UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

NOVEL USER-CENTRIC ARCHITECTURES FOR FUTURE GENERATION CELLULAR NETWORKS: DESIGN, ANALYSIS AND PERFORMANCE OPTIMIZATION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

BY

UMAIR SAJID HASHMI
Norman, Oklahoma
2019

NOVEL USER-CENTRIC ARCHITECTURES FOR FUTURE GENERATION CELLULAR NETWORKS: DESIGN, ANALYSIS AND PERFORMANCE OPTIMIZATION

A DISSERTATION APPROVED FOR THE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

BY

Dr. Ali Imran, Chair

Dr. Thordur Runolfsson

Dr.Hazem Refai

Dr. Kam Wai Clifford Chan

Dr. Curt Adams

Dedicated to **“Ammi Jee”**

Acknowledgments

First of all, I bow down before the Almighty for his blessings on me throughout my life in general, and the last 4-5 years of my life as a PhD student in particular. I owe deepest gratitude to my parents, without whose prayers and encouragement, I would not have achieved this success. Then, my supervisor, Dr. Ali Imran, who has not only guided me throughout my research work but has also been a role model academician and professional. I will forever remain indebted to Prof. Imran, who instilled in me "the researcher aptitude". I will always remember and try my best to follow his advice on how to excel in personal and professional life. He has always encouraged and supported me to follow my dreams, and allowed me the flexibility to work on my thesis alongside gaining useful experience at multiple cellular and research organizations within US. I would like to mention my project mentor Dr. Syed Ali Raza Zaidi from University of Leeds. Dr. Zaidi's mentorship and guidance has been critical in refining and producing quality research work during my PhD. I consider myself fortunate to work alongside some of the brightest young researchers in the AI4NETWORKS Lab. The positive feedback, support and a congenial work environment has helped me achieve my degree targets over the last 4 years.

I would like to mention 2 people that have lived this PhD journey with me, through all its highs and lows. I thank my wife, Dr. Ayesha Umair, and son, Arham Umair Hashmi, who patiently and supportively stood by me over the last 5 years. Our family received immense love and kindness from the OU-Tulsa community. I will cherish the time I spent serving OU in non academic roles, in ISA, OUTSGA, I-CCEW and OUTPAC. I would specially like to mention Renee Wagenblatt, Lee Lawson and Krista Pettersen who always provided timely support and feedback for administrative and non academic activities throughout my degree. As we leave

OU-Tulsa and head back to Pakistan, we are taking loads of fond memories and leaving a part of ourselves behind. As the graduation certificate read: "No matter where you go, You'll always be a SOONER!".

Table of Contents

List of Figures

List of Tables

Abstract

Ambitious targets for aggregate throughput, energy efficiency (EE) and ubiquitous user experience are propelling the advent of ultra-dense networks. Inter-cell interference and high energy consumption in an ultra-dense network are the prime hindering factors in pursuit of these goals. To address this challenge, we investigate the idea of transforming network design from being base station-centric to user-centric. To this end, we develop mathematical framework and analyze multiple variants of the user-centric networks, with the help of advanced scientific tools such as stochastic geometry, game theory, optimization theory and deep neural networks. We first present a user-centric radio access network (RAN) design and then propose novel base station association mechanisms by forming virtual dedicated cells around users scheduled for downlink. The design question that arises is what should the ideal size of the dedicated regions around scheduled users be? To answer this question, we follow a stochastic geometry based approach to quantify the area spectral efficiency (ASE) and energy efficiency (EE) of a user-centric Cloud RAN architecture. Observing that the two efficiency metrics have conflicting optimal user-centric cell sizes, we propose a game theoretic self-organizing network (GT-SON) framework that can orchestrate the network between ASE and EE focused operational modes in real-time in response to changes in network conditions and the operator's revenue model, to achieve a Pareto optimal solution. The designed model is shown to outperform base-station centric design in terms of both ASE and EE in dense deployment scenarios. Taking this user-centric approach as a baseline, we improve the ASE and EE performance by introducing flexibility in the dimensions of the user-centric regions as a function of data requirement for each device. So instead of optimizing the network-wide ASE or EE, each user device competes for a user-centric region based on its data requirements. This

competition is modeled via an evolutionary game and a Vickrey-Clarke-Groves auction. The data requirement based flexibility in the user-centric RAN architecture not only improves the ASE and EE, but also reduces the scheduling wait time per user.

Offloading dense user hotspots to low range mmWave cells promises to meet the enhance mobile broadband requirement of 5G and beyond. To investigate how the three key enablers; i.e. user-centric virtual cell design, ultra-dense deployments and mmWave communication; are integrated in a multi-tier Stienen geometry based user-centric architecture. Taking into account the characteristics of mmWave propagation channel such as blockage and fading, we develop a statistical framework for deriving the coverage probability of an arbitrary user equipment scheduled within the proposed architecture. A key advantage observed through this architecture is significant reduction in the scheduling latency as compared to the baseline user-centric model. Furthermore, the interplay between certain system design parameters was found to orchestrate the ASE-EE tradeoff within the proposed network design. We extend this work by framing a stochastic optimization problem over the design parameters for a Pareto optimal ASE-EE tradeoff with random placements of mobile users, macro base stations and mmWave cells within the network. To solve this optimization problem, we follow a deep learning approach to estimate optimal design parameters in real-time complexity. Our results show that if the deep learning model is trained with sufficient data and tuned appropriately, it yields near-optimal performance while eliminating the issue of long processing times needed for system-wide optimization.

The contributions of this dissertation have the potential to cause a paradigm shift from the reactive cell-centric network design to an agile user-centric design that enables real-time optimization capabilities, ubiquitous user experience, higher

system capacity and improved network-wide energy efficiency.

List of Key Acronyms

3GPP	3rd Generation Partnership Project
5G	fifth generation wireless networks
AP	access point
APG	access point group
ANN	artificial neural network
ARPU	average revenue per user
ASE	area spectral efficiency
BBU	base band unit
BS	base station
CBS	control base station
CAPEX	capital expenditures
C-RAN	cloud radio access network
CDSA	control-data separation architecture
CoMP	coordinated multi-point
CQI	channel quality indicator
DNN	deep neural network
EE	energy efficiency
FDM	frequency division multiplexing
GA	genetic algorithm
GAN	generative adversarial network
GT-SON	game theoretic self-organizing network
IoT	internet-of-things
KPI	key performance indicator
LoS	line-of-sight
LTE	long term evolution
LTE-A	long term evolution-advanced
MBS	macro base station
MIMO	multiple-input-multiple-output
MNO	mobile network operator
MU	mobile user
MRT	maximal ratio transmission
NLoS	non-line-of-sight
NBS	Nash bargaining solution
OPEX	operational expenditures
PPP	Poisson point process
PRB	physical resource block
QoE	quality of experience
QoS	quality of service
RAN	radio access network
RLF	radio link failure

RSRP	reference signal received power
RSRQ	reference signal received quality
RRH	remote radio head
SBS	small base station
SIR	signal-to-interference ratio
SNR	signal-to-noise ratio
SINR	signal-to-interference-plus-noise ratio
SON	self-organizing network
S-cell	Stienen cell
TDD	time-division duplexing
TDM	time division multiplex
TTI	transmission time interval
UC-RAN	user-centric cloud radio access network
UE	user equipment
UDN	ultra-dense networks
VCG	Vickrey-Clarke-Groves

CHAPTER 1

Introduction

1.1 Motivation for user-centric architectures in future cellular networks

Since the advent of LTE / LTE-A at the start of this decade, we have seen an explosion of internet services that include online multimedia streaming, mobile social networking and internet gaming. As the number of users and the complexity of mobile applications increased exponentially, so did the mobile traffic growth in recent years. According to the latest visual network index report from Cisco [6], global mobile traffic will rise from 7.2 Exabytes per month in 2016 to reach 49.0 Exabytes per month by 2021. In addition to mobile data growth, use cases for fifth generation of cellular networks (also known as "5G") which include augmented / virtual reality, vehicular communication and inclusion of sensor devices as part of internet-of-things (IoT) further adds to the complexity of data requirements.

So how will mobile operators provide gigabit experience and zero latency to support 5G services? As shown in Fig.1.1 [1], from the different possible avenues, it has been well established in both academia as well as wireless network industry that the major contribution will come from impromptu network densification. However, network operators are facing numerous challenges arising from the dense small deployment, high inter-cell interference being the primary culprit. Deployment wise, network densification increases the total cost of ownership (TCO) which includes capital and operational expenditures (CAPEX, OPEX). As the average revenue per user (ARPU) remains virtually flat, network operators are fearing a

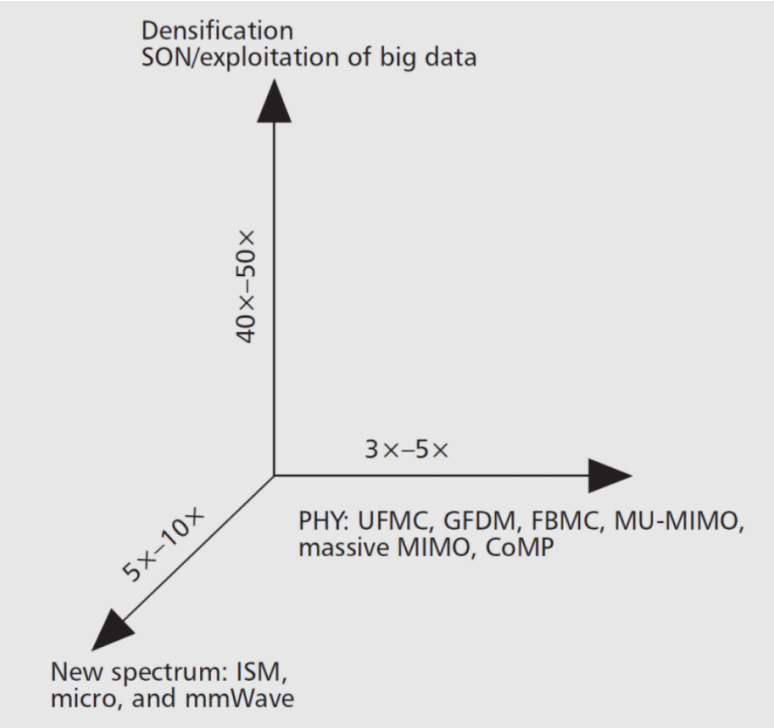


Fig. 1.1: Impact on new 5G methodologies / technologies on capacity gain [1].

crisis situation where rising expenses may overcome the dwindling profit margins [7]. To aggravate things further, impromptu cell deployments by mobile users (MUs) render traditional cell planning strategies inept. The aforementioned idiosyncrasies of ultra-dense small cell networks call for a paradigm shift in network design.

There was a time when it was believed that signal-to-interference ratio (SIR) distribution in dense wireless networks is independent of the base station (BS) density assuming that the BSs are spatially distributed as Poisson point process (PPP) with a power-law propagation model [8]. That implied that network can be densified infinitely to provide capacity gains without any bound. As exciting as it may sound, the density invariance property seemed way too optimistic. Subsequent works demonstrated the effect of exponential pathloss on SIR and showed the coverage probability to approach zero at very high base station (BS) deployment

densities. The multi slope pathloss model demonstrated that the area spectral efficiency saturates at high BS densities to a finite value [9]. The interference-limited behavior of ultra-dense networks (UDN) is not the only factor that requires a paradigm shift in the network design. Other contributing factors of the UDN that are pain points for a network operator are summarized below:

1. **Cell Edge Problem in UDN:** The "cell edge" refers to the region in the service area of mobile networks which receives cellular signals with significant reference signal received power (RSRP) from nearby BSs. This reduces the overall SIR for mobile users (MUs) in this region and hence limits the data rate for these MUs. The reduction in inter-site distances in UDNs will only worsen the cell-edge problem, as a multitude of nearby transmitting BSs will increase manifold the interference at an arbitrary MU. To provide ubiquitous user experience regardless of user location and movement, a new design architecture must enable uniform quality of experience (QoE) across the service region of the mobile network.

2. **Deployment and Operational Costs of UDN:** Overlaying dense small cell networks is associated with high deployment, operational and maintenance costs. The economical and ecological toll on the society in general, and network operators in particular is a cause of concern. A mobile network operator (MNO) has to undergo several processes for deployment of a small BS (SBS). This includes gaining site and equipment approvals, price negotiations with the city or landlord, deployment and provisioning of the BS, ensuring efficiency backhaul and power availability, and confirming to aesthetic and environmental regulations. Additionally, operational costs for dense small cell deployments mainly come from power consumption in power amplification and baseband processing as shown from project EARTH's anal-

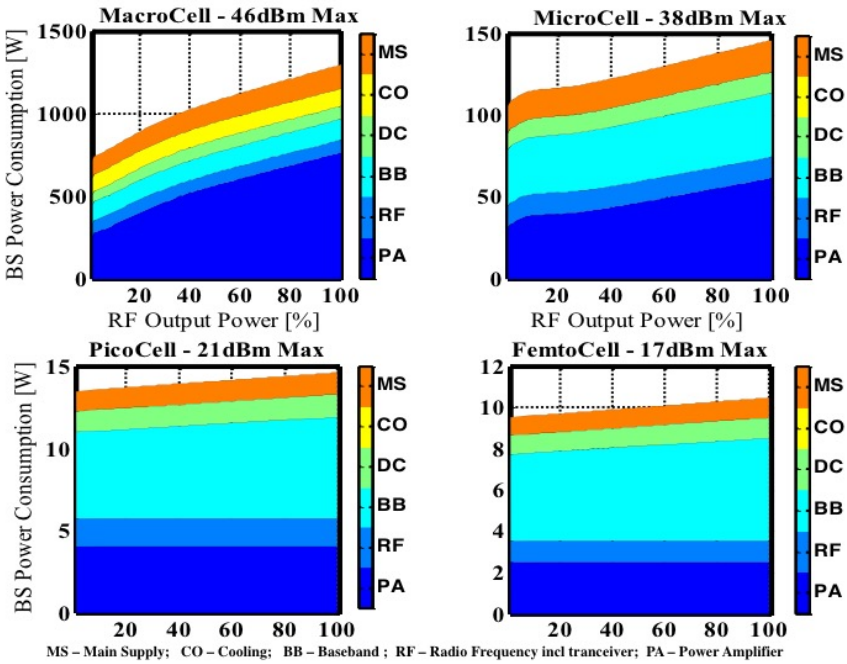


Fig. 1.2: Measured component level power consumption.

ysis [10, 11, 12, 13, 14, 15, 16, 17] in Fig. 1.2. Clearly, we need system designs that can overcome the challenge of high CAPEX and OPEX for UDN and allow the network to operate at a level which balances the benefits of ultra-dense deployment and additional deployment and network operating costs.

3. **Differentiated user quality of service:** The success of 5G will be dependent on the provisioning of quality of service (QoS) for applications whose demands, and nature are highly heterogenous. With the plethora of diverse mobile applications, ranging from messaging services, such as WhatsApp to social media such as Twitter and high definition video streaming, the quality of service requirement between MUs is highly non-uniform. Additionally, the low bitrate internet-of-things sensor devices is further adding to the mix of data requirement types. The future radio access network (RAN) design should be well equipped to handle the volume and variety of the traffic classes. Eventually, the aim is to re-design networks to provide high qual-

ity of experience to all connected devices, at the optimal network operating costs.

4. **From reactive to proactive mode of operation:** Current network optimization approaches to deal with user traffic diversity are reactive in nature. For instance, when there is a high data requirement, additional small BSs are deployed in that spatio-temporal region. Considering the acute dynamics of the future 5G networks, such approaches yield sub-optimal capacity and energy efficiency gains. Different architectures such as Cloud RAN (C-RAN) and control data separation architecture (CDSA) allow the flexibility to turn off small BSs ON and OFF according to traffic variation. However, to move towards truly proactive cellular networks, tools from artificial intelligence and machine learning can be leveraged to design algorithms for SBS discovery and selection, and radio resource allocation by switching ON/OFF SBS proactively instead of reactively to jointly maximize both sum capacity and energy efficiency without compromising QoE.

1.2 Research Objectives

In light of the discussion in section 1.1, the research presented in this dissertation provides answers to the following questions:

i) How do we characterize a user-centric architecture, supported by C-RAN and CDSA, in terms of ASE and EE while taking into account all dimensions of ASE-EE interplay, for instance user-centric service zone sizes, SBS, macro BS (MBS) and MU density, transmission powers and power consumption models, QoE constraints, and spatio-temporal traffic variations, thereby enabling not

only its design optimization but also a systematic and fair comparison with the legacy architectures?

ii) Given that the optimal user-centric service zone size that maximizes SE is different from that which maximizes EE (see Fig. 1.3 [18]), what combination of user-centric service regions and other design parameters can be optimized jointly to maximize both ASE and EE? How can we design a self-organizing framework to dynamically adjust the user-centric design parameters such as service zone size, SBS density, transmission powers and maximize both ASE and EE while also leveraging spatio-temporal variations in user traffic?

iii) How feasible is a distributive architecture in which UEs compete for user-centric service regions through collaborative or non-collaborative game models? How would the ASE, EE and user scheduling delay performance metrics change if flexible sized service zones are employed in the user-centric design?

iv) Will user-centric architectures also be applicable at higher frequency mmWave spectrum and yield similar ASE, EE performance gains as in the sub-6 GHz spectrum? What is the optimal geometry of user-centric service zone regions, for instance, circular, Voronoi regions, Stienen regions etc.?

v) From the analytical models for optimization of the network efficiency tradeoff modelling in user-centric networks, what is the computational efficiency for real-time optimization of design parameters in cellular networks? Can we design artificial neural networks (ANNs) that can be trained to learn the map between system parameters and corresponding efficiency optimization utility function? How can the trained ANN be refined to improve the performance? How close will the ANN performance be to the global optimum, and what gains in real-time computation can be achieved using this approach?

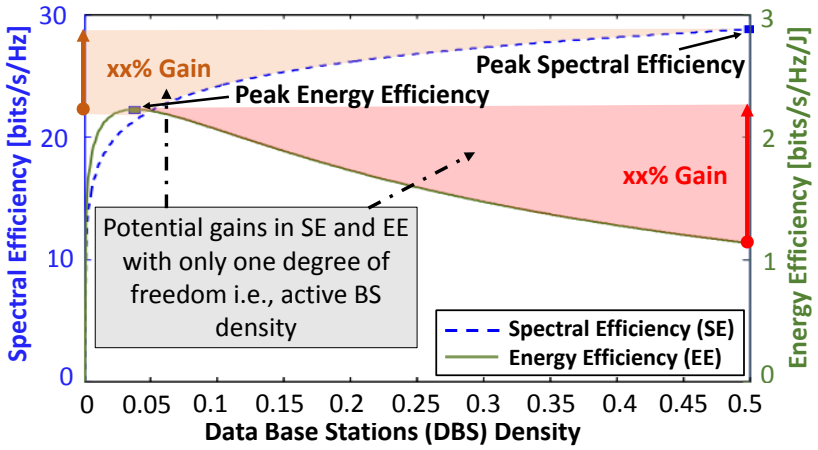


Fig. 1.3: EE-SE gain tradeoff characteristics in user-centric networks

1.3 Contributions

This dissertation addresses the aforementioned research questions. Analytical models are developed, and 3GPP-compliant rigorous simulation studies are carried out to find and validate the answers to the above questions. The key contributions of the dissertation are outlined in the following section.

- This dissertation contributes by presenting a novel user-centric architecture for dense small cell deployment in a C-RAN enabled cellular network. A comprehensive statistical framework is developed for computation of the area spectral and energy efficiency of a large-scale user-centric cloud radio access network. We also propose a user-centric RRH clustering algorithm which enables dynamic coverage extension and shrinkage by activating a single remote radio head (RRH) within a specified demarcation around a scheduled user based on max SIR gain criteria. The user-centric architecture ensures uniform coverage and no cell-edge degradation for all the users irrespective of their physical location. We demonstrate that there exists an optimal cluster radius which maximizes the area spectral efficiency of the UC-RAN.

It was also demonstrated that this optimal cluster radius is coupled with the user density and hence must be adapted by a self-organization mechanism. The link level performance is then employed to perform cost-benefit analysis of the proposed protocol. More specifically, the amount of power dissipated in the association process under the proposed protocol is considered as the cost of obtaining the throughput gains. The throughput-cost ratio is hence essentially the energy efficiency of the user-centric C-RAN. It was shown that there exists another optimal cluster radius which maximizes the energy efficiency of the proposed network model. However, this is larger to the one obtained under area spectral efficiency criterion. Consequently, the throughput-energy consumption trade-off manifests itself in terms of dimensioning of the cluster radius in the user-centric model. Using a game theoretic framework, we demonstrate that a self-organizing network (SON) engine within the centralized base band unit (BBU) pools may be employed to dynamically configure the optimal cluster size. Simulation results indicated that: i) the SON mechanism allows more than 100x efficiency variation through real-time adjustment in the Nash bargaining solution (NBS) bias parameter, ii) in comparison to current always-ON RRH deployments, selective RRH activation offers high area spectral and energy efficiency gains, and iii) significant SINR gains can be realized in both ASE and EE operating modes by virtue of interference-free RRH cluster zones around each scheduled user.

- Next, we develop a game theoretic based user-centric model that attempts to optimize the network efficiency metrics by taking into account the diversity of user data demands within different spatio-temporal regions in the network. More specifically, we propose a QoE-centric elastic framework for a dense multi-tier cellular network deployment. The framework leverages the control

and data plane separation architecture (CDSA) for enabling selective SBS activation within user equipment (UE)-centric virtual cells (also referred to as service zones). The allocation of these virtually elastic service zones around selected UEs is conducted via a central control base station (CBS) and modeled through two game techniques, namely evolutionary and auction games. Both the games are based on a utility minimization problem which is a function of weighted mean UE throughput and usage-based UE service demands. To illustrate the trade-offs between the game models, network level performance is compared in terms of aggregate throughput, energy efficiency, algorithm convergence speed and mean UE scheduling probabilities.

- The next contribution of this dissertation is analytical quantification and analysis of a novel user-centric Stienen cell model for dense mmWave cellular networks. In the proposed architecture, at most one remote radio head is activated within non-overlapping user equipment-centric Stienen cells (S-cells) generated within the Voronoi region around each UE. Under the presented framework, we derive analytical models for the three key performance indicators (KPIs): i) SINR distribution (used as an indicator for quality of service (QoS)), ii) area spectral efficiency (ASE), and iii) energy efficiency (EE) as a function of the three major design parameters in the proposed architecture, namely UE service probability, S-cell radius coefficient and RRH deployment density. The analysis is validated through extensive Monte Carlo simulations. The simulation results provide practical design insights into the interplay among the three design parameters, tradeoffs among the three KPIs, sensitivity of each KPI to the design parameters as well as optimal range of the design parameters. Results show that compared to current non-user-centric architectures, the proposed architecture not only offers sig-

nificant SINR gains, but also the flexibility to meet diverse UE specific QoS requirements and trade between EE and ASE by dynamically orchestrating the design parameters.

- Building on the earlier contributions of the dissertation, an optimization framework for modeling the ASE-EE tradeoff in a multi-tier user-centric Stienen network 130 model is designed using principles from stochastic optimization. The framework integrates the randomness in the MBS, mmWave RRH and UE locations and also penalizes rapid fluctuation in the formulated utility function with changes in the network snapshots. It is seen that the utility employed to model the ASE-EE tradeoff is highly non-convex with the system parameters under consideration namely UE service probability, S-cell radius coefficient and RRH activation percentage. The non-convexity arises from randomness in the MBS, RRH and UE locations along with the piecewise pathloss functions to differentiate between UEs having line-of-sight (LoS) and non-line-of-sight (NLoS) connectivity with their respective serving RRHs. A deep neuro-evolution approach is then followed to map the optimization parameters to the network statistics and system utility value. A deep neural network (DNN) is trained on a set of data containing optimization parameters, network statistics containing mean UE, MBS and RRH Poisson point process (PPP) intensities, and the output utility value. The trained DNN is tested on new network statistics and is shown to predict very similar optimal parameters as the time intensive brute force method and consequentially yields near-optimal utility values. The real-time DNN optimization is further enhanced with hyperparameter tuning using genetic algorithm (GA) method. Another refinement is made by data enrichment on training data sub-space having higher sensitivity to the output utility. Our

overarching goal is to demonstrate that real-time optimization of user-centric networks is possible through a centralized SON engine, without significant loss in performance.

1.4 Dissemination and Publications

Throughout the course of preparation for this dissertation, several dissemination activities were carried out. These activities have resulted in the following presentations and peer reviewed (accepted or pending) articles.

Journals:

J1. U. Hashmi, S. A. R. Zaidi and A. Imran, "Enhancing Downlink QoS and Energy Efficiency through a User-Centric Stienen Cell Architecture for mmWave Networks", IEEE Transactions on Green Communications and Networking, 2019, (submitted).

J2. M. Hartmann, U. Hashmi and A. Imran, "Edge Computing in Smart Health Care Systems: Review, Challenges and Research Directions", Transactions on Emerging Telecommunications Technologies, 2019, (accepted for publication).

J3. U. S. Hashmi, S. A. R. Zaidi and A. Imran, "User-Centric Cloud RAN: An Analytical Framework for Optimizing Area Spectral and Energy Efficiency," in IEEE Access, vol. 6, pp. 19859-19875, 2018. doi: 10.1109/ACCESS.2018.2820898

J4. U. Hashmi, U. Masood and A. Imran, "On real-time optimization of multi-tier user-centric Stienen networks: A sequential deep learning approach", IEEE IEEE Transactions on Vehicular Technology, 2019 (to be submitted).

Conferences:

C1. U. S. Hashmi, A. Rudrapatna, Z. Zhao, M. Rozwadowski, J. Kang, R. Wuppalapati and A. Imran, "Towards Real-time User QoE assessment via Machine Learning on LTE network data", 2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall), Honolulu, HI, USA, 2019.

C2. U. B. Farooq, U. S. Hashmi, J. Qadir, A. Imran and A. N. Mian, "User Transmit Power Minimization through Uplink Resource Allocation and User Association in HetNets," 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates, 2018, pp. 1-6. doi: 10.1109/GLOCOM.2018.8647409

C3. U. S. Hashmi, A. Islam, K. M. Nasr and A. Imran, "Towards User QoE-Centric Elastic Cellular Networks: A Game Theoretic Framework for Optimizing Throughput and Energy Efficiency," 2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Bologna, 2018, pp. 1-7. doi: 10.1109/PIMRC.2018.8580747

C4. U. S. Hashmi, S. A. R. Zaidi, A. Darbandi and A. Imran, "On the Efficiency Tradeoffs in User-Centric Cloud RAN," 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, 2018, pp. 1-7. doi: 10.1109/ICC.2018.8422228

C5. S. Cetinkaya, U. S. Hashmi and A. Imran, "What user-cell association algorithms will perform best in mmWave massive MIMO ultra-dense HetNets?," 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, 2017, pp. 1-7.

doi: 10.1109/PIMRC.2017.8292248

C6. U. S. Hashmi, A. Darbandi and A. Imran, "Enabling proactive self-healing by data mining network failure logs," 2017 International Conference on Computing, Networking and Communications (ICNC), Santa Clara, CA, 2017, pp. 511-517. doi: 10.1109/ICCNC.2017.7876181

1.5 Organization

The dissertation is structured as follows. Chapter 2 presents the background on the different techniques utilized in the dissertation. These include: i) user-centric architectures for 5G, and ii) application of game theory for optimization wireless networks.Chapter 3 focuses on a non-elastic user-centric architecture in a Cloud RAN deployment for dense networks. After detailing the system model, we explain the user-centric RRH clustering algorithm and the one-to-one UE-RRH association mechanism. We then quantify the area spectral and energy efficiencies of the 1st tier elastic user-centric network. Integration of the SON framework for optimal balance between the two efficiency metrics is also discussed before presenting the numerical results. In chapter 4, our focus shifts on the user-level elasticity in the user-centric networks where the pay off function and service zone modeling around high priority UEs is characterized. The game models employed in the work as UEs compete for the service zones are elaborated in detail. Following that, we present analytical and simulation results for both the elastic models are presented which show the performance enhancement over the non-elastic user-centric models. In chapter 5, we propose and analyze a user-centric Stienen cell architecture for densely deployed remote RRHs operating in mmWave frequency spectrum. We develop a statistical framework for deriving the coverage probability of an arbitrary UE scheduled within the proposed architecture. Quantification of the ASE and EE is done both for single tier networks, with mmWave RRH coverage

only, and multi-tier networks that have deployment of both sub-6 GHz MBS and mmWave RRHs. The user-centric scheduling in both single-tier and multi-tier scenarios is also presented. Numerical results and tradeoff analysis is done to emphasize on the need of a SON engine to yield Pareto optimal tradeoff between the network efficiency parameters. We extend this work in chapter 6, in which we discuss the convexity of the proposed optimization formulation for a multi-tier user-centric Stienen network. We then proceed to discuss the sequential learning based approach to select optimal parameters and discuss its performance results in comparison to the global optima. Refinements to the neural network model and the improvement in performance is also presented. Finally, chapter 7 discusses the conclusions and future work, and it thus concludes the dissertation.

CHAPTER 2

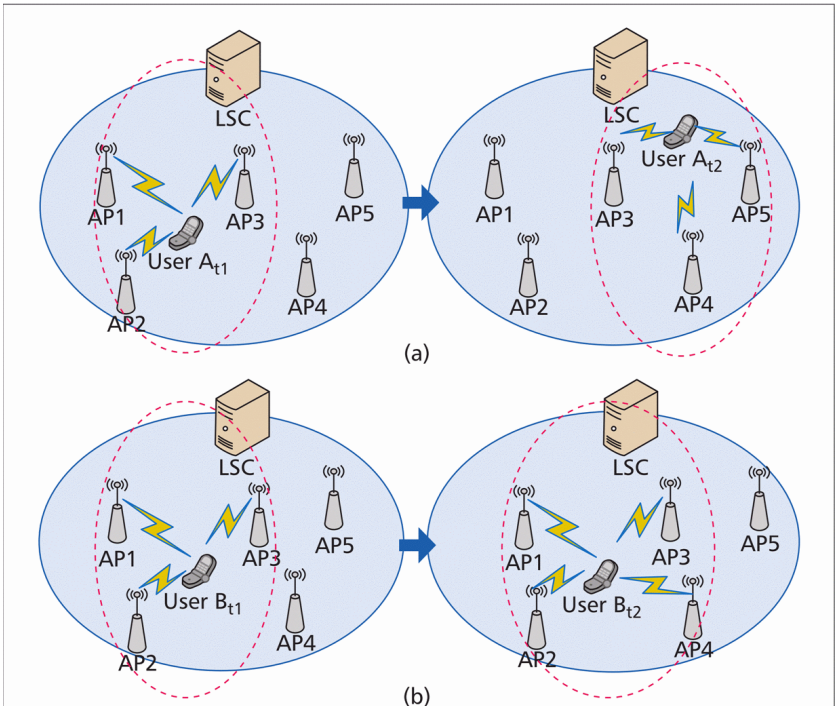
Background and Preliminaries

Over the last decade or so, research community and experts in telecom industry alike have unanimously agreed that frequency re-use through densification of the network is the prime source to meet the exponential data demands in 5G and beyond cellular networks. Around the same time, researchers from Ericsson started discussion on why it is viable to consider 3-300 GHz spectrum for mobile broadband applications [19]. Higher frequency band offers multiple advantages, including high spectrum availability and small component sizes for mobile applications. A few years along the line, efforts began in trying to replicate the abilities of human brain to extract the abstract characteristics from the feature data via a computer device [20]. Deep learning, based on neural networks design, is able to find hotspots, interference distribution and spectrum availability from a large amount of network parameters. In this chapter, we will discuss some of the background principles and relevant works in literature within the areas covered in this dissertation. This includes: i) user-centric networks in 5G and beyond, and ii) application of game theory for resource optimization in wireless networks.

2.1 User-centric architectures for future cellular networks

The main concept behind a user-centric cellular architecture is a paradigm shift from the base-station centric design to the philosophy of the network serving the user. This means that the whole network optimization, be it interference management, mobility management or resource management, will now be performed to

enhance the user experience at individual basis [2]. The shift from network controlling user to the network service user involves the decoupling of local service and network service. This can be performed by decoupling of user and control plane, where large range low frequency macro base stations act on the control plane, and dense deployment of small cells provide the data plane for the users. One of the prime challenges in the design of user-centric service is assignment of an access point group (APG) to the user based on the mobility patterns. The access point group assignment changes with every interval based on the speed and direction of the user. An example of such grouping for two user terminals in consecutive time slots is given as Fig. 2.1 [2].



	Time point	AP1	AP2	AP3	AP4	AP5
User A's APG	t1	√	√	√		
	t2			√	√	√
User B's APG	t1	√	√	√		
	t2	√	√	√	√	

Fig. 2.1: Access point grouping in consecutive time intervals [2].

User-centric architectures have been shown to improve rate-per-user in different scenarios, such as the cell free massive multiple-input-multiple-output (MIMO), wherein a large number of distributed access points (APs) simultaneously and jointly serve a small number of user devices [21]. In comparison with co-located massive MIMO, the distributed massive MIMO where multiple user devices are served by multiple-antenna APs using time-division duplexing (TDD). This cell-free approach alleviates the cell-edge issues and improves the likely per-user throughput [22]. The user-centric approach further improves not only the per-user data rate but also reduces the backhaul requirements [23]. Other works have also shown that the user-centric approach outperforms the cell free design in terms of per-user throughout and optimal power allocation [24, 25].

User-centric architectures in mmWave spectrum need to take into account the effects from blockages and different small-scale fading distributions (Nakagami, Rayleigh and no fading). At low mmWave small cell deployment and high blockage parameter scenarios, the system behaves as noise-limited and provides considerable ergodic capacity gain [26]. In another paper, a user-centric adaptive clustering method is proposed to improve the outage capacity and raise the tenth percentile SINR by 12 dB [27]. The user-centric approach has also been employed as a resource-aware energy-saving technique which aims at minimizing the total power consumption of the system while keeping the data rate requirement at every user as a constraint in the optimization framework [28]. In [29], a virtual user-centric Cloud RAN cell design is proposed with the user at the center of circular virtual cell served by multiple remote radio heads (RRHs) simultaneously. Optimal power control in the form of maximal ratio transmission (MRT) is used at the RRHs. The virtual size can be fluctuated to yield the maximum system capacity in at a given RRH density. A power control strategy is presented from the user-centric

perspective in [30]. All cooperative access points within a cooperation region will receive power control requests from the user terminals. The size of the cooperating region is the optimization parameter that maximizes the system spectral efficiency. Finally, in [31], a self-organized, load-aware user-centric coordinated multi-point (CoMP) clustering algorithm is presented to maximize spectral efficiency and load balancing efficiency. Additional capacity generated by increasing the cluster size is utilized for load balancing and a novel re-clustering algorithm is proposed to distribute the traffic from highly loaded cells to neighbor cells with light load, reducing the number of unsatisfied user equipments (UEs) significantly.

2.2 Game theory applications in wireless networks

Game theory has been around since mid-20th century as a science where multiple agents interact to achieve their respective objectives [32]. Game theory has been shown to have numerous applications in the fields of economics, social science, engineering and computer science. There are multiple text books that present comprehensive discussion advanced game theory [33, 34]. Game theory also finds its application in communications engineering, particularly for distributed resource management. Some of these include power control [35, 36, 37], optimal pricing strategies [38], congestion control [39], and load balancing [40]. Most of the literature on game frameworks in wireless networks applies statics games in which the players make decisions simultaneously, without knowledge of the other players' strategies. Examples of static games in wireless networks include power allocation policies in wireless networks [41, 42, 37], and optimal resource allocation in cognitive radio networks [43, 44]. On the other hand, in stochastic games, multiple players take actions based on their respective state space and transition probabilities in an attempt to maximize their utilities. For instance, stochastic games,

when applied to access network selection in [45] has shown to achieve higher total utility per user.

[46] is a relatively recent book on application of game theory in wireless networks. The book first elaborates different game types, including non-cooperative games [47], Bayesian games [48], differential games [47], evolutionary games [49], cooperative games [32] and auction theory [50]. It then highlights different wireless network topologies which are catered by game theory. These include: i) cellular and broadband wireless networks, ii) multi-hop networks, iii) cooperative transmission networks, iv) cognitive radio networks, and v) internet networks. In Table 2.1, we summarize some major works which are relevant to the tasks undertaken in this dissertation and elaborate their scope and methodology.

Table 2.1: Some related works on Game Theory application in wireless networks

Reference	Methodology	Scope	Results
Shih et al. [51]	An economic framework for incentivising resource sharing by femto BS owners to other users through monetory rewards from the operator in a dense FBS deployment scenario.	Co-existence of a framework with users and femtocells cooperating for optimal resource sharing.	Two-stage sequential game shows better revenue for femto cell operators.
Semasinghe et al. [52]	Evolutionary game theory (EGT) to model the distributed subcarrier and power level selection schemes for a multi tier network.	Distributed resource allocation in self-organizing small cell networks underlaying macrocell networks.	The distributed algorithm shows close to optimal performance with clear reduction in processing time.
Kang et al. [53]	The spectrum sharing problem in a heterogeneous network is modelled as a resource allocation price based Stackelberg game.	Interference control schemes for both the uplink and downlink transmissions in spectrum-sharing femtocell networks.	Low complexity solution using both uniform and non-uniform pricing strategies.
Pantisano et al. [54]	A coalition game framework is used where femto base stations(FBSs) have the choice of suppressing interference in a cooperative manner within coalitions.	Decision making at the FBS level based on the tradeoff between the cooperation gains, in terms of an increased revenue, and the cost in terms of power for information exchange.	The average payoff per femtocell is enhanced for the proposed methodology.
Zheng et al. [55]	The interference mitigation game has been modelled as an exact potential game with the potential function that maximizes the potential of players (FBSs) locally or globally.	For optimal channel selection that is distributed and fully dynamic in nature.	Yields better performance than reinforcement learning and linear-reward learning.
Hashmi et al. [56]	A distributed utility minimization problem is developed and solved using evolutionary game theory and Vickrey Clark Groves (VCG) auction.	Incorporating the effect of non-uniform user demands across and within spatio-temporal zones in user-centric networks.	The proposed model outperforms static user-centric models in terms of system throughout and energy efficiency.

CHAPTER 3

Analysis of non elastic user-centric architectures in dense Cloud RAN

High energy consumption in ultra-dense small cell networks is a paramount challenge that is a pain point for MNOs all over the globe. While different energy saving schemes have been proposed, an additional challenge faced by the operators is to ensure ubiquitous service level to its customers. In this chapter, we discuss the first contribution of the dissertation which is based on created non elastic service regions around UEs for downlink scheduling. This 1st tier elasticity in user-centric networks relies upon creating non-overlapping circular service zones around high priority UEs. This results in a one-to-one UE-BS association for each service zone such that a single BS is activated at max per service zone. The service zone radius is employed as a control parameter to realize the desired compromise between EE and ASE. The main feature of this macro level elasticity is that the service regions around UEs are of uniform size in any given time instance. Extensive system-level simulations demonstrate the high QoE possible with such a user-centric framework. The tradeoff between the network efficiency parameters enables the operator to choose the mode of operation according to it's business objectives.

3.1 Introduction

3.1.1 Background

As the mobile data transmission is expected to grow 7-fold from 2016 to 2021 [6], network densification through a conglomeration of diverse technologies (HetNets) seems to be the viable way forward to 5G. Network operators are facing numerous challenges arising from the dense small deployment, high inter-cell interference being the primary culprit. Deployment wise, network densification increases the total cost of ownership (TCO) which includes capital and operational expenditures (CAPEX, OPEX). As the average revenue per user (ARPU) remains virtually flat, network operators are fearing a crisis situation where rising expenses may overcome the dwindling profit margins [7]. To aggravate things further, impromptu cell deployments by mobile users (MUs) render traditional cell planning strategies inept. The aforementioned idiosyncrasies of ultra-dense small cell networks call for a paradigm shift in network design.

Certain promising disruptive 5G technologies such as massive MIMO and mmWave are being considered for higher average user throughput in 5G. However, both of these technologies, while offering higher network wide capacity, are likely to cost more in terms of energy efficiency and location-independent uniform user Quality of Experience (QoE). In this work, we investigate a User-centric Cloud Radio Access Network (UC-RAN) architecture that has the potential to address the aforementioned challenges. Conventional C-RAN allows centralizing and sharing of the baseband processing between several small cells in a virtual baseband processing unit (BBU) pool [57][58]. By separating baseband units from the radio access units, the C-RAN architecture: (i) reduces the capital and operational expenditure [58]; (ii) provides huge energy saving (due to centralized air-conditioning etc.)

and (iii) provisions implementation of sophisticated coordination mechanisms for reducing the co-channel interference [59]. However, there is one key 5G requirement that conventional C-RAN still fails to address, i.e. QoE. In conventional C-RAN, the Quality of Service (QoS) varies significantly from cell center to cell edge, same way it does in legacy networks, leading to poor QoE. UC-RAN on the other hand has potential to virtually remove cell edges by shifting the pivot of the cell design from the base station (BS) to the mobile user (MU) [2] [60]. The key distinct feature of UC-RAN is that, a cell is built around a user and not around the RRH or BS as in current networks [61]. This enables dynamic coverage as well as higher gains at the user terminals through spatial diversity from having several RRHs available to serve a user [59], [62].

The system design of user-centric architectures for small cell based networks has sparked interest for research in this area, which includes but is not limited to access point grouping mechanisms [63] [2], transmit power control strategies [30], interference alignment [64], RAN selection [65], dynamic load balancing [31] and optimal cluster dimensioning [59] [62] [61]. However, to the best of authors' knowledge, the analytical characterization of the area spectral and energy efficiencies and analysis of the impact different network parameters have on these efficiency metrics remains terra incognita. To this end, in this article, we address some fundamental design questions and propose a novel RRH clustering technique for designing efficient large scale UC-RANs. Furthermore, we present a game theoretic framework to trade-off between ASE and EE in dynamic fashion. This framework allows to retain a pareto-optimal performance while accommodating varying network load and operator's priority between ASE and EE.

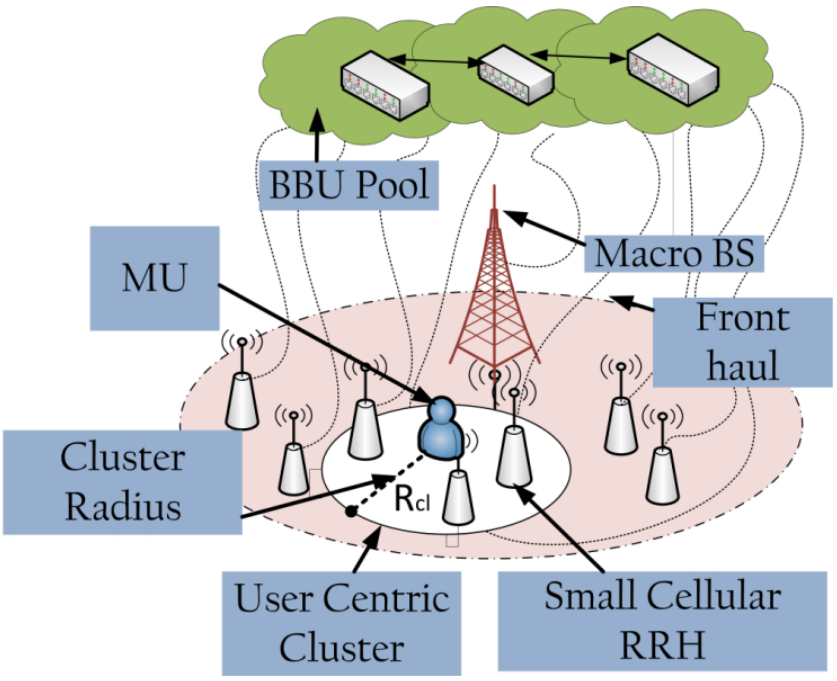


Fig. 3.1: UC-RAN architecture with activation disc of radius R_{cl} for a served user.

3.1.2 Architectural Overview

With the assumption of ultra-dense RRH deployment, which is a plausible scenario in future networks, inter-cell interference and network management issues due to close BS-user spatial proximity become prominent. UC-RAN addresses these by provisioning demand based baseband processing to RRHs and seamless coverage to the users. In a UC-RAN architecture, the BBU pool turns ON the RRHs that are required to serve a desired user at a certain QoS. The dynamic clustering of RRHs allows users to be served seamlessly and provides uniform service experience regardless of user location and movement [59]. The user-centric cluster size serves as a proxy to the minimum spatial separation between an arbitrary user and its closest interfering RRH, thereby improving the average SINR at the user. Furthermore, contrary to the traditional small cell networks where the energy consumption and the handovers both increase with the density of the small cell

RRHs, UC-RAN mitigates this problem by providing on demand coverage, i.e., by selectively turning ON as opposed to always turned ON RRHs. Additionally, the throughput gains provided by distributed diversity alleviate the overhead of cooperation.

Fig. 3.1 provides a graphical illustration of a UC-RAN with RRH clustering. The RRHs are connected to the pool of BBUs via flexible front haul. The front haul is usually an optical fiber where signaling is done using radio-over-fiber (RoF) or common public radio interface (CPRI). Most of the signal processing at baseband level is delegated to the BBUs. The RRH deployment is expected to be very high density by leveraging the existing infrastructure (e.g. street lamp posts, poles, side of buildings etc.). The key idea here is to dynamically select the best RRH (in terms of SIR) within a circular area (virtual cell) of pre-defined radius around selected users (based on scheduling priority) during each scheduling interval. All other RRHs within the circle here after called cluster are kept OFF thereby minimizing the interference. The aforementioned UC-RAN architecture provides two-fold benefits: i) on-demand centralized processing at the BBU pools caters to non-uniform user traffic that subsequently enables OPEX reduction by as much as 30% [66], ii) user-centric RRH clustering reduces the number of nearby interfering RRHs and eliminates cell-edge coverage issues, hence improving the overall user experience.

3.1.3 Research Objectives

UC-RAN functions on resource pooling and virtual cell formation around scheduled users. The centralized user-centric RRH clustering not only reduces frequent handovers but also increases the signal-to-interference-plus-noise-ratio (SINR) that subsequently reduces the outage probability in dense networks. This in turn paves

the way for increase in the system wide spectral efficiency. Additionally, selective RRH activation enables reduced power consumption, hence making the network more energy efficient. Effectively, the user-centric RRH clustering:

1) empowers UC-RAN to provide demand based coverage, i.e., the coverage can dynamically extend or shrink based on user density by intelligently switching RRHs ON/OFF;

2) enables energy savings as an RRH is only turned ON when required to serve a nearby user;

3) provides an efficient way to control the signal strength which is a function of RRH cluster size (through both the maximum pathloss incurred and the number of serving RRHs providing selection diversity gain);

4) enables effective interference protection to an MU by inducing repulsion between RRH clusters (i.e., clusters are not allowed to overlap spatially which induces a natural guard-zone for a scheduled MU).

A critical design parameter in UC-RAN is the RRH cluster size around an arbitrary user. The cluster size may be defined in terms of number of nearby RRHs or an area covered by a circular disk around the user. In our model which is the latter case, the RRHs falling within the circular disk are designated to the corresponding user in a given time slot. Subsequent SIR based RRH activation is performed from amongst the RRHs within the cluster around the user. Increasing the cluster size offers following gains: i) larger distances between user and interfering RRHs results in larger link SINR and thus better link throughput; and ii) a larger cluster yields high macro diversity gain through selection among the larger number of RRHs in the cluster. However, the down side of a larger cluster is reduced spectrum reuse and a lower number of users that can be served simultaneously. This in turn

reduces system level capacity. Hence with a larger cluster, there are fewer higher quality links as opposed to many low bit rate links (which occur with smaller cluster radius). Another dimension of the trade-off that cluster radius offers is the energy efficiency. Higher cluster radius keeps more RRHs off as compared to lower cluster radius. In the back drop of these insights the goal of this paper is to investigate following research questions:

- What is the optimal RRH cluster size that maximizes a key performance indicator of capacity, i.e., area spectral efficiency?
- What is the cluster radius that yields optimal performance in terms of energy efficiency?
- What parameters are crucial in defining the optimal cluster sizes that maximize these system efficiencies (ASE, EE)? How sensitive are the efficiencies to variations in these parameters?
- Can we design a self-organizing framework to dynamically adjust the user-centric RRH cluster size and trade between ASE and EE in UC-RAN to cope with the spatio-temporal variations in user traffic?

In this contribution, we take the first step towards analytical treatment of the above mentioned design issues and answering the key research question at hand, i.e., what is the optimal cluster size around a scheduled user? Amongst recent works, studies in [59] and [62] are most relevant. However, our analysis differs in three key aspects: 1) [59] and [62] leverage user-centric architectures to optimize virtual cluster radius that maximizes the system capacity. On the other hand, we present a framework to simultaneously analyze ASE and EE in a UC-RAN. 2) Unlike [59] where the proposed clustering is overlapping (scenarios where a

single RRH may simultaneously serve multiple MUs), our model builds on non-overlapping user-centric clusters resulting in a one-to-one RRH-MU association during a given time slot. 3) Contrary to analysis in prior studies, we take into account variations in user density. By employing principles from stochastic geometry to model the thinned user and RRH densities in a particular time slot, we analyze the overall system efficiency more accurately. This allows investigation of relationship between key design parameters such as pathloss exponent and SINR threshold on ASE and EE for given user and RRH densities.

3.2 Network Model

3.2.1 Spatial Model of the Network

We consider a cloud radio access network under-laid within a large-scale cellular network. Both the small cell RRHs and MUs are assumed to be spatially distributed across the macrocells (see Fig. 3.2). The spatial distribution of the RRHs and the MUs is captured by two independent stationary Poisson point processes (SPPPs): $\Pi_{CLR} \in \mathbb{R}^2$ and $\Pi_{MU} \in \mathbb{R}^2$ with intensities λ_{CLR} and λ_{USR} respectively. Specifically, at an arbitrary time instant, the probability of finding $n_i \in \mathbb{N}, i \in \{RRH, MU\}$ RRHs/MUs inside a typical macro-cell with area footprint $\mathcal{A} \subseteq \mathbb{R}^2$ follows the Poisson law with mean measure $\Lambda_i(\mathcal{A}) = \lambda_i v_2(\mathcal{A})$. The mean measure is characterized by the average number of RRHs/MUs per unit area (i.e. λ_{CLR} and λ_{USR}) and the Lebesgue measure [67] $v_2(\mathcal{A}) = \int_{\mathcal{A}} d\boldsymbol{x}$ on $\mathbb{R}^2$, where if $\mathcal{A}$ is a disc of radius r then $v_2(\mathcal{A}) = \pi r^2$ is the area of the disc.

3.2.2 Channel Model

The channel between a UC-RAN RRH $\boldsymbol{x} \in \Pi_{CLR}$ and an arbitrary MU $\boldsymbol{y} \in \Pi_{MU}$ is modeled by $h_{xy}l(||\boldsymbol{x} - \boldsymbol{y}||)$. Here $h_{xy} \in \mathcal{E}(1)$ is a unit mean exponential random variable which captures the impact of a Rayleigh fading channel between an RRH and an MU. The small-scale Rayleigh fading is complemented by a large-scale pathloss modeled by $l(||\boldsymbol{x} - \boldsymbol{y}||) = K||\boldsymbol{x} - \boldsymbol{y}||^{-\alpha}$ power-law function. Here $||\boldsymbol{x} - \boldsymbol{y}||$ is the distance between $\boldsymbol{x}$ and $\boldsymbol{y}$, K is a frequency dependent constant and $\alpha \geq 2$ is an environment/terrain dependent pathloss exponent. The fading channel gains are assumed to be mutually independent and identically distributed (i.i.d.). Without any loss of generality, we will assume $K = 1$ for the rest of this discussion. It is assumed that the communication is interference limited and hence the thermal noise is negligible. Furthermore, we assume that all RRHs employ the same transmit power P_{CLR}.

3.3 User-centric Clustering in UC-RAN

In this article, we propose a user-centric clustering mechanism for the UC-RANs. More specifically, we envision a scenario where out of the multitude of small cell RRHs deployed in close proximity of an intended MU, a single RRH that provides the best channel gain (and consequently the highest signal-to-interference ratio (SIR)) is activated to serve that MU. The proximity or neighborhood of an MU is characterized by the cluster radius R_{cl}. The proposed user-centric clustering mechanism (Algorithm 3.1, Fig. 3.2) yields Π'_{MU} and Π'_{CLR} which is the set of scheduled MUs and activated RRHs during a particular time slot respectively.

As specified by Algorithm 3.1, the macro-cell or the BBU data center assigns a mark/tag $p_{USR} \sim \mathcal{U}(0, 1)$ to each MU. These marks correspond to the downlink

scheduling priority of the MUs. More specifically, the lower the value of the mark, the higher is the priority of the user to be served by the RRHs. Effectively, these marks can be thought of as the timers corresponding to each MU which are decremented on each time slot where service to this MU is deferred. A MU is scheduled for a downlink transmission iff it has highest scheduling priority in its neighborhood. In other words, there is no other MU in a disc of radius R_{cl} centered at MU with a higher priority. This round robin scheduling scheme ensures fair DL scheduling among MUs[1]. Notice that this disc also characterizes the size of the RRH cluster from which MU is being served. For a fixed R_{cl}, the percentage of MUs served in a given transmission time interval (TTI) is a function of relative RRH and MU PPP densities, i.e., if $\lambda_{USR} >> \lambda_{CLR}$, the average wait time before an arbitrary MU is served will be longer as compared to the scenario with same order MU and RRH densities.

The activation of RRHs is coupled with the user-centric scheduling mechanism (Algorithm 3.1). Only the RRHs which lie in the neighborhood of the scheduled users and provide the best propagation channel gain to their respective MUs are activated by the macro base station (MBS) (or BBU pool). This implies that each scheduled MU has a set of nearby RRHs that defines its user-centric RRH cluster. From this cluster of RRHs, only one that yields the highest SINR at the user is activated. Consequentially, there is at max one activated RRH that lies within a user-centric circular disk of area πR_{cl}^2. Effectively, activation of RRHs is on demand basis which provides UC-RAN capability of self-organizing the coverage to cope with the spatio-temporal variations of the user demography.

One might argue that such a non overlapping user-centric clustering scheme may result in service holes, i.e. there may exist MUs that are not associated with any

[1]The case with MUs having non-uniform scheduling priorities will be covered in future extensions of this work.

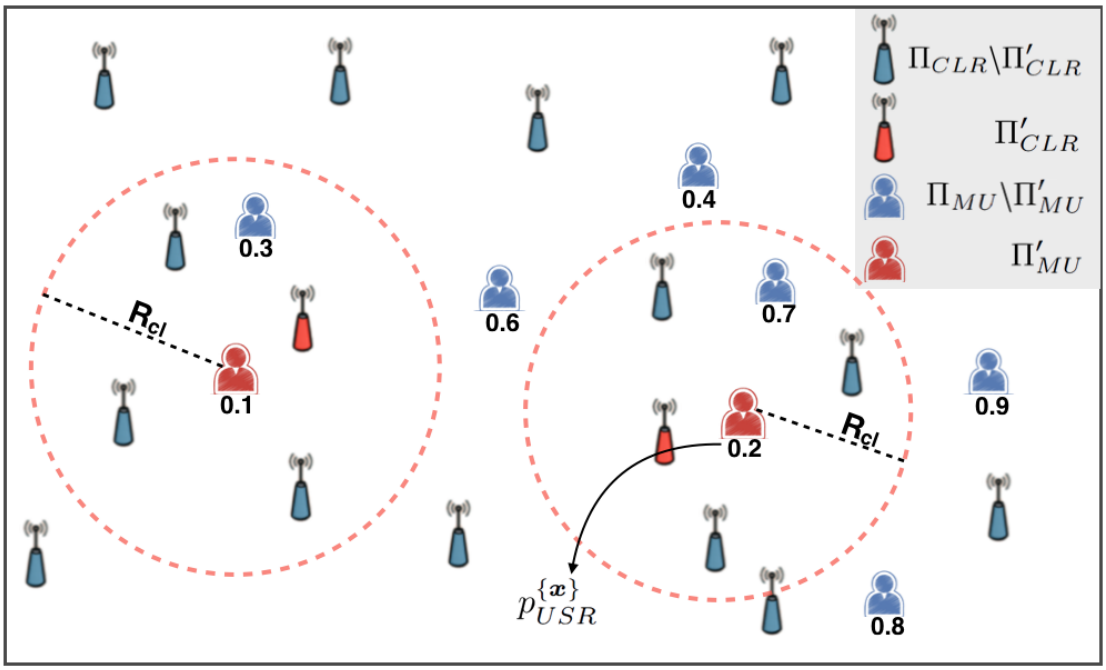


Fig. 3.2: User-centric RRH clustering in a UC-RAN. Each scheduled user is served by a single RRH in its respective cluster that maximizes its received SIR. The user-centric scheduling is based on $p_{USR}^{\{x\}}$ values with lower marks corresponding to high scheduling priorities.

RRHs due to empty RRH clusters around those MUs. Since we are considering dense small cell deployments with comparable λ_{CLR} and λ_{USR}, user-centric RRH clusters with realistic R_{cl} will hardly be void. In the unlikely scenario of a void cluster though, user clustering strategies [68] may be employed where nearby MUs are grouped together and optimization is performed on the MU clusters rather than individual MUs[2]. Furthermore, it is known that best RRH activation with a proximity constraint provides dual benefits of low outage probability and high power efficiency in dense deployment scenarios [69].

[2]In the interest of space, detailed discussion and evaluation of MU clusters will be presented in future publications.

Algorithm 3.1 User-centric clustering for C-RAN

Data: Π_{MU},Π_{CLR},R_{cl}
Result: Π'_{MU},Π'_{CLR}
$\Pi'_{MU} \leftarrow \emptyset$,$\Pi'_{CLR} \leftarrow \emptyset$;
foreach $\boldsymbol{x} \in \Pi_{MU}$ **do**
  $F_{PRIO}[\boldsymbol{x}] \leftarrow \mathcal{U}(0,1)$
end
foreach $\boldsymbol{x} \in \Pi_{MU}$ **do**
  $p_{USR}^{\{\boldsymbol{x}\}} \leftarrow F_{PRIO}[\boldsymbol{x}]$ $p_{SCH}^{\{\boldsymbol{x}\}} \leftarrow 1$ **foreach** $\boldsymbol{y} \in \Pi_{MU}$ **do**
    if $\boldsymbol{y} \neq \boldsymbol{x}$ **then**
      if $\boldsymbol{y} \in b(\boldsymbol{x}, 2R_{cl})$ *and* $p_{USR}^{\{\boldsymbol{y}\}} > p_{USR}^{\{\boldsymbol{x}\}}$ **then**
        $p_{SCH}^{\{\boldsymbol{x}\}} \leftarrow 0$
      else
        continue
      end
    else
      continue
    end
  end
  foreach $\boldsymbol{r} \in \Pi_{CLR}$ **do**
    if $\boldsymbol{r} \in b(\boldsymbol{x}, R_{cl})$ *and* $p_{SCH}^{\{x\}} == 1$ **then**
      if $h_{rx} l(||\boldsymbol{r} - \boldsymbol{x}||) > h_{r'x} l(||\boldsymbol{r'} - \boldsymbol{x}||), \forall \boldsymbol{r'} \in \Pi_{CLR}, \boldsymbol{r'} \in b(\boldsymbol{x}, R_{cl}), \boldsymbol{r'} \neq \boldsymbol{r}$ **then**
        $\Pi'_{CLR} \cup \{\boldsymbol{r}\}$
      end
    end
  end
  if $p_{SCH}^{\{x\}} == 1$ **then**
    $\Pi'_{MU} \cup \{\boldsymbol{x}\}$
  end
end

3.4 Quantifying the Area Spectral Efficiency of a UC-RAN

In the previous section, we presented an outline of a user-centric clustering algorithm for a UC-RAN. As is obvious from the algorithm, the size of the cluster employed for scheduling is a critical system design parameter. Optimal dimensioning of the R_{cl} is necessitated by the fact that:

1. The cluster size determines the number of the active RRHs at any given time. In turn the density of active RRHs shapes the co-channel interference experienced by a scheduled MU.

2. The radius of the cluster, also characterizes the number of concurrently scheduled MUs per unit area.

3. The dimensions of a cluster also determine the number of RRHs serving a scheduled MUs. This in turn determines the diversity gain experienced due to spatially distributed RRHs.

The area spectral efficiency of a UC-RAN network is strongly coupled with these three factors. In this section, our objective is two fold:

- To characterize the area spectral efficiency of a large scale UC-RAN.
- To investigate the optimal dimensioning of the cluster radius for maximizing the throughput potential of the UC-RAN.

3.4.1 Signal Model

Consider a scheduled user $\boldsymbol{x} \in \Pi'_{MU}$. Let $\mathcal{S}_{cop}(\boldsymbol{x}, R_{cl}) = \Pi'_{CLR} \cap (\boldsymbol{x}, R_{cl})$ be the unique RRH which is fed by the same BBU as $\boldsymbol{x}$ and selected to serve $\boldsymbol{x}$ on the basis of the scheduling criteria (Algorithm 1). Furthermore, let $\Pi_I = \Pi'_{CLR} \backslash \mathcal{S}_{cop}(\boldsymbol{x}, R_{cl})$ be the set of RRHs which are concurrently scheduled to serve $\boldsymbol{y} \neq \boldsymbol{x}, \forall \boldsymbol{y} \in \Pi'_{MU}$. In this article, we assume that the UC-RAN employs the RRH selection to serve its intended MU. Under RRH selection transmission the received signal at $\boldsymbol{x}$ can be written as (3.1), where $\max_{\boldsymbol{i} \in \mathcal{S}_{cop}} h_{\boldsymbol{ix}} l(||\boldsymbol{x} - \boldsymbol{i}||)$ is the channel gain between the serving RRH and the MU $\boldsymbol{x}$, $\max_{\boldsymbol{j} \in \Pi'_{CLR} \cap (\boldsymbol{y}, R_{cl})} h_{\boldsymbol{xj}} l(||\boldsymbol{x} - \boldsymbol{j}||)$ is the interference

$$r_x = \sqrt{P_{CLR} \max_{\boldsymbol{i} \in \mathcal{S}_{cop}} h_{\boldsymbol{ix}} l(||\boldsymbol{x} - \boldsymbol{i}||)} s_{\boldsymbol{x}} + \sum_{\boldsymbol{y} \in \Pi'_{MU}, \boldsymbol{y} \neq \boldsymbol{x}} \sqrt{P_{CLR} \max_{\boldsymbol{j} \in \Pi'_{CLR} \cap (\boldsymbol{y}, R_{cl})} h_{\boldsymbol{jx}} l(||\boldsymbol{x} - \boldsymbol{j}||)} s_{\boldsymbol{y}} + \varphi_{\boldsymbol{x}}. \tag{3.1}$$

experienced at $\boldsymbol{x}$ due to RRH $\boldsymbol{j}$ serving MU $\boldsymbol{y}$ and $s_{\boldsymbol{k}}$ is the message signal transmitted to MU $\boldsymbol{k}$ by its selected RRH. P_{CLR} is the transmit power employed by the RRH and $\varphi_{\boldsymbol{x}}$ is the additive white Gaussian noise (AWGN) at the receiver front end [3].

3.4.2 Probe Cluster

In order to characterize the area spectral efficiency of a UC-RAN, quantification of the success probability for a scheduled MU is essential. The key hurdle in characterizing the performance arises from the fact that unlike Π_{MU} the point process of the scheduled users Π'_{MU} is non-stationary. A closer inspection of Π'_{MU} reveals that it is indeed a modified version of a type II Matern hardcore process [67]. Fortunately, for such processes it is well established that they can be approximated by an equidense SPPP with appropriate modified intensity [70], [71][4]. Consequently, Π'_{MU} can be approximated by a SPPP with the intensity

$$\lambda_{USR}^{\{EF\}} = \frac{1 - \exp(-4\pi\lambda_{USR} R_{cl}^2)}{4\pi R_{cl}^2}. \tag{3.2}$$

[3]We notice that even with the induced spatial repulsion between MUs which consequently thins the PPPs of serving MUs and activated RRHs, the noise power at served MUs is negligible as compared to the aggregate interference. In other words, the network operates in a saturated conditions and the links are interference limited.

[4]This is validated by employing Ripley's K function [67] for both PPPs. It is observed that the K function of equi-dense PPP forms a lower bound on the Ripley's K function of scheduled MUs. The bound is very tight as the results are indistinguishable. In subsequent discussion, the equi-dense approximation is further validated by comparing theoretical area spectral efficiency of C-RAN against Monte-carlo simulation results.

Notice that since the user-centric clusters are considered non over lapping, the minimum distance between any two user-centric clusters should be $2R_{cl}$. Exploiting the stationary characteristics of the resultant SPPP, it is sufficient to focus on a typical MU. According to Silvnyak's theorem [67], the law of the SPPP does not change by addition of a single point. Hence we add a probe MU at origin. Moreover, the received signal ($r_{x(o)}$) in Eq. (3.1) can now be simplified with $r_i = ||\boldsymbol{i}-\boldsymbol{o}||$ and $r_y = ||\boldsymbol{y}-\boldsymbol{o}||$. For the sake of compactness, we will drop the index o for the rest of the discussion (e.g., $h_{\boldsymbol{io}} = h_{\boldsymbol{i}}$).

3.4.3 Lower-bound on the Success Probability of Scheduled MU

From Eq. (3.1) the received SIR at the probe MU can be expressed as

$$\text{SIR} = \Gamma_{MU} = \frac{\max_{\boldsymbol{i}\in\mathcal{S}_{cop}} h_{\boldsymbol{i}} l(r_{\boldsymbol{i}})}{\sum_{\boldsymbol{j}\in\Pi_I} h_{\boldsymbol{j}} l(r_{\boldsymbol{j}})}. \tag{3.3}$$

Notice that $\mathcal{S}_{cop}$ is a function of the non-stationary Poisson point process Π'_{CLR}.

Proposition 3.1 (Moments of Aggregate Interference). *The mean and variance of the aggregate interference experienced by a typical MU during a user-centric algorithm can be approximated as follows*

$$\kappa_1 = \mathbb{E}(I) = \frac{2\pi\lambda_{CLR}[1-\exp(-[1-\exp(-4\pi\lambda_{USR}R_{cl}^2]/4)]}{(\alpha-2)(R_{cl})^{\alpha-2}(\lambda_{CLR}\pi R_{cl}^2)}, \tag{3.4}$$

$$\kappa_2 = \mathbb{V}(I) = \frac{\pi\lambda_{CLR}[1-\exp(-[1-\exp(-4\pi\lambda_{USR}R_{cl}^2]/4)]}{(\alpha-1)(R_{cl})^{2(\alpha-1)}(\lambda_{CLR}\pi R_{cl}^2)},$$

where λ_{CLR} is the density of the UC-RAN RRHs, α is the pathloss exponent and R_{cl} is the radius of UC-RAN cluster.

Proof: Consider the SPPP Π_{CLR}, then under user-centric clustering algorithm, for each scheduled user, only a single RRH which resides in the vicinity as well

$$
\begin{aligned}
p_{ACT} &\overset{(a)}{=} \Pr\left[\Pi'_{MU} \cap b(\boldsymbol{r}, R_{cl}) \neq \emptyset | \boldsymbol{r} \in \Pi'_{CLR}\right].\{h_{\boldsymbol{r}} l(r_{\boldsymbol{r}}) > h_{\boldsymbol{j}} l(r_{\boldsymbol{j}}) | \boldsymbol{j} \in \Pi'_{CLR}, \boldsymbol{j} \neq \boldsymbol{r}\}\right], \\
&= \left[1 - \Pr\{\Pi'_{MU} \cap b(\boldsymbol{r}, R_{cl}) = \emptyset | \boldsymbol{r} \in \Pi'_{CLR}\}\right].\Pr\{h_{\boldsymbol{r}} l(r_{\boldsymbol{r}}) > h_{\boldsymbol{j}} l(r_{\boldsymbol{j}}) | \boldsymbol{j} \in \Pi'_{CLR}, \boldsymbol{j} \neq \boldsymbol{r}\}, \\
&= \left[1 - \exp(-\lambda_{USR}^{\{EF\}} \pi R_{cl}^2)\right].(1/[\lambda_{CLR} \pi R_{cl}^2]), \\
&= \frac{1 - \exp(-[1 - \exp(-4\pi\lambda_{USR} R_{cl}^2]/4)}{\lambda_{CLR} \pi R_{cl}^2}.
\end{aligned}
\tag{3.5}
$$

---

as provides maximum channel gain to that user is activated by the macro-cell. A natural implication of this policy is that the resulting PPP Π'_{CLR} is non-stationary. However, like Π'_{MU} it can be approximated with an equivalent SPPP with modified density $\lambda_{CLR}.p_{ACT}$. Here p_{ACT} is the activation probability for the RRH and can be computed as (3.5), where (a) follows from the fact that a BS is only activated if: i) there is a scheduled user within distance R_{cl}, and ii) there is no other BS within this distance of that user that can provide better channel gain. Now noticing that $\Pi_I = \Pi'_{CLR} \backslash \mathcal{S}_{cop}(\boldsymbol{o}, R_{cl})$, we can precisely describe $\Pi_I = \Pi'_{CLR} \backslash b(\boldsymbol{o}, R_{cl})$. Hence the mean and the variance can be computed using Campbell's theorem [67] as follows

$$
\begin{aligned}
\kappa_1 = \mathbb{E}(I) &= \mathbb{E}\left(\sum_{\boldsymbol{j} \in \Pi'_{CLR} \backslash b(\boldsymbol{o}, R_{cl})} h_{\boldsymbol{j}} l(r_{\boldsymbol{j}})\right), \\
&= 2\pi\lambda_{CLR} p_{ACT} \int_{Rcl}^{\infty} \mathbb{E}(H) r^{1-\alpha} dr,
\end{aligned}
\tag{3.6}
$$

and similarly

$$
\kappa_2 = 2\pi\lambda_{CLR} p_{ACT} \int_{Rcl}^{\infty} \mathbb{E}(H^2) r^{1-2\alpha} dr. \tag{3.7}
$$

Substituting $\mathbb{E}(H) = \mathbb{E}(H^2) = 1$ in (3.6) and (3.7) concludes the proof.

■

Remarks

1. From (3.4), we notice that the average aggregate interference experienced by an MU increases with an increase in the user density. For the fixed density of RRH, the only parameter that designer can adjust to compensate for the increase in the user density is to reduce the size of the cluster. While reducing the cluster size will increase the number of RRHs activated by accommodating more users, it also reduces the interference protection available to each MU link. More specifically, the small number of large clusters or large number of small clusters may lead to a similar co-channel interference environment.

2. The average interference experienced by an MU decreases with an increase in pathloss. This follows from the fact that with an increase in pathloss, signals attenuate more rapidly and hence the aggregate interference power is reduced. However, the signal strength is also reduced for the same reason.

Proposition 3.2 (Link success probability for a scheduled MU). *The link success probability of the probe MU served under the proposed user centric clustering and RRH selection scheme algorithm can be lower-bound as follows*

$$\mathbb{P}_{suc}(\gamma_{th}, R_{cl}^2) \geq 1 - \exp\left(-\frac{\lambda_{CLR}\pi\delta}{\gamma_{th}^{\delta}\kappa_1^{\delta}}\gamma(\delta, \gamma_{th}\kappa_1 R_{cl}^{\alpha})\right), \tag{3.8}$$

where γ_{th} is the MU's desired SIR threshold, $\delta = \frac{2}{\alpha}$ and $\gamma(a,b) = \int_0^b t^{\alpha-1}\exp(-t)dt$ is the lower incomplete Gamma function.

Proof: Consider the probe MU scheduled under the proposed clustering mechanism, the link success probability for this MU is given by

$$\begin{aligned}\mathbb{P}_{suc}(\gamma_{th}, R_{cl}^2) &= \Pr\{\Gamma_{MU} > \gamma_{th}\},\\ &= 1 - \Pr\{\Gamma_{MU} \leq \gamma_{th}\},\\ &= 1 - \mathbb{E}_I[\underbrace{\Pr\{\max_{\boldsymbol{i}\in\mathcal{S}_{cop}} h_{\boldsymbol{i}} l(r_{\boldsymbol{i}}) \leq I\gamma_{th}\}}_{A_1}].\end{aligned} \tag{3.9}$$

The term $A_1 = \Pr\{\max_{\boldsymbol{i}\in\mathcal{S}_{cop}} h_{\boldsymbol{i}} l(r_{\boldsymbol{i}}) \leq I\gamma_{th}\}$ can be computed by noticing the fact the $\mathcal{S}_{cop}$ is a SPPP inside a finite area $b(\boldsymbol{o}, R_{cl})$ and we can construct a Marked PPP by assigning the fading marks to each $\boldsymbol{i} \in \mathcal{S}_{cop}$ [5]. Additional Bernoulli or indicator marks are assigned to the PPP such that the intensity of modified process[6] can be expressed as

$$\lambda_S(r, h) = \lambda_{CLR} 2\pi r \mathbb{1}(hl(r) \geq I\gamma_{th}) f_H(h). \tag{3.10}$$

Now A_1 can be computed by the void probability of the modified point process as

$$A_1 = \exp\left(- \underbrace{\int_0^{\infty}\int_0^{R_{cl}} \lambda_s(r, h) dr dh}_{\Lambda_s} \right), \tag{3.11}$$

The mean measure Λ_s can be evaluated by

$$\begin{aligned}\Lambda_S &= \lambda_{CLR} 2\pi \int_0^{R_{cl}}\int_0^{\infty} r\mathbb{1}(hl(r) \geq I\gamma_{th}) f_H(h) dr dh,\\ &\overset{(a)}{=} \lambda_{CLR} 2\pi \int_0^{R_{cl}} r Pr(h \geq I\gamma_{th} r^{\alpha}) dr,\\ &= \lambda_{CLR} 2\pi \int_0^{R_{cl}} r \exp(-I\gamma_{th} r^{\alpha}) dr,\\ &= \frac{\lambda_{CLR}\pi\delta}{\gamma_{th}^{\delta} I^{\delta}} \gamma(\delta, \gamma_{th} I R_{cl}^{\alpha}),\end{aligned} \tag{3.12}$$

where (a) follows from the CDF of the exponential function. Employing (3.11) and (3.12), we obtain

[5] A detailed discussion on the Marked PPP is beyond the scope of this article. Interested readers should refer to [67].

[6] The modified intensity corresponds to the dependently thinned point process.

$$\mathbb{P}_{suc}(\gamma_{th}, R_{cl}^2) = 1 - \mathbb{E}_I\left[\exp\left(-\frac{\lambda_{CLR}\pi\delta}{\gamma_{th}^{\delta} I^{\delta}}\gamma(\delta, \gamma_{th} I R_{cl}^{\alpha})\right)\right],$$
$$\overset{(b)}{\geq} 1 - \exp\left(-\frac{\lambda_{CLR}\pi\delta}{\gamma_{th}^{\delta}\kappa_1^{\delta}}\gamma(\delta, \gamma_{th}\kappa_1 R_{cl}^{\alpha})\right).$$

where $\kappa_1 = \mathbb{E}_I(I)$ from (3.4) and (b) follows from the Jensen's inequality.

■

The area spectral efficiency of the large scale UC-RAN is defined as the number of bits/s which can be transmitted over a unit Hertz bandwidth per second in the area of 1 square meter. In other words, the area spectral efficiency measures the amount of information that is flowing through a unit area when one Hertz of bandwidth is employed. The lower bound on the link success probability (which is equivalent to the coverage probability) can be employed to establish a lower bound on the area spectral efficiency of the UC-RAN as

$$\mathcal{T}_{CLR} = \lambda_{USR}^{\{EF\}} \log_2(1 + \gamma_{th})\mathbb{P}_{suc}(\gamma_{th}, R_{cl}^2), \tag{3.13}$$

where $\lambda_{USR}^{\{EF\}}$ is the effective density of the scheduled user defined in (3.2). As is clear from (3.13), the area spectral efficiency of the UC-RAN is strongly coupled with the cluster size. Intuitively, increasing the cluster size decreases the effective number of scheduled users. However it also increases both the SIR (due to lower number of nearby interfering RRHs) and the interference protection margin. Essentially, this implies that there exists an optimal radius for the cluster which will balance these two opposite effects to maximize the attainable area spectral efficiency.

3.5 Energy Efficiency Analysis

In previous sections, we focused on the spectral performance of the proposed UC-RAN. While the proposed user-centric mechanism exploits centralized processing

in cloud to harness the distributed diversity gains, an important issue from network operator/designers perspective is the cost associated with these gains. More specifically, from an energy consumption perspective the cost-benefit analysis can be formulated in terms of energy efficiency. The network or link level energy efficiency characterizes the number of bits that can be transmitted per unit usage of available spectrum at the expense of one Joule in one second.

Due to a large spatio-temporal variance in user traffic, energy efficiency can be significantly improved in dense urban environment through efficient ON/OFF activation [72]. In order to quantify the energy consumption-throughput trade-off, our prime focus here is the energy consumption associated in discovering the best RRH for the association. To that end, we only focus on this additional energy which is required for the discovery purpose and can be considered as the overhead incurred for harnessing the diversity gain. Note that during the discovery process, each RRH is required to estimate the channel gain from the scheduled MU which comes at the expense of energy dissipation.

The power consumption of a standalone RRH was investigated in the project EARTH [73]. The proposed power consumption model provides accurate estimates of dissipated power in different building blocks such as antenna interface, cooling, power amplifiers and baseband processing. The model was extended by parameterization with the C-RAN efficiency in [74]. In this article, our primary interest is to compute the total power consumed in the discovery process in each user cluster. Thus, we propose a modified power consumption model which is inspired by [74] and [73]. Mathematically, the power consumption of the discovery process can be quantified as:

$$P_{CRAN} = \omega_{CRAN}(\mathrm{M}, \theta) P_0 + \Delta_u P_u + P_{ou}, \tag{3.14}$$

where P_u is the transmit power employed by the MU, P_o is the fixed power consumption of the RRH in listening mode, Δ_u is the radio frequency dependent component of power consumption at the MU, $\omega_{CRAN}(\mathrm{M}, \theta)$ is the the UC-RAN coefficient and P_{ou} is the fixed circuit power consumed at the MU. The UC-RAN coefficient is coupled with the number of RRHs in each cluster (denoted by M) and a parameter θ which parameterizes the implementation efficiency. More specifically, $\omega_{CRAN}(\mathrm{M}, \theta) \leq \mathrm{M}$ captures the performance gains due to consolidated architecture of UC-RAN. The lower the value of $\omega_{CRAN}(\mathrm{M}, \theta)$, the lesser is the amount of power dissipated in each cluster. A simple parameterization of the efficiency coefficient from can be obtained as follows:

$$\omega_{CRAN}(\mathrm{M}, \theta) = \theta \mathrm{M}, \quad 0 \leq \theta \leq 1 \tag{3.15}$$

where $\theta = 1$ captures the least efficient UC-RAN implementation. The mathematical expression for determining average number of RRHs in each cluster (M) is given in Lemma1.

Lemma III.I: The average number of RRHs within each user-centric cluster, i.e. M, is the complement of the void probability of the RRHs, i.e. $\mathrm{M} = 1 - e^{-\lambda_{CLR}\pi R_{cl}^2}$.

Proof: Consider that Π_{CLR} is an SPPP with intensity λ_{CLR}, then under user-centric scheme, the average number of RRHs within a circular area of radius R_{cl} is given by $\lambda_{CLR}\pi R_{cl}^2$. Since each user-centric cluster can have at most one RRH, the average number of RRHs is the complement of the probability that an arbitrary cluster would at least one RRH within its foot-prints, i.e.

$$\begin{aligned}
\mathrm{M} &= \Pr\{\Pi_{CLR} \cap b(\boldsymbol{x}, R_{cl}) \neq \emptyset | \boldsymbol{x} \in \Pi'_{MU}\}, \\
&= 1 - \Pr\{\Pi_{CLR} \cap b(\boldsymbol{x}, R_{cl}) = \emptyset | \boldsymbol{x} \in \Pi'_{MU}\}, \\
&= 1 - \exp\{-\pi \lambda_{CLR} R_{cl}^2\}.
\end{aligned}$$

■

The average power consumption of each cluster can then be written as

$$P_{CRAN} = \omega_{CRAN}(1 - \exp\{-\pi\lambda_{CLR}R_{cl}^2\}, \theta)P_o + \Delta_u P_u + P_{ou}. \tag{3.16}$$

Notice that in this analysis we are mainly focusing on the power consumed at MU for re-broadcasting the pilot signal and the power consumed at RRHs to estimate channel from this pilot. We do not consider the power consumption at macro BS for initial transmission of pilot signal, since this cannot be regarded as an energy overhead. Such transmission is part of the macro BS operation even in the traditional cellular networks.

The network wide average energy efficiency is defined to be as the ratio of sustainable throughput for each scheduled MU and the average power consumption times the number of scheduled users. Mathematically

$$\eta_{EE} = \frac{\mathrm{B}\log_2(1 + \Gamma_{CRAN}^I)}{\omega_{CRAN}(1 - \exp[-\lambda_{CLR}\pi R_{cl}^2], \theta)P_o + \Delta_u P_u + P_{ou}}, \tag{3.17}$$

where B is the employed bandwidth (assumed unity for subsequent discussion) and Γ_{CRAN}^I is the effective SIR defined in (18).

Remarks:

1. The per user throughput scales as $\mathcal{O}(\lambda_{CLR}^{1/\delta-1})$ while the average power consumption of each cluster involved in discovery process scales as $\mathcal{O}(1-e^{-\lambda_{CLR}})$. This implies that both the user throughput and the power consumption are increasing functions of the RRH density. However, as λ_{CLR} increases, the power consumption quickly saturates to:

   $P_o + \Delta_u P_u + P_{ou}$ as $\omega_{CRAN}(M, \theta) \to 1$.

2. Similar to RRH density, it is obvious that the throughput and power consumption are monotonically increasing functions of the cluster radius (R_{cl}). Due to the saturation of the power consumption though, the optimal cluster radius which maximizes the energy efficiency of the UC-RAN would be the maximum possible cluster size as per network operator's design specifications.

3. These two observations lead to an important design question, i.e., how different the EE-optimal cluster size is as compared to the cluster radius which maximizes the network wide area spectral efficiency? Also, since an energy efficient design would prefer a larger cluster size compromising on the area spectral efficiency, is there a way to work out a balance between these two parameters. The rest of our discussion will be formed across this design issue.

3.6 SON Framework for RRH Cluster Size Optimization

The GT based SON engine is embedded within the centralized BBU pool for real-time adjustment of R_{CLR} to optimize a system level efficiency parameter of interest with respect to terrain environment, user demographics, RRH deployment scenario and network operator's spatio-temporal revenue model (see Fig.3.3). The variation in the cluster size models the dynamic tradeoff between ASE and EE in our bargaining game model. The proposed GT-SON framework with the sequence of steps in dynamic cluster size adjustment for modeling the ASE-EE tradeoff is given in Fig. 3.3.

To analytically express the ASE-EE tradeoff, we formulate a two-player cooperative bargaining game where both ASE and EE are modelled as virtual game players

that independently estimate the best cluster size for maximizing their respective utility functions. We will see later that due to a large dissimilarity in cluster size preferences of the players, each player's payoff is affected by the cluster size selection made by the other player. However, both players can mutually benefit through the cooperative game where they negotiate for the R_{CLR} that achieves optimal ASE-EE tradeoff. Using Nash's axiomatic model, it is well known that the Nash bargaining solution (NBS) achieves a pareto-optimal solution, i.e. the optimal tradeoff in the utilities of the players in such cooperative games [46].

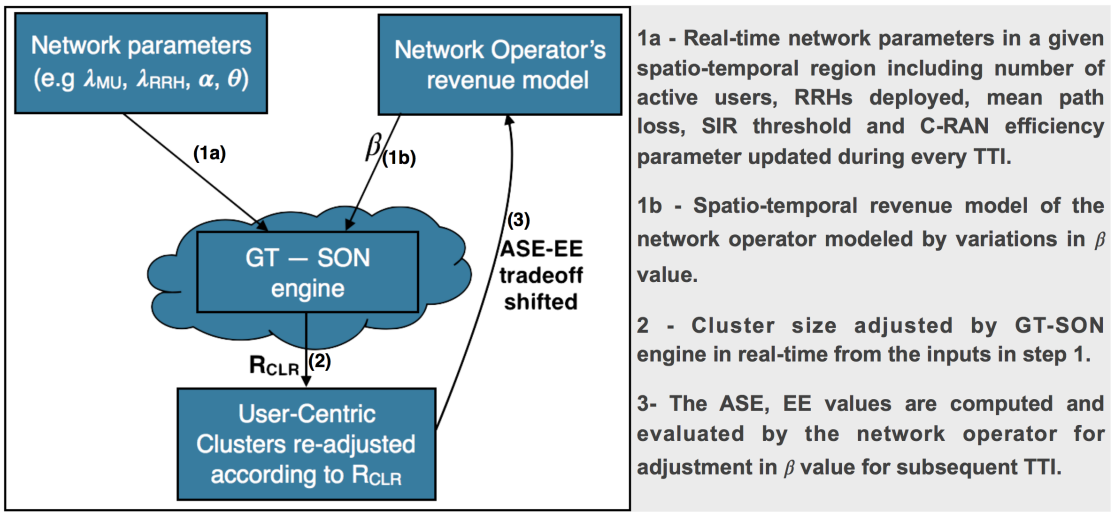


Fig. 3.3: GT-SON Framework in UCRAN

If the players can be denoted by the set $\mathbb{N} = \{1, 2\}$, where player $i = 1$ denotes ASE, player $i = 2$ denotes EE and S_i denotes the set of all feasible payoffs to an arbitrary MU i as

$$S_i = \{U_i | U_i = U_i(R_{CLR}), R_{CLR} \in \mathbb{R} : R_{CLR} > 0\}. \tag{3.18}$$

Let us define the space $\mathcal{S}$ as the set of all feasible payoffs that players $i \in \mathbb{N}$ can achieve when they collaborate, i.e.

$$\mathcal{S} = \{U = (u_1, u_2) | u_1 \in S_1, u_2 \in S_2\} \tag{3.19}$$

where $u_1(x_1)$ is the utility of the first player and $u_2(x_2)$ is the utility of the second player such that

$$s_1 = u_1(x_1) = [\mathcal{AS}\mathcal{E}(R_{CLR})]^{\beta}, \tag{3.20}$$

$$s_2 = u_2(x_2) = [\mathcal{E}\mathcal{E}(R_{CLR})]^{1-\beta} \tag{3.21}$$

and $x_1 = x_2 = R_{CLR} \in \mathbb{R} : R_{CLR} > 0$. $\beta \in [0, 1]$ is the exponential bias factor in NBS that defines the bargaining power (or the tradeoff) division between the two players. We also define the disagreement space $\mathcal{D} \in \mathcal{S}$ as the set of the two disagreement points $d = (d_1, d_2)$ where $d_1 = u_1(\mathcal{D})$ and $d_2 = u_2(\mathcal{D})$ represent the payoffs for the two players if the bargaining process fails and no outcome is reached. For our game, we set $d = (0, 0)$ thus giving both players uniform leeway to improve their utilities. [18] shows that the NBS in such parametric cooperative games exists only if the utility functions for the players form convex and compact sets.

Proposition 3.3. *The utility and disagreement spaces in the proposed GT-SON framework constitute a two-player bargaining problem defined by* $(\mathcal{S}, d)$ *where* $\mathcal{S} \in \mathbb{R}^2$, $d \in \mathcal{S}$ *and the resulting unique bargaining outcome is pareto-optimal.*

Proof: A bargaining problem can be defined as the pair $(\mathcal{S}, d)$ if: i) $\mathcal{S}$ is a convex and compact set, ii) There exists some $s \in \mathcal{S}$ such that $s > d$, i.e. $s_1 > d_1$ and $s_2 > d_2$. It is quite obvious that $\mathcal{S}$ is compact and since $d = (0, 0)$, positive utilities for our players satisfies the 2nd condition. This leaves behind the question whether $\mathcal{S}$ is convex which holds true if: $\forall \epsilon : 0 \leq \epsilon \leq 1$, if $U^a = (u_1^a, u_2^a) \in S_1$ and $U^b = (u_1^b, u_2^b) \in S_2$, then $\epsilon U^a + (1 - \epsilon)U^b \in \mathcal{S}$. From (3.13), we see that $\epsilon u_1^a + (1 - \epsilon)u_1^b = [\bar{\lambda}_{MU} \log_2(1 + \gamma_{th})\bar{P}]^{\beta}$ where $\bar{P} = [\epsilon(\mathbb{P}_{suc}^a)^{\beta} + (1 - \epsilon)(\mathbb{P}_{suc}^b)^{\beta}]$ and

since we know that $0 \leq \mathbb{P}^a_{suc}, \mathbb{P}^b_{suc}, \beta \leq 1$, the sum in (3.22) forms a convex set, i.e.

$$\epsilon u_1^a + (1-\epsilon)u_1^b \in S_1. \tag{3.22}$$

Similarly, from (3.17), we see that $\epsilon u_2^a + (1-\epsilon)u_2^b = [\frac{\bar{\lambda}_{MU}\log_2(1+\gamma_{th})\bar{P}}{\lambda_{RRH}p_{\bar{ACT}}(M\theta P_0+\Delta_u P_u)}]^{(1-\beta)}$, where the numerator is convex from (3.22) and denominator is convex since $p_{\bar{ACT}} = \epsilon(p^a_{ACT})^{1-\beta} + (1-\epsilon)(p^b_{ACT})^{1-\beta}$ and $0 \leq p^a_{ACT}, p^b_{ACT}, \beta \leq 1$. Therefore,

$$\epsilon u_2^a + (1-\epsilon)u_2^b \in S_2. \tag{3.23}$$

From (3.22) and (3.23), we conclude that $\epsilon U^a + (1-\epsilon)U^b \in \mathcal{S}$ which satisfies the conditions for convexity for set $\mathcal{S}$. According to Nash's axiomatic approach [46], there exists a unique solution for the two-player bargaining problem which is the pair of utilities (s_1^*, s_2^*) that solves the following optimization problem:

$$\max_{(s_1,s_2)} (s_1-d_1)(s_2-d_2), (s_1,s_2) \in \mathcal{S} \geq (d_1,d_2). \tag{3.24}$$

■

Proposition 3 implies that for an arbitrary MU $\boldsymbol{x}$, the optimal cluster size "$R^{opt}_{CLR,x}$" is obtained through the solution of a convex optimization problem (also known as Nash Product (NP)) which for our model can be given by

$$R^{opt}_{CLR,\boldsymbol{x}} = \max_{R_{CLR,\boldsymbol{x}}} [\mathcal{AS}\mathcal{E}(R_{CLR,\boldsymbol{x}})]^{\beta}[\mathcal{E}\mathcal{E}(R_{CLR,\boldsymbol{x}})]^{1-\beta}. \tag{3.25}$$

Notice that the computational complexity of the GT-SON engine is a function of the cluster size granularity, i.e. $\mathcal{O}(N_{CLR})$ where N_{CLR} denotes the number of distinct cluster sizes over which the optimization in (3.25) is performed. As the processing times are independent of MU or RRH densities, real-time implementa-

tion of the GT-SON optimization framework is practically realizable and scalable throughout the network.

3.7 Performance Analysis

In this section, we discuss the analytical trends and Monte Carlo simulation results by employing a 3GPP standard compliant LTE network simulator. The first tier elasticity is performed by optimization of the mathematical developed in Chapter 3.6. For simulation, we consider a two tier HetNet with a tri-sector hexagonal MBS of radius 500 m. We consider a single sector of the MBS covering an area of 73850 m^2 where MUs and small cell RRHs are uniformly distributed according to their independent SPPPs. Without loss of generality, the channel power gains between all MUs and RRHs are assumed unity. We assume uniform transmit power of 30 dBm for all RRHs. Other power consumption parameters are taken from [18]. Simulation results are averaged over 1000 Monte Carlo trials.

3.7.1 ASE, EE performance under varying β

Table 3.1: Cluster sizes and efficiency loss at different bargaining weightage

	β	R_{CLR}^{NBS}(m)	$\%ASELoss$	$\%EELoss$
A	0,0.25,0.5,0.75,1	51,33,22,13,6	87,71,45,19,0	0,9,36,68,91
B	0,0.25,0.5,0.75,1	22,15,9,6,5	84,66,39,6,0	0,9,36,64,77
C	0,0.25,0.5,0.75,1	42,28,19,12,6	81,60,38,17,0	0,8,33,67,92
D	0,0.25,0.5,0.75,1	19,12,9,6,5	78,53,28,6,0	0,8,31,54,62

From the analytical results in (3.13), (3.17) and (3.25), we investigate the variation in the optimal cluster size and the efficiency metrics as β is shifted between ASE-optimal ($\beta = 1$) and EE-optimal ($\beta = 0$) points. The GT-SON engine optimizes R_{CLR} on the following fixed network parameters: $\lambda_{MU} = 10^{-2}/\text{m}^2$, $\theta = 0.5$,

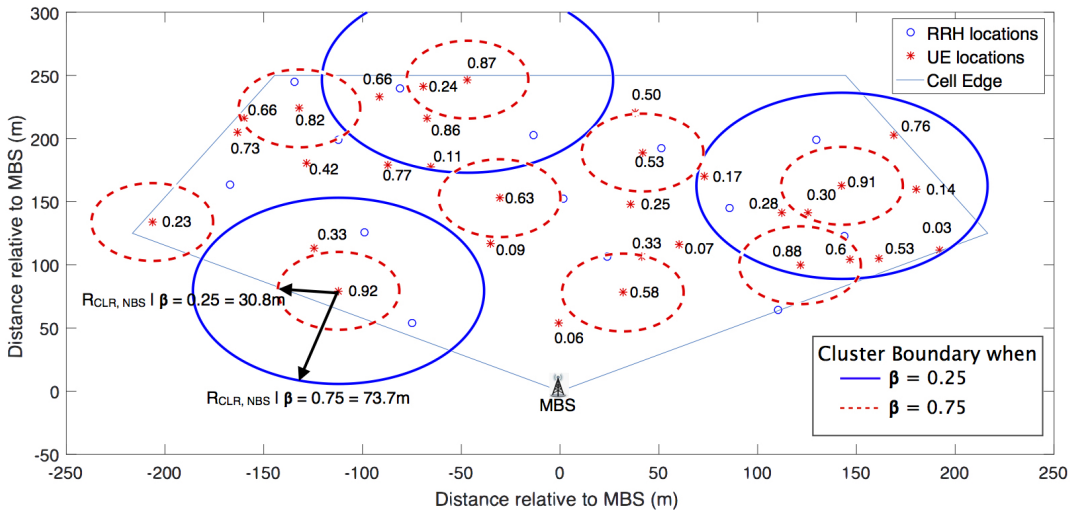


Fig. 3.4: User-centric RRH clustering under different β values.

$\gamma_{th} = 4$ dB, and $0 < R_{CLR} \leq 100$ m. An illustration of the user-centric clusters when different β values are employed is given in Fig. 3.4. Table 3.1 shows that while lower β values result in larger user-centric clusters corresponding to high average SINR, higher β ensures that moer number of users are simultaneously scheduled reducing the average latency for a UE.

The ASE results in Fig. 3.5(a) indicate around the same ASE-optimal cluster size of 5m for variations in pathloss exponent and RRH deployment densities. It is seen that higher RRH densities yield superior system throughput which is understandable considering p_{ACT} is expected to increase with λ_{RRH}. It is also noted that $\alpha = 4$ yields more than two-fold increase in ASE as compared to $\alpha = 3$. Since mmWave network propagation studies [75] have indicated higher pathloss due to blocking effects, the UCRAN is expected to yield more system capacity at mmWave spectrum by virtue of relatively larger MU-interfering RRH distances. EE results in Fig. 3.5(b) indicate optimal R_{CLR} to be the highest possible cluster size because of the combined effect of increased throughput and reduced power consumption with increase in R_{CLR}. Like ASE, the maximum EE is achieved

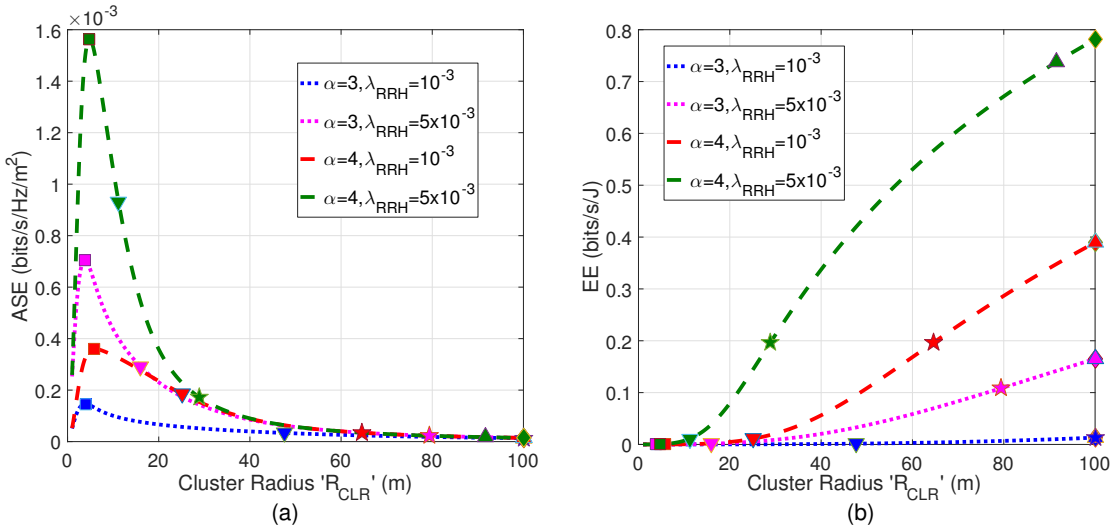


Fig. 3.5: (a) ASE and (b) EE at different β values when varying the λ_{RRH} and α. The NBS for each use case is shown separately for ASE and EE at β values of 0, 0.25, 0.5, 0.0.75 and 1 and denoted by ♦, ▲, ★ , ▼ and ■ respectively.

at higher RRH density and pathloss exponents. This implies that the GT-SON engine will most effectively utilize the ASE-EE tradeoff with gain variations of over 100% through appropriate β adjustment in ultra-dense mmWave networks.

3.7.2 User QoE Performance Analysis

Users' QoE analysis is conducted through SINR distribution between MUs with network parameters: $\lambda_{MU} = 10^{-2}/\text{m}^2$, $\lambda_{RRH} = 10^{-3}/\text{m}^2$, α=4, $\theta = 0.5$, $\gamma_{th} = 4$ dB and bandwidth B=1 Hz. Both the MU and RRH deployments are performed using uniform PPPs and average performance results are obtained via Monte Carlo simulations. We use two variants of the proposed user-centric approach: i) RRH cluster size deployment that maximizes ASE henceforth referred as UC(ASE), and ii) cluster size deployment that maximizes EE henceforth referred as UC(EE). To compare the performance with a standard non user-centric PPP deployment, we follow the approach in [8] and represent it as NUC. Results in Fig. 3.6 show that even with the most data throughput efficient user-centric design, we obtain a SINR

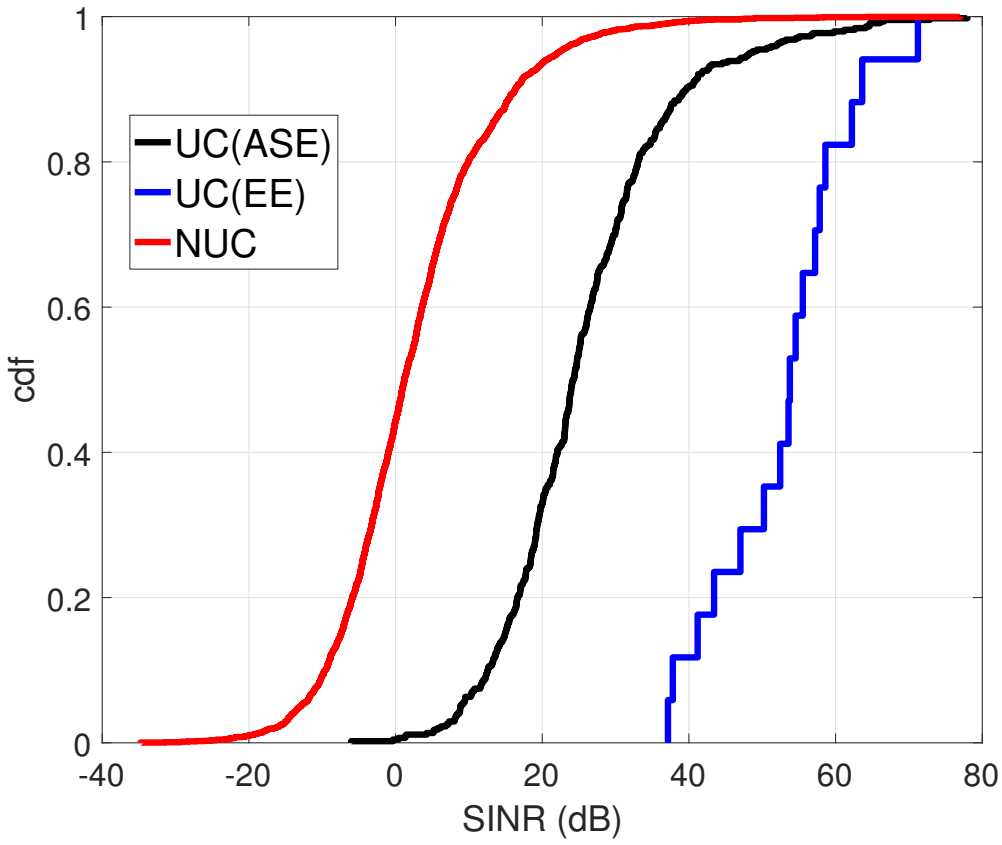


Fig. 3.6: Downlink SINR cdf comparison between user-centric and non user-centric approaches

gain of over 20 dB for almost 50% of the users. The ruggedness in the cdf graph of UC(EE) in comparison to the other two CDFs is because of lower number of users in the thinned PPP $\bar{\lambda}_{MU}$ which is a direct consequence of the larger cluster sizes in EE optimization. The 5 percentile SINR performance (for the cell-edge users with worst SINR in conventional networks) is also significantly improved for user-centric approaches with about 20 dB and 40 dB gain with UC(ASE) and UC(EE) respectively. Clearly the user-centric approach eliminates cell-edge degradation and guaranteed high QoE for every user regardless of its physical location.

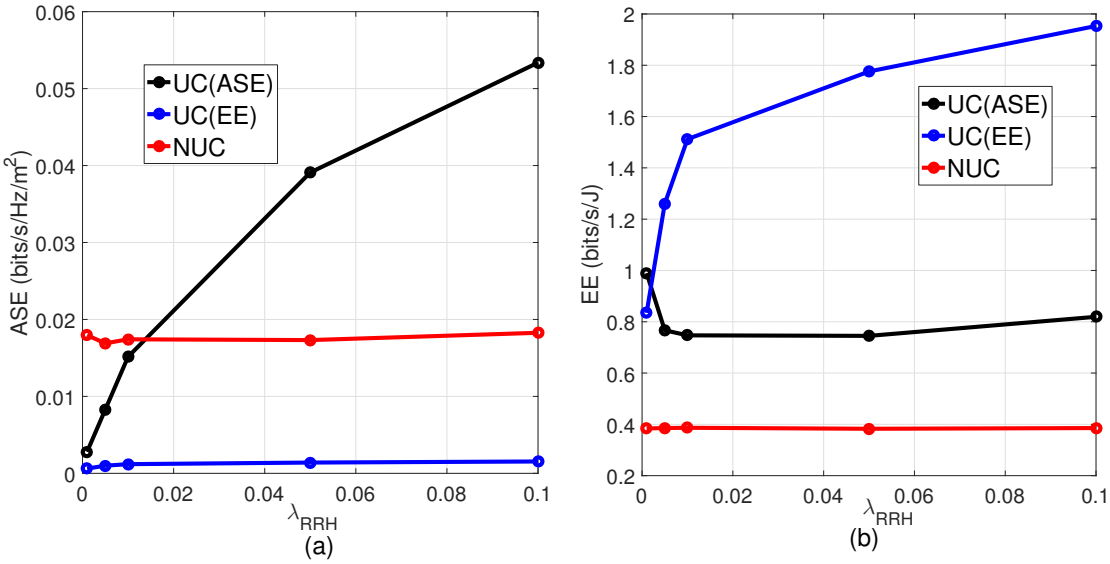


Fig. 3.7: (a) ASE and (b) EE comparison of UC(ASE), UC(EE) and NUC approaches with different RRH densities

3.7.3 Performance comparison with non user-centric networks

Fig. 3.7 compares the system wide ASE and EE of the user-centric approaches with the baseline scheme at different RRH densities and $\lambda_{MU} = 10^{-2}/\text{m}^2$, α=4, $\theta = 0.5$ and $\gamma_{th} = 4$ dB. Fig. 3.7(a) reveals that as the RRH deployment density increases, UC(ASE) emerges as the most data efficient scheme. While NUC exhibits uniform ASE, UC(ASE) by virtue of increased $\mathbb{P}_{suc}$ exhibits highest system capacity. On the other hand, UC(EE), though not throughput efficient by any regards, yields more than 5 times power efficient network as compared to NUC approach (fig. 3.7(b)). This observation highlights the inherent ASE-EE tradeoff available to the network operator by adjusting β via the GT-SON and choosing the appropriate RRH cluster size.

3.7.4 Scheduling rate under varying RRH and UE densities

In order to study the impact of dense RRH deployment, Fig. 3.8 depicts the mean scheduling success rate for the MUs under different β and deployment density

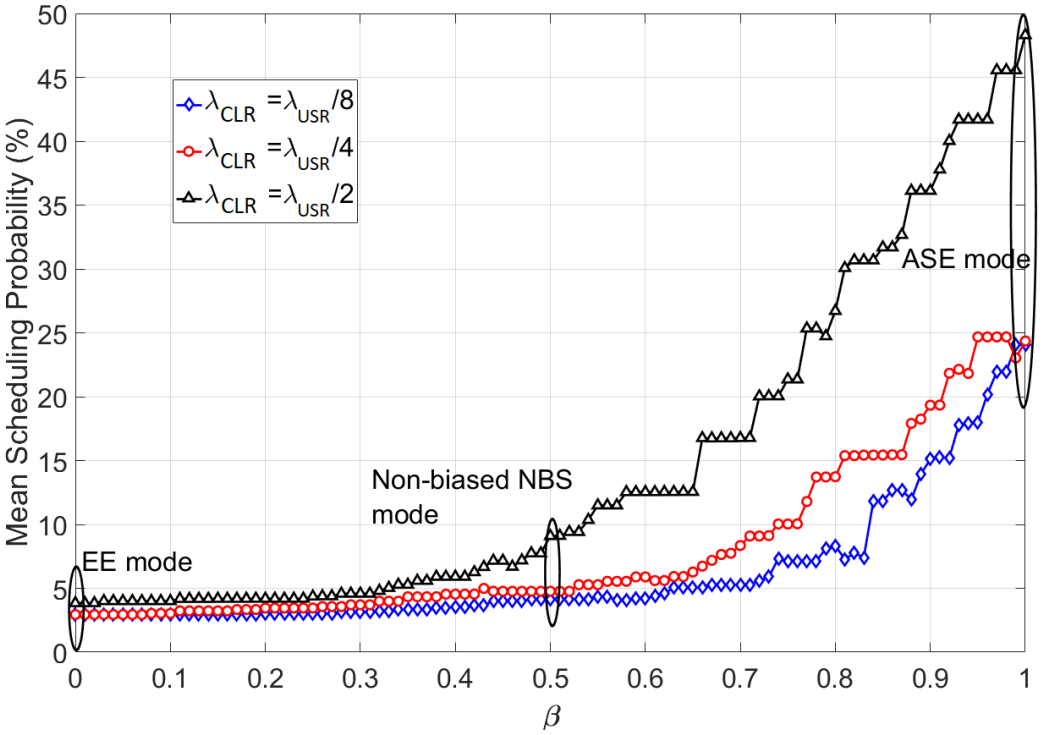


Fig. 3.8: Mean scheduling success rate versus β under different $\lambda_{CLR}/\lambda_{USR}$.

$(\frac{\lambda_{CLR}}{\lambda_{USR}})$ scenarios. For each scenario, we simulate 1000 consecutive transmission time intervals (TTIs) for $\lambda_{USR} = 4\text{x}10^{-4}$. Using user-centric RRH clustering (Chapter 3.2), we update the p_{USR} and RRH clusters for each TTI which allows us to estimate the mean number of MUs that are scheduled for DL transmission under each simulation scenario. The results in Fig. 5.5 show rapid increase in the scheduling probability as we move towards ASE regime. However, we notice that as RRH deployment density increases to $\lambda_{USR}/2$, the scheduling success doubles to 48% as compared to 24% in case of $\lambda_{USR}/4$ and $\lambda_{USR}/8$ for $\beta = 1$. This implies that increasing small cell deployment density in ASE mode allows higher number of concurrent DL transmissions that consequently reduces main scheduling delay for an arbitrary user in UC-RAN.

3.8 Conclusion

In this chapter, we developed a comprehensive statistical framework for computation of the area spectral and energy efficiency of a large-scale user-centric cloud radio access network (UC-RAN). We proposed a user-centric RRH clustering algorithm which enables dynamic coverage extension and shrinkage by activating a single remote radio head within a specified demarcation around a scheduled user based on max SIR gain criteria. The user-centric architecture ensures uniform coverage and no cell-edge degradation for all the users irrespective of their physical location.We demonstrated that there exists an optimal cluster radius which maximizes the area spectral efficiency of the UC-RAN. It was also demonstrated that this optimal cluster radius is coupled with the user density and hence must be adapted by a self-organization mechanism.

The link level performance was then employed to perform cost-benefit analysis of the proposed protocol. More specifically, the amount of power dissipated in the association process under the proposed protocol is considered as the cost of obtaining the throughput gains. The throughput-cost ratio is hence essentially the energy efficiency of the UC-RAN. It was shown that there exists another optimal cluster radius which maximizes the energy efficiency of the UC-RAN. However, this is larger to the one obtained under area spectral efficiency criterion. Consequently, the throughput-energy consumption trade-off manifests itself in terms of dimensioning of the cluster radius in UC-RAN. Using a game theoretic framework, we demonstrated that a SON engine within the centralized BBU pools may be employed to dynamically configure the optimal cluster size. Simulation results indicated that: i) the SON mechanism allows more than 100x efficiency variation through real-time adjustment in the NBS bias parameter, ii) in comparison to

current always-ON RRH deployments, selective RRH activation in UC-RAN offers high area spectral and energy efficiency gains, particularly when $\lambda_{CLR} > \lambda_{USR}$, and iii) significant SINR gains can be realized in both ASE and EE operating modes by virtue of interference-free RRH cluster zones around each scheduled user.

CHAPTER 4

QoE based Elastic user-centric networks

In the chapter 3, we discussed the macro level elasticity in user-centric networks. The main question we will discuss in this chapter is what if UEs have different data demands, which is the actual case in real networks. Will a one size fit all scheme work better for the overall network performance? In this chapter, we introduce the idea of a 2nd tier elastic user-centric cellular networks, where the optimization for user-centric service zones is carried out individually for all UEs. In summary, we will model elastic user-centric service zones with variable sizes to model the disparity in the data demands of the UEs. Game theory would be applied in tandem with concepts from stochastic geometry to model the bargain of the S-Zone modeling between UEs.

4.1 Introduction

4.1.1 Background

Network densification through the use of small cells is considered a viable solution for meeting the capacity targets in future cellular networks (i.e. 5G). While densification is inevitable, it has a couple of major associated problems that persist as a bottle neck in network planning: i) high inter-cell interference and ii) low energy efficiency [1]. Though upcoming 5G technologies such as massive multiple input multiple output (MIMO) and millimeter wave (mmWave) offer promising prospects for increased system capacity, they may not address the aforementioned

issues. Redesigning the network orchestration in cell planning from the traditional base station (BS)-centric to a user equipment (UE)-centric approach [2] has been recently envisioned as a first step to address these challenges [62]. This user-centric (UEC) architecture guarantees a higher energy efficiency (EE) along with location-independent uniform Quality of Experience (QoE). Analysis in [18] has shown that the cell density that yields optimal EE is different than that which yields maximum outage capacity. In [4], we leverage this analysis to propose a UEC network deployment solution where an ultra-dense network is deployed, and is then orchestrated between EE and area spectral efficiency (ASE) optimal modes by intelligently switching OFF/ON small cells. This switching ON and OFF is done by creating non-overlapping exclusion zones around high priority UEs within which only one small cell is turned ON during each scheduling instant. The size of the exclusion zone is then used as a control parameter to realize the desired compromise between EE and ASE. However, due to diversity in mobile data usage trends, a static network wide optimal exclusion zone size does not offer the elasticity to optimize individual user's QoE in different spatio-temporal zones.

4.1.2 Architectural Overview

In this chapter, we present and introduce a second tier of elasticity within UEC systems that integrates non-uniform exclusion zones centered around users. These non-uniform exclusion zones, which we call as service zones (S-Zones), cater for data demand disparity between spatio-temporal zones as well as the diversity of data requirements from user applications (for instance HD video streaming v/s whatsapp messaging) within a single spatio-temporal zone. The basic premise behind deploying virtual flexible service zones around mobile users is to control the interference limit that a user can experience while still getting the throughput

sufficient for its data needs. For instance, high definition real-time gaming applications will require a high throughput and low latency communication link, which can be guaranteed if the signal to interference ratio (SINR) is sufficiently high. To ensure that, the controller will assign this user a large service zone to not only assign it larger number of resources, but also to reduce interference from concurrent downlink transmissions for other users. The same is true other way round for IoT based sensor devices that require low throughput transmission, and hence a relatively moderate SINR. Consequently, a smaller service zone would suffice for their data requirements.

This demand based UEC scheme is an ideal candidate for implementation in a control-data separation architecture (CDSA) [76] where small cells referred to as data base stations (DBSs) provide data services to UEs while macro base stations also referred to as control base stations (CBSs) provide necessary control and signaling. While CBSs provide the essential coverage, intelligent activation/deactivation of the DBSs enables potential for significant energy savings in CDSA. In addition to this, CDSA can offer better spectral efficiency mainly because of selection diversity that stems from large number of DBSs in dense deployments. Centralized coordination at the CBS solves the cell discovery problem for DBSs in a conventional BS-centric architecture. In the proposed UEC framework, this allows for turning DBSs ON/OFF, depending upon an individual UE's S-Zone size and the propagation link quality between that UE and the DBSs within its S-Zone.

The analysis of strategies UEs may adapt while competing for downlink (DL) resources to meet their data requirements in a UEC CDSA is the focus of this chapter. To this end, we investigate the application of game theoretic techniques which have been well known for resource management and interference mitiga-

tion in dense small cell networks [77] [52]. In particular, we apply auction and evolutionary game techniques (referred to as AGT and EGT respectively), with users as game players adapting strategies to secure DL scheduling within virtually interference free S-Zones. While there has been significant research in user-centric networks in recent times [78],[79], to the best of our knowledge, analysis of second tier elasticity in user-centric CDSA that caters to non-uniform user demands remains a terra incognita. This work is a first attempt to analyze the tradeoff of game theoretic techniques for UE level demand based CDSA architecture.

4.1.3 Research Objectives

Our prime objective in this part of the dissertation is to evaluate the performance of distributive game theory models in the context of resource allocation for dense user-centric networks. Using the CDSA as the enabler for the technological framework, we present a system model that links the activation of DBSs to user requirements as well as the level of interference in the environment surrounding the scheduled users. The research questions that this chapter aims to answer are the following:

- What is a plausible utility model for distributed games in user-centric networks?
- How will the user scheduling look like for downlink scheduling in demand oriented user-centric architectures?
- Can we identify game models that yield higher network efficiency than its competitors, and under what network conditions do they show superior performance?

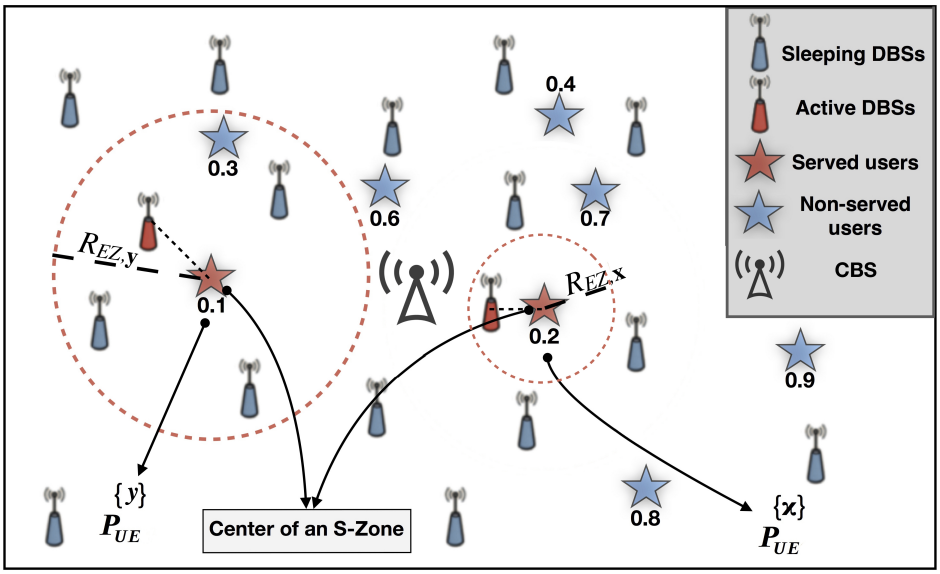


Fig. 4.1: S-Zones concept illustration in UE-centric CDSA architectures.

To answer these questions, a detailed comparative analysis is presented using two game theoretic techniques, namely evolutionary game and auction theory. Both the games are based on a distributed utility minimization problem. The evolutionary game involves iterative action strategy adjustment by the UEs, whereas the auction game comprises UEs bidding their true valuation with the aim of winning the auction and securing virtual S-Zones for DL scheduling.

4.2 Network Model

Fig. 4.1 shows a UE-centric based CDSA network model where users having a high payoff (used interchangeably with utility) are served by a single DBS providing best channel quality (e.g. CQI (channel quality indicator) measure) within their virtual service zones respectively. Each scheduled UE is the center of an S-Zone with an active DBS. The remaining DBSs within and outside the S-Zones are turned OFF to reduce inter-cell interference as well as lower overall power consumption. The area of S-Zones around scheduled UEs is adjustable based on individual UE's

throughput requirements. This two-tier CDSA with elastic service zone model consists of a central CBS providing essential control and signaling functionalities to the UEs while the DBSs serve the UEs with DL data transmissions. Based on the channel feedback by the UEs, the CBS allocates S-Zones to scheduled users based on outcome of game models (Section 4.3) during each transmission time interval (or TTI).

Borrowing from well established tools in stochastic geometry [67], we model the spatial distributions of DBSs and UEs using two independent stationary Poisson point processes (SPPPs): $\Pi_{\text{DBS}} \in \mathbb{R}^2$ and $\Pi_{\text{UE}} \in \mathbb{R}^2$ with intensities λ_{DBS} and λ_{UE} respectively. Specifically, at an arbitrary time instant the probability of finding $n_i \in \mathbb{N}, i \in \{DBS, UE\}$ DBSs/UEs inside a typical macro-cell with area foot-print $\mathcal{A} \subseteq \mathbb{R}^2$ follows the Poisson law with mean measure $\Lambda_i(\mathcal{A}) = \lambda_i v_2(\mathcal{A})$ [62]. The mean measure is characterized by the average number of DBSs/UEs per unit area (i.e., $\lambda_{\text{DBS}} \backslash \lambda_{\text{UE}}$) and the Lebesgue measure [67] $v_2(\mathcal{A}) = \int_{\mathcal{A}} d\boldsymbol{x}$ on $\mathbb{R}^2$, where if $\mathcal{A}$ is a disc of radius r then $v_2(\mathcal{A}) = \pi r^2$ is the area of the disc.

4.2.1 Channel Model

The channel between a DBS $\boldsymbol{x} \in \Pi_{\text{DBS}}$ and an arbitrary UE $\boldsymbol{y} \in \Pi_{\text{UE}}$ is modeled by $h_{\boldsymbol{xy}} l(||\boldsymbol{x}-\boldsymbol{y}||)$. Here $h_{\boldsymbol{xy}} \sim \mathcal{E}(1)$ is a unit mean exponential random variable which captures the impact of a Rayleigh block-fading channel between $\boldsymbol{x}$ and $\boldsymbol{y}$. The small-scale Rayleigh fading is complemented by a large-scale path-loss modeled by $l(||\boldsymbol{x} - \boldsymbol{y}||) = K||\boldsymbol{x} - \boldsymbol{y}||^{-\alpha}$ power-law function. $||\boldsymbol{x} - \boldsymbol{y}||$ is the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{y}$, K is a frequency dependent constant and $\alpha \geq 2$ is an environment/terrain dependent path-loss exponent. The fading channel gains are assumed to be mutually independent and identically distributed (i.i.d.). Without any loss of generality, we assume $K = 1$ for the rest of this discussion. Furthermore,

we assume that all DBSs employ the same transmit power.

4.2.2 UE-Centric DBS Scheduling

The first step in DBS scheduling is identification of high priority users to be scheduled in a given TTI. The second step is creation of non-overlapping circular S-Zones centered around each UE selected to be served in that TTI. The size of the S-Zone is the parameter of optimization and will be discussed later. Each UE is then served by a DBS within its S-Zone that provides strongest received signal power. The remaining DBSs are switched off. The size of circular S-Zone around a scheduled UE $\boldsymbol{x}$ is characterized by a variable radius $R_{\mathrm{SZ},\boldsymbol{x}}$ which is a function of $\boldsymbol{x}$'s data requirements and the interference from nearby active DBSs. The CBS is responsible for control signaling to all the UEs within its footprints. In addition, the CBS also assigns scheduling priorities in the form of a mark/tag $p_{\mathrm{UE}} \sim \mathcal{U}(0,1)$ to each UE. The marked PPP [67] formed as a result of user-centric scheduling impacts the downlink scheduling priority of the UEs. More specifically, the lower the value of the mark, the higher is the priority of the UE to be scheduled. Effectively, these marks can be thought of as the timers corresponding to each UE that are decremented in each time slot where DL service to this UE is deferred. Based on the channel quality measure between the DBSs and the UEs, the CBS decides and activates the relevant DBSs for DL transmission to the scheduled UEs. The advantages of such UEC scheduling is two-fold: firstly, due to non-overlapping S-Zones, the interference experienced by a scheduled UE is considerably reduced and secondly, on-demand activation of DBSs provides the network self-organizing capabilities to cope with spatio-temporal variations in user demography.

One might argue that such a one-to-one UE-DBS association within a non overlapping user-centric S-Zone scheme may result in service holes, i.e. there may exist

UEs that are not associated with any DBSs due to empty UE-centric S-Zones. Since we are considering dense small cell deployments with λ_{DBS} and λ_{UE} of the same order, UE-centric S-Zones with realistic areas will hardly be void. In the unlikely scenario of a void S-zone though, user clustering strategies [68] may be employed where nearby UEs are grouped together and optimization is performed on the UE clusters rather than individual UEs. Furthermore, it is known that best DBS activation with a proximity constraint provides dual benefits of low outage probability and high power efficiency in dense deployment scenarios [69].

4.3 Payoff functuon and S-Zone Modeling

Payoff function: We model the payoff of a UE $\boldsymbol{x}$ as $u_{\boldsymbol{x}} = \delta(\bar{\tau} - \tau_{\boldsymbol{x}}(R_{\mathrm{SZ},\boldsymbol{x}})) + (1-\delta)(\tau_{d,\boldsymbol{x}} - \tau_{\boldsymbol{x}}(R_{\mathrm{SZ},\boldsymbol{x}}))$; where $\tau_x(R_{\mathrm{SZ},\boldsymbol{x}})$ is the achievable data throughput for $\boldsymbol{x}$, $\bar{\tau} = \frac{1}{N}\sum_{n=1}^{N}\tau_{\boldsymbol{n}}$ is the mean achievable user throughput (considering N active UEs) of the spatio-temporal zone estimated through a central entity (such as the CBS) and $\tau_{d,\boldsymbol{x}}$ is $\boldsymbol{x}$'s variable throughput demand based on the application usage. $0 \leq \delta \leq 1$ is a weight priority index, controllable via CBS, with factor δ enforcing uniform throughput regardless of service demand disparity while $1-\delta$ allows users with high scheduling priority to selfishly meet their data demands at the expense of non-priority users.

The first component of the payoff measures the utility of a UE in terms of the penalty (positive / negative) depending upon how much lesser / greater the UE's achievable throughput is to the mean achievable throughput of all UEs. Similarly, the second component of the payoff determines the penalty associated with how deviant the achievable throughput is to the UE's actual service based throughput demand at a given S-Zone radius. Using this novel characterization of payoff, we can formulate the optimization problem to be solved by the game models for a UE

$\boldsymbol{x}$'s payoff as a function of $R_{\mathrm{SZ},\boldsymbol{x}}$ and given as

$$\min_{R_{\mathrm{SZ},\boldsymbol{x}}} |\delta(\bar{\tau} - \tau_{\boldsymbol{x}}(R_{\mathrm{SZ},\boldsymbol{x}})) + (1-\delta)(\tau_{d,\boldsymbol{x}} - \tau_{\boldsymbol{x}}(R_{\mathrm{SZ},\boldsymbol{x}}))|. \tag{4.1}$$

S-Zone radius: The achievable throughput for a UE $\boldsymbol{x}$ is expressed using Shannon's theorem as

$$\tau_x(R_{\mathrm{SZ},\boldsymbol{x}}) = \log_2(1 + \mathrm{SINR}(R_{\mathrm{SZ},\boldsymbol{x}})), \tag{4.2}$$

where the S-Zone size dependent received signal to interference ratio (SINR) at $\boldsymbol{x}$ when served by DBS $\boldsymbol{y}$ can be written as follows:

$$\mathrm{SINR}(R_{\mathrm{SZ},\boldsymbol{x}}) = \frac{max_{\boldsymbol{y}\in\Pi_{\mathrm{DBS}}\cap(\boldsymbol{x},R_{\mathrm{SZ},\boldsymbol{x}})} h_{\boldsymbol{x}\boldsymbol{y}} l(||\boldsymbol{x}-\boldsymbol{y}||)}{N_o + \sum_{\boldsymbol{z}\in\Pi_I} h_{\boldsymbol{x}\boldsymbol{z}} l(||\boldsymbol{x}-\boldsymbol{z}||)}. \tag{4.3}$$

$\Pi_{\mathrm{DBS}} \cap (\boldsymbol{x}, R_{\mathrm{SZ},\boldsymbol{x}})$ is the thinned PPP representing the DBSs within the UE-centric virtual circular cell of area $\pi R^2_{\mathrm{SZ},\boldsymbol{x}}$ around $\boldsymbol{x}$, Π_I denotes the thinned PPP of interfering DBSs, i.e. the active DBSs in S-Zones other than that centered around $\boldsymbol{x}$, and N_o is the variance of the additive white Gaussian noise at $\boldsymbol{x}$.

To ensure that the UE-centric S-Zones are within practical dimensions, we tried several mathematical formulations of $R_{\mathrm{SZ},\boldsymbol{x}}$ as a function of λ_{DBS}, λ_{UE} and $\gamma_{\boldsymbol{x}}$. Drawing insights from extensive simulation based experiments with different DBS and UE densities, we propose the following model to characterize the S-Zone area around a UE $\boldsymbol{x}$ served by a DBS $\boldsymbol{y}$[1]:

$$R_{\mathrm{SZ},\boldsymbol{x}} = \frac{||\boldsymbol{x}-\boldsymbol{y}|| \ln(\lambda_{\mathrm{DBS}})}{(1-\gamma_{\boldsymbol{x}}) \ln(\lambda_{\mathrm{UE}})}, \tag{4.4}$$

where $0.1 \leq \gamma_{\boldsymbol{x}} \leq 0.9$ is the application based variable UE demand with normal distribution $\mathbb{N}(0.5, 0.1)$. The limits on the UE demand ensures avoiding circum-

[1]Note that the formulation for $R_{\mathrm{SZ},\boldsymbol{x}}$ is done for the parameters considered in Section 4.4. A different formulation may be required for suburban and sparsely deployed DBS regions.

stances when UEs with extremely high/low service demands request impractically high/low S-Zone radii.

4.4 Game Formulations for UE-Centric Zone Scheduling

We leverage both evolutionary (EG) and auction (AG) games to determine optimal S-Zone sizes by solving the optimization problem in (4.1). Let $\mathbf{N}$={1,2,...,N} denote the set of UEs participating in each game iteration. Each UE $\boldsymbol{x}$ demands a certain throughput $\tau_{d,\boldsymbol{x}}$, which is a function of the variable demand variable $\gamma_{\boldsymbol{x}}$, such that $\tau_{d,\boldsymbol{x}} \propto \gamma_{\boldsymbol{x}}$. For this work, we consider a linear relationship, $\tau_{d,\boldsymbol{x}} = K\gamma_{\boldsymbol{x}}+\epsilon$, where ϵ is the UE specific noise in the throughput demand. The throughput demand $\tau_{d,\boldsymbol{x}}$ in turn determines the S-Zone radius $R_{\mathrm{SZ},\boldsymbol{x}}$ through iterative update in $\gamma_{\boldsymbol{x}}$ until convergence is achieved for the utility in (4.1). The CBS calculates and communicates $\bar{\tau}$ to every UE within its coverage so that they may adjust $\gamma_{\boldsymbol{x}}$ with the objective of solving (4.1).

4.4.1 Evolutionary Game

In the context of an evolutionary game for S-Zone size optimization, each UE adapts its demand strategy according to the received payoff in (4.1). This evolution of the game allows the population states to evolve over time. For this work, we are considering two UE population states: over-served and under-served expressed as $\mathrm{UE_{OS}}$ and $\mathrm{UE_{US}}$ respectively. An over-served UE is characterized by a negative utility in (4.1) indicating surplus resources in terms of higher achievable throughput as compared to the mean throughput and/or the application based throughput demand. The action strategy for over-served UEs is an adjustment in their S-Zone size through a prefixed step reduction in $\gamma_{\boldsymbol{x}}$. Similarly, the under-

served UEs with positive utilities increase their S-Zone areas through a prefixed step increase in $\gamma_{\boldsymbol{x}}$, within the demand constraints (Step 1, Algorithm 2).

The evolutionary game in Algorithm 2 is governed by principles of replicator dynamics [52], according to which the number of UEs selecting a particular strategy will increase if that action yields close to zero payoff. The frequency for a particular action strategy is given by $X_s = \frac{N_s}{N}$. N_s is the number of UEs selecting strategy s where $s \in S = \{\text{increase } \gamma_{\boldsymbol{x}},\ \text{reduce } \gamma_{\boldsymbol{x}}\}$.

Algorithm 4.1 EG algorithm

1: Each UE chooses a throughput demand $\tau_{d,\boldsymbol{x}}$ based on the application usage. The disparity in $\tau_{d,\boldsymbol{x}}$ is modeled by assuming that the UE demand $\gamma_{\boldsymbol{x}}$ has a normal distribution, i.e. $\gamma_{\boldsymbol{x}} \sim \mathbb{N}(0.5, 0.1)$, and constraints $0.1 \leq \gamma_{\boldsymbol{x}} \leq 0.9$. Iteration counter is set to $i = 1$.
2: Based on the channel gains and $\gamma_{\boldsymbol{x}}$ values, the CBS calculates $R_{\mathrm{SZ},\boldsymbol{x}}$, creates virtual S-Zones around high priority UEs and turns ON a single DBS per S-Zone. Elaborating mathematically, a UE $\boldsymbol{x}$ is scheduled iff $p_{\mathrm{UE}}^{\{\boldsymbol{x}\}} < p_{\mathrm{UE}}^{\{\boldsymbol{x'}\}}; \forall \boldsymbol{x} \in \Pi_{\mathrm{UE}}, \boldsymbol{x'} \in \Pi_{\mathrm{UE}} \cap (\boldsymbol{x}, R_{\mathrm{SZ},\boldsymbol{x}}), \boldsymbol{x'} \neq \boldsymbol{x}$.
3: The scheduled UEs observe $\mathrm{SINR}(R_{\mathrm{SZ},\boldsymbol{x}})$ and subsequently the achievable throughput $\tau_x(R_{\mathrm{SZ},\boldsymbol{x}})$ which is sent to the CBS.
4: The CBS calculates $\bar{\tau}$ and broadcasts it to all UEs.
5: Each scheduled UE computes its utility $u_{\boldsymbol{x}}$ and adjusts $\gamma_{\boldsymbol{x}}$. If $u_{\boldsymbol{x}} > 0$, then $\gamma_{\boldsymbol{x}} = \min(\gamma_{\boldsymbol{x}} + 0.05, 0.9)$; if $u_{\boldsymbol{x}} < 0$, then $\gamma_{\boldsymbol{x}} = \max(\gamma_{\boldsymbol{x}} - 0.05, 0.1)$; and left unchanged otherwise.
6: Update $p_{\mathrm{UE}}^{\{\boldsymbol{x}\}}$, increment i and go to step 2 while $i < i_{\max}$.

$i_{\max}$ is the maximum number of simulation iterations for a given network configuration. The EG based strategy adaptation algorithm functions in a distributive manner where each UE adapts its individual strategy to optimize (4.1). Additionally, the algorithm relieves the CBS from centralized optimization computation, making its implementation scalable throughout the network.

4.4.2 Auction Game

Auction theory allows players to intelligently select their strategies in order to gain maximum resources. For our work, we adapt the Vickrey Clark Groves (VCG) auction mechanism which is known for ensuring assurance of truthfulness from the players as well as maximization of fairness [80]. Also called the "sealed-bid second-price auction", this auctioning scheme awards the bid to the highest bidder who pays an amount equivalent to the second highest bid. Contrary to the first price auctions, VCG auctions prevent selfish players from cheating because bidding the true valuation is the weakly dominant strategy in this model [46]. This guarantees that under most general circumstances, VCG will yield bid winners as players with highest valuations.

In contrast to EG where S-Zone assignment was dependent upon $p_{\mathrm{UE}}^{\{\boldsymbol{x}\}}$ alone, in AG (Algorithm 3), we integrate the utility $u_{\boldsymbol{x}}$ within the bidders' (or UEs') valuation structure as

$$b_{\boldsymbol{x}} = \frac{1}{p_{\mathrm{UE}}^{\{\boldsymbol{x}\}}[\delta(\bar{\tau} - \tau_x(R_{\mathrm{SZ},\boldsymbol{x}})) + (1-\delta)(\tau_{d,\boldsymbol{x}} - \tau_x(R_{\mathrm{SZ},\boldsymbol{x}}))]}. \tag{4.5}$$

As seen from (4.5), the UEs with optimal utilities are rewarded with higher bid values. Each iteration in the AG is a new game as the winners of the conducted auctions are assigned S-Zones by the CBS and barred from further bidding. Because we analyze system level performance metrics, the cost paid by UEs after winning the bids is irrelevant for this work. However, we employ a VCG auction due to its relevance to wireless networks [81]. Moreover, the existing framework can be analyzed in extensions of this work and include the effect of UE cost when modeling the network over large number of TTIs.

Algorithm 4.2 AG algorithm

Steps 1-4 same as EG algorithm.
5: Each UE participating in the auction calculates its bid value $b_{\boldsymbol{x}}$ and sends it to the CBS.
6: The CBS chooses a player as the winner of the auction if its bid is not lower than any other player that can form a non-overlapping S-Zone with the existing S-Zones, i.e. a UE $\boldsymbol{x}$ is the auction winner iff $b_{\boldsymbol{x}} > b_{\boldsymbol{x'}}; \forall \boldsymbol{x}, \boldsymbol{x'} \in \Pi'_{\mathrm{UE}}$, where Π'_{UE} refers to the UEs whose $R_{\mathrm{SZ},\boldsymbol{x}}$ (and $R_{\mathrm{SZ},\boldsymbol{x'}}$) allow non-overlapping S-Zones with past auction winners.
7: The CBS removes the auction winner in current iteration from future bidding. New bidding round (step 5) continues until there is no new winner.
8: CBS schedules auction winners, updates $p_{\mathrm{UE}}^{\{\boldsymbol{x}\}}$ and i. Go to step 3 and re-evaluate the metrics for existing players. Continue until $i < i_{\max}$.

Table 4.1: Simulation parameters

Parameter	**Value**
Simulation area dimensions ($\|A\|$)	100 m x 100 m
$\lambda_{\mathrm{UE}}\|A\|$	400
$\lambda_{\mathrm{DBS}}\|A\|$	50, 100, 150
α	4
δ	0, 0.25, 0.5, 0.75, 1
Power consumption parameters P_o, P_u, Δ_u and P_{ou}	6.8, 1, 4 and 4.3 W
θ	0.5
$i_{\max}$	100
No. of Monte Carlo realizations	1000

4.5 Performance Analysis

We will now discuss the simulation results for a range of efficiency parameters to evaluate the performance of the game theoretic techniques in question, i.e. EGT and AGT within the 2nd tier elastic CDSA framework elaborated in Chapter 4.4. The basic simulation parameters are given in Table 4.1.

4.5.1 Aggregate Throughput performance

The aggregate throughput is calculated numerically as the sum of the achievable throughput of the served UEs within a TTI, i.e. $\lambda_{\mathrm{DBS,Act}}|A| \sum_{1}^{\lambda_{\mathrm{DBS,Act}}|A|} \tau_{\boldsymbol{x}}$ where $\lambda_{\mathrm{DBS,Act}}|A|$ denotes the number of activated DBSs (or served UEs) in the network. The results in Fig. 4.2 reveal contrasting trends with varying δ and DBS densities under considered game theoretic algorithms. While the EGT demonstrates reduction in throughput for the range of DBS densities considered as δ increases, the AGT shows increasing throughput trends for denser DBS deployments with an increase in δ. The trends are disruptive for $\delta = 1$ which is the pure fairness centric scheduling scheme and takes no consideration of users' QoE requirements.

As far the deployment density is concerned, EGT is the superior scheme for $\lambda_{\mathrm{UE}}/8$ DBS density. For denser networks with mean DBS deployment density of $3\lambda_{\mathrm{UE}}/8$, AGT is clearly the preferred game model. The findings highlight the necessity of a SON implementation within the CBS for intelligent and dynamic adaptation of game model as a function of both λ_{DBS} and δ to maximize the system throughput.

4.5.2 Energy Efficiency performance

To estimate the EE of an elastic CDSA under UE-centric architecture, we take inspiration from the work of award winning European project EARTH [82] and apply relevant modifications to construct a power consumption model for a UE-centric CDSA given as

$$P_{\mathrm{CDSA}} = \lambda_{\mathrm{DBS,Act}}|A|(\theta P_o + \Delta_u P_u + P_{ou}), \tag{4.6}$$

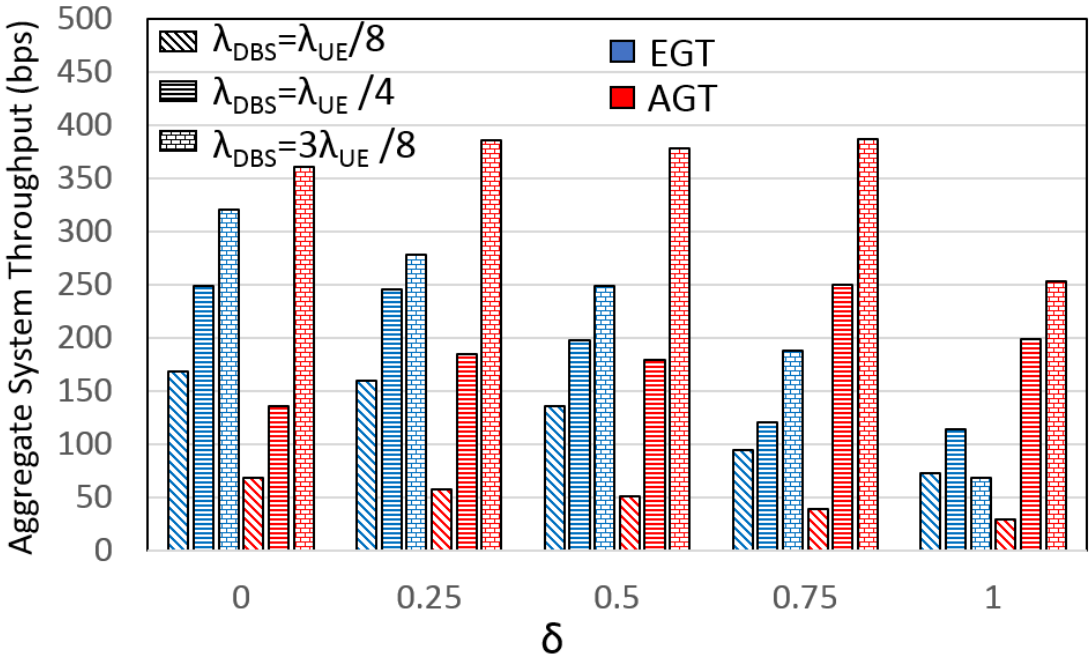


Fig. 4.2: System throughput comparison of EGT and AGT.

where P_o, P_u and P_{ou} denote fixed power consumption of an active DBS, transmit power of a UE terminal and circuit power consumed at UE terminal during discovery respectively [83]. $0 \leq \theta \leq 1$ is a system deployment efficiency parameter with $\theta = 1$ capturing least energy efficient deployment. The power consumption of the CDSA as seen from Eq. (4.6) is an increasing function of the DBS density and a decreasing function of the average UE S-Zone area.

Although AGT (Fig. 4.3) demonstrates comparable EE performance for $\lambda_{\text{DBS}} = \lambda_{\text{UE}}/4$ and $\delta \leq 0.5$; for most of the simulated scenarios, EGT is a clear winner. This can be attributed to a larger average S-Zone area for the EGT implementation which reduces the average number of concurrent DL transmissions and hence more DBSs are deactivated.

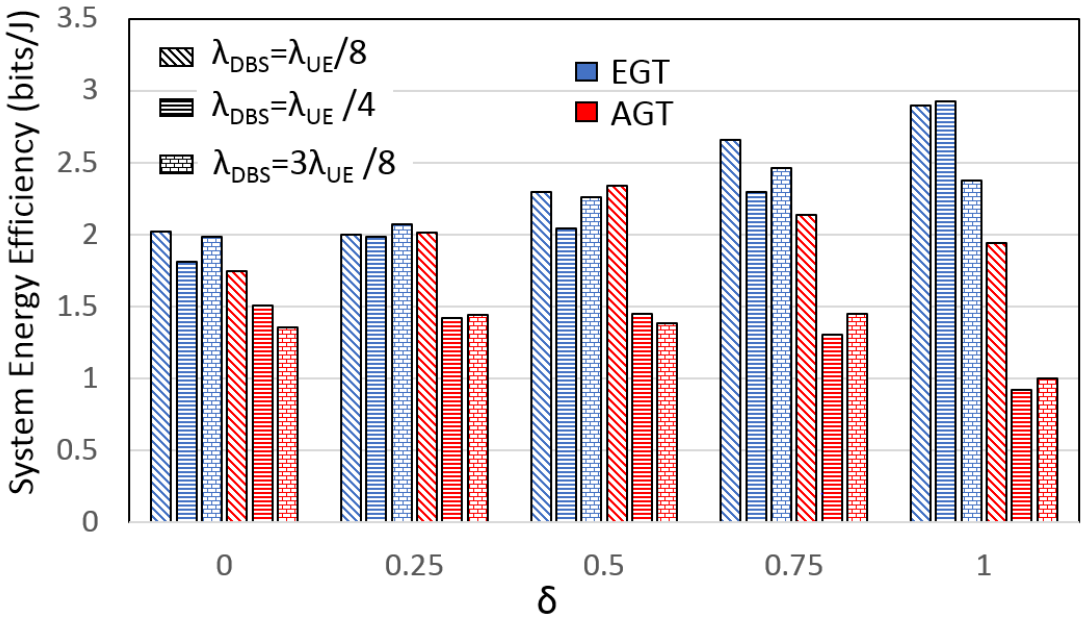


Fig. 4.3: Energy Efficiency comparison of EGT and AGT.

4.5.3 Convergence Analysis

To analyze the convergence of the algorithms, we plot the average payoff received by UEs using EGT and AGT (Fig. 4.4). It can be seen that the system converges to equilibrium relatively faster with AGT. Particularly at low DBS densities, for instance at $\lambda_{\text{DBS}} = \lambda_{\text{UE}}/8$ in fig. 5.8b, the mean UE utility converges to 0 almost instantly. Not only does the AGT outperform EGT in convergence speed, but also in achieving optimal UE utility, i.e. by minimizing $|u_{\boldsymbol{x}}|$. Negative utilities for EGT indicate UEs receiving sufficiently high SINR to attain achievable throughputs exceeding the desired levels needed to optimize (4.1). The root cause for the non-ideal UE utility distribution with EGT can be traced back to UE demand constraints ($0.1 \leq \gamma_{\boldsymbol{x}} \leq 0.9$) which bars the UEs with high negative utilities to further reduce the S-Zone size and increase their utility.

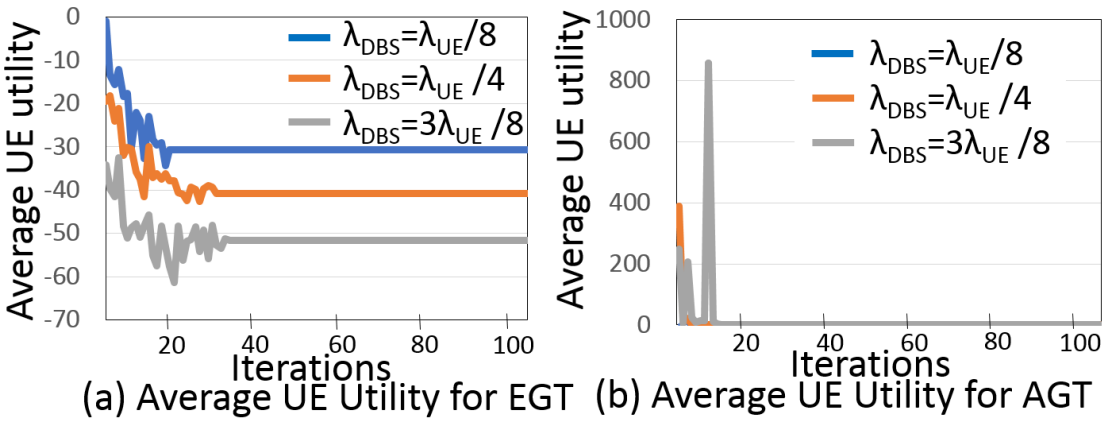


Fig. 4.4: Convergence of (a) EGT and (b) AGT algorithms for an elastic UEC CDSA network.

4.5.4 UE Scheduling success

The expected delay for a UE to be scheduled is analyzed by plotting the mean served UE ratio ($\frac{\lambda_{\text{DBS,Act}}}{\lambda_{\text{UE}}}$) under variable DBS densities and δ values (Fig. 4.5). The served UE ratio represents the expected number of scheduled UEs after equilibrium is achieved for the EGT and AGT games. The smallest wait time is observed for AGT scheme with $\lambda_{\text{DBS}} = 3\lambda_{\text{UE}}/8$ where an arbitrary UE is expected to be re-scheduled after every 5th or 6th TTI. For $\lambda_{\text{DBS}} = \lambda_{\text{UE}}/8$, EGT marginally outperforms AGT while AGT exhibits higher scheduling success for $\delta \geq 0.5$ when $\lambda_{\text{DBS}} = \lambda_{\text{UE}}/4$. The simulation results in fig. 5.9 once again reiterate the practicality of a SON engine capable of adjusting δ and alternating between game models to yield desired service delay times within UE-centric CDSA.

4.5.5 Performance Comparison with first tier user-centric elasticity

Fig. 4.6 shows the performance gains in terms of aggregate system throughput (Fig. 4.6(a)) and energy efficiency (Fig. 4.6(b)) for the proposed UE-centric elastic CDSA in comparison to the uniform user-centric service regions proposed

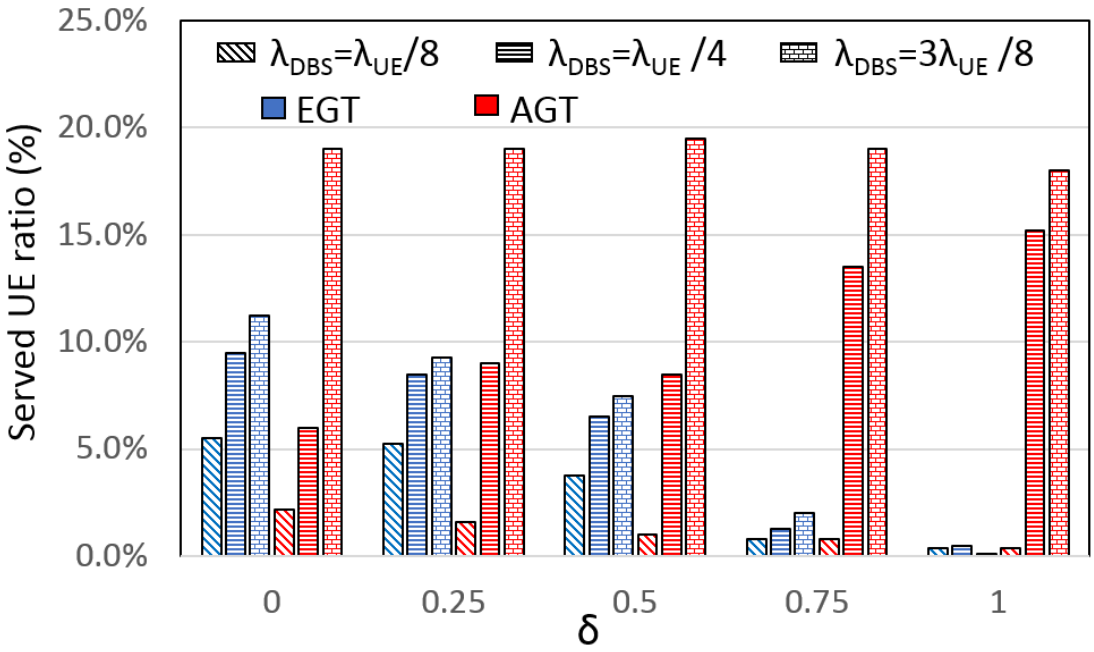


Fig. 4.5: UE Scheduling Success probabilities with EGT and AGT algorithms.

in Chapter 3 [3]. The network models with fixed user-centric regions in [3] to maximize ASE and EE are referred to as FS(ASE) and FS(EE) respectively. The variable-sized QoE-centric service zones proposed in this work with EGT and AGT implementations are referred to as VS(EGT) and VS(AGT) respectively. Clearly, the proposed model outperforms "one-size fits all" strategy both in terms of system throughout and EE by virtue of assigning flexible user-centric service zones that are appropriately sized to meet an arbitrary UE's data requirements. While AGT yields higher data throughput, particularly at high λ_{DBS}, the EGT is more energy efficient. Once again, this result reiterates the need for an intelligent SON enabled CBS that can switch the game models to support a higher data throughout (or energy efficiency).

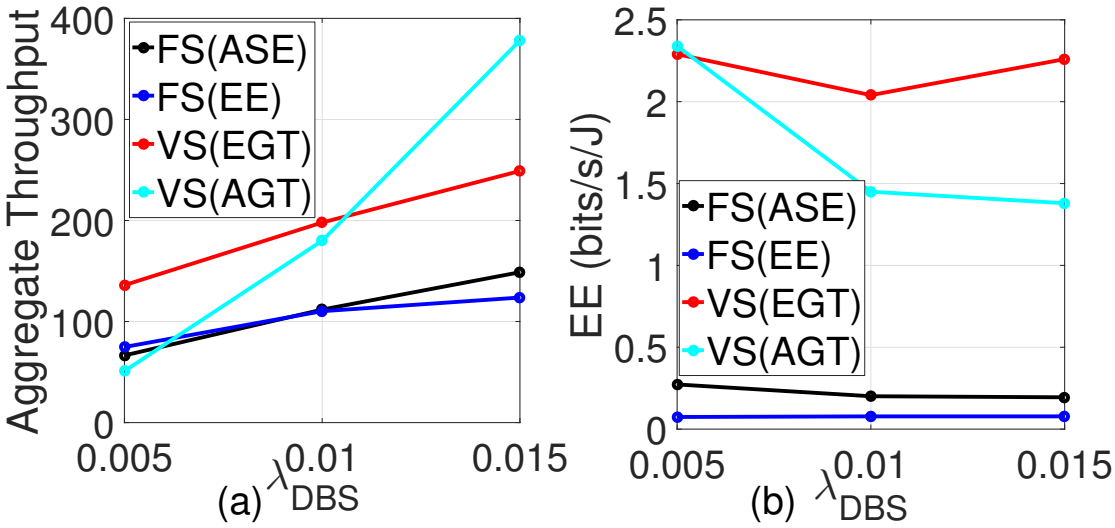


Fig. 4.6: (a) System throughput and (b) EE comparison of uniform user-centric [3] and QoE-centric service zone approaches with different DBS densities.

4.6 Conclusion

In this chapter, we presented an elastic cellular network framework capable of catering to individual UE QoE requirements. The QoE flexibility is realized through virtual interference free service zones centered around scheduled UEs. We proposed a distributed utility minimization problem to model appropriate S-Zone formations around the UEs. To evaluate the optimization of S-Zone allotment to UEs, we conducted a detailed comparative analysis using evolutionary and auction based game implementations at a centralized CBS. We investigated different key performance indicators including aggregate network throughput, energy efficiency, mean UE scheduling probability and algorithm convergence speed. Simulation results were presented for performance evaluation of the algorithms in terms of aggregate system throughput, energy efficiency, user scheduling ratio and mean algorithm convergence time. It was demonstrated that the proposed QoE based user-centric service zone provisioning yields better performance as compared to a UE-centric network with static system-wide service zone area. Our analysis advo-

cates integration of an intelligent self-organizing network (SON) engine [84] within the proposed UEC CDSA network architecture. The SON engine would optimize a network efficiency metric by dynamically shifting game strategies with respect to network dynamics and spatio-temporally varying operator's business model.

CHAPTER 5

Stienen model based mmWave user-centric networks

In chapters 3 and 4, we presented user-centric models with circular service regions around UEs with high scheduling priorities (in chapter 2) and those UEs that won distributive games (in chapter 3). However, in both these central and distributed game mechanisms, the common feature was the circular geometer of the service region. The obvious drawback of the circular geometry is lower number of possible non overlapping service regions. In the scheduling algorithm, this corresponds to a longer wait time for an arbitrary UE. For future 5G, in which ultra reliable low latency communication (URLLC) is an integral component, a longer scheduling wait time is not desirable. This chapter presents an analytical framework for performance characterization of a novel Stienen cell based user-centric architecture operating in millimeter wave spectrum. In the proposed architecture, at most one remote radio head (RRH) is activated within non overlapping user equipment (UE)-centric Stienen cells (S-cells) generated within the Voronoi region around each UE. Under the presented framework, we derive analytical models for the three key performance indicators (KPIs): i) SINR distribution (used as an indicator for quality of service (QoS)), ii) area spectral efficiency (ASE), and iii) energy efficiency (EE) as a function of the three major design parameters in the proposed architecture, namely UE service probability, S-cell radius coefficient and RRH deployment density. The analysis is validated through extensive Monte Carlo simulations. The simulation results provide practical design insights into the interplay among the three design parameters, tradeoffs among the three KPIs, sensitivity of each KPI to the design parameters as well as optimal range of the

design parameters.

5.1 Introduction

5.1.1 Background

Ultra-dense deployment of small cells using higher frequency bands, such as millimeter wave (mmWave), is being widely accepted in both academia and industry as the prime course to meet the ever growing data demands in future cellular networks, vis-a-vis 5G and beyond. Unlike some earlier studies, it has now been established that densification alone does not yield linear gains in coverage probability. In fact, the coverage probability shows a decreasing trend at high base station (BS) densities as the network transitions from a noise-limited to an interference-limited system [85]. To aggravate matters further, high operational expenditures and energy consumption associated with dense deployments take a toll on the network operators by further reducing the already dwindling profit margins.

Designing and operating the network in a user equipment (UE)-centric fashion instead of the traditional cell-centric design has gained traction recently as a viable strategy for 5G and beyond [2]. Shifting the network design pivot from the BS to the UE not only ensures ubiquitous coverage and UE specific differentiated quality of service (QoS) to the UEs in dense deployments, but also provides a mechanism for selective BS activation that ensures reduced energy consumption. Recent works have quantified the area spectral efficiency (ASE) and energy efficiency (EE) in a user-centric Cloud radio access network (C-RAN) [62, 3]. The network design in both these works relies upon creating non-overlapping circular service zones around high priority UEs. This results in a one-to-one (1-1) UE-BS association for each service zone such that a single BS is activated at max per service zone. The service

zone radius is employed as a control parameter to realize the desired tradeoff between EE and ASE. However, the disjoint circular user-centric service zones in the mentioned works cause high latency, specially within dense user hotspots, due to longer wait times in downlink (DL) scheduling.

5.1.2 Architectural Overview

In this chapter, a novel user-centric architecture based on Stienen cells around UEs is analyzed for downlink scheduling in ultra-dense deployment scenarios. The Stienen model, introduced in 1982 [86] considers the maximal ball inscribed within the Poisson Voronoi cell and centered at the generating point of the Voronoi cell. In simple terms, Stienen model is a union of non overlapping circular cells with a diameter that has the exact same distribution as the nearest neighbor distance of a Poisson point process. In the proposed architecture, all the BSs within a UE-centric Stienen cell (S-cell) are associated to its centroid UE. The Stienen cells are bounded by the Voronoi tessellation generated through the UE positions on the x-y plane (see Fig. 1). The Stienen model offers manifold advantages compared to the already conceived user-centric architectures: i) it integrates the randomness in BS deployment as well as user locations, ii) it enables flexibility to capture the effect of polygonal user-centric service regions, something that the one-size-fits-all strategy fails to do, and iii) it gives an opportunity for all the UEs to be scheduled, provided the BS deployment is sufficiently dense to ensure presence of at least one BS within each UE-centric S-cell.

By employing stochastic geometry principles, we obtain the analytical framework to determine the coverage bounds in the proposed user-centric Stienen architecture. While deployment of a large number of BSs is capable of enhancing coverage for an arbitrary UE, the energy consumption becomes significant. Our model tack-

les the dual problem of maintaining high throughput as well as energy efficiency through the selective BS activation mechanism. We consider operating activated BSs within S-cells at high frequency mmWave band on DL because: i) experimental trials have indicated suitability of mmWave communications due to high spectrum availability per channel [87]; ii) higher frequencies allow implementation of small-sized antenna arrays to facilitate narrow directional beams and longer transmission ranges; and iii) larger free space pathloss and directional DL transmissions at mmWave frequencies reduce the unwanted interference from nearby BSs. Additionally, the UE-BS pair proximity within the Stienen model ensures resilience to blockage effects that occur at larger UE-BS distances. Furthermore, since the mmWave cells do not interfere with existing sub-6 GHz BSs, they can be deployed within existing deployments and enable architectures such as Cloud RAN enabled heterogeneous networks and control-data separation architecture (CDSA) [76].

5.1.3 Research Objectives

User-centric architectures have been well investigated for both mmWave and sub-6 GHz deployments. However, analytical characterization and performance analysis of the Stienen based model in user-centric mmWave network remains terra incognita. In this chapter, we investigate the relevance of this system model in future ultra dense 5G networks, thereby reducing this research gap. To this end, the main research questions investigated are summarized as follows:

- How would DL scheduling within a Stienen based user-centric model look like in dense mmWave wireless networks?
- How does directional communication in mmWave impact the inter-cell in-

terference, particularly for the user-centric Stienen cell model?

- Can we formulate closed form expressions for the efficiency parameters, i.e. ASE and EE for the proposed user-centric S-cell model?
- What are the advantages of the proposed model in comparison with the earlier proposed circular shaped
- How do the efficiency formulations change to incorporate a second tier (MBS) within the network?
- Does there exist some ASE-EE tradeoff by varying a parameter of interest within the user-centric S-cell network>

For this purpose, we validate the formulated expressions through extensive Monte Carlo simulations. The simulation results provide practical design insights into the interplay among the three design parameters, tradeoffs among the three KPIs, sensitivity of each KPI to the design parameters as well as optimal range of the design parameters.

5.2 Network Model

5.2.1 Spatial Model

We consider the downlink of a two-tier ultra-dense network consisting of one sub-6 GHz MBS that has mmWave RRHs and UEs spatially distributed across its foot-prints. Both the RRHs and UEs are assumed to be outdoors. This is to limit number of parameters in our model for simplicity. For indoor users, building penetration losses can be accounted for by appropriate scaling of the signal and interference powers. The spatial distributions for RRHs and UEs are modelled using

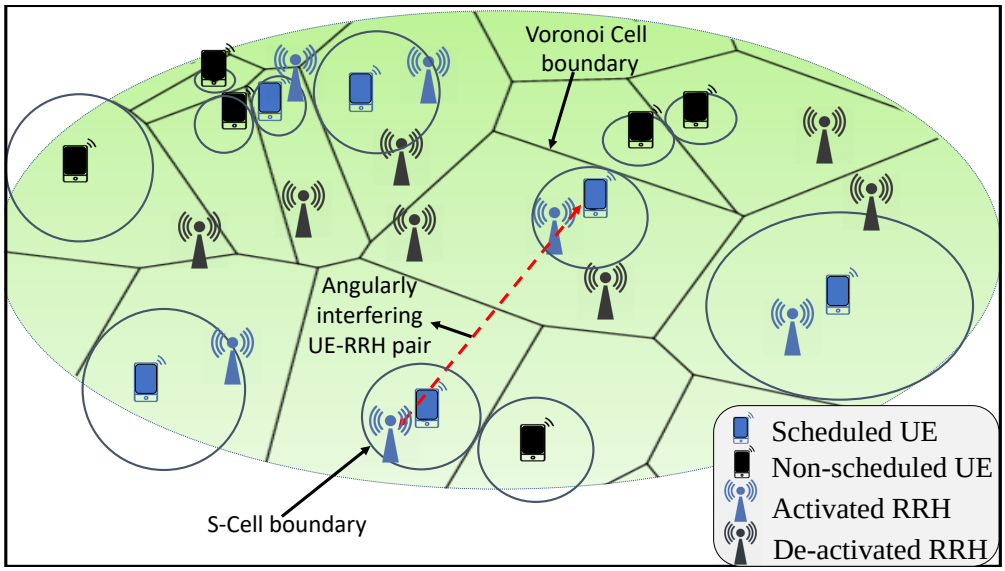


Fig. 5.1: The UE-centric S-cell architecture. The Voronoi tessellation of the plane is formed by Φ_{UE}. The circles around UEs represent S-cell edges at $\zeta = 1/2$. S-cell boundaries are not drawn to scale.

two independent homogeneous Poisson point processes (HPPPs) Φ_{RRH} and Φ_{UE} with intensities λ_{RRH} and λ_{UE} respectively. The UE locations act as generating points for the UE-centric Voronoi tessellation. Each RRH is associated to a UE based on its physical location on the Voronoi plane. This implies that each RRH can at the most serve a single UE which is spatially closest to the RRH during a transmission time interval (TTI).

5.2.2 User-Centric Stienen Cell Geometry

To visualize a Stienen based user-centric cell design, consider the UEs as generating points of a 2-dimensional Poisson-Voronoi Tessellation. Every point in the convex polygons generated by the UEs is closer to its generating UE than to any other. Now a UE-centric Stienen cell can be formed by constructing around each UE a circular disk with a radius that is less than or equal to half of the distance between the UE and its closest neighbor (see Fig. 16.2). The resulting S-cell regions form

a Poisson hard sphere model, which by definition is formed when the interiors of the disks centered at the generating points of the parent Poisson process do not overlap almost surely. The radius of the S-cell around a UE $u_j \in \Phi_{\text{UE}}$ is given as:

$$R_j = \min_{u_k} \zeta \mid\mid u_j - u_k \mid\mid; u_j, u_k \in \Phi_{\text{UE}}, j \neq k, \tag{5.1}$$

where $0 < \zeta \leq 1/2$ is the S-cell radius coefficient that models the flexibility in S-cell size. As $\zeta \to 0$, the S-cell size around UEs becomes negligible. For $\zeta = 0.5$, we obtain the largest possible S-cell sizes around each UE without overlap with adjacent S-cells. The coefficient provides control on modeling the S-cell sizes around UEs and is particularly useful in load balancing between RRHs and MBS in multi-tier networks. Note that $\zeta = 0.5$ creates the nearest neighbor model proposed by Stienen in [86].

We can observe from Fig. 6.2 that the user-centric Voronoi cells can be divided into two regions; the one within the circular Stienen cells and the other which is outside. This demarcation is particularly useful to ensure that as long as the serving RRH is within the Stienen cell, a UE will not have any interfering RRH in a neighboring cell, that is spatially closer to the UE than its serving RRH.

5.2.3 Dual Slope LOS Ball Pathloss Model

Propagation in mmWave band in known to be severely impacted by blocking, atmospheric attenuation and low diffraction around obstacles. To effectively exhibit the blockage effects at high frequencies, a tri-state model is commonly used [88], according to which, a UE-RRH link can be in a line-of-sight (LOS), non-line-of-sight (NLOS) or outage state. A LOS link occurs when there is no blockage between a UE and its serving RRH. A NLOS link, on the other hand, occurs

when the UE-serving RRH link is blocked, but the UE still receives sufficient signal strength through multipath components. An outage state refers to infinite pathloss, i.e., when the UE-RRH spatial separation is high enough to an extent that communication is not possible. The probability distribution for the tri-state model is based on experimental trials [87] and expressed as (5.2).

$$\begin{aligned} p_{\mathrm{LOS}}(r) &= \left[1 - \max\left(0, 1 - 181.27\exp(-r/30)\right)\right]\exp(-r/67.1), \\ p_{\mathrm{NLOS}}(r) &= \left[1 - \max\left(0, 1 - 181.27\exp(-r/30)\right)\right]\left(1 - \exp(-r/67.1)\right), \\ p_{\mathrm{OUT}}(r) &= 1 - p_{\mathrm{LOS}}(r) - p_{\mathrm{NLOS}}(r). \end{aligned} \tag{5.2}$$

For analytical tractability, we use an equivalent LOS ball approximation where the pathloss is expressed as a Bernoulli random variable [89] with the assumption that all UE-RRH links are LOS within a distance constraint. If $p_{\mathrm{LOS}}(r)$ and $p_{\mathrm{NLOS}}(r)$ are the probabilities of a LOS and NLOS link between a UE and mmWave RRH at distance r respectively, then for the LOS ball radius of R_o,

$$\begin{aligned} p_{\mathrm{LOS}}(r) &= 1; \text{ if } r \leq R_o, \text{ and} \\ p_{\mathrm{NLOS}}(r) &= 1; \text{ if } r > R_o. \end{aligned} \tag{5.3}$$

The described LOS ball model has been shown to simplify mathematical derivation in the system-level analysis at the cost of only a minor difference from the actual SINR distribution [90]. Although $p_{\mathrm{LOS}}(r)$ for different UE-RRH links is not independent, it has been shown that the dependence is weak [91]; therefore we ignore potential blockage correlations in our model. Log-normal shadowing may be considered, but is not used to maintain the tractability of the model. From (5.3), the distinct pathloss behavior for LOS and NLOS links in our work is expressed as:

$$
\begin{aligned}
PL(r) &= A_{\mathrm{LOS}} r^{\alpha_{\mathrm{LOS}}};\ \text{if } r \leq R_o, \text{ and} \\
&= A_{\mathrm{NLOS}} r^{\alpha_{\mathrm{NLOS}}};\ \text{if } r > R_o,
\end{aligned}
\tag{5.4}
$$

where α_{LOS} and α_{NLOS} are the terrain and operating frequency dependent pathloss exponents for LOS and NLOS links respectively. The intercepts A_{LOS} and A_{NLOS} may be assumed identical if same closed-in reference distance is employed [92]. Note that high α_{NLOS} values result in sufficiently large $PL(r)$, effectively causing the UE-RRH links to be in outage as r increases.

5.2.4 mmWave Beamforming

It is assumed that both the UEs and RRHs are equipped with antenna arrays to perform directional beamforming. We assume a sectorized antenna gain pattern to allow for constant array gains within the main lobe and side lobe. We also consider perfect channel knowledge between a UE and its serving RRH which enables them to adjust their beam steering orientation to achieve maximum directionality gain. For the sake of simplicity, we do not consider errors in channel estimation and synchronization (time and/or frequency) in this work. Given $\mathrm{M} \in \{\mathrm{UE}, \mathrm{RRH}\}$, let G_{M}, g_{M} and θ_{M} denote the main lobe gains, side lobe gains and half power beamwidths (HPBWs) respectively of the UEs and RRHs. Then the directivity gain for a desired UE-RRH link is $G_{\mathrm{UE}} G_{\mathrm{RRH}}$. Assuming that the angle of arrival of an interfering beam at a typical UE is independent and uniformly distributed between $(0, 2\pi]$, the directivity gain G_{I} is a discrete random variable with the probability distribution given in [89] and mean interference gain for an arbitrary UE expressed as:

$$\mathbb{E}(G_{\mathrm{I}}) = \frac{\theta_{\mathrm{UE}}\theta_{\mathrm{RRH}}}{4\pi^2} G_{\mathrm{UE}} G_{\mathrm{RRH}} + \frac{\theta_{\mathrm{UE}}}{2\pi}\frac{1-\theta_{\mathrm{RRH}}}{2\pi} G_{\mathrm{UE}} g_{\mathrm{RRH}} + \frac{1-\theta_{\mathrm{UE}}}{2\pi}\frac{\theta_{\mathrm{RRH}}}{2\pi} g_{\mathrm{UE}} G_{\mathrm{RRH}} + \frac{(1-\theta_{\mathrm{UE}})(1-\theta_{\mathrm{RRH}})}{4\pi^2} g_{\mathrm{UE}} g_{\mathrm{RRH}}. \tag{5.5}$$

5.2.5 Channel Model

Due to the limited scattering behavior of mmWave signals, the Rayleigh fading model commonly used for sub-6 GHz band is not applicable [92]. Therefore, we assume independent Nakagami fading for each UE-RRH link with N_{L} and N_{N} representing the LOS and NLOS parameters respectively. The small-scale fading in signal power given by $|h|$ under the Nakagami assumption is a normalized Gamma random variable. We assume N_{L} and N_{N} to be positive integers. Furthermore, shadowing is not assumed for the sake of tractability.

5.3 Downlink Association Scheme user-centric S-cell networks

A typical UE is associated to an RRH within its S-cell using the smallest pathloss criteria. The smallest $PL(r)$ criteria ensures the maximum average SINR to each of the scheduled UEs. To cater for system limitations and interference management, we introduce a UE selectivity parameter $p_{\mathrm{UE}} \in (0, 1]$ that is a random thinning factor denoting the percentage of UEs that will be considered for DL service in a given TTI. Here, $p_{\mathrm{UE}} = 1$ implies that all the UEs that have at least one RRH in their S-cell will be scheduled for DL. Consequently, the served UEs form a thinned PPP $\Phi_{\mathrm{UE}}^{\mathrm{EF}}$ characterized by an intensity that is a function of p_{UE} as well as the RRH and UE densities and given by

$$\lambda_{\mathrm{UE}}^{\mathrm{EF}} = p_{\mathrm{UE}}(1 - p_{\mathrm{emp}})\lambda_{\mathrm{UE}}, \tag{5.6}$$

where p_{emp} is the probability that a S-cell has no RRH to serve the associated UE. To evaluate the empty S-cell probability, we first express the probability density function (pdf) of the number of RRHs within the S-cell of a UE (see Lemma 2 in [93]) as

$$f_{n_{\text{RRH}}}(n_{\text{RRH}}) = \frac{\lambda_{\text{UE}}\zeta^{-2}}{\lambda_{\text{UE}}\zeta^{-2} + \lambda_{\text{RRH}}} \left(\frac{\lambda_{\text{RRH}}}{\lambda_{\text{RRH}} + \lambda_{\text{UE}}\zeta^{-2}} \right)^{n_{\text{RRH}}}. \tag{5.7}$$

The empty S-cell probability can simply be calculated by putting $n_{\text{RRH}} = 0$ in (5.7), i.e.

$$p_{\text{emp}} = \frac{\lambda_{\text{UE}}\zeta^{-2}}{\lambda_{\text{UE}}\zeta^{-2} + \lambda_{\text{RRH}}}. \tag{5.8}$$

To illustrate the effect of variable S-cell sizes, Fig. 5.2 shows the circular areas around UEs with different ζ values. As ζ increases, the number of RRHs within S-cells also increases, thereby increasing the intensity of the PPP representing interfering RRHs. This user scheduling scheme within the user-centric S-cell architecture is summarized as algorithm 1. Here, $b(\mathbf{x}, r)$ denotes a ball of radius r centered at a point $\mathbf{x}$.

From the practical implementation perspective, in the event of a RRH-free S-cell, user clustering strategies [68] may be employed where nearby UEs are grouped together and optimization is performed on the UE clusters rather than individual UEs. Such strategies are beyond the scope of this work and can be a topic of future extensions of this work.

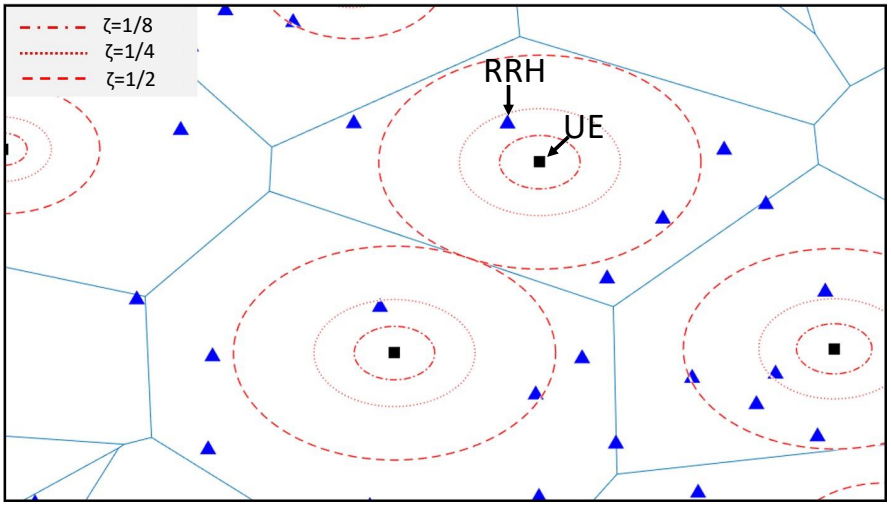


Fig. 5.2: The UE-centric Stienen cell sizes with different ζ values.

5.4 Distance Distribution to Angularly Interfering RRHs

In this chapter, we characterize the distribution of the distance between a typical UE scheduled on DL and the activated RRHs outside the UE's S-cell that have completely aligned antenna beams with the UE's directional beam. As discussed in section II-E, the probability of such an event occurring is $\frac{\theta_{\mathrm{UE}}\theta_{\mathrm{RRH}}}{4\pi^2}$. We focus on the angularly interfering RRHs, also called "angular interferers", because they contribute the largest share of interference at an arbitrary UE due to maximum directivity gain. Exploiting the well-known fact that the distance between the nearest neighbors in a 2-D Poisson process is Rayleigh distributed [8], we can write the distribution of the distance between an arbitrary UE and its serving RRH given as r_o in Fig. 6.2 as:

$$f_{r_o}(r) = 2\pi r p_{\mathrm{UE}} \lambda_{\mathrm{UE}} \exp(-\pi r^2 p_{\mathrm{UE}} \lambda_{\mathrm{UE}}). \tag{5.9}$$

The tight packing in a Voronoi cell structure, especially in user hotspots, will inevitably give rise to scenarios with interfering RRHs co-located in close vicinity

Algorithm 5.1 UE scheduling algorithm in a user-centric S-cell network

Inputs: Φ_{RRH}, Φ_{UE}, p_{UE}, ζ

Outputs: Φ'_{RRH}, Φ'_{UE}

1: Initialize the set of scheduled UEs and the RRHs serving within the user-centric S-cells at any given time slot as Φ'_{UE}, $\Phi'_{\text{RRH}} \leftarrow \emptyset$.

2: Update $\Phi^{\text{EF}}_{\text{UE}}$ by thinning Φ_{UE} with a factor of p_{UE}.

3: For each $\boldsymbol{x} \in \Phi^{\text{EF}}_{\text{UE}}$, estimate the size of S-cell as $R_{\boldsymbol{x}} = \zeta d_{\boldsymbol{x}}$, where $d_{\boldsymbol{x}}$ is the distance to the nearest UE in $\Phi^{\text{EF}}_{\text{UE}}$.

4: Update Φ'_{RRH} and Φ'_{UE} for the current time slot using the following conditions:

foreach $\boldsymbol{x} \in \Phi^{EF}_{UE}$ **do**
  if $\boldsymbol{x} \cap b(\boldsymbol{x}, R_{\boldsymbol{x}}) \neq \emptyset$, $\forall \boldsymbol{x} \in \Phi^{EF}_{UE}$ **then**
    $\Phi'_{\text{UE}} \cup \{\boldsymbol{x}\}$
    foreach $\boldsymbol{y} \in \Phi_{RRH}$ **do**
      if $\boldsymbol{y} \in b(\boldsymbol{x}, R_{\boldsymbol{x}})$ **then**
        if $PL_{\boldsymbol{x},\boldsymbol{y}} < PL_{\boldsymbol{x},\boldsymbol{y'}}, \forall \boldsymbol{y'} \in \Phi_{RRH}, \boldsymbol{y'} \in b(\boldsymbol{x}, R_{\boldsymbol{x}}), \boldsymbol{y'} \neq \boldsymbol{y}$ **then**
          $\Phi'_{\text{RRH}} \cup \{\boldsymbol{y}\}$
        end
      end
    end
  else
    continue.
  end
end

5: Serve all the scheduled users Φ'_{UE} from the associated RRHs in Φ'_{RRH}.

6: Go to step 1.

to the serving RRH. However, the requirement for narrow beam directionality in mmWave systems reduces the chances of exact alignment between a UE and an interfering RRH. Additionally, inherent to the Voronoi cell design, a UE-angularly interfering RRH pair must lie on opposite sides of the RRH serving that UE. This induces a mean minimum repulsion distance between a UE and an angularly interfering RRH which is equivalent to $\mathbb{E}(f_{r_o}(r))$.

Proposition 5.1. The distribution of the distance between a typical UE and its ith nearest angular interferer in a UE-centric S-cell network can be characterized

as

$$f_{r_i}(r) = (r^{2i-1}) \left(\frac{p_{\mathrm{UE}} \lambda_{\mathrm{UE}} (1 - p_{\mathrm{emp}}) \theta_{\mathrm{UE}} \theta_{\mathrm{RRH}}}{4\pi} \right)^i \frac{2}{\Gamma(i)} \exp \left(\frac{-r^2 p_{\mathrm{UE}} \lambda_{\mathrm{UE}} (1 - p_{\mathrm{emp}}) \theta_{\mathrm{UE}} \theta_{\mathrm{RRH}}}{4\pi} \right). \tag{5.10}$$

Proof: The probability that the distance between an arbitrary UE and its ith nearest angular interferer is at least r is essentially the probability that there exist exactly $i-1$ angular interferers inside the circular region of area πr^2 around that UE. Mathematically, it is expressed as

$$F_{r_i}(r) = 1 + \frac{(\lambda_{\mathrm{AIRRH}} \pi r^2)}{(i-1)!} \exp(-\lambda_{\mathrm{AIRRH}} \pi r^2), \tag{5.11}$$

where λ_{AIRRH} is the mean intensity of the PPP Φ_{AIRRH} representing angularly interfering RRHs around an arbitrary UE. For characterization of λ_{AIRRH}, we first define Φ_{IRRH} which is the PPP representing interfering RRHs, i.e. activated RRHs outside a UE's S-cell. Due to 1-1 UE-RRH association, the distribution of interfering RRHs is identical to the served UEs. Using Slivnyak's theorem [67], we can express the mean intensity of Φ_{IRRH} represented by λ_{IRRH} as $\lambda_{\mathrm{IRRH}} = \lambda_{\mathrm{UE}}^{\mathrm{EF}} = p_{\mathrm{UE}}(1 - p_{\mathrm{emp}})\lambda_{\mathrm{UE}}$.

Now, the angularly interfering RRHs is simply a subset of interfering RRHs containing the RRHs having completely aligned antenna beams with the considered UE. This implies that Φ_{AIRRH} is a thinned version of Φ_{RRH}. Stating more precisely, the mean intensity of the number of angularly interfering RRHs to an arbitrary UE is dependent on the following factors:

- The number of UEs scheduled for service per unit area, which is controllable by an adjustable parameter p_{UE}.

- The probability that the Stienen cell around a UE would have an RRH within its covered area. The size of the Stienen cell is adjustable via ζ.
- The probability that the antenna bore-sights of a UE and an activated RRH outside its S-cell are completely aligned.

Keeping the conditions stated above in consideration, the mean intensity of Φ_{AIRRH} is expressed as

$$\lambda_{\text{AIRRH}} = \frac{p_{\text{UE}}\lambda_{\text{UE}}(1-p_{\text{emp}})\theta_{\text{UE}}\theta_{\text{RRH}}}{(2\pi)^2}. \tag{5.12}$$

Substituting (5.12) in (5.11) and differentiating to find the probability density function yields (5.10). Note that for omni-directional transmission, Φ_{AIRRH} essentially converges to Φ_{IRRH}, which is the union of angularly interfering RRHs and interfering RRHs with misaligned antenna beams with UE.

□

5.5 Coverage Probability and Area Spectral Efficiency in a UE-Centric S-cell

Now that we have discussed the distance distribution of the angular interferers, we will proceed towards characterization of the coverage probability of a UE within a UE-centric S-cell design. Consider a scheduled UE $\mathbf{x} \in \Phi_{\text{UE}}$. Let $\mathbf{o} \in (\Phi_{\text{RRH}} \cap S(\mathbf{x}))$ be the RRH that yields minimum pathloss and is selected for DL service within $\mathbf{x}$'s S-cell area "$S(\mathbf{x})$" which is mathematically given as $S(\mathbf{x}) = b(\mathbf{x}, R_{\mathbf{x}})$. We consider the aggregate interference from both the angularly and non-angularly interfering RRHs. We have already seen in chapter 5.4 that the mean intensity of the PPP representing the interfering RRHs is given by $\lambda_{\text{IRRH}} = \lambda_{\text{UE}}(1-p_{\text{emp}})$. Without

loss of generality, we use the Silvnyak's theorem [67] and focus our analysis on the arbitrary UE $\mathbf{x}$ assumed to be located at the origin. With this assumption, the downlink SINR is given as:

$$\mathrm{SINR} = \frac{\max_{\boldsymbol{o}\in(\Phi_{\mathrm{RRH}}\cap S(\mathbf{x}))} h_{\boldsymbol{o}} G_{\mathrm{UE}} G_{\mathrm{RRH}} PL(r_{\boldsymbol{o}})}{\sigma^2 + \sum_{\boldsymbol{i}\in\Phi_{\mathrm{IRRH}}} h_{\boldsymbol{i}} \mathbb{E}(G_{\mathrm{I}}) PL(r_{\boldsymbol{i}})}, \tag{5.13}$$

where $r_{\boldsymbol{o}}$ and $r_{\boldsymbol{i}}$ are the relative distances of UE $\mathbf{x}$ with its DL scheduled and interfering RRHs respectively. σ^2 is the variance of the additive white Gaussian noise (AWGN) at the UE front end.

Once the interference is characterized, we can approximate the link success probability which represents the percentage of users with adequate link channel quality with the connected RRHs for DL. We can represent the QoS demands represented numerically through an SINR threshold γ_{th}. In this case, the coverage probability while taking into account the distinct fading characteristics and pathloss behaviors of LOS and NLOS links and is given by Theorem 5.1.

Theorem 5.1: *The link coverage probability of an arbitrary UE served under the proposed user-centric S-cell design and a one-to-one UE-RRH association scheme can be expressed as*

$$\begin{aligned} P_{cov}(\gamma_{th}) = &\sum_{n=1}^{N_{\mathrm{L}}} (-1)^{n+1} \binom{N_{\mathrm{L}}}{n} \int_0^{R_o} \exp\left(\frac{-n\eta_{\mathrm{L}}\gamma_{th}\sigma^2 r^{\alpha_{\mathrm{LOS}}}}{G_{\mathrm{UE}}G_{\mathrm{RRH}}}\right) \exp(-I_{\mathrm{LL}}(\gamma_{th}, r)) \exp(-I_{\mathrm{LN}}(\gamma_{th}, r)) f_{r_o}(r)\mathrm{d}r \\ &+ \sum_{n=1}^{N_{\mathrm{N}}} (-1)^{n+1} \binom{N_{\mathrm{N}}}{n} \int_{R_o}^{\infty} \exp\left(\frac{-n\eta_{\mathrm{N}}\gamma_{th}\sigma^2 r^{\alpha_{\mathrm{NLOS}}}}{G_{\mathrm{UE}}G_{\mathrm{RRH}}}\right) \exp(-I_{\mathrm{NL}}(\gamma_{th}, r)) \exp(-I_{\mathrm{NN}}(\gamma_{th}, r)) f_{r_o}(r)\mathrm{d}r. \end{aligned} \tag{5.14}$$

The terms I_{LL}, I_{LN}, I_{NL} and I_{NN} in (5.14) can be evaluated from equations (5.15)-(5.18).

$$I_{\mathrm{LL}}(\gamma_{th}, r) = 2\pi\lambda_{\mathrm{IRRH}} \sum_{k=1}^{4} b_k \int_{r}^{R_o} F\left(N_{\mathrm{L}}, \frac{n\eta_{\mathrm{L}}\bar{a_k}\gamma_{th} r^{\alpha_{\mathrm{LOS}}}}{N_{\mathrm{L}} t^{\alpha_{\mathrm{LOS}}}}\right) t\mathrm{dt}. \tag{5.15}$$

$$I_{\mathrm{LN}}(\gamma_{th}, r) = 2\pi\lambda_{\mathrm{IRRH}} \sum_{k=1}^{4} b_k \int_{R_o}^{\infty} F\left(N_{\mathrm{N}}, \frac{n\eta_{\mathrm{L}}\bar{a_k}\gamma_{th} r^{\alpha_{\mathrm{LOS}}}}{N_{\mathrm{N}} t^{\alpha_{\mathrm{NLOS}}}}\right) t\mathrm{dt}. \tag{5.16}$$

$$I_{\mathrm{NL}}(\gamma_{th}, r) = 0. \tag{5.17}$$

$$I_{\mathrm{NN}}(\gamma_{th}, r) = 2\pi\lambda_{\mathrm{IRRH}} \sum_{k=1}^{4} b_k \int_{r}^{\infty} F\left(N_{\mathrm{N}}, \frac{n\eta_{\mathrm{N}}\bar{a_k}\gamma_{th} r^{\alpha_{\mathrm{NLOS}}}}{N_{\mathrm{N}} t^{\alpha_{\mathrm{NLOS}}}}\right) t\mathrm{dt}. \tag{5.18}$$

Eq. (5.17) is a direct consequence of the Ball LOS approximation model according to which if a UE is being served by an RRH in NLOS region (i.e. $r_o > R_o$), no interfering RRH will have a LOS link with that UE. The quantities $\bar{a_k}$, b_k, $F(N, x)$ and η_i used in equations (5.15)-(5.18) are given by equations (5.19)-(5.22) respectively.

$$\bar{a_k} = \left[\frac{G_{\mathrm{UE}}G_{\mathrm{RRH}}}{G_{\mathrm{UE}}G_{\mathrm{RRH}}}, \frac{G_{\mathrm{UE}}g_{\mathrm{RRH}}}{G_{\mathrm{UE}}G_{\mathrm{RRH}}}, \frac{g_{\mathrm{UE}}G_{\mathrm{RRH}}}{G_{\mathrm{UE}}G_{\mathrm{RRH}}}, \frac{g_{\mathrm{UE}}g_{\mathrm{RRH}}}{G_{\mathrm{UE}}G_{\mathrm{RRH}}}\right]. \tag{5.19}$$

$$\begin{aligned} b_k = [&\frac{\theta_{\mathrm{UE}}\theta_{\mathrm{RRH}}}{(2\pi)^2}, \frac{\theta_{\mathrm{UE}}}{2\pi}(1 - \frac{\theta_{\mathrm{RRH}}}{2\pi}), \\ &(1 - \frac{\theta_{\mathrm{UE}}}{2\pi})\frac{\theta_{\mathrm{RRH}}}{2\pi}, (1 - \frac{\theta_{\mathrm{UE}}}{2\pi})(1 - \frac{\theta_{\mathrm{RRH}}}{2\pi})]. \end{aligned} \tag{5.20}$$

$$F(N,x) = 1 - \frac{1}{(1+x)^N}. \tag{5.21}$$

$$\eta_i = N_i(N_i!)^{-\frac{1}{N_i}}. \tag{5.22}$$

Proof: See Appendix A.

□

The system-wide area spectral efficiency is measured as the number of bits which can be transmitted per Hertz bandwidth per second within 1 squared meter area. We discuss the ASE for two different DL transmission mechanisms:

Fixed Rate Transmission

In a fixed rate transmission, all UEs which meet the SINR criteria for DL transmission are scheduled with uniform data rate links. The potential throughput for a UE-centric S-cell architecture in this scenario can be quantified as

$$\mathrm{ASE}_{\mathrm{FR}} = \lambda_{\mathrm{UE}}^{\mathrm{EF}} \log_2(1+\gamma_{th}) P_{cov}(\gamma_{th}). \tag{5.23}$$

Adaptive Rate Transmission

In this system, each UE is provided a DL data rate which is proportional to its SINR, subject to meeting the SINR threshold criteria. Mathematically, the ASE

in this situation may be expressed as

$$\mathrm{ASE}_{\mathrm{AR}} = \frac{1}{\mathbb{A}} \sum_{u \in \Phi_{\mathrm{UE}}^{\mathrm{EF}}} \log_2(1 + \mathrm{SINR}_u);\ \text{if } \mathrm{SINR}_u > \gamma_{th}, \tag{5.24}$$

where $\mathbb{A}$ is the total network area under consideration. As is clear from (5.23) and (6.5), the area spectral efficiency of the UE-centric S-cell is strongly coupled with p_{UE}. Intuitively, a higher p_{UE} increases the effective number of scheduled users. However it also lowers the SINR due to shorter mean distance of a UE to its closest interfering RRH. In the results section, through evaluation of the attainable area spectral efficiency, we will investigate which of the two opposite effects is dominant in the proposed architecture.

5.6 UE-Centric S-cell architecture in a Multi-tier Network

Until now, we have analyzed how a UE-centric Stienen cell architecture is modeled for a single tier network where UEs are only served if there resides at least one mmWave RRH within its S-cell area as depicted in Algorithm 1. Needless to point out, this gives rise to longer wait times and even results in network outage (due to larger UE-serving RRH distance) in sparse RRH deployment regions. It is therefore pertinent to include the analysis of a multi-tier system with MBS tier serving the UEs which do not have any RRH within their S-cell at sub-6 GHz. There is a two-fold advantage of this approach: firstly, operating the MBS and RRH tier in different frequency bands avoids any co-tier interference, and secondly, it is well known that propagation loss on sub-6 GHz band is less severe as compared to mmWave, allowing MBS to serve UEs at larger distances. With respect to UE association, the mechanism will have slight modifications as compared to Algorithm 1. All the UEs that are void of RRHs within their S-cell are connected to

their respective MBS for DL coverage. Let us consider that Φ_{MBS} and $\Phi_{\text{UE,MBS}}$ are PPPs, with mean intensities λ_{MBS} and $\lambda_{\text{UE,MBS}}$ respectively, and represent MBS deployment density and UEs served by MBSs respectively. Unlike RRH deployment which is impromptu and hence is modelled well by PPP, MBS deployment is likely to be well planned. Therefore, to model realistic MBS deployment, we induce repulsion between the PPP representation of MBS tier. This is done by modeling it as a type II Matern hardcore process [67] where we choose a subset of the original PPP with a distance constraint. Mathematically, the thinned PPP representation of MBS is given by

$$\lambda_{\text{MBS}}^{\text{EF}} = \frac{1 - \exp(-4\pi\lambda_{\text{MBS}} R_{\text{MBS}}^2)}{4\pi R_{\text{MBS}}^2}. \tag{5.25}$$

(5.25) implies that the minimum allowable distance between adjacent MBSs is $2R_{\text{MBS}}$. The analytical characterization of network level KPIs for the proposed UE-centric S-cell under the multi-tier network is presented below:

5.6.1 SINR

The SINR at a UE $\mathbf{x}$ connected to a RRH or MBS is dependent on the presence of an RRH within its S-cell and is given by a piece-wise function in (5.26). For the UEs connected to MBSs, we consider Rayleigh fading environment. This implies that when $\boldsymbol{x} \in \Phi_{\text{UE,MBS}}$, the channel gain ($h_o$ and h_i) is assumed to be a unit mean exponential random variable and pathloss at $\mathbf{x}$ from an RRH $\mathbf{y}$ is modeled by $l(||\boldsymbol{x} - \boldsymbol{y}||) = ||\boldsymbol{x} - \boldsymbol{y}||^{-\alpha_{\text{MBS}}}$ power-law function.

$$\text{SINR}_{\text{MT}} = \begin{cases} \frac{\max_{\boldsymbol{o} \in \Phi_{\text{RRH}} \cap S(\mathbf{x})} h_{\boldsymbol{o}} G_{\text{UE}} G_{\text{RRH}} PL(r_{\boldsymbol{o}})}{\sigma^2 + \sum_{\boldsymbol{i} \in \Phi_{\text{IRRH}}} h_{\boldsymbol{i}} \mathbb{E}(G_{\text{I}}) PL(r_{\boldsymbol{i}})} & \boldsymbol{x} \in \Phi'_{\text{UE}} \\ \frac{\max_{\boldsymbol{o} \in \Phi_{\text{MBS}}} h_{\boldsymbol{o}} PL(r_{\boldsymbol{o}})}{\sigma^2 + \sum_{\boldsymbol{i} \in \Phi_{\text{MBS}'}} h_{\boldsymbol{i}} PL(r_{\boldsymbol{i}})} & \boldsymbol{x} \in \Phi_{\text{UE,MBS}} \end{cases} \tag{5.26}$$

5.6.2 Coverage Probability

For a multi-tier network, the coverage probability is dependent on both the SINR threshold γ_{th} and the probability of a UE being served by an RRH, which is a function of ζ and p_{emp}. Mathematically, the coverage probability is given by

$$P_{cov,\text{MT}}(\gamma_{th}) = (1 - p_{\text{emp}})P_{cov}(\gamma_{th}) + p_{\text{emp}}P_{cov,\text{MBS}}(\gamma_{th}), \tag{5.27}$$

where $P_{cov}(\gamma_{th})$ is the coverage probability of a UE connected to a mmWave RRH given by (5.14) while $P_{cov,\text{MBS}}(\gamma_{th})$ is the probability that the SINR at an arbitrary UE served by an MBS exceeds the QoS threshold, i.e. SINR $\geq \gamma_{th}$. The coverage probability $P_{cov,\text{MBS}}(\gamma_{th})$ is inspired from the work in [8] for the Rayleigh fading scenario and expressed as

$$P_{cov,\text{MBS}}(\gamma_{th}) = \pi\lambda_{\text{MBS}} \int_0^\infty \exp\left(-\pi\lambda_{\text{MBS}} r \left[1 + \gamma_{th}^{2/\alpha_{\text{MBS}}} \int_{\gamma_{th}^{-2/\alpha_{\text{MBS}}}}^\infty \frac{1}{1 + v^{\alpha_{\text{MBS}}/2}} dv\right] - \gamma_{th}\sigma^2 r^{\alpha_{\text{MBS}}/2}\right) dr, \tag{5.28}$$

where α_{MBS} is the pathloss exponent for the propagation on MBS tier.

5.6.3 Area Spectral Efficiency

In the same spirit as Section 5.5, the area spectral efficiency for a multi-tier network can be evaluated for fixed rate and adaptive rate transmissions as (5.29) and (5.30) respectively:

$$\text{ASE}_{\text{FR,MT}} = \lambda_{\text{UE}} p_{\text{UE}} \log_2(1 + \gamma_{th})[(1 - p_{\text{emp}})P_{cov}(\gamma_{th}) + p_{\text{emp}}P_{cov,\text{MBS}}(\gamma_{th})]. \tag{5.29}$$

$$\mathrm{ASE}_{\mathrm{AR,MT}} = \frac{1}{\mathbb{A}}[\sum_{u\in\Phi_{\mathrm{UE}}^{\mathrm{EF}}} \log_2(1+\mathrm{SINR}_u) + \sum_{v\in\Phi_{\mathrm{UE,MBS}}} \log_2(1+\mathrm{SINR}_v)], \qquad (5.30)$$

where SINR_u and SINR_v are the SINR quantities for RRH and MBS connected UEs respectively and given by (5.26).

5.7 Energy Efficiency Analysis

The power consumption of a stand-alone small cell RRH is investigated in the project EARTH [73]. This model was extended further to integrate the benefits of centralized processing in [74]. Taking inspiration from these works, we formulate the power consumption for both the MBSs and small cell RRHs as a linear combination of fixed power and load dependent power consumption components. Mathematically, the total power consumption of the two-tier network can be simplified as

$$P = |\mathbb{A}|\{\underbrace{\lambda_{\mathrm{MBS}}^{\mathrm{EF}}\left(A_{\mathrm{MBS}}P_{\mathrm{MBS,Tx}} + B_{\mathrm{MBS}}\right)}_{\textbf{MBS-Tier Power}} + \underbrace{\lambda_{\mathrm{UE}}p_{\mathrm{UE}}(1-p_{\mathrm{emp}})P_{cov}(\gamma_{th})\left(A_{\mathrm{RRH}}P_{\mathrm{RRH,Tx}} + B_{\mathrm{RRH}}\right)}_{\textbf{Activated RRH-Tier Power}}$$
$$+ \underbrace{B_{\mathrm{RRH}}\left(\lambda_{\mathrm{RRH}} - \lambda_{\mathrm{UE}}p_{\mathrm{UE}}(1-p_{\mathrm{emp}})P_{cov}(\gamma_{th})\right)}_{\textbf{De-activated RRH-Tier Power}}\}. \qquad (5.31)$$

B_{MBS} (and B_{RRH}) denotes the fixed power consumption of an MBS (and RRH). This is the energy cost which is bore by the network regardless of the number of UEs requesting DL service. The coefficients A_{MBS} (and A_{RRH}) lump together frequency dependent response of a power amplifier and several other factors within MBS (and RRH). The coefficients A_{i} and B_{i}, where $\mathrm{i} \in \{\mathrm{MBS,RRH}\}$ are expressed as

$$A_{\mathrm{i}} = \frac{1}{\eta_{\mathrm{i}}^{\mathrm{PA}}(1-\sigma_{\mathrm{i}}^{\mathrm{feed}})(1-\sigma_{\mathrm{i}}^{\mathrm{MS}})(1-\sigma_{\mathrm{i}}^{\mathrm{DC}})(1-\sigma_{\mathrm{i}}^{\mathrm{cool}})}, \text{ and} \tag{5.32}$$

$$B_{\mathrm{i}} = \frac{P_{\mathrm{i,RF}} + P_{\mathrm{i,BB}}}{(1-\sigma_{\mathrm{i}}^{\mathrm{MS}})(1-\sigma_{\mathrm{i}}^{\mathrm{DC}})(1-\sigma_{\mathrm{i}}^{\mathrm{cool}})}. \tag{5.33}$$

Note that we are considering an always ON MBS deployment to avoid coverage holes and provide uninterrupted control/signaling to the UEs. The power saving in the proposed architecture will thus come from intelligently turning OFF mmWave RRHs that are not providing DL data services to UEs. For detailed explanation of the power consumption parameters, readers are referred to [73]. The network wide EE is analyzed for adaptive rate transmission scenario in our work and expressed mathematically as

$$EE = \frac{\mathbb{A}[\mathrm{ASE}_{\mathrm{AR,MT}}]}{P}. \tag{5.34}$$

EARTH's segmentation of the power consumption employed in this work allows us to analyze the amount of energy saving possible when an RRH is turned OFF. It is well known that the major chunk of power consumption in a cellular BS takes place within the power amplifier. By dynamic shutting down of the power amplifier and the associated transmission unit, there is significant energy saving, especially in dense deployments. In the next section, we will look at the potential power saving in terms of EE variation with adjustments in three network parameters: i) served UE percentage, ii) RRH deployment density, and iii) user-centric S-cell sizes.

Table 5.1: Simulation parameters

Parameter	Value
Simulation area dimensions ($\|\mathbb{A}\|$)	100 m x 100 m
$\lambda_{\mathrm{UE}}\|\mathbb{A}\|$; $\lambda_{\mathrm{RRH}}\|\mathbb{A}\|$; $\lambda_{\mathrm{MBS}}\|\mathbb{A}\|$	400;400;16
p_{UE}	[0.25 0.5 0.75 1]
$\theta_{\mathrm{UE}}, \theta_{\mathrm{RRH}}$	$[7^o\ 10.9^o]$
G_{UE}; G_{RRH};g_{UE}; g_{RRH}	10 dB; 10 dB; -3 dB; -3 dB;
R_o (from (5.2)); N_{L}; N_{N}	46.5 m; 3; 2
Pathloss exponents: α_{LOS}; α_{NLOS}; α_{MBS}	2.4; 4.7; 3
A_{LOS}, A_{NLOS}	1
R_{MBS}	50 m
$P_{\mathrm{MBS,Tx}}$; $P_{\mathrm{RRH,Tx}}$	10 W; 1 W
$\eta_{\mathrm{RRH}}^{\mathrm{PA}}$,$\sigma_{\mathrm{RRH}}^{\mathrm{feed}}$,$\sigma_{\mathrm{RRH}}^{\mathrm{MS}}$,$\sigma_{\mathrm{RRH}}^{\mathrm{DC}}$,$\sigma_{\mathrm{RRH}}^{\mathrm{cool}}$,$\mathrm{P}_{\mathrm{RRH,RF}}$, $P_{\mathrm{RRH,BB}}$	0.0025,0,0.1,0.08,0, 0.4 W,1.2 W
$\eta_{\mathrm{MBS}}^{\mathrm{PA}}$,$\sigma_{\mathrm{MBS}}^{\mathrm{feed}}$,$\sigma_{\mathrm{MBS}}^{\mathrm{MS}}$,$\sigma_{\mathrm{MBS}}^{\mathrm{DC}}$,$\sigma_{\mathrm{MBS}}^{\mathrm{cool}}$,$\mathrm{P}_{\mathrm{MBS,RF}}$,$P_{\mathrm{MBS,BB}}$	0.388,0,0.07,0.06,0.09, 10.9 W,14.8 W
A_{RRH}; B_{RRH}; A_{MBS}; B_{MBS}	23.22; 1.932; 3.24; 32.3
No. of Monte Carlo realizations	100000

5.8 Performance Analysis

In this chapter, we present numerical as well as Monte Carlo simulation results to evaluate the validity of the developed model. After ensuring accuracy of the models, we compare the performance of the proposed UE-centric S-cell architecture in mmWave networks with the conventional architecture. We also discuss the interplay between the network parameters and their impact on ASE and EE in a multi-tier user-centric S-cell network. Unless otherwise specified, the basic simulation parameters used in our analysis are given in Table 5.1.

5.8.1 Model Validation

We validate the expression for SINR coverage probability derived in Theorem 5.1 in Fig. 5.3. The plot shows close agreement between the simulations and the derived analytical results, particularly at high γ_{th} regimes. We observe that the analytical model holds true for variations in both ζ (Fig. 5.3 (a)) and p_{UE} (Fig.

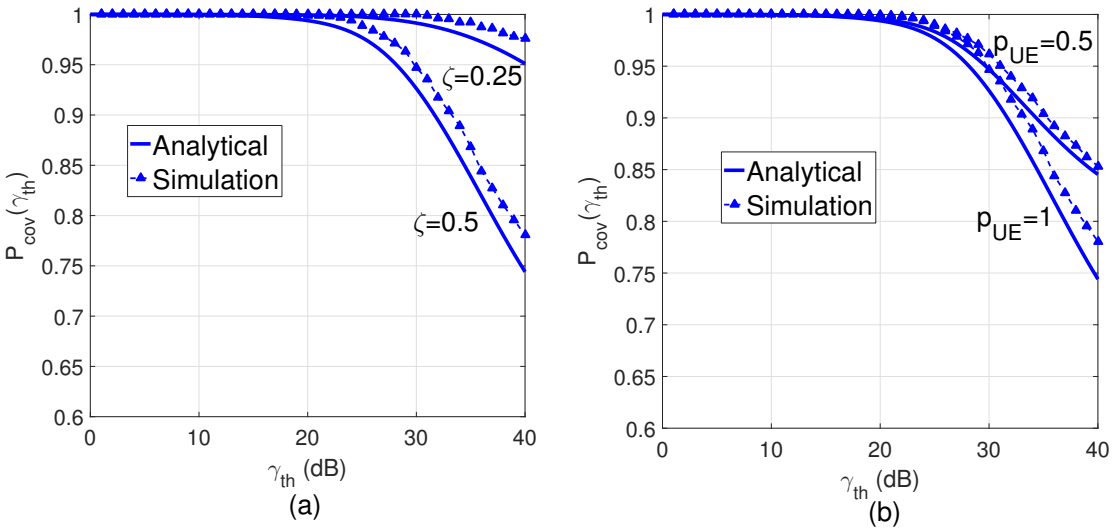


Fig. 5.3: SINR coverage probability at: (a) different ζ and p_{UE}=1, (b) different p_{UE} and $\zeta = 0.5$.

5.3(b)), keeping the other parameter fixed. Fig. 5.3 also depicts a decrease in coverage probability at higher ζ and p_{UE} for the same SINR threshold. This is intuitive because a larger p_{UE} reduces the average UE-interfering RRH distance. As a consequence, the average interference increases which reduces the coverage probability of an arbitrary UE. Similarly, a larger ζ means larger S-cell area which increases the probability of a UE being served due to presence of at least one RRH within its S-cell (see Fig. 5.2). In other words, higher ζ results in greater number of activated RRHs which increases the overall interference for an arbitrary UE.

To validate the distance distribution of angularly interfering RRHs derived in section 5.4, we compare the analytical and Monte Carlo simulation results for λ_{IRRH} in Fig. 5.4. Results show the formulated model to be quite accurate for variations in both p_{UE} and ζ for a range of RRH densities. Another interesting observation from Fig. 5.4. is the high sensitivity of interfering RRH density to ζ. Increasing ζ by 2 at $p_{\text{UE}} = 1$ yields an interfering RRH density increase of about 220% while the same increase in p_{UE} at $\zeta = 0.5$ yields only 35% increase in interfering RRHs.

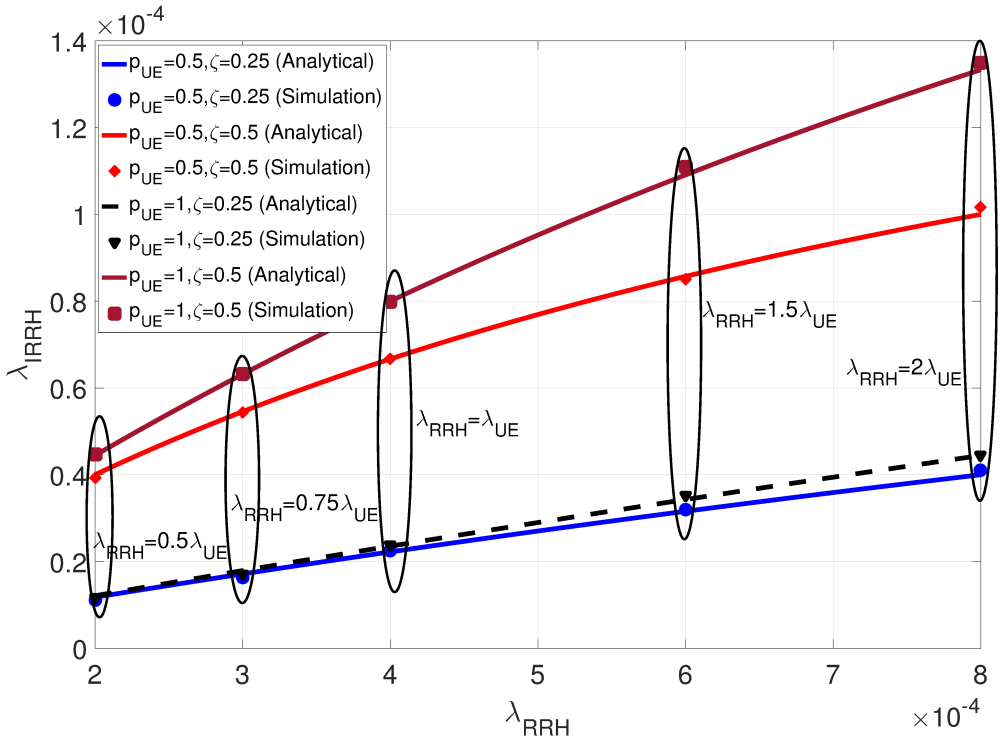


Fig. 5.4: Interfering RRH density validation at different ζ and p_{UE} values.

5.8.2 Sensitivity Analysis

Next, we analyze the gradient of coverage probability with respect to UE selection probability and the S-cell size in Fig. 5.5. Keeping uniform intervals for the possible range of p_{UE} and ζ, we note that the rate of change of coverage probability is far more sensitive to an interval change in ζ. Negative gradient is observed for the entire range of p_{UE} and ζ which makes sense because an increase in either of these parameters results in a larger number of interfering RRHs and a subsequent reduction in coverage probability. The peak absolute value for $\frac{\mathrm{d}P_{cov}}{\mathrm{d}p_{\mathrm{UE}}}$ and $\frac{\mathrm{d}P_{cov}}{\mathrm{d}\zeta}$ occur at $p_{\mathrm{UE}} = 0.34$ and $\zeta = 0.35$ respectively. However, the rate of coverage change per interval variation in ζ is almost 3 times as compared to when p_{UE} is varied. This is in agreement with Fig. 5.4 which demonstrated higher increase in interfering RRHs with a unit increase in ζ.

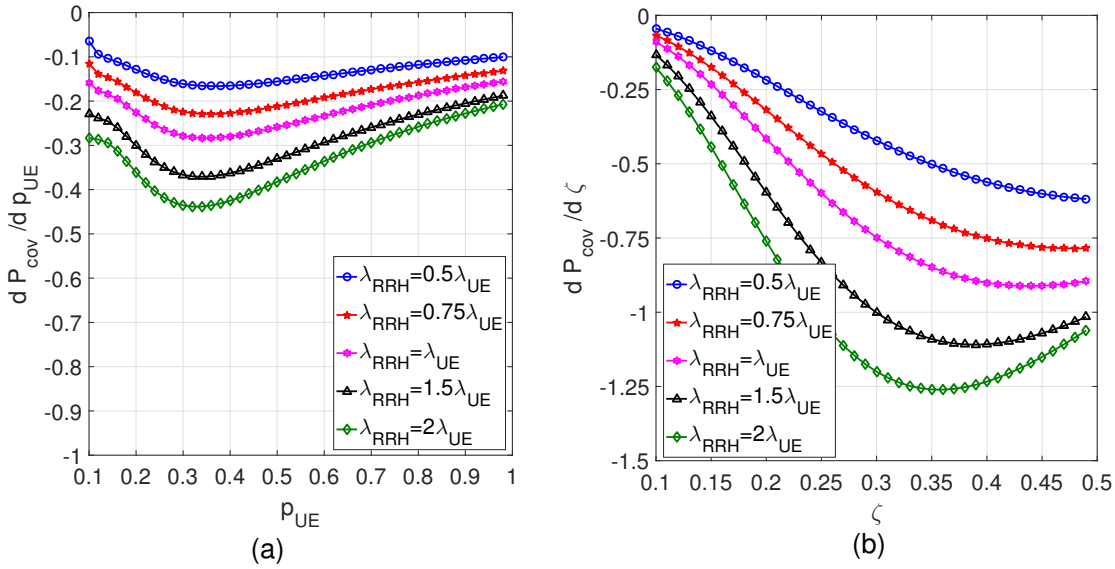


Fig. 5.5: Gradient of coverage probability with respect to: (a) p_{UE} and (b) ζ.

Both Figs. 5.5(a) and 5.5(b) demonstrate that the impact on coverage probability reduction is found to be less severe in sparse RRH deployments. The results provide design insights for the proposed UE-centric S-cell network, for instance, with regards to choosing between UE selection parameter adjustment and S-cell size adjustment or appropriate combination of the two for optimizing coverage.

5.8.3 QoE Enhancement in User-Centric S-cell Network

Users' QoE analysis is conducted through SINR distribution between UEs at different p_{UE} and ζ values in Fig. 5.6. To compare the QoE performance with a standard non user-centric PPP deployment, we follow the approach in [90] and represent it as "NUC". Results in Fig. 5.6 (a) for a single tier mmWave network show that the UE-centric S-cell approach yields an SINR gain ranging from 40 dB to 65 dB for almost 50% of the users. Following the same trend from Figs. 5.4 and 5.5, the SINR observed a marked gain from a decrease in ζ. As discussed earlier, the increase in SINR at lower ζ values is a result of thinning of the PPP Φ_{IRRH}

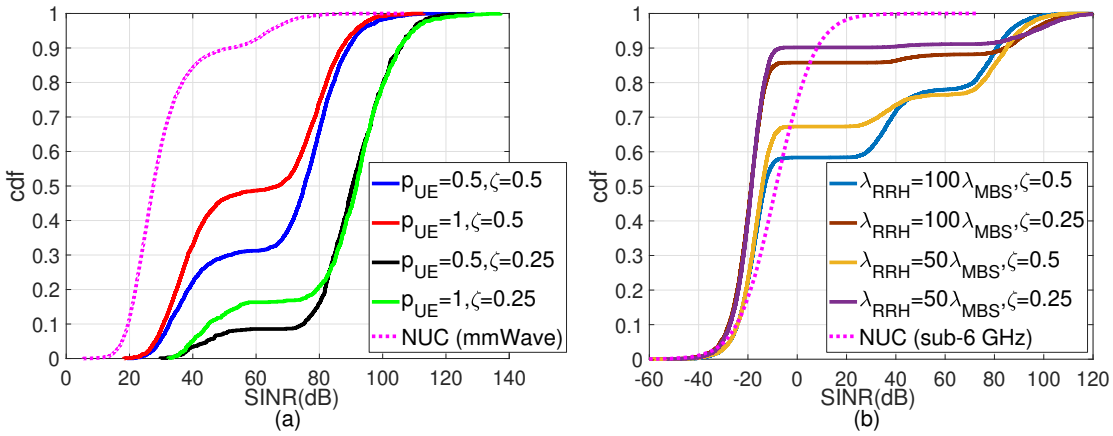


Fig. 5.6: Downlink SINR cdf comparison between user-centric S-cell and non user-centric approaches in (a) single-tier and (b) multi-tier networks.

and a consequent increase in the average interferer distance.

The multi-tier SINR cumulative distribution function (cdf) in Fig. 5.6(b) shows some interesting trends. We observe a clear distinction between the SINR of the UEs connected to the MBS and mmWave RRH tiers. More specifically, three distinct regions can be identified from the multi-tier SINR distribution plot in Fig. 5.6(b): i) the majority of UEs having SINR less than -10 dB SINR for all the simulation cases; ii) a set of UEs with SINR between 25 dB and 50 dB for $\zeta = 0.5$; and iii) a set of UEs with SINR above 70 dB for $\zeta = 0.5$. These jumps of 35 dB and 20 dB are observed as we transition from sub-6 GHz MBS connected UEs to NLOS mmWave RRH connected UEs, and then from NLOS to LOS mmWave RRH connected UEs respectively.

For the same S-cell size, a denser RRH deployment pushes more UEs to the RRH tier. Although the number of UEs connected to mmWave RRH tier increases, the resulting average SINR for the tier is lower as compared to a sparse RRH deployment due to a higher number of interfering mmWave RRHs. This allows the network operator to fluctuate the design parameters and choose between a small number of UEs connected to mmWave RRHs with extremely high user QoE

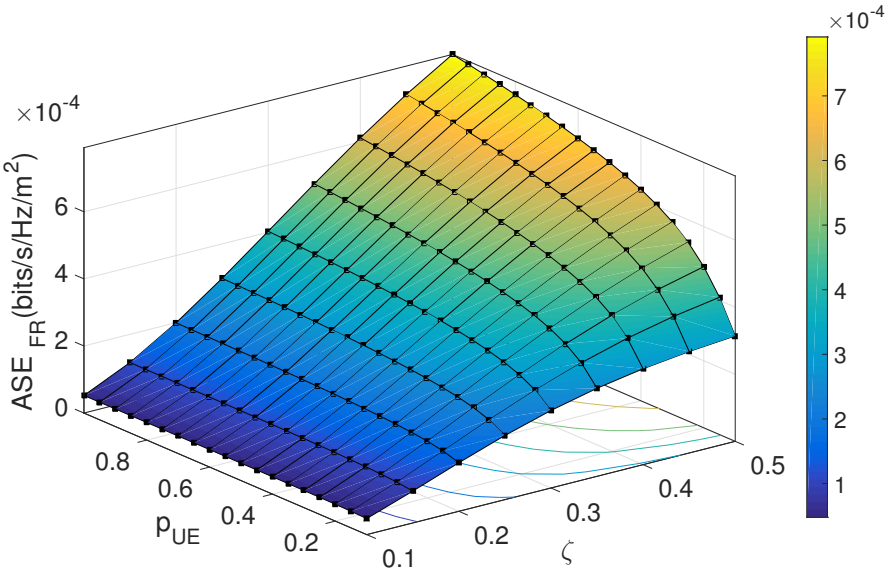


Fig. 5.7: ASE v/s p_{UE} and ζ in a UE-centric S-cell network

or a larger number of UEs connected to mmWave RRHs with moderately lower QoE.

5.8.4 ASE, EE Performance in Single-tier User-Centric S-cell Networks

In this chapter, we investigate the system wide ASE performance in fixed rate as well as adaptive rate transmission scenarios and the impact of p_{UE}, ζ, θ_{RRH} and λ_{RRH} on ASE. Results in Fig. 5.7 reveal that fixed rate ASE is a monotonically increasing function of both ζ and p_{UE}. This implies that the increase in $\lambda_{\mathrm{UE}}^{\mathrm{EF}}$ accompanied by a higher p_{UE} dominates the decrease in the SINR coverage probability observed in Fig. 5.3. Similar to the earlier presented results in Fig. 5.5, $\mathrm{ASE}_{\mathrm{FR}}$ is more sensitive to ζ with the steepest gradient at $p_{\mathrm{UE}} = 1$.

Fig. 5.8 presents the Monte Carlo simulation results for the adaptive rate area spectral efficiency and energy efficiency for the single tier UE-centric S-cell network. We observe a significant increase in the ASE (Fig. 5.8 (a)) as the density

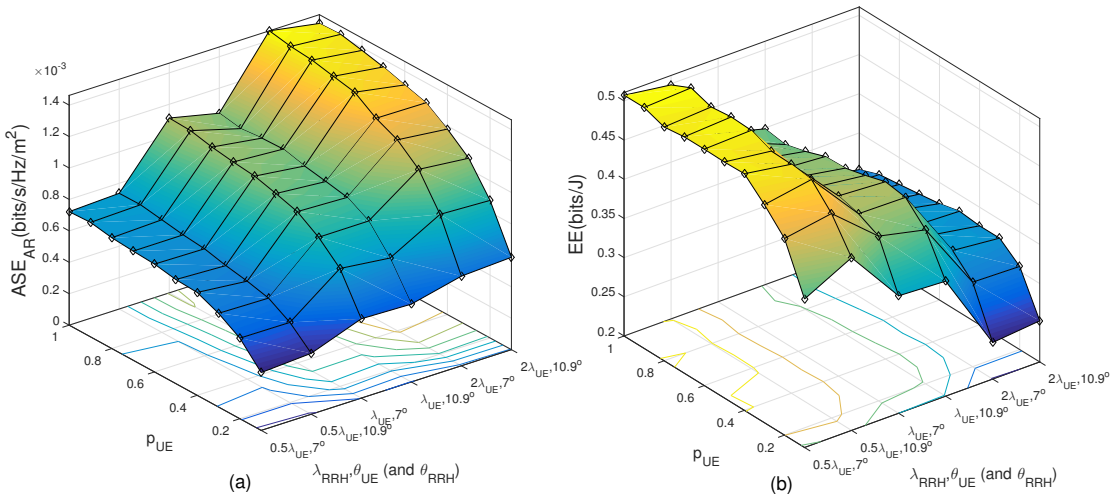


Fig. 5.8: ASE and EE variation with UE selection parameter, RRH densities and transmission beamwidths.

of RRH deployments increases. The overall trend for ASE with regards to p_{UE} is monotonic increase, with the increment being almost 46% as p_{UE} goes from 0.1 to 0.2. For $p_{\text{UE}} > 0.5$, we only observe a marginal gain in ASE upon further increments in p_{UE}, keeping all other network parameters constant. Additionally, we observe a marginal ASE reduction with wider antenna beamwidths as a result of higher λ_{AIRRH} when θ_{UE} (and / or) θ_{RRH} are increased.

The network wide EE plotted as Fig. 5.8 (b) peaks on average for $0.4 \leq p_{\text{UE}} \leq 0.6$. As expected, the EE results show opposite trend for dense RRH deployments as a high number of activated mmWave RRHs contribute towards additional energy costs for the operator. There is no observable link between EE and the UE (and RRH) transmission beamwidths. We observe a slight increase in EE for wider beamwidths when $\lambda_{\text{RRH}} = \frac{1}{2}\lambda_{\text{UE}}$ and $\lambda_{\text{RRH}} = 2\lambda_{\text{UE}}$, but a decrease in the case of $\lambda_{\text{RRH}} = \lambda_{\text{UE}}$. From a network operator's perspective, the findings highlight the necessity of a SON implementation in order to determine the right balance between maximizing net throughput and minimizing cost per bit.

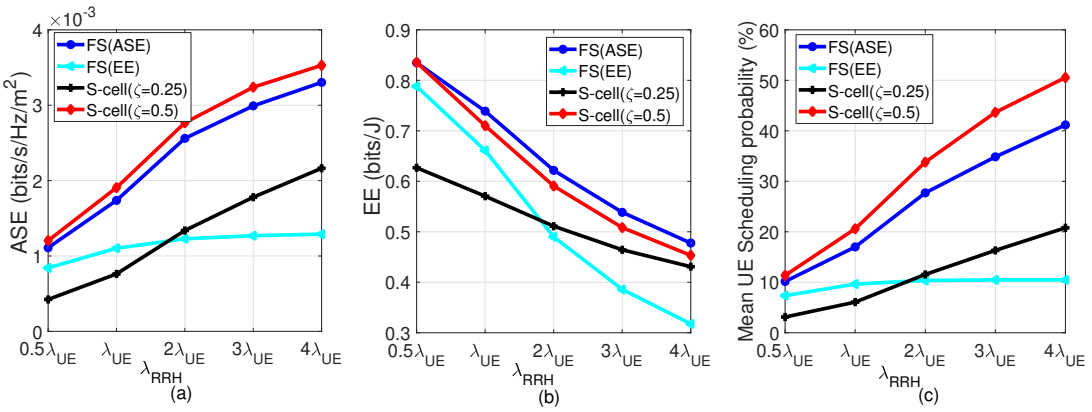


Fig. 5.9: (a) ASE, (b) EE and (c) UE scheduling ratio comparison of non-elastic user-centric [4] and proposed user-centric S-cell approaches under different RRH densities.

5.8.5 Performance Comparison with Fixed Size User-Centric Networks

Fig. 5.9 shows the performance gains in terms of adaptive rate area spectral efficiency (Fig. 5.9 (a)), energy efficiency (Fig. 5.9 (b)) and mean UE scheduling ratio (Fig. 5.9 (c)) for the proposed UE-centric S-cell architecture in comparison to the non-elastic fixed sized circular user-centric service regions proposed in earlier works [4]. The network models with fixed user-centric regions in [4] to maximize ASE and EE are referred to as FS(ASE) and FS(EE) respectively. The S-cell user-centric network used for comparison is configured at $p_{\text{UE}} = 1$ and $\zeta = \{0.25, 0.5\}$. Clearly, the proposed model at $\zeta = 0.5$ outperforms both extremes of the earlier "one-size fits all" strategy in terms of system throughout at dense mmWave RRH deployment scenarios at the cost of marginal EE loss as compared to FS(ASE). The user-centric S-cell at $\zeta = 0.25$ exhibits higher efficiency than FS(EE) for all the three measures at $\lambda_{\text{RRH}} \geq 2\lambda_{\text{UE}}$. This is because denser deployments reduce empty S-cells and due to shorter UE-RRH spatial distance, higher throughput is achieved which increases both ASE and EE. In addition to higher aggregate throughput, Fig. 5.9 (c) shows that the user-centric S-cell network also reduces

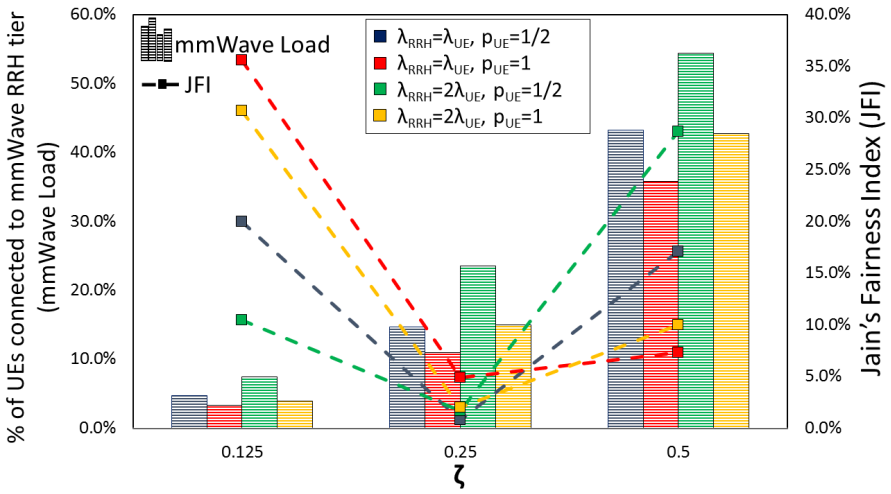


Fig. 5.10: UE load distribution between MBS and RRH tiers and achievable throughput fairness index in a two-tier user-centric S-cell network.

mean waiting time for an arbitrary UE from 2.5 TTIs in FS(ASE) to 2 TTIs when $\zeta = 0.5$. This can be traced to the non-conflicting nature of the proposed user-centric S-cell design where all the UEs that have at least one mmWave RRH within their S-cell are scheduled on DL. On the other hand, scheduling success in the non-flexible user-centric model in [4] depends upon both the probability of cell overlap with nearby UEs and probability of presence of a serving mmWave RRH within the user-centric cell.

5.8.6 ASE, EE Performance in Multi-Tier User-Centric S-cell Networks

In Fig. 5.10, we compare the inter-tier load distribution for three different S-cell sizes at $\zeta = 1/8$, $\zeta = 1/4$ and $\zeta = 1/2$. $\zeta = 1/2$ yields the most proportionate UE distribution between the MBS and mmWave RRH tiers. Apparently, setting a high ζ value seems the most obvious choice for offloading a congested MBS-tier. This is because a higher ζ pushes a large number of UEs from the MBS to mmWave RRH tier. A large ζ value not only allows more UEs connected served by the RRH tier

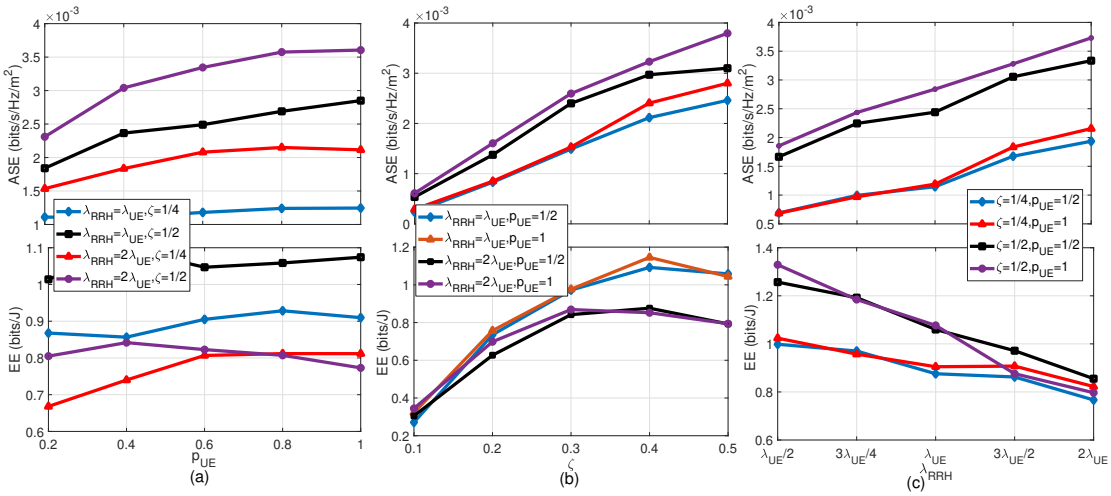


Fig. 5.11: ASE and EE trends with different p_{UE}, ζ and λ_{RRH} values.

to avail the wider mmWave spectrum, but also reduces the average interference at a typical UE due to highly directional transmission in mmWave RRH tier as opposed to omni-directional transmission from interfering MBSs in the sub-6 GHz MBS tier.

To assess the uniformity in the achievable throughput amongst UEs, we apply the Jain's fairness equation [94] on the UE throughput in a given TTI. The Jain's fairness index (JFI) values plotted in Fig. 5.10 exhibit highest values at $\zeta = 1/8$ when a large proportion of UEs is connected to the MBS tier. The JFI dips below 10 % for all the considered RRH deployment and p_{UE} scenarios at $\zeta = 1/4$ when the UEs in mmWave connectivity are mostly within LOS boundary due to smallness of the UE-centric S-cell regions. As a result, there is a large disproportion between SINRs of UEs connected to mmWave RRHs and sub-6 GHz MBS tier. Nevertheless, the fairness index improves at $\zeta = 1/2$ when the number of UEs connected to mmWave RRHs increases, particularly in the NLOS regions. Another interesting observation from the JFI results is that while a dense RRH deployment and sparse UE selection maximizes fairness at $\zeta = 1/2$, the vice versa exhibits highest fairness at $\zeta = 1/8$. This trend can be explained from the inter-tier

UE distribution graph. The JFI is maximized (and UE throughput disparity is minimized) when we have the least mmWave load percentage at $\zeta = 1/8$. However, at $\zeta = 1/2$, higher mmWave tier percentage results in more UEs experiencing similar QoS, and hence increasing overall JFI.

Hitherto, we have seen how the network level ASE and EE behave with variations in the three adjustable network parameters, i) S-cell size (characterized via ζ), ii) UE selection probability (p_{UE}), and iii) mmWave RRH deployment density (λ_{RRH}). The question that begs further analysis is whether there is an optimal combination of these parameters which simultaneously maximizes the network wide ASE and EE. Fig. 5.11 plots the adaptive rate ASE and EE for the multi-tier user-centric S-cell network under variation, within practical range, of each parameter while keeping the other two constant. The overall trend for ASE is positive for increments in each of the three parameters. However, the percentage ASE gain resulting from a unit increase in p_{UE} is far less as compared to the ASE gains achieved from increasing ζ and λ_{RRH}. Moving on to multi-tier EE, we observe no visible trend with p_{UE}. Just like the single tier network, the EE behaves in the exact opposite fashion as compared to ASE with variation in RRH density. As the S-cell size increases from $\zeta = 0.1$ to $\zeta = 0.4$, we observe almost 4 times gains in EE. However, on further increment to $\zeta = 0.5$, there is a decrease in EE for all four network configurations involving different p_{UE} and λ_{RRH} combinations. Although larger S-cell sizes increase the sum throughput due to a high proportion of UEs connected to mmWave RRHs, this is accompanied by a higher percentage of NLOS mmWave links. The high pathloss associated with NLOS mmWave connections limits the throughput gains which is eventually overshadowed by the additional power consumption due to a higher number of activated RRHs. The results in Fig. 5.11 once again re-emphasize the need of an optimization framework within

an intelligent SON engine integrated in a central entity to dynamically adjust ζ, p_{UE} and λ_{RRH} and maximize the formulated network wide ASE-EE tradeoff utility. The next chapter will be dedicated to designing such an optimization framework, and evaluation of its convexity and ability to yield global optima for a pre-fixed range of investigated design parameters.

5.8.7 SON framework for network-wide efficiency tradeoff

In the preceding chapters, we observed the SINR gains offered by our user-centric S-cell design. On one hand, the average user SINR (or QoE) decreases with an increase in p_{UE}, however on the other, the sum throughput increases monotonically with p_{UE}, particularly at high ζ regimes. This is because a higher p_{UE} corresponds to a larger population of UEs being candidates for DL scheduling in a specific time instance. Similarly, a larger ζ and λ_{RRH} increases the sum throughput by virtue of a lower p_{emp}. Due to mmWave beamforming and interferer thinning, increased RRH deployment density does not impact the interference environment to the same extent as omni-directional sub-6 GHz Rayleigh environment and the network continues to operate in noise limited mode. While an operator may achieve high throughput targets through dense mmWave network deployment with our proposed model, it severely compromises on the network energy efficiency. Additionally, one cannot ignore the increased capital expenditures (CAPEX) due to large scale RRH deployment and operating expenditure (OPEX) on account of power consumption of the denser mmWave tier. A centralized SON engine can be leveraged to provide real-time optimal network parameters, i.e. RRH deployment density, S-cell size and the user selection probability according to the operator's desired throughput-cost balance. An example of such framework is presented in Fig. 5.12 that may operate at different modes: (i) high throughput mode dur-

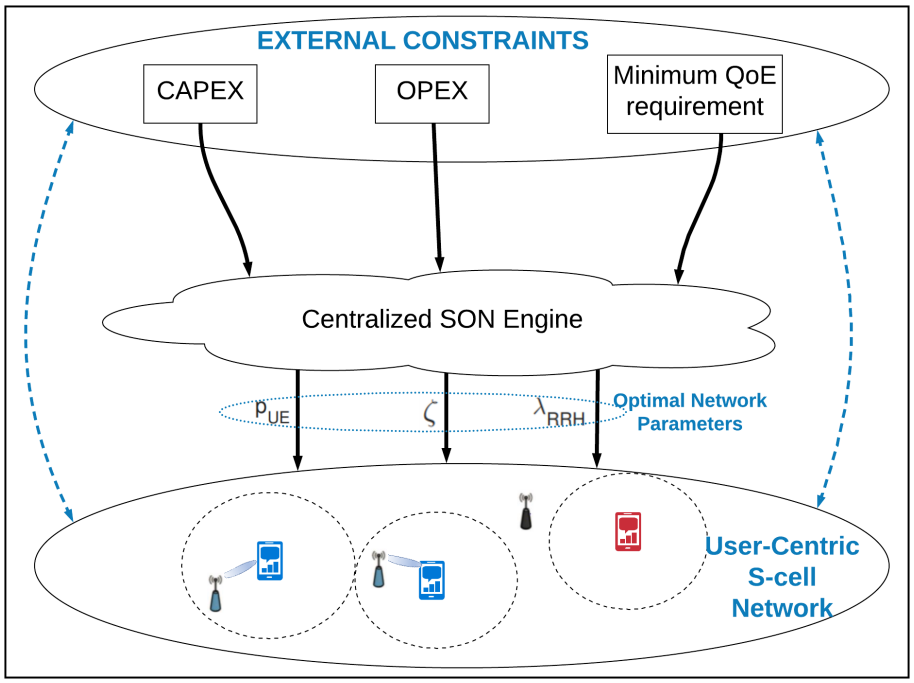


Fig. 5.12: SON Framework for dynamic network operating mode adjustment in a user-centric S-cell network.

ing peak traffic durations, and (ii) energy efficient mode during night hours. The SON engine shall also take into account all related costs including mmWave cell deployment expenditure, power consumption, and maintenance costs within the optimization framework.

5.9 Conclusion

In chapter 5 of this dissertation, we proposed a user-centric Stienen cell network architecture capable of offering higher system capacity and improved received signal quality in dense deployment scenarios, compared to non user-centric conventional cellular architectures. Since current models for cellular network performance analysis are not applicable to the proposed architecture, we developed a comprehensive statistical framework for analytical characterization of the area spectral efficiency

and energy efficiency of the novel cellular architecture. We presented user-centric scheduling schemes for two network layouts: i) a single-tier network where each UE is served by a single mmWave RRH that resides within its S-cell, and ii) a multi-tier network with a sub-6 GHz macro base station tier overlaid on top of the mmWave RRH tier. While mmWave RRHs are dedicated to serve users in close proximity, the users with void Stienen cells are connected to the macro BSs in the multi-tier design, resulting in a higher average user quality of experience. We further characterized the distance distribution between an arbitrary UE and the angularly interfering RRHs that cause significant interference within mmWave networks.

Our analysis validates the usefulness of the proposed architecture in the form of large SINR gains achieved by virtue of minimizing interference within virtual Stienen zones around scheduled users. We also observed higher network level capacity and lower scheduling delays when compared with a benchmark user-centric architecture from literature. Numerical results based on the derived expressions reveal practical design insights by characterizing the interplay among three design parameters namely S-cell size, user service probability and RRH density; and the network wide KPIs; i.e. ASE and EE. Extensive simulation results show that from the three design parameters, the KPIs are more sensitive to the Stienen cell size and RRH deployment density. While the ASE is optimized in a dense RRH deployment scenario, the EE deteriorates with increase in the RRH density. Similarly, the ASE and EE show contrasting results with respect to Stienen cell size. Setting the Stienen cell size to its maximum possible limit allows a larger number of users to be connected to mmWave RRHs, thereby increasing the sum throughput of the network. However, the EE decreases if the Stienen cell size is increased beyond a certain optimal value due to higher network power consumption caused

by additional mmWave RRH activations.

To fully optimize a network efficiency metric, we advocate a SON enabled entity that is capable of dynamic adaptation of the modeling parameters to offer higher throughput or optimal energy utilization, whichever is desired by the network operator in a given spatio-temporal region. The following chapter carries this framework ahead and includes formulation for concurrent optimization of throughput and power consumption cost in the proposed network, as well as analysis of the Pareto optimal tradeoff between the two network efficiency measures.

CHAPTER 6

DNN-Based Learning Approach for Pareto Optimal ASE-EE Tradeoff

6.1 Introduction

In the preceding chapter, we analyzed a user-centric Stienen cell architecture for both single-tier and multi-tier cellular networks. We analyzed the network performance for area spectral efficiency and energy efficiency while optimizing with three network-level parameters: UE selection probability, S-cell size factor and RRH deployment density. It was demonstrated that while RRH density has completely contrasting trends for ASE and EE, the same is not true for the other parameters. In this chapter, we investigate methods to design utility optimization for a multi-tier Stienen cell architecture for the same three parameters. We first design a stochastic optimization problem and then investigate the real-time computational efficiency and complexity of the exhaustive search space optimization. We then propose an approach based on modular deep neural networks (DNN) to perform near-optimal network parameter selection with real-time computational ability.

6.1.1 Background

A multitude of aspects in wireless communications, such as fading, shadowing and user mobility, cause the optimization of one or more network-wide KPIs to be uncertain. In addition to this, varying user data demands add another layer of

non-deterministic elements to the optimization of wireless networks' performance. One possible solution is to design the network for peak data demands. However, this solution is unfeasible when we consider the associated costs to the network operator, as a result of additional network deployment or the utilization of a higher bandwidth spectrum for its service. Furthermore, due to the constant mobility of user terminals and the dynamic propagation environments, wireless channel gains are stochastic and dynamic. A significant amount of high data rate communication in future networks is expected to take place in the high bandwidth mmWave spectrum. The deterministic optimization of resource planning for peak data demands is thus infeasible.

Stochastic optimization is a tool used to minimize (or maximize) an objective function in the presence of randomness in the optimization process. The randomness from the perspective of wireless networks is caused by random users, small BS locations and the mobility factor. The uncertain parameter may be modeled as a random variable with a known distribution, and the expected value of the random variable may be minimized (or maximized) in the objective function. For worst-case optimization, the mean value of the random variable is replaced by the supremum and infimum for minimization and maximization problems respectively. Several mathematical techniques to solve problems associated with uncertainty have been discussed in the existing literature. In this section, we discuss some examples of those techniques. The first technique is the networked-Markov decision process, which optimizes model-based problems. The randomness is modeled via Markov chains, whereby the future state of the system is independent of the past state if the present state is given [95]. Mathematically, given the state space $\mathbb{S}$, action space $\mathbb{A}$ and reward / cost function $R(.)$, at the end of each time slot, the system jumps from the current state to the next state $s' \in \mathbb{S}$ with the prob-

ability $\Pr(s|s,a)$ and the associated reward/cost, i.e. $R(a,s,s)$. Over time, the Markov decision process will either minimize the cost or maximize the reward. The other mathematical tool is called discrete stochastic optimization, which is used when the input to the optimization function is a sequence of random variables. One way to solve these problems is by exhaustive search procedure, which is also known as simulation optimization [96]. In the exhaustive search method, the empirical average for the objective function is taken for a large number of possible network scenarios. By law of large numbers [97], the optimal solution is the one that maximizes the mean of the objective function. However, the exhaustive search procedure has some disadvantages. Firstly, it requires an extremely large amount of computation, which is time- and resource-intensive. Secondly, due to the time-varying nature of wireless channels and the dynamic nature of future wireless networks, where small cell RRHs can be deployed on an impromptu basis, a set of optimal parameters will perform sub-optimally almost immediately after being ascertained. Some techniques to reduce the complexity have been proposed in the existing literature, such as ranking and selection [98], multiple comparison procedures [99] and random search [100]. Our focus in this chapter is on a relatively new paradigm of online learning to solve such optimization problems in real-time with a close-to-optimal performance.

Due to limitations in capacity, wireless resource management has always been a significant challenge for network operators around the world. The main problem with numerical optimization is its high computational complexity, which prevents its real-time implementation. To overcome this barrier, online learning may be employed whereby the input and output of a resource allocation algorithm are fed into a DNN [101]. The DNN then learns the non-linear mapping between the two and approximates it for any new condition. If this mapping can be learned with

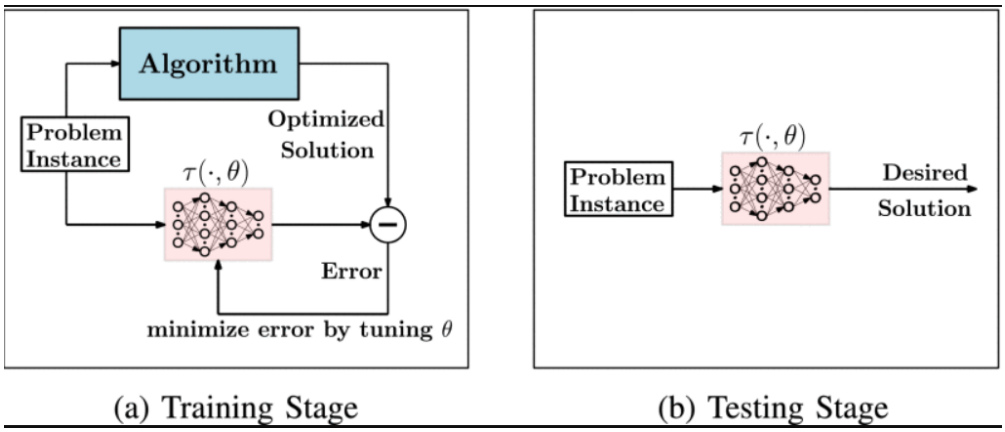


Fig. 6.1: Modular DNN approach for real-time Optimization [5]

high accuracy, real-time resource allocation could be made possible. Thus instead of following the iterative process in which the algorithm takes network parameters, such as user density, channel gains and SINR as inputs and runs through multiple iterations to yield the optimal resource allocation strategy, a trained DNN will output the optimal strategy in a few simple steps when a highly accurately trained model is used. DNN-enabled resource optimization in wireless networks provides two advantages: (i) real-time optimization is possible due to significantly lower computational requirements for retrieving the optimal parameters from a trained model; and (ii) model training can be performed offline by running the optimization algorithm on a simulated data set, hence the training process does not interfere with the network's operation. Recently, significant interest in the research community has been devoted to the application of DNNs to approximate iterative optimization problems in the wireless domain [102, 103, 104, 105, 106, 107].

As illustrated in Fig. 6.1 [5], this online training is a two-step process. In the first step, which is the training stage, the DNN model is trained for instances of network parameters and the optimized solution is generated through an already established algorithm. These algorithms vary with the type of application; for instance, the well-known iterative water-filling algorithms [108, 109] are used for

power control. Other examples include the WMMSE algorithm [110] and the semi-definite relaxation scheme [111] for transmitting and receiving beamforming optimization. Similarly, sparse optimization schemes [112] may be employed for BS clustering in the case of the user-centric architectures described in chapters 3 to 5 of this dissertation. In the second stage of this process, which is the testing stage, an unknown network instance is provided as input to the trained DNN to yield the desired optimal configuration. Several studies have applied DNNs for various problems in communications [113, 114, 115, 116, 117]. However, with regard to the optimization problem at hand, there are multiple related questions that must be answered. Some examples of these questions are as follows: (i) Is the optimization problem "learnable"? In other words, can it be approximated arbitrarily well by a DNN?, ii) How many layers and neurons per layer are needed to approximate an algorithm? At what level does the DNN start to overfit and yield poor test stage results?, and iii) What is the sensitivity of the trained model to real data? Hyperparameter tuning that includes adjustment in the depth and width of the employed DNN must also be performed to analyze under the dimensions under which a DNN minimize the approximation error rate.

6.1.2 Architectural Overview

In this chapter, we apply the learning based approach to a stochastic optimization problem that aims to find the Pareto optimal balance between ASE and EE in a multi-tier user-centric S-cell network. The modular learning approach starts by generating a utility for a given representation of input parameters. The utility value is a function of both the first and second moments of the optimization function. We simulate multiple network snapshots to find the optimal parameters for each network scenario. Once sufficient data has been generated, it is used to

train a DNN that aims to produce utility value for unknown (test) scenarios. If a set containing all possible combinations of network parameters passes through the trained DNN model, we will receive utility for each combination. Simply applying the maximum function will yield the output parameters from the DNN approach. This is the benchmark deep learning process that we employ in our study.

To improve upon this benchmark DNN output, we apply further refinements to the basic DNN model. The first refinement is hyperparameter tuning. In simple terms, hyperparameters are the variables that determine and shape of the DNN structure. Some of these hyperparameters include the number of hidden layers, the learning rate and the number of neurons in a given hidden layer. Another design question concerns how the hyperparameter tuning should be performed. In this study, we implement hyperparameter tuning using the genetic algorithm, (GA). The GA is a meta-heuristic technique to solve NP-hard optimization problems [118]. Other meta-heuristic techniques include simulated annealing [119, 120, 121], particle swarm optimization [122], Taguchi's method [123], and ant colony optimization [124], which yield near optimal solutions for selected sets of parameters. Another refinement involves analyzing the marginal contribution of each attribute to the output, i.e. the final utility value of the optimization function. In the context of this study, if each feature is considered as a player with the prediction being the payout, then according to coalition game theory, the Shapley value determines the fairness in payout distribution among the features [125, 126]. From the feature value contributions, we not only obtain the features that have a higher relative influence on the utility function in comparison to others, but also obtain regions within the important feature that have a greater influence on the predicted value. Synthetic data generation using generative adversarial networks (GANs) is performed in those high sensitivity regions. GANs consist of two neural networks,

called the generator and discriminator, which have adversarial goals in the iterative learning process. GANs have recently been used to augment available dataset to improve the model performance in both regression and classification use cases [127, 128]. The cumulative data including the original data and the synthetic data from GANs, is passed through the same process, i.e., first through a simple DNN and then through the hyperparameter tuned DNN model. Synthetic data creation enables us to avoid the laborious process of data generation using network-wide simulations for a large number of Monte Carlo distribution scenarios to yield the optimal parameter set for a single set of network features.

6.1.3 Research Objectives

Our prime objective in this part of the dissertation is to evaluate the performance of DNN-based techniques in order to optimize user-centric Stienen cell networks at a reduced computational complexity. We first present a stochastic optimization framework and then formulate a modular approach to the sequential improvement of the trained DNN model. The research questions that this chapter aims to answer as follows:

- What is a plausible utility optimization framework that yields parameter values for Pareto optimal balance between ASE and EE in a user-centric S-cell network?
- Can an online learning-based approach be employed to enable real-time optimization of the designed problem?
- What techniques can be used to improve the performance of the baseline DNN model?

- Which design parameters have a greater influence on the output utility?
- How close is the performance of the DNN model-based approach to the global optima?

To answer these questions, detailed analysis is presented in the following chapters, starting with a stochastic optimization framework. We then perform convexity and computational complexity analyses. A modular DNN approach is presented that starts with a DNN and is further refined with hyperparameter tuning and data augmentation using GANs.

6.2 Network Model

6.2.1 Spatial Model

We consider the downlink of a two-tier ultra-dense network consisting of sub-6 GHz MBSs, mmWave RRHs and UEs spatially distributed across its foot-prints. The spatial distributions for MBSs, RRHs and UEs are modeled using independent homogeneous Poisson point processes (HPPPs) Φ_{MBS}, Φ_{RRH} and Φ_{UE} with intensities λ_{MBS}, λ_{RRH} and λ_{UE} respectively. The UE locations act as generating points for the UE-centric Voronoi tessellation. Each RRH is associated with a UE based on its physical location on the Voronoi plane. This implies that each RRH can, at most, serve a single UE that is spatially closest to the RRH during a transmission time interval (TTI). Furthermore, to model realistic MBS deployment, we induce repulsion between MBS locations in the considered network. This is done by modeling it as a type II Matern hardcore process [67], whereby we choose a subset of the original PPP with a distance constraint. Mathematically, the thinned PPP representation of MBS is given by

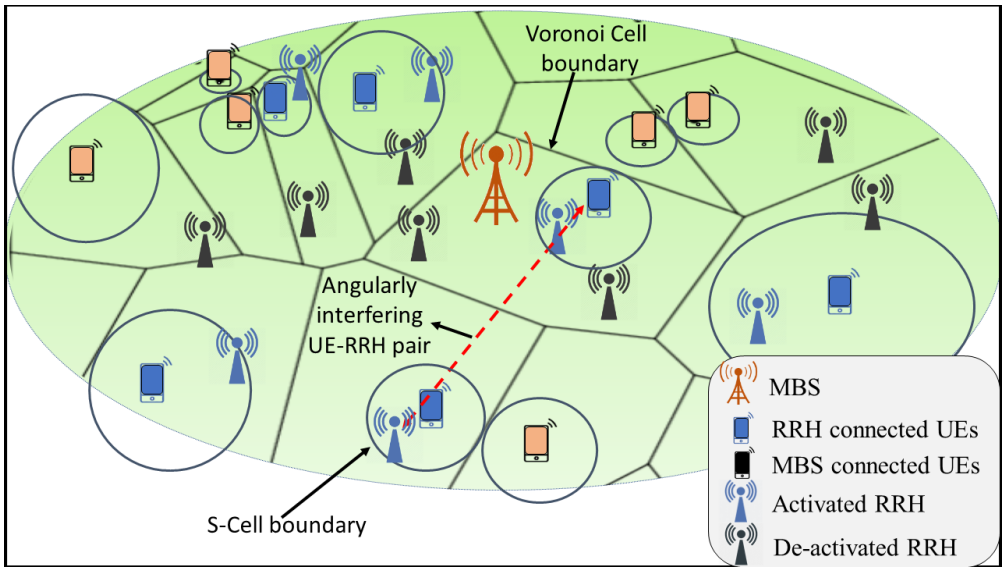


Fig. 6.2: The two-tier UE-centric S-cell architecture. A UE is served by a mmWave RRH in the case of non-empty S-cells, and is otherwise served by the closest MBS.

$$\lambda_{\text{MBS}}^{\text{EF}} = \frac{1 - \exp(-4\pi\lambda_{\text{MBS}} R_{\text{MBS}}^2)}{4\pi R_{\text{MBS}}^2}. \tag{6.1}$$

6.2.2 Pathloss Model

Based on the system model explained in the previous section, UEs may either be connected to sub-6 GHz MBSs or mmWave RRHs. However, the UEs connected to mmWave RRHs will have severe blockage effects in comparison to those connected to the MBSs. To model the mmWave blockage, we utilize a LOS ball model that assumes all UEs within a ball of a specified radius around a mmWave RRH to be in the LOS region; and the UEs outside the ball region to be in the NLOS region. The pathloss model for the overall system is given by:

$$
\begin{aligned}
PL(r) &= A_{\mathrm{LOS}} r^{\alpha_{\mathrm{LOS}}};\ \text{if UE connected to mmWave RRH and } r \leq R_o, \text{ and} \\
&= A_{\mathrm{NLOS}} r^{\alpha_{\mathrm{NLOS}}};\ \text{if UE connected to mmWave RRH, and } r > R_o, \\
&= A_{\mathrm{MBS}} r^{\alpha_{\mathrm{MBS}}};\ \text{if UE connected to MBS}
\end{aligned} \tag{6.2}
$$

where α_{LOS}, α_{NLOS} and α_{MBS} are the terrain and operating frequency-dependent pathloss exponents, LOS mmWave links, NLOS mmWave links and MBS connection links respectively. The intercepts A_{LOS}, A_{NLOS} and A_{MBS} are all assumed to be unity. Note that high α_{NLOS} values result in sufficiently large $PL(r)$, effectively causing the UE-RRH links to be in outage as r increases.

6.2.3 Antenna Gain Model

Due to free space attenuation and sensitivity to blockages, directional transmission is essential for mmWave communication. Given $\mathrm{M} \in \{\mathrm{UE}, \mathrm{RRH}\}$, let G_{M}, g_{M} and θ_{M} denote the main lobe gains, side lobe gains and half power beamwidths (HPBWs), respectively, of the UEs and RRHs. Thus, the directivity gain for a desired UE-RRH link is $G_{\mathrm{UE}} G_{\mathrm{RRH}}$. Assuming sectorized antenna beam patterns at both UEs and RRHs, the probability of exact alignment for the uniform distribution of antenna beams is $\frac{\theta_{\mathrm{UE}} \theta_{\mathrm{RRH}}}{4\pi^2}$. The other extreme is when both of the UE or RRH beams are misdirected, the probability of which is $\frac{(1-\theta_{\mathrm{UE}})(1-\theta_{\mathrm{RRH}})}{4\pi^2}$. The directivity gain G_{I} for an arbitrary RRH connected UE and an interfering RRH is a discrete random variable with the probability distribution given in [89] and expressed as:

$$\mathbb{E}(G_{\mathrm{I}}) = \frac{\theta_{\mathrm{UE}}\theta_{\mathrm{RRH}}}{4\pi^2} G_{\mathrm{UE}} G_{\mathrm{RRH}} + \frac{\theta_{\mathrm{UE}}}{2\pi}\frac{1-\theta_{\mathrm{RRH}}}{2\pi} G_{\mathrm{UE}} g_{\mathrm{RRH}} + \frac{1-\theta_{\mathrm{UE}}}{2\pi}\frac{\theta_{\mathrm{RRH}}}{2\pi} g_{\mathrm{UE}} G_{\mathrm{RRH}} + \frac{(1-\theta_{\mathrm{UE}})(1-\theta_{\mathrm{RRH}})}{4\pi^2} g_{\mathrm{UE}} g_{\mathrm{RRH}}. \tag{6.3}$$

The sub-6 GHz MBSs have omni-directional antennas with consistent antenna gain in all directions. Beamforming is not required for communication on sub-6 GHz spectrum.

6.2.4 Small Scale Fading

For MBS connected links, we assume Rayleigh fading with unit mean exponential channel gain. For mmWave links, we assume independent Nakagami fading for each UE-RRH link, with N_{L} and N_{N} representing the LOS and NLOS parameters respectively. The small-scale fading in signal power given by $|h|$ under the Nakagami assumption is a normalized Gamma random variable. Shadowing is ignored for the sake of analytical tractability.

6.2.5 User Association Mechanism

Each UE has an associated S-cell with a diameter equal to the distance of the neighboring UE. If the S-cell contains at least one mmWave RRH, the UE is connected to the RRH based on the smallest pathloss criterion. In case there is no RRH within the S-cell, the UE is connected to an MBS, also on the basis of the smallest pathloss. The smallest $PL(r)$ criterion ensures the maximum average SINR in each of the scheduled UEs. Note that the MBS distribution considered for UE connections is $\Phi_{\mathrm{MBS}}^{\mathrm{EF}}$, which is the Matern hardcore PPP that represents the finally deployed MBSs with a minimum distance constraint from the closest MBS.

6.3 Optimization Framework

6.3.1 SINR

From the user-association mechanism elaborated in chapter 6.2.5, we can divide the UE population into those that are connected to mmWave RRHs and the ones that are connected to sub-6 GHz MBSs. We represent these two disjoint user populations as PPPs given by $\Phi_{\text{UE,MBS}}$ and $\Phi_{\text{UE,RRH}}$ for MBS and RRH connected UEs respectively. The downlink SINR equation for an arbitrary UE will thus be given by:

$$\begin{aligned}\text{SINR} &= \frac{h_{\boldsymbol{oRRH}} G_{\text{UE}} G_{\text{RRH}} (r^{-\alpha_o})}{kTB_{\text{RRH}} + \sum_{\boldsymbol{i}\in\Phi_{\text{IRRH}}} h_{\boldsymbol{iRRH}} \mathbb{E}(G_{\text{I}})(r^{-\alpha_i})}; \text{if } \mathbf{x} \in \Phi_{\text{UE,RRH}}, \\ &= \frac{h_{\boldsymbol{oMBS}} (r^{-\alpha_{MBS}})}{kTB_{\text{MBS}} + \sum_{\boldsymbol{i}\in\Phi^{\text{EF}}_{\text{MBS}}} h_{\boldsymbol{iMBS}} (r^{-\alpha_{MBS}})}; \text{if } \mathbf{x} \in \Phi_{\text{UE,MBS}}.\end{aligned} \tag{6.4}$$

α_o and α_i take the value of α_{LOS} and α_{NLOS} in case of $r \leq R_o$ and $r > R_o$ respectively. T is the Kelvin temperature and B_i where $i \in$ MBS,RRH is the bandwidth for MBS and mmWave spectrum respectively.

6.3.2 ASE

The area spectral efficiency is a function of the downlink SINR at the UEs and the SINR threshold criteria set by the operator. Mathematically, the ASE in this adaptive rate transmission situation may be expressed as

$$\text{ASE} = \frac{1}{\mathbb{A}} \sum_{u\in\Phi_{\text{UE}}} \log_2(1 + \text{SINR}_u); \text{ if SINR}_u > \gamma_{th}, \tag{6.5}$$

where $\mathbb{A}$ is the total network area under consideration. We assume a consistent SINR threshold γ_{th} for both the RRH and MBS connected UEs.

6.3.3 EE

We recap the mathematical expressions for the power consumption and energy efficiency of a multi-tier Stienen cell network given in chapter 5.7. We leverage work from project EARTH [73] to express the total power consumption of the two-tier network as

$$\boldsymbol{P} = \sum_{\Phi_{\mathrm{MBS}}^{\mathrm{EF}}} P_{\mathrm{MBS}} + \sum_{\Phi_{\mathrm{RRH,act}}} P_{\mathrm{RRH,act}} + \sum_{\Phi_{\mathrm{RRH,deact}}} P_{\mathrm{RRH,deact}}. \tag{6.6}$$

$\Phi_{\mathrm{RRH,act}}$ and $\Phi_{\mathrm{RRH,deact}}$ are the RRHs in activated and de-activated state respectively depending upon their presence inside a UE's S-cell and subsequent selection for service via the smallest PL criterion. P_{MBS}, $P_{\mathrm{RRH,act}}$ and $P_{\mathrm{RRH,deact}}$ are the consumed power by an MBS, activated RRH and de-activated RRH respectively. Note that we are considering an always ON MBS deployment to avoid coverage holes and provide uninterrupted control/signaling to the UEs. The power saving in the proposed architecture will thus come from intelligently turning OFF mmWave RRHs that are not providing DL data services to UEs.

Now the individual power consumption in each state is given by the following equations:

$$P_{\mathrm{MBS}} = A_{\mathrm{MBS}} P_{\mathrm{MBS,Tx}} + B_{\mathrm{MBS}}. \tag{6.7}$$

$$P_{\mathrm{RRH,act}} = A_{\mathrm{RRH}} P_{\mathrm{RRH,Tx}} + B_{\mathrm{RRH}}. \tag{6.8}$$

$$P_{\mathrm{RRH,deact}} = B_{\mathrm{RRH}}. \tag{6.9}$$

B_{MBS} (and B_{RRH}) denotes the fixed power consumption of an MBS (and RRH). This is the energy cost which is bore by the network regardless of the number of UEs requesting DL service. The coefficients A_{MBS} (and A_{RRH}) lump together frequency dependent response of a power amplifier and several other factors within MBS (and RRH). The coefficients A_{i} and B_{i}, where $\mathrm{i} \in \{\mathrm{MBS,RRH}\}$ are expressed as

$$A_{\mathrm{i}} = \frac{1}{\eta_{\mathrm{i}}^{\mathrm{PA}}(1-\sigma_{\mathrm{i}}^{\mathrm{feed}})(1-\sigma_{\mathrm{i}}^{\mathrm{MS}})(1-\sigma_{\mathrm{i}}^{\mathrm{DC}})(1-\sigma_{\mathrm{i}}^{\mathrm{cool}})}, \text{ and} \tag{6.10}$$

$$B_{\mathrm{i}} = \frac{P_{\mathrm{i,RF}} + P_{\mathrm{i,BB}}}{(1-\sigma_{\mathrm{i}}^{\mathrm{MS}})(1-\sigma_{\mathrm{i}}^{\mathrm{DC}})(1-\sigma_{\mathrm{i}}^{\mathrm{cool}})}. \tag{6.11}$$

For detailed explanation of the power consumption parameters, readers are referred to [73]. The network wide EE is analyzed for adaptive rate transmission scenario in our work and expressed mathematically as

$$EE = \frac{\mathbb{A}[\mathrm{ASE}]}{\boldsymbol{P}}. \tag{6.12}$$

6.3.4 Problem Formulation

From our analysis in chapter 5, we noticed that ASE and EE in a multi-tier user-centric S-cell network have contrasted trends with at least the RRH deployment density. Considering that ASE and EE are game players in a cooperative game, we model their tradeoff as a Nash product (NP), which is mathematically expressed

as

$$\text{NP} = [\text{ASE}]^{\beta}[EE]^{1-\beta}. \tag{6.13}$$

So we converted a multi-objective optimization problem which consisted of maximizing both ASE and EE into a single objective optimization problem with $\beta \in [0, 1]$ assigning weight to each objective. $\beta = 0$ reduces the optimization problem into EE-maximization while $\beta = 1$ corresponds to ASE-maximization. Now, the Nash product given in Eq. (6.13) if optimized on the three design parameters, i.e. p_{UE}, 2ζ and $p_{\text{RRH,act}}$; will only be useful for a particular snapshot of the UE, RRH and MBS placements within the network. p_{UE} denotes the percentage of UEs which participate in the DL scheduling process, hence determining the average geometry of the S-cell network. 2ζ is the S-cell size factor with a max value of $2\zeta = 1$ that occurs when the S-cell is inscribed within the user-centric Voronoi cell. Finally, $p_{\text{RRH,act}}$ is the percentage of RRHs that are not sleeping, i.e. they are discoverable by the UEs. In other words, this is a percentage determining how many of the total mmWave RRHs contribute towards the RRH-tier density within the network. Note that we have a consistent range of the optimization parameters, i.e. $0 < p_{\text{UE}}, 2\zeta, p_{\text{RRH,act}} \leq 1$.

To cater for uncertainties that arise from different UE, RRH and MBS positioning as well as constant change in channel dynamics and user mobility, we employ a stochastic optimization framework. In this framework, we take several snapshots of a single network parameter set and optimize a combination of the mean utility with a penalty to the fluctuations in the utility. This mean-variance model also describes the magnitude of risk aversion, i.e. how much impact does a high variation in utility have to the overall optimization framework. The utility maximization

problem is given as:

$$[p_{\mathrm{UE}}^{opt}, 2\zeta^{opt}, p_{\mathrm{RRH,act}}^{opt}] = \max_{p_{\mathrm{UE}}, 2\zeta, p_{\mathrm{RRH,act}}} \underbrace{[\mathbb{E}(\mathrm{NP})] - \eta[\mathbb{E}(\mathrm{NP}^2) - \mathbb{E}(\mathrm{NP})]}_{\textbf{Utility}}. \tag{6.14}$$

The optimized parameter set would not only maximize the function in Eq. (6.14), but also ensure that the function value does not have rapid fluctuations with different snapshots of the network with same mean UE, RRH and MBS densities. η is the degree of penalty that a particular parameter configuration would have on fluctuations in the objective function. In the subsequent chapters, we will discuss the properties of this optimization problem and a learning approach for efficient solution to this problem.

6.4 Convexity and Computational Complexity Analysis

Determining whether an optimization function is convex is a critical aspect of determining the optimal parameters in a multi-objective non-linear optimization problem, such as Eq. ((6.14)). This is because the result of convexity analysis will determine if the problem will be solved in a feasible time frame or not. A convex optimization problem is characterized by convexity of both the objective and constraint functions in the optimization variables [129, 130]. Every stationary point in the objective function of a convex optimization problem will be a global minima. This property ensures polynomial time complexity in the number of variable.

Due to diversity in constraints and optimization parameters, most of the capacity and resource optimization problems in wireless networks are inherently non-convex in nature. Multiple heuristic techniques can be applied to solve non-convex opti-

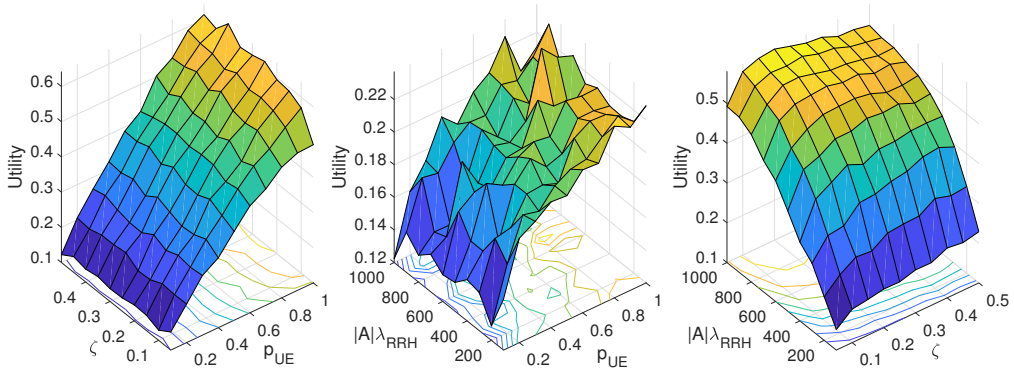


Fig. 6.3: Objective function utility with optimization parameters at $\beta = 0.2$.

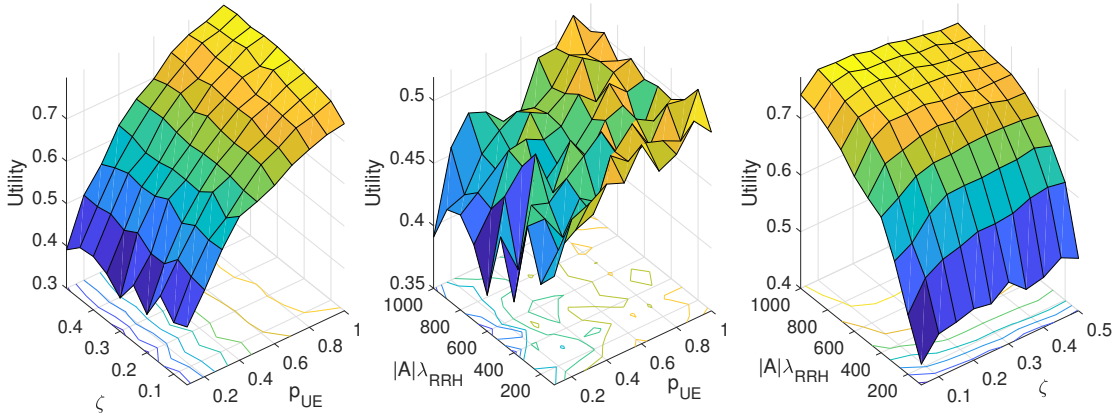


Fig. 6.4: Objective function utility with optimization parameters at $\beta = 0.8$.

mization problems. Some of these include genetic algorithms [131], particle swarm optimization [132], sequential quadratic programming and pattern search [133]. Genetic algorithms are known to be immune to initialization points and also have transient properties from non-feasible to feasible zones. However, this algorithm takes longer to converge with increase in search space, making real-time optimization impossible in large systems.

From the plot of objective function value of Eq. ((6.14)) in Figs. 6.3 and 6.4 for a search space sampling of 10 intervals for each optimization parameter, it is clearly seen that the solution space is a combination of multiple hills and valleys, thereby implying non-convexity. The non-convex behaviour can be traced back to

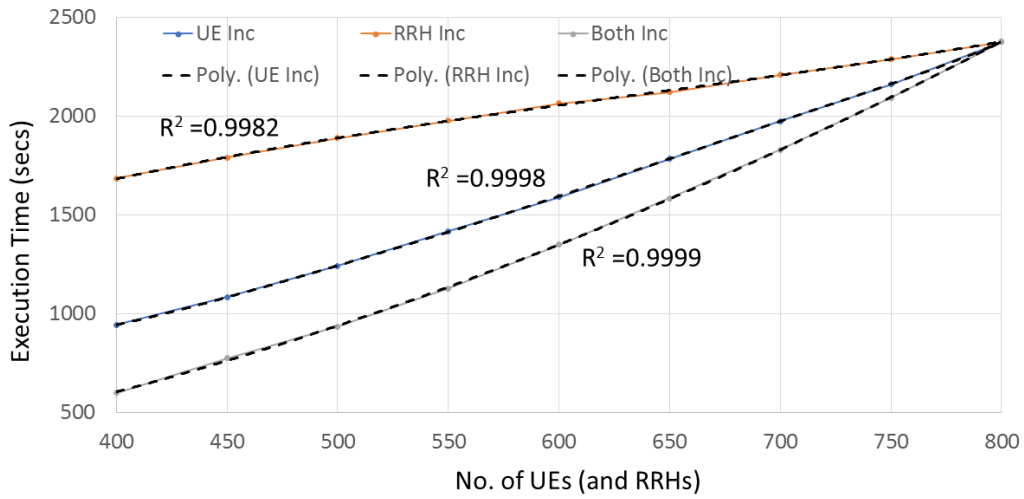


Fig. 6.5: Time complexity for a single parameter and network simulation scenario.

two major reasons: i) the piece-wise SINR function in Eq. (6.4) for the mmWave connected UEs results in abrupt shift in the user SINR, and consequentially, the objective function value, as a UE moves from the LoS region to a NLoS region, and vice vera; ii) the randomness is UE, RRH and MBS placements, in particular the MBS placements, over-shadows the overall trend of the utility with an optimization parameter. Although we see a clear increasing trend in the utility with increase in p_{UE}, the fluctuations that arise with step changes in p_{UE} give rise to local maximas. These fluctuations are independent of the number of Monte Carlo scenarios of different UE, RRH and MBS placements. The non-convexity is seen to be more severe during the interaction of ζ and $p_{\text{RRH,act}}$ at constant p_{UE} for both $\beta = 0.2$ and $\beta = 0.8$ values.

To assess the time complexity for finding the utility value for one of search space configurations in Eq. ((6.14)), we analyze the execution time for three different scenarios: i) number of UEs in the system are increasing with a constant pace keeping number of RRHs fixed (at 800 in the given figure), ii) number of RRHs

is increasing while keeping number of UEs constant (at 800), and iii) both UEs and RRHs are increasing simultaneously (from 400 to 800 in our case). All the scenarios exhibit polynomial time complexity. The execution time for the two scenarios when only UEs or RRHs fits well with a high R-squared value in case of quadratic regression. This implies that the time complexity when only UEs and RRHs in the system are changing is $\mathcal{O}([|A|\lambda_{\mathrm{UE}}]^2)$ and $\mathcal{O}([|A|\lambda_{\mathrm{RRH}}]^2)$ respectively. In case when both UEs and RRHs are increasing at the same rate simultaneously, it is seen that the increase in execution time is higher than the earlier scenarios. Polynomial curve fitting of degree 4 shows very high regression reliability. This makes sense because the time complexity in this case will be $\mathcal{O}(|A|^2[\lambda_{\mathrm{UE}}]^2[\lambda_{\mathrm{RRH}}]^2)$. Although polynomial time complexity is manageable for processing, one has to keep in mind that this is just once configuration of search space. For brute force optimization which calculates utility for all the configurations in search space, which in our case is 10x10x10 = 1000, this process is repeated for the entire search space. Therefore, brute force method will not yield real-time optimization. The DNN based approach that aims to overcome this issue is discussed in the succeeding chapters.

6.5 Modular DNN-Based Approach for Real-Time Optimization

In this section, we elaborate on the modular DNN approach that we follow to generate a reliable model for real-time optimization of the Pareto optimal ASE-EE tradeoff problem in Eq. (6.14). The steps are presented in sequential order in the following paragraphs:

- In the data preparation phase, we perform Monte Carlo simulations for a given network environment and generate utility function values for all of

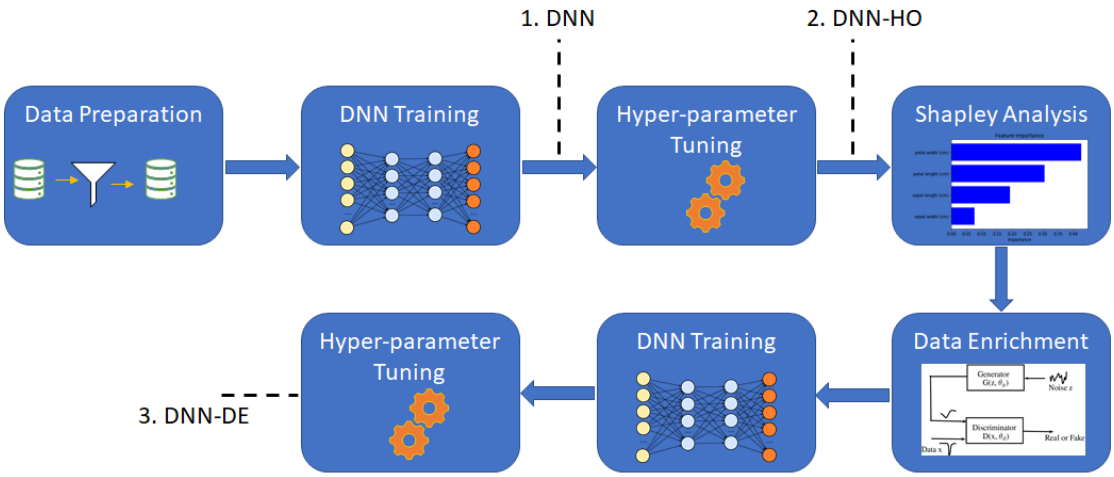


Fig. 6.6: Proposed Modular DNN methodology.

the possible network parameter configuration possibilities. As previously discussed, we have divided the intervals of the three optimization parameters p_{UE}, 2ζ and $p_{\text{RRH,act}}$ into 10 discrete levels, making the search space size 1000. The brute force (BF) analysis is conducted for a large number of network scenarios. The variables that define the network scenario are $|A|\lambda_{\text{UE}}$, $|A|\lambda_{\text{RRH}}$, $|A|\lambda_{\text{MBS}}$ and β. The aim is to have sufficient data for the DNN to learn which parameter configuration yields the best utility, for instance at low RRH densities versus high deployment density.

- Once we have the simulated the data set in order, we feed the input and output utility in a DNN to allow it to learn the non-linear mapping between the network parameters and the optimization parameters for the utility value. The motivation behind this is to enable real-time optimization in a few matrix manipulation steps, instead of the time consuming stochastic optimization, which requires hundreds of Monte Carlo simulations for each scenario in entire search space. In this initial DNN design, we propose an artificial neural network with six hidden layers of 32 neurons each. We use

a rectified linear unit (ReLU) activation function for the hidden layers and a linear activation for the output layer.

- The performance of an ANN not only depends on the type of application and quality of training data, but also on the hyperparameters chosen to train the ANN. These hyperparameters may include design variables such as the number of hidden layers, the activation functions, the training batch size and the number of neurons in each layer. Selecting suitable structure and hyperparameters improves the learning rate and accuracy of an ANN. The exhaustive grid search technique can be used to evaluate the ANN performance at all combinations of the hyperparameters and select the best combination [134]. However, the disadvantage of grid search is its long computational time [135]. Other techniques such as random search [136] and GAs [137] have been used to yield hyperparameters with similar accuracy at a much lower time complexity. In this work, we utilize genetic algorithms, that are based on the principle of natural selection, to determine the optimal hyperparameters with lower computational overheads.

- The next step in our analysis involves understanding the feature importance in the DNN learning and the interaction of important features within the DNN. We use SHAP (SHapley Additive exPlanations), a model that is used to provide an importance value for each feature of a utility prediction. SHAP utilizes additive feature attribution methods to unify six models for feature importance determination and outputs results that are similar to human intuition [138, 139, 140]. The SHAP analysis not only provides insights regarding feature importance in the DNN model but also identifies a range of the important features that affect the output utility value to a significant extent.

- Data generation through extensive Monte Carlo simulations for a given UE, RRH and MBS distribution, as well as the ASE-EE tradeoff factor for the entire optimization search space is a computationally intensive process. This makes it virtually impossible to train the DNN over an exhaustive range of UE, MBS and RRH densities, and β combinations. To overcome this issue, we employ different synthetic data generation techniques to augment the simulated data. The first method is known as generative adversarial networks, or GANs. A GAN consists of two neural networks, a generator and a discriminator [141]. The main aim of the generator ANN is to produce data instances that are similar to those in the real dataset. The closer the generated data distribution is to the real data distribution, the more likely it is to fool the discriminator as it attempts to identify the false data instances in the dataset. The generator and discriminator are trained in tandem, even though they have conflicting goals. The GAN is trained when it achieves a saddle point in the optimization of the two ANNs. Generative adversarial networks have been used in wireless networks, for instance to learn the channel model when channel state information is unavailable [142, 143, 144]. Other studies have focused on improving the system performance when synthetic data is augmented to an already available dataset and fed to a DNN to yield higher prediction accuracy [128, 145]. Apart from GANs, we also utilize the analysis from SHAP to simulate more data instances in regions that correspond with greater effects on the utility value. More specifically, if we know which features have the greatest effects on the utility value, and if we can identify range of values for which the utility value have the highest sensitivity, we can generate data specifically for that highly sensitive parameter sub-space and augment it to the already simulated data.

- The new data set that contains the originally simulated dataset and the GAN-generated synthetic data is fed to a new DNN, which is first trained and then optimized using GA based hyperparameter tuning. It is expected that the new model will outperform the optimized DNN model from the simulated data alone.

The performance of the modular DNN approach explained above is evaluated at different stages of the process. The first output is taken after training the basic DNN without any hyperparameter tuning. The utility values obtained are referred to as "DNN" output. The second output is taken from the DNN with hyperparameter tuning and is referred to as "DNN-HO". The final utility output is yielded from the DNN that is trained and optimized on the larger data set that contains the simulated and synthetic datasets. This output termed "DNN-DE" in our analysis. In the next section, we evaluate the performance of these three utility outputs in comparison to the global optima from the brute force exhaustive search, referred to as "BF".

6.6 Performance Analysis

The training data set fed to the DNN at different stages in the modular process explained in the previous section contains seven features and one output. From the seven features, three are optimization variables in the optimization problem expressed in Eq. (6.14). The parameters have ten distinct values given by p_{UE}, 2ζ, $p_{\mathrm{RRH,act}} \in 0.1, 0.2, ..., 1$. These variables are represented as "pUE", "2zeta" and "RRHactperc" in the training dataset respectively. Next we have the mean UE, RRH and MBS densities with values between 400 and 800 for UE and RRH densities and between 5 and 20 for MBS densities, i.e. $400 \leq |A|\lambda_{\mathrm{UE}}, |A|\lambda_{\mathrm{RRH}} \leq 800$;

pUE	2zeta	RRHactperc	Num_UE	Num_RRH	Num_MBS	beta	UTILITY
0.1	0.1	0.1	762	451	19	0.81472	0.42485
0.1	0.1	0.2	439	511	13	0.63236	0.16494
0.1	0.1	0.3	786	463	20	0.95751	0.77926
0.1	0.1	0.4	594	720	7	0.95717	0.41067
0.1	0.1	0.5	766	717	19	0.42176	0.09222
0.1	0.1	0.6	414	740	19	0.65574	0.23344
0.1	0.1	0.7	703	697	11	0.67874	0.20462
0.1	0.1	0.8	468	682	5	0.65548	0.22294
0.1	0.1	0.9	418	439	17	0.27692	0.10858
0.1	0.1	1	527	780	6	0.69483	0.28347
0.1	0.2	0.1	553	706	17	0.43874	0.12406
0.1	0.2	0.2	596	578	15	0.18687	0.18176
0.1	0.2	0.3	702	510	15	0.70936	0.33828
0.1	0.2	0.4	465	448	12	0.6551	0.32886
0.1	0.2	0.5	536	634	8	0.95974	0.82201
0.1	0.2	0.6	502	602	15	0.75127	0.46749
0.1	0.2	0.7	784	619	7	0.8909	0.72777
0.1	0.2	0.8	503	736	9	0.14929	0.38165
0.1	0.2	0.9	497	772	10	0.81428	0.62356
0.1	0.2	1	500	646	12	0.1966	0.31813

Fig. 6.7: Training dataset sample.

and $5 \leq |A|\lambda_{\text{MBS}} \leq 20$. These quantities are labelled as "Num_UE", "Num_RRH" and "Num_MBS" respectively. The ASE-EE tradeoff parameter $0 \leq \beta \leq 1$ is represented as "beta" in our dataset. Finally, the output measure which is the objective function value for Eq. (6.14) with the input optimization parameters and the network parameters is labelled as "UTILITY". Other system parameters for generating the simulation data are same as Table 5.1. A sample of the training dataset is given as Fig. 6.7.

6.6.1 DNN Training Results

Fig. 6.8 presents the training and validation losses for the basic DNN model trained on the training dataset. The DNN model has 6 hidden layers with 32 neurons per hidden layer and 64 batch size for each training round. The results show

that both the training and validation errors reduce with each epoch. The training stops when the validation loss stops improving for a few pre-defined number of epochs.

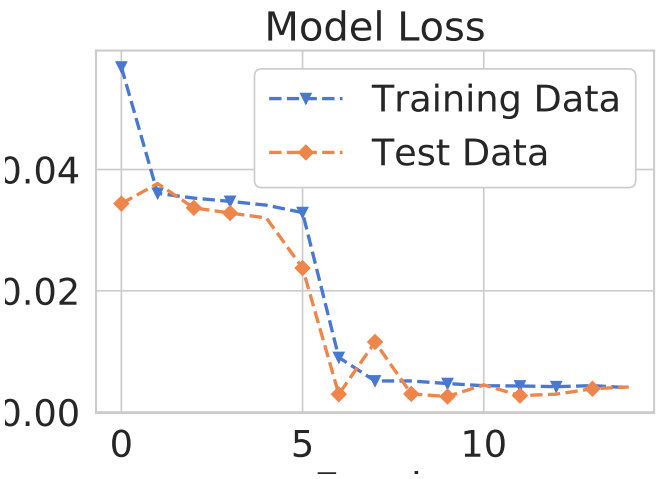


Fig. 6.8: Training and Test Error Loss.

6.6.2 SHAP Analysis

Fig. 6.9 plots the feature importance in the trained model. The SHAP value magnitudes are shown over all the training data samples. In addition to this, the color representation also shows the relationship between the feature and the output utility value. From the plot, we can see that the S-cell size and ASE-EE tradeoff measure, represented by "2zeta" and "beta" respectively, have te highest impact on the utility measure. Low values of ζ are shown to negatively impact the utility. The next most importance feature is β, whose lower values yield high utility values. This is because low values correspond to EE maximization and since EE values bits/J are higher than the ASE values in bits/sec/Hz/m^2, we get high utility as $\beta \to 0$. Lowe values of RRH activation percentage has an adverse impact on the utility. This means that when we have lower number of available mmWave RRHs, the utility will go down because the number the UEs connected

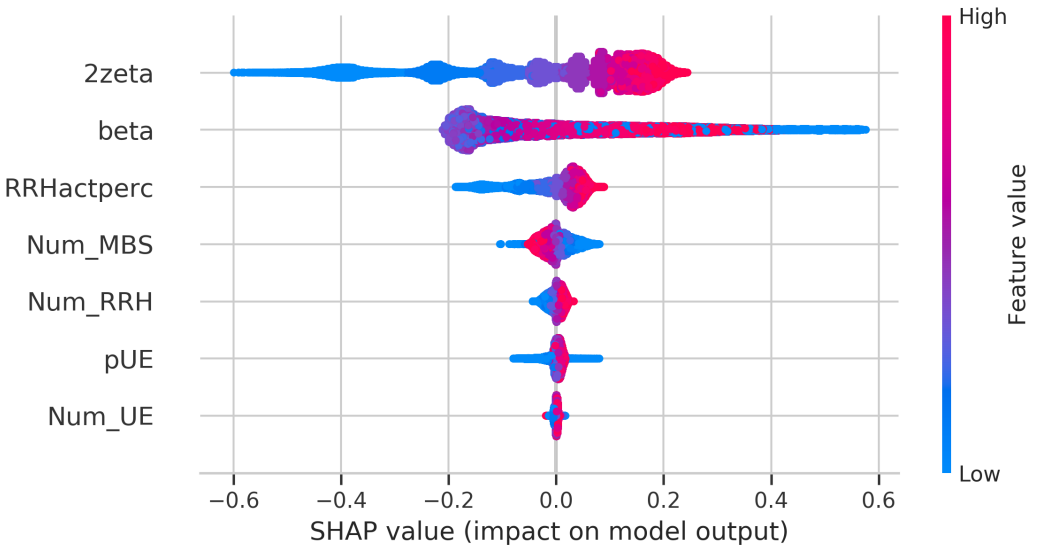


Fig. 6.9: Feature Importance in the trained model.

to the MBSs will increase.

The bar plot in Fig. 6.10 depicts the mean absolute value of the SHAP values for each feature in the model. Also known as the SHAP Summary plot (or the feature importance plot), it shows the mean importance of each feature in the variability of the model output. This plot is particularly useful for a system level control as it shows that what control knob (or optimization variable) needs to be played the most for tuning network configuration to get optimal performance. As in Fig. 6.9, the Stienen cell size and ASE-EE tradeoff factor have the highest average impact on the output utility. Since "2zeta" is an optimization parameter in the stochastic optimization problem, we will analyze the utility variations with β in the succeeding analysis.

SHAP Dependency plots, on the other hand, show the behavior of feature importance (or SHAP value) with respect to the value of its corresponding feature and its interaction with the most dependent feature. This plot is useful for observing the range of values for a pair of features which have the highest impact on the

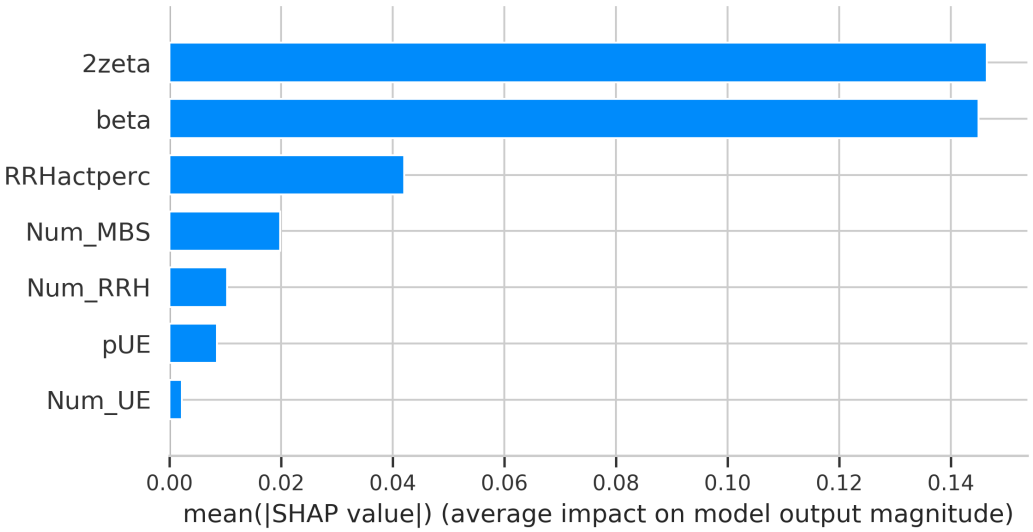


Fig. 6.10: Mean absolute SHAP values.

model output. For Example, Fig. 6.11 shows that the model β and 2ζ have he highest impact on the model output when $0 < \beta < 0.1$, $0.8 < 2\zeta < 0.1$ and $0.8 < \beta < 1$, $0.1 < 2\zeta < 0.3$. This information can be exploited, as usually the process of enriching training data is costly, so a selective enrichment of training data can provide a lower cost/benefit ratio as compared to a uniform or random enrichment of training data. It would therefore be interesting to see if data augmentation at high and low β ranges would improve model performance in terms of achieving global optimal utility values.

6.6.3 GAN Training and Performance

We evaluate the performance of GAN training by plotting the generator and discriminator training loss after few iterations. The GAN consists of a discriminator and a generator with the following settings:

- Generator has 40 input neurons (noise dimension), 4 hidden layers, and 8 output neurons (data dimensions). Batch normalization is also done between

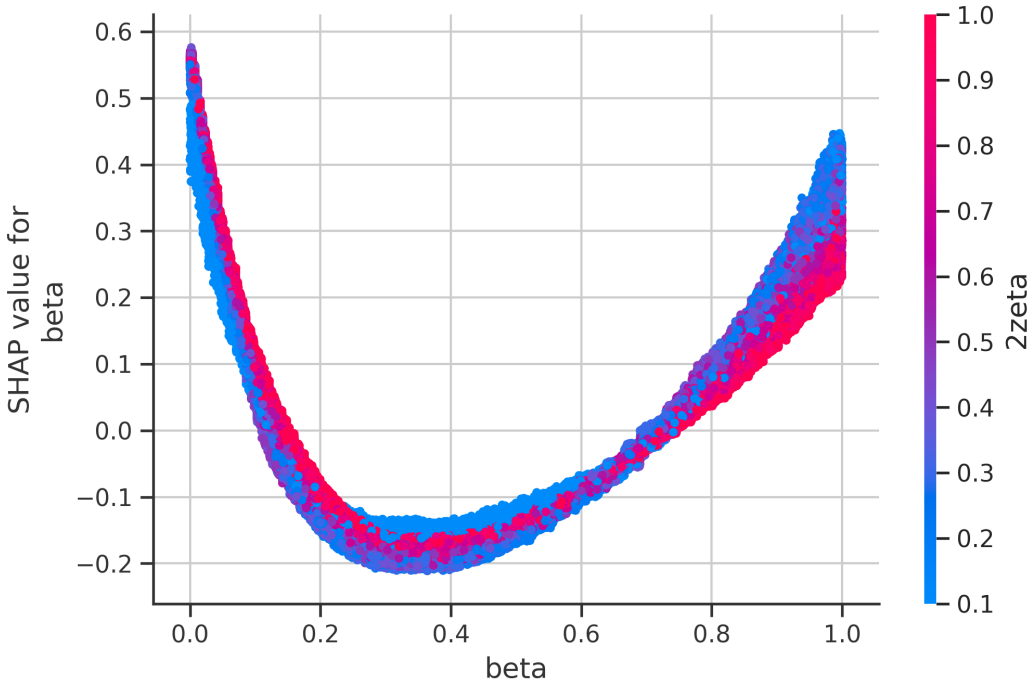


Fig. 6.11: SHAP Dependence plot for β.

each layer to make the model faster, more stable and regularized during training.

- Discriminator on the other hand has 8 input neurons (Data dimension), 4 hidden layers and only 1 output neuron with a sigmoid activation function, as it's purpose is to classify the input measurement as real or fake.

- Binary Cross Entropy loss function is used as a loss function in both of these networks, as the output variable is either 0 (fake) or 1 (real).

Additionally, data standardization is performed before training and a manually tuned optimizer (learning rate) to avoid oscillations during the training. We observe in Fig. 6.12 that the training loss for both the neural networks decreases gradually and attains an equilibrium state towards the end of training.

The plot in Fig. 6.12 shows that the GAN has trained well on the input data,

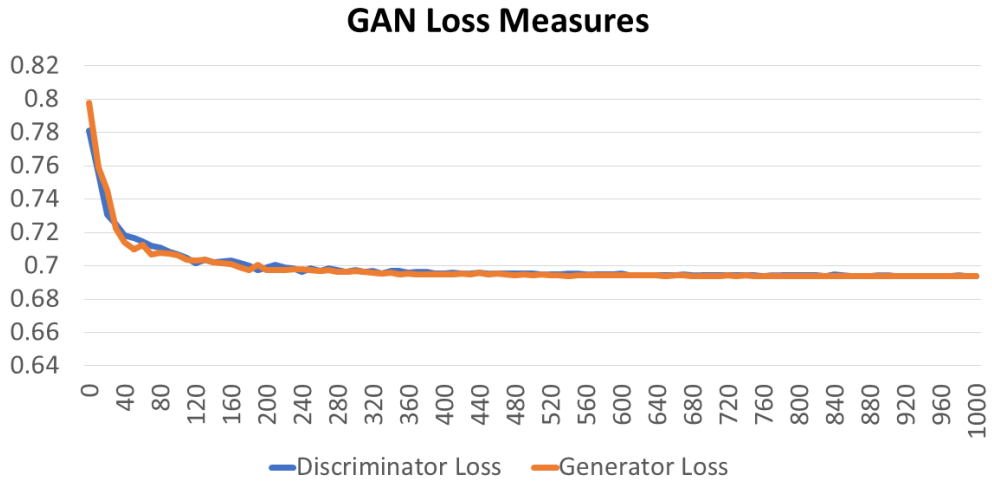


Fig. 6.12: Generator and Discriminator Training Loss.

which is the same training dataset to the baseline DNN model. Now to visualize the distribution of the original dataset and the GAN generated synthetic dataset, we plot the box plots of the output utility in the two datasets in Fig. 6.13. The box plots show that the synthetic data has a higher mean utility. As we will see in upcoming results, this will skew the performance of DNN when trained on the augmented dataset and increase the error measures.

Another variation of synthetic data generation that we use in our work is through creating more simulated data scenarios, but on targeted feature sub-space. More specifically, we generate data samples for the β sensitive ranges discussed in previous section. This is because we are only interested in augmenting data which has a larger significance to the overall model.

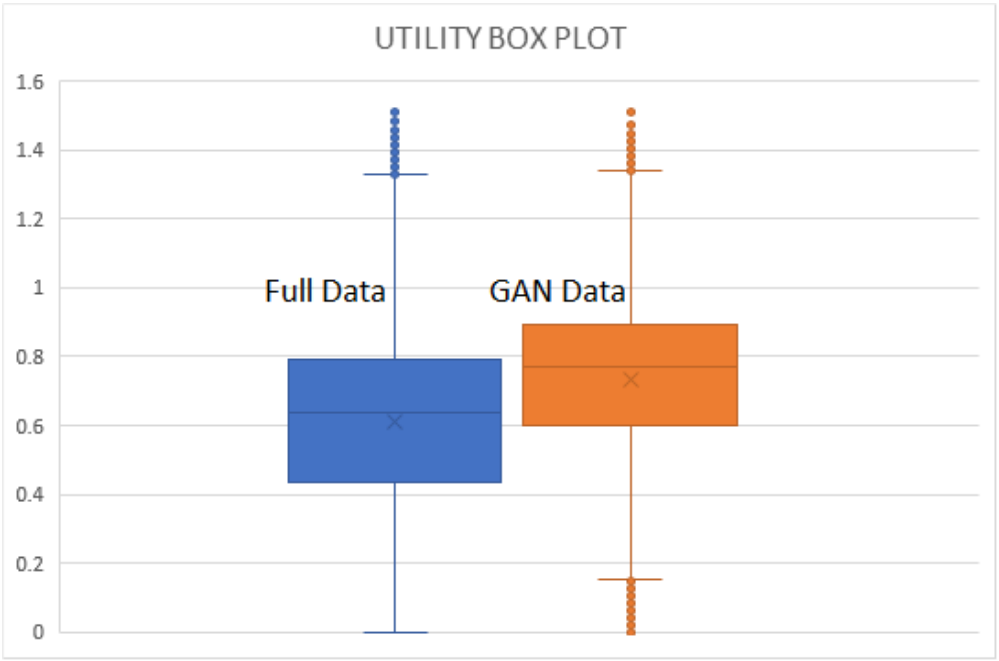


Fig. 6.13: Original vs Synthetic dataset Utility Boxplot.

Table 6.1: RMSE Test Data set Results

Methodology	DNN	DNN-HO	DNN-GAN	DNN-DE
Utility RMSE	0.173	0.031	0.164	0.039

6.6.4 Performance Comparison

In this section, we will analyze the performance of the DNN models in different stages of its sequential improvement process overlain if Fig. 6.6. We start by presenting the performance of the models when they are fed with new data samples, which is the test dataset. The root mean square error (RMSE) in the utility value prediction from the input features is performed for: i) baseline DNN model (DNN), ii) GA based hyperparameter optimized DNN (DNN-HO), iii) DNN trained and optimized on original data set augmented with synthetic data samples from GAN (DNN-GAN), and iv) DNN trained and optimized on original dataset augmented with new simulated data in the high importance β regions (DNN-DE). The hyperparameters that are tuned using Genetic Algorithm are:

- Number of neurons in each layer, from a pre-defined set (64, 128, 256, 512, 1024)
- Number of hidden layers on the model, varied from 1 to 9
- Training batch size, selected from 32,64,128,256
- Activation functions used in each layer, selected from ReLU, ELU, tanh and sigmoid.
- Optimizer from a predeifned set ('rmsprop', 'adam', 'sgd', 'adagrad', 'adadelta', 'adamax', 'nadam')

Genetic Algorithm is run for 4 generations and the optimal hyperparameter combination with the lowest RMSE is the following: 'nb_neurons': 64, 'nb_layers': 9, 'batch_size': 64, 'activation': 'relu', and 'optimizer': 'adam'.

The errors for DNN-HO and DNN-DE are considerably lower than baseline DNN and DNN trained on GAN augmented dataset. The reason for poor DNN-GAN performance is the mismatch in the utility distribution in the original training dataset and synthetic data generated via GAN. We will therefore not discuss the DNN-GAN performance in the succeeding sections.

We analyze the accuracy of output utility of the DNN approaches in comparison with the exhaustive search ('BF') method in Fig. 6.14. The BF represents the global optimal solution, therefore has the highest utility at any point in the cumulative distribution function (cdf) graph. The baseline DNN performance yields the worst output utility from its predicted optimal p_{UE}, 2ζ, and $p_{\mathrm{RRH,act}}$. The utility performance gap between BF and DNN is substantially reduced by the DNN-HO and DNN-DE models. The results not only show the effectiveness of learning models for optimal network parameter configuration, but also emphasizes on how

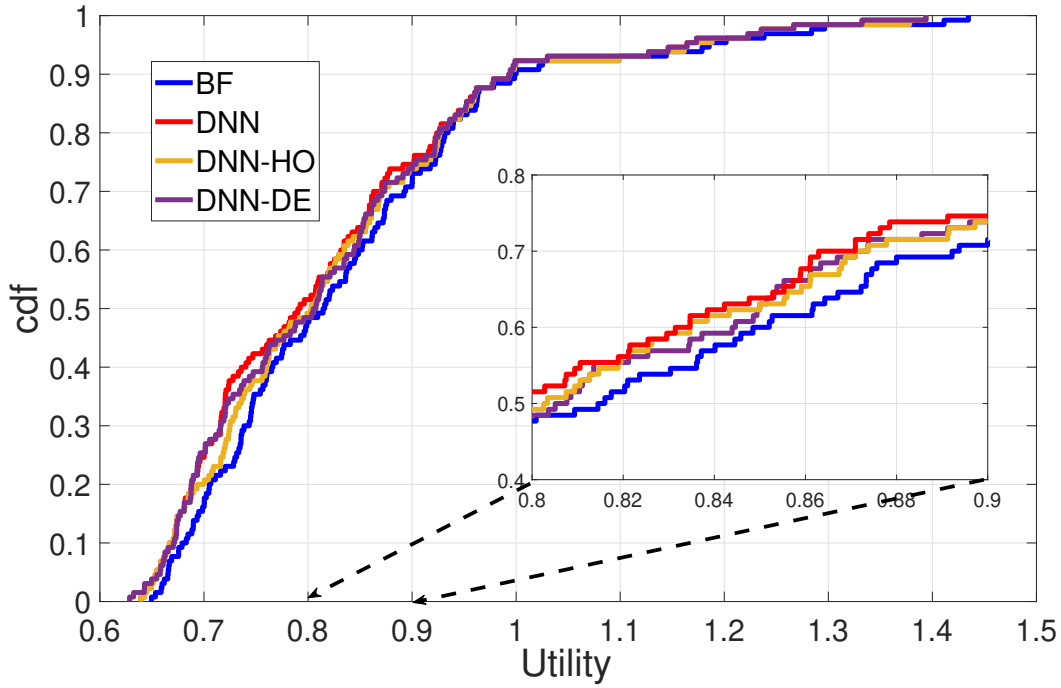


Fig. 6.14: Performance comparison of modular DNN in terms of Utility maximization.

efficient DNN model tuning can yield very close to global optimal performance. The negligible performance cost is overshadowed by the immense computational efficiency from the DNN approaches. The optimization is performed in real-time from the trained DNN model, and is thus suitable for implementation in real wireless networks.

We also compare the achieved utility with the DNN approaches with variations in β when the test dataset is given to the trained models. Each dataset point has the results from the three DNN model variations and the global optimal utility from exhaustive search space optimization. The utility comparison in Fig. 6.15 shows that utility achieved from ASE maximization is lower as compared to EE maximization. Secondly, the utility from 'BF' is always the highest for all test dataset samples. Another interesting observation is that the performance gradient between baseline 'DNN' and 'BF' is larger for high β regime. This implies that for ASE maximization, the optimization parameters from the DNN model would

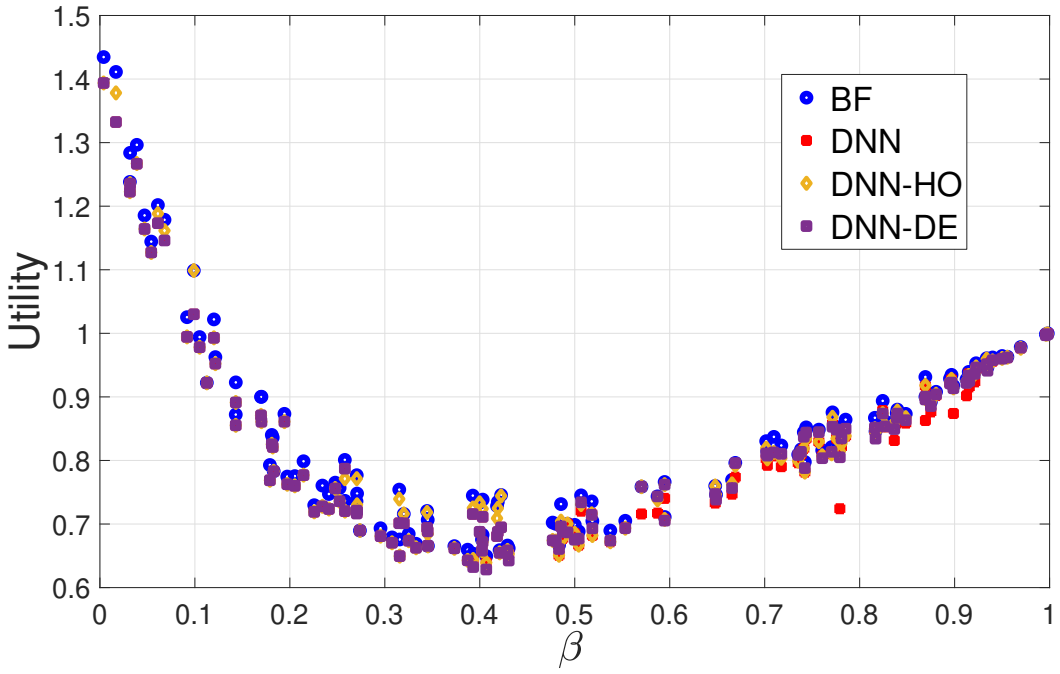


Fig. 6.15: Output Utility from the DNN approaches v/s global optima.

be more contrasting to the true optimization parameters.

6.7 Conclusion

In this chapter, we discussed a learning0based approach to optimize the tradeoff utility between the area spectral and energy efficiencies in a two-tier user-centric Stienen cell architecture with real-time computational efficiency. By integrating the dynamic placement of UEs, mmWave RRHs and sub-6 GHz MBSs, we developed a stochastic optimization problem to determine the Pareto optimal operating point between the conflicting ASE and EE efficiency measures. We demonstrated that the formulated optimization objective function is highly non-convex due to the random positioning of MBSs and the piece-wise nature of SINR at the UEs. The discontinuity for SINR occurs at two levels: one for the abrupt change as a UE connects to a sub-6 GHz MBS from a mmWave RRH and vice versa, and

secondly, due to change in pathloss exponent as a UE transitions from a LOS region to a NLOS region and vice versa. To find the global optimal solution to the problem, we performed hundreds of Monte Carlo simulations for a single network configuration to determine the utility value of the objective function. This process was repeated for the entire search space, causing the problem to have NP-hard time complexity.

In our proposed approach, we simulated a large number of network configurations, and UE, MBS and RRH distribution scenarios, and used the simulated data to train a DNN. The DNN was further optimized by adjusting its depth and design features to accurately learn the non-linear relationship between the set of design parameters and network population statistics with the optimization function utility. Shapley analysis was performed to identify the features that are most important in determining the output utility, and the regions that are the most critical regions for those influential parameters. We also discussed two synthetic data augmentation techniques to improve the DNN's learning efficiency. The first of these approaches trains a generative adversarial network with simulated dataset in order to generate more data samples with near-identical distribution. The second approach is based on leveraging insights from the SHAP analysis to generate simulation data only within intervals that have a high SHAP value for most influential input features. Our results demonstrate that the DNN tuned with a genetic algorithm estimates utility values for a given set of parameter configuration and network statistics with a very low RMSE. As a result, the performance loss that results from the learning-based approach is substantially reduced and near-optimal ASE-EE tradeoff is performed with real-time computational complexity.

CHAPTER 7

Conclusions and Future Work

7.1 Conclusions

With the exponentially increasing data demands and ever more complex mobile and IoT requirements, ultra-dense deployment of small cells becomes inevitable. However, while ultra-dense deployment manages to somewhat alleviate the data demands, it still does not ensure ubiquitous service level to all users and devices. In this dissertation, we investigate a relatively recent concept in radio access network design. Known as user-centric or cell less architecture, it deviates from current cell-centric design in the sense that the devices dictate UE-BS associations, cell activations and handover preferences. In other words, the network follows the devices, and no device faces the cell edge issues which are a major concern for current deployments. Our work has focused on novel user-centric RAN architectures, user association mechanisms therein, and leveraging tools from game theory and deep learning to optimize the intertwined network wide efficiency metrics while ensuring adequate user QoE.

The first contribution of this dissertation focuses on a game theoretic based user-centric architecture in a cloud radio access network (C-RAN) deployment. The key idea here is to deploy ultra-dense network consisting of radio remote heads. A network orchestration algorithm is then designed to dynamically orchestrate the network and select the best RRH within a pre-defined radius (virtual user-centric instead of base station-centric cell) around the high priority user equipments (UEs) selected for downlink transmission during each TTI. This user-centric architecture

allows switching on/off of RRHs that guarantees a higher energy efficiency along with location-independent uniform Quality of Experience (QoE). The intelligent deactivation of BSs not serving any UE in the user-centric network results in two-fold benefits: first, the radius of the circular exclusion zones around every scheduled UE serves as a proxy for the minimum spatial separation for the closest interfering BS. This combined with the thinning of the BSs due to user-centric scheduling results in reduced interference at an arbitrary UE. Secondly, a key goal of 5G, i.e. energy efficiency is enhanced as a result of reduction in power consumption through deactivation of BSs. Our goal in this work is to develop analytical characterization of the area spectral and energy efficiency for such a user-centric network architecture supported via Cloud RAN. We observe that the cell density that yields optimal EE is different than that which yields maximum area spectral efficiency (ASE). The size of the exclusion zone is then used as a control parameter to realize the desired compromise between EE and ASE. The tradeoff analysis is performed using a two-player bargaining game with the ASE and EE negotiating for the Pareto optimal tradeoff (Nash equilibrium) dictated by a network operator weightage parameter specifying the business model within a spatio-temporal region.

In consideration of the future IoT paradigm with billions of connected devices with large disparity in data demands, the next contribution of the dissertation introduces a second tier of elasticity within user-centric systems that integrates non-uniform exclusion zones centered on UEs. These non-uniform service zones cater for data demand disparity between spatio-temporal zones as well as the diversity of data requirements from user applications (for instance HD video streaming v/s whatsapp messaging) within a single spatio-temporal zone. The allocation of these virtually elastic service zones around selected UEs is conducted via a central

control base station (CBS) and modeled through two game techniques, namely evolutionary and auction games. Both the games are based on a utility minimization problem which is a function of weighted mean UE throughput and individual UE service demands. To illustrate the tradeoffs between the game models, network level performance is compared in terms of aggregate throughput, energy efficiency, algorithm convergence speed and mean UE scheduling probabilities.

To further improve upon the performance of user-centric designs., we propose and analyze a novel Stienen cell user-centric architecture operating in the mmWave spectrum, which is expected to be a major frequency usage region in future wireless networks. Taking into account the idiosyncrasies of mmWave blocking and propagation characteristics, a statistical framework is developed for deriving the coverage probability of an arbitrary user equipment (UE) scheduled within the proposed architecture. Numerical results show that, by virtue of selective mmWave RRH activation and consequently thinning of the interfering RRHs, the proposed architecture offers higher user Quality of Experience (QoE) compared to legacy base station centric architectures. Furthermore, as compared to the user-centric model designed in first dissertation contribution, we observe an increase in area spectral efficiency (ASE) and reduction in scheduling wait times at negligible cost in energy efficiency (EE). Yet another key insight from this work is the existence of an exploitable tradeoff between ASE and EE that can be exactly dictated by three key design parameters namely: 1) the active UE population, 2) Stienen cell size factor and 3) mmWave RRH deployment density. The tradeoff can be leveraged in practice through a self-organizing network (SON) engine that can orchestrate these parameters dynamically to achieve a Pareto optimal performance.

The final contribution of this dissertation is a continuation of the Stienen cell model, where first the optimization of the utility in terms of the three design

parameters is shown to be highly non-convex. Then, a stochastic optimization problem is defined to tradeoff between the ASE and EE in the proposed system model. Thereafter, a deep neural network (DNN) based approach is utilized as a substitute for brute force and conventional meta heuristic techniques. The key idea is that if the non-linear mapping between system utility and a combination of network parameters and design parameters can be learned accurately by a DNN of moderate size, then identification of the optimal design parameters can be done in near real-time. It is shown that at a negligible cost in system utility, the DNN based approach can indeed be applied on this stochastic optimization problem to find optimal system design parameters. The baseline DNN performance is improved by hyperparameter tuning and data enrichment using generative adversarial networks (GANs).

To summarize, this dissertation introduces many novel user-centric designs, the corresponding user-centric user scheduling schemes and also lays down the statistical framework necessary to quantify the network-wide efficiency parameters. The work will hopefully pave the way for deploying efficient networks, that follow the user, in 5G and beyond cellular networks.

7.2 Future Works

The work undertaken in this dissertation could be enhanced on multiple fronts. The most obvious margin of improvement is integration of the latency factor in the optimization framework. Since 5G networks target for latency of the order of milliseconds, the latency induced due to non-overlap user-centric service regions would be detrimental in achieving this goal. It would be interesting to analyze if user-clustering strategies and flexible frame structure introduced in 5G New

Radio standards can provide the 1ms end-to-end latency. Additionally, instead of the proposed time division multiplex (TDM) scheduling scheme, we can also study how frequency division multiplexing (FDM) will impact the latency as well as the capacity and energy efficiency measures.

Another feature that can add value to this dissertation is inclusion of the mobility factor to allow proactivity in service zone assignment to high priority users. It is well known that typical human mobility features 93% average predictability [146]. This can be leveraged to ascertain a particular user's future location co-ordinates and activate the RRH that is anticipated to reside within its future user-centric region for DL scheduling if required.

With respect to C-RAN design, it is pertinent that the backhaul latency i slow to allow LTE-A features such as carrier aggregation and coordinated multipoint transmission. Since ultra-dense deployment means a high number of RRHs will be simultaneously transmitting to BBU pools, the data volume along with the real-time optimization of user-centric regions will put a toll on the backhaul links connected the RAN to the core network. To meet the high demands on backhaul, mobile network operators usually consider dark fiber as the viable solution. However, this significantly increases the CAPEX as large-scale dedicated fiber deployment is an expensive and time consuming process. Future extensions of our optimization framework can include CAPEX and OPEX in the objective function with the end goal being maximization of the revenue per user, also known as ARPU.

Finally, we have not studied the performance enhancement possible from leveraging databases at RRH and MBSs. The databases may include standard measurements such as reference signal received power (RSRP), reference signal received quality (RSRQ), signal-to-noise ratio (SNR), signal-to-interference-plus-noise-ratio

(SINR), channel quality indicator (CQI), physical resource block (PRB) usage, mobility traces and radio link failure (RLF) reports. An artificial intelligence engine trained on the database can be used to predict location of de-activated RRHs and also proactively define user-centric regions and consequently activate/deactivate RRHs and optimally allocate spectrum and energy resources to jointly optimize ASE and EE.

Bibliography

[1] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5G: how to empower SON with big data for enabling 5G," *IEEE Network*, vol. 28, no. 6, pp. 27–33, Nov 2014.

[2] S. Chen, F. Qin, B. Hu, X. Li, and Z. Chen, "User-centric ultra-dense networks for 5G: challenges, methodologies, and directions," *IEEE Wireless Communications*, vol. 23, no. 2, pp. 78–85, April 2016.

[3] U. Hashmi, S. A. R. Zaidi, A. Darbandi, and A. Imran, "On the efficiency tradeoffs in User-Centric cloud RAN," in *IEEE ICC 2018 Next Generation Networking and Internet Symposium (ICC'18 NGNI)*, Kansas City, USA, May 2018.

[4] U. S. Hashmi, S. A. R. Zaidi, and A. Imran, "User-Centric Cloud RAN: An Analytical framework for Optimizing Area Spectral and Energy Efficiency," *IEEE Access*, pp. 1–1, 2018.

[5] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for wireless resource management," in *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, July 2017, pp. 1–6.

[6] CISCO, "Cisco visual networking index: Global mobile data traffic forecast, 2016-2021 - whitepaper," *Tech. Rep.*, 2017.

[7] J. Research, "Press release: Mobile network operator revenues," June 2011. [Online]. Available: http://juniperresearch.com/viewpressrelease.php?pr=245

[8] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A Tractable Approach to Coverage and Rate in Cellular Networks," *IEEE Transactions on Communications*, vol. 59, no. 11, pp. 3122–3134, November 2011.

[9] A. AlAmmouri, J. G. Andrews, and F. Baccelli, "Analysis of dense cellular networks with stretched exponential path loss," in *2017 51st Asilomar*

Conference on Signals, Systems, and Computers, Oct 2017, pp. 1837–1847.

[10] M. Gruber, O. Blume, D. Ferling, D. Zeller, M. A. Imran, and E. C. Strinati, "EARTH - Energy Aware Radio and Network Technologies," in *2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sep. 2009, pp. 1–5.

[11] G. Auer, O. Blume, V. Giannini, I. Godor, M. A. Imran, Y. Jading, E. Katranaras, M. Olsson, D. Sabella, P. Skillermark, and W. Wajda, "D2.3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown," pp. 1–68, 12 2010.

[12] O. K. Rayel, G. Brante, J. L. Rebelatto, R. D. Souza, and M. A. Imran, "Energy Efficiency-Spectral Efficiency Trade-Off of Transmit Antenna Selection," *IEEE Transactions on Communications*, vol. 62, no. 12, pp. 4293–4303, Dec 2014.

[13] D. Kaltakis, M. A. Imran, and C. Tzaras, "Information theoretic capacity of cellular multiple access channel with shadow fading," *IEEE Transactions on Communications*, vol. 58, no. 5, pp. 1468–1476, May 2010.

[14] E. Katranaras, M. A. Imran, and R. Hoshyar, "Sum Rate of Linear Cellular Systems with Clustered Joint Processing," in *VTC Spring 2009 - IEEE 69th Vehicular Technology Conference*, April 2009, pp. 1–5.

[15] E. Katranaras, M. A. Imran, and C. Tzaras, "Uplink capacity of a variable density cellular system with multicell processing," *IEEE Transactions on Communications*, vol. 57, no. 7, pp. 2098–2108, July 2009.

[16] G. Auer, V. Giannini, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, C. Desset, and O. Blume, "Cellular energy efficiency evaluation framework," in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, May 2011, pp. 1–6.

[17] C. Desset, B. Debaillie, V. Giannini, A. Fehske, G. Auer, H. Holtkamp, W. Wajda, D. Sabella, F. Richter, M. J. Gonzalez, H. Klessig, I. GÃş-dor, M. Olsson, M. A. Imran, A. Ambrosy, and O. Blume, "Flexible power

modeling of lte base stations," in *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2012, pp. 2858–2862.

[18] B. Romanous, N. Bitar, S. A. R. Zaidi, A. Imran, M. Ghogho, and H. H. Refai, "A Game Theoretic Approach for Optimizing Density of Remote Radio Heads in User Centric Cloud-Based Radio Access Network," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.

[19] F. Khan and Z. Pi, "mmWave mobile broadband (MMB): Unleashing the 3-300GHz spectrum," in *34th IEEE Sarnoff Symposium*, May 2011, pp. 1–6.

[20] Q. Mao, F. Hu, and Q. Hao, "Deep Learning for Intelligent Wireless Networks: A Comprehensive Survey," *IEEE Communications Surveys Tutorials*, vol. 20, no. 4, pp. 2595–2621, Fourthquarter 2018.

[21] S. Buzzi, C. D'Andrea, and A. Zappone, "User-Centric 5G Cellular Networks: Resource Allocation and Comparison with the Cell-Free Massive MIMO Approach," *CoRR*, vol. abs/1803.02261, 2018.

[22] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-Free Massive MIMO Versus Small Cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, March 2017.

[23] S. Buzzi and C. DâĂŹAndrea, "Cell-Free Massive MIMO: User-Centric Approach," *IEEE Wireless Communications Letters*, vol. 6, no. 6, pp. 706–709, Dec 2017.

[24] S. Buzzi and C. D'Andrea, "User-Centric Communications versus Cell-free Massive MIMO for 5G Cellular Networks," in *WSA 2017; 21th International ITG Workshop on Smart Antennas*, March 2017, pp. 1–6.

[25] S. Buzzi and A. Zappone, "Downlink power control in user-centric and cell-free massive mimo wireless networks," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Oct 2017, pp. 1–6.

[26] J. Shi, C. Pan, W. Zhang, and M. Chen, "Performance Analysis for User-Centric Dense Networks With mmWave," *IEEE Access*, vol. 7, pp. 14 537–

14 548, 2019.

[27] V. Garcia, Y. Zhou, and J. Shi, "Coordinated Multipoint Transmission in Dense Cellular Networks With User-Centric Adaptive Clustering," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4297–4308, Aug 2014.

[28] A. M. Ghaleb, A. M. Mansoor, and R. B. Ahmad, "An Energy-Efficient User-Centric Approach High-Capacity 5G Heterogeneous Cellular Networks," 2018.

[29] Yingxiao Zhang and Ying Jun Zhang, "User-centric virtual cell design for Cloud Radio Access Networks," in *2014 IEEE 15th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2014, pp. 249–253.

[30] H. Zhang, Z. Yang, Y. Liu, and X. Zhang, "Power Control for 5G User-Centric Network: Performance Analysis and Design Insight," *IEEE Access*, vol. 4, pp. 7347–7355, 2016.

[31] S. Bassoy, M. Jaber, M. A. Imran, and P. Xiao, "Load Aware Self-Organising User-Centric Dynamic CoMP Clustering for 5G Networks," *IEEE Access*, vol. 4, pp. 2895–2906, 2016.

[32] J. Nash, "Two-Person Cooperative Games," *Econometrica*, vol. 21, no. 1, pp. 128–140, 1953. [Online]. Available: http://www.jstor.org/stable/1906951

[33] M. J. Osborne and A. Rubinstein, *A course in game theory.* Cambridge, USA: The MIT Press, 1994, electronic edition.

[34] K. Binmore, *Fun and Games: A Text on Game Theory.* D.C. Heath, 1992. [Online]. Available: https://books.google.com/books?id=tPOZQgAACAAJ

[35] T. Heikkinen, "A potential game approach to distributed power control and scheduling," *Computer Networks*, vol. 50, no. 13, pp. 2295 – 2311, 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128605002859

[36] F. Meshkati, Mung Chiang, H. V. Poor, and S. C. Schwartz, "A game-theoretic approach to energy-efficient power control in multicarrier CDMA systems," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 6, pp. 1115–1129, June 2006.

[37] G. Scutari, D. P. Palomar, and S. Barbarossa, "Asynchronous Iterative Water-Filling for Gaussian Frequency-Selective Interference Channels: A Unified Framework," in *2006 IEEE 7th Workshop on Signal Processing Advances in Wireless Communications*, July 2006, pp. 1–5.

[38] T. Heikkinen, "Distributed Scheduling via Pricing in a Communication Network," in *Proceedings of the Second International IFIP-TC6 Networking Conference on Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; and Mobile and Wireless Communications*, ser. NETWORKING '02. London, UK, UK: Springer-Verlag, 2002, pp. 850–862. [Online]. Available: http://dl.acm.org/citation.cfm?id=647077.714482

[39] E. Altman, T. Basar, and R. Srikant, "Nash equilibria for combined flow control and routing in networks: asymptotic behavior for a large number of users," *IEEE Transactions on Automatic Control*, vol. 47, no. 6, pp. 917–930, June 2002.

[40] D. Grosu and A. T. Chronopoulos, "Algorithmic mechanism design for load balancing in distributed systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 77–84, Feb 2004.

[41] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb 2005.

[42]

[43] N. Nie and C. Comaniciu, "Adaptive Channel Allocation Spectrum Etiquette for Cognitive Radio Networks," *Mob. Netw. Appl.*, vol. 11, no. 6, pp. 779–797, Dec. 2006. [Online]. Available: http://dx.doi.org/10.1007/s11036-006-0049-y

[44] M. Felegyhazi, M. Cagalj, S. S. Bidokhti, and J. . Hubaux, "Non-Cooperative Multi-Radio Channel Allocation in Wireless Networks," in *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, May 2007, pp. 1442–1450.

[45] X. Li, Q. Huang, and D. Wu, "A repeated stochastic game approach for strategic network selection in heterogeneous networks," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2018, pp. 88–93.

[46] Z. Han, D. Niyato, W. Saad, T. Baar, and A. Hjrungnes, *Game Theory in Wireless and Communication Networks: Theory, Models, and Applications*. Cambridge University Press, 2012.

[47] T. Basar and G. Olsder, *Dynamic Noncooperative Game Theory: Second Edition*, ser. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 1999. [Online]. Available: https://books.google.com/books?id=nry8U3CfF-gC

[48] G. He, M. Debbah, and S. Lasaulce, "K-player Bayesian waterfilling game for fading multiple access channels," in *2009 3rd IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec 2009, pp. 17–20.

[49] R. A. Fisher and J. H. Bennett, *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press, 1999. [Online]. Available: http://books.google.com/books?id=sT4lIDk5no4C

[50] L. Anderegg and S. Eidenbenz, "Ad hoc-VCG: A Truthful and Cost-efficient Routing Protocol for Mobile Ad Hoc Networks with Selfish Agents," in *Proceedings of the 9th Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '03. New York, NY, USA: ACM, 2003, pp. 245–259. [Online]. Available: http://doi.acm.org/10.1145/938985.939011

[51] Y. Shih, A. Pang, M. Tsai, and C. Chai, "A rewarding framework for network resource sharing in co-channel hybrid access femtocell networks," *IEEE Transactions on Computers*, vol. 64, no. 11, pp. 3079–3090, Nov. 2015. [Online]. Available: doi.ieeecomputersociety.org/10.1109/TC.2015.2394453

[52] P. Semasinghe, E. Hossain, and K. Zhu, "An Evolutionary Game for Distributed Resource Allocation in Self-Organizing Small Cells," *IEEE Transactions on Mobile Computing*, vol. 14, no. 2, pp. 274–287, Feb 2015.

[53] X. Kang, R. Zhang, and M. Motani, "Price-based resource allocation for spectrum-sharing femtocell networks: A stackelberg game approach," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 538–549, April 2012.

[54] F. Pantisano, M. Bennis, W. Saad, M. Debbah, and M. Latva-aho, "Interference alignment for cooperative femtocell networks: A game-theoretic approach," *IEEE Transactions on Mobile Computing*, vol. 12, no. 11, pp. 2233–2246, Nov 2013.

[55] J. Zheng, Y. Cai, and A. Anpalagan, "A Stochastic Game-Theoretic Approach for Interference Mitigation in Small Cell Networks," *IEEE Communications Letters*, vol. 19, no. 2, pp. 251–254, Feb 2015.

[56] U. S. Hashmi, A. Islam, K. M. Nasr, and A. Imran, "Towards User QoE-Centric Elastic Cellular Networks: A Game Theoretic Framework for Optimizing Throughput and Energy Efficiency," in *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2018, pp. 1–7.

[57] R. Wang, H. Hu, and X. Yang, "Potentials and Challenges of C-RAN Supporting Multi-RATs Toward 5G Mobile Networks," *IEEE Access*, vol. 2, pp. 1187–1195, 2014.

[58] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for Mobile Networks: A Technology Overview," *IEEE Communications Surveys Tutorials*, no. 1, pp. 405–426, Firstquarter 2014.

[59] Y. Zhang and Y. J. Zhang, "User-centric virtual cell design for Cloud Radio Access Networks," in *Signal Processing Advances in Wireless Communications (SPAWC), 2014 IEEE 15th International Workshop on*, June 2014, pp. 249–253.

[60] D. Huang, T. Xing, and H. Wu, "Mobile cloud computing service models: a user-centric approach," *Network, IEEE*, vol. 27, no. 5, pp. 6–11, September 2013.

[61] S. Zaidi, S. Affes, U. Vilaipornsawai, L. Zhang, and P. Zhu, "Wireless Access Virtualization Strategies for Future User-Centric 5G Networks," in *2016 IEEE Globecom Workshops (GC Wkshps)*, Dec 2016, pp. 1–7.

[62] S. Zaidi, A. Imran, D. McLernon, and M. Ghogho, "Characterizing Coverage and Downlink Throughput of Cloud Empowered HetNets," *Communications Letters, IEEE*, vol. PP, no. 99, pp. 1–1, 2015.

[63] B. Dai and W. Yu, "Sparse Beamforming and User-Centric Clustering for Downlink Cloud Radio Access Network," *IEEE Access*, vol. 2, pp. 1326–1339, 2014.

[64] C. Li, J. Zhang, M. Haenggi, and K. B. Letaief, "User-Centric Intercell Interference Nulling for Downlink Small Cell Networks," *IEEE Transactions on Communications*, vol. 63, no. 4, pp. 1419–1431, April 2015.

[65] A. Awad, A. Mohamed, and C. F. Chiasserini, "Dynamic Network Selection in Heterogeneous Wireless Networks: A user-centric scheme for improved delivery," *IEEE Consumer Electronics Magazine*, vol. 6, no. 1, pp. 53–60, Jan 2017.

[66] C. Mobile, "C-RAN: the road towards green RANs," *White Paper, ver, vol. 2*, 2011.

[67] J. M. D. Stoyan, W. S. Kendall and L. Ruschendorf, *Stochastic geometry and its applications*. Wiley Chichester, 1995, vol. 2.

[68] A. Imran, M. A. Imran, A. Abu-Dayya, and R. Tafazolli, "Self Organization of Tilts in Relay Enhanced Networks: A Distributed Solution," *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 764–779, February 2014.

[69] H. He, J. Xue, T. Ratnarajah, F. A. Khan, and C. B. Papadias, "Modeling and Analysis of Cloud Radio Access Networks Using Matern Hard-Core

Point Processes," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 4074–4087, June 2016.

[70] M. Haenggi, "Mean Interference in Hard-Core Wireless Networks," *Communications Letters, IEEE*, vol. 15, no. 8, pp. 792–794, August 2011.

[71] Z. Chen, C. X. Wang, X. Hong, J. S. Thompson, S. A. Vorobyov, X. Ge, H. Xiao, and F. Zhao, "Aggregate Interference Modeling in Cognitive Radio Networks with Power and Contention Control," *IEEE Transactions on Communications*, vol. 60, no. 2, pp. 456–468, February 2012.

[72] A. Taufique, M. Jaber, A. Imran, Z. Dawy, and E. Yacoub, "Planning Wireless Cellular Networks of Future: Outlook, Challenges and Opportunities," *IEEE Access*, vol. 5, pp. 4821–4845, 2017.

[73] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *Wireless Communications, IEEE*, vol. 18, no. 5, pp. 40–49, October 2011.

[74] R. Gupta, E. Calvanese Strinati, and D. Ktenas, "Energy efficient joint DTX and MIMO in cloud Radio Access Networks," in *Cloud Networking (CLOUDNET), 2012 IEEE 1st International Conference on*, Nov 2012, pp. 191–196.

[75] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband Millimeter-Wave Propagation Measurements and Channel Models for Future Wireless Communication System Design," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3029–3056, Sept 2015.

[76] A. Mohamed, O. Onireti, M. A. Imran, A. Imran, and R. Tafazolli, "Control-Data Separation Architecture for Cellular Radio Access Networks: A Survey and Outlook," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 446–465, Firstquarter 2016.

[77] J. Zheng, Y. Cai, and A. Anpalagan, "A Stochastic Game-Theoretic Approach for Interference Mitigation in Small Cell Networks," *IEEE Communications Letters*, vol. 19, no. 2, pp. 251–254, Feb 2015.

[78] J. Cao, T. Peng, Z. Qi, R. Duan, Y. Yuan, and W. Wang, "Interference Management in Ultra-Dense Networks: A User-Centric Coalition Formation Game Approach," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2018.

[79] H. Zhang, Y. Chen, and Y. Liu, "Spatial correlation based analysis of power control in user-centric 5G networks," *IET Communications*, vol. 12, no. 3, pp. 326–333, 2018.

[80] W. Vickrey, "Counterspeculation, Auctions, and Competitive Sealed Tenders," *The Journal of Finance*, vol. 16, no. 1, pp. 8–37, 1961.

[81] M. A. Alavijeh, B. Maham, Z. Han, and W. Saad, "Truthful spectrum auction for efficient anti-jamming in cognitive radio networks," in *2017 IEEE Symposium on Computers and Communications (ISCC)*, July 2017, pp. 742–747.

[82] [Online]. Available: https://www.ict-earth.eu/

[83] R. Gupta, E. C. Strinati, and D. Ktenas, "Energy efficient joint DTX and MIMO in cloud Radio Access Networks," in *2012 IEEE 1st International Conference on Cloud Networking*, Nov 2012, pp. 191–196.

[84] O. G. Aliu, A. Imran, M. A. Imran, and B. Evans, "A survey of self organisation in future cellular networks," *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 336–361, 2013.

[85] J. G. Andrews, X. Zhang, G. D. Durgin, and A. K. Gupta, "Are we approaching the fundamental limits of wireless network densification?" *IEEE Communications Magazine*, vol. 54, no. 10, pp. 184–190, October 2016.

[86] J. Stienen, *Die Vergroeberung von Karbiden in reinen Eisen-Kohlenstoff-Staehlen.* na, 1982.

[87] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter Wave Channel Modeling and Cellular Capacity Evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, June 2014.

[88] M. D. Renzo, "Stochastic Geometry Modeling and Analysis of Multi-Tier Millimeter Wave Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 9, pp. 5038–5057, Sept 2015.

[89] T. Bai and R. W. Heath, "Coverage and Rate Analysis for Millimeter-Wave Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 2, pp. 1100–1114, Feb 2015.

[90] J. G. Andrews, T. Bai, M. N. Kulkarni, A. Alkhateeb, A. K. Gupta, and R. W. Heath, "Modeling and analyzing millimeter wave cellular systems," *IEEE Transactions on Communications*, vol. 65, no. 1, pp. 403–430, Jan 2017.

[91] O. Onireti, A. Imran, and M. A. Imran, "Coverage, Capacity and Energy Efficiency Analysis in the Uplink of mmWave Cellular Networks," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2017.

[92] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband Millimeter-Wave Propagation Measurements and Channel Models for Future Wireless Communication System Design," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3029–3056, Sept 2015.

[93] R. Hernandez-Aquino, S. A. R. Zaidi, M. Ghogho, D. McLernon, and A. Swami, "Stochastic geometric modeling and analysis of non-uniform two-tier networks: A stienen's model-based approach," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3476–3491, June 2017.

[94] R. Jain, D.-M. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," *CoRR*, vol. cs.NI/9809099, 1998.

[95] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, 1st ed. New York, NY, USA: John Wiley & Sons, Inc., 1994.

[96] J. R. Swisher, P. D. Hyden, S. H. Jacobson, and L. W. Schruben, "A survey of simulation optimization techniques and procedures," in *2000 Winter Simulation Conference Proceedings (Cat. No.00CH37165)*, vol. 1, Dec 2000, pp. 119–128 vol.1.

[97] K. Yao and J. Gao, "Law of Large Numbers for Uncertain Random Variables," *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 3, pp. 615–621, June 2016.

[98] D. Goldsman and B. L. Nelson, "Ranking, selection and multiple comparisons in computer simulation," in *Proceedings of Winter Simulation Conference*, Dec 1994, pp. 192–199.

[99] J. Hsu, *Multiple Comparisons: Theory and Methods*, ser. Guilford School Practitioner. Taylor & Francis, 1996. [Online]. Available: https://books.google.com/books?id=8AK8PUbw3lsC

[100] S. AndradÃşttir, "A Global Search Method for Discrete Stochastic Optimization," *SIAM Journal on Optimization*, vol. 6, no. 2, pp. 513–530, 1996. [Online]. Available: https://doi.org/10.1137/0806027

[101] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," *Nature*, vol. 521, pp. 436–44, 05 2015.

[102] K. Gregor and Y. LeCun, "Learning Fast Approximations of Sparse Coding," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ser. ICML'10. USA: Omnipress, 2010, pp. 399–406. [Online]. Available: http://dl.acm.org/citation.cfm?id=3104322.3104374

[103] A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009. [Online]. Available: https://doi.org/10.1137/080716542

[104] J. R. Hershey, J. L. Roux, and F. Weninger, "Deep Unfolding: Model-Based Inspiration of Novel Deep Architectures," *CoRR*, vol. abs/1409.2574, 2014.

[105] P. Sprechmann, R. Litman, T. B. Yakar, A. M. Bronstein, and G. Sapiro, "Supervised Sparse Analysis and Synthesis Operators," in *NIPS*, 2013.

[106] M. Andrychowicz, M. Denil, S. G. Colmenarejo, M. W. Hoffman, D. Pfau, T. Schaul, and N. de Freitas, "Learning to learn by gradient descent by

gradient descent," in *NIPS*, 2016.

[107] K. Li and J. Malik, "Learning to optimize," *CoRR*, vol. abs/1606.01885, 2016. [Online]. Available: http://arxiv.org/abs/1606.01885

[108] A. Rashdi, N. M. Sheikh, and A. U. H. Sheikh, "Distributed centralized power control algorithms for Very High-Speed Digital Subscriber Lines (VDSL) upstream transmission," in *2005 13th European Signal Processing Conference*, Sep. 2005, pp. 1–4.

[109] G. Scutari, D. P. Palomar, and S. Barbarossa, "Optimal Linear Precoding Strategies for Wideband Noncooperative Systems Based on Game TheoryâĂŤPart I: Nash Equilibria," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1230–1249, March 2008.

[110] Q. Shi, M. Razaviyayn, Z. Luo, and C. He, "An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[111] Z. Luo, W. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite Relaxation of Quadratic Optimization Problems," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, May 2010.

[112] M. Hong, R. Sun, H. Baligh, and Z. Luo, "Joint Base Station Clustering and Beamformer Design for Partial Coordinated Transmission in Heterogeneous Networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 226–240, February 2013.

[113] T. J. O'Shea, T. C. Clancy, and R. McGwier, "Recurrent Neural Radio Anomaly Detection," *CoRR*, vol. abs/1611.00301, 2016.

[114] N. Farsad and A. J. Goldsmith, "Detection Algorithms for Communication Systems Using Deep Learning," *CoRR*, vol. abs/1705.08044, 2017.

[115] N. E. West and T. O'Shea, "Deep architectures for modulation recognition," in *2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, March 2017, pp. 1–6.

[116] T. J. O'Shea, T. Erpek, and T. C. Clancy, "Deep Learning Based MIMO Communications," *CoRR*, vol. abs/1707.07980, 2017.

[117] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, 2017.

[118] H. Yang, J. Wang, X. Song, Y. Yang, and M. Wang, "Wireless base stations planning based on GIS and genetic algorithms," in *2011 19th International Conference on Geoinformatics*.

[119] E. Yaacoub and Z. Dawy, "LTE BS Placement Optimization Using Simulated Annealing in the Presence of Femtocells," in *European Wireless 2014; 20th European Wireless Conference*, May 2014, pp. 1–5.

[120] S. Hurley, "Planning effective cellular mobile radio networks," *IEEE Transactions on Vehicular Technology*, vol. 51, no. 2, pp. 243–253, March 2002.

[121] H. Peyvandi, A. Imran, M. A. Imran, and R. Tafazolli, "A Target-Following Regime using Similarity Measure for Coverage and Capacity Optimization in Self-Organizing Cellular Networks with Hot-Spot," in *European Wireless 2014; 20th European Wireless Conference*, May 2014, pp. 1–6.

[122] D. Tsilimantos, D. Kaklamani, and G. Tsoulos, "Particle swarm optimization for UMTS WCDMA network planning," in *2008 3rd International Symposium on Wireless Pervasive Computing*, May 2008, pp. 283–287.

[123] A. Awada, B. Wegmann, I. Viering, and A. Klein, "Optimizing the Radio Network Parameters of the Long Term Evolution System Using Taguchi's Method," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 8, pp. 3825–3839, Oct 2011.

[124] V. Berrocal-Plaza, M. A. Vega-Rodriguez, J. A. Gomez-Pulido, and J. M. Sanchez-Perez, "Artificial Bee Colony Algorithm applied to WiMAX network planning problem," in *2011 11th International Conference on Intelligent Systems Design and Applications*, Nov 2011, pp. 504–509.

[125] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning Important Features

Through Propagating Activation Differences," *CoRR*, vol. abs/1704.02685, 2017. [Online]. Available: http://arxiv.org/abs/1704.02685

[126] A. Datta, S. Sen, and Y. Zick, "Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems," in *2016 IEEE Symposium on Security and Privacy (SP)*, May 2016, pp. 598–617.

[127] C. Bowles, L. Chen, R. Guerrero, P. Bentley, R. N. Gunn, A. Hammers, D. A. Dickie, M. del C. Valdés Hernández, J. M. Wardlaw, and D. Rueckert, "GAN Augmentation: Augmenting Training Data using Generative Adversarial Networks," *CoRR*, vol. abs/1810.10863, 2018. [Online]. Available: http://arxiv.org/abs/1810.10863

[128] B. Hughes, S. Bothe, H. Farooq, and A. Imran, "Generative Adversarial Learning for Machine Learning empowered Self Organizing 5G Networks," in *2019 International Conference on Computing, Networking and Communications (ICNC)*, Feb 2019, pp. 282–286.

[129] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, March 2004. [Online]. Available: http://www.amazon.com/exec/obidos/redirect?tag=citeulike-20&path=ASIN/0521833787

[130] D. Bertsekas, *Nonlinear Programming*, ser. Athena scientific optimization and computation series. Athena Scientific, 2016. [Online]. Available: https://books.google.com/books?id=TwOujgEACAAJ

[131] L. Xia, W. Li, H. Zhang, and Z. Wang, "A Cell Outage Compensation Mechanism in Self-Organizing RAN," in *2011 7th International Conference on Wireless Communications, Networking and Mobile Computing*, Sep. 2011, pp. 1–4.

[132] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, Oct 1995, pp. 39–43.

[133] W. Hart, "Evolutionary pattern search algorithms," 9 1995.

[134] K. Fernandes, P. Vinagre, and P. Cortez, "A Proactive Intelligent Decision

Support System for Predicting the Popularity of Online News," in *EPIA*, 2015.

[135] I. Syarif, A. Prugel-Bennett, and G. Wills, "SVM Parameter Optimization using Grid Search and Genetic Algorithm to Improve Classification Performance," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 14, p. 1502, 12 2016.

[136] J. Bergstra and Y. Bengio, "Random Search for Hyper-parameter Optimization," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 281–305, Feb. 2012. [Online]. Available: http://dl.acm.org/citation.cfm?id=2503308.2188395

[137] M. Mitchell, *An Introduction to Genetic Algorithms*. Cambridge, MA, USA: MIT Press, 1998.

[138] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4765–4774. [Online]. Available: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[139] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why Should I Trust You?": Explaining the Predictions of Any Classifier," in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: ACM, 2016, pp. 1135–1144. [Online]. Available: http://doi.acm.org/10.1145/2939672.2939778

[140] E. Štrumbelj and I. Kononenko, "Explaining prediction models and individual predictions with feature contributions," *Knowledge and Information Systems*, vol. 41, no. 3, pp. 647–665, Dec 2014. [Online]. Available: https://doi.org/10.1007/s10115-013-0679-x

[141] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2672–2680. [Online]. Available: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

[142] T. J. O'Shea, T. Roy, N. West, and B. C. Hilburn, "Physical Layer Communications System Design Over-the-Air Using Adversarial Networks," in *2018 26th European Signal Processing Conference (EUSIPCO)*, Sep. 2018, pp. 529–532.

[143] T. J. O'Shea, T. Roy, and N. West, "Approximating the Void: Learning Stochastic Channel Models from Observation with Variational Generative Adversarial Networks," in *2019 International Conference on Computing, Networking and Communications (ICNC)*, Feb 2019, pp. 681–686.

[144] H. Ye, G. Y. Li, B. F. Juang, and K. Sivanesan, "Channel Agnostic End-to-End Learning Based Communication Systems with Conditional GAN," in *2018 IEEE Globecom Workshops (GC Wkshps)*, Dec 2018, pp. 1–5.

[145] B. Bollepalli, L. Juvela, and P. Alku, "Generative adversarial network-based glottal waveform model for statistical parametric speech synthesis," in *INTERSPEECH*, 2017.

[146] C. Song, Z. Qu, N. Blumm, and A.-L. BarabÃąsi, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010. [Online]. Available: http://www.barabasilab.com/pubs/CCNR-ALB_Publications/201002-19_Science-Predictability/201002-19_Science-Predictability.pdf

[147] H. Alzer, "On some inequalities for the incomplete gamma function," *Mathematics of Computation*, vol. 66, no. 218, pp. 771–778, 1997. [Online]. Available: http://www.jstor.org/stable/2153894

Funding Acknowledgements

The projects undertaken in this dissertation are based upon work supported by the National Science Foundation under Grant Numbers 1619346, 1559483, 1718956 and 1730650. For more details about these projects, please visit www.ai4networks.com.

Appendix A

The SINR coverage probability for an arbitrary UE can be expressed as

$$P_{cov}(\gamma_{th}) = \mathbb{P}(\text{SINR} > \gamma_{th}) = \mathbb{P}(h_o > \frac{\gamma_{th}(\sigma^2 + \mathbb{I})}{G_{\text{UE}}G_{\text{RRH}}PL(r_{\boldsymbol{o}})}), \tag{7.1}$$

where $\sigma^2 + \mathbb{I}$ is the summation of noise and interference given as denominator in (5.13). Since h_o is considered to be a normalized Gamma random variable, we employ Alzer's Lemma [147] to modify (7.1) as

$$P_{cov}(\gamma_{th}) = \sum_{n=1}^{N}(-1)^{n+1}\binom{N}{n}\mathbb{E}\left(\exp(\frac{-n\eta\gamma_{th}(\sigma^2 + \mathbb{I})}{G_{\text{UE}}G_{\text{RRH}}PL(r_{\boldsymbol{o}})})\right), \tag{7.2}$$

where $\eta = N(N!)^{-\frac{1}{N}}$ and N is the parameter for h_o and takes the value of N_{L} (or N_{N}) depending upon whether the serving RRH is within the LOS (or NLOS) region. The noise and interference components can be treated distinctly as

$$\begin{aligned} P_{cov}(\gamma_{th}) &= \sum_{n=1}^{N}(-1)^{n+1}\binom{N}{n}\mathbb{E}\left(\exp(\frac{-n\eta\gamma_{th}(\sigma^2)}{G_{\text{UE}}G_{\text{RRH}}PL(r_{\boldsymbol{o}})})\right)\mathbb{E}\left(\exp(\frac{-n\eta\gamma_{th}(\mathbb{I})}{G_{\text{UE}}G_{\text{RRH}}PL(r_{\boldsymbol{o}})})\right), \\ &\overset{(a)}{=} \sum_{n=1}^{N}(-1)^{n+1}\binom{N}{n}\mathbb{E}\left(\exp(\frac{-n\eta(\sigma^2)}{G_{\text{UE}}G_{\text{RRH}}PL(r_{\boldsymbol{o}})})\right)L_{\mathbb{I}}\left(\frac{-n\eta\gamma_{th}}{G_{\text{UE}}G_{\text{RRH}}PL(r_{\boldsymbol{o}})}\right), \end{aligned} \tag{7.3}$$

where (a) follows from the Laplace functional of the interference, i.e. $L_{\mathbb{I}}(s) = \mathbb{E}\left(\exp(-s\mathbb{I})\right)$. To evaluate $L_{\mathbb{I}}(s)$, we can split the interfering RRHs into LOS and NLOS considering the distribution of distance from an arbitrary UE to LOS

interferers and NLOS interferers is only weakly dependent [91]. Hence, if we assume Φ_{ILOS} and Φ_{INLOS} to be the PPPs of the interfering RRHs within LOS and NLOS regions respectively, then by applying the independence property, (7.3) becomes

$$P_{cov}(\gamma_{th}) = \sum_{n=1}^{N}(-1)^{n+1}\binom{N}{n}\mathbb{E}\left(\exp(\frac{-n\eta\gamma_{th}(\sigma^2)}{G_{\mathrm{UE}}G_{\mathrm{RRH}}PL(r_{\boldsymbol{o}})})\right) L_{\mathbb{I}_{\mathrm{L}}}(\frac{-n\eta_{\mathrm{L}}\gamma_{th}}{G_{\mathrm{UE}}G_{\mathrm{RRH}}PL(r_{\boldsymbol{o}})})L_{\mathbb{I}_{\mathrm{N}}}(\frac{-n\eta_{\mathrm{N}}\gamma_{th}}{G_{\mathrm{UE}}G_{\mathrm{RRH}}PL(r_{\boldsymbol{o}})}). \tag{7.4}$$

Let us consider that the arbitrary UE under consideration is being served by an RRH within its LOS region (i.e. $r \leq R_o$). Now, the interfering RRHs could be in either the LOS or the NLOS region. For interferers in the LOS region, the Laplace functional of the interference in (7.4) is calculated as

$$\begin{aligned} L_{\mathbb{I}_{\mathrm{L}}}(\frac{-n\eta_{\mathrm{L}}\gamma_{th}}{G_{\mathrm{UE}}G_{\mathrm{RRH}}PL(r_{\boldsymbol{o}})}) &= \mathbb{E}\left(\exp(\frac{-n\eta_{\mathrm{L}}\gamma_{th}r^{\alpha_{\mathrm{LOS}}}\sum_{i\in\Phi_{\mathrm{ILOS}}}|h_i|^2G_ir_i^{-\alpha_{\mathrm{LOS}}}}{G_{\mathrm{UE}}G_{\mathrm{RRH}}})\right) \\ &\overset{(b)}{=} \exp\left(-2\pi\lambda_{\mathrm{IRRH}}\sum_{k=1}^{4}b_k\int_{r}^{R_o}[1-\mathbb{E}_{h_i}(\exp\{-n\eta_{\mathrm{L}}\gamma_{th}h_i\bar{a_k}(r/t)^{\alpha_{\mathrm{LOS}}}\})t\mathrm{dt}]\right), \end{aligned} \tag{7.5}$$

where the directivity gain of the interfering RRH G_i is evaluated by considering it as a discrete random variable [90]. Note that we have assumed the pathloss intercepts for both LOS and NLOS scenarios as unity. Further, (b) follows from computing the Laplace functional of Φ_{ILOS}. Finally, by computing the moment generating function of the normalized gamma random variable h_i, we obtain

$$L_{\mathbb{I}_{\mathrm{L}}}(\frac{-n\eta_{\mathrm{L}}\gamma_{th}}{G_{\mathrm{UE}}G_{\mathrm{RRH}}PL(r_{\boldsymbol{o}})}) = \exp\left[-2\pi\lambda_{\mathrm{IRRH}}\sum_{k=1}^{4}b_k\int_{r}^{R_o}F\left(N_{\mathrm{L}},\frac{n\eta_{\mathrm{L}}\bar{a_k}\gamma_{th}r^{\alpha_{\mathrm{LOS}}}}{N_{\mathrm{L}}t^{\alpha_{\mathrm{LOS}}}}\right)t\mathrm{dt}\right] = \exp\left[-I_{\mathrm{LL}}(\gamma_{th},r)\right]. \tag{7.6}$$

In a similar manner, the Laplace functional of the NLOS interfering RRHs for this

UE is given by

$$L_{\mathbb{I}_{\mathrm{N}}}(\frac{-n\eta_{\mathrm{N}}\gamma_{th}}{G_{\mathrm{UE}}G_{\mathrm{RRH}}PL(r_{\boldsymbol{o}})}) = \exp\left[-2\pi\lambda_{\mathrm{IRRH}}\sum_{k=1}^{4} b_k \int_{r}^{R_o} F\left(N_{\mathrm{N}}, \frac{n\eta_{\mathrm{L}}\bar{a_k}\gamma_{th}r^{\alpha_{\mathrm{LOS}}}}{N_{\mathrm{N}}t^{\alpha_{\mathrm{NLOS}}}}\right) t\mathrm{dt}\right] = \exp\left[-I_{\mathrm{LN}}(\gamma_{th}, r)\right]. \tag{7.7}$$

Now for a UE served by an RRH in the NLOS region, it is intuitive to observe that $\Phi_{\mathrm{ILOS}} = \emptyset$, therefore, the Laplace functional

$$L_{\mathbb{I}_{\mathrm{L}}}(\frac{-n\eta_{\mathrm{L}}\gamma_{th}}{G_{\mathrm{UE}}G_{\mathrm{RRH}}PL(r_{\boldsymbol{o}})}) = \exp[-I_{\mathrm{LN}}(\gamma_{th}, r)] = 1. \tag{7.8}$$

Similar to (7.7), we derive the Laplace functional of the NLOS interferes as

$$L_{\mathbb{I}_{\mathrm{N}}}(\frac{-n\eta_{\mathrm{N}}\gamma_{th}}{G_{\mathrm{UE}}G_{\mathrm{RRH}}PL(r_{\boldsymbol{o}})}) = \exp\left[-2\pi\lambda_{\mathrm{IRRH}}\sum_{k=1}^{4} b_k \int_{r}^{\infty} F\left(N_{\mathrm{N}}, \frac{n\eta_{\mathrm{N}}\bar{a_k}\gamma_{th}r^{\alpha_{\mathrm{NLOS}}}}{N_{\mathrm{N}}t^{\alpha_{\mathrm{NLOS}}}}\right) t\mathrm{dt}\right] = \exp\left[-I_{\mathrm{LN}}(\gamma_{th}, r)\right]. \tag{7.9}$$

Finally, integrating over $f_{r_o}(r)$ for LOS and NLOS regions and by summation of coverage probabilities for each region, we obtain (5.14).