ENERGY-EFFICIENT BASE STATION SLEEP-MODE STRATEGIES FOR ULTRA-DENSE CELLULAR NETWORKS

by

Johnson Opadere

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Electrical Engineering

Charlotte

2019

Approved by:

_____
Dr. Tao Han

_____
Dr. Asis Nasipuri

_____
Dr. Thomas P. Weldon

_____
Dr. Yu Wang

ABSTRACT

JOHNSON OPADERE. Energy-efficient Base Station Sleep-mode Strategies for Ultra-dense Cellular Networks. (Under the direction of DR. TAO HAN)

Future mobile cellular networks will be characterized by massive densification of base stations (BSs) and a reshape of network functionalities, which include the potentials of advanced technologies like the network function virtualization, network slicing, and resource sharing among multiple operators. In the legacy networks, sleep-mode strategies have been proven to significantly increase the network energy efficiency.

In this dissertation, BS sleep-mode schemes are formulated and proposed for the two most prevalent ultra-dense network types in the 5G mobile cellular networks. First, a multi-operators cellular network, where BSs are closely located is considered. Sleep-mode with Efficient Beamformers and Spectrum-sharing (SEBS) strategy, which minimizes BS power consumption of cooperative multi-operators is proposed. The licensed bandwidth of each operator is partitioned into private and shared bands to avail the active BSs sufficient spectrum resources for the support of all UEs.

A mobile edge computing network is also considered, where densely deployed BSs are equipped with computation resources to process users offloaded computation-intensive tasks. In order to jointly tackle the issue of power consumption and latency interplay, the active number of BSs, uplink and downlink beamforming vectors, computation resource allocation, and task completion latency are formulated as an optimization problem, with the aim of reducing the network power consumption while satisfying the latency requirement. To efficiently solve the resulting joint optimization problem, a framework that first selects the active BSs based on communication and computation power-aware selection rule is proposed, and thus the remaining BSs can be switched off. The computation resources and dual-link beamformers are subsequently optimized for further network energy savings.

The proposed energy efficiency strategies in this dissertation are endowed with active BSs selection, beamforming vectors optimization, latency minimization and bandwidth sharing, which makes this research applicable to various cellular network types. Therefore, this research will provide important insights into the development of energy efficiency strategies of future mobile cellular networks.

ACKNOWLEDGEMENTS

I wish to thank and acknowledge the contributions of several people who guided, encouraged and contributed to this dissertation.

First and foremost, I express my gratitude to my advisor, Dr. Tao Han, for giving me an excellent opportunity to join his very innovative Wireless Networks and Systems (WinSys) research group. I am particularly grateful for his patience with me, his guidance and invaluable encouragement throughout the development of this dissertation.

I would also like to thank my committee members: Dr. Asis Nasipuri, Dr. Thomas P. Weldon, and Dr. Yu Wang for their advice and time. I am indebted to many professors who have taught and guided me during my PhD education in classes, and other areas of my academic life: Dr. Aba Ebong, Dr. Shree Bhattarai, Dr. Linda Jiang Xie, Dr. Miri Mehdi, Dr. Bruce Willis, Dr. Yogendra Kakad, Dr. Benny Rodriquez-Medina, Dr. Jiancheng Jiang, and Dr. Jim Conrad.

I am grateful to my colleagues and friends: Qiang Liu, Dr. Kang Li, Dr. Prince Anyaba, Dr. Saeed Mohajeryami, Dr. Yawo Amengonu, Nazmus Sakib, and many others. I appreciate the help I have received from them.

My appreciation also goes to my mother, and my brothers: Samuel Alade, Olajide, and Yinka Femi, who have over the years supported and encouraged me. Finally, I would like to give a big thanks to my wife, Dianarose, for her patience and unwavering support through this journey.

DEDICATION

To the loving memory of my father,

Prince Abel Opadere Olaosebikan

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

η_k	Zero mean i.i.d complex-symmetric additive Gaussian noise
γ_k^D	SINR of the k-th UE downlink transmission
γ_k^U	SINR of the k-th UE uplink transmission
$(\cdot)^{\mathsf{H}}$	Hermitian transpose
$\mathbf{h}_{kv}$	Channel vector from v-th RRH to the k-th UE
$\mathbf{m}$	Uplink beamforming vector
$\mathbf{w}$	Downlink beamforming vector
$\mathbf{w}_{vk}^{pr}$	Intra-operator beamforming vector
$\mathbf{w}_{vk}^{sh}$	Inter-operator beamforming vector
$\mathcal{I}$	Number of APs
Υ_k^{pr}	SINR ratio for the k-th UE served in private spectrum
Υ_k^{sh}	SINR ratio for the k-th UE served in shared spectrum
A	Total number of RRHs in the region
a_m	RRH mode indicator
B_m	Operator m total allocated bandwidth
B_{pr}^m	Operator m private bandwidth
B_{sh}	Combined shared bandwidth of cooperating operators
D_k	The size of the input data of the computation task from the UE k
f_k	Allocated CPU cycles by AP to process UE k's task

K The set of UEs in the region

M Number of cooperating operators

N Number of RRH antennas

P^{bh} The backhaul power consumption

P_m The total BS power consumption of operator m

P_{comm} Total communication power of an AP

P_{comp} Computation power of an AP

P_T The joint total BS power consumption of the cooperating operators

$P^f_{v,sleep}$ The power consumed in the fronthaul link when RRH v is inactive

$P^{rrh}_{v,sleep}$ The power consumed by RRH in sleep mode

$P^f_{v,static}$ The power consumed in the fronthaul link when RRH v is active

$P^{rrh}_{v,static}$ Static power consumption of RRH

P_v RRH transmission power

P^f_v The power consumed in the fronthaul link of the RRH v

P^{rrh}_v RRH power consumption

R^D_k The downlink achievable rate for user k

R^{pr}_k Achievable data rate of the UE served in the private spectrum

R^{sh}_k Achievable data rate of the UE served in the spectrum spectrum

R^U_k The uplink achievable rate for user k

s_k the k-th user data symbol

$T_{i,k}^{exe}$ Execution time UE of AP server i for procession UE k's task

T_k^{tot} Total computation latency for processing UE k's task remotely

T_k^U Transmission time of user k for offloading the task size to AP

U_k The total number of CPU cycles required to complete the task from the UE k

V The set of active RRHs

y_k The signal received by the k-th UE

y_k^{pr} The signal received by the k-th UE using private spectrum

y_k^{sh} The signal received by the k-th UE using shared spectrum

Z The set of inactive RRHs

CHAPTER 1: INTRODUCTION

As we are on the brink of manifestation of 5G networks where at least 100 billion devices will be supported, and a 1,000-fold surge in capacity compared to the legacy networks, the trade-off between the imminent spectral increase and energy efficiency (EE) has become one of the main focuses of 5G network researchers. The increased capacity demand is in response to the continued rise in the mobile traffic. The global mobile data traffic is forecasted to reach 49 exabytes per month, or a run rate of 587 exabytes annually [2]. On the part of mobile network operators (MNOs), the massive densification of base stations (BSs) is one of the capacity building measures to satisfy the coverage needs of the mobile traffic growth. Incidentally, most power consumed in cellular networks occurs in the BSs. The energy consumption at the BSs represents about 60% of the total energy consumed in a cellular network (see Figure 1.1). The high power dissipation emanates mainly from inefficient power conversion, especially at the RF power amplifier section. The power amplifier is responsible for 60 to 70% comsumption of the supplied power to the BS [1]. From the remaining supplied power, considerable quantity is dissipated as heat, and only a small fraction is utilized for transmission output. With the MNOs' densification approach to meet the enormous growth in data rates demand and market penetration, high BS energy consumption will continually result in substantial energy bills, constitutes a significant part of MNOs' operational expenditures (OPEX).

Moreover, the BS deployment expansion will yield increased CO_2 emission since fossil fuels are one of the primary sources for producing electrical energy supplying the BS sites. The studies conducted in [3,4] estimate the contribution of mobile networks to be 0.2 of the global CO_2 emissions in 2007 and projected to be 0.4 in 2020. Thus,

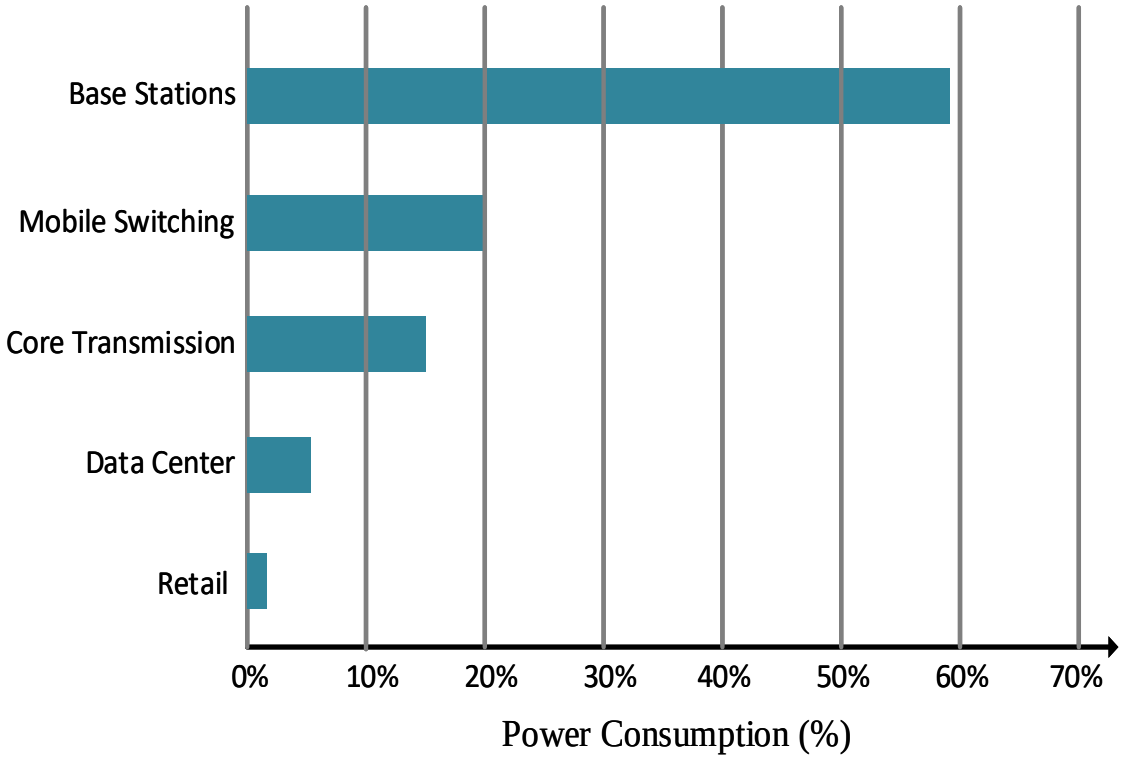

Figure 1.1: Percentages of power consumption in cellular network infrastructure [1]

improving the energy-efficiency of the BSs can both achieve OPEX savings and CO_2 emission reduction.

One way to reduce BS power consumption is by designing more energy efficient power amplifiers since the most percentage of power dissipation is attributed to the power amplifiers as aforementioned. However, this method has an insignificant effect of energy efficiency if the BS is always active [5]. BS sleep-mode has therefore been considered as a more efficient approach for high energy saving. Cellular network traffic exhibits variations during a typical daily cycle. As the traffic fluctuates temporally and spatially, under-utilized BSs can be dynamically put on sleep-mode ro conserve power. For instance, the application of BS sleep-mode scheme has been saving an estimated power of is 36 million kWh per year for China Mobile since 2009 [6]. As the BS deployment get densier, and more emerging paradigm incorporated into mobile networks, there is even more needs to design sleep-mode strategies to meet the

requirements of the imminent 5G and future networks.

1.1 Sleep-mode Schemes in cellular Networks

Many strategies have been proposed to activate BS sleep-mode to reduce overall network power consumption. The schemes are delineated in Figure 2.1. In this section, we discuss the aforementioned strategies and their efficiency in providing energy savings at the the cellular networks base stations.

1.1.1 Specific cell(s) switching

Specific BSs are powered off to conserve energy while the neighboring BSs take over radio support to the mobile users (MUs). Sleep-mode is restricted only to the particular BSs while the rest always remain active. In [7], switching off only the central cell is proposed. The technique is premised on cluster cells deployment with overlapping coverage. The merit of this method is that is relatively simple. It is only targeted at the spatially center cell. However, the simplicity may be inefficient in areas where there is temporal traffic fluctuation. The algorithm heavily relies on BS selection based on spatial location instead of the traffic weight. Therefore, the algorithm may not be optimal because the most central cell does not necessarily imply the one with the least traffic.

1.1.2 Load Adaptiveness

One of the ideas of sleep-mode is to disengage some BSs in a cluster when the traffic load is reduced. Temporal variation of traffic load and its relationship to BS power consumption is investigated in [8]. The authors consider load fluctuation at different hour in a 24-hour period in a certain location. The study in [9] details the spatial variation of traffic load, using Europe as a case study. For both cases, some BSs are switched to sleep mode according to some load threshold. However, these two studies are only reliable for areas where there is one of temporal and spatial traffic fluctuation. In certain locations such as campus, the variation could be either way

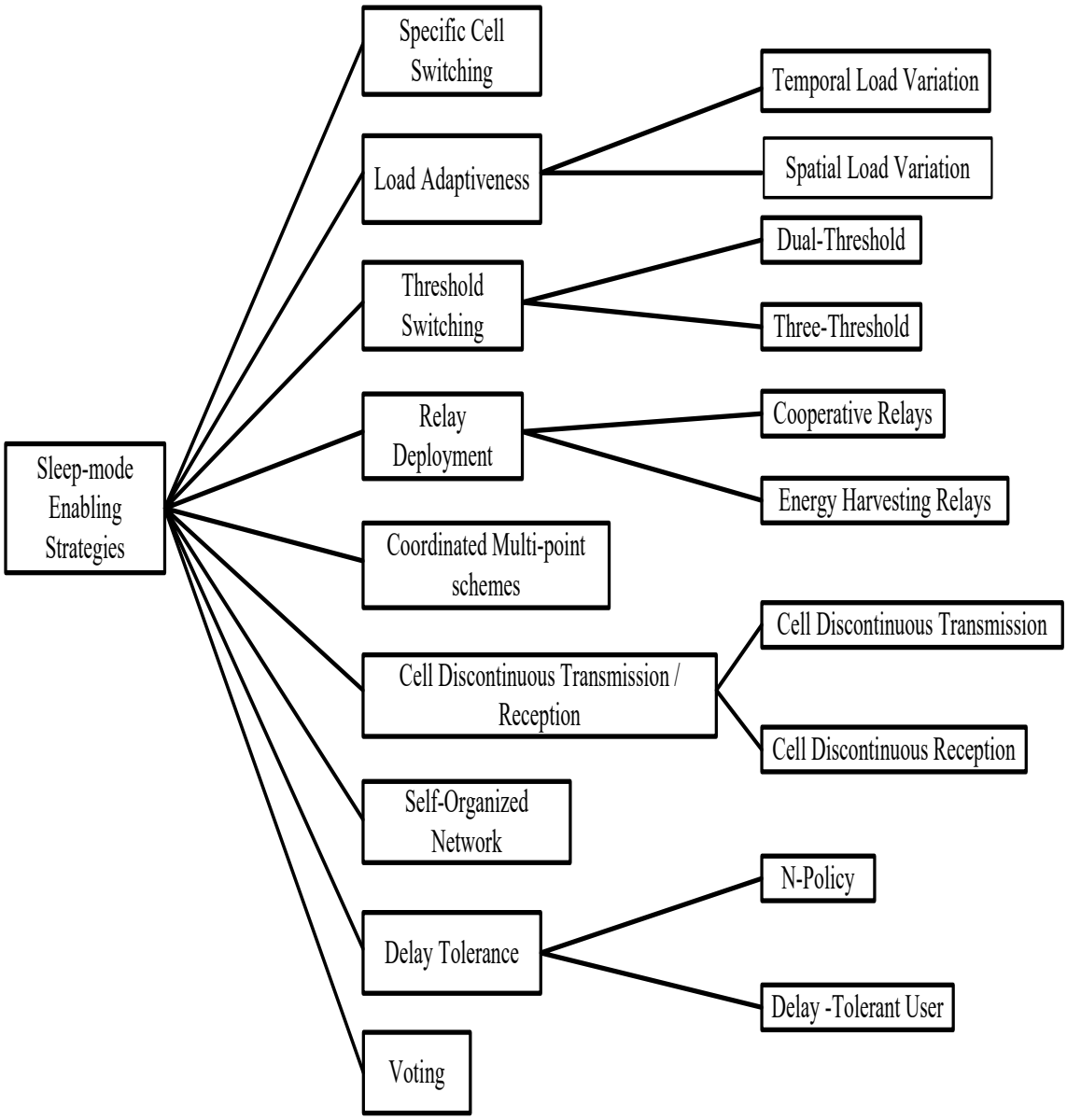

Figure 1.2: Categories of sleep-mode enabling strategies

intermittently.

1.1.3 Switching Threshold

In some studies BS energy savings based on specific network QoS are proposed. The proposal made in [10] provides network QoS based threshold for putting eNBs

to sleep mode. A retransmission tolerance range determined by QoS, such as delay and throughput, is used to select which nodes to switch off. Beside the advantage of having low complexity, the load level of each BS is also considered in the sleep activation threshold. A sleep mode triggered by dual thresholds is proposed in [11]. The triggering thresholds for switching are traffic load and cell-edge QoS constraint.

1.1.4 Relay Deployment

Combination of relays and BS switching has been presented to improve energy efficiency. For instance, sleep mode with optimal placement of relays in a network is considered in [12]. Simulation results from numerical example confirm power savings with optimal relay location when combined with BS sleep-mode at low loads. A similar approach has also been proposed in [13], but with more elaborate schemes. The algorithm minimizes the coverage area and power consumption, and optimally places relay stations for the uncovered users. Multi-hop and cooperative relays are used with BS sleep-mode in [14]. The use of Energy harvesting (EH) relays with MBS sleep mode has been investigated in [15]. Different densities of BSs were analyzed. Without application of sleep mode, the energy efficiency (EE) of the network ultimately declines. The results give stable and better EE with the combination of EH relays and BS sleep-mode.

1.1.5 Coordinated multi-point (CoMP) schemes

CoMP with either cell zooming or sleep mode has been been noticed to improve SINR in [16, 17]. The simulation results in [16] further show that CoMP with sleep mode produces more energy savings than CoMP with cell zooming. The study adds that, at sleep mode, zoom out by active cells gives improved performance at minimizing coverage holes. The performance is achieved with the cooperative characteristic of CoMP by allowing MUs situated in coverage holes and cell edges communicate with multiple BSs. The results obtained in [18] affirmed than CoMP with sleep mode

gives 48% energy efficiency over Non CoMP with all active BSs.

1.1.6 Cell discontinuous transmission

Cell discontinuous transmission is the deactivation of some hardware components of a BS during the empty transmission time intervals (TTIs). This can be referred to as intermittent short term BS sleep-mode. Some of the early investigations on DTX [19, 20] highlighted the energy saving capability of employing DTX in network planning. When included in the network planning, a 42% energy reduction is reported by [21] as compared to when BSs are put to long term idle mode. A major drawback of employing DTX in network planning is high cost. For the aforementioned increase in [21], the cost surges by 110%. For the corresponding energy saving concept at MUs, the discontinuous reception (DRX) potentials are noted in [22, 23].

1.1.7 Self-organizing networks

The abstraction of Self-organizing network (SON) was introduced in the 3GPP standard (3GPP TS 32.521). The SON control algorithms can be used to put some BSs to sleep mode for energy savings. In [24], MUs access are categorized into registered and non-registered users. Priority access to the network is allowed only by registered users, and the BSs with no such users can be put to sleep mode. While in sleep mode, only registered users can activate the sleeping BS. SON utilization for energy savings by timed sleep mode is proposed in [25] to coordinate and invoke sleep mode at pre-defined time interval. Coverage demand is met by adjacent BSs.

1.1.8 Delay tolerance

A trade-off between energy consumption and delay in dense cellular networks has been established [26]. Therefore the trade-off can be characterized in ways that some resources could be saved by compromising on certain QoS performance. To formulate strategies hinging on delay tolerance, BS is modeled as an M/G/1 vacation queue in [26]. BS is put to sleep-mode if there is no network access request from MUs and

remains in this mode until N MUs queue builds up (N-Policy). The result shows some limit where increasing the number of MUs leads to decrease in mean power consumption decrease. Also, above the limit monotonic decrease of power is not achieved. Thus, there exist delay bound in which the BSs energy savings is increased.

1.1.9 Voting

The authors in [27], explore the application of voting to enable sleep mode of BS. In the study, a cluster is assumed to be comprised of one MBS per cell, and multiple SBSs within the cell. Sleep mode mechanism is applied to the MBS. When compared with the neighboring MBSs, the MBS with the lowest traffic load value is put to sleep. Each BS 'votes' for the BS with the lower load metric value. The vote is updated as the load information is shared among neighbors. In meeting with the network QoS requirement, the number of current users is taken into account. Therefore the metric value, with which comparison is made, is determined as the ratio of the number of the current associated MUs to the number of votes. This technique has the advantage of selection of active BSs at each cycle of pre-voting. This prevents inclusion of inactive BSs as candidates to be voted to sleep. The study, however, does not consider the wake-up mechanism for the strategy.

1.2 Wake-up Schemes

In the literature, some consider and include the schemes to wake up the sleeping BSs at the end of the idle mode to resume normal operations. However, others do not. Therefore, we present, in this section, the sleep mode wake-up techniques proposed in the literature.

No technique is one-size-fits-all. Stand-alone self activation offers simple and low overhead wake-up scheme, but it is prone to energy inefficiency when more than necessary number of BSs self-wake. Activation by MBS provides the advantages of central coordination from the SBS, which precludes the downside of the self-activation.

However, some form of its implementation may trigger all supported SBSs instead of the required number for the traffic. This is elaborated later in this section. Though it is not suitable for non-delay tolerant users, queuing based activation harnesses the energy-delay tradeoff(EDT), which guarantees required level of users to be supported before switching the BSs. Wake-up by rewards offers the possibility of longer sleep duration and prompts wake-up due to the assigned rewards to the modes. Reward process is prone to high complexity.

1.2.1 Stand-alone Self Activation

In heterogeneous networks where MBS are always put on, small cell self activation is proposed in [28]. The sleeping SBS has its uplink receiver turned on to intermittently measure the interference plus noise $(I + N)$. The detection of an increase in $(I + N)$ when a connection request is made from MU to MBS will trigger the SBS to wake up and connect to the MU if it has higher $(I + N)$ than the MBS.

1.2.2 Activation by Macro Base Station

MBS is used to wake up sleeping SBSs in the same study [28]. When the number of MUs requesting connection with the MBS is greater than a pre-determined threshold, the MBS wakes up all the SBSs in the coverage area. However, the strategy does not select which SBS is woken up to assist and instead all are woken. This may ultimately not be energy efficient. The authors also explore the use of an indicator, called Timing Advance (TA) for showing how far the MU is from its serving BS. Each MU's uplink signal is transmitted with time TA ahead of the sub frame boundary. The TA of each SBS is also intermittently received. At MUs connection request, the MBS is able to determine the distance with the TA and wake the corresponding SBS closest to the MU. The method precludes unnecessary switching on all the SBSs in the coverage area.

1.2.3 Queuing Based Activation

For a case where sleep mode is formulated as a queue model, BS is put to sleep when the network is 'empty' and waits for some time. While waiting, the BS is woken up when a user arrives. The pattern of sleep is referred to as hysteresis sleep. As proposed in [29], the wake up schemes can be implemented in three forms as shown in Fig. **??**. In Single Sleep (SS) scheme, the BS wakes up from sleep mode after a certain pre-determined time. This is simple to implement, but it is insensitive to QoS degradation that may arise from 'blind' sleeping. In multiple Sleep (MS) scheme, the sleeping BS wakes up when it discovers a waiting user, as it listens to the network status at a predefined time interval. When no waiting user is detected when it listens, it continues sleeping. This approach solves the setback of SS scheme, but also requires additional power for periodical listening to the network status. The third approach is the N-limited scheme, where the BS wakes up when there are N users waiting in the network. This gives potential of higher energy efficiency due to EDT, but additional network element is required to keep the counts.

1.2.4 Wake-up by Access Reward

Game theoretical approach is applied in [30] where assess to network by MUs is rewarded. To wake up an idle BS, reward assignment is used, with higher reward assigned to the MUs requesting access. Conversely, the BS in sleep mode is given less revenue relative to the active ones. Thus, MUs requiring service are encouraged to wake up the idle BS and the BS is equally motivated.

1.3 Summary

The discussed achemes have proven to improve energy efficiency in conventional cellular networks, though each method has merits and drawbacks. Switching specific cells has the advantage of simplicity in implementation, but does not guarantee an optimal choice of the cells to be switched off. Load adaptiveness method takes the

network traffic profile into account and ensure the least loaded BS are put into sleep mode. However, this method is prone to creating coverage holes [31]. Another load dependent approach utilized is switching threshold. The method has low complexity, but not applicable where the traffic profile is erratic. Power savings by relay deployment interestingly incurs low operating cost, but proper network planning is required to ensure optimal placement.

Coordinated multi-point (CoMP) schemes are great with minimizing coverage holes. Cell discontinuous transmission and self-organizing networks give high energy savings, however high cost is incurred in incorporating them in network planning [21]. Application of some delay, to achieve lower BS power consumption, is flexible and it does not require hardware cost. The setback is that it is not applicable where or when non-delay tolerant users are serviced. Voting strategy ensures the best candidate BS is put to sleep. The method, however, is not immune to challenges like coverage holes, if it is not combined with other algorithms. Network function virtualization gives relatively better energy control with high potential savings. The application is still nascent, but promising and evolving.

1.4 Overview of the Proposed Research

Many studies and the schemes discussed have achieved considerable results in minimizing power consumption at the BSs. However, most of the proposed methods are targeted and applicable to the pre-5G legacy networks. 5G networks will be characterized by massively dense BS deployment, and a reshape of network functionalities, which includes the potentials of advanced technologies like, network function virtualization, network slicing and resource sharing. [32]. It is therefore imperative to develop BS energy-saving schemes compliant with the emerging 5G network paradigms. In this research work, we focus on BS power consumption reduction strategies in tandem with the network functionalities of the 5G.

CHAPTER 2: RELATED WORKS

In this chapter, the related work in the proposed strategies to minimize BS power consumption in ultra-dense mobile cellular networks.

2.1 Existing Sleep-mode Strategies in Multi-operator Network

In the first two decades of cellular mobile network roll-out, the network infrastructure was based on exclusive ownership and utilization by individual mobile network operators [33]. The RF planning and locations of BSs of different operators are usually implemented independently of each other. However, the market coverage competition and increased number of subscribers sometimes lead to colocation of BSs within an overlap region of coverage for different MNOs. The participation of third-party mobile tower companies has also led to increased incidence of the BS colocation. In the US for instance, a strong presence of colocation is fostered by companies which sublease space from independent landlords to deploy BSs belonging to more than one MNO at a single location [34]. While the colocation helps an individual MNO to hold its share of subscribers in the region, there are redundancy and under-utilization of power-hungry BS components, especially at low traffic [35].

Recent studies have focused on minimizing OPEX resulting from under-utilization and high power consumption of colocated BSs belonging to different operators. Many works mainly centered on infrastructure sharing, such as site sharing, tower sharing, Radio Access Network (RAN) sharing, and core network sharing [36]. The RAN sharing approach to energy saving has received much research attention because BSs are an integral part of RAN. A study on the adoption of active RAN sharing in [37] reveals potential savings up to $60 Billion in 5 years. The drawback of the inter-operator

spectrum sharing through the active RAN sharing is the constraint in separating both data and control planes among the different participating operators, and the flexibility of inclusion of individual operator's requirements [38]. Network virtualization is introduced to address these needs such that services are separated from their underlying infrastructure [39]. In this paper, we posit MNOs to pool their RAN resources to form a virtual network, and a mobile operator can rely on the coverage of other operators' networks to serve its users in the region of overlapping coverage. Therefore, in the low traffic scenario, some MNOs can switch off their BSs to conserve power while other operators support the users of all MNOs in the region. The BS energy saving by the sleep-mode technique is hinged on the cooperation among BSs either in a single-operator network or a multi-operators network. When some BSs are put into the sleep-mode, these inactive BSs release their spectrum resources for the active neighboring BSs to be utilized for the coverage expansion to support all users [40]. Among existing works on BSs energy saving in cellular networks, most of them focus on a single MNO. The studies [41–43] focus on the user association, where an optimal association of users and BSs is sought to yield the required number of BSs to support the users, and then to switch off other BSs. In [44], a sleep-mode strategy is used to improve the energy efficiency of a wireless network, where an interference alignment technique is used to manage interference among users. Energy saving is further enhanced by the introduction of transmission-mode adaptation, which implies reallocation of transmitted power of active users following user sleep-mode control. However, the sleep-mode strategy is only applied to the users.

The BS sleep-mode by self-organizing network, where BS configurations can be automatically adjusted to adapt to traffic condition, was applied in [45, 46]. In [47], energy efficiency (EE) of a network is improved via zero and partial zero forcing based approaches, where beamforming and power allocation are jointly optimized in the presence of inter- and intra-cell interference. However, most of the proposed sin-

gle operator strategies are targeted and applicable to the pre-5G legacy networks. 5G networks will be characterized by a reshape of network functionalities, which include the potentials of advanced technologies like the network function virtualization, network slicing, and resource sharing among multiple operators [32]. It is imperative to develop BS energy-saving schemes compliant with the emerging 5G network paradigms. Therefore, recent research directions are geared towards multi-operators cellular networks.

Collaboration among wireless network operators has been shown to reduce OPEX as elicited in cooperation between a cellular operator and Internet Service providers (ISPs) [48–50], and in multi-operators cellular networks [51]. Cooperation among multiple operators for the goal of energy saving has been researched in recent studies. In [52], the BS sleep-mode strategy based on a game theory algorithm has been proposed to select active BSs between two cooperating MNOs while incorporating the fairness concerning the inter-operator roaming cost. The drawback of this work is the limitation to two network operators. Multi-operators cases are studied in [53–55]. In [53], a 24-hour time profile BS switch-off pattern has been proposed for cooperating MNOs. The pattern, however, may not be feasible in practice as all but one network are switched off at a time. Consequently, the only active BS may be overburdened at the expense of network QoS. BS sleep-mode schemes using game theory are also studied in a cooperative multi-operators network, such as in [54, 55], where network stability is considered. The schemes are extended to a heterogenous multi-operators network in [56]. Most of the discussed works mainly focus on optimized solutions solely based on network dynamics akin to the case of a single operator. However, to be feasible in 5G networks, a step further is required to capture the integration of modern paradigms, such as the network virtualization and dynamic resource pooling, in actualizing collaborative BS energy saving. In the comparative study on multi-operator systems done in [35], the BS sleep-mode strategy in a virtual-

ized network has achieved significant energy savings as compared to the conventional network. However, the scheme is applicable for offline applications in a region where the number of BSs and user statistics are known *a priori*.

2.2 Existing Sleep-mode Strategies in Mobile Edge Networks

One of the fundamental techniques for boosting network capacity is network densification, where base stations are deployed in an ultra-dense fashion. Also, incorporation of cloud computing into radio access networks (RANs), facilitated by Cloud-RAN, is geared towards meeting the requirement of system capacity efficiently. Cloud-RAN is based on centralization of base station baseband processing into a pool with data traffic conveyed between the pool and densely deployed remote radio heads (RRHs) over a fronthaul.

However, Cloud-RAN and network densification come with some challenges, such as fronthaul constraints [57] and increase in energy budget due to the constant operation of the network devices [58]. To alleviate these constraints, mobile edge and fog-computing based RANs are proposed to alleviate the Cloud-RAN fronthaul challenge. These computing based RAN architectures move cloud services such as computing and storage to the RRHs, closer to the user equipment (UEs). Depending on the architecture and the functionalities, the computation equipped RRHs have been termed enhanced RRHs (eRRHs) [59], fog-computing Access Points (F-APs) [60], and edge-computing Access Points (EC-APs) [61]. The term Access Points (APs) is adopted in this paper to describe these base stations capable of supporting radio and edge node computing services.

With edge and fog computing, the densely deployed APs can provide radio resource to the UEs, and capability of execution of computation tasks. However, such AP network deployment exhibits significant traffic variations [58]. In a trace study conducted in [58], the traffic profile significantly differs between weekdays and weekends. Hourly fluctuation of traffic is also observed in the same study. Consequently, some APs are

either idle or underutilized at low traffic. Higher fluctuating traffic profile is expected in 5G wireless systems, where densely deployed edge devices, such as APs, can be idle at every time instant [62]. Therefore, it will be considerably energy-efficient to put the optimal number of APs into operation to meet the communication and computation demands of the UEs and switch off other APs.

Besides the benefit of energy efficiency, offloading of UEs' computation workloads to APs provides lower latency for task processing. The conventional way of offloading computation to the remote cloud could cause intolerable delay [63]. For instance, computation processing request round-trip times (RTTs) to remote cloud servers could amount to the order of tens of milliseconds, while edge servers could proffer less than 10ms RTT [64]. Thus, computation offloading to the proximate servers is increasingly becoming attractive.

The need to optimally select active number of APs in ultra-dense networks with fluctuating traffic profile can be intuitively illustrated. A higher number of active APs leads to lower computation latency as more edge servers are available for computation processing. However, the network power consumption rises due to the high number of APs in operation. Thus, we are motivated by the interplay between computation completion latency at the edge servers and the number of active APs to seek optimal AP subset to support radio and computation traffic demands with the ultimate aim of minimizing network energy consumption and satisfying delay constraint.

Energy efficiency strategies in the downlink Cloud-RAN systems commonly entail power-aware AP selection, efficient radio resource allocation and optimal precoding methods on AP–UE channel. In Cloud-RAN's extensions such as network architectures for mobile edge and fog-computing based RANs, more energy minimization approaches are introduced due to additional network capabilities like caching and computation. Besides selecting power efficient cache hardware to improve AP energy efficiency [65], storing of most frequently requested contents at the edge server

is prevalent in recent studies [66, 67]. These studies focus on minimizing AP and fronthaul power consumption costs by caching the popular contents at the AP server and optimizing the network-wide beamforming vectors. However, the studies do not consider the scenarios and application where APs engage in computation tasks processing for the UEs.

There exist only a few works on energy efficient computation at the edge servers. The works like [68–71] focus on optimizing resource allocation for a reduction in AP power consumption. However, optimal selection of AP is not considered in these works. While [72] considers optimal AP mode that will yield efficient power consumption, beamforming vectors are not included in the optimization strategy. AP downlink energy transmit beamformer is jointly optimized with CPU frequencies in [73], but the sleep-mode strategy is not considered, thus making it inefficient for the ultra-dense network if all APs are active at all times regardless of traffic fluctuations.

In contrast to the above works, this paper jointly considers reducing network system power dissipated by communication and computation infrastructure and resources by selecting an optimal number of active APs and minimizing the dual link beamforming vectors, and optimally allocating computation resource for offloaded tasks processing at the network edge.

CHAPTER 3: PROPOSED SLEEP-MODE WITH EFFICIENT BEAMFORMERS AND SPECTRUM-SHARING IN A MULTI-OPERATORS NETWORK

3.1 Introduction

In this research, we present the Sleep-mode with efficient beamformers and equal spectrum pooling strategy for energy-efficient BS operations in the collaborative multi-operator cellular system. Specifically, based on the traffic condition, an optimal number of BSs is selected to serve the users in the region, and other BSs are put to the sleep-mode. We leverage on the network virtualization to facilitate the inter-operator spectrum resource pooling and access for mutual users support with the ultimate goal of joint network energy saving. The virtualized RAN (VRAN) applied in our work hinges on the cloud-RAN (C-RAN) architecture. The VRAN involves virtualization of the baseband unit (BBU) pool, which is separated from BS radio head units (RRHs) and run as virtualized vetwork functions (VNFs) on commodity hardware in data center [74].

In contrast to other works, the spectrum pooling in our work serves two purposes. First, to avail the active BSs sufficient resource for the support of their own and inter-operator handed over users. The second objective is to motivate the MNOs to cooperate. This motivation is a form of incentive for lightly loaded MNOs to trade the spectrum for the users support. Moreover, to accommodate for a system of more than two participating MNOs, we model the spectrum band of each MNO as a combination of spectrum blocks based on carrier aggregation (CA) model introduced in LTE-Advanced. In 3GPP release 13, 18 pairs of inter-band non-contiguous carrier aggregation of LTE bands are feasible [75]. The forms of carrier aggregation is shown in Figure 3.1. For the use case of user equipment (UEs) in the multi-operator shared

spectrum in particular, the inter-band non-contiguous aggregation can be utilized. Therefore our inter-operator model is conceptualized on the cooperation among multiple MNOs where their carrier aggregated components are dynamically shared into inter-band non-contiguous aggregated private and shared spectrum sub-bands.

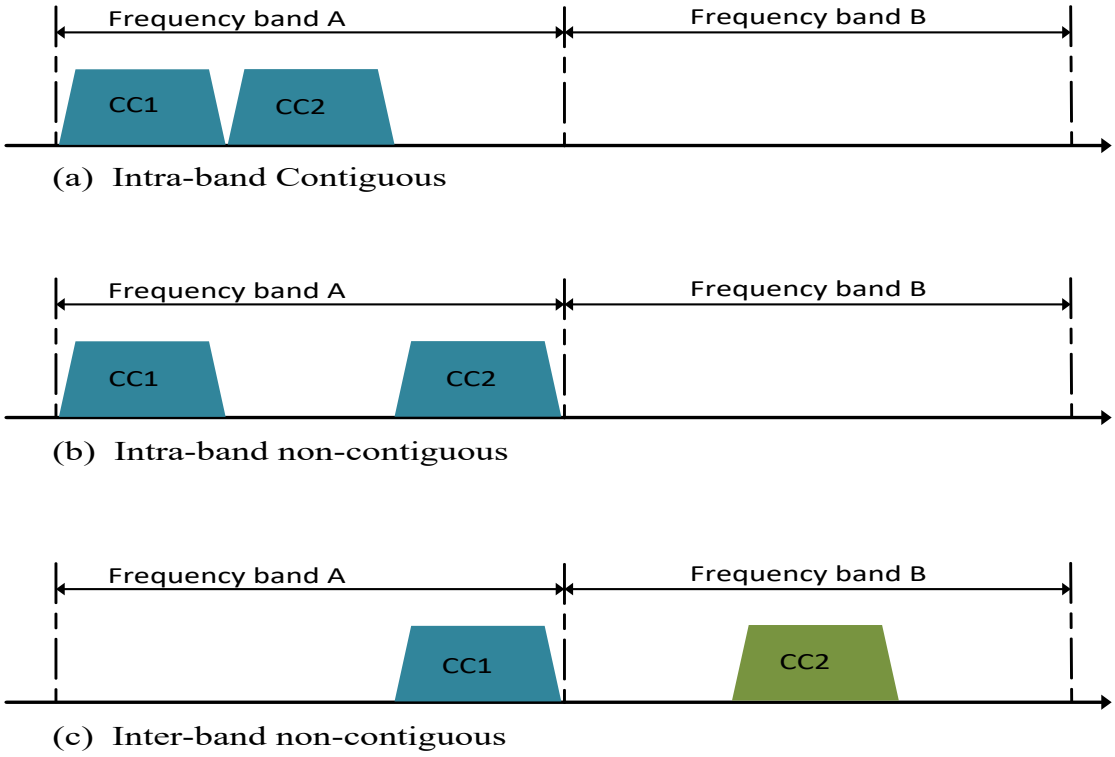

Figure 3.1: Illustration of the different ways of carriers can be aggregated.

In addition to energy gains achieved by switching off some BSs, energy is further saved by optimizing the transmit power of the active BSs. An optimization problem is formulated to obtain optimal beamforming vectors for power efficient transmissions. The optimal beamforming vectors also ensure a better signal reception by all UEs; thus the coverage is improved. Since the active BSs are to serve own and inter-operator UEs, both intra- and inter-operator beamforming vectors are optimized.

3.1.1 Network Function Virtualization in Cellular Networks

NFV is an emerging technology that enables network functions to run on industry standard hardware through software virtualization techniques. NFV Industry Specification Group of the European Telecommunications Standards Institute (ETSI) proposes NFV use cases for cellular networks [74]. The cases relevant to this study are Virtualized Radio Access Network (VRAN) and Virtualized Evolved Packet Core (VEPC). The VRAN is conceived from the C-RAN architecture and it provides virtualization of baseband unit (BBU) pool running in the data centers, and remotely located BSs' radio head units (RRHs). In addition, the NFV can enable multiple operators to share resources [76]. Fig 1 shows VRAN and VEPC of two cooperating MNOs sharing network resources and UEs access.

3.1.2 Inter-operator Spectrum Sharing

The aforementioned dramatic increase in mobile traffic has led to a demand surge in spectrum. Inter-operator spectrum sharing is one of the areas researchers, such as in [77–81], are looking into for supplementary spectrum access. A complementary spectrum access method, called co-primary shared access (CoPSS) has been introduced to enable multiple MNOs fully or partly share their licensed spectrum [82].

The technology of carrier aggregation introduced in LTE-Advanced is also a harbinger to inter-operator spectrum sharing. The scenarios of the use of carrier aggregation are intra-band contiguous, intra-band non-contiguous and inter-band non-contiguous. 3GPP release 13 allows inter-band non-contiguous carrier aggregation in 18 pairs of LTE bands [75]. For the use case of UEs in the multi-operator shared spectrum, inter-band non-contiguous aggregation can be harnessed. In this study, we conceptualize a cooperation between multiple MNOs where their carrier components are dynamically shared into inter-band non-contiguous aggregated private and shared spectrum sub-bands.

3.2 System and Power Model

3.2.1 Spectrum Access and Sharing Model

We consider M closely located identical cellular network systems based on LTE/LTE-Advanced network, with identical spectrum license, RANs, fronthaul and backhaul infrastructure. Each considered network belongs to a different operator. Following [83], network virtualization is applied to their backhauls for mutual cooperation. Thus, their backhauls lead to a joint Virtualized Evolved Packet Core (VEPC). We follow the concept of [80,81] in which a certain number of component carriers are available for inter-operator dynamic sharing. To forestall inter-band interference, we submit that each operator has its carrier components laid on inter-band non-contiguous pair bands to enable dynamic spectrum partitioning into private and shared sub-bands. Thus, interference between private and shared sub-bands is ignored in this model. For the M cooperating operators each having an original licensed spectrum allocation of B_m, the operator's shared spectrum sub-band is $\frac{\mathrm{B}_m}{M}$ and the private sub-band is $\mathrm{B}_{pr}^m = \mathrm{B}_m - \frac{\mathrm{B}_m}{M}$. This interprets that $\mathrm{B}_{sh} = \sum_{m=1}^{M} \frac{\mathrm{B}_m}{M}$ is the shared spectrum.

3.2.2 System Model

Each operator exclusively uses its private spectrum, while all cooperating operators can access and utilize the shared band. A UE's access is mutually exclusive. For example, a UE registered with operator 1 is served in B_{pr}^1 when served by its MNO. When served by other operator's BS, B_{sh} is utilized. Consequently, the received signal by a UE could be emanating from intra-RAN or inter-operator RAN. Inter-operator RAN backhaul routing is not discussed in this work as the focus is on energy efficiency.

Let the region of interest consist of A identical remote radio heads (RRHs) belonging to different operators. Each RRH is equipped with N antennas. The set of operator m's RRH is denoted by L_m such that $ML_m = A$. Our objective is to switch

off some BSs and transfer their loads to other active BSs. When sleep-mode strategy is applied, the active BSs set is denoted with V and the ones in sleep-mode is Z such that $V \cup Z = A$ and $V \cap Z = \emptyset$. The active RRH of operator m is represented with RRH_m^v, and the inactive with RRH_m^z. We assume close proximity of BSs in subsets V and Z for feasibility of load transfer. Since the UE's access is mutually exclusive, when operator m's UE is served by RRH_m^v private spectrum is utilized and intra-RAN precoding is considered. Inter-RAN precoding is imperative when UE associated with RRH_m^z is transferred to any other MNO in subset V.

The signal received by the k-th UE registered with a MNO m in the private spectrum is

$$y_k^{pr} = \sum_{v \in V_m} \mathbf{h}_{kv}^{\mathsf{H}} \mathbf{w}_{vk}^{pr} s_k + \sum_{i \neq k} \sum_{v \in V_m} \mathbf{h}_{kv}^{\mathsf{H}} \mathbf{w}_{vi}^{pr} s_i + \eta_k, \tag{3.1}$$

where the complex scalar s_k represents the k-th UE data symbol and $\mathbf{w}_{vk}^{pr} \in \mathbb{C}^N$ denotes the intra-operator beamforming vector at RRH_m^v for the k-th UE. $\mathbf{h}_{kv} \in \mathbb{C}^N$ is the channel vector responsible for CSI from RRH_m^v to UE k and $\eta_k \sim \mathbb{C}\mathcal{N}(0, \sigma^2)$ is the zero mean i.i.d complex-symmetric additive Gaussian noise at the receiver.

A UE k registered with $m \in Z$ when served by $m \in V$ in the shared spectrum receives a signal

$$y_k^{sh} = \sum_{v \in V_m} \mathbf{h}_{kv}^{\mathsf{H}} \mathbf{w}_{vk}^{sh} s_k + \sum_{i \neq k} \sum_{v \in V_m} \mathbf{h}_{kv}^{\mathsf{H}} \mathbf{w}_{vi}^{sh} s_i + \eta_k. \tag{3.2}$$

We note that the inter-operator beamforming vector $\mathbf{w}^{sh}$ is obtained from inter-RAN precoding and backhaul routing. Essentially, utilization of $\mathbf{w}_{vi}^{sh}$ and $\mathbf{w}_{vi}^{sh}$ is facilitated by the virtualized core pooling.

It follows from (3.1) and (3.2) that the corresponding signal-to-interference-plus-noise ratio (SINR) for the k-th UE served in the private and shared bandwidth,

respectively is given by

$$\Upsilon_k^{pr} = \frac{\left|\sum_{v\in V_m} \mathbf{h}_{kv}^{\mathsf{H}} \mathbf{w}_{vk}^{pr}\right|^2}{\sum_{i\neq k}\left|\sum_{v\in V_m} \mathbf{h}_{kv}^{\mathsf{H}} \mathbf{w}_{vi}^{pr}\right|^2 + \sigma_k^2}, \tag{3.3}$$

$$\Upsilon_k^{sh} = \frac{\left|\sum_{v\in V_m} \mathbf{h}_{kv}^{\mathsf{H}} \mathbf{w}_{vk}^{sh}\right|^2}{\sum_{i\neq k}\left|\sum_{v\in V_m} \mathbf{h}_{kv}^{\mathsf{H}} \mathbf{w}_{vi}^{sh}\right|^2 + \sigma_k^2}. \tag{3.4}$$

Therefore the achievable data rate of the UE served in the private bandwidth is

$$R_k^{pr} = B_{pr}\log_2(1+\Upsilon_k^{pr}). \tag{3.5}$$

Similarly, the UE served in the shared spectrum achieves a data rate of

$$R_k^{sh} = B_{sh}\log_2(1+\Upsilon_k^{sh}). \tag{3.6}$$

3.2.3 Power Consumption Model

To capture the joint energy consumption of all cooperating MNOs, we consider the power consumption at the RRHs of each operator. Also, power dissipation in the fronthaul links of each operator's Cloud Radio Access Network (C-RAN) following [65] is included, and power consumed in the backhaul [84].

3.2.3.1 RRH Power Consumption Model

An active RRH supports both UEs from its own MNO and others. Thus, the transmit power is not only a function intra-RAN precoding but also inter-RAN. Each active RRH has the following transmit power constraint

$$P_v = \sum_k \|\mathbf{w}_{vk}^{pr}\|_2^2 + \sum_k \|\mathbf{w}_{vk}^{sh}\|_2^2 \leq P_v^{max}. \tag{3.7}$$

We use the following empirical model for RRH power consumption:

$$P_v^{rrh} = \begin{cases} P_{v,static}^{rrh} + \frac{1}{\eta}\sum_k \|\mathbf{w}_{vk}^{pr}\|_2^2 + \frac{1}{\eta}\sum_k \|\mathbf{w}_{vk}^{sh}\|_2^2, & P_v \neq 0 \\ P_{v,sleep}^{rrh}, & P_v = 0, \end{cases} \tag{3.8}$$

where P_{static}^{rrh} is the static power consumption of RRH. η is the efficiency of the RF power, which is dependent on the number of transmitter antenna [85]. The RRH power consumed in sleep-mode is denoted by $P_{v,sleep}^{rrh}$.

3.2.3.2 Fronthaul Power Consumption Model

The virtualized RAN hinges on the legacy C-RAN architecture, which introduces physical location separation between RRHs and BBUs. The BBUs are combined into a centralized BBU pool at the data center, creating fronthaul links between the pool data center location and the multiple remote RRHs. The power consumed in each fronthaul link is expressed as

$$P_v^f = \begin{cases} P_{v,static}^f & P_v \neq 0 \\ P_{v,sleep}^f & P_v = 0, \end{cases} \tag{3.9}$$

where $P_{v,static}^f$ captures the power consumed in the fronthaul link while conveying RRH transmission and $P_{v,sleep}^f$ is the fronthaul link power when its BS is in sleep-mode.

3.2.3.3 Intra-operator Power Consumption Model

Here we present the outlook of power consumed by an individual MNO entity in the region of interest. When RRH v of operator m is active, its private bandwidth is used for its UEs, while the pooled shared bandwidth is used for the UEs belonging to other operators. The power consumed by a BS is evaluated as follows.

- When active:

$$P_m = \frac{1}{\eta} \sum_k \|\mathbf{w}_{vk}^{pr}\|_2^2 + \frac{1}{\eta} \sum_k \|\mathbf{w}_{vk}^{sh}\|_2^2 + P_{active} + P_m^{bh}, \tag{3.10}$$

  where $P_{active} = P_{v,static}^{rrh} + P_{v,static}^{f}$. The backhaul power consumption is represented by P_m^{bh}.

- When in sleep-mode:

$$P_m = P_{sleep} + P_m^{bh}, \tag{3.11}$$

  where $P_{sleep} = P_{v,sleep}^{rrh} + P_{v,sleep}^{f}$.

The total BS power consumption at any mode can be expressed as:

$$P_m = a_m \left[\frac{1}{\eta} \sum_k \|\mathbf{w}_{vk}^{pr}\|_2^2 + \frac{1}{\eta} \sum_k \|\mathbf{w}_{vk}^{sh}\|_2^2 + P_{active} - P_{sleep} \right] + \\ + P_{sleep} + P_m^{bh}, \tag{3.12}$$

where $a_m \in \{0, 1\}$ is a mode indicator. $a_m = 1$ when RRH v is transmitting, and 0 if otherwise.

3.2.3.4 Inter-Operator Joint Power Consumption Model

To obtain the overall power consumption of the colocated BSs of the cooperating operators in the region, we consider the summation of individual total BS power consumption. The inter-operator joint power consumption is

$$P_T = \sum_{m=1}^{M} P_m. \tag{3.13}$$

Since identical networks is assumed, $P_{static}^{rrh} = P_{v,static}^{rrh} = P_{z,static}^{rrh}$ is the static power of the transmitting RRH encompassed in the P_{active} formulation. Similarly, $P_{sleep}^{rrh} = P_{v,sleep}^{rrh} = P_{z,sleep}^{rrh}$ is the power of a BS in sleep-mode included in P_{sleep}. It can also be assumed that $\sum_{m=1}^{M} P_m^{bh} = M P_m^{bh}$ because identical networks are considered. How-

ever, $\sum_k \|\mathbf{w}_{vk}^{pr}\|_2^2 \neq \sum_k \|\mathbf{w}_{vk}^{sh}\|_2^2$ due to difference in intra- and inter-RAN precoding.

3.3 Transmit Power Optimization

3.3.1 Optimization Formulation

As mentioned in the previous section, intra- and inter-operator precoding implementation is required for the diverse UE support. Also, each BS transmit power depends on intra- and inter-RAN beamforming vectors. Seeking the optimal beamforming vectors can be formulated as an optimization problem. The objective is to find respective optimal inter- and intra-operator beamforming vectors which minimize inter-operator total power consumption while satisfying relevant network constraints.

$$\begin{aligned} \min_{\mathbf{w}_{vk}^{pr}, \mathbf{w}_{vk}^{sh}} \quad & \sum_{m=1}^{M} P_m(\mathbf{w}_{vk}^{pr}, \mathbf{w}_{vk}^{sh}) \\ \text{s.t.} \quad & R_k^{min} \leq R_k^{pr}, \forall k \in K_m \\ & R_k^{min} \leq R_k^{sh}, \forall k \in K_{-m} \\ & a_m \left[\sum_k \|\mathbf{w}_{vk}^{pr}\|_2^2 + \sum_k \|\mathbf{w}_{vk}^{sh}\|_2^2 \right] \leq P_v^{max} \\ & a_m \in \{0, 1\} \end{aligned} \quad . \tag{3.14}$$

The required minimum UE data rates in private and shared spectrum are stated in first and second constraints, respectively. The third constraint gives the RRH power limit. A registered UE of active operator m is $k \in K_m$. A UE belonging to another operator, but being served by MNO m is $k \in K_{-m}$. Thus, $K_m \cup K_{-m} = K$ represents the total number of UEs in the considered region.

3.3.2 Optimization Problem Reformulation

The power optimization problem (14) is non-convex due to non-convexity of R_k^{pr} and R_k^{sh}. Since the phases responsible for the complex components of $\mathbf{w}_{vk}^{pr}$ and $\mathbf{w}_{vk}^{sh}$ have no effect on the objective function [86] and the constraints, only their magnitude

parts are considered. Thus, we take the approach of [65,86] to obtain the convex form:

$$\|\mathrm{r}_k^{pr}\|_2 \leq \sqrt{1+1/(2^{R_k^{\min}/B_{pr}}-1)}\mathrm{Re}\{D_{kk}^{pr}\}, \forall k, \tag{3.15}$$

where $r_k^{pr} = [D_{k1}, D_{k2}, ..., D_{kk}, \sigma_k]^T$,
$D_{ki} = \sum_{v \in V_m} \mathbf{h}_{kv}^{\mathsf{H}} \mathbf{w}_{vi}^{pr}$.

Similarly, for the rate with the shared bandwidth,

$$\|\mathrm{r}_k^{sh}\|_2 \leq \sqrt{1+1/(2^{R_k^{\min}/B_{sh}}-1)}\mathrm{Re}\{D_{kk}^{sh}\}, \forall k, \tag{3.16}$$

where $r_k^{sh} = [D_{k1}, D_{k2}, ..., D_{kk}, \sigma_k]^T$,
$D_{ki} = \sum_{v \in V_{-m}} \mathbf{h}_{kv}^{\mathsf{H}} \mathbf{w}_{vi}^{sh}$.

Therefore, the optimization problem becomes:

$$\begin{aligned}
\min_{\mathbf{w}_{vk}^{pr}, \mathbf{w}_{vk}^{sh}} \quad & \sum_{m=1}^{M} P_m(\mathbf{w}_{vk}^{pr}, \mathbf{w}_{vk}^{sh}) \\
\text{s.t.} \quad & \|\mathrm{r}_k^{pr}\|_2 \leq \sqrt{1+1/(2^{R_k^{\min}/B_{pr}}-1)}\mathrm{Re}\{D_{kk}^{pr}\}, \forall k \in K_m \\
& \|\mathrm{r}_k^{sh}\|_2 \leq \sqrt{1+1/(2^{R_k^{\min}/B_{sh}}-1)}\mathrm{Re}\{D_{kk}^{sh}\}, \forall k \in K_{-m}. \\
& \sum_k \|\mathbf{w}_{vk}^{pr}\|_2^2 + \sum_k \|\mathbf{w}_{vk}^{sh}\|_2^2 \leq a_m P_v^{max} \\
& a_m \in \{0,1\}
\end{aligned} \tag{3.17}$$

Problem (17) can be solved by a typical convex optimization tool, such as CVX, since the problem is now a second order cone programming (SOCP) problem.

3.4 Sleep-mode with Efficient Beamformers and Spectrum-sharing (SEBS) Algorithms

In this section, we present the algorithms that make up the proposed BS energy-saving strategies. Based on equal spectrum sharing formulation among operators, the shared bandwidth is obtained from individually licensed bandwidth by equal spectrum pooling. The load transfer condition (LTC) algorithm is triggered by load threshold L_{th} to ascertain UE–BS channel quality relative to required minimum channel gain L_{th}. With the active set V and inactive set Z determined, the inter-and intra-operator beamforming vectors for optimal transmit power are obtained, as discussed in section V, while taking R_k and P_{max} into consideration. The power saving (PS) algorithm is invoked with an active BS set, optimal transmit power, and partitioned bandwidths. The scheme is shown in Figure 3.2.

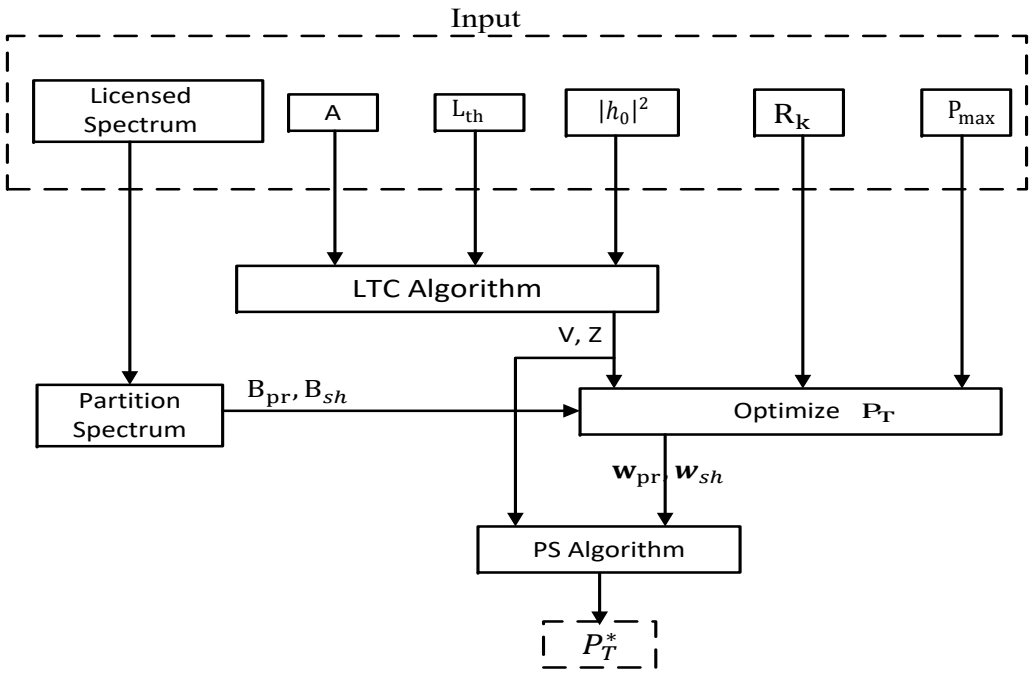

Figure 3.2: Outline of BS power-saving scheme

3.4.1 Load transfer Condition Algorithm

A load transfer condition is checked to ensure all UEs in the region have good channel quality with the associated active BSs. The steps are depicted in Algorithm 1. The combined UE traffic L_T in the region is evaluated and used to compute the required number of BS $|V|$ to support the UEs efficiently. The individual BS load limit L_{max} is taken into consideration in calculating $|V|$. Since the BSs are identical, L_{max} is assumed equal for all BSs. The BSs are sorted into set A, in descending order, according to their current associated traffic.

A $K \times |V|$ dimension matrix H is formed with K depicting the total number of UEs and $|V|$ the number of active BSs. The UE–BS channel quality is checked exhaustively between each UE and BS by comparing the channel power gain, $|h_{nk}|^2$, against the minimum required threshold, $|h_0|^2$. When the channel power gain declines below the threshold, the transmitter does not send any bits and outage ensues [87]. A good UE–BS channel state is indicated by 1 in matrix H, and 0 if otherwise. A l_0-norm of each row of matrix H is computed to yield a $K \times 1$ dimension vector U. A product of all array elements of vector U indicates the load transfer check result. $\prod_{k=1}^{K} U_k = 1$ means the load transfer criterion is satisfied. If otherwise, an additional BS is added to set V, and the steps are repeated until $V' \subset A \neq \emptyset$.

3.4.2 Power Savings with Equal Spectrum Pooling (ESP)

We propose an inter-operator power-saving algorithm where cooperative colocated BSs conserve joint power consumption by putting the least loaded BS(s) into sleep-mode. The algorithm is contingent on spectrum pooling, traffic demand, and channel quality. The steps are presented in Algorithm 2. Each of the M cooperating operators with B_m contributes equal among of bandwidth to the shared bandwidth pool. That is, individual MNO apportions

$$\mathrm{B}_{sh}^{m} = \frac{\mathrm{B}_m}{M} \tag{3.18}$$

Algorithm 1: Load Transfer Condition Algorithm

Form a sorted set A containing the BSs in descending order of their load weight.
Compute required active BSs number $|V| = ceil\left(\frac{L_T}{L_{max}}\right)$
Select the first $|V|$ BSs from set A to form set V
if $V' \subset A \neq \emptyset$ **then**
    Generate matrix H with dimension $K \times |V|$.
    for *each UE (row)* **do**
        Check the channel quality for each UE against all BSs (columns) and and note in matrix H, i.e $H_{k,v} = \begin{cases} 1 & |h_{vk}|^2 \geq |h_0|^2, v \in V, k \in K \\ 0 & \text{otherwise} \end{cases}$.
        Generate a vector $U = [U_1, U_2, \cdots, U_K]$, where $U_k = \|\sum_v H_{k,v}\|_0$.
        if $\prod_{k=1}^{K} U_k = 1$ **then**
            Load transfer condition is met
        end
        else Add an additional BS from subset $V' \subset A$, i.e $|V| = |V| + 1$, and go to Step 3.
        break
    end
end
else
    Load transfer condition is not met.
end

to the pool. The private sub-band is therefore

$$\mathrm{B}_{pr}^{m} = \mathrm{B}_{m} - \frac{\mathrm{B}_{m}}{M}. \tag{3.19}$$

This interprets that the total pooled shared bandwidth is $\mathrm{B}_{sh} = \sum_{m=1}^{M} \frac{\mathrm{B}_{m}}{M}$.

The scheme is triggered by traffic demand. A combined load threshold L_{th} is determined. When the joint traffic falls below the threshold, the load transfer criterion of Algorithm 1 is checked to ascertain a good channel quality for each UE. If the criterion is met, the optimal number of active BSs to support the load is computed. Each MNO partitions its private and shared bands according to (18) and (19), respectively. For efficient transmit power optimization problem (17) is solved to obtain optimal intra- and inter-operator beamforming vectors. The BSs not selected to be active have their UEs transferred to the active BSs, and therefore put into the sleep-mode. The active BSs support own UEs on their respective private bands, while they support the UEs of the inactive BSs using the pooled shared spectrum. On the other hand, if the load transfer criterion is not met, each MNO maintains the status quo, which is to continue to support its respective associated UEs in its full unpartitioned bandwidth.

3.5 Performance Evaluation

The energy saving performance of the proposed algorithm for multiple cooperating MNOs is evaluated. We consider 10 identical RRHs (representing 10 MNOs) closely deployed in the region area. Each RRH is equipped with 4 transmit antennas, and one antenna to every UE. The total bandwidth of each MNO's BS is 10 MHz. The P_{static}^{rrh}, and P_{static}^{f} is 22.5 Watts and 3.5 Watts, respectively. The power values of 12 Watts and 1.2 Watts are used for P_{sleep}^{rrh} and P_{sleep}^{f}. 10 Watt is taken as the maximum transmit power of the RRH. The η value of 0.36 is used. The noise power is $\sigma^2 = -94$ dBm. We adopt the pathloss model of $140 + 36.7\log_{10}(d)$ for distance d, as in [65,88],

Algorithm 2: Power-saving (PS) Algorithm

```
1  Evaluate the total load L_T in the region.
2  if L_T < L_th then
3  |  if Load transfer condition (Algorithm 1) is met then
4  |  |  Partition spectrum into B_pr^m and B_sh^m using (3.18) and (3.19).
5  |  |  Set a_m in (3.12) for all selected active BS set V to 1 and a_m = 0 for
   |  |   other BSs.
6  |  |  Obtain w_vk^pr and w_vk^sh by solving optimization problem (3.17).
7  |  |  BSs in set V support their registered UEs on their respective licensed
   |  |   spectrum B_pr^m, and other UEs on B_sh.
8  |  end
9  |  else
10 |  |  Go to Step 14.
11 |  end
12 end
13 else
14 |  Each BS continue to support its associated UEs on its respective licensed
   |   spectrum B_m.
15 end
```

to give attenuation between the RRH and UE. The required minimum UE data rate is 0.5 Mbps.

The default bandwidth partition into private and public sub-bands are as discussed in section IIA. The maximum RRH power is chosen to be 10 Watts. The default number of UEs is 10. A coverage radius of 200 meters is chosen for each RRH. The UEs are randomly distributed in the area.

To evaluate the relative performance of the proposed algorithm, we performed comparative simulations with respect to some network schemes. The schemes are defined below.

- *Sleep-mode with No Cooperation (SNC)*: In this scheme, the MNOs are not in cooperation. There is no shared bandwidth. Each MNO solely uses its license bandwidth. However, sleep mode is applied to a BS with no load. An idle BS only becomes active at an arrival of a user.

- *No Sleep-mode and No Cooperation (NSNC)*: Similar to SNC, there is no inter-

MNO cooperation. No spectrum sharing. In contrast to SNC, this scheme does not include putting the BS with no load on sleep-mode. All BSs are in active mode all the time.

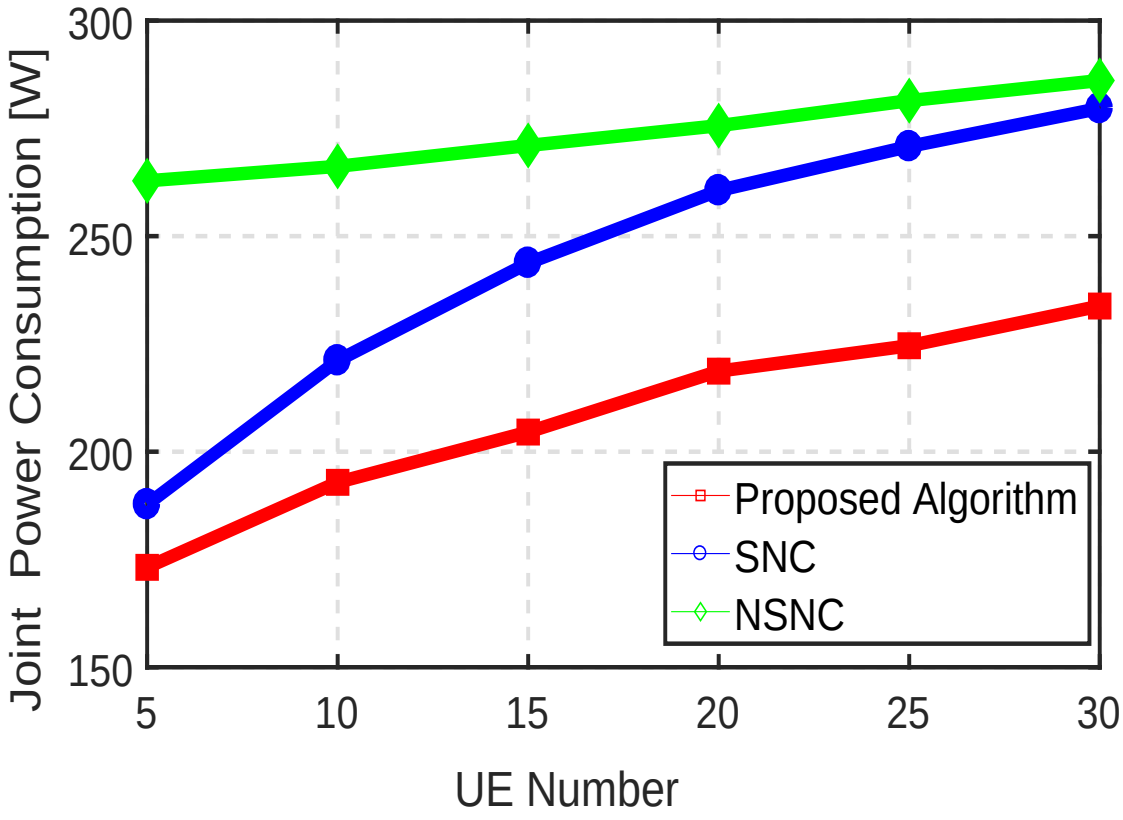

Figure 3.3: Total inter-operator Power Consumption versus number of UEs

Figure 3.3 shows the higher energy saving performance of the proposed algorithm over non-cooperative schemes. NSNC reminds of the waste of power without a sleep-mode strategy in cellular networks. The power consumption is evidently high even at low load. Despite the inclusion of BS sleep-mode in SNC, the proposed algorithm yields better energy efficiency due to inter-operator cooperation. Moreover, the rate of power dissipation in SNC as the traffic grows is higher than that of the proposed algorithm. The rate is due to relatively less number of BSs put to sleep mode in

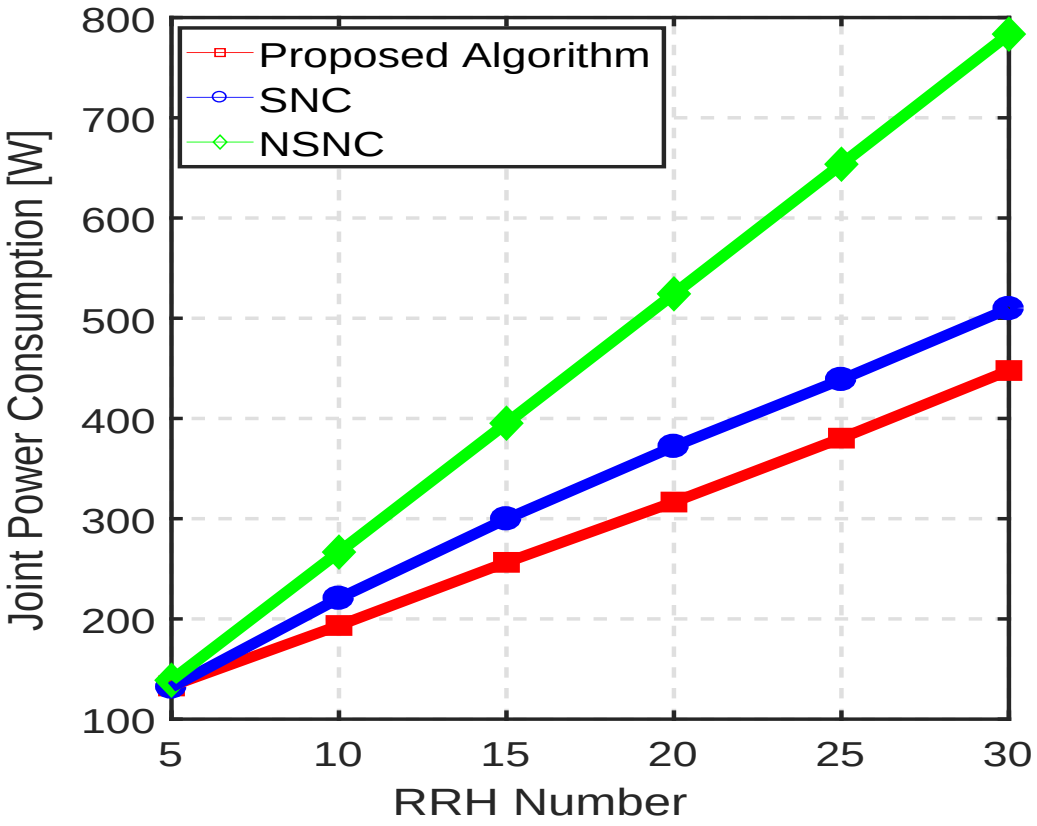

Figure 3.4: Total inter-operator Power Consumption versus number of RRHs

SNC as the traffic increases. More power is consumed in SNC than in the proposed algorithm as the traffic grows. The growth of traffic load as evident by the increase in the number of UEs causes the total power consumption to rise. At low load more BSs can be switched off, which implies the proposed algorithm gives the best performance at low load.

The algorithm is also evaluated with increases in each MNO BS number, with each belonging to different operators. The UE number is fixed at the default value. The objective is to assess the impact of the number of the cooperating MNOs on the amount of energy saved. The result of the evaluation is shown in Figure 3.4. With a few number of cooperating BSs, all the schemes seem to converge as all the BSs are engaged; none is lightly loaded to be put on sleep-mode. As expected, increasing

the number of BSs raises the total power consumption. However, increasing BS number creates an opportunity for load transfer from some BSs and hence the power consumption of the proposed algorithm is clearly lower than the other schemes. Even with sleep-mode strategy in SNC, the proposed algorithm is still more energy efficient.

3.6 Contribution

In this chapter. I proposed a BS energy-saving strategy, which reduces the power consumption of colocated BSs of multiple cooperative operators. Given the BS maximum transmit power limit and target data rate per UE, the scheme partitions each operator's spectrum into private and shared bands, finds the intra- and inter-RAN beamformers, and puts lightly loaded BSs into the sleep-mode. The scheme involves four steps. First, when the traffic in the region falls below a certain threshold, the number of BSs required to support the traffic optimally is determined. Second, the joint MNOs' BS power consumption is formed into a convex optimization problem, which is solved to obtain efficient intra- and inter-operator beamforming vectors. Third, the bandwidths are partitioned into private and shared bands, and all the shared bands are pooled together for joint access. Finally, the selected BSs support own UEs on their private bandwidths, and other UEs are served in the pooled shared bandwidth. The remaining BSs are put into the sleep-mode. The main contributions of this research are summarized as follows.

1. I proposed a sleep-mode and spectrum-sharing strategy to minimize the BS power consumption for cooperative multiple operators by leveraging on virtualized radio access and core networks. The scheme jointly involved inter-operator spectrum sharing, optimization of intra-and inter-RAN beamforming vectors, and sleep-mode control.

2. I presented multiple methods for dynamic inter-operator spectrum sharing cognizant of inter-RAN traffic support volume to motivate MNOs to cooperate

to achieve energy efficiency in their RANs. Based on the proposed spectrum–sharing formulations, inter-operator joint optimization problem was formulated to obtain power efficient intra- and inter-RAN beamforming vectors for supplementary energy gains and improved UE signal reception.

3. Equal spectrum pooling among cooperative MNOs was shown, through numerical simulations, to improve network energy efficiency significantly.

3.7 Summary

In the work of this chapter, I harnessed the VRAN use case to investigate energy saving feasibility for multiple cooperating MNOs. I have also utilized the concept of dynamic spectrum allocation and carrier aggregation to partition an operator's licensed bandwidth into private and shared bands in which UEs access is mutually exclusive. To achieve energy efficient inter-operator precoding, I formulated an optimization problem to obtain optimal intra-and inter-operator beamforming vectors while taking the maximum RRH power limit, and minimum data rates in private and shared spectrum into consideration. I proposed inter-operator load transfer condition algorithm, which considers the channel quality and current traffic load of the cooperating BSs. The second proposed algorithm, energy saving by sleep mode scheme in inter-operator power consumption shows a high reduction in the total power consumed in the RANs.

CHAPTER 4: PROPOSED SLEEP-MODE WITH SPECTRUM-POWER CONSUMPTION TRADING IN A MULTI-OPERATORS NETWORK

4.1 Introduction

Although power saving by equal spectrum pooling, presented in the previous chapter, proves to save energy considerably, some operators may not be willing to participate due to unequal gains from the scheme. All operators contribute equally to the shared spectrum. However, the MNOs with fewer subscribed UEs may benefit more from the joint energy-saving scheme. They are less likely to be selected (from the sorted BS list in Algorithm 1) to support the joint UEs, thus saving more power. The spectral efficiency of the active BSs may also negatively impact BSs with high traffic due to the release of a part of their spectrum for sharing.

Many authors [89–91] have validated the tradeoff between spectral and energy efficiencies in mobile cellular networks. The EE–SE (or EE–throughput) tradeoff has also been pertinent in resource sharing in cellular networks [92], specifically when a spectrum is shared [93]. In this section, we utilize this trade-off as an incentive to motivate MNOs to cooperate for the overall power consumption reduction.

4.2 System Model

4.2.1 Spectrum Access and Sharing Model

Consider a region consisting of M colocated cellular mobile networks. The considered MNOs are assumed to have equal spectrum bandwidths. Their RANs are also assumed identical for tractable power consumption modeling. Each network is owned by a different MNO. The multi-operators system leverages the virtualized RAN and core networks for spectrum sharing and inter-RAN UE support. Virtu-

alization of RAN and core networks facilitates shared network by allowing capacity and resources, such as spectrum, to be decoupled from the underlying physical resources. Therefore, the work relies on the assumption that the radio resources are virtualized. The system model is illustrated in Figure 4.1 with an example of 4 cooperating MNOs, each having a BS colocated with others in the region of interest. We follow the concept of [80, 81] in which a certain number of component carriers in the licensed spectrum are available for inter-operator dynamic sharing. To preclude inter-band interference, we propose that each operator has its carrier components composed of inter-band non-contiguous bands. Thus, the spectrum can be dynamically partitioned into private and shared sub-bands with non-contiguous bands. Since there is no contiguity between private and shared bandwidths, interference between private and shared sub-bands is inconsequential in this model, and thus ignored.

4.2.2 Transmission Model

Each operator has exclusive access to its private spectrum after partitioning. On the other hand, the pooled shared bands can be accessed and utilized by all the MNOs in cooperation. Only one BS supports one UE at a time, and the UE's access to the partitioned bandwidths is mutually exclusive. That is, a UE is served in its MNO's private spectrum when its BS is active. While the BS whose MNO the UE is registered with is in the sleep-mode, the UE is supported in the pooled shared spectrum by another active BS. Therefore, a signal received by a UE in the region could be emanated from intra-RAN or inter-operator RAN. In our work, an ideal case of full CSI sharing amongst MNOs, similar to other studies of multi-operators MIMO systems [94,95], is assumed. We, however, note that in practice full CSI sharing amongst multiple operators is not easily realizable due to the extra transmission overhead [96], the requirement of high data rate backhaul links [95], and the need to exchange the information at a reasonable time scale smaller than the channel coherence time [97, 98]. The signal processing theories of inter-operator RAN backhaul

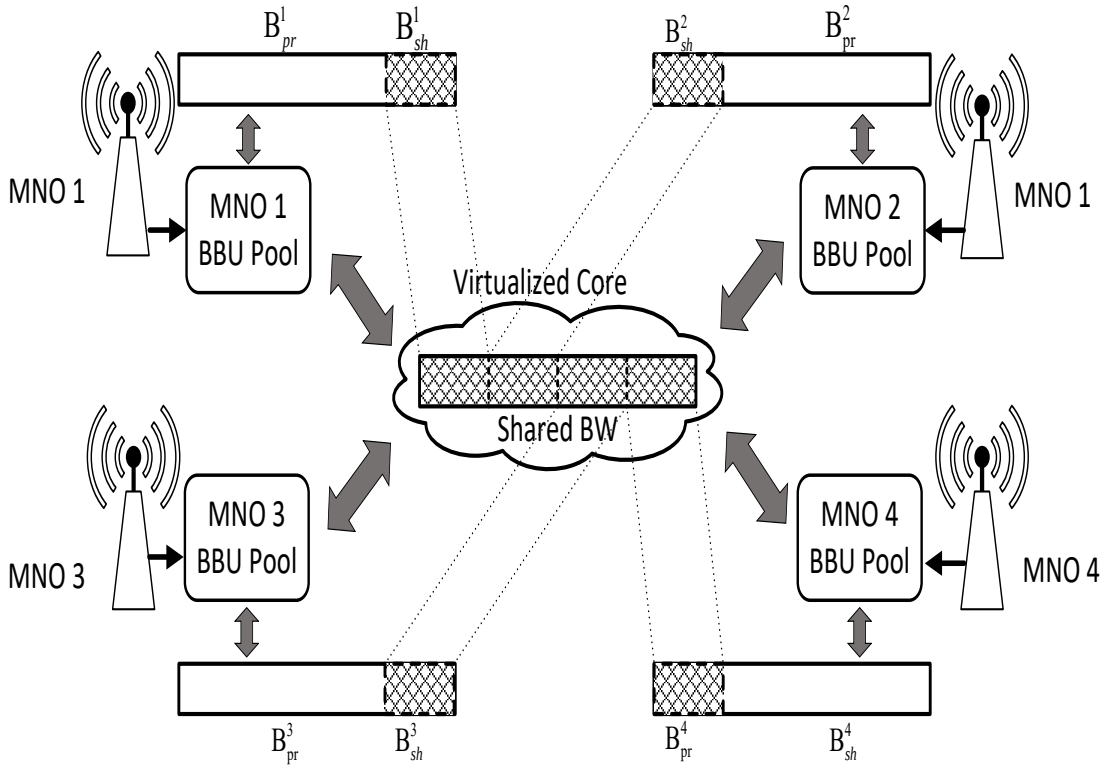

Figure 4.1: Virtualized RAN and Core of 4 cooperating operators with colocated BSs

routing are not discussed here as the energy efficiency is the focus of this work.

Let the region of interest consists of $|A|$ identical and colocated remote radio heads (RRHs), each belonging to a different operator. Each RRH is equipped with N antennas. Our objective is to minimize joint operators' power consumption by switching off some BSs with low traffic and transferring their UEs to other active BSs. When the sleep-mode strategy is applied, the set of active RRHs is denoted by V and the ones in the sleep-mode are depicted by Z, such that $V \cup Z = A$ and $V \cap Z = \emptyset$. The cardinality of set V is denoted by $|V|$. Since a UE can either be supported in the private bandwidth B_{pr}^{m} or shared bandwidth B_{sh}^{m}, both intra- and inter-RAN precoding are considered.

4.2.3 Power Model and Optimization Problem formulation

The network components and parameters of Chapter 3 are adopted here. Therefore, the total BS power consumption at any mode can be expressed as:

$$P_m = a_m \left[\frac{1}{\eta} \sum_k \|\mathbf{w}_{vk}^{pr}\|_2^2 + \frac{1}{\eta} \sum_k \|\mathbf{w}_{vk}^{sh}\|_2^2 + P_{active} - P_{sleep} \right] + \\ + P_{sleep} + P_m^{bh}, \tag{4.1}$$

where $a_m \in \{0, 1\}$ is a mode indicator. $a_m = 1$ when RRH v is transmitting, and 0 if otherwise. The inter-operator joint power consumption is $P_T = \sum_{m=1}^{M} P_m$.

Similarly, the transmission power optimization problem, consisting of intra-and inter-RAN beamforming vectors, can be written as:

$$\begin{aligned} \min_{\mathbf{w}_{vk}^{pr}, \mathbf{w}_{vk}^{sh}} \quad & \sum_{m=1}^{M} P_m(\mathbf{w}_{vk}^{pr}, \mathbf{w}_{vk}^{sh}) \\ \text{s.t.} \quad & \|\mathrm{r}_k^{pr}\|_2 \leq \sqrt{1 + 1/(2^{R_k^{\min}/B_{pr}} - 1)}\mathrm{Re}\{D_{kk}^{pr}\}, \forall k \in K_m \\ & \|\mathrm{r}_k^{sh}\|_2 \leq \sqrt{1 + 1/(2^{R_k^{\min}/B_{sh}} - 1)}\mathrm{Re}\{D_{kk}^{sh}\}, \forall k \in K_{-m}. \\ & \sum_k \|\mathbf{w}_{vk}^{pr}\|_2^2 + \sum_k \|\mathbf{w}_{vk}^{sh}\|_2^2 \leq a_m P_v^{max} \\ & a_m \in \{0, 1\} \end{aligned} \tag{4.2}$$

4.2.4 Spectrum Trading Sharing Model

In this section, we take advantage of the trade-off between spectrum and power consumption to create an incentive for cooperation for spectrum sharing, and ultimately power saving. We propose dynamic partitioning of B_{pr}^m and B_{sh}^m based on symmetric fair spectrum–power consumption trading model. The amount of traffic belonging to a BS embarking on the sleep-mode is used as the *numeraire* for spectrum trading with the supporting BSs. The operators are categorized into two: spectrum donors are the *Suppliers*, while those accepting spectrum in exchange for UE support

are the *Demanders*.

Let the fraction of the individually licensed spectrum allocated to the shared spectrum based on trading for each operator be denoted by δ_m. Since the number of associated UEs per operator may not be equal, δ_m may differ across MNOs. For clarity, δ_m is further denoted as δ_m^d for a demander and δ_m^s for a supplier.

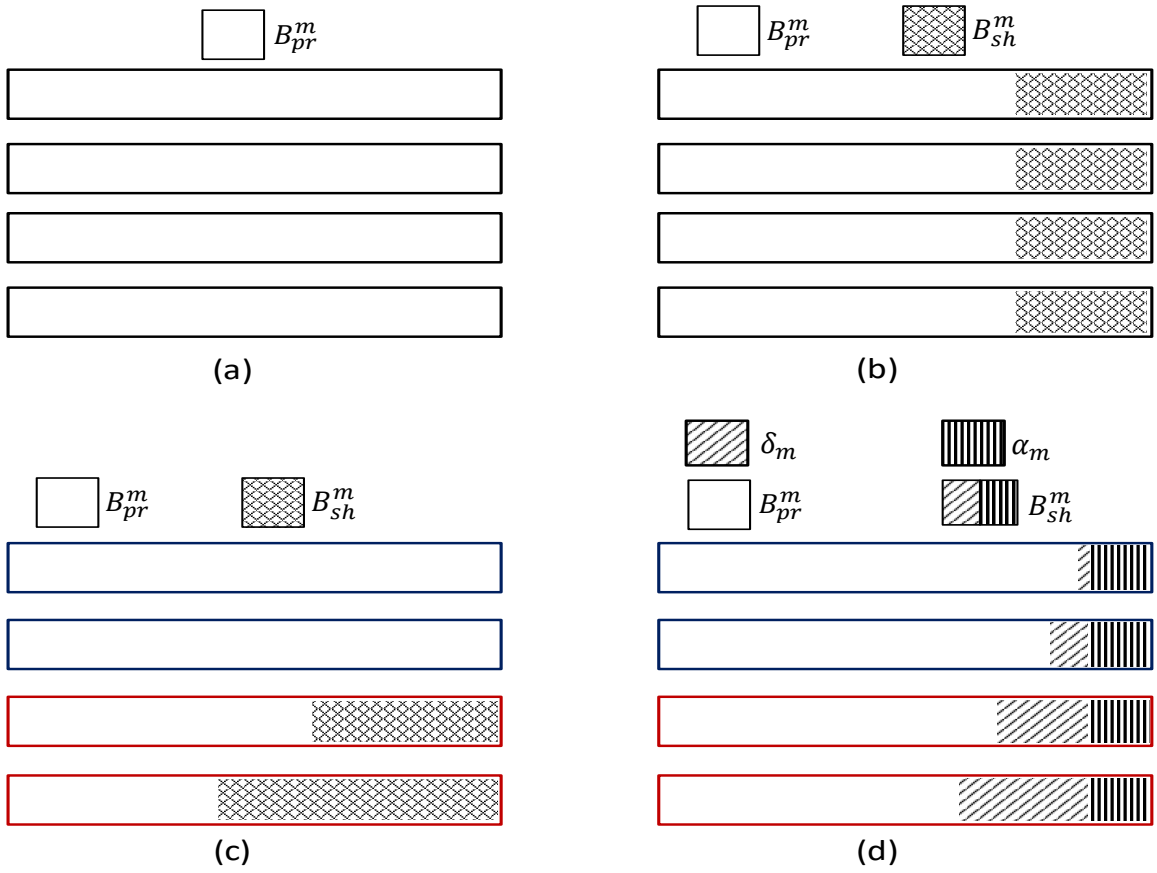

Figure 4.2: Spectrum Sharing partitioning scenarios. (a) No spectrum sharing; each MNO exclusively uses its licensed spectrum. (b) All operators contribute equally to the shared bandwidth. In this case, α_m is fixed. The spectrum pooling scheme is used in SESP where $\alpha_m = \frac{B_m}{M}$. (c) Spectrum sharing here is based on Spectrum–Power consumption Trading as in SSPT. Bandwidths belonging to demanders are in red, and the suppliers are delineated in blue. Here, $\alpha_m = 0$. (d) This is also a SSPT scheme with $\alpha_m \neq 0$.

BS power consumption, however, does not solely depend on the number of UEs

served. For example, power is also consumed by an active fronthaul regardless of the number UEs. Moreover, the BS power consumption in varying traffic is not the same for all BS types. The impact of the number of UEs on BS power consumption is higher in a macro BS than a micro BS, even much higher than in pico and femto [85]. We represent the factors that cannot be captured by δ_m by a flexible constant α_m.

Therefore, $\alpha_m \geq 0$ represents the agreed spectrum fraction each operator m contributes to the shared spectrum regardless of current and future traffic condition. This amount is equal for all operators, i.e $\alpha_1 = \alpha_2 = ... = \alpha_M$. In this work, we take the total pooled shared bandwidth to be $B_{sh} = B_m$, and hence $B_{sh} \geq M\alpha_m$.

For demander $m \in V$, the inherited traffic load to be supported is L_m^d such that the combined load of all demanders is depicted as $\sum_{m=1}^{Z} L_m^d$. The spectrum numeraire of demander m is

$$\delta_m^d = \frac{L_m^d}{\sum_{m=1}^{Z} L_m^d} \left(B_{sh} - M\alpha_m\right). \tag{4.3}$$

The private bandwith of demander m becomes

$$B_{pr}^{m,d} = (B_m - \alpha_m) + \delta_m^d, \tag{4.4}$$

and its total contribution to the pooled shared bandwidth is simply

$$B_{sh}^{m,d} = B_m - B_{pr}^{m,d}. \tag{4.5}$$

For each supplier, the spectrum numeraire in exchange for UE support is

$$\delta_m^s = \frac{L_m^s}{\sum_{m=1}^{Z} L_m^s} \left(B_{sh} - M\alpha_m\right), \tag{4.6}$$

where L_m^s is the number of other operators $m \in Z$ with their traffic supported by the

active BSs. Therefore, its contribution to the pooled shared bandwidth is

$$B_{sh}^{m,s} = \alpha_m + \delta_m^s, \tag{4.7}$$

and the private bandwidth of supplier m can be simply expressed as

$$B_{pr}^{m,s} = B_m - B_{sh}^{m,s}. \tag{4.8}$$

The illustration of various spectrum sharing and partitioning methods is shown in Figure 4.2. The ESP is described in Figure 4.2(b), while SPT methods are depicted in Figure 4.2(c) and (d). The simplified steps of the ESP and SPT spectrum sharing approaches are shown in Figure 4.3.

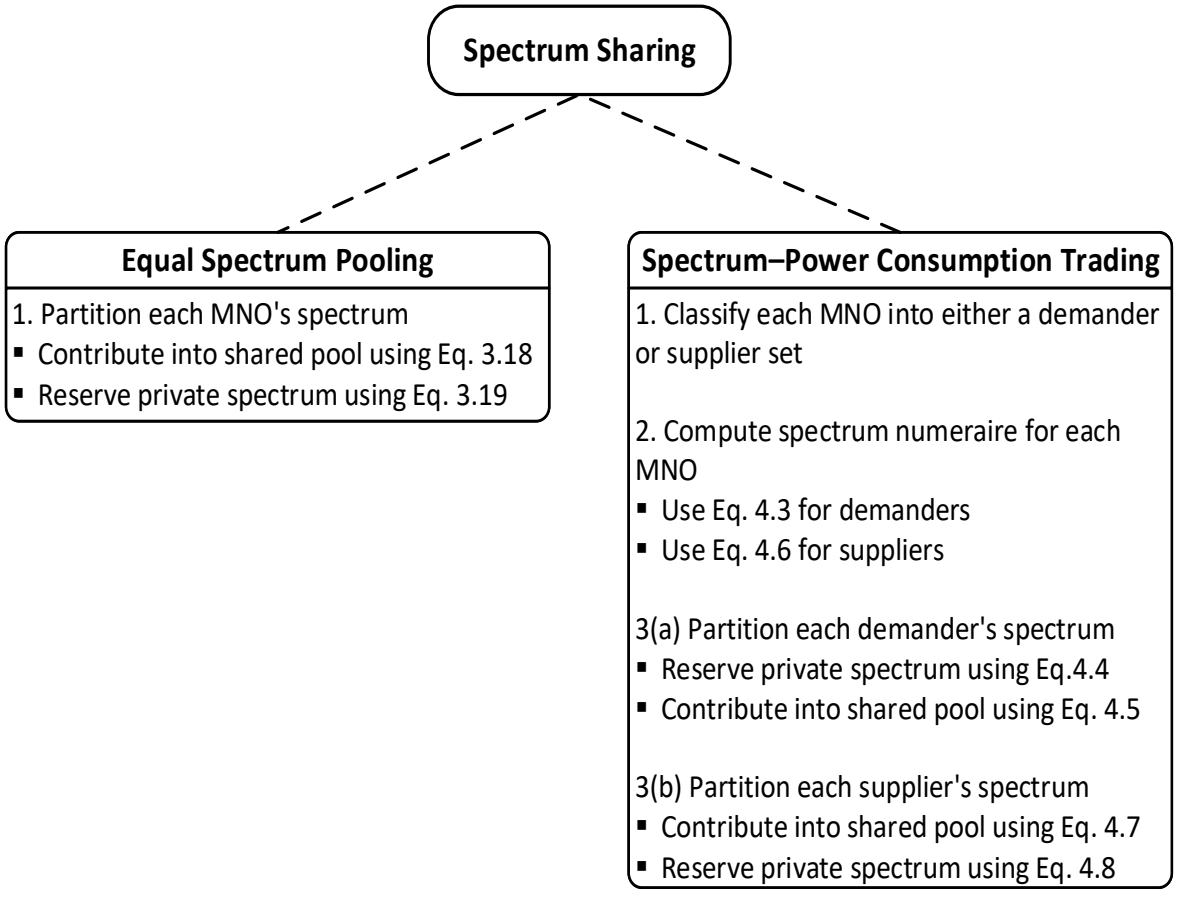

Figure 4.3: Simplified illustration of the spectrum sharing algorithms.

4.2.5 Joint Power Savings by Spectrum–Power Consumption Trading

Here, we present the energy-saving strategy by BS sleep-mode based on spectrum trading for UE support. The scheme is illustrated in Algorithm 3 and in Figure 4.4. While the strategy can be extended to be triggered by a joint load threshold, the algorithm allows MNOs to make desired trade side requests. Both sets of demanders and suppliers are sorted in descending order of their associated traffic. The set of potential active BSs, V, is initialized by the demanders and the UE–BS channel quality is checked by Steps 5 – 11 of Algorithm 1. A bad overall channel condition outcome leads to rejection of the first supplier's request and inclusion in set V. The step is repeated until a good channel quality is attained for all UEs.

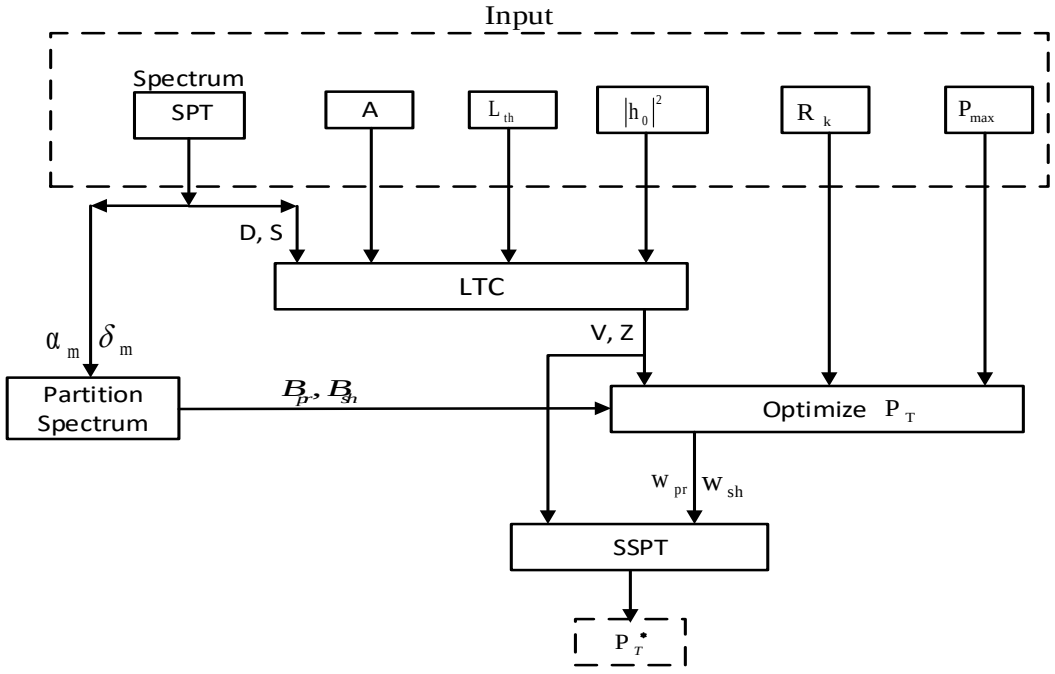

Figure 4.4: Outline of BS power-saving Scheme by Spectrum–Power Consumption Trading

4.3 Performance Evaluation

In this section, we evaluate the performance of the proposed schemes. We show SEBS strategies can achieve power savings and sum rate increase in multiple operator

Algorithm 3: Sleep-mode with Spectrum–Power consumption Trading (SSPT) Algorithm

1 Group the BSs based on requests: set of spectrum demanders and set of suppliers.
2 Sort the 2 BS sets in descending order of their load weight; D= sorted spectrum demanders and S = spectrum suppliers.
3 Concatenate sets D and S to form set A, i.e $A = \{D, S\}$.
4 Form set $V \subset A$ and initialize V with subset D.
5 **while** $V' \subset A \neq \emptyset$ **do**
6 Check for all UE–BS channel quality using **Algorithm 1** Steps 5–11.
7 **if** *channel condition is satisfied* **then**
8 For $\forall m \in V$ apportion $B_{pr}^{m} = B_{pr}^{m,d}$ using (4.4) and $B_{sh}^{m} = B_{sh}^{m,d}$ using (4.5). For $\forall m \in V'$ partition $B_{sh}^{m} = B_{sh}^{m,s}$ using (4.7) and $B_{pr}^{m} = B_{pr}^{m,s}$ using (4.8).
9 Solve problem (4.2) to obtain the optimal $\mathbf{w}_{vk}^{pr}$ and $\mathbf{w}_{vk}^{sh}$.
10 Each BS $\forall m \in V$ supports its own UEs in B_{pr}^{m} and offloaded UEs from any V' on B_{sh}, and put $V' \subset A$ into sleep-mode.
11 **end**
12 **else**
13 Include the most loaded BS from suppliers, i.e $|V| = |V| + 1$ from $V' \subset A$, and go to Step 6.
14 **end**
15 **end**
16 Each BS supports own UEs on its respective licensed spectrum B_m.

Table 4.1: Simulation Paramters

Parameter	**Value**
Coverage radius	200 meters
Original MNO bandwidth	10 MHZ [88]
Minimum achievable data rate	0.5 Mbps
Pathloss model	$140 + 36.7\log_{10}(d)$ [65]
Noise power, σ^2	-94 dBm
RRH maximum transmission power	10 W
RRH static power, P^{rrh}_{static}	22.5 W
RRH sleep-mode power, P^{rrh}_{sleep}	12 W
Fronthaul link static power, P^{f}_{static}	3.5 W
Fronthaul link sleep-mode power, P^{f}_{sleep}	1.2 W

scenarios. In the simulation, the following are utilized as the default scenarios and parameters. 10 identical colocated RRHs are deployed, each belonging to a separate MNO and a uniform distribution of UEs in the region of interest of coverage radius of 200 meters. Each RRH is equipped with 4 transmit antennas, and one antenna to every UE. The total licensed bandwidth of each MNO is 10 MHz. P^{rrh}_{static} and P^{f}_{static} are 22.5 Watts and 3.5 Watts, respectively. The power consumption values used for P^{rrh}_{sleep} and P^{f}_{sleep} are 12 Watts and 1.2 Watts, respectively. The maximum transmission power of the RRH is 10 Watts and η value of 0.36 is adopted. The noise power is $\sigma^2 = -94$ dBm. We adopt the pathloss model of $140 + 36.7\log_{10}(d)$ for distance d, as in [65], to give attenuation between the RRH and UE. Other simulation parameters are presented in Table II. For SSPT, the number of intitial spectrum demanders and suppliers are $|D| \sim \left(\left[ceil\left(\frac{|A|}{4}\right), ceil\left(\frac{3|A|}{4}\right)\right]\right)$ and $|S| = |A| - |D|$, respectively. 200 simulation runs are performed for all schemes.

4.3.1 Benchmarks

To evaluate the relative energy efficiency performance of the proposed schemes, we performed comparative simulations of some network BSs operation in cooperative and non-cooperative modes. The comparative schemes are defined below.

- *Sleep-mode with No Cooperation (SNC)*: In this scheme, BS sleep-mode strat-

egy is employed. BSs with no loads are switched off, but powered on upon arrivals of UEs. However, the MNOs are not in cooperation. Thus, there is no inter-operator bandwidth sharing. Each operator solely uses its total licensed bandwidth exclusively for its UEs.

- *No Sleep-mode and No Cooperation (NSNC)*: Unlike the SNC, this scheme does not incorporate the sleep-mode strategy. All BSs are in the active mode all the time. Also, there is no inter-operator cooperation in this scheme, hence no spectrum sharing.

4.3.2 Performance Evaluation with Traffic Variation

The performance of SEBS schemes is evaluated in various traffic scenarios. The schemes are compared with the benchmarks mentioned above in the simulation. The number of RRHs is held constant at 10, while the number of UEs in the region is varied.

In the SESP strategy, the cooperating MNOs partition their licensed spectra into shared bands according to (18), and private bands using (19). For fair inter-operator spectrum sharing by the SEPT scheme, $\alpha_m = 0$ is used to model the BS's contribution to the shared bandwidth for power consumption analysis since identical BSs are assumed. Therefore, spectrum pooling is only dependent on the number of UEs traded.

Figure 4.5 shows the higher energy saving performance of the proposed schemes over non-cooperative benchmark schemes. The inter-operator joint power consumption increases with traffic in all schemes. The non-cooperative strategies, NSNC and SNC, exhibit smaller EE than the proposed cooperative schemes. NSNC confirms the energy inefficiency of BS operation without a sleep-mode strategy in cellular networks. The power consumption is undoubtedly high even at low traffic loads. Despite the inclusion of the BS sleep-mode in SNC, the proposed algorithms yield better EE due to

inter-operator cooperation. Moreover, SNC yields a higher rate of power dissipation as the traffic grows than the proposed schemes. The rate is due to the relatively less number of BSs put into the sleep-mode in SNC as the traffic increases. SSPT is more power efficient than the other proposed cooperative scheme, SESP. Unlike in SESP where the active BSs are constrained to support UEs with the fixed pre-apportioned shared bandwidth, SSPT allows some spectrum slices from the inactive BSs to be used to support UEs. More availability of spectrum to active BSs results in the edge in EE over SESP.

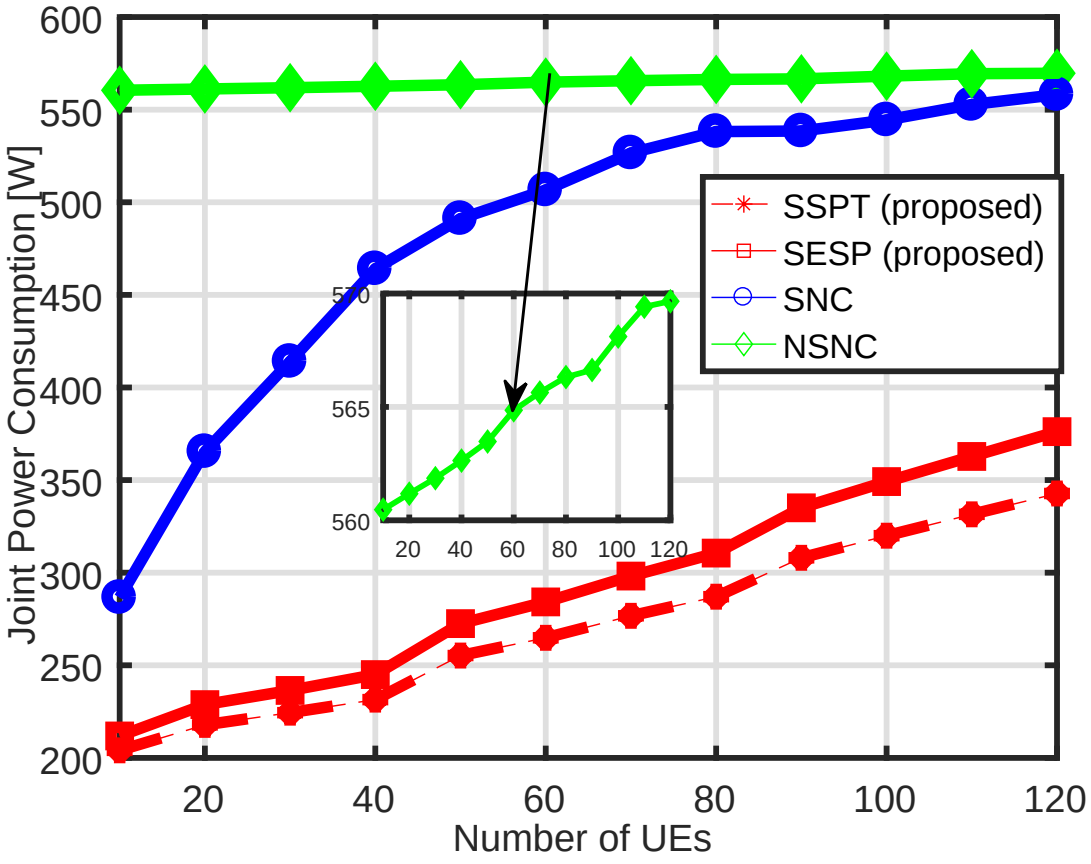

Figure 4.5: Total inter-operator power consumption versus number of UEs

4.3.3 Impact of the Number of Cooperative MNOs

The number of participating MNOs is modeled here with the variation in the number of RRHs since each RRH may belong to a separate operator. The UE number is fixed at the default value of 10. Shared bandwidth pooling is facilitated by $\alpha_m = 0$ for the fair spectrum-sharing of the SSPT scheme. The objective is to assess the impact of the number of cooperating MNOs on the amount of energy saved. The simulation of the performance is shown in Figure 4.6.

With a few RRHs, all the schemes seem to converge as all the RRHs are actively engaged; none is lightly loaded to be put into the sleep-mode. Increasing the number of RRHs raises the total power consumption in all the schemes mainly from P^{rrh}_{static} for the active RRHs, and $P^{rrh}_{v,sleep}$ for RRHs on the sleep-mode. On the other hand, the increase in the number of MNOs creates an opportunity for traffic handover from some BSs. Thus, the power consumption is lower in sleep-mode strategies than NSNC. However, with sleep-mode strategy in SNC, the proposed schemes (SESP and SSPT) are still more power efficient. This validates that inter-operator cooperation is more energy efficient even with the increase in the number of participants. Similar to the previous analysis, fair spectrum-sharing of the SSPT scheme provides better energy-saving performance than SESP.

4.3.4 Impact of Channel Power Gain

The BS–UE channel quality evaluation of the LTC (Algorithm 1) is the pivot of the two proposed algorithms. Therefore, this evaluation tests the performance of SEBS in different channel states conditioned on the threshold $|h_0|^2$.

Figure 4.7 shows the power consumption incurred by the SEBS schemes for different channel conditions while the numbers of UEs and RRHs are held constant. Similar to the previous results, SSPT yields higher EE than SESP. However, the power consumed, in both schemes, is not linear as the channel threshold increases. At lower

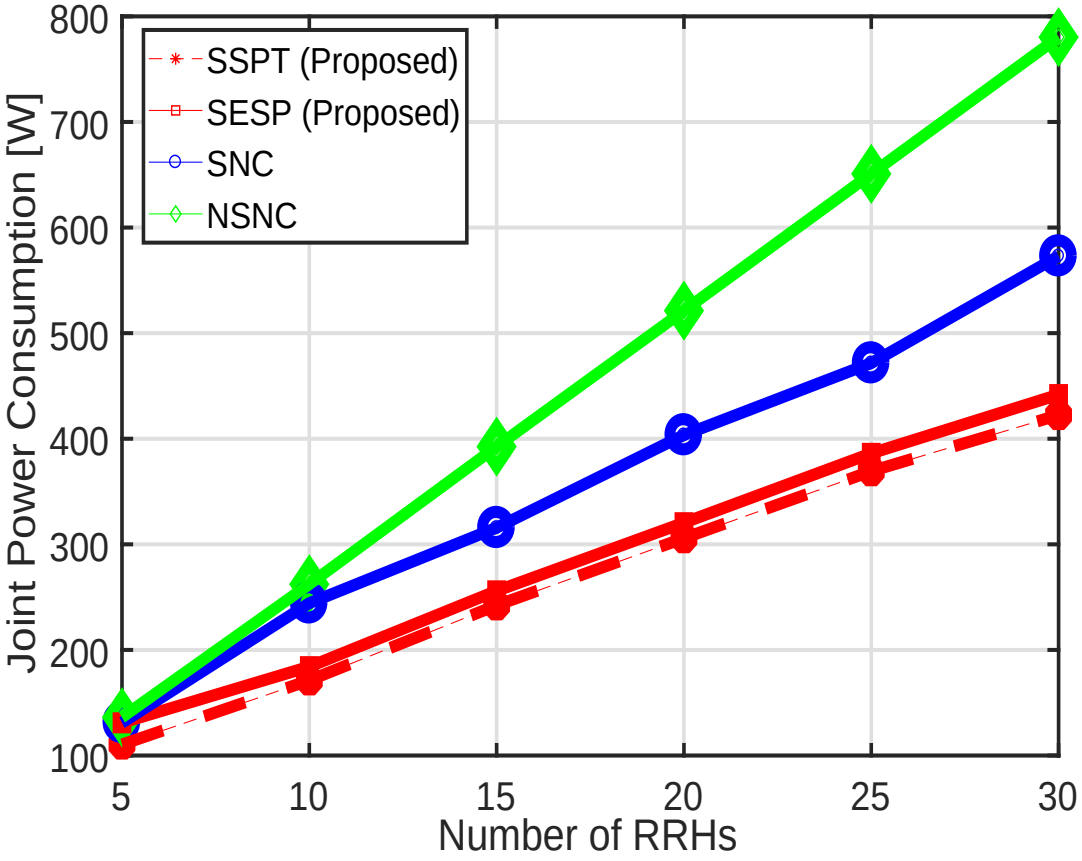

Figure 4.6: Total inter-operator power consumption versus number of RRHs

thresholds, fewer BSs are engaged, but heavily laden. This results in increasing the dynamic power part, and the evident initial surge in power consumption. The system energy consumed drops with the channel threshold increase to some optimal point. However, as the threshold is tightened, it becomes more difficult for more BSs to satisfy LTC, thus leading to an increase in more active BSs. With more BSs in operation, the aggregate of more P^{rrh}_{static} and $P^{f}_{v,static}$ causes the system energy consumption to rise.

4.3.5 Impacts of Sharing Ratios on Sum Rate

Spectrum Efficiency (SE) - EE tradeoff is one of the fundamental tradeoffs in wireless networks [99]. In this subsection, we validate the performance of SEBS on the sum rate of all UEs of all the colocated BSs. For a critical evaluation, we examine

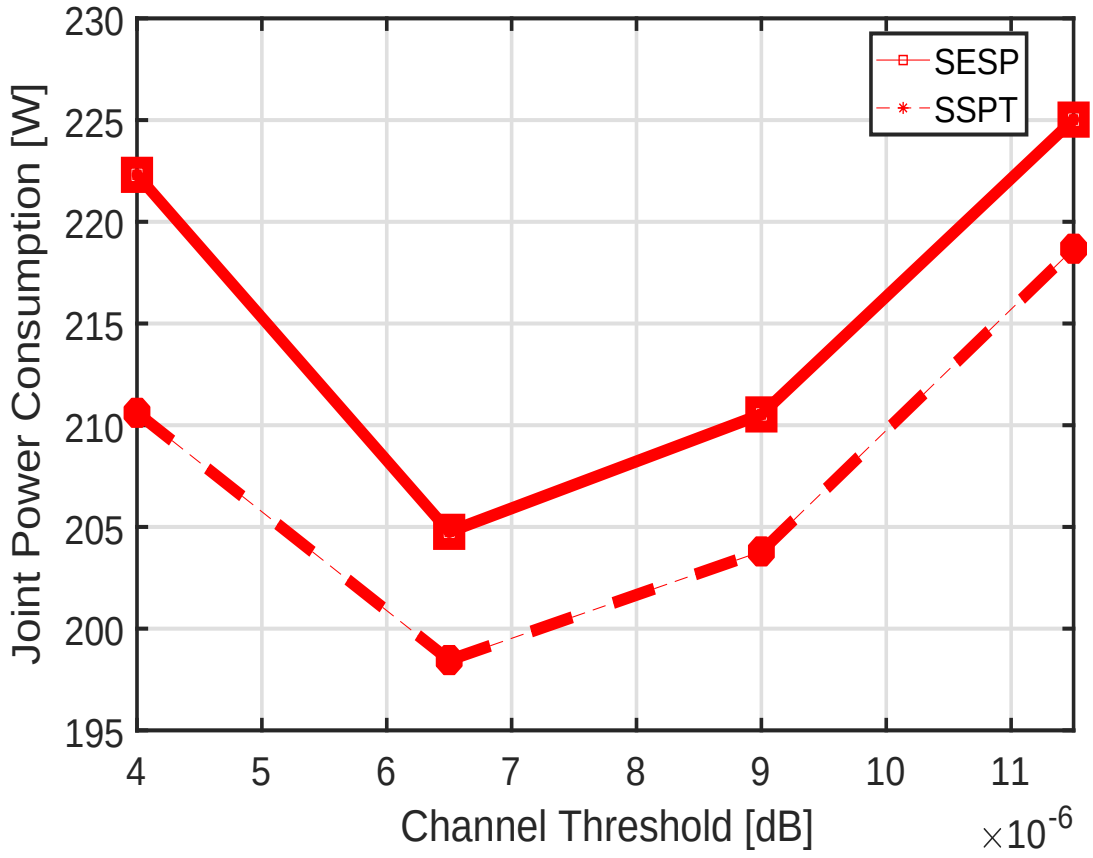

Figure 4.7: Channel quality threshold versus power consumption

various spectrum sharing scenarios depicted in Figure 4.5. This avails the opportunity to examine the impact of different bandwidth sizes on the SE–EE relationship.

Different shared bandwidth partitioning ratios characterize the various SSPT curves in Figure 4.8. Strictly equal spectrum sharing is represented by SESP, and SNC is the non-cooperative scheme. All schemes involve the sleep-mode strategy for unbiased analysis. SNC clearly results in a relatively much lower sum rate.

A further probe into the result shows that not all sharing methods outperform the fixed spectrum sharing method as per spectral efficiency. This also provides an insight on the impact of support of inherited UEs on the SE. Specifically, at $\alpha_m = 0$ and $\alpha_m = 0.5B_m/M$, more bandwidth is available to support the inherited UEs. This ultimately yields a higher sum rate than the fixed equal spectrum sharing method,

SESP. However, a lower sum rate is achieved at $\alpha_m = 1.5B_m/M$. Here, inherited UEs are supported with less bandwidth, thus culminating in a lower sum rate.

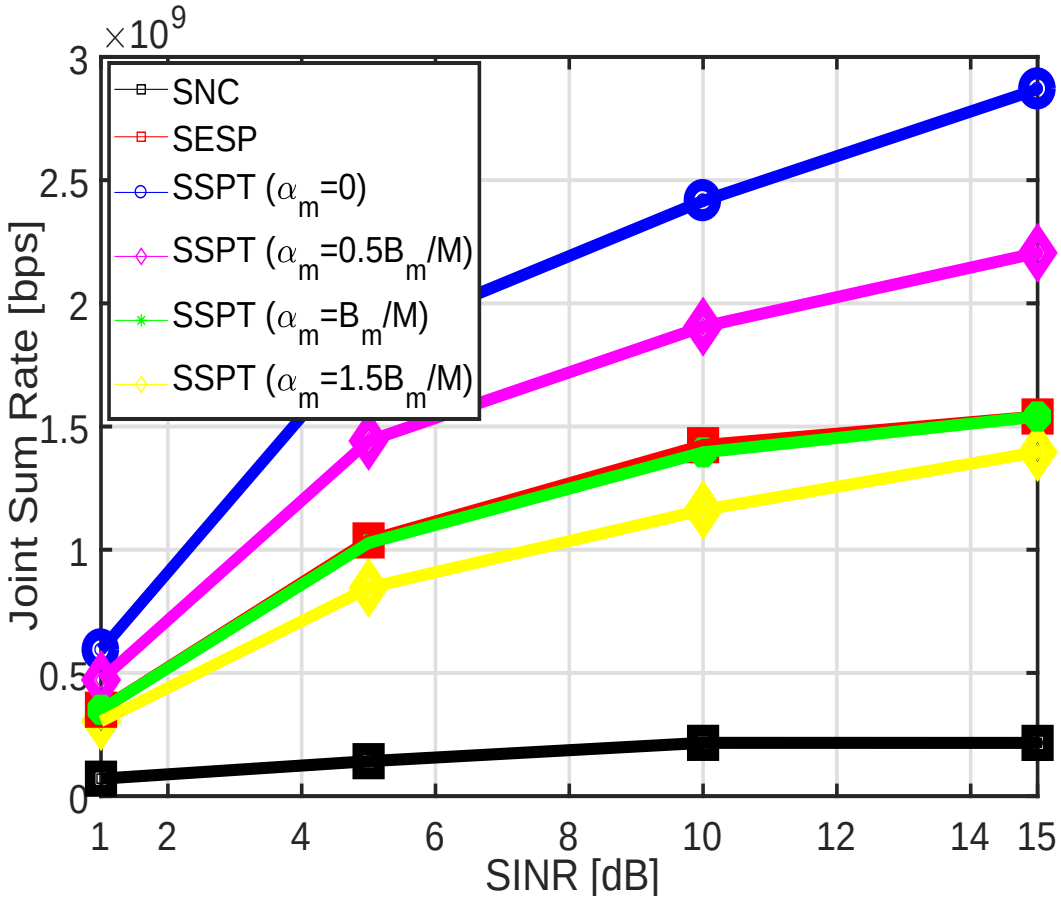

Figure 4.8: Joint sum rate of various energy saving schemes

4.3.6 Contributions

We propose the SEBS strategy, which reduces the power consumption of colocated BSs of multiple cooperative operators. Given the BS maximum transmission power limit and target data rate per UE, SEBS partitions each operator's spectrum into private and shared bands, finds the intra- and inter-RAN beamformers, and puts lightly loaded BSs into the sleep-mode. The SEBS scheme involves four steps. First, when the traffic in the region falls below a certain threshold, the number of BSs required to support the traffic is optimally determined. Second, the joint MNOs'

BS power consumption is formulated into a convex optimization problem, which is solved to obtain efficient intra- and inter-operator beamforming vectors. Third, the bandwidths are partitioned into private and shared bands, and all the shared bands are pooled together for joint access. Finally, the selected BSs support their own UEs on their private bandwidths, and other UEs are served in the pooled shared bandwidth. The remaining BSs are put into the sleep-mode. To motivate MNOs to cooperate, SEBS is extended to accommodate the spectrum–power consumption trading, where an operator can trade part of its licensed spectrum for serving its UEs. The main contributions of this paper are summarized as follows.

1. I proposed a sleep-mode and spectrum-sharing strategy, which traded spectrum for user support,to minimize the BS power consumption for cooperative multiple operators by leveraging on virtualized radio access and core networks.

2. I presented multiple methods for dynamic inter-operator spectrum sharing cognizant of inter-RAN traffic support volume to motivate MNOs to cooperate to achieve energy efficiency in their RANs. Based on the proposed spectrum–sharing formulations, inter-operator joint optimization problem was formulated to obtain power efficient intra- and inter-RAN beamforming vectors for supplementary energy gains and improved UE signal reception.

3. I showed, by numerical simulations, the better energy efficiency performance of the SEBS strategy as compared to the non-cooperative scheme. I also showed that the EE by spectrum trading may not always boost the spectral efficiency. When the bandwidth allocated for inter-RAN UE support exceeded a specific ratio, the intra-RAN UE support bandwidth wasvreduced, thus resulting in lower UE data rates.

4.4 Summary

In this work, we have presented the Sleep-mode with Efficient Beamformers and Spectrum-sharing (SEBS) strategy, which reduces the power consumption of colocated BSs of multiple cooperative operators. We have utilized the concept of dynamic spectrum allocation and carrier aggregation to partition an operator's licensed bandwidth into private and shared bands in which a UE's access is mutually exclusive. To achieve energy efficient transmission power, we have formulated an optimization problem to obtain the optimal intra- and inter-RAN beamforming vectors while taking the maximum RRH power limit, and intra- and inter-operator UE minimum data rates into consideration. We have also proposed Sleep-mode with Efficient Beamformers and Spectrum-sharing (SEBS) strategies based on the load transfer condition algorithm to achieve gains in energy efficiency. The load transfer condition algorithm considers the channel quality and current traffic load of the cooperating BSs. The proposed inter-operator joint energy saving schemes include a spectrum–user support trading to motivate cooperation among identical operators. We have also analyzed the impacts of the schemes on the joint sum rates of UEs by exploring different bandwidth sharing formulations. The results show that SEBS strategies achieve significantly a higher energy efficiency than non-cooperative schemes. It is also revealed that spectrum–user support trading yields higher joint sum rate than in equal spectrum sharing when appropriate bandwidth sharing ratios are used.

CHAPTER 5: PROPOSED SLEEP-MODE STRATEGY IN MOBILE EDGE COMPUTING NETWORK

5.1 Introduction

One of the fundamental techniques for boosting network capacity is network densification, where base stations are deployed in an ultra-dense fashion. Also, incorporation of cloud computing into radio access networks (RANs), facilitated by Cloud-RAN, is geared towards meeting the requirement of system capacity efficiently. Cloud-RAN is based on centralization of base station baseband processing into a pool with data traffic conveyed between the pool and densely deployed remote radio heads (RRHs) over a fronthaul.

However, Cloud-RAN and network densification come with some challenges, such as fronthaul constraints [57] and increase in energy budget due to the constant operation of the network devices [58]. To alleviate these constraints, mobile edge and fog-computing based RANs are proposed to alleviate the Cloud-RAN fronthaul challenge. These computing based RAN architectures move cloud services such as computing and storage to the RRHs, closer to the user equipment (UEs). Depending on the architecture and the functionalities, the computation equipped RRHs have been termed enhanced RRHs (eRRHs) [59], fog-computing Access Points (F-APs) [60], and edge-computing Access Points (EC-APs) [61]. The term Access Points (APs) is adopted in this paper to describe these base stations capable of supporting radio and edge node computing services.

With edge and fog computing, the densely deployed APs can provide radio resource to the UEs, and capability of execution of computation tasks. However, such AP network deployment exhibits significant traffic variations [58]. In a trace study conducted

in [58], the traffic profile significantly differs between weekdays and weekends. Hourly fluctuation of traffic is also observed in the same study. Consequently, some APs are either idle or underutilized at low traffic. Higher fluctuating traffic profile is expected in 5G wireless systems, where densely deployed edge devices, such as APs, can be idle at every time instant [62]. Therefore, it will be considerably energy-efficient to put the optimal number of APs into operation to meet the communication and computation demands of the UEs and switch off other APs.

Besides the benefit of energy efficiency, offloading of UEs' computation workloads to APs provides lower latency for task processing. The conventional way of offloading computation to the remote cloud could cause intolerable delay [63]. For instance, computation processing request round-trip times (RTTs) to remote cloud servers could amount to the order of tens of milliseconds, while edge servers could proffer less than 10ms RTT [64]. Thus, computation offloading to the proximate servers is increasingly becoming attractive. The focus of this research is on minimizing energy consumption of densely deployed APs processing radio and delay sensitive computation task. We propose to achieve the goal by optimally selecting active APs while maintaining the system QoS constraints, especially latency.

The need to optimally select active number of APs in ultra-dense networks with fluctuating traffic profile can be intuitively illustrated. A higher number of active APs leads to lower computation latency as more edge servers are available for computation processing. However, the network power consumption rises due to the high number of APs in operation. Thus, we are motivated by the interplay between computation completion latency at the edge servers and the number of active APs to seek optimal AP subset to support radio and computation traffic demands with the ultimate aim of minimizing network energy consumption and satisfying delay constraint.

5.2 System Model

Without loss of generality, we consider a densely deployed edge computing cellular network based on C-RAN architecture. The network consists of I identical RRHs equipped with N antennas each. Each RRH is endowed with a computation server. Thus, the coupling of RRH-server is referred to as AP as noted earlier. The APs set in the region is indexed by $\mathcal{I} = \{1, 2, ..., I\}$. Let $\mathcal{L} \subseteq \mathcal{I}$ denote the set of active APs and $\mathcal{Z}$ denote the set of APs in sleep mode, such that $\mathcal{L} \cup \mathcal{Z} = \mathcal{I}$. Each UE is equipped with a single antenna. The UEs set in the region is indexed by $\mathcal{K} = \{1, 2, ..., K\}$. Due to limited computational capacity, each UE processes a fraction of its computation task locally and offload the remaining to its associated AP for processing. Unlike the legacy Cloud-RAN where the fronthaul is inefficiently burdened by centralized baseband processing at the BBU pool, signal and computation processing are performed in a distributed fashion. Therefore, the functionalities at the AP include the capabilities of the radio frequency, the physical layer, and the upper layers procedures.

5.2.1 Communication Model

We consider the uplink and downlink transmission in the model. Let the vector consisting of the channels from all the APs to the k-th UE be $\mathbf{h}_k^{\mathsf{H}} = \left[\mathbf{h}_{1k}^{\mathsf{H}}, \cdots, \mathbf{h}_{Ik}^{\mathsf{H}}\right]$. Channel reciprocity is assumed in the uplink and downlink channels such that with time-division duplex (TDD), the channel vector in the uplink is merely the transpose of that in the downlink [101].

For the k-th UE uploading its computation task to the AP, we denote the uplink transmission normalized symbol as $s_k^U \in \mathbb{C}$, such that $\mathbb{E}\left[|s_k|^2\right] = 1$. All UEs are assumed to transmit with an identical power p^U. Thus, the transmit signal from the UE k is

$$x_k^U = \sqrt{p^U} s_k^U, \quad \forall k \in \mathcal{K}. \tag{5.1}$$

Therefore, the received uplink signal at all APs from the k-th UE is

$$\mathbf{y_k} = \sum_{k \in \mathcal{K}} \mathbf{h}_{ik} \sqrt{p^U} s_k^U + \boldsymbol{\eta}_k^U, \tag{5.2}$$

where $\boldsymbol{\eta}_k^U \in \mathbb{C}^N$ is the receiver noise vector at all APs consisting of circularly symmetric complex Gaussian random variables each distributed as $\mathcal{CN}(0, \sigma^2)$.

Let $\mathbf{m}_k^U$ denote the receiver beamforming vector used to decode s_k^U from the kth UE. The signal-to-interference-plus-noise ratio (SINR) of the k-th UE uplink transmission after applying $\mathbf{m}_k^U$ is given by

$$\gamma_k^U = \frac{p^U \left| \left(\mathbf{m}_k^U\right)^T \mathbf{h}_k \right|^2}{\sum_{k \in \mathcal{K}, j \neq k} p^U \left| \left(\mathbf{m}_k^U\right)^T \mathbf{h}_j \right|^2 + \left\| \mathbf{m}_k^U \right\|_2^2 \sigma^2}, \tag{5.3}$$

where $\mathbf{m}_k^U = \left[\left(\mathbf{m}_{1k}^U\right)^T, ..., \left(\mathbf{m}_{Ik}^U\right)^T \right]^T$. Also, $\mathbf{m}_{ik}^U \in \mathbb{C}^N$ represents the beamforming vector for the k-th UE received at the i-th AP, where the array of RRH antennas acts as the receiver for the K independent streams of data transmitted from the UEs. The uplink achievable rate for k-th UE is expressed as

$$R_k^U = W \log_2(1 + \gamma_k^U). \tag{5.4}$$

Let $\mathbf{w}_{ik}^D \in \mathbb{C}^N$ denotes the downlink transmission beamforming vector from AP i to UE k and the downlink transmitted normalized symbol denoted by s^D. The transmitted signal is given as

$$x_i^D = \sum_{k \in \mathcal{K}} \mathbf{w}_{ik}^D s_k^D, \quad \forall i \in \mathcal{I}. \tag{5.5}$$

The aggregated beamforming vectors from all APs to the k-th UE is therefore denoted as $\mathbf{w}_k^D = \left[\left(\mathbf{w}_{1k}^D\right)^T, ..., \left(\mathbf{w}_{Ik}^D\right)^T \right]^T$. Thus, we can write the received signal $y_k^D \in \mathbb{C}^N$ by

the k-th UE as

$$y_k^D = \mathbf{h}_k^{\mathsf{H}} \mathbf{w}_k^D s_k^D + \sum_{j \neq k}^{K} \mathbf{h}_k^{\mathsf{H}} \mathbf{w}_j^D s_j^D + \eta_k^D, \tag{5.6}$$

where $\eta_k^D \sim \mathbb{CN}(0, \sigma^2)$ is the additive white Gaussian noise at the k-th UE . The downlink signal-to-interference-plus-noise ratio (SINR) for the UE k can be expressed as

$$\Upsilon_k^D = \frac{\left|\mathbf{h}_k^{\mathsf{H}} \mathbf{w}_k^D\right|^2}{\sum_{k \in \mathcal{K}, j \neq k} \left|\mathbf{h}_k^{\mathsf{H}} \mathbf{w}_j^D\right|^2 + \sigma^2}. \tag{5.7}$$

The downlink achievable rate of the k-th UE is therefore

$$R_k^D = W \log_2(1 + \Upsilon_k^D). \tag{5.8}$$

5.2.2 Computation Model

We consider a computational model where the UE executes a fraction of its computation locally and offload the remaining [102]. The UE's decision is dependent on the device energy, computation capacity and the task's latency constraint [103]. There exist recent works that focus on energy minimization of the UE, such as [66, 104]. Since our interest is in the reduction of the mobile operator's high energy cost, in this paper, we focus on minimizing power consumption of the computation processing MEC network under the latency constraints of the offloaded task. The interest to focus only on the energy efficiency of the APs has also been demonstrated by studies, such as [68, 103, 105].

Each UE offloaded portion of its task to its associated AP is described by tuple $\langle D_k, U_k, T_k \rangle$. D_k denotes the size of the input data (in bits) of the computation task from the k-the UE to the i-th AP, which may include program codes and input parameters. U_k represents the total number of AP server CPU cycles required to complete the task, and T_k denotes the task completion deadline (in seconds). Each task is atomic and cannot be partitioned into subtasks; hence a task cannot be offloaded to

more than one AP. In principle, multiple APs could cooperatively serve a UE, but such a scenario would complicate computation offload analysis and management Thus . Similar to studies in [106, 107] we assume each UE is served by a single AP.

We can compute the transmission time of k-th UE for offloading the task size D_k to the i-th AP as

$$T_{i,k}^{U} = \frac{D_k}{R_k^U}. \tag{5.9}$$

5.2.2.1 AP Server Computing Cost

The total energy consumption at the i-th AP due to CPU computation execution, denoted by E_i^{comp} is given as [62]

$$E_i^{comp} = \sum_{k \in \mathcal{K}} \kappa U_k f_{i,k}^2, \tag{5.10}$$

where κ is a hardware architecture related constant. $f_{i,k}$ is the computation capacity in cycles per second of the AP server allocated to k-th UE's task. The computation at each server is limited by $\sum_{k \in \mathcal{K}} f_{i,k} \leq F_i^{max}$, which implies that the maximum computation capacity in cycles per second of i -th AP server is denoted by F_i^{max}. The overhead in terms of time for executing the k-th UE's task by the i-th AP server is expressed as

$$T_{i,k}^{exe} = \frac{U_k}{f_{i,k}}. \tag{5.11}$$

5.2.2.2 Computation latency

Similar to many studies [108–110] we ignore the latency for computation result delivery from AP server to the UEs because, in many applications, the computation outcome is much smaller than the input task size [111]. The total remote computation latency for the offloaded k-th UE's task at the i-th AP is therefore

$$T_{i,k}^{tot} = T_{i,k}^{U} + T_{i,k}^{exe}. \tag{5.12}$$

5.2.3 Power Model

In this paper, we focus on the energy efficiency of the computation processing MEC network, and aim to minimize the network power consumption while satisfying the latency constraints of the offloaded computation tasks from the UEs. The power consumption includes both the conventional BS power and computation server power.

5.2.3.1 Communication Power Consumption Model

Here, we consider the power consumption of a conventional BS (excluding the computation server). To make a distinction between the power consumed by the conventional BS components and the added server, we term the power consumed by the former as *communication power*, and the power consumption of the latter as *computation power*. We note that the communication power includes the BS transmission power and the static power required to keep the BS auxiliary site equipment in operation.

We represent the operation power (excluding transmission power) of the i-th AP and its wired fronthaul link while on the active mode as P_i^{active}. The power consumed by the i-th AP and its fronthaul link on the sleep-mode is denoted by P_i^{sleep}. Therefore, the total communication power consumption can be expressed as

$$P_{comm} = \eta^{-1} \sum_{i \in \mathcal{L}} P_{tr} + \sum_{i \in \mathcal{L}} P_i^{active} + \sum_{i \in \mathcal{Z}} P_i^{sleep}.. \tag{5.13}$$

Since $\mathcal{L} \cup \mathcal{Z} = \mathcal{I}$, then $\sum_{i \in \mathcal{Z}} P_i^{sleep} = \sum_{i \in \mathcal{I}} P_i^{sleep} - \sum_{i \in \mathcal{L}} P_i^{sleep}$. Also, substituting $\sum_{i \in \mathcal{L}} P_i^{d} = \sum_{i \in \mathcal{L}} P_i^{active} - \sum_{i \in \mathcal{L}} P_i^{sleep}$, and $P_i^{tr} = \|\mathbf{w}_{ik}\|_2^2$, the communication power can be rewritten as

$$P_{comm} = \eta^{-1} \sum_{i \in \mathcal{L}} \|\mathbf{w}_{ik}\|_2^2 + \sum_{i \in \mathcal{L}} P_i^{d} + \sum_{i \in \mathcal{I}} P_i^{sleep}, \tag{5.14}$$

where η is the RF power efficiency, which depends on the number of transmitter antenna [112].

5.2.3.2 Computation Power Consumption Model

Using (5.10) and (5.11), we can compute the total computation power at AP server i as

$$P_i^{comp} = \sum_{k \in \mathcal{K}} \frac{\kappa U_k f_{i,k}^2}{T_{i,k}^{exe}} = \sum_{k \in \mathcal{K}} \kappa f_{i,k}^3. \tag{5.15}$$

Thus, the network computation power is

$$P_{comp}(\mathbf{f}) = \sum_{i \in \mathcal{L}} \sum_{k \in \mathcal{K}} \kappa f_{i,k}^3. \tag{5.16}$$

It can be seen that in AP sleep-mode P_i^{comp} is 0, as only AP $i \in \mathcal{L}$ is involved in workload computation. Thus we propose putting the server into the sleep-mode with the RRH for optimal power savings.

5.2.3.3 Total Network Power

The total network power consumption can now be aggregated as $P_{tot} = P_{comm}(\mathcal{L}, \mathbf{w}) + P_{comp}(\mathbf{f})$, explicitly as

$$P_{tot} = \sum_{i \in \mathcal{L}} \left[\eta^{-1} \sum_{k \in \mathcal{K}} \|\mathbf{w}_{ik}\|_2^2 + P_i^d + \sum_{k \in \mathcal{K}} \kappa f_{ik}^3 \right]. \tag{5.17}$$

Since the last term $\sum_{i \in \mathcal{I}} P_i^{sleep}$ of (5.14) is a constant involving all APs, it is trivial to optimal active AP selection problem, and it is therefore omitted in (5.17).

5.3 Problem Formulation

In times of low traffic, an optimal number of active APs can be sought. The remaining APs are put into the sleep-mode, while their associated UEs are transferred to the active APs for communication and remote computation support. This is illustrated in Figure 5.1. Our goal is to select the optimal number of active APs that

minimize the overall network power consumption. Also, efficient beamforming vectors are jointly sought between the UEs and the active APs while satisfying task computation deadline and other QoS constraints. The constraints are transmission power constraints, and uplink and downlink transmissions minimum data rates. The dual link constraints imply optimization of the uplink beamforming vectors in addition to the downlink beamformers.

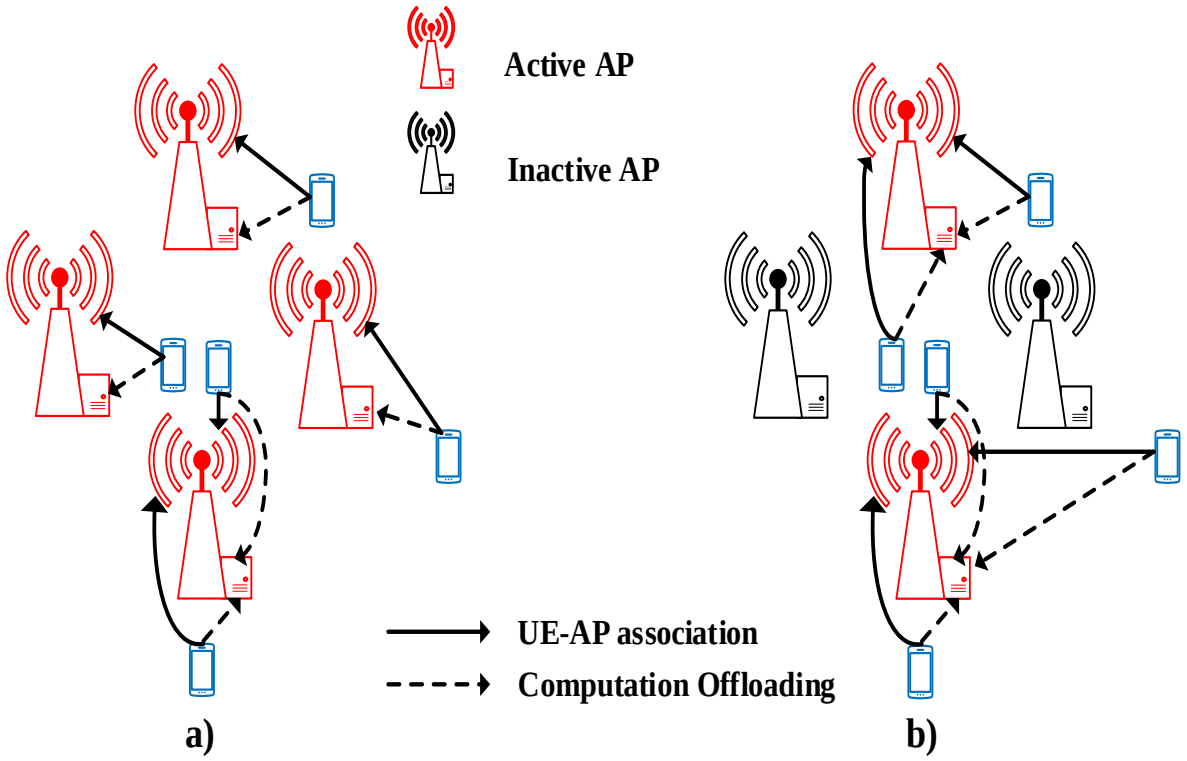

Figure 5.1: Communication and computation load support with active AP set selection. a) All APs are active. b) Some APs are put on sleep-mode and their UEs transferred to active APs for communication and remote computation task processing.

The optimization problem can be formulated as

$$
\begin{aligned}
\mathscr{P}: \quad & \min_{\mathbf{L},\mathbf{w},\mathbf{m},\mathbf{f}} \quad P_{tot} \\
& \text{s.t.} \qquad \mathcal{C}1: \sum_{k} \|\mathbf{w}_{ik}\|_2^2 \leq P_i^{max}, \quad \forall i \in \mathcal{L} \\
& \qquad\quad \mathcal{C}2: \sum_{k} \|\mathbf{w}_{ik}\|_2^2 = 0, \quad \forall i \in \mathcal{Z} \\
& \qquad\quad \mathcal{C}3: R_k^U \geq R_{k,min}^U, \quad \forall k \in \mathcal{K} \\
& \qquad\quad \mathcal{C}4: R_k^D \geq R_{k,min}^D, \quad \forall k \in \mathcal{K} \\
& \qquad\quad \mathcal{C}5: \sum_{k \in \mathcal{K}} f_{ik} \leq F_i^{max}, \quad \forall i \in \mathcal{L} \\
& \qquad\quad \mathcal{C}6: \sum_{k \in \mathcal{K}} f_{ik} = 0, \quad \forall i \in \mathcal{Z} \\
& \qquad\quad \mathcal{C}7: T_{i,k}^{tot} \leq T_{i,k}^{max}
\end{aligned}
\qquad . \tag{5.18}
$$

where $\mathbf{f} = \left[\mathbf{f}_1^T, ..., \mathbf{f}_K^T\right]^T$, $\mathbf{f}_k = [f_{1,k}, ..., f_{I,k}]$ and $\mathbf{w} = \left[\mathbf{w}_{1k}^{\mathsf{H}}, \mathbf{w}_{2k}^{\mathsf{H}}, \cdots, \mathbf{w}_{Ik}^{\mathsf{H}}\right]^{\mathsf{H}}$. $\mathcal{C}1$ gives the upper limit of the AP transmit power. $\mathcal{C}2$ enforces $\mathbf{w}_{ik} = 0$ for any k-th UE not associated with any i-th AP . The uplink and downlink minimum data rates are given in $\mathcal{C}3$ and $\mathcal{C}4$, respectively. $\mathcal{C}5$ gives the upper limit of CPU cycles to be allocated to the computation tasks by the AP, and $\mathcal{C}6$ ensures the computation power of an AP having no offloaded tasks is 0. The total allowable deadline for each offloaded task is given in $\mathcal{C}7$.

Constraints Transformation

The uplink and downlink data rates constraints of problem $\mathscr{P}$ are non-convex. First, we rewrite the downlink data rate constraint into second-order cone equivalent. Since the phase of part of the aggregated beamformers has no effect on the objective function, the downlink data rate constraint is given as [113]

$$
\mathcal{C}3\text{a}: \quad \|\mathrm{r}_k^D\|_2 \leq \sqrt{1 + 1/(2^{R_{k,\min}^D/W} - 1)}\mathrm{Re}\{u_{kk}\}, \forall k, \tag{5.19}
$$

where $r_k^D = [u_{k1}, u_{k2}, ..., u_{kk}, \sigma_k]^T$,

$u_{kj} = \sum_{i \in \mathcal{L}} \mathbf{h}_k^{\mathsf{H}} \mathbf{w}_j$.

For the uplink transmission rate constraint, we transform the uplink channel into a virtual downlink channel using uplink and downlink duality. The duality implies that same minimum transmit power is required to meet the downlink SINR target as in the uplink target, where the uplink channel is actualized by downlink's input and output reversal [114]. After application of uplink and downlink duality, uplink beamforming vectors $\mathbf{m}_k^U$ can be rewritten as $\mathbf{v} = \left[(\mathbf{v}_{1k})^T, ..., (\mathbf{v}_{Ik})^T \right]^T$, where $\mathbf{v}_{ik} \in \mathbb{C}^{IN}$. Thus, the second-order cone form is

$$\mathcal{C}4\text{a}: \quad \|\mathrm{r}_k^U\|_2 \leq \sqrt{1 + 1/(2^{R_{k,\min}^U/W} - 1)} \mathrm{Re}\{d_{kk}\}, \forall k, \tag{5.20}$$

where $r_k^U = [d_{k1}, d_{k2}, ..., d_{kk}, \sigma_k]^T$,

$d_{kj} = \sum_{i \in \mathcal{L}} \mathbf{h}_k^{\mathsf{H}} \mathbf{v}_j$. The optimization problem becomes

$$\begin{aligned} \mathscr{P}5.1: \quad & \min_{\mathbf{L},\mathbf{w},\mathbf{v},\mathbf{f}} \quad P_{tot} \\ & \text{s.t.} \qquad \mathcal{C}1, \mathcal{C}2, \mathcal{C}3\text{a}, \mathcal{C}4\text{a}, \mathcal{C}5 - \mathcal{C}7. \end{aligned} \tag{5.21}$$

5.4 Problem Analysis

Despite the transformation to yield second-order cone (SOC) form in $\mathcal{C}3$ and $\mathcal{C}4$, the problem $\mathscr{P}5.1$ is still not convex due to the latency constraint, $\mathcal{C}7$. Using (5.9), (5.11) and (5.12), $\mathcal{C}7$ can be rewritten as

$$\frac{D_k}{R_k^U} + \frac{U_k}{f_{i,k}} \leq T_{i,k}^{max}. \tag{5.22}$$

Therefore $\mathcal{C}7$ is reformulated as

$$\mathcal{C}7\text{a:} \quad R_k^U \geq \frac{D_k}{T_{i,k}^{max} - \frac{U_k}{f_{i,k}}}. \tag{5.23}$$

Yet, (5.23) is not tractable because it is embedded with two decision (optimization) variables $\mathbf{v}$ in R_k^U, and $f_{i,k}$. Thus, we start out by fixing $f_{i,k}$ and optimizing other variables. One way to fix $f_{i,k}$ is to allocate it based on its U_k relative to all others at the AP. Let each U_k of D_k data arrived at AP i be denoted by U_{ik}, the number CPU cycles allocated by AP i for task k is

$$\widetilde{f}_{i,k} = \frac{U_{ik}}{\sum_{k\in\mathcal{K}} U_{ik}} F_i^{max}, \quad \forall k \in \mathcal{K}, \forall i \in \mathcal{L}. \tag{5.24}$$

Similar to (20), $\mathcal{C}7a$ can be transformed to so SOC form

$$\mathcal{C}7\text{b:} \quad \|\mathrm{r}_k^U\|_2 \leq \sqrt{1 + 1/(2^{\delta_k/W} - 1)}\mathrm{Re}\{d_{kk}\}, \forall k, \tag{5.25}$$

where $\delta_k = \frac{D_k}{T_{ik}^{max} - \frac{U_k}{\widetilde{f}_{i,k}}}$.
With fixed $\widetilde{f}_{i,k}$, the problem is restated as

$$\begin{aligned} \mathscr{P}5.2: \quad & \min_{\mathbf{L},\mathbf{w},\mathbf{v}} \quad P_{tot} \\ & \text{s.t.} \quad \mathcal{C}1, \mathcal{C}2, \mathcal{C}3\text{a}, \mathcal{C}4\text{a}, \mathcal{C}5, \mathcal{C}6, \mathcal{C}7b. \end{aligned} \tag{5.26}$$

To solve problem $\mathscr{P}5.2$, we consider a case of a given active APs set $\mathcal{L}$. Hence, the AP transmission power constraint $\mathcal{C}1$ can be rewritten as

$$\mathcal{C}1a: \quad \sum_k \|\mathbf{L}_{ik}\mathbf{w}_{ik}\|_2^2 \leq P_i^{max}, \quad \forall i \in \mathcal{L}, \tag{5.27}$$

where $\mathbf{L}_{ik} \in \mathbb{C}^{LN\times LN}$ is a block diagonal matrix having i-th main diagonal identity matrix $\mathbf{I}_N$ and zeros elsewhere. With a given active AP set $\mathcal{L}$, the problem can be

restated as

$$\mathscr{P}5.3\,(\mathcal{L}) :\min_{\mathbf{w},\mathbf{v}} \quad \sum_{i\in\mathcal{L}}\left[\eta^{-1}\sum_{k\in\mathcal{K}}\|\mathbf{L}_{ik}\mathbf{w}_{ik}\|_2^2 + P_i^d + \sum_{k\in\mathcal{K}}\kappa\widetilde{f}_{i,k}^3\right] \tag{5.28}$$
$$\text{s.t.} \quad \mathcal{C}1\text{a}, \mathcal{C}2, \mathcal{C}3\text{a}, \mathcal{C}4\text{a}, \mathcal{C}5, \mathcal{C}6, \mathcal{C}7b.$$

It is possible to solve problem $\mathscr{P}5.3\,(\mathcal{L})$ by the interior method, and subsequently search over all AP sets that will minimize network power consumption, but this approach will be computationally expensive [115]. Instead, we will apply AP selection algorithm similar to the greedy algorithm to select optimal active AP set.

5.4.1 Active AP selection

The selection rule is akin to the *minimum-increase rule* in the successive thinning algorithm in [116] for sparse filter design applied for backward selection in [115]. In contrast to our holistic power computation rule, the selection in [115] is limited to the communication resource. Hence, it is inapplicable to delay-sensitive networks. With the set of inactive APs initially set to null, the problem $\mathscr{P}5.3$ is iteratively solved and the AP that yields maximum total power consumption reduction, when switched off, in each iteration is removed from the active set and added to the inactive set until the optimal active set is achieved. At each iteration, while the problem is feasible, $\mathbf{w}$, $\mathbf{v}$, and $\mathbf{f}$ are re-optimized for the remaining set of APs. We assume the feasible region of $\mathscr{P}5.3\left(\mathcal{L}^{[j]}\right)$ is nonempty. As shown in Algorithm 4, $y^{[j]}$ denotes the AP at iteration j, obtained as $y^{[j]} = \arg\min_{i\in\mathbf{L}^{[j]}} \mathscr{P}5.3\left(\mathcal{L}^{[j]}\right)$ that yields minimum power consumption when its AP is switched off and thereafter added to the inactive set $\mathcal{Z}^{[j+1]}$. The removal in the procedure is without replacement. The optimal active AP set is denoted by $\mathcal{L}^*$, and we depict the optimal $\mathbf{w}$, and $\mathbf{v}$ of this procedure as $\mathbf{w}^*$, and $\mathbf{v}^*$, respectively.

5.4.2 Joint Optimization Algorithm

To achieve efficient the network power consumption, the joint optimization of the AP active set, the uplink and downlink beamforming vectors, and the AP allocated

Algorithm 4: Active AP Set Selection Algorithm

1 Initialize $j = 0$, $\mathcal{L}^{[0]} = \{1, ..., I\}$ and $\mathcal{Z}^{[0]} = \emptyset$.
2 With $\mathbf{f} = \widetilde{\mathbf{f}}$, solve problem $\mathscr{P}5.3\left(\mathcal{L}^{[j]}\right)$, and obtain $\mathscr{P}^*5.3\left(\mathcal{L}^{[j]}\right)$
3 **if** *feasible* **then**
4 Solve $y^{[j]} = \arg\min_{i \in \mathbf{L}^{[j]}} \mathscr{P}1.3\left(\mathcal{L}^{[j]}\right)$. Update $\mathcal{Z}^{[j+1]} = \mathcal{Z}^{[j]} \cup y^{[j]}$, and $\mathcal{L}^{[j+1]} = \mathcal{L}^{[j]} / y^{[j]}$, $j \leftarrow j + 1$, and go to step 3.
5 **end**
6 **else**
7 Go to Step 9
8 **end**
9 Output optimal AP active set $\mathcal{L}^{*[j]}$ if $j = 0$ or $\mathcal{L}^*[j-1]$ if $j \geq 1$, and the optimal $\mathbf{w}^*$, and $\mathbf{v}^*$

CPU cycles per task is thereafter implemented. With optimal number of APs realized from Algorithm 4, $\mathcal{L}^*$, $\mathbf{w}^*$, and $\mathbf{v}^*$ are obtained using $\mathscr{P}5.3\left(\mathcal{L}\right)$. The CPU cycles are then unfixed, and the optimal number of cycles is obtained by solving $\mathscr{P}5.1\left(\mathcal{L}^*, \mathbf{w}^*, \mathbf{v}^*\right)$. The overall approach is given in Algorithm 5.

Algorithm 5: Overall Algorithm for the Joint Optimization problem

1 Transform constraints $\mathcal{C}3$, $\mathcal{C}4$ and $\mathcal{C}7$ using (5.19), (5.20) and (5.25), respectively
2 Fix $\mathbf{f} = \widetilde{\mathbf{f}}$ using (5.24)
3 Solve $\mathscr{P}5.3\left(\mathcal{L}\right)$ using Algorithm 4 and obtain $\mathcal{L}^*$, $\mathbf{w}^*$, and $\mathbf{v}^*$
4 Unfix $\mathbf{f}$. Solve problem $\mathscr{P}5.1\left(\mathcal{L}^*, \mathbf{w}^*, \mathbf{v}^*\right)$ and obtain $\mathbf{f}^*$
5 Output: $\mathcal{L}^*$, $\mathbf{w}^*$, $\mathbf{v}^*$ and $\mathbf{f}^*$

5.5 Performance Evaluation

In this section, we evaluate the performance of the proposed joint optimization algorithms with numerical simulation. We consider 5 - 25 APs randomly deployed with inter-site distance (ISD) of 200 meters, within which UEs are uniformly distributed. Each AP is equipped with 4 transmit antennas, and each UE is endowed with one antenna. A pathloss model of 140 + 36.7log10(d) is assumed for the link distance d between APs and UEs. The utilized noise power σ^2 is -94dBm. A total network bandwidth of 20 MHZ is used. Similar to [115], we set $P_i^{max} = 4W$, $P_i^d = 5.6W$,

$P_i^{sleep} = 5.05W$, and $\eta = 0.36$. The required minimum uplink and downlink data rate is 5 Mbps and 10 Mbps, respectively.

For the computation tasks offloaded to the APs, we use $D_k \sim \text{Unif}([0, 2d_{avg}])$, where d_{avg} = 1kbits [68]. 330 cycles/byte, typical for Gzip, is adopted for U_k [70]. F_i^{max} is 10 Mega-cycles/second, and T_{ik}^{max} is 1 second.

Since radio and computation resources are jointly optimized, we consider benchmarks incorporating scenarios involving various selection methods for APs and the number of CPU cycles, including CPU cycles allocation using (24). The $f_{i,k}$ allocation as in (5.24) is termed *Disjoint Resource Allocation* [105]. For clarity of simulation results, the benchmarks and our proposed joint optimization algorithms are defined below.

- *All-AP active and Disjoint Resource Allocation (ADRA)*: All APs are in operation. Also, each task's $f_{i,k}$ at the AP is assigned using (5.24), i.e the allocated number of CPU cycles is set proportional to the amount of the computational load of each UE.

- *Optimal-AP number and Disjoint Resource Allocation (ODRA)*: Optimal number of APs are selected for UEs' communication and computation support, while others are put into the sleep-mode. Disjoint Resource Allocation is applied to each task's $f_{i,k}$.

- *Proposed Algorithm*: Joint optimization of the active AP number, the beamforming vectors, and the computation resource allocation.

In Figure 5.2, we show the performance of the joint optimization algorithm relative to the other discussed benchmarks as the traffic grows. The ODRA consumes much less power than the ADRA evidently due to the fewer number of operating APs. Considerable power is saved from the swicthed off RRHs. The proposed joint optimization algorithm, in addition to the switching off of the unselected APs, also

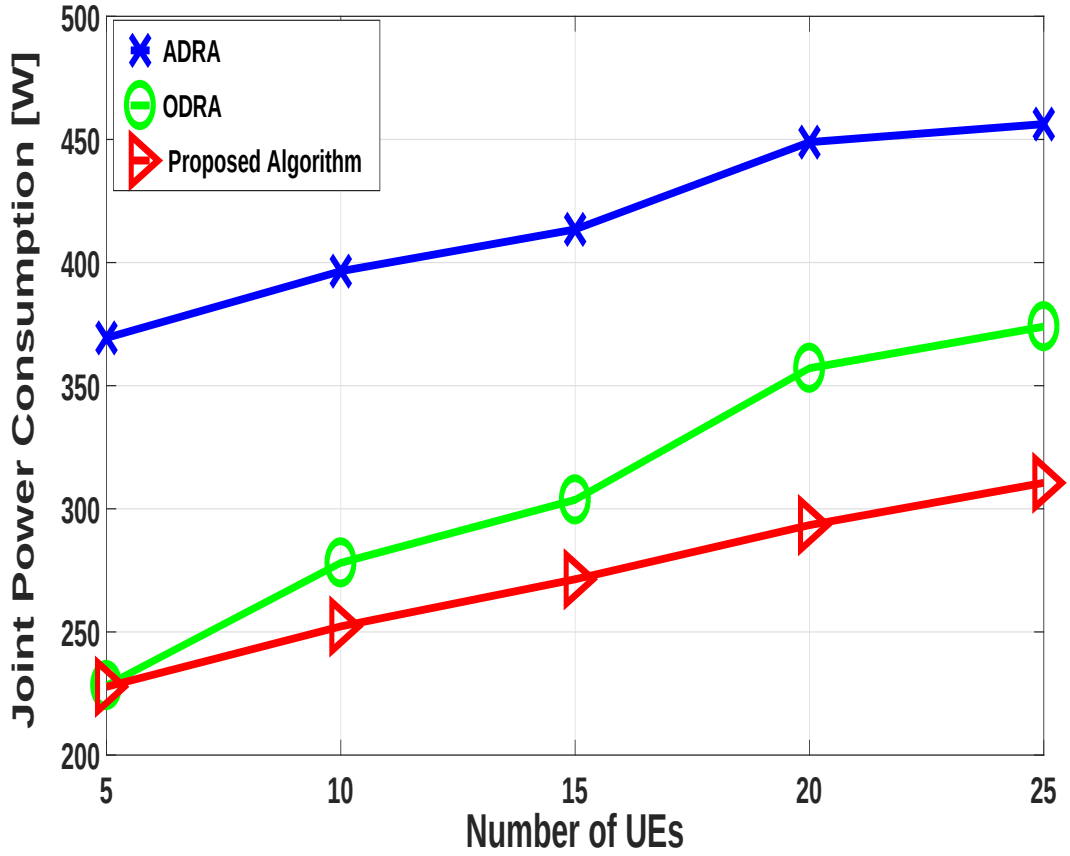

Figure 5.2: Total Power Consumption versus number of UEs

entails putting the computation servers of the inactive APs into the sleep-mode. Due to the efficient transmission power by the optimization of the precoding vectors, the proposed joint optimization saves even more power than the ODRA.

In order to evaluate the energy efficiency performance of our proposed scheme with network densification, we fix the number of UEs at 25 and increase the number of APs. The result is presented in Figure 5.3. The ODRA and the joint optimization algorithm yield significant energy savings even with increased AP densification.

The impact of optimizing the computational resource is isolated for illustration in Figure 5.4. Here, we compare the AP CPU cycles allocation by Disjoint Resource Allocation with the proposed joint optimization. As expected, the ADRA dissipates more computation power than the ODRA despite both schemes use the same CPU

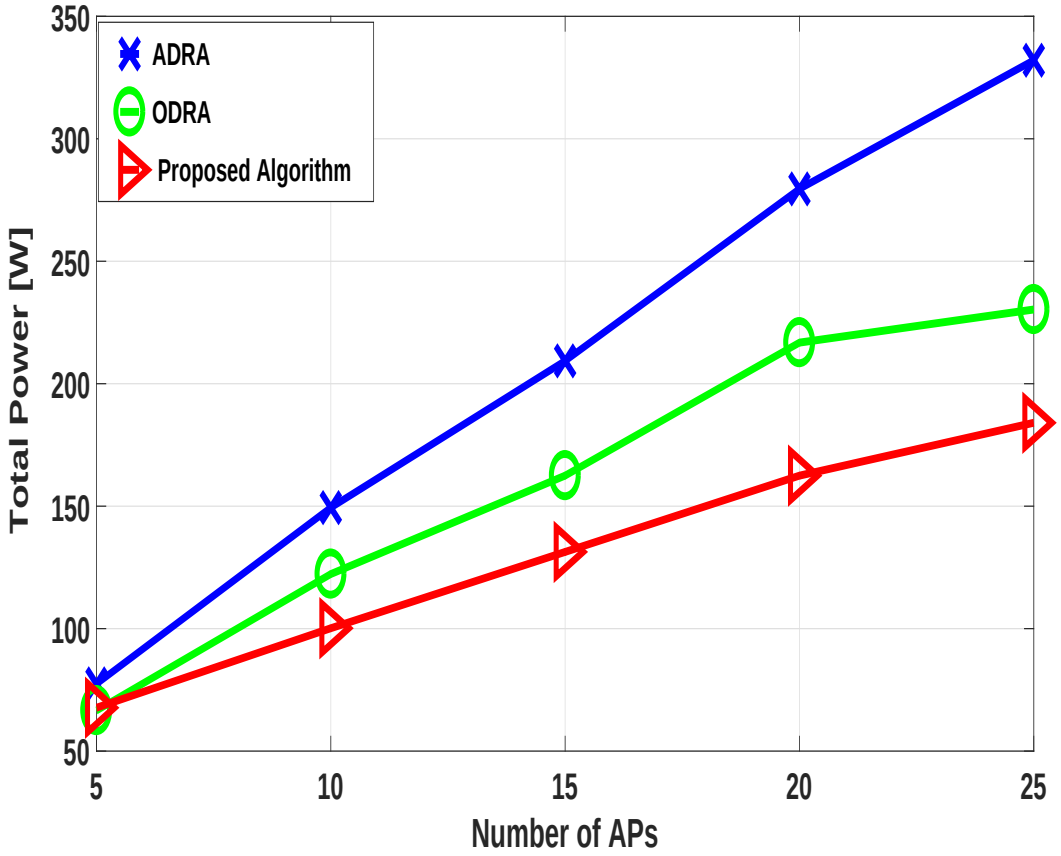

Figure 5.3: Total power consumption versus number of APs

cycles allocation method. This is as a result of having a smaller number of APs in the active mode in the ODRA scheme. The computation power of the proposed joint optimization shows more power is conserved in processing offloaded tasks than in the ODRA even though the same number of APs are engaged.

We extend the evaluation to the impact of increasing the CPU cycles at the APs on the computation power and the average latency of the computation tasks. This is performed using the proposed joint optimization scheme, and shown in Figure 5.5. The number of offloaded tasks and APs are fixed, each at 15. With increased available CPU cycles for tasks processing, the computation completion latency drops. However, with more computation resources used, more power is consumed. Interestingly, the computation power increase and the corresponding decrease in task processing

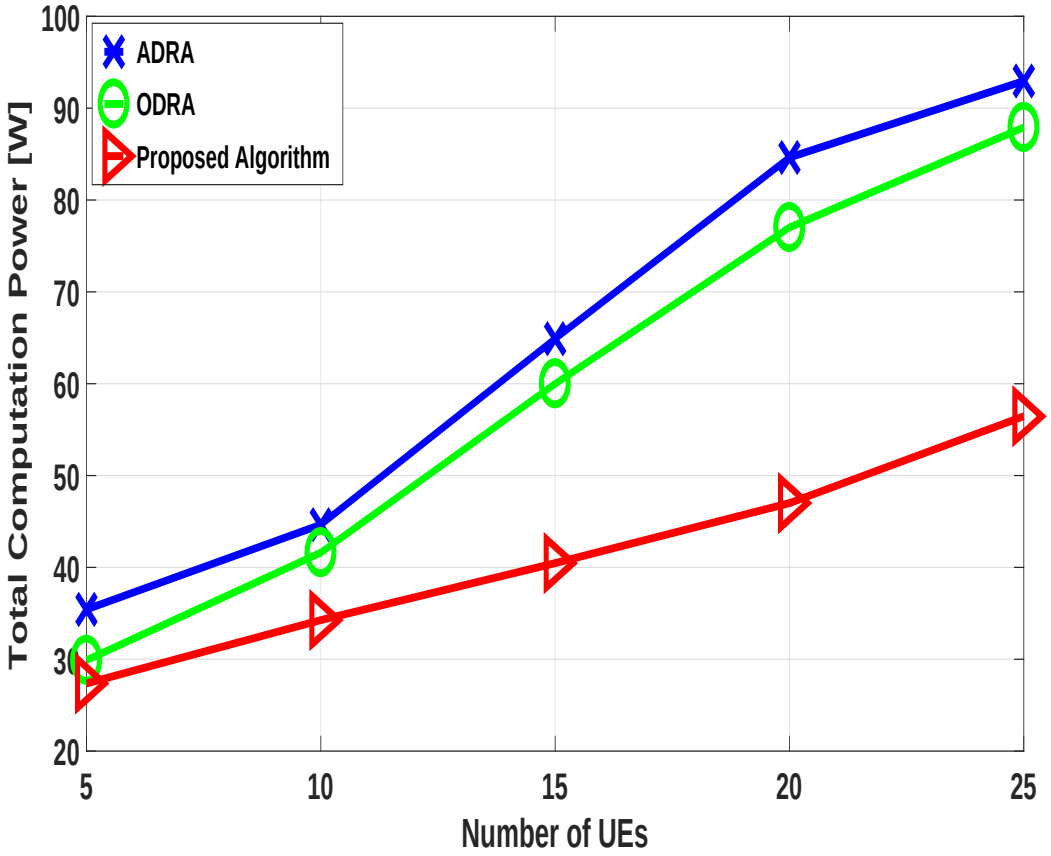

Figure 5.4: Total computation power consumption versus number of UEs

latency are within the specified constraints limits due to joint optimization of the communication and computation resources. For instance, each task latency is within the 1 second deadline regardless of the available AP maximum CPU cycles.

5.6 Contributions

The objective of this work is to minimize the energy consumption in edge computing based ultra-dense cellular networks while satisfying the UEs' service requirement in terms of latency. The main contributions of this paper are summarized in the following.

- I jointly optimized both uplink and downlink beamforming vectors, the edge server CPU cycles and active AP subset. To the best of our knowledge, this

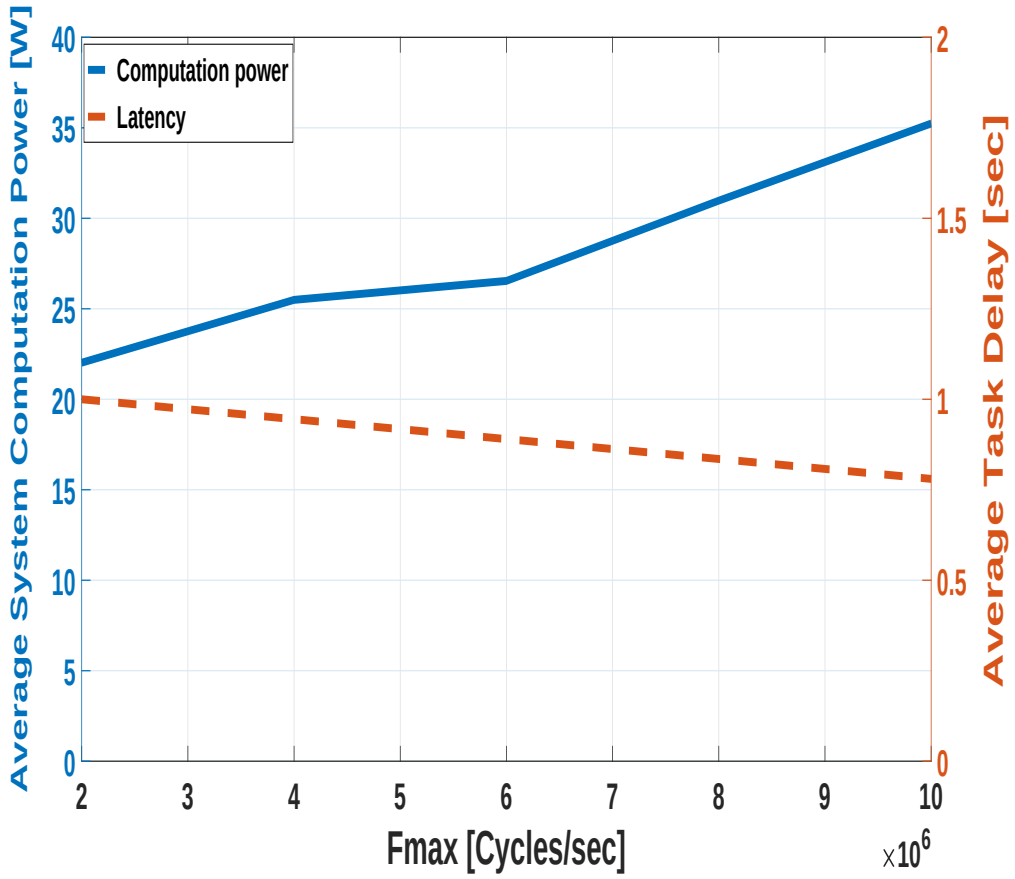

Figure 5.5: Impact of maximum AP CPU cycles on computation power consumption and task processing latency

is the first work that jointly considered the dual link beamforming vectors, CPU cycles, and optimal AP selection. The uplink beamforming vector was incorporated in the joint optimization to achieve efficient uplink data rate for low uplink latency.

- I selected optimal APs based on holistic energy selection rule. The power consumption rule encompassed computation processing power compared to the ubiquitous radio resources only power consumption.

- I evaluated the performance of our proposed radio and computation resources optimization algorithm in the emerging ultra-dense network scenario with CPU cycles and latency constraints. Simulation results showed considerable power

savings of our proposed scheme as compared to the relevant benchmarks.

5.7 Summary

In this chapter, I proposed a joint optimization strategy for achieving energy efficient network in which UEs' computation-intensive tasks are processed at the network edge while satisfying the computation latency requirement. The optimization method entailed selection of active APs, sufficient to support the communication and computation load in the network based on the contribution of the individual AP power dissipation to the overall network power consumption. The downlink and uplink beamforming vectors were also optimized for efficient transmission power and low computation latency, respectively. The CPU capacity allocated to each task was optimized to reduce power consumed in computation processing. The simulation results showed significant power savings by active AP selection and further savings by joint optimization of CPU cycles and the dual link beamforming vectors.

CHAPTER 6: CONCLUSION AND FUTURE WORKS

6.1 Conclusion

In this dissertation, BS sleep-mode strategies for multi-operators cellular networks were proposed, where BSs are closely located. Leveraging carrier aggregation for inter-operator bandwidth sharing, active BS selection was performed to minimize BS power consumption of cooperative multi-operators.

I have also proposed energy-saving strategy for a mobile edge computing network, where densely deployed BSs are equipped with computation resources to process users offloaded computation-intensive tasks. The trade-off between power consumption and latency was addressed by jointly formulating active number of BSs, uplink and downlink beamforming vectors, computation resource allocation, and task completion latency as an optimization problem. Subsequently, an active BSs selection framework based on communication and computation power-aware selection rule was proposed. The computation resources and dual link beamformers have been subsequently optimized for further network energy savings.

In all the cellular networks considered in this dissertation, the proposed sleep-mode strategies have been validated by numerical simulations to yield significant energy gains.

6.2 Future Works

Future works will be directed towards application of sleep-mode techniques in renewable energy powered base stations. Studies will be conducted in the following areas.

- Renewable energy powered densely deployed base stations. In practice, there exist many remote region without access to the power grid. Energy management with sleep-mode strategy in such application will be investigated.

- Techno-economic analysis of base stations with hybrid energy sources, i.e power grid and renewable energy sources will be investigated. Energy cost saving with sleep-mode strategy and net metering in hybrid powered ultra-dense base stations will be investigated.

REFERENCES

[1] L. Suarez, L. Nuaymi, and J.-M. Bonnin, "An overview and classification of research approaches in green wireless networks," *Eurasip journal on wireless communications and networking*, vol. 2012, no. 1, p. 142, 2012.

[2] C. V. networking Index, "Forecast and methodology, 2016-2021, white paper," *San Jose, CA, USA*, vol. 1, 2016.

[3] G. I. Gartner and I. Green, "The new industry shockwave, presentation at symposium," in *ITXPO conference, April*, 2007.

[4] J. Malmodin, Å. Moberg, D. Lundén, G. Finnveden, and N. Lövehagen, "Greenhouse gas emissions and operational electricity use in the ict and entertainment & media sectors," *Journal of Industrial Ecology*, vol. 14, no. 5, pp. 770–790, 2010.

[5] F. Han, Z. Safar, and K. R. Liu, "Energy-efficient base-station cooperative operation with guaranteed qos," *IEEE Transactions on Communications*, vol. 61, no. 8, pp. 3505–3517, 2013.

[6] M. Feng, S. Mao, and T. Jiang, "Base station on-off switching in 5g wireless networks: Approaches and challenges," *arXiv preprint arXiv:1703.09875*, 2017.

[7] M. M. Raikar, P. Doddagoudar, and M. S. Maralapannavar, "An algorithmic perspective of base station switching in dense cellular networks," in *Eco-friendly Computing and Communication Systems (ICECCS), 2014 3rd International Conference on*, pp. 177–182, IEEE, 2014.

[8] M. Ismail and W. Zhuang, "Network cooperation for energy saving in green radio communications," *IEEE Wireless Communications*, vol. 18, no. 5, pp. 76–81, 2011.

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

[10] Q. Zhao and D. Grace, "Transfer learning for qos aware topology management in energy efficient 5g cognitive radio networks," in *5G for Ubiquitous Connectivity (5GU), 2014 1st International Conference on*, pp. 152–157, IEEE, 2014.

[11] S. Wang and W. Guo, "Energy and cost implications of a traffic aware and quality-of-service constrained sleep mode mechanism.," *IET Communications*, vol. 7, no. 18, pp. 2092–2101, 2013.

[12] S. Zhou, A. J. Goldsmith, and Z. Niu, "On optimal relay placement and sleep control to improve energy efficiency in cellular networks," in *2011 IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, 2011.

[13] C. Gao, W. Zhang, J. Tang, C. Wang, S. Zou, and S. Su, "Relax, but do not sleep: A new perspective on green wireless networking," in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pp. 907–915, IEEE, 2014.

[14] L.-J. Wang and T. Han, "The relay selection algorithm based on minimizing the user terminal energy consumption," in *Wireless Communication and Sensor Network (WCSN), 2014 International Conference on*, pp. 270–275, IEEE, 2014.

[15] Z. Zhang, Y. Li, K. Huang, S. Zhou, and J. Wang, "Energy efficiency analysis of cellular networks with cooperative relays via stochastic geometry," *China Communications*, vol. 12, no. 9, pp. 112–121, 2015.

[16] S. K. A. Landou and A. N. Barreto, "Use of comp in 4g cellular networks for increased network energy efficiency," in *Telecommunications (IWT), 2015 International Workshop on*, pp. 1–6, IEEE, 2015.

[17] T. Han and N. Ansari, "On greening cellular networks via multicell cooperation," *IEEE Wireless Communications*, vol. 20, no. 1, pp. 82–89, 2013.

[18] S. Han, C. Yang, G. Wang, and M. Lei, "On the energy efficiency of base station sleeping with multicell cooperative transmission," in *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1536–1540, IEEE, 2011.

[19] P. Frenger, P. Moberg, J. Malmodin, Y. Jading, and I. Gódor, "Reducing energy consumption in lte with cell dtx," in *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*, pp. 1–5, IEEE, 2011.

[20] K. Hiltunen, "Total power consumption of different network densification alternatives," in *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications-(PIMRC)*, pp. 1401–1405, IEEE, 2012.

[21] S. Tombaz, S.-w. Han, K. W. Sung, and J. Zander, "Energy efficient network deployment with cell dtx," *IEEE Communications Letters*, vol. 18, no. 6, pp. 977–980, 2014.

[22] S. Baek and B. D. Choi, "Analysis of discontinuous reception (drx) with both downlink and uplink packet arrivals in 3gpp lte," in *Proceedings of the 6th International Conference on Queueing Theory and Network Applications*, pp. 8–16, ACM, 2011.

[23] A. Anisimov, S. Andreev, A. Lokhanova, and A. Turlikov, "Energy efficient operation of 3gpp lte-advanced and ieee 802.16 m downlink channel," in *Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), 2011 3rd International Congress on*, pp. 1–5, IEEE, 2011.

[24] I. Ashraf, F. Boccardi, and L. Ho, "Sleep mode techniques for small cell deployments," *IEEE Communications Magazine*, vol. 49, no. 8, pp. 72–79, 2011.

[25] S. Bhaumik, G. Narlikar, S. Chattopadhyay, and S. Kanugovi, "Breathe to stay cool: adjusting cell sizes to reduce energy consumption," in *Proceedings of the first ACM SIGCOMM workshop on Green networking*, pp. 41–46, ACM, 2010.

[26] Z. Niu, X. Guo, S. Zhou, and P. R. Kumar, "Characterizing energy–delay tradeoff in hyper-cellular networks with base station sleeping control," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 4, pp. 641–650, 2015.

[27] P. Ren and M. Tao, "A decentralized sleep mechanism in heterogeneous cellular networks with qos constraints," *IEEE Wireless Communications Letters*, vol. 3, no. 5, pp. 509–512, 2014.

[28] J. Manssour, P. Frenger, L. Falconetti, S. Moon, and M. Na, "Smart small cell wake-up field trial: Enhancing end-user throughput and network energy performance," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, pp. 1–5, IEEE, 2015.

[29] X. Guo, Z. Niu, S. Zhou, and P. Kumar, "Delay-constrained energy-optimal base station sleeping control," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 5, pp. 1073–1085, 2016.

[30] J.-B. Seo, "Rewarded access to a sleeping base station," *IEEE Communications Letters*, vol. 19, no. 9, pp. 1612–1615, 2015.

[31] C.-Y. Chang, W. Liao, H.-Y. Hsieh, and D.-S. Shiu, "On optimal cell activation for coverage preservation in green cellular networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 11, pp. 2580–2591, 2014.

[32] A. Mayoral, R. Vilalta, R. Casellas, R. Martinez, and R. Munoz, "Multi-tenant 5g network slicing architecture with dynamic deployment of virtualized tenant management and orchestration (mano) instances," in *ECOC 2016; 42nd European Conference on Optical Communication; Proceedings of*, pp. 1–3, VDE, 2016.

[33] J. Kibiłda, P. Di Francesco, F. Malandrino, and L. A. DaSilva, "Infrastructure and spectrum sharing trade-offs in mobile networks," in *IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), 2015*, pp. 348–357.

[34] "`https://www.cell-phone-towers.com/Cell-Tower-Co-location.html`."

[35] M. J. Farooq, H. Ghazzai, E. Yaacoub, A. Kadri, and M.-S. Alouini, "Green virtualization for multiple collaborative cellular operators," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 3, pp. 420–434, 2017.

[36] T. GSMA, "Mobile infrastructure sharing," *Report, September*, 2012.

[37] R. Sharing, "Necâs approach towards active radio access network sharing," *White Paper*, 2013.

[38] X. Costa-Pérez, J. Swetina, T. Guo, R. Mahindra, and S. Rangarajan, "Radio access network virtualization for future mobile carrier networks," *IEEE Communications Magazine*, vol. 51, no. 7, pp. 27–35, 2013.

[39] C. Liang and F. R. Yu, "Wireless network virtualization: A survey, some research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 358–380, 2015.

[40] J. Wu, Y. Zhang, M. Zukerman, and E. K.-N. Yung, "Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey," *IEEE communications surveys & tutorials*, vol. 17, no. 2, pp. 803–826, 2015.

[41] Y. S. Soh, T. Q. Quek, M. Kountouris, and H. Shin, "Energy efficient heterogeneous cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 5, pp. 840–850, 2013.

[42] K. Son, H. Kim, Y. Yi, and B. Krishnamachari, "Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks," *IEEE journal on selected areas in communications*, vol. 29, no. 8, pp. 1525–1536, 2011.

[43] K. Dufková, M. Bjelica, B. Moon, L. Kencl, and J.-Y. Le Boudec, "Energy savings for cellular network with evaluation of impact on data traffic performance," in *Wireless Conference (EW), 2010 European*, pp. 916–923, IEEE, 2010.

[44] N. Zhao, F. R. Yu, and H. Sun, "Adaptive energy-efficient power allocation in green interference-alignment-based wireless networks.," *IEEE Trans. Vehicular Technology*, vol. 64, no. 9, pp. 4268–4281, 2015.

[45] S. Bhaumik, G. Narlikar, S. Chattopadhyay, and S. Kanugovi, "Breathe to stay cool: adjusting cell sizes to reduce energy consumption," in *Proceedings of the first ACM SIGCOMM workshop on Green networking*, pp. 41–46, ACM, 2010.

[46] M. F. Hossain, K. S. Munasinghe, and A. Jamalipour, "Toward self-organizing sectorization of lte enbs for energy efficient network operation under qos constraints," in *Wireless Communications and Networking Conference (WCNC), 2013 IEEE*, pp. 1279–1284, IEEE, 2013.

[47] J. Tang, A. Shojaeifard, D. K. So, K.-K. Wong, and N. Zhao, "Energy efficiency optimization for comp-swipt heterogeneous networks," *IEEE Transactions on Communications*, vol. 66, no. 12, pp. 6368–6383, 2018.

[48] M. Yousefvand, T. Han, N. Ansari, and A. Khreishah, "Distributed energy-spectrum trading in green cognitive radio cellular networks," *IEEE Transactions on Green Communications and Networking*, vol. 1, no. 3, pp. 253–263, 2017.

[49] T. Han and N. Ansari, "Enabling mobile traffic offloading via energy spectrum trading," *IEEE transactions on wireless communications*, vol. 13, no. 6, pp. 3317–3328, 2014.

[50] N. Ansari, T. Han, and M. Taheri, "Gate: Greening at the edge," *arXiv preprint arXiv:1508.06218*, 2015.

[51] D.-E. Meddour, T. Rasheed, and Y. Gourhant, "On the role of infrastructure sharing for mobile network operators in emerging markets," *Computer Networks*, vol. 55, no. 7, pp. 1576–1591, 2011.

[52] A. Bousia, E. Kartsakli, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Game theoretic approach for switching off base stations in multi-operator environments," in *Communications (ICC), 2013 IEEE International Conference on*, pp. 4420–4424, IEEE, 2013.

[53] M. A. Marsan and M. Meo, "Energy efficient wireless internet access with cooperative cellular networks," *Computer Networks*, vol. 55, no. 2, pp. 386–398, 2011.

[54] L. Militano, A. Molinaro, A. Iera, and Á. Petkovics, "Introducing fairness in cooperation among green mobile network operators," in *Software, Telecommunications and Computer Networks (SoftCOM), 2012 20th International Conference on*, pp. 1–5, IEEE, 2012.

[55] A. Bousia, E. Kartsakli, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Game-theoretic infrastructure sharing in multioperator cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3326–3341, 2016.

[56] M. Oikonomakou, A. Antonopoulos, L. Alonso, and C. Verikoukis, "Cooperative base station switching off in multi-operator shared heterogeneous network," in *Global Communications Conference (GLOBECOM), 2015 IEEE*, pp. 1–6, IEEE, 2015.

[57] J. Peng, P. Hong, and K. Xue, "Stochastic analysis of optimal base station energy saving in cellular networks with sleep mode," *IEEE Communications Letters*, vol. 18, no. 4, pp. 612–615, 2014.

[58] Y. Chen, X. Wen, Z. Lu, H. Shao, and W. Jing, "Cooperation-enabled energy efficient base station management for dense small cell networks," *Wireless Networks*, vol. 23, no. 5, pp. 1611–1628, 2017.

[59] S.-H. Park, O. Simeone, and S. S. Shitz, "Joint optimization of cloud and edge processing for fog radio access networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7621–7632, 2016.

[60] H. Zhang, Y. Qiu, X. Chu, K. Long, and V. Leung, "Fog radio access networks: Mobility management, interference mitigation and resource optimization," *arXiv preprint arXiv:1707.06892*, 2017.

[61] M. Peng and K. Zhang, "Recent advances in fog radio access networks: Performance analysis and radio resource allocation," *IEEE Access*, vol. 4, pp. 5003–5009, 2016.

[62] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "Mobile edge computing: Survey and research outlook," *arXiv preprint arXiv:1701.01090*, 2017.

[63] T. X. Tran, A. Hajisami, P. Pandey, and D. Pompili, "Collaborative mobile edge computing in 5g networks: New paradigms, scenarios, and challenges," *IEEE Communications Magazine*, vol. 55, no. 4, pp. 54–61, 2017.

[64] E. Cuervo, A. Balasubramanian, D.-k. Cho, A. Wolman, S. Saroiu, R. Chandra, and P. Bahl, "Maui: making smartphones last longer with code offload," in *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pp. 49–62, ACM, 2010.

[65] C. Liu, B. Natarajan, and H. Xia, "Small cell base station sleep strategies for energy efficiency," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1652–1661, 2016.

[66] D. Chen, S. Schedler, and V. Kuehn, "Backhaul traffic balancing and dynamic content-centric clustering for the downlink of fog radio access network," in *Signal Processing Advances in Wireless Communications (SPAWC), 2016 IEEE 17th International Workshop on*, pp. 1–5, IEEE, 2016.

[67] M. Tao, E. Chen, H. Zhou, and W. Yu, "Content-centric sparse multicast beamforming for cache-enabled cloud ran," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6118–6131, 2016.

[68] Y. Mao, J. Zhang, K. Ben Letaief, Y. Wang, J. T. Y. Kwok, Q. Yao, L. NI, L.-H. Xiong, X. He, J. Xia, *et al.*, "Joint task offloading scheduling and transmit power allocation for mobile-edge computing systems," *Biotechnology for Biofuels*, vol. 10, no. 115, 2017.

[69] C. You, K. Huang, H. Chae, and B.-H. Kim, "Energy-efficient resource allocation for mobile-edge computation offloading," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1397–1411, 2017.

[70] T. Q. Dinh, J. Tang, Q. D. La, and T. Q. Quek, "Offloading in mobile edge computing: Task allocation and computational frequency scaling," *IEEE Transactions on Communications*, 2017.

[71] P. Zhao, H. Tian, C. Qin, and G. Nie, "Energy-saving offloading by jointly allocating radio and computational resources for mobile edge computing," *IEEE Access*, vol. 5, pp. 11255–11268, 2017.

[72] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental trade-offs on green wireless networks," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 30–37, 2011.

[73] X. Wang, A. V. Vasilakos, M. Chen, Y. Liu, and T. T. Kwon, "A survey of green mobile networks: Opportunities and challenges," *Mobile Networks and Applications*, vol. 17, no. 1, pp. 4–20, 2012.

[74] G. ETSI, "Network functions virtualisation (nfv); use cases," *V1*, vol. 1, pp. 2013–10, 2013.

[75] "`http://www.3gpp.org/DynaReport/SpecVsWi--36101.htm`."

[76] R. Guerzoni *et al.*, "Network functions virtualisation: an introduction, benefits, enablers, challenges and call for action, introductory white paper," in *SDN and OpenFlow World Congress*, 2012.

[77] J. S. Panchal, R. D. Yates, and M. M. Buddhikot, "Mobile network resource sharing options: Performance comparisons," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4470–4482, 2013.

[78] P. Spapis, K. Chatzikokolakis, N. Alonistioti, and A. Kaloxylos, "Using sdn as a key enabler for co-primary spectrum sharing," in *Information, Intelligence, Systems and Applications, IISA 2014, The 5th International Conference on*, pp. 366–371, IEEE, 2014.

[79] R. Kokku, R. Mahindra, H. Zhang, and S. Rangarajan, "Cellslice: Cellular wireless resource slicing for active ran sharing," in *Communication Systems and Networks (COMSNETS), 2013 Fifth International Conference on*, pp. 1–10, IEEE, 2013.

[80] F. Mazzenga, M. Petracca, R. Pomposini, F. Vatalaro, and R. Giuliano, "Performance evaluation of spectrum sharing algorithms in single and multi operator scenarios," in *Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd*, pp. 1–5, IEEE, 2011.

[81] Y. Teng, Y. Wang, and K. Horneman, "Co-primary spectrum sharing for denser networks in local area," in *Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM), 2014 9th International Conference on*, pp. 120–124, IEEE, 2014.

[82] B. Singh, "Repeated games for inter-operator spectrum sharing," *arXiv preprint arXiv:1505.04041*, 2015.

[83] Z. J. *et al.*, "Adaptive and robust signal processing in multi-user and multicellular environments (initial) d3. 1a," tech. rep., tech. rep., SAPHYRE, 2012.

[84] M. W. Arshad, A. Vastberg, and T. Edler, "Energy efficiency gains through traffic offloading and traffic expansion in joint macro pico deployment," in *Wireless Communications and Networking Conference (WCNC), 2012 IEEE*, pp. 2203–2208, IEEE, 2012.

[85] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, *et al.*, "How much energy is needed to run a wireless network?," *IEEE Wireless Communications*, vol. 18, no. 5, 2011.

[86] A. Wiesel, Y. C. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed mimo receivers," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 161–176, 2006.

[87] J. Li, X. Wu, and R. Laroia, *OFDMA mobile broadband communications: A systems approach*. Cambridge University Press, 2013.

[88] O. Tirkkonen, *Dynamic Inter-operator Spectrum Sharing Between Co-located Radio Access Networks Using Cooperation Transmission*. PhD thesis, Aalto University, 2014.

[89] S. Khakurel, L. Musavian, and T. Le-Ngoc, "Trade-off between spectral and energy efficiencies in a fading communication link," in *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th*, pp. 1–5, IEEE, 2013.

[90] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-and spectral-efficiency tradeoff in downlink ofdma networks," *IEEE transactions on wireless communications*, vol. 10, no. 11, pp. 3874–3886, 2011.

[91] H. Pervaiz, L. Musavian, and Q. Ni, "Energy and spectrum efficiency trade-off for green small cell networks," in *Communications (ICC), 2015 IEEE international conference on*, pp. 5410–5415, IEEE, 2015.

[92] Y. Wu, Y. Chen, J. Tang, D. K. So, Z. Xu, I. Chih-Lin, P. Ferrand, J.-M. Gorce, C.-H. Tang, P.-R. Li, *et al.*, "Green transmission technologies for balancing the energy efficiency and spectrum efficiency trade-off," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 112–120, 2014.

[93] P. Luoto, *Co-primary multi-operator resource sharing for small cell networks*. PhD thesis, University of Oulu, 2017.

[94] O. Aydin, D. Aziz, and E. Jorswieck, "Radio resource sharing among operators through mimo based spatial multiplexing in 5g systems," in *2014 IEEE Globecom Workshops (GC Wkshps)*, pp. 1063–1068, IEEE, 2014.

[95] J. Luo, J. Lindblom, J. Li, R. Mochaoura, A. Kortke, E. Karipidis, M. Haardt, E. Jorswieck, and E. G. Larsson, "Transmit beamforming for inter-operator spectrum sharing: From theory to practice," in *2012 International Symposium on Wireless Communication Systems (ISWCS)*, pp. 291–295, IEEE, 2012.

[96] R. Gangula, D. Gesbert, J. Lindblom, and E. G. Larsson, "On the value of spectrum sharing among operators in multicell networks," in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, IEEE, 2013.

[97] J. Lindblom and E. G. Larsson, "Does non-orthogonal spectrum sharing in the same cell improve the sum-rate of wireless operators?," in *2012 IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 6–10, IEEE, 2012.

[98] P. Bhat, S. Nagata, L. Campoy, I. Berberana, T. Derham, G. Liu, X. Shen, P. Zong, and J. Yang, "Lte-advanced: an operator perspective," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 104–114, 2012.

[99] Y. Chen, S. Zhang, S. Xu, and G. Y. Li, "Fundamental tradeoffs on green wireless networks," *arXiv preprint arXiv:1101.4343*, 2011.

[100] J. Opadere, Q. Liu, and T. Han, "Energy-efficient rrh sleep mode for virtual radio access networks," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pp. 1–6, IEEE, 2017.

[101] S. Luo, R. Zhang, and T. J. Lim, "Downlink and uplink energy minimization through user association and beamforming in c-ran," *IEEE Transactions on Wireless Communications*, vol. 14, no. 1, pp. 494–508, 2015.

[102] F. Wang and X. Zhang, "Dynamic interface-selection and resource allocation over heterogeneous mobile edge-computing wireless networks with energy harvesting," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 190–195, IEEE, 2018.

[103] S. Sardellitti, G. Scutari, and S. Barbarossa, "Distributed joint optimization of radio and computational resources for mobile cloud computing," in *Cloud Networking (CloudNet), 2014 IEEE 3rd International Conference on*, pp. 211–216, IEEE, 2014.

[104] F. Guo, L. Ma, H. Zhang, H. Ji, and X. Li, "Joint load management and resource allocation in the energy harvesting powered small cell networks with mobile edge computing," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 299–304, IEEE, 2018.

[105] S. Sardellitti, G. Scutari, and S. Barbarossa, "Joint optimization of radio and computational resources for multicell mobile-edge computing," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 1, no. 2, pp. 89–103, 2015.

[106] S. Sardellitti, S. Barbarossa, and G. Scutari, "Distributed mobile cloud computing: Joint optimization of radio and computational resources," in *Globecom Workshops (GC Wkshps), 2014*, pp. 1505–1510, IEEE, 2014.

[107] C. Wang and S. Hu, "Efficient radio resource management for wireless cellular networks with mobile edge computing," *arXiv preprint arXiv:1706.09091*, 2017.

[108] X. Lyu, H. Tian, C. Sengul, and P. Zhang, "Multiuser joint task offloading and resource optimization in proximate clouds," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 4, pp. 3435–3447, 2017.

[109] T. X. Tran and D. Pompili, "Joint task offloading and resource allocation for multi-server mobile-edge computing networks," *arXiv preprint arXiv:1705.00704*, 2017.

[110] C. Wang, C. Liang, F. R. Yu, Q. Chen, and L. Tang, "Computation offloading and resource allocation in wireless cellular networks with mobile edge computing," *IEEE Transactions on Wireless Communications*, 2017.

[111] X. Chen, "Decentralized computation offloading game for mobile cloud computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 4, pp. 974–983, 2015.

[112] J. Opadere, Q. Liu, and T. Han, "Energy-efficient rrh sleep mode for virtual radio access networks," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pp. 1–6, IEEE, 2017.

[113] Q. Liu, T. Han, and G. Wu, "Computing resource aware energy saving scheme for cloud radio access networks," in *Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom), 2016 IEEE International Conferences on*, pp. 541–547, IEEE, 2016.

[114] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE transactions on wireless communications*, vol. 9, no. 5, 2010.

[115] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green cloud-ran," *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2809–2823, 2014.

[116] T. Baran, D. Wei, and A. V. Oppenheim, "Linear programming algorithms for sparse filter design," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1605–1617, 2010.