navrati Saxena. A novel static clustering approach for CoMP. In *7th International Conference on Computing and Convergence Technology (ICCCT)*, pages 757–762. IEEE, 2012.

[21] Isiaka Ajewale Alimi, António Luís Teixeira, and Paulo Pereira Monteiro. Toward an efficient C-RAN optical fronthaul for the future networks: A tutorial on technologies, requirements, challenges, and solutions. *IEEE Communications Surveys & Tutorials*, 20(1):708–769, 2017.

[22] Osianoh Glenn Aliu, Ali Imran, Muhammad Ali Imran, and Barry Evans. A survey of self organisation in future cellular networks. *IEEE Communications Surveys & Tutorials*, 15(1):336–361, 2013.

[23] NGMN Alliance. Small cell backhaul requirements. *White Paper, June*, 2012.

[24] Talal Alsedairy, Yinan Qi, Ali Imran, Muhammad Ali Imran, and Barry Evans. Self organising cloud cells: A resource efficient network densification strategy. *Transactions on Emerging Telecommunications Technologies*, 26(8):1096–1107, 2015.

[25] Jeffrey G Andrews, Sarabjot Singh, Qiaoyang Ye, Xingqin Lin, and Harpreet S Dhillon. An overview of load balancing in HetNets: Old myths and open problems. *IEEE Wireless Communications*, 21(2):18–25, 2014.

[26] Krzysztof R Apt and Tadeusz Radzik. Stable partitions in coalitional games. *arXiv preprint cs/0605132*, 2006.

[27] Krzysztof R Apt and Andreas Witzel. A generic approach to coalition formation. *International Game Theory Review*, 11(03):347–367, 2009.

[28] Paolo Baracca, Federico Boccardi, and Volker Braun. A dynamic joint clustering scheduling algorithm for downlink CoMP systems with limited CSI. In *International Symposium on Wireless Communication Systems (ISWCS)*, pages 830–834. IEEE, 2012.

[29] Alan Barbieri, Peter Gaal, Stefan Geirhofer, Tingfang Ji, Durga Malladi, Yongbin Wei, and F. Xue. Coordinated downlink multi-point communications in heterogeneous cellular networks. In *Information Theory and Applications Workshop (ITA)*, pages 7–16. IEEE, 2012.

[30] Selcuk Bassoy, Hasan Farooq, Muhammad A Imran, and Ali Imran. Coordinated multi-point clustering schemes: A survey. *IEEE Communications Surveys & Tutorials*, 2017.

[31] Selcuk Bassoy, Muhammad Ali Imran, Shufan Yang, and Rahim Tafazolli. A load-aware clustering model for coordinated transmission in future wireless networks. *IEEE Access*, 7:92693–92708, 2019.

[32] Selcuk Bassoy, Mona J Jaber, Muhammad A Imran, and Pei Xiao. Load aware self-organising user-centric dynamic CoMP clustering for 5G networks. *IEEE Access*, pages 2895–2906, 2016.

[33] Anthony Beylerian and Tomoaki Ohtsuki. Coordinated non-orthogonal multiple access (CO-NOMA). In *Global Communications Conference (GLOBECOM)*. IEEE, 2016.

[34] Anthony Beylerian and Tomoaki Ohtsuki. Multi-point fairness in resource allocation for C-RAN downlink CoMP transmission. *EURASIP Journal on Wireless Communications and Networking*, 2016(1):12, 2016.

[35] Thorsten Biermann, Luca Scalia, Changsoon Choi, Holger Karl, and Wolfgang Kellerer. Improving CoMP cluster feasibility by dynamic serving base station reassignment. In *22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 1325–1330. IEEE, 2011.

[36] Thorsten Biermann, Luca Scalia, Changsoon Choi, Wolfgang Kellerer, and Holger Karl. How backhaul networks influence the feasibility of coordinated multipoint in cellular networks. *IEEE Communications Magazine*, 51(8):168–176, 2013.

[37] Emil Bjornson, Eduard Axel Jorswieck, Mérouane Debbah, and Bjorn Ottersten. Multiobjective signal processing optimization: The way to balance conflicting metrics in 5G systems. *IEEE Signal Processing Magazine*, 31(6):14–23, 2014.

[38] Federico Botta, Helen Susannah Moat, and Tobias Preis. Quantifying crowd size with mobile phone and twitter data. *Royal Society Open Science*, 2(5):150162, 2015.

[39] Stefan Brueck, Lu Zhao, Jochen Giese, and M Awais Amin. Centralized scheduling for joint transmission coordinated multi-point in LTE-Advanced. In *International ITG Workshop on Smart Antennas (WSA)*, pages 177–184. IEEE, 2010.

[40] Stefano Buzzi, I Chih-Lin, Thierry E Klein, H Vincent Poor, Chenyang Yang, and Alessio Zappone. A survey of energy-efficient techniques for 5G networks and challenges ahead. *IEEE Journal on Selected Areas in Communications*, 34(4):697–709, 2016.

[41] Francesco Calabrese, Massimo Colonna, Piero Lovisolo, Dario Parata, and Carlo Ratti. Real-time urban monitoring using cell phones: A case study in Rome. *IEEE Transactions on Intelligent Transportation Systems*, 12(1):141–151, March 2011.

[42] Dongxu Cao, Sheng Zhou, Chao Zhang, and Zhisheng Niu. Energy saving performance comparison of coordinated multi-point transmission and wireless relaying. In *Global Telecommunications Conference (GLOBECOM)*, pages 1–5. IEEE, 2010.

[43] Lei Cao, Xuanguang Hu, Meng Zhang, Xianling Wang, and Xin Zhang. Interactive CoMP with user-centric clustering based on load balancing in 5G dense networks. In *International Conference on Communications Workshops (ICC Workshops)*, pages 1–6. IEEE, 2018.

[44] Veronique Capdevielle, Pantelis Monogioudis, Carl Weaver, Jean-Michel Pugeat, and Steve Myers. Learning based spectral clustering for LTE downlink CoMP systems. In *14th Annual Consumer Communications & Networking Conference (CCNC)*, pages 791–795. IEEE, 2017.

[45] Jou-Yi Chen, Kuang-Hao Stanley Liu, and Szu-Lin Su. Performance of network-centric clustering for coordinated joint transmission with irregular cluster topology. In *88th Vehicular Technology Conference (VTC-Fall)*, pages 1–5. IEEE, 2018.

[46] Min Chen and Adlen Ksentini. Cache in the air: Exploiting content caching and delivery techniques for 5G systems. *IEEE Communications Magazine*, page 132, 2014.

[47] Shuyi Chen, Tianyu Zhao, Hsiao-Hwa Chen, Zhiping Lu, and Weixiao Meng. Performance analysis of downlink coordinated multipoint joint transmission in ultra-dense networks. *IEEE Network*, 31(5):106–114, 2017.

[48] Shuyi Chen, Tianyu Zhao, Hsiao-Hwa Chen, and Weixiao Meng. Downlink coordinated multi-point transmission in ultra-dense networks with mobile edge computing. *IEEE Network*, 33(2):152–159, 2018.

[49] Siyun Chen, Ye Wang, Jia Yu, Niannian Wang, and Yan Yan. User association in cache-enabled ultra dense network with JT CoMP. In *IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pages 964–968. IEEE, 2018.

[50] Yi-Han Chiang, Wanjiun Liao, and Yusheng Ji. RELISH: Green multicell clustering in heterogeneous networks with shareable caching. In *Global Communications Conference (GLOBECOM)*, pages 1–7. IEEE, 2018.

[51] I. Chih-Lin, Corbett Rowell, Shuangfeng Han, Zhikun Xu, Gang Li, and Zhengang Pan. Toward green and soft: A 5G perspective. *Communications Magazine, IEEE*, 52(2):66–73, 2014.

[52] Gencer Cili, Halim Yanikomeroglu, and F. Richard Yu. Cell switch off technique combined with coordinated multi-point (CoMP) transmission for

energy efficiency in beyond-LTE cellular networks. In *International Conference on Communications (ICC)*, pages 5931–5935. IEEE, 2012.

[53] Cisco. Cisco visual networking index: Forecast and trends, 2017-2022. *White Paper*, Nov. 2018.

[54] Mehammed Daoui, Abdelaziz M'zoughi, Mustapha Lalam, Malika Belkadi, and Rachida Aoudjit. Mobility prediction based on an ant system. *Computer communications*, 31(14):3090–3097, 2008.

[55] Alexei Davydov, Gregory Morozov, Ilya Bolotin, and Apostolos Papathanassiou. Evaluation of joint transmission CoMP in C-RAN based LTE-A HetNets with large coordination areas. In *Globecom Workshops*, pages 801–806. IEEE, 2013.

[56] Matha Deghel, Ejder Bastug, Mohamad Assaad, and Merouane Debbah. On the benefits of edge caching for MIMO interference alignment. In *16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 655–659. IEEE, 2015.

[57] Melike Erol-Kantarci and Sukhmani Sukhmani. Caching and computing at the edge for mobile augmented reality and virtual reality (AR/VR) in 5G. In *Ad Hoc Networks*, pages 169–177. Springer, 2018.

[58] Si Feng, Wei Feng, Hongliang Mao, and Jianhua Lu. Overlapped clustering for CoMP transmissions in massively dense wireless networks. In *International Conference on Communication Systems (ICCS)*, pages 308–312. IEEE, 2014.

[59] Michal Ficek and Lukas Kencl. Inter-call mobility model: A spatio-temporal refinement of call data records using a Gaussian mixture model. In *In Proc. of IEEE INFOCOM*, pages 469–477, Mar. 2012.

[60] Vanessa Frias-Martinez, Cristina Soguero, and Enrique Frias-Martinez. Estimation of urban commuting patterns using cellphone network data. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, pages 9–16. ACM, 2012.

[61] Sébastien Gambs, Marc-Olivier Killijian, and Miguel Núñez del Prado Cortez. Next place prediction using mobility markov chains. In *Proceedings of the First Workshop on Measurement, Privacy, and Mobility*, pages 3:1–3:6. ACM, 2012.

[62] Pimmy Gandotra, Rakesh Kumar Jha, and Sanjeev Jain. Green communication in next generation cellular networks: A survey. *IEEE Access*, 5:11727–11758, 2017.

[63] Ian Dexter Garcia, Naoki Kusashima, Kei Sakaguchi, Kiyomichi Araki, Shoji Kaneko, and Yoji Kishi. Impact of base station cooperation on cell planning. *EURASIP Journal on Wireless Communications and Networking*, 2010(1):5, 2010.

[64] Virgile Garcia, Yiqing Zhou, and Jinglin Shi. Coordinated multipoint transmission in dense cellular networks with user-centric adaptive clustering. *IEEE Transactions on Wireless Communications*, 13(8):4297–4308, Aug. 2014.

[65] Huaining Ge, Xiangming Wen, Wei Zheng, Zhaoming Lu, and Bo Wang. A history-based handover prediction for LTE systems. In *International Symposium on Computer Network and Multimedia Technology (CNMT)*, pages 1–4. IEEE, 2009.

[66] Xiaohu Ge, Song Tu, Guoqiang Mao, Cheng-Xiang Wang, and Tao Han. 5g ultra-dense cellular networks. *IEEE Wireless Communications*, 23(1):72–79, 2016.

[67] Amitabha Ghosh, Andreas Maeder, Matthew Baker, and Devaki Chandramouli. 5G evolution: A view on 5G cellular technology beyond 3GPP release 15. *IEEE Access*, 7:127639–127651, 2019.

[68] Anastasios Giovanidis, Jonatan Krolikowski, and Stefan Brueck. A 0–1 program to form minimum cost clusters in the downlink of cooperating base stations. In *Wireless Communications and Networking Conference (WCNC)*, pages 940–945. IEEE, 2012.

[69] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.2, 01 2020. [Available online, `http://cvxr.com/cvx/`].

[70] Francesco Guidolin, Leonardo Badia, and Michele Zorzi. A distributed clustering algorithm for coordinated multipoint in LTE networks. *IEEE Wireless Communications Letters*, 3(5):517–520, Oct 2014.

[71] Vu Nguyen Ha, Long Bao Le, et al. Coordinated multipoint transmission design for cloud-RANs with limited fronthaul capacity constraints. *IEEE Transactions on Vehicular Technology*, 65(9):7432–7447, 2015.

[72] Shima Haddadi and Abdorasoul Ghasemi. Coordinated multi-point joint transmission evaluation in heterogenous cloud radio access networks. In *2017 Iranian Conference on Electrical Engineering (ICEE)*, pages 1938–1943. IEEE, 2017.

[73] Ahmed M. Hamza and Jon W. Mark. A clustering scheme based on timing requirements in coordinated base-stations cooperative communications. In *Wireless Communications and Networking Conference (WCNC)*, pages 3764–3769. IEEE, 2013.

[74] Shengqian Han, Chenyang Yang, Gang Wang, and Ming Lei. On the energy efficiency of base station sleeping with multicell cooperative transmission. In *22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 1536–1540. IEEE, 2011.

[75] Zhu Han, Dusit Niyato, Walid Saad, Tamer Başar, and Are Hjørungnes. *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*. Cambridge University Press, 2012.

[76] Ekram Hossain, Dong In Kim, and Vijay K Bhargava. *Cooperative cellular wireless networks*. Cambridge University Press, 2011.

[77] Ronghui Hou, Jingfeng Cai, and King-Shan Lui. Distributed cache-aware CoMP transmission scheme in dense small cell networks with limited backhaul. *Computer Communications*, 138:11–19, 2019.

[78] Fan Huang, Yafeng Wang, Jian Geng, Mei Wu, and Dacheng Yang. Clustering approach in coordinated multi-point transmission/reception system. In *72nd Vehicular Technology Conference Fall (VTC 2010-Fall)*, pages 1–5. IEEE, 2010.

[79] Ali Imran, Elias Yaacoub, Muhammad Ali Imran, and Rahim Tafazolli. Distributed load balancing through self organisation of cell size in cellular systems. In *23rd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 1114–1119. IEEE, 2012.

[80] Ali Imran, Ahmed Zoha, and Adnan Abu-Dayya. Challenges in 5G: How to empower SON with Big Data for enabling 5G. *IEEE Network*, 28(6):27–33, 2014.

[81] Muhammad Ali Imran, Syed Ali Raza Zaidi, and Muhammad Zeeshan Shakir. *Access, Fronthaul and Backhaul Networks for 5G and Beyond*. Institution of Engineering and Technology, 2017.

[82] Ralf Irmer, Heinz Droste, Patrick Marsch, Michael Grieger, Gerhard Fettweis, Stefan Brueck, H.-P. Mayer, Lars Thiele, and Volker Jungnickel. Coordinated multipoint: Concepts, performance, and field trial results. *Communications Magazine, IEEE*, 49:102–111, 2011.

[83] ITU-R M.2135.1. Guidelines for evaluation of radio interface technologies for IMT-Advanced. *ITU-R, Geneva, Switzerland*, Dec. 2009.

[84] Mona Jaber, Muhammad Imran, Rahim Tafazolli, and Anvar Tukmanov. An adaptive backhaul-aware cell range extension approach. In *IEEE International Conference on Communication Workshop (ICCW)*, pages 74–79. IEEE, 2015.

[85] Mona Jaber, Muhammad Ali Imran, Rahim Tafazolli, and Anvar Tukmanov. 5G backhaul challenges and emerging research directions: A survey. *IEEE Access*, 4:1743–1766, 2016.

[86] Nihar Jindal and Angel Lozano. A unified treatment of optimum pilot overhead in multipath fading channels. *IEEE Transactions on Communications*, 58(10):2939–2948, 2010.

[87] Volker Jungnickel, Konstantinos Manolakis, Wolfgang Zirwas, Berthold Panzner, Volker Braun, Moritz Lossow, Mikael Sternad, Rikke Apelfroojd, and Tommy Svensson. The role of small cells, coordinated multipoint, and massive MIMO in 5G. *IEEE Communications Magazine*, 52(5):44–51, 2014.

[88] Michail Karavolos, Vasileios I Tatsis, Dimitrios N Skoutas, Nikolaos Nomikos, Demosthenes Vouyioukas, and Charalabos Skianis. A dynamic hybrid clustering scheme for LTE-A networks employing CoMP-DPS. In *22nd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pages 1–5. IEEE, 2017.

[89] Raja Karmakar, Samiran Chattopadhyay, and Sandip Chakraborty. A learning-based dynamic clustering for coordinated multi-point (CoMP) operation with carrier aggregation in LTE-advanced. In *10th International Conference on Communication Systems & Networks (COMSNETS)*, pages 283–290. IEEE, 2018.

[90] Efstathios Katranaras, Muhammad Ali Imran, and Mehrdad Dianati. Energy-aware clustering for multi-cell joint transmission in LTE networks. In *International Conference on Communications Workshops (ICC)*, pages 419–424. IEEE, 2013.

[91] Paulo Valente Klaine, Muhammad Ali Imran, Oluwakayode Onireti, and Richard Demo Souza. A survey of machine learning techniques applied to self-organizing cellular networks. *IEEE Communications Surveys & Tutorials*, 19(4):2392–2431, 2017.

[92] Narjes Lassoued, Noureddine Boujnah, and Ridha Bouallegue. Enhancing energy efficiency in wireless heterogeneous networks using coordinated multipoint and eNB parameters tuning. In *15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, pages 1650–1655. IEEE, 2019.

[93] Daewon Lee, Hanbyul Seo, Bruno Clerckx, Eric Hardouin, David Mazzarese, Satoshi Nagata, and Krishna Sayana. Coordinated multipoint transmission and reception in LTE-Advanced: Deployment scenarios and operational challenges. *Communications Magazine, IEEE*, 50(2):148–155, 2012.

[94] Juho Lee, Younsun Kim, Hyojin Lee, Boon Loong Ng, David Mazzarese, Jianghua Liu, Weimin Xiao, and Yongxing Zhou. Coordinated multipoint transmission and reception in LTE-Advanced systems. *Communications Magazine, IEEE*, pages 44–50, 2012.

[95] Juho Lee, Erika Tejedor, Karri Ranta-aho, Hu Wang, Kyung-Tak Lee, Eliane Semaan, Eiman Mohyeldin, Juyeon Song, Christian Bergljung, and Sangyeob Jung. Spectrum for 5G: Global status, challenges, and enabling technologies. *IEEE Communications Magazine*, 56(3):12–18, 2018.

[96] Bingxue Leng, Xinyu Gu, Lei Zhang, and Lin Zhang. Energy efficiency optimization in 3D dense urban scenarios with coordinate multi-point transmission and sleep strategy. In *29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 1–7. IEEE, 2018.

[97] Yann Léost, Moussa Abdi, Robert Richter, and Michael Jeschke. Interference rejection combining in LTE networks. *Bell Labs Technical Journal*, 17(1):25–49, 2012.

[98] Chang Li, Jun Zhang, Martin Haenggi, and Khaled Ben Letaief. User-centric intercell interference nulling for downlink small cell networks. *IEEE Transactions on Communications*, pages 1–1, 2015.

[99] Hongbin Li, Hui Tian, Cheng Qin, and Yushan Pei. A novel distributed cluster combination method for CoMP in LTE-A system. In *15th International*

Symposium on Wireless Personal Multimedia Communications (WPMC), pages 614–618. IEEE, 2012.

[100] Liying Li, Chungang Yang, Mbazingwa E Mkiramweni, and Lei Pang. Intelligent scheduling and power control for multimedia transmission in 5G CoMP systems: A dynamic bargaining game. *IEEE Journal on Selected Areas in Communications*, 2019.

[101] Qian Clara Li, Huaning Niu, Apostolos Tolis Papathanassiou, and Geng Wu. 5G network capacity: Key elements and technologies. *IEEE Vehicular Technology Magazine*, 9:71–78, March 2014.

[102] Yun Li, Yafei Ma, Yong Wang, and Weiliang Zhao. Base station sleeping with dynamical clustering strategy of CoMP in LTE-Advanced. In *in Proc. IEEE International Conference on Green Computing and Communications*, pages 157–162, August 2013.

[103] Yuzhou Li, Yu Zhang, Kai Luo, Tao Jiang, Zan Li, and Wei Peng. Ultra-dense HetNets meet big data: Green frameworks, techniques, and approaches. *IEEE Communications Magazine*, 56(6):56–63, 2018.

[104] Wei-Shun Liao, Mirza Golam Kibria, Gabriel Porto Villardi, Ou Zhao, Kentaro Ishizu, and Fumihide Kojima. Coordinated multi-point downlink transmission for dense small cell networks. *IEEE Transactions on Vehicular Technology*, 68(1):431–441, 2018.

[105] An Liu and Vincent Lau. Exploiting base station caching in MIMO cellular networks: Opportunistic cooperation for video streaming. *IEEE Transactions on Signal Processing*, 63(1):57–69, January 2015.

[106] Chun-Hung Liu and Po-Chia Chen. Load-aware coordinated multipoint joint transmission in dense heterogeneous networks: Downlink coverage and throughput limits. In *International Conference on Communications (ICC)*, pages 1–7. IEEE, 2017.

[107] Chun-Hung Liu and Di-Chun Liang. Heterogeneous networks with power-domain NOMA: Coverage, throughput, and power allocation analysis. *IEEE Transactions on Wireless Communications*, 17(5):3524–3539, 2018.

[108] Jingxin Liu and Dongming Wang. An improved dynamic clustering algorithm for multi-user distributed antenna system. In *International Conference on Wireless Communications & Signal Processing, WCSP*, pages 1–5. IEEE, 2009.

[109] Ling Liu, Yiqing Zhou, Virgile Garcia, Lin Tian, and Jinglin Shi. Load aware joint CoMP clustering and inter-cell resource scheduling in heterogeneous ultra dense cellular networks. *IEEE Transactions on Vehicular Technology*, 67(3):2741–2755, 2018.

[110] Mingqian Liu, Bingbing Li, and Zhao Tong. Cell clustering and resource allocation scheme for CoMP SU-MIMO. *Wireless Personal Communications*, 69(4):1399–1411, April 2013.

[111] David López-Pérez, Ming Ding, Holger Claussen, and Amir H Jafari. Towards 1 Gbps/UE in cellular systems: Understanding ultra-dense small cell deployments. *Communications Surveys & Tutorials, IEEE*, 17(4):2078–2101, 2015.

[112] George R MacCartney and Theodore S Rappaport. Millimeter-wave base station diversity for 5G coordinated multipoint (CoMP) applications. *IEEE Transactions on Wireless Communications*, 18(7):3395–3410, 2019.

[113] Vincenzo Mancuso and Sara Alouf. Reducing costs and pollution in cellular networks. *Communications Magazine, IEEE*, 49(8):63–71, 2011.

[114] Andrea Marotta, Dajana Cassioli, Cristian Antonelli, Koteswararao Kondepu, and Luca Valcarenghi. Network solutions for CoMP coordinated scheduling. *IEEE Access*, 7:176624–176633, 2019.

[115] Andrea Marotta, Francesco Giannone, Koteswararao Kondepu, Dajana Cassioli, Cristian Antonelli, Luca Valcarenghi, and Piero Castoldi. Reducing CoMP control message delay in PON backhauled 5G networks. In *European Wireless 2017; 23th European Wireless Conference*, pages 1–5. VDE, 2017.

[116] Andrea Marotta, K Kondepu, F Giannone, S Doddikrinda, Dajana Cassioli, Cristian Antonelli, L Valcarenghi, and P Castoldi. Performance evaluation of CoMP coordinated scheduling over different backhaul infrastructures: A real use case scenario. In *IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, pages 1–5. IEEE, 2016.

[117] Patrick Marsch and Gerhard Fettweis. Static clustering for cooperative multi-point (CoMP) in mobile communications. In *Proceedings of IEEE International Conference on Communications*, pages 1–6. IEEE, 2011.

[118] Zoltán Mayer, Jingya Li, Agisilaos Papadogiannis, and Tommy Svensson. On the impact of backhaul channel reliability on cooperative wireless networks. In *International Conference on Communications (ICC)*, pages 5284–5289. IEEE, 2013.

[119] China Mobile. C-RAN: The road towards green RAN. *White Paper*, 2011.

[120] Abdelrahim Mohamed, Oluwakayode Onireti, Muhammad Imran, Ali Imran, and Rahim Tafazolli. Control-data separation architecture for cellular radio access networks: A survey and outlook. *IEEE Communications Surveys & Tutorials*, 2015.

[121] Jung-Min Moon and Dong-Ho Cho. Formation of cooperative cluster for coordinated transmission in multi-cell wireless networks. In *Consumer Communications and Networking Conference (CCNC)*, pages 528–533. IEEE, 2013.

[122] Hafiz Atta Ul Mustafa, Muhammad Ali Imran, Muhammad Zeeshan Shakir, Ali Imran, and Rahim Tafazolli. Separation framework: An enabler for cooperative and D2D communication for future 5G networks. *IEEE Communications Surveys & Tutorials*, 18(1):419–445, 2016.

[123] Omer Narmanlioglu and Engin Zeydan. Mobility-aware cell clustering mechanism for self-organizing networks. *IEEE Access*, 6:65405–65417, 2018.

[124] Kien-Giang Nguyen, Quang-Doanh Vu, Markku Juntti, and Le-Nam Tran. Globally optimal beamforming design for downlink CoMP transmission with limited backhaul capacity. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3649–3653. IEEE, 2017.

[125] Zhisheng Niu, Sheng Zhou, Yao Hua, Qian Zhang, and Dongxu Cao. Energy-aware network planning for wireless cellular system with inter-cell cooperation. *IEEE Transactions on Wireless Communications*, 11(4):1412–1423, April 2012.

[126] 3GPP TSG-RAN WG1 Meeting no:64. R1-110651, Deployment and backhaul constraints for CoMP evaluations. Technical report, 2011.

[127] 3GPP TSG-RAN WG1 Meeting no:64. R1-111174, backhaul modelling for CoMP. Technical report, 2011.

[128] 3GPP TSG-RAN WG1 Meeting no:65. R1-111628, Phase 1 CoMP simulation evaluation results and analysis for full buffer. Technical report, 2011.

[129] Nokia Networks. Looking ahead to 5G. *White Paper*, July. 2014.

[130] Edward J Oughton, Zoraida Frias, Sietse van der Gaast, and Rudolf van der Berg. Assessing the capacity, coverage and cost of 5G infrastructure strategies: Analysis of the Netherlands. *Telematics and Informatics*, 37:50–69, 2019.

[131] P. Marsch and G. Fettweis. *Coordinated Multi-Point in Mobile Communications.From Theory to Practice*. Cambridge University Press, 1st edition edition, 2011.

[132] Cunhua Pan, Maged Elkashlan, Jiangzhou Wang, Jinhong Yuan, and Lajos Hanzo. User-centric C-RAN architecture for ultra-dense 5G networks: Challenges and methodologies. *IEEE Communications Magazine*, 56(6):14–20, 2018.

[133] Agisilaos Papadogiannis and George C. Alexandropoulos. The value of dynamic clustering of base stations for future wireless networks. In *International Conference on Fuzzy Systems (FUZZ)*, pages 1–6. IEEE, 2010.

[134] Agisilaos Papadogiannis, David Gesbert, and Eric Hardouin. A dynamic clustering approach in wireless networks with multi-cell cooperative processing. In *International Conference on Communications, 2008. ICC'08*, pages 4033–4037. IEEE, 2008.

[135] Javier Pastor-Perez, Felip Riera-Palou, and Guillem Femenias. Clustering and subband allocation for CoMP-based MIMO-OFDMA networks with soft frequency reuse. In *European Wireless 2017; 23th European Wireless Conference*, pages 1–8. VDE, 2017.

[136] Javier Atanasio Pastor-Pérez, Felip Riera-Palou, and Guillem Femenias. Analytical network-wide optimization of CoMP-aided MIMO-OFDMA irregular networks with frequency reuse: A multiobjective approach. *IEEE Transactions on Communications*, 67(3):2552–2568, 2018.

[137] Javier Atanasio Pastor Pérez, Felip Riera-Palou, and Guillem Femenias. Multi-objective optimization of coordinated multipoint-aided MIMO-OFDMA systems with frequency reuse. *IEEE Access*, 5:15448–15467, 2017.

[138] Venkatadheeraj Pichapati and Parul Gupta. Practical considerations in cluster design for coordinated multipoint (CoMP) systems. In *International Conference on Communications (ICC)*, pages 5860–5865. IEEE, 2013.

[139] Hakimeh Purmehdi, Robert C. Elliott, Witold A. Krzymień, and Jordan Melzer. Rotating cluster mechanism for coordinated heterogeneous MIMO cellular networks. *EURASIP Journal on Wireless Communications and Networking*, 2018(1):59, Mar. 2018.

[140] Tongqing Qiu, Junlan Feng, Zihui Ge, Jia Wang, Jun Xu, and Jennifer Yates. Listen to me if you can: Tracking user experience of mobile network on social media. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*, pages 288–293. ACM, 2010.

[141] Qualcomm. How CoMP can extend 5G NR to high capacity and ultra-reliable communications. 2018. Available online, "https://www.qualcomm.com/documents/how-comp-can-extend-5g-nr-high-capacity-and-ultra-reliable-communications".

[142] Buvaneswari A Ramanan, Lawrence M Drabeck, Mark Haner, Nachi Nithi, Thierry E Klein, and Chitra Sawkar. Cacheability analysis of HTTP traffic in an operational LTE network. In *Wireless Telecommunications Symposium (WTS), 2013*, pages 1–8. IEEE, 2013.

[143] Sean A. Ramprashad, Giuseppe Caire, and Haralabos C. Papadopoulos. A joint scheduling and cell clustering scheme for MU-MIMO downlink with limited coordination. In *International Conference on Communications (ICC)*, pages 1–6. IEEE, 2010.

[144] Jaya B. Rao and Abraham O. Fapojuwo. A survey of energy efficient resource management techniques for multicell cellular networks. *IEEE Communications Surveys & Tutorials*, 16(1):154–180, 2014.

[145] Ana Santos Rodríguez and Ana García Armada. Analysis of the cluster size in coordinated multipoint transmission. In *19th Signal Processing and Communications Applications Conference*, pages 1197–1200. IEEE, 2011.

[146] H. Sanneck S. Hämäläinen and C. Sartori. *LTE Self-Organising Networks (SON): Network Management Automation for Operational Efficiency*. John Wiley-Sons, 2012.

[147] Walid Saad, Zhu Han, Merouane Debbah, and Are Hjorungnes. A distributed coalition formation framework for fair user cooperation in wireless

networks. *IEEE Transactions on Wireless Communications*, 8(9):4580–4593, September 2009.

[148] Walid Saad, Zhu Han, Mérouane Debbah, Are Hjorungnes, and Tamer Basar. Coalitional game theory for communication networks. *Signal Processing Magazine, IEEE*, 26(5):77–97, 2009.

[149] Walid Saad, Zhu Han, Rong Zheng, Merouane Debbah, and H Vincent Poor. A college admissions game for uplink user association in wireless small cell networks. In *INFOCOM, 2014 Proceedings IEEE*, pages 1096–1104. IEEE, 2014.

[150] Ali Safa Sadiq, K Abu Bakar, Kayhan Zrar Ghafoor, and Alberto J Gonzalez. Mobility and signal strength-aware handover decision in mobile IPv6 based wireless LAN. In *International MultiConference of Engineers and Computer Scientists, Hong Kong*, 2011.

[151] Dragan Samardzija and Howard Huang. Determining backhaul bandwidth requirements for network MIMO. In *Proc. 17th European Signal Processing Conference (EUSIPCO)*, 2009.

[152] Mamoru Sawahashi, Yoshihisa Kishiyama, Akihito Morimoto, Daisuke Nishikawa, and Motohiro Tanno. Coordinated multipoint transmission/reception techniques for LTE-Advanced [Coordinated and Distributed MIMO]. *Wireless Communications, IEEE*, 17:26–34, 2010.

[153] Mansoor Shafi, Andreas F Molisch, Peter J Smith, Thomas Haustein, Peiying Zhu, Prasan De Silva, Fredrik Tufvesson, Anass Benjebbour, and Gerhard Wunder. 5G: A tutorial overview of standards, trials, challenges, deployment, and practice. *IEEE journal on selected areas in communications*, 35(6):1201–1221, 2017.

[154] Tareq M Shami, David Grace, Alister Burr, and Muhammad D Zakaria. User-centric JT-CoMP clustering in a 5G cell-less architecture. In *29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 177–181. IEEE, 2018.

[155] Hidekazu Shimodaira, Gia Khanh Tran, Kiyomichi Araki, Kei Sakaguchi, Satoshi Konishi, and Shinobu Nanba. Diamond cellular network - optimal combination of small power base stations and CoMP cellular networks. In *24th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC Workshops)*, pages 163–167. IEEE, 2013.

[156] Guochao Song, Wei Wang, Da Chen, and Tao Jiang. KPI/KQI-driven coordinated multipoint in 5G: Measurements, field trials, and technical solutions. *IEEE Wireless Communications*, 25(5):23–29, 2018.

[157] Accenture Strategy. SMARTer2030: ICT solutions for 21st century challenges. Technical report, Global e-Sustainability Initiative, 2015.

[158] Huan Sun, Xiaobo Zhang, and Wei Fang. Dynamic cell clustering design for realistic coordinated multipoint downlink transmission. In *22nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 1331–1335. IEEE, 2011.

[159] Shaohui Sun, Qiubin Gao, Ying Peng, Yingmin Wang, and Lingyang Song. Interference management through CoMP in 3GPP LTE-advanced networks. *IEEE Wireless Communications*, 20(1):59–66, 2013.

[160] Xiaofang Sun, Nan Yang, Shihao Yan, Zhiguo Ding, Derrick Wing Kwan Ng, Chao Shen, and Zhangdui Zhong. Joint beamforming and power allocation in downlink NOMA multiuser MIMO networks. *IEEE Transactions on Wireless Communications*, 17(8):5367–5381, 2018.

[161] Karthikeyan Sundaresan, Mustafa Y. Arslan, Shailendra Singh, Sampath Rangarajan, and Srikanth V. Krishnamurthy. Fluidnet: A flexible cloud-based radio access network for small cells. In *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, pages 99–110. ACM, 2013.

[162] Tae-hyong Kim and Jae-woo Kim. Handover optimization with user mobility prediction for femtocell-based wireless networks. In *International Journal of Engineering and Technology (IJET), Vol.5 No.2*, pages 1829–1837, 2013.

[163] Ardian Ulvan, Melvi Ulvan, and Robert Bestak. The enhancement of handover strategy by mobility prediction in broadband wireless access. In *Proceedings of the Networking and Electronic Commerce Research Conference(NAEC)*, pages 266–276, 2009.

[164] Ingo Viering, Martin Dottling, and Andreas Lobinger. A mathematical perspective of self-optimizing wireless networks. In *International Conference on Communications (ICC)*, pages 1–6. IEEE, 2009.

[165] Cheng-Xiang Wang, Fourat Haider, Xiqi Gao, Xiao-Hu You, Yang Yang, Dongfeng Yuan, Hadi M Aggoune, Harald Haas, Simon Fletcher, and Erol

Hepsaydir. Cellular architecture and key technologies for 5G wireless communication networks. *IEEE Communications Magazine*, 52(2):122–130, 2014.

[166] Jingjing Wang and Bhaskar Prabhala. Periodicity based next place prediction. In *Nokia Mobile Data Challenge 2012 Workshop. p. Dedicated task*, volume 2. Citeseer, 2012.

[167] Qixing Wang, Dajie Jiang, Jing Jin, Guangyi Liu, Zhigang Yan, and Dacheng Yang. Application of BBU+RRU based CoMP system to LTE-Advanced. In *IEEE International Conference on Communications Workshops*, pages 1–5. IEEE, 2009.

[168] Ralf Weber, Andrea Garavaglia, Matthias Schulist, Stefan Brueck, and Armin Dekorsy. Self-organizing adaptive clustering for cooperative multipoint transmission. In *73rd Vehicular Technology Conference (VTC Spring)*, pages 1–5. IEEE, 2011.

[169] Qingqing Wu, Geoffrey Ye Li, Wen Chen, Derrick Wing Kwan Ng, and Robert Schober. An overview of sustainable green 5G networks. *IEEE Wireless Communications*, 24(4):72–80, 2017.

[170] Wei Xu, Jian Huang, Fei Yang, and Hua Zhang. Dynamic cell clustering for CoMP in LTE-A and its calibrated system level performance evaluation. In *5th International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications (MAPE)*, pages 71–77. IEEE, 2013.

[171] Miao Yao, Munawwar Sohul, Vuk Marojevic, and Jeffrey H Reed. Artificial intelligence defined 5G radio access networks. *IEEE Communications Magazine*, 57(3):14–20, 2019.

[172] Qiaoyang Ye, Beiyu Rong, Yudong Chen, Mazin Al-Shalash, Constantine Caramanis, and Jeffrey G. Andrews. User association for load balancing in heterogeneous cellular networks. *IEEE Transactions on Wireless Communications*, 12(6):2706–2716, 2013.

[173] Ya-Ju Yu, Tzu-Yang Hsieh, and Ai-Chun Pang. Millimeter-wave backhaul traffic minimization for CoMP over 5G cellular networks. *IEEE Transactions on Vehicular Technology*, 68:4003–4015, 2019.

[174] Ya-Ju Yu, Wei-Chieh Tsai, and Ai-Chun Pang. Backhaul traffic minimization under cache-enabled CoMP transmissions over 5G cellular systems. In

2016 IEEE Global Communications Conference (GLOBECOM), pages 1–7. IEEE, 2016.

[175] Cui Zeng, Pinyi Ren, Chao Zhang, and Gangming Lv. A hybrid clustering approach in coordinated multi-point transmission system. In *7th International ICST Conference on Communications and Networking in China (CHINACOM)*, pages 627–631. IEEE, 2012.

[176] Deze Zeng, Jie Zhang, Lin Gu, Song Guo, and Jiangtao Luo. Energy-efficient coordinated multipoint scheduling in green cloud radio access network. *IEEE Transactions on Vehicular Technology*, 67(10):9922–9930, 2018.

[177] Yinghui Zhang, Yang Liu, and Jing Gao. Optimal design of energy efficient two-tier HetNets with massive MIMO by coordinated multipoint beamforming. *Physical Communication*, 28:183–189, 2018.

[178] Zhe Zhang, Ning Wang, Jiankang Zhang, Xiaomin Mu, and Kon Max Wong. Cooperation resource efficient user-centric clustering for QoS provisioning in uplink CoMP. In *18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5. IEEE, 2017.

[179] Zhengquan Zhang, Guang Yang, Zheng Ma, Ming Xiao, Zhiguo Ding, and Pingzhi Fan. Heterogeneous ultradense networks with NOMA: System architecture, coordination framework, and performance evaluation. *ieee vehicular technology magazine*, 13(2):110–120, 2018.

[180] Jian Zhao and Zhongding Lei. Clustering methods for base station cooperation. In *Wireless Communications and Networking Conference (WCNC)*, pages 946–951. IEEE, 2012.

[181] Jian Zhao, Tony QS Quek, and Zhongding Lei. Semi-distributed clustering method for CoMP with limited backhaul data transfer. In *Wireless Communications and Networking Conference (WCNC)*, pages 3776–3781. IEEE, 2013.

[182] Jingjing Zhao, Tiankui Zhang, Zhimin Zeng, Qiubin Gao, and Shaohui Sun. An overlapped clustering scheme of coordinated multi-point transmission for LTE-A systems. In *14th International Conference on Communication Technology (ICCT)*, pages 479–484. IEEE, 2012.

[183] Zhongyuan Zhao, Mugen Peng, Zhiguo Ding, Chonggang Wang, and H Vincent Poor. Cluster formation in cloud-radio access networks: Performance

analysis and algorithms design. In *IEEE International Conference on Communications (ICC)*, pages 3903–3908. IEEE, 2015.

[184] Yi Zhong, Xiaohu Ge, Howard H Yang, Tao Han, and Qiang Li. Traffic matching in 5G ultra-dense networks. *IEEE Communications Magazine*, 56(8):100–105, 2018.

[185] Fasheng Zhou, Lisheng Fan, Ning Wang, Gaoyong Luo, Jie Tang, and Wei Chen. A cache-aided communication scheme for downlink coordinated multipoint transmission. *IEEE Access*, 6:1416–1427, 2017.

[186] Anatolij Zubow, Ahmad Rostami, and Suzan Bayhan. On practical cooperative multi point transmission for 5G networks. *Computer Networks*, page 107105, 2020.