Spring 2020

# Topology Control, Scheduling, and Spectrum Sensing in 5G Networks

Prosanta Paul
*Old Dominion University*, ppaul001@odu.edu

# TOPOLOGY CONTROL, SCHEDULING, AND SPECTRUM SENSING IN 5G NETWORKS

by

Prosanta Paul
B.Sc. March 2009, Khulna University of Engineering & Technology
M.Sc. August 2014, South Dakota School of Mines & Technology

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

ELECTRICAL & COMPUTER ENGINEERING

OLD DOMINION UNIVERSITY
May 2020

Approved by:

Chunsheng Xin (Director)

Hongyi Wu (Member)

Oscar Gonzalez (Member)

Yanxiao Zhao (Member)

Frederic (Rick) D. McKenzie (Member)

# ABSTRACT

TOPOLOGY CONTROL, SCHEDULING, AND SPECTRUM SENSING IN 5G NETWORKS

Prosanta Paul
Old Dominion University, 2020
Director: Dr. Chunsheng Xin

The proliferation of intelligent wireless devices is remarkable. To address phenomenal traffic growth, a key objective of next-generation wireless networks such as 5G is to provide significantly larger bandwidth. To this end, the *millimeter wave* (mmWave) band (20 GHz - 300 GHz) has been identified as a promising candidate for 5G and WiFi networks to support user data rates of multi-gigabits per second. However, path loss at mmWave is significantly higher than today's cellular bands. Fortunately, this higher path loss can be compensated through the antenna beamforming technique–a transmitter focuses a signal towards a specific direction to achieve high signal gain at the receiver. In the beamforming mmWave network, two fundamental challenges are *network topology control* and *user association and scheduling*. This dissertation proposes solutions to address these two challenges. We also study a spectrum sensing scheme which is important for spectrum sharing in next-generation wireless networks.

Due to beamforming, the network topology control in mmWave networks, i.e., how to determine the number of beams for each base station and the beam coverage, is a great challenge. We present a novel framework to solve this problem, termed *Beamforming Oriented tOpology coNtrol* (BOON). The objective is to reduce total downlink transmit power of base stations in order to provide coverage of all users with a minimum quality of service. BOON smartly groups nearby user equipment into clusters to dramatically reduce interference between beams and base stations so that we can significantly reduce transmit power from the base station. We have found that on average BOON uses only 10%, 32%, and 25% transmit

power of three state-of-the-art schemes in the literature.

Another fundamental problem in the mmWave network is the *user association and traffic scheduling*, i.e., associating users to base stations, and scheduling transmission of user traffic over time slots. This is because base station has a limited power budget and users have very diverse traffic, and also require some minimum quality of service. User association is challenging because it generally does not rely on the user distance to surrounding base stations but depends on if a user is covered by a beam. We develop a novel framework for user association and scheduling in multi-base station mmWave networks, termed the *clustering Based dOwnlink user assOciation, Scheduling, beamforming with power allocaTion* (BOOST). The objective is to reduce the downlink network transmission time of all users' traffic. On average, BOOST reduces the transmission time by 37%, 30%, and 26% compared with the three state-of-the-art user scheduling schemes in the literature.

At last, we present a wavelet transform based spectrum sensing scheme that can simultaneously sense multiple subbands, even without knowing how the subbands are divided, i.e., their boundaries. It can adaptively detect all active subband signals and, thus, discover the residual spectrum that can be used by unlicensed devices.

# ACKNOWLEDGEMENTS

I am thankful to many people in my life who helped me throughout this long Ph.D. journey. Without their help and support, I could never accomplish this work.

At first, I would like to pay my special regards to my advisor, Dr. ChunSheng Xin, for allowing me to work in his research group and advised me with great understanding and patience during my entire Ph.D. study. I deeply thank Dr. Xin for his endless support and guidance, which always encouraged me to keep myself on the right path of my destination. I am honored to have such a humble and knowledgeable person as my Ph.D. advisor.

I would like to express my sincere gratitude to my co-advisor, Dr. Hongyi Wu, for helping me with countless technical discussions, research ideas, and guidance throughout my Ph.D. study at ODU. Without his guidance, it was almost impossible to complete my research. I also would like to thank my committee members Dr. Oscar Gonzalez, Dr. Yanxiao Zhao, and Dr. Rick McKenzie for their valuable time and consideration to advise me and review my dissertation. My special thanks to Dr. Yanxiao Zhao, my M.S. advisor, for encouraging me to pursue a Ph.D. in ODU.

I want to thank the Electrical and Computer Engineering Department and Old Dominion University Research Foundation for financial support during my Ph.D. study.

My special thanks to my labmate, Peng Jiang, for his true friendship, reliable support, and research discussions. It would be boring and tiresome in the lab without him. I also want to thank my fellow labmates Saygen Baksi, Md Sharif Ullah, and Md Moin Uddin Chowdhury for always being nice to me, giving technical and mental supports and helping me whenever I needed. I am thankful to all friends and families in the Bangladeshi community for extending their helping hands when I needed and making my life enjoyable in Norfolk.

Last, but not least, I wish to acknowledge the support and great love of my parents, wife, and siblings. They always believed in me.

# TABLE OF CONTENTS

Page

Chapter Page

# LIST OF TABLES

# LIST OF FIGURES

Figure Page

# CHAPTER 1

# INTRODUCTION

In recent years we have witnessed a remarkable proliferation of intelligent wireless devices. At the same time, mobile broadband services such as HD video streaming and virtual/augmented reality will continue driving the demand for higher data rates. For instance, by the Internet of Things (IoT) forecast of Ericsson, there will be 29 billion IoT devices by 2022, and mobile data traffic (especially video traffic) will grow by 35 percent annually through 2024 [8]. Next generation wireless networks such as *fifth-generation cellular wireless* (5G) will address the phenomenal growth of traffic demands by providing significantly higher wireless system capacity and reliable communications. 5G is more accurately a collection of technologies that connects a high volume of mobile users and provides seamless support for a massive number of applications. Fig. 1 illustrates a triangular use case diagram developed by the International Mobile Telecommunications (ITU) as 2020 performance targets, where 5G applications are categorized into three major classes: Enhanced Mobile Broadband (eMBB), Massive Machine-type Communications (mMTC), Ultra-reliable and Low-latency Communications (URLLC). Each class is seated in a corner of the triangle and can support different services with similar performance requirements. For example, industrial automation and mission-critical communications both require ultra-low-latency, smart city devices require massive machine type connectivity, and high quality entertainment and media need enhanced mobile broadband connections. Due to the broad range of technical and business opportunities, 5G is highly important for our society and economy. For instance, by the estimation of IHS Markit-a global information provider, the 5G value chain alone will drive $3.6 trillion of economic output and support 22.3 million jobs by 2035 [9]. To ensure ubiquitous, very high data rate, and faster communications, 5G needs to provide enormous system capacity.

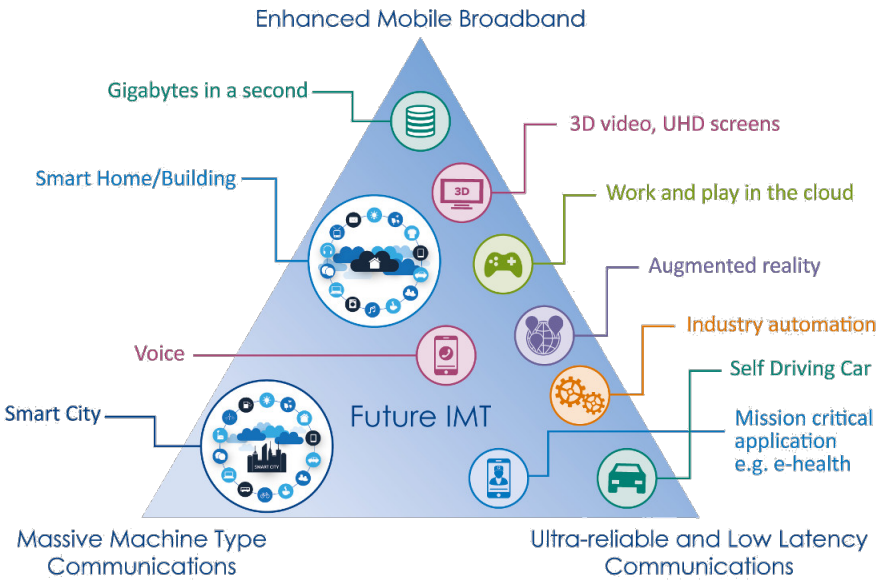
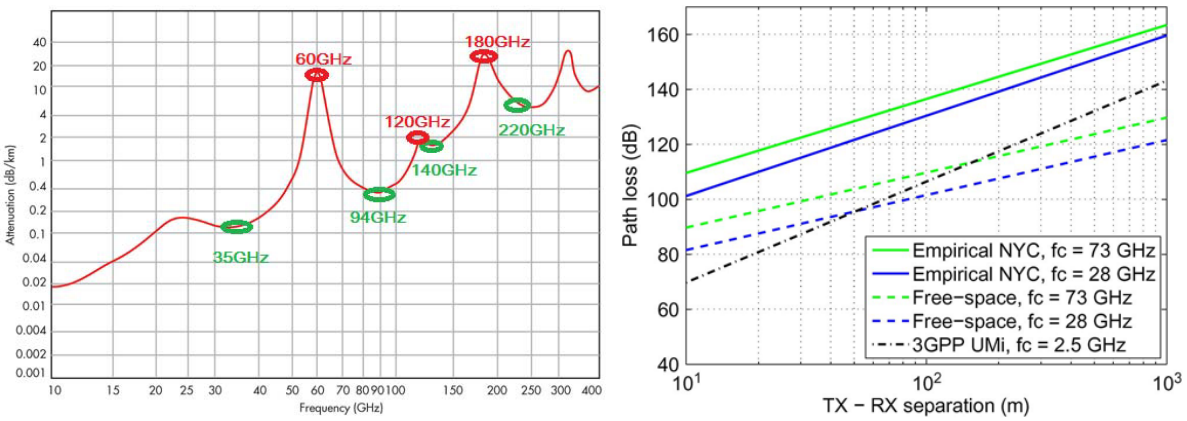



**Figure 1.** International Mobile Telecommunications (IMT)-2020 Performance Targets [2]

## 1.1 MILLIMETER WAVE (mmWave) FREQUENCY

More spectrum needs to be pressed into service to dramatically increase the system capacity in 5G. This is because while the wireless spectrum efficiency has improved continuously, such advances cannot meet the requirement for drastic growth in wireless capacity. Today's wireless systems mainly operate in the sub-6 GHz microwave spectrum, which is experiencing a severe shortage and has become a precious resource. To this end, massive spectrum between 20 and 300 GHz, known as *millimeter wave* (mmWave) band, has been identified as a potential candidate for 5G to support user data rates of multi-gigabit per second and a thousand-fold increase in system capacity [10–17].

While the mmWave band promises much higher bandwidth, it also raises a technical challenge. As illustrated in Fig. 2(a), signals at mmWave frequencies, especially 60 GHz, 120 GHz and 180 GHz with attenuation peak, suffer higher attenuation caused by the atmospheric oxygen absorption. The attenuation caused by atmospheric absorption is 0.012 dB




(a) (b)

**Figure 2.** Channel characteristics at mmWave frequencies: (a) Atmospheric and molecular absorption at mmWave frequencies in dB/km [3, 4] (b) Comparison of distance-based path loss between"Empirical NYC" at 28 and 73 GHz, "free space propagation" at 28 and 73 GHz, and "3GPP Urban Micro (UMi) model" for 2.5 GHz [5].

over 200 m at 28 GHz and 0.016 dB over 200 m at 38 GHz. Frequencies from 70 to 100 GHz and 125 to 160 GHz also have small loss [10]. Also, transmission at mmWave frequency suffers from significantly higher path loss because it is proportional to the square of the frequency in accordance with the Friis equation (power received $\propto$ power transmitted/frequency$^2$). For instance, as shown in Fig. 2(b), path loss measure at 73 GHz is about $30 - 40$ dB higher than today's cellular bands at 2.5 GHz [11, 12]. Furthermore, penetration loss, diffraction loss, etc., are also very high at mmWave frequency. For instance, signal attenuation from mortar, brick, and concrete is $40 - 80$ dB, the human body and furniture are up to $20 - 35$ dB. The rain attenuation at 28 GHz has an attenuation of 7 dB/km for very heavy rainfall of 25 mm/hr (about 1 inch per hour). If cell coverage regions are 200 m in radius, the rain attenuation will reduce to 1.4 dB [10]. For the above reasons, mmWave communication links are typically intermittent and the transmission range of the mmWave system is significantly shorter than the current microwave system. Therefore, we have to adopt new techniques to reduce losses in the channel and exploit the potentials in the mmWave system.

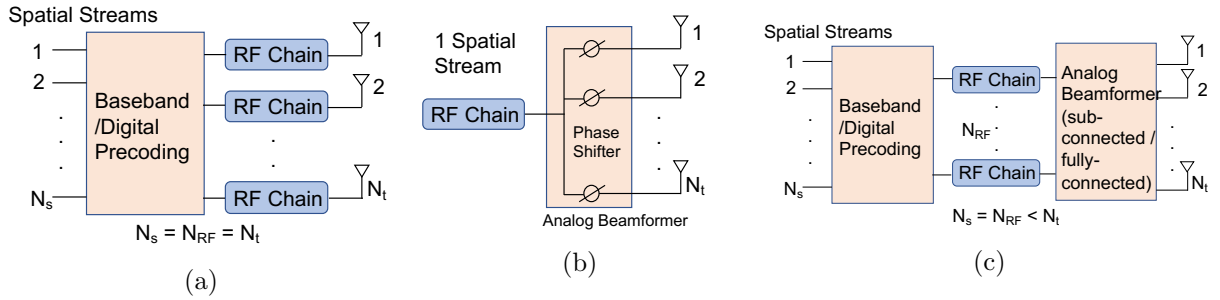



**Figure 3.** Beamforming architectures: (a) Fully-digital beamforming, (b) Analog beamforming, (c) Hybrid beamforming.

## 1.2 BEAMFORMING IN mmWave

To make mmWave communications practical, i.e., achieve a practical range under such a severe path loss, a critical technology, *beamforming*, is needed. It uses a reconfigurable antenna array and controls the amplitude and phase of the signal at each antenna element to concentrate the transmission power on a narrow beam toward the receiver, to result in a high signal gain. At the receiver, all received signals are coherently combined using a different scale factor. As a result, it dramatically increases the *signal to interference and noise ratio* (SINR) at the receiver, thus, mitigates the higher path loss at the mmWave band. This gain in SINR in antenna array systems is called beamforming gain. Owing to the short wavelengths in the mmWave system, it is possible to pack a large number of antennas in a small physical dimension (as compared to the antenna arrays at current 3G or LTE frequencies), thus forming narrow beam to achieve high signal gain and very small interference in the network. Depending on the physical design of the trans-receiver architecture, beamforming can be categorized into three major categories: Digital, Analog, and Hybrid beamforming. Next, we will briefly discuss these beamforming architectures and their use cases.

### 1.2.1 DIGITAL BEAMFORMING

Digital beamforming architecture is illustrated in Fig. 3(a), where each antenna element in the array requires a dedicated baseband-to-RF chain (simply RF chain). The number of spatial streams (beams) is equal to the number of RF chains. This beamforming architecture is highly flexible to form a sharp beam for very high data rate applications. However, due to a large number of the RF chain, this architecture is infeasible because cost, power consumption, and complexity are very high and scale up with the number of antenna elements and operating bandwidth [18]. Note that each RF chain has an analog-digital converter (ADC) whose costs also scale up with its resolution (specified in bits to represent the ADC measurement) [19–21]. Channel estimation for digital beamforming is also complex, especially for a large number of antenna elements, because every antenna pair between the transmitter and receiver requires us to estimate the channel for digital precoding [22, 23]. Therefore, for a wide bandwidth system such as mmWave and large scale antenna arrays such as massive MIMO (multiple-input and multiple-output), fully digital beamforming is not suitable due to high cost and substantial power consumption.

### 1.2.2 ANALOG BEAMFORMING

Analog beamforming architecture is shown in Fig. 3(b). Unlike digital beamforming, one RF chain is shared among all antenna elements in the array that can form only one beam. The signal amplitude at each element is constant, and only phase changes to change the direction of the beam. This architecture shapes the output beam using the digital phase shifters connected to each antenna. Electronic switches or lens antenna can be also be used to replace the digital phase shifters (see [24] more details). This technique is extensively studied in the MIMO system and already established in many existing standards including mmWave, WLAN, WPAN, etc.

On the positive side, the requirement for instantaneous channel information at antenna

arrays can be waived in analog beamforming. Generally, a finite size codebook is designed to cover certain directions and spatial beam search techniques are used to find the optimal beam direction between transmitter and receiver [25–27]. Analog beamforming is simple to design with very low complexity but performance is lower than digital beamforming. In this technique, only one beam can be formed at a time and user multiplexing within a beam is not possible, i.e., analog beamforming provides only directivity gain.

### 1.2.3 HYBRID BEAMFORMING

Hybrid beamforming is a two-stage hybrid (digital and analog precoding) beamforming architecture as illustrated in Fig. 3(c). This is a promising beamforming design for large scale arrays in the mmWave system, where the number of RF chains is less than the number of antenna elements. Here, precoding is performed at both radio frequency (analog) and baseband (digital) domains. This architecture is well-adopted in mmWave based next generation mobile networks including 5G. Hybrid beamforming is further sub-divided into two types according to the design in the RF domain; sub-connected hybrid beamforming and fully connected hybrid beamforming. In sub-connected hybrid beamforming, a separate antenna array (a subset of an array) is connected to individual the RF chain, i.e., each antenna element belongs to only one RF chain. On the other hand, each RF chain is connected to all antennas in the fully connected hybrid beamforming architecture (see [24,28,29] for more details). In this architecture, digital precoding uses the channel matrix and analog beamforming weight to achieve an effective channel gain at the receiver. Here, the analog RF domain provides directivity gain in each spatially separated beam which helps to compensate the severe path loss in mmWave systems. With the help of advanced signal processing techniques, interference between adjacent RF chains is taken care of in the digital baseband domain. The throughput performance of hybrid beamforming can be very close to the digital beamforming, with fewer RF chains [30]. Due to the reduced energy consumption and processing overhead, hybrid beamforming architecture is a popular and widely adopted choice

for large scale applications. This dissertation has adopted this beamforming architecture as well. However, in the rest of the dissertation, "hybrid beamforming" is referred to as simply "beamforming".

While beamforming can significantly increase the reach of *base stations* (BSs), it raises a great challenge for forming network topology, in particular for multi-BS mmWave networks. In this dissertation, *topology control* in the context of a beamforming network is referred to as the discovery of the number of beams in each BS, the number of *user equipment* (UE) per beam, and guaranteeing the coverage of all UEs while maintaining a minimum quality of service at each UE. Topology control is not a trivial task due to different intermingled factors including harmful interference between beams in BS. In this dissertation, we first identify the unique challenges in topology control and then discuss a heuristic, yet efficient, solution for multi-BS networks. Next, we discuss *UE association and scheduling* in dense mmWave networks, where there are a large number of UEs but BSs have insufficient resources (e.g. transmit power and beamformer in each BS) to cover all UEs simultaneously. In this case, BSs need to smartly associate UEs and pick a set of UEs for each BS to schedule their traffic in each time slot. The UE association and scheduling jointly affect the traffic scheduling performance but are highly challenging tasks due to important factors including beam interference, diverse UE traffic, and their distribution, etc. We discuss a novel framework for UE association and traffic scheduling in the second part of the dissertation. In the last part of the dissertation, we discuss a novel spectrum sensing technique for a subdivided licensed band to enable efficient spectrum sharing in 5G massive and heterogeneous applications. This technique allows an unlicensed device to smartly utilize the residual spectrum in the licensed band. We are going to introduce the details of each problem in the next few sections.

## 1.3 TOPOLOGY CONTROL IN BEAMFORMING NETWORKS

Topology control is challenging because we have to design beams to cover all UEs while avoiding strong interference to each UE from other beams of the same BS or different BSs.

This results in a joint beamforming optimization problem among BSs, including UE association among BSs and inter-beam and inter-BS interference reduction. This problem is fundamentally challenging. Without a careful design, the system performance such as the sum rate can be poor while consuming significant resources such as transmit power.

Topology control is basically a framework for discovery of the number of beams for each BS and UE coverage by each beam, and multi-BS beamforming in a mmWave network. In traditional cellular networks, the omnidirectional antenna is used and the coverage of a BS is a circular/hexagonal area (cell) based on the communication range of the BS. The topology control is thus straightforward since it is easy to determine the preferred BS for a UE as well as the preferred UEs for a BS. Unlike traditional cellular networks, a cell or the coverage area of a BS in mmWave networks is complicated and usually has an irregular shape. The coverage area of a BS is essentially the area covered by the beams formed by the BS. The network connection between a BS and a UE relies on whether there is a beam from the BS to cover the UE, rather than the distance between them. On the other hand, where and how to form beams depends on several factors such as the UE spatial distribution, interference between beams, power budget, etc. Hence, topology control in mmWave networks is fundamentally challenging. This dissertation discusses a unique topology control framework to solve these problems.

## 1.4 UE ASSOCIATION AND SCHEDULING

While beamforming significantly increases the reach of a BS in the mmWave networks, it raises a great challenge for *UE association and scheduling*, which is a critically important problem. In mmWave networks, the UE association generally does not rely on the distance to surrounding BSs, but depends on if a UE is covered by a beam. This makes the UE association and beamforming a joint optimization problem. The problem is even more challenging as a BS usually cannot support all beams to cover its UEs simultaneously. Instead, it has to schedule its UEs/beams across multiple time slots, because of the limitation on the power

budget, as well as the limited number of beamformers due to the high cost of RF chains. Note that traditional cellular networks also schedule UEs across time slots, but that is due to the limited spectrum bandwidth at sub-6 GHz cellular bands.

There are also other critical factors we need to consider during the UE association and scheduling such as UE traffic load, the quality of service requirement for UEs, beam design, interference, power and spectrum allocation among UEs in a beam, etc. Those factors are crucial and intermingled, which makes the UE association and scheduling a fundamentally challenging joint optimization problem. Here, we develop a unique framework for the UE association and scheduling, with an objective to reduce the overall downlink latency or traffic transmission time in the network. Traffic latency is an important feature in 5G as a large number of applications (e.g. industry automation, self-driving car, mission-critical applications, etc.) are delay sensitive and require ultra-low latency communications. The overall transmission time in a network is the total time required to deliver all UEs traffic. However, a UE traffic transmission time is simply the ratio of its traffic size and data rate which further is a function of spectrum allocation and received signal to interference and noise ratio (SINR) at UE. Moreover, a UE SINR relies on the received beamforming gain and interference, and power allocation in the beam, etc. Therefore, a UE traffic transmission time highly relies on its association and scheduling in a BS. In the second part of the dissertation, a novel framework is designed for the UE association and scheduling, which objective is to reduce the overall transmission time in the network while maintaining a minimum SINR at each UE.

## 1.5 SUBDIVIDED BAND SPECTRUM SENSING

One of the biggest challenges for massive connectivity in 5G is the lack of available spectrum, especially for emerging large scale IoT devices and machine type communications as illustrated in Fig. 1. Dynamic spectrum sharing between different technologies can relieve (partially) spectrum scarcity in 5G. Spectrum sensing can be used for spectrum sharing,

which is usually implemented in the *cognitive radio* (CR) platform. In CR, *secondary users* (SUs) do not have a license to use the spectrum, but they sense the spectrum environment and dynamically adapt the transmission or reception parameters, and channel access method to access the partially or completely unused spectrum without causing harmful interference to *primary users* (PUs) [31–35]. If the PU band is *subdivided*, it is challenging for SUs to detect which sub-bands are occupied by PUs, especially when SUs do not have the knowledge of the subdivision. For instance, 2.4 GHz WiFi band is divided into 14 channels each with 22 MHz. In fact, many wireless communication systems, such as Bluetooth and cellular systems, subdivided its band into a number of channels [36].

Most of the existing spectrum sensing techniques focus on opportunistic access of SUs, i.e., SUs access the PU band when it is absent. The great advantage of a subband spectrum sensing system is that SUs can sense active subbands by a PU and utilize the unused subbands to avoid frequent channel switching which can significantly disrupt the SU communications. However, classical spectrum sensing techniques do not work well in such subdivided licensed bands. In the last part of the dissertation, we discuss a spectrum sensing technique that addresses unique challenges in subdivided licensed band spectrum sensing.

## 1.6 DISSERTATION CONTRIBUTIONS

This dissertation is comprised of three major contributions for future wireless communication systems such as 5G. **The first contribution** is that a novel topology control framework, termed *Beamforming Oriented tOpology coNtrol* (BOON), is devised to provide reliable coverage to all UEs in multi-BS mmWave networks. The objective of BOON is to reduce the total downlink transmit power and interference between beams, and achieve high sum rate with a minimum SINR at each UE. BOON smartly groups nearby UE into clusters, constructs sets from UE clusters, and associates UE to BSs and beams. Here, we develop a beamforming oriented UE clustering algorithm to group UEs with respect to each BS in the network. The objective of the algorithm is to reduce transmit power and interference

between beams. Later, we devise a set construction scheme and an interference aware set covering algorithm that addresses the unique challenges in beamforming to smartly associate UEs to beams and BSs. Finally, we design an approach to form beams for all BSs with the objective to reduce the total transmit power and inter-BS interference, subject to coverage of all UEs and meeting a minimum quality of service.

**The second contribution** is that we devise a novel framework for UE association and scheduling in BS power limited dense mmWave networks, termed the *clustering Based dOwnlink UE assOciation, Scheduling, beamforming with power allocaTion* (BOOST). The objective is to reduce the transmission time for all UE traffic, subject to the power budget of BSs, number of beamformers of BSs, UE traffic loads, and the required minimum SINR at each UE. BOOST jointly considers UE association among multiple BS, UE scheduling, and beamforming altogether, to reduce the transmission time of the entire network. In the BOOST framework, at first, we develop a clustering algorithm that focuses on unique features of beamforming to group UEs into clusters to reduce interference for beamforming. Next, we design a novel UE association scheme that effectively reduces interference and balances UE traffic loads between clusters and BSs, to decrease the transmission time. Finally, a unique scheme is developed for joint beamforming, power allocation, and UE traffic scheduling to reduce the transmission time, subject to the BS power budget, UEs traffic loads, and the minimum SINR requirement at UEs.

**The third contribution** is that a novel spectrum sensing technique is devised to use in unlicensed devices for discovering residual spectrum in a subdivided license band, called *exponentially moving average based multi-scale summation* (EMAMS). The objective of EMAMS is to detect all active sub-bands in a PU (license) band so that SUs (unlicensed users) can utilize the residual spectrum in the PU band to significantly increase spectrum efficiency in 5G. EMAMS does not depend on the number of used sub-bands and adaptively detects active signals in a spectrum band measured by a spectrum analyzer. Instead of using synthetic PU signals such as raised cosine, rectangular pulse, sinusoidal, we use the real world signals

measure by a spectrum analyzer for performance evaluation.

## 1.7 ORGANIZATION OF THE DISSERTATION

The remainder of the dissertation is organized as follows. In Chapter 2, we discuss the BOON framework to minimize the downlink transmit power for UE coverage in mmWave systems. This chapter presents a detailed architecture of the BOON framework that includes several heuristic approaches for UE coverage with an optimal beamforming scheme. At the end of this chapter, the performance of BOON is evaluated and compared with the three state-of-the-art techniques. Chapter 3 presents the details of the BOOST framework, which is designed for UE association and scheduling to jointly reduce the overall traffic transmission time in dense multi-BS mmWave networks. This chapter demonstrates the formulation of the joint optimization problem, several heuristic algorithms such as UE clustering, traffic balancing for UE association, and UE traffic scheduling with optimal beamforming. Lastly, we evaluate the performance of BOOST and compare the results with the three state-of-the-art schemes in the literature. In Chapter 4, we present the EMAMS scheme that is designed for spectrum sensing in the subdivided license band. Initially, we discuss several classical spectrum sensing techniques and their performances for detecting active subbands within a licensed band. Later, the EMAMS algorithm is discussed in details and a comparative study is performed with an existing solution using measured-real world signals by a spectrum analyzer. Finally, Chapter 5 summarizes the dissertation with conclusions and future research directions.

# CHAPTER 2

# BEAMFORMING ORIENTED TOPOLOGY CONTROL

In a beamforming based mmWave network, we refer to *topology control* as the discovery of the number of beams for each BS and UE coverage by each beam, and multi-BS beamforming. In this chapter, we discuss a topology control framework, called *Beamforming Oriented tOpology coNtrol* (BOON). The objective of BOON is to reduce the total downlink transmit power of BSs and interference between beams while covering all UEs in the network and maintaining a minimum signal to noise and interference ratio (SINR) at each UE receiver. Four components of BOON are UE clustering, set construction from cluster, a weighted set covering, and computation of interference aware beamforming weight vector at the BS antenna. Each component has a specific role in BOON to reduce the overall transmit power and beam interference. We discuss each component and develop two algorithms: one for clustering and the other for the set covering. Toward the end, we evaluate the overall performance of BOON in terms of transmit power, network sum rate, signal to interference and noise ratio, and computation complexity. We compare BOON with three existing topology control schemes.

The rest of the chapter is organized as follows. Section 2.1 discusses the fundamentals and background of topology control in the context of mmWave networks. Section 2.2 presents an overview of the BOON framework. Section 2.3 describes the system model. The BOON framework is presented in Section 2.4 along with in-depth analysis. The numerical results are presented in Section 2.5. Finally, Section 2.6 concludes the chapter.

## 2.1 TOPOLOGY CONTROL FUNDAMENTALS

Topology control is a challenging task in beamforming based mmWave networks. It is straightforward in traditional cellular networks as BS coverage is usually a circular or

hexagonal area which is basically determined by the BS transmit power and the distance between BS and UE. On the contrary, BS coverage in mmWave networks is complicated and usually has an irregular shape. A BS coverage area is essentially the area covered by its beams, which depends on a number of factors including transmit power in a beam, UE spatial distribution, interference between beams, etc. Therefore, topology control in mmWave networks is fundamentally a challenging problem.

A straightforward approach to form network topology for mmWave networks is to construct a beam to each UE through *multi-user beamforming* (MUB) [37–41]. However, forming one beam for each UE can result in significant interference between nearby UEs, thus degrading system performance such as the *signal to interference and noise ratio* (SINR) at the UEs. Furthermore, as we discussed earlier in Section 1.2, each beam needs a RF chain or *beamformer*, which is prohibitively costly or impossible in large scale mmWave networks. Another straightforward approach for topology control is to construct one beam to cover all UEs associated with a BS, through *multicast beamforming* (MCB) [42–44]. However, this approach cannot form sharp beams, resulting in wasteful energy consumption. Fig. 4 illustrates the MCB and MUB beam patterns for two UEs with $5^o$ angular separation with respect to a BS, and their average SINR as a function of their separation angle. The BS transmit power is 4 watt, the UE-BS distance is 100 meters, and the number of antenna elements in BS is 16. Clearly, for MUB, a large separation angle results in a good SINR, while a small separation angle results in a poor SINR, due to the strong interference between the two beams. In contrast, for MCB, a small separation angle results in a good SINR, while a large separation angle results in a poor SINR because beams are not formed to concentrate power on UEs.

The BOON framework is designed to combine the benefits of MUB and MCB based topology control approaches. BOON smartly clusters UEs into groups and form beams based on the groups. Thus, nearby UEs are covered using the same beam to avoid interference, while UEs sufficiently separated, i.e., from different groups, are covered by different beams, to

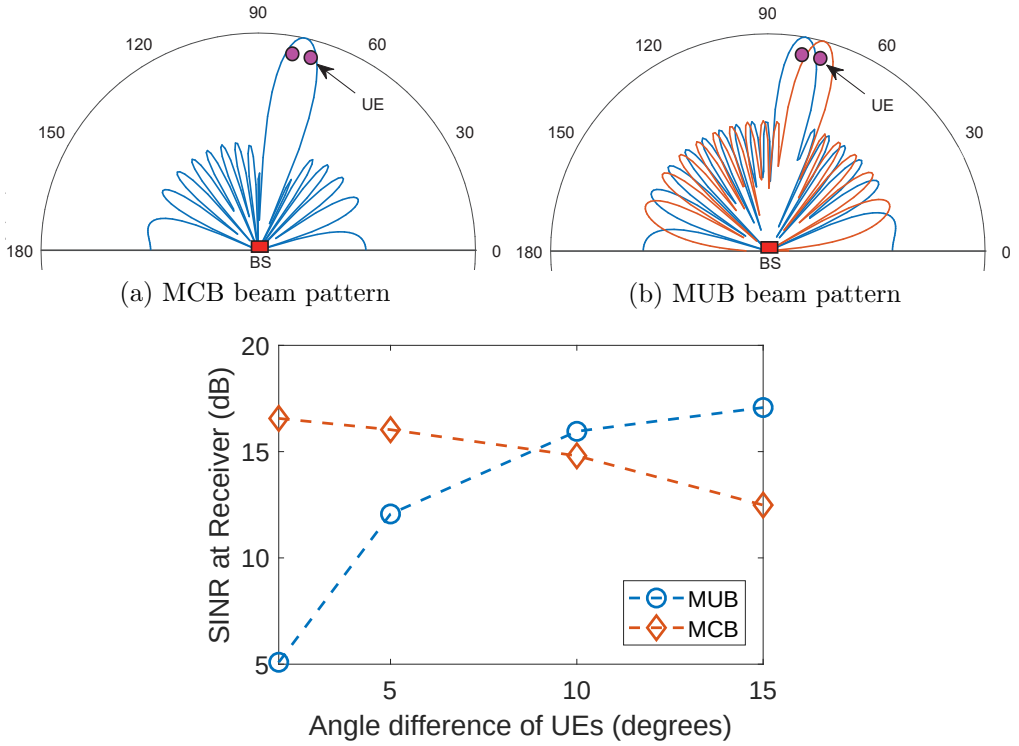



**Figure 4.** Beam patterns and average SINR: (a) MCB beam pattern and (b) MUB beam pattern for two UEs with $5^o$ angle difference with respect to the BS (c) average SINR of UEs as a function of their angular separation relative to the BS [1].

avoid wasteful energy consumption. Furthermore, BOON smartly associates UEs to different BSs and beams, and finds appropriate beamforming vectors for all BSs and beams, with the objective to reduce the interference between BSs and beams. The objective of BOON is to reduce the total transmit power. BOON can significantly reduce mutual interference between nearby UEs and between BSs, resulting in lower power consumption, and can reduce system costs by using significantly fewer beams, compared with the MUB based topology control. On the other hand, it can form sharp beams to focus energy on the UEs to reduce power consumption, compared with the MCB based topology control.

There have been several studies on topology control or UE selection in MIMO systems

in the literature. In [37], the authors studied a UE selection method based on channel orthogonality between different UEs, to select a group of UEs with orthogonal channel vectors, which are then served using the MUB scheme. In [38], the authors discussed a user selection scheme to maximize the harmonic sum of UE SINR, which prioritizes cell-edge users to increase throughput, where a UE can receive multiple spatial streams from different BSs. In [45], UEs are partitioned into UE groups, with each group having similar channel covariances. All UEs in a group are served simultaneously using one precoding vector through an MCB scheme. In [46], the authors studied a simpler beamforming scheme called switched beamforming. In [39–41, 47], the UEs with uncorrelated or orthogonal channels are grouped together, and then served using the MUB scheme. In [48, 49], a BS forms a beam in a random direction, which can cover UEs in that direction. Most of those works focused on beamforming with one BS, while the UE association among multiple BSs assumed a simple scheme based on the strongest received signal power. Here BOON not only considers UE selection for a single BS, but also aims to optimize UE association among BSs, with the objective to reduce interference between beams of different BSs.

## 2.2 BOON OVERVIEW

BSs are deployed by wireless operators and are generally connected to a cloud or a backend management system, through either wired or wireless backhaul connections. Hence, the BSs information, including locations, the number of antennas, etc., is known to the network operator. We assume a low mobility UE network where the updating of UE locations at the BS or the cloud can be on the order of seconds. For a walking speed mobile UE, the channel fading characteristics remain unchanged for several seconds [45,46]. This means UE tracking can be done at a slow rate. The message exchange delay is expected to be small, e.g., in the order of 1 millisecond in 5G networks. In a low mobility network, the BOON algorithm can be re-run at a frequency in the order of seconds. Both the message exchange delay and running time of BOON are acceptable compared with such an interval. On the other hand,

mmWave networks are not well suited for high mobility UE networks [45]. This is because the Doppler shift in wireless channels increases linearly with the carrier frequency and UE mobility, and the expected angular spread of UE signals at the BS is also higher in mmWave frequency [14, 45].

In general, there are two approaches for a BS to discover UEs. The first approach is to utilize *beam sweeping* [50, 51]. Each BS transmits initial signals in random directions to sweep the whole angular space, to discover UEs. This approach has a lower system cost, but may result in a large delay. The second approach utilizes co-existing macrocells for UE discovery, e.g., LTE towers, to achieve low delay [18]. Given that major cellular operators all have deployed LTE systems, this approach is a feasible solution. Both LTE towers and mmWave BSs are connected to a cloud. The mmWave system off-loads data transport, while the LTE system is used for control and data transport for UEs that cannot be covered by mmWave BSs. The discovery of UE locations results in some overhead. Nevertheless, this overhead is not significant, as the message exchange between LTE towers and UEs is fast.

Through UE discovery, the UEs information, including locations, the *angle of arrival* (AoA) of UE signals, etc., is also known to the mmWave network operator. BOON resides in the cloud of the operator to coordinate topology control among BSs and UEs. It finds the optimal number of beams for each BS, assigns UEs to be covered by each beam, and computes the beamforming vectors for all beams. The objective is to reduce the total transmit power and interference between beams and between BSs, while covering all UEs and meeting a minimum quality of service. As illustrated in Fig. 5, there are four components in BOON: 1) *beamforming oriented UE clustering* (BOC), 2) *cluster decomposition based set construction* (CDSC), 3) *interference aware set covering* (IASC), and 4) *multi-BS beamforming*. BOC carries out fine granularity UE clustering for each BS among all UEs that can be reached through beamforming by the BS. Both the angles of UEs relative to the BS and the distance between UEs are considered in the clustering similarity metric. CDSC decomposes each cluster formed by BOC into a list of overlapping sets, which can be viewed as a list of

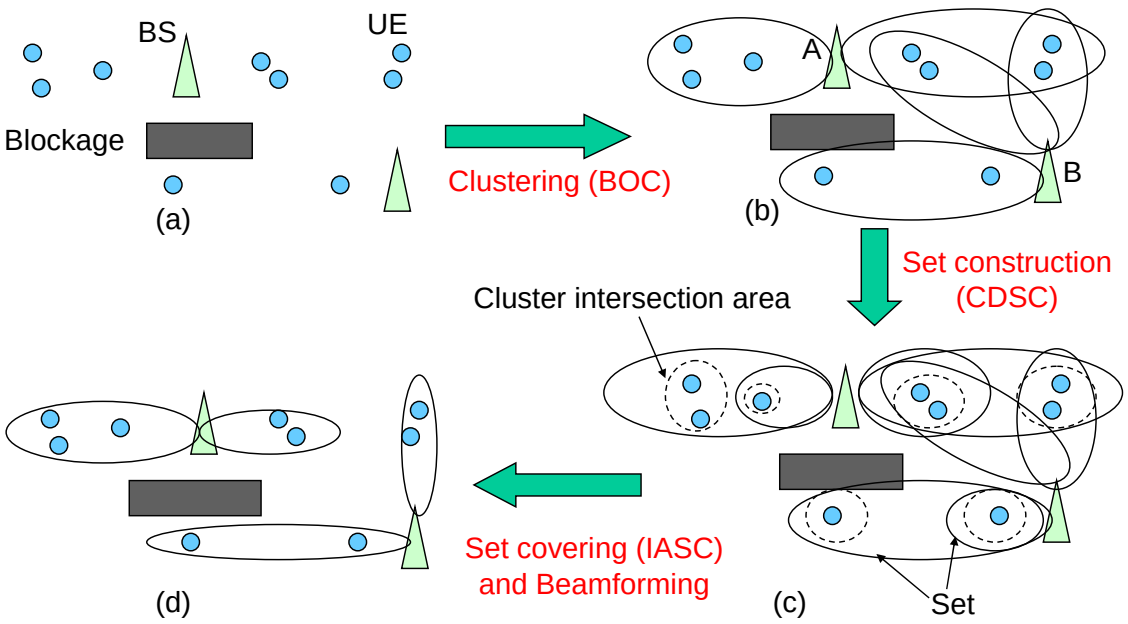



**Figure 5.** BOON: (a) UE clustering through the BOC algorithm, (b) set construction (CDSC), (c) set covering to optimally select UE-BS association and select sets for beamforming through the IASC algorithm, (d) beamforming for all BSs based on the selected sets [1].

overlapping sectors, starting from a BS outward. IASC addresses the unique challenges in beamforming and solves a *weighted set covering* problem to select a list of sets among all sets constructed by CDSC, to jointly cover all UEs, with the objective to reduce interference between beams and between BSs. The resulted sets from IASC determine the UE association and BS beam coverage. At last, one beam is formed for each set selected by IASC, with the objective to reduce the total transmit power of all BSs. The BOC algorithm can be run either distributedly at each BS, or at the cloud. The CDSC, IASC, and beamforming can be centrally run in the cloud. However, they can also be implemented in a distributed mode by exchanging the UE clustering information between BSs.

## 2.3 SYSTEM MODEL

In this section, we describe the system model for BOON, including mmWave channel

**Table 1.** Important notations in BOON

| | |
|---|---|
| $m, n, k$ | index for BS, UE and beam, respectively |
| $B$,$\mathcal{U}, C, \mathcal{B}$ | set of BSs, UEs, beams, and UE-clusters, respectively |
| $\langle m, k \rangle$ | the $k$-th beam of the $m$-th BS |
| $\theta_{mn}, \mathbf{a}_{\theta_{mn}}$ | angle of arrival of UE $n$ signal at BS $m$ array and the corresponding steering vector, respectively |
| $\mathbf{h}_{mn}$ | channel gain vector between BS $m$ and UE $n$ |
| $\mathbf{w}_{mk}$ | beamforming weight vector of beam $\langle m, k \rangle$ |
| $\phi_{mk}^{hp}$ | half power beamwidth of beam $\langle m, k \rangle$ |
| $p_{mk}^{t}$ | transmit power of beam $\langle m, k \rangle$ |
| $p_{n,m,k}^{r}$ | received power at UE $n$ from beam $\langle m, k \rangle$ |
| $K$ | maximum number of beamformers at a BS |
| $I_{n,m,k}$, $r_{n,m,k}$, $\gamma_{n,m,k}$ | interference, rate, and SINR at UE $n$ when it is served by beam $\langle m, k \rangle$ |
| $\gamma_o$ | minimum required SINR for UEs |

modeling, transmit power computation, and UE downlink data rate computation. Table 1 lists major notations used in this section.

### 2.3.1 mmWave CHANNEL

Let $\mathcal{U}$, $B$ and $C_m$ be the set of UEs, BSs, and beams of BS $m$ ($m \in B$), respectively. Let $\mathcal{U}_m \subset \mathcal{U}$ be the set of UEs that can be covered by BS $m$, with the maximum range decided by a beam with the minimum beamwidth and maximum BS transmit power. Throughout this dissertation, we assume each BS is equipped with an antenna array with $L$ antenna elements, and each UE has a single antenna. Note that BOON can be extended to accommodate UEs with multiple antennas.

Let $K$ be the number of beamformers for each BS. Each beamformer can form one beam. Note that the required processing power, design complexity and fabrication cost of a BS grows with the number of beamformers; hence a BS can only have a limited number of beamformers. For each beam $k \in C_m$, let $\mathcal{U}_{mk} \subseteq \mathcal{U}_m$ be the set of UEs that are inside beam $k$ of BS $m$, called beam $\langle m, k \rangle$.

Many works on mmWave channel modeling consider both LOS and *non-line-of-sight* (NLOS) paths [5, 10–12, 45]. Nevertheless, multiple channel measurement experiments on mmWave frequency showed that the LOS path dominates the NLOS paths. For instance, the NLOS path gain is typically 20 dB weaker than the LOS path gain [5, 49, 52, 53] and the NLOS path exponent was found as large as 5.76 in downtown New York City [10]. In BOON, we focus on the LOS paths to streamline the algorithm development. NLOS paths can be addressed to extend this work in future.

Let $d_{mn}$ and $\rho_{mn}$ denote the distance, and average LOS path loss, respectively, between BS $m$ and UE $n$. Note that $\rho_{mn}$ is often modeled as proportional to $d_{mn}^{-\eta}$, where $\eta$ is the path loss exponent of the LOS path. Also let $\theta_{mn}$ be the normalized direction of the LOS path to reach UE $n$ from BS $m$, as illustrated in Fig. 6, and $\alpha_{mn}$ be the complex gain of the LOS path which is modeled using a complex Gaussian distribution, i.e., $\alpha \sim \mathcal{CN}(0, 1)$. Now the channel vector between the $m$-th BS with $L$ antennas and the $n$-th UE ($n \in \mathcal{U}_{mk}$) with a single antenna is given as [5, 48, 49]

$$\mathbf{h}_{mn} = \sqrt{\frac{L}{\rho_{mn}}} \alpha_{mn} \mathbf{a}_{\theta_{mn}}, \tag{1}$$

where $\mathbf{a}_\theta$ is the response vector of the BS antenna array for the signal path in the $\theta$ direction. The response vector for a *uniform linear array* (ULA) at a BS is expressed as [5, 48, 49]

$$\mathbf{a}_\theta = \frac{1}{\sqrt{L}} \left[1, \quad e^{-j\pi\theta}, \ldots e^{-j\pi(L-1)\theta}\right]^T, \tag{2}$$

where $(\bullet)^T$ denotes transpose of a vector. Note that the normalized direction $\theta$ is a function of the physical *angle of departure* of the BS antenna array, denoted by $\phi \in [-\pi/2, \pi/2]$, as $\theta = \frac{2\Delta \sin(\phi)}{\lambda}$, where $\lambda$ is the channel frequency wavelength and $\Delta$ is the spacing between two adjacent antenna elements in a ULA. Concatenated channel matrix formed by all UEs in BS $m$ is therefore written as

$$\mathbf{H}_m = [\mathbf{h}_{m1}, \mathbf{h}_{m2}, \ldots, \mathbf{h}_{mn}], \quad n \in \mathcal{U}_m. \tag{3}$$

The received signals in all UEs under BS $m$ is given by

$$\mathbf{Y}_m = \mathbf{H}_m^H \mathbf{W}_m \mathbf{x}_m + \mathbf{z}, \tag{4}$$

where $(\bullet)^H$ denotes the complex conjugate transpose, $\mathbf{x}_m \in \mathbb{C}^{|\mathcal{U}_m| \times 1}$ is the vector of transmitted signals for UEs in BS $m$, $\mathbf{z} \in \mathbb{C}^{|\mathcal{U}_m| \times 1}$ is the vector of Additive White Gaussian Noise (AWGN) with zero mean and unit variance, $\mathbf{W}_m \in \mathbb{C}^{L \times |C_m|}$ is the beamforming weight matrix formed by concatenating weight vectors, $\mathbf{w} \in \mathbb{C}^{L \times 1}$, in BS $m$, i.e., $\mathbf{W}_m = \left[\mathbf{w}_{m1}, \mathbf{w}_{m2}, \ldots, \mathbf{w}_{m|C_m|}\right]$.

### 2.3.2 DOWNLINK TRANSMIT POWER

In BOON, each BS constructs a set of unit-power beams, denoted by $\{\mathbf{w}_{m1}, \ldots, \mathbf{w}_{m|C_m|}\}$, i.e., the power in each beam $||\mathbf{w}||^2 = 1$. As illustrated in Fig. 6, each beam serves a group of UEs. Let $\hat{\theta}_{mk}$ denote the main lobe direction (boresight angle) of the beam $\langle m, k \rangle$ formed by $\mathbf{w}_{mk}$ to serve all UEs in $\mathcal{U}_{mk}$. The effective channel gain of UE $n \in \mathcal{U}_{mk}$ from beam $\langle m, k \rangle$ can be written as [48, 49]

$$|\mathbf{h}_{mn}^H \mathbf{w}_{mk}|^2 \approx \frac{|\alpha_{mn}|^2}{\rho_{mn}} F_L(\hat{\theta}_{mk} - \theta_{mn}), \tag{5}$$

where $(\bullet)^H$ denotes the complex conjugate transpose, $\alpha_{mn}$ is the complex gain of the LOS

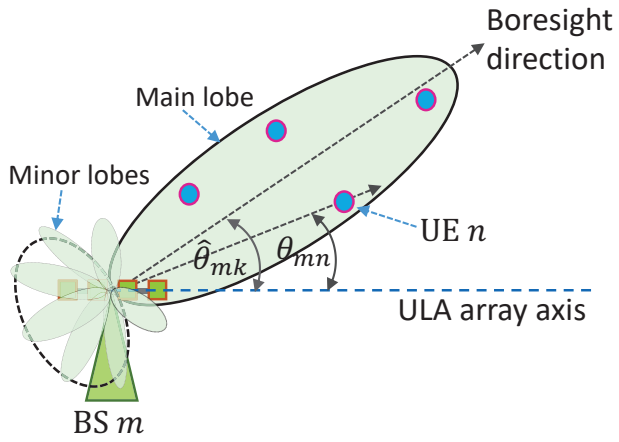



**Figure 6.** One beam to cover a group of nearby UEs. $\hat{\theta}_{mk}$ and $\theta_{mn}$ denote the boresight direction of beam $\langle m, k\rangle$ and the direction of UE $n$, respectively [1].

path between BS $m$ and UE $n$, $\rho_{mn}$ is the average LOS path loss between BS $m$ and UE $n$, and $F_L(x)$ is the Fejér kernel [48]. If $(\hat{\theta}_{mk} - \theta_{mn}) \to 0$, the UE channel gain is at maximum.

Let $p^t_{mk}$ be the transmit power of beam $\langle m, k\rangle$. The received signal power at UE $n$ from beam $\langle m, k\rangle$, denoted as $p^r_{n,m,k}$, is written as

$$p^r_{n,m,k} = p^t_{mk} |\mathbf{h}^H_{mn} \mathbf{w}_{mk}|^2, \quad \forall n \in \mathcal{U}_{mk}. \tag{6}$$

Let $I_{n,m,k}$ denote the total interference at UE $n$ from all beams, if it is covered by beam $\langle m, k\rangle$. Let $\sigma^2$ be the thermal noise at a UE, which is modeled as $\sigma^2 = N_o + 10log(w) + \text{NF}$, where $w$ is the system bandwidth, $N_o$ is the noise power spectral density and NF denotes the noise figure at the UE receiver. The SINR at UE $n$ from beam $\langle m, k\rangle$ is written as

$$\gamma_{n,m,k} = \frac{p^r_{n,m,k}}{I_{n,m,k} + \sigma^2}. \tag{7}$$

To achieve a certain quality of service for the UE, the SINR at the UE has to be greater

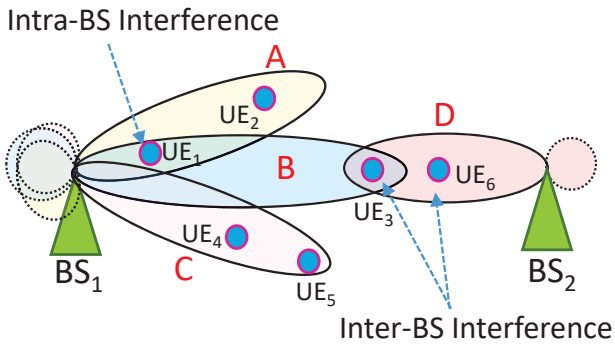



**Figure 7.** Inter-BS interference at UE3 and intra-BS interference at UE1 [1].

than or equal to a minimum SINR $\gamma_o$. Thus, we need to have

$$\frac{\gamma_{n,m,k}}{\gamma_o} \geq 1, \quad \forall n \in \mathcal{U}_{mk}. \tag{8}$$

In order to maintain the minimum $\gamma_0$ at all UEs in a beam, the required transmit power $p^t_{mk}$ is obtained using (6)-(26) as

$$p^t_{mk} \geq \frac{(I_{n,m,k} + \sigma^2)\gamma_o}{|\mathbf{h}^H_{mn}\mathbf{w}_{mk}|^2}, \quad \forall n \in \mathcal{U}_{mk}. \tag{9}$$

The transmit power vector for all beams of BS $m$ is given by

$$\mathbf{p}^t_m = \left[p^t_{m1}, p^t_{m2}, \ldots, p^t_{m|C_m|}\right], \tag{10}$$

where $|x|$ denote the cardinality of set $x$. Our objective is to reduce the total transmit power

$$\sum_{m\in B, i\in C_m} p^t_{mi}. \tag{11}$$

### 2.3.3 DOWNLINK UE RATE

The beam radiation patterns in the real world are complicated. Fig. 7 depicts a network with 2 BSs and 4 beams, with beam A transmitting to UE2, and beam B transmitting to UE1 and UE3. While UE1 is covered by beam B, it may receive *intra-BS interference* from beam A, e.g., if the null of beam A is not perfectly aligned with UE1. UE3 has *inter-BS interference* from beam D.

We quantify both intra-BS interference and inter-BS interference for a UE. For UE $n$ served by beam $\langle m, k\rangle$, the total interference $I_{n,m,k}$ is the summation of intra-BS interference $I'_{n,m,k}$ and inter-BS interference $I''_{n,m,k}$ given as follows.

$$\begin{aligned} I_{n,m,k} = I'_{n,m,k} + I''_{n,m,k} &= \sum_{i \in C_m \setminus k} p^r_{n,m,i} + \sum_{j \in B \setminus m} \sum_{l \in C_j} p^r_{n,j,l} \\ &= \sum_{i \in C_m \setminus k} p^t_{mi} |\mathbf{h}^H_{mn} \mathbf{w}_{mi}|^2 + \sum_{j \in B \setminus m} \sum_{l \in C_j} p^t_{jl} |\mathbf{h}^H_{jn} \mathbf{w}_{jl}|^2 \end{aligned} \tag{12}$$

In (25), the first term $I'_{n,m,k}$ is the summation of powers received by UE $n$ from all beams in $C_m$ except the serving beam $k$. The second term $I''_{n,m,k}$ is the sum of powers received from the beams of other BSs, i.e., except the serving BS $m$. Note that we compute the received power for a UE using (6).

The throughput capacity of the downlink channel has to be shared among all UEs in a beam. UEs in the same beam can use any multiplexing scheme to share resources, e.g., TDMA (time division multiple access), OFDMA (orthogonal frequency division multiple access), or NOMA (non-orthogonal multiple access). In the ensuing discussions, we assume the OFDMA system, but it can be similarly done for NOMA and TDMA. Similar to LTE we assume a fair resource allocation i.e., the OFDM sub-carriers are shared equally among UEs in a beam. If UE $n$ is scheduled in beam $\langle m, k\rangle$, the rate of UE $n$ in each downlink frame is given as

$$r_{n,m,k} = w_n \log_2(1 + \gamma_{n,m,k}), \tag{13}$$

where $w_n$ is the sum of the frequency spectrum from all downlink sub-carriers assigned to UE $n$ in the beam, and $\gamma_{n,m,k}$ is the corresponding SINR given by (24). If $w$ is the total system bandwidth, $w_n = w/|\mathcal{U}_{mk}|$, where $|\mathcal{U}_{mk}|$ denotes the number of UEs in the beam $\langle m, k\rangle$. The sum rate of the network is given as

$$\sum_{m\in B, 1\leq k\leq K, n\in\langle m,k\rangle} r_{n,m,k} \tag{14}$$

## 2.4 BEAMFORMING ORIENTED TOPOLOGY CONTROL (BOON)

The objective for topology control is to reduce the total transmit power in (29), subject to coverage of all UEs, maintaining the minimum SINR at each UE by (27), and using no more than the maximum number of beamformers at each BS. Moreover, for a given transmit power that meets those constraints, we aim to increase the sum rate in (14). We take a heuristic approach to achieve these two objectives. To reduce the transmit power, we first group UEs together through a *beamforming oriented UE clustering* (BOC) algorithm. This clustering is with regard to each BS assuming the maximum range and the clusters are formed with a radial shape with the center at a BS. The clusters of different BSs may overlap with each other, i.e., some UEs are included in clusters of different BSs. To associate those overlapped UEs to the right BSs, we divide each cluster into finer granularity sets through a *cluster decomposition based set construction* (CDSC) algorithm. Next, we associate these sets to the BSs through an *interference aware set covering* (IASC) algorithm, which selects a list of sets that cover all UEs, but do not overlap. Such UE/set association thus avoids or reduces interference between BSs, which reduces the transmit power of the BSs and increases the sum rate. At last, one beam is formed to cover each set by its associated BS. In the following, we describe each of those steps, which form the four components of BOON.

### 2.4.1 BEAMFORMING ORIENTED UE CLUSTERING (BOC)

From the topology control perspective, clustering is a process of organizing UEs into groups based on a similarity or cost metric. Clustering is in general NP-hard [54]. In this section, we develop a heuristic algorithm to find the number of clusters and their UEs for a BS, with the objective to reduce the transmit power of the BS. UEs are grouped based on their angles relative to the BS and the UE-UE distance.

Algorithm 1 illustrates the BOC algorithm that iteratively clusters UEs for a BS. BOC is a heuristic greedy algorithm which iteratively reduces the total clustering cost in each successive stage. It starts with each UE as one cluster. Two or more clusters are merged into one cluster in each successive stage, until the number of clusters is up to the number of beamformers $K$ of a BS or the cost does not decrease anymore. The output is the optimal number of clusters and UE set $\mathcal{U}_{mk}$ in each cluster. In order to assess if we should merge two clusters $i$ and $j$ into a new cluster $(i,j)$ in an iteration, we compute the cost metric for the new cluster. The cost metric determines the performance of BOC. Here we use a cost metric based on *energy efficiency* (EE) in bit/joule. EE measures the amount of downlink rate that a beam can deliver with per unit energy in joules it consumes. The clustering cost metric is the inverse of EE since the goal is to reduce transmit power in the network. Specifically, the cost of cluster $(i,j)$ is defined as

$$c_{(i,j)} = \frac{p^t_{(i,j)}}{\sum_{n\in\mathcal{U}_{(i,j)}} w_n \log_2\left(1+\frac{p^t_{(i,j)}|\mathbf{h}_n^H\mathbf{w}_{(i,j)}|^2}{I_{n,(i,j)}+\sigma^2}\right)}, \tag{15}$$

where $p^t_{(i,j)}$ is the transmit power, $I_{n,(i,j)}$ is the received interference at UE $n$ given by (25), and $\mathbf{w}_{(i,j)}$ is the weight vector.

While it is straightforward to consider all possible pairs of clusters for merging, in practice the beamforming gain is negligible when the main lobe is very wide. For instance, the approximate beamforming gain is $\pi^2/(\phi^{hp})^2$, where $\phi^{hp}$ is the solid angle for both the azimuth

and elevation plane [55]. We choose a feasible value of $\varphi$ beyond which the UE channel gain is very small and beamforming becomes inefficient. Hence, in line 8 of Algorithm 1, we actually do not need to consider all possible pairs. Instead, for a cluster, we only consider the next few clusters within angle $\varphi$ in the angular space for possible merging. At last, we choose one pair to merge into one cluster which yields the lowest cost.

A challenge for BOC is that the UE set in the merged cluster $(i, j)$ is not necessarily the combined UEs from clusters $i$ and $j$ together. This is different from classic clustering. Let $\mathcal{U}_{(i,j)}$ denote the UE set of the new cluster $(i, j)$. For classic clustering, if an object $x \notin \mathcal{U}_i$ and $x \notin \mathcal{U}_j$, then we must have $x \notin \mathcal{U}_{(i,j)}$. However, in BOC, there may exist a UE $x$ such that $x \notin \mathcal{U}_i$ and $x \notin \mathcal{U}_j$, but $x \in \mathcal{U}_{(i,j)}$. As an example, we consider three clusters (or beams) A, B, and C, in the network illustrated in Fig. 7. Suppose clusters A and C are selected to be merged. Then to avoid beams overlapping, we have to let the new cluster (A,C) to serve UE3, even though it is not in cluster A or C. Therefore, in line 10, we check if a cluster in between needs to be merged with clusters $(i, j)$.

A benefit of the above cost metric is that it can help to automatically find the optimal number of clusters, which corresponds to the lowest clustering cost. There is a trade-off between the number of UEs in a cluster and the total number of clusters for a BS. More UEs in a cluster reduce the spectral share per UE, which results in a higher clustering cost. On the other hand, more clusters mean fewer UEs in a cluster, which increases the UE spectral share. Nevertheless, this in turn increases the interference between beams, also resulting in a higher clustering cost. By 'minimizing' the clustering cost, BOC automatically selects the optimal number of clusters. This is done through continuously merging clusters even after the number of clusters reaches $K$, the number of beamformers at a BS. The algorithm terminates when the total cost $\sum \mathbf{c}$ does not decrease anymore, i.e., the clustering cost of the current iteration $\sum \mathbf{c}$ is higher than the one of the previous iteration $\sum \bar{\mathbf{c}}$ in line 4.

If $N$ UEs are distributed separately from each other in the angular space with regards to the BS, such that the difference between two neighbor UEs is greater than $\varphi$, each UE

**Algorithm 1:** Beamforming oriented clustering (BOC) for a BS [1]

**Input:** UE set $\mathcal{U}$ and maximum number of beams $K$
Let each UE be a cluster, with the set of clusters $\mathcal{B} = \{\{1\}, \{2\}, \ldots \{|\mathcal{U}|\}\}$
Let $v = K$, and Let cost vectors $\mathbf{c} = \mathbf{0}$ and $\bar{\mathbf{c}} = \mathbf{0}$
**while** $|\mathcal{B}| \geq v$ *or* $\sum \mathbf{c} < \sum \bar{\mathbf{c}}$ **do**
    $\bar{\mathcal{B}} \leftarrow \mathcal{B}$, $\bar{\mathbf{c}} \leftarrow \mathbf{c}$
    **if** $|\mathcal{B}| = 1$ **then**
        break out of loop
    For each cluster $s$ and the next clusters $k$ within angle $\varphi$ in the angular space, compute $c_{(s,k)}$ using (15)
    Select two clusters $(i, j) = \underset{s,k}{\operatorname{argmin}}\, c_{(s,k)}$
    Update $\mathcal{B}$: create a new cluster $h$ for $\mathcal{B}$ to replace clusters $i, j$ in $\mathcal{B}$. If the beam to cover UEs in cluster $h$ also covers a UE in a different cluster (not $i$ or $j$), then merge that cluster into cluster $h$
    Update $\mathbf{c}$ to be the cost vector of clusters in $\mathcal{B}$
Output $\bar{\mathcal{B}}$ as the set of clusters

forms one cluster, where $K$ is the number of beamformers for the BS, and $\varphi$ is the threshold for the main lobe width of a beam. BOC only needs to compute the clustering cost for each of the $N$ UEs once. Thus, the time complexity is $\Theta(N)$. However, in the worst case, BOC may need to run $N - 1$ iterations for merging clusters. In each iteration, BOC needs to compute $\Theta(N_j)$ cost metrics, where $N_j$ is the number of clusters in the $j$th iteration, since for each cluster $s$, we only need to consider the next few clusters, as discussed above. Thus, the worst-case time complexity of BOC is $\mathcal{O}(N^2)$.

### 2.4.2 CLUSTER DECOMPOSITION BASED SET CONSTRUCTION (CDSC)

The BOC algorithm groups UEs into clusters for each BS independently. In a multi-BS network, we cannot form beams directly based on those clusters, because clusters from different BSs can overlap with each other, which causes significant inter-BS interference. Therefore, we develop a *cluster decomposition based set construction* (CDSC) scheme to decompose each cluster into a number of sets. The objective is to avoid inter-BS interference for formed beams. For instance, in Fig. 8(a), UE $a$ can be covered by both BS1 and BS2, and thus is clustered by both BS1 and BS2. If we form beams directly based on clusters, the two beams interfere with each other. Through constructing sets as in Fig. 8(c) or (d), we can let BS1 form a beam to cover set $B$, which then eliminates interference between BS1 and BS2.

Different set construction schemes produce different beams which result in different transmit power. A naive approach is the *exhaustive set construction* (ESC). With ESC, we start from the closest UE of a cluster in the radial direction and contiguously construct one set per UE outward, with each set being a true superset of the previous set. For instance, suppose there are three UEs $(1, 2, 3)$ in a cluster for a BS, with UE 1 the closest to, and UE 3 the farthest from the BS. The ESC scheme constructs three sets from this cluster, $C = \{1\}$, $B = \{1, 2\}$, and $A = \{1, 2, 3\}$. While ESC is easy to implement, it may generate a large number of sets, which results in a large running time for set covering in the next step of BOON. Hence, we develop another scheme, called *shrinkable set construction* (SSC), to reduce the number of sets generated by set construction.

Fig. 8 illustrates ESC and SSC with six UEs $(a, b, c, d, e, f)$ forming four clusters with four BSs. For each cluster of a BS, we first identify the *intersections* between this cluster and the clusters of other BSs, by looking at the common UEs between them. Then the cluster is partitioned into intersection areas and non-intersection areas, as illustrated in Fig. 8(b). The intuition behind partitioning UEs into intersection and non-intersection areas is that UEs in

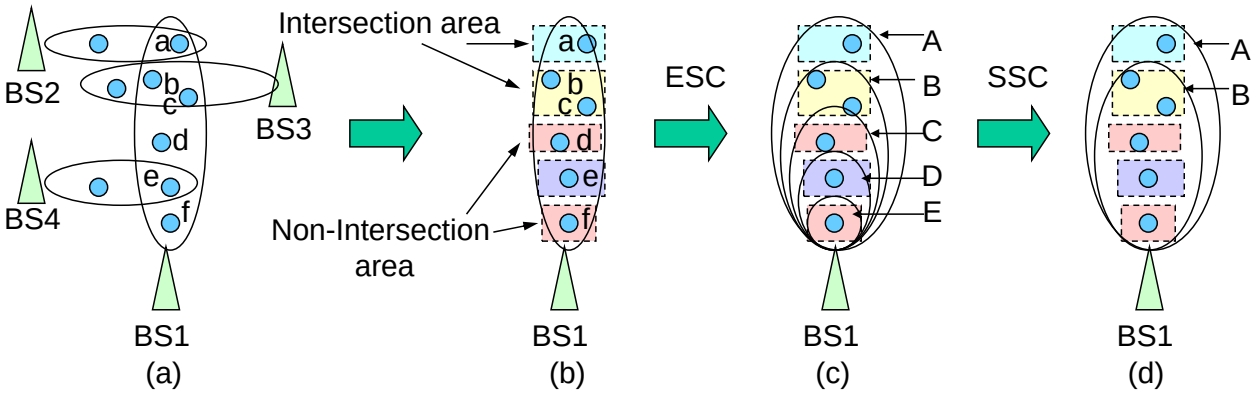



**Figure 8.** Exhaustive set construction (ESC) and shrinkable set construction (SSC) [1].

the intersection areas can be associated with either BS, but UEs in the non-intersection area can be associated with only one BS. For example, in Fig. 8, UE $a$ forms an intersection area and can be associated with either BS1 or BS2. UE $d$ forms a non-intersection area, and can be associated with BS1 only. We group the UEs in an intersection area into an *atomic set*. The UEs in a non-intersection area are also grouped into an atomic set. For example, there are five atomic sets in Fig. 8(b). Three atomic sets $\{a\}, \{b, c\}, \{e\}$ are formed by intersection areas and two atomic sets $\{d\}, \{f\}$ are formed by non-intersection areas. We then treat each atomic set as a UE, and apply the ESC scheme to construct sets, as illustrated in Fig. 8(c). Note that each set is constructed by concatenating a sequence of contiguous atomic sets starting from the one closest to the BS and going outward.

The intersection between clusters can be more complicated than the case in Fig. 8. We illustrate three complicated intersection cases: *partial intersection*, *crossing intersection*, and *overlapping intersection* in Fig. 9. With partial intersection, we can extend the intersection area to become a full intersection, illustrated by the extended dotted line in Fig. 9(a). This is done by letting the BS slightly increase the transmit power to cover the UEs in the dotted line. The scenario with crossing intersection is treated the same as full intersection. For the overlapping intersection with two or more other BSs, as illustrated in Fig. 9(c), we start from the intersection area closest to the BS, and let all UEs in this area be an atomic set.

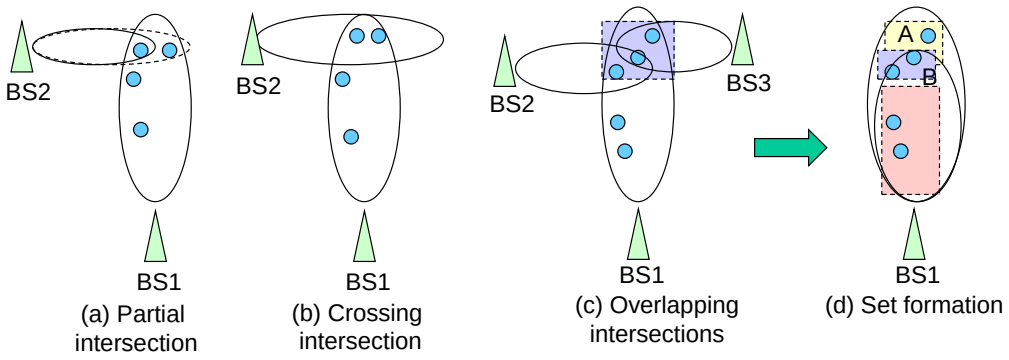



**Figure 9.** Intersection areas [1].

We remove the overlapped part of this intersection area with the next intersection area, and let the remaining UEs in the next intersection area be the next atomic set, as illustrated in Fig. 9(d).

The main idea of the SSC scheme is to remove some sets formed by ESC to reduce the total number of sets, as illustrated in Fig. 8(d). This may significantly reduce the running time of the set covering algorithm in the next section. First of all, we define the concept 'shrinkable'. A set $Y$ is *shrinkable* to its direct subset $X$, iff the atomic set $s = Y \setminus X$ is an intersection area. The core idea of SSC is that if $Y$ is *shrinkable* to $X$, then $X$ can be removed, without affecting the set covering result. We explain the intuition through an example illustrated in Fig. 8. As discussed earlier, an atomic set in a non-intersection area is covered by only one BS, but an atomic set in an intersection area can be covered by more than one BS. If the set covering algorithm (to be discussed in the next section) first selects the outermost atomic set $\{a\}$ of set $A$ to be covered by BS2, then for BS1, set $A$ will automatically shrink to set $B$. Hence, in the very beginning, we do not need to have set $B$ in the initial sets for the set covering algorithm. That is, the set covering result does not change even if we remove set $B$ in the initial sets. On the other hand, if the set covering algorithm selects atomic set $\{e\}$ in set $C$ to be covered by BS4, then set $C$ cannot shrink to set $D$ or $E$ because the atomic set $\{d\}$ in $C$ has to be covered by BS1. In summary, a set

$Y$ is *shrinkable* to its direct subset $X$, iff the atomic set $s = Y \setminus X$ is an intersection area. The following theorem states whether a set can be removed before feeding the sets to the set covering algorithm.

**Theorem 1.** *Consider a set $Y$ and its direct subset $X$ constructed from the same cluster, i.e., $Y \setminus X$ contains only one atomic set. We can remove $X$ without affecting the set covering results, if any of the following conditions are satisfied.*

1. *$X$ is not shrinkable.*

2. *$X$ is shrinkable, but $Y$ is not shrinkable.*

*Proof.* First of all, with condition 1, we must have that set $Y$ is shrinkable. It is not possible that both $X$ and $Y$ are not shrinkable, based on the construction of beamforming sets. This is because if this is the case, then the outmost two atomic sets in set $Y$ are both in non-intersection areas, which is in conflict with the partitioning of a cluster into intersection and non-intersection areas. We use an example to illustrate $X$ and $Y$. Consider sets $B$ and $C$ in Fig. 8(c), which correspond to $Y$ and $X$. Furthermore, let set $Z$ denote the set from another cluster that intersects with the outmost atomic set of set $Y$. With set $Y$ as shrinkable, if the set covering algorithm selects set $Z$ first (before selecting $Y$), set $Y$ can shrink to set $X$. That is, even if set $X$ was removed, shrinking $Y$ essentially re-creates set $X$, which could be selected by the set covering algorithm later. Thus the set covering is not affected even if set $X$ was removed. On the other hand, if the set covering algorithm selects set $Y$ first, we do not have to keep set $X$ since set $X$ would not be selected anyway by the set covering algorithm, because selecting set $Y$ implies the UEs in set $X$ are already covered. Hence condition 1 is a sufficient condition to remove set $X$.

With condition 2, the outmost atomic set $s$ in $Y$, where $s = Y \setminus X$, is in a non-intersection area. Hence the UEs in the atomic set $s$ cannot be served by another BS, so the BS that set $Y$ belongs to has to cover atomic set $s$, which implies that set $Y$ is covered. This in turn

implies set $X$ would not be selected by the set covering algorithm later on. It is possible that if we do not remove set $X$, the set covering algorithm may select set $X$ before selecting set $Y$. However, eventually the set covering algorithm has to select set $Y$ or its superset, to cover the atomic set $s$. Thus, set $X$ would be removed since its superset is selected. That is, we can remove set $X$ in the beginning, without affecting the set covering result. ■

### 2.4.3 INTERFERENCE AWARE SET COVERING (IASC)

While set covering is a classic problem, the inter-beam interference raises a great challenge to applying set covering for beamforming. A major difference is that, in classic set covering, the objects in the selected set in a given step are removed from the remaining sets, which is a key technique to optimize performance in classic set covering. Unfortunately, IASC cannot use this technique, as removing UEs from the remaining sets may cause strong interference in the later beamforming stage. In fact, with IASC, there is only one scenario where we can remove a UE from the remaining sets. That is, if a remaining set is *shrinkable*, we can remove a UE from it. This is because, when the set *shrinks*, the cost of the remaining set is recalculated and updated. If it results in a strong interference, the cost will be high. Hence this shrunk set is unlikely to be selected in future steps of set covering.

Algorithm 2 illustrates the IASC algorithm. Let $\mathcal{S}_i$ denote the $i$th SSC set after applying the CDSC scheme in BOON. Note that $\bigcup_i \mathcal{S}_i = \mathcal{U}$. All SSC sets of all BSs are denoted as $\mathcal{F} = \{\mathcal{S}_i\}$. We use (15) to compute the cost for set $\mathcal{S}_i$, denoted as $c(\mathcal{S}_i)$. The objective of IASC is to find an index subset $J \subseteq \{1, 2, \ldots, |\mathcal{F}|\}$ such that $\sum_{i \in J} c(\mathcal{S}_i)$ is minimum, subject to $\bigcup_{j \in J} \mathcal{S}_j = \mathcal{U}$. Initially, the algorithm lets $\mathcal{C}$ be empty. IASC picks a set $\mathcal{S}$ with the minimum cost in each iteration, and removes UEs of this set from $\mathcal{U}$ (note: not from other remaining sets in $\mathcal{F}$). The selected set $\mathcal{S}$ is then removed from $\mathcal{F}$. Next, we check whether any $x \in \mathcal{F}$ is "shrinkable" if we remove UEs of $\mathcal{S}$ from $x$. If $x$ can "shrink", we remove the common UEs in both $\mathcal{S}$ and $x$, i.e., $x = x \backslash \mathcal{S}$, and update the cost $c(x)$. If $x$

**Algorithm 2:** Interference aware set covering (IASC) [1]

1 **Input:** Set family $\mathcal{F}$, set costs $c(\mathcal{S})$, UE set $\mathcal{U}$
2 $\mathcal{C} = \emptyset$
3 **while** $\mathcal{U} \neq \emptyset$ **do**
4  $\mathcal{S} = \text{argmin}_{x \in \mathcal{F}} c(x)$
5  Remove UEs of $\mathcal{S}$ from $\mathcal{U}$
6  $\mathcal{F} \leftarrow \mathcal{F} \setminus \mathcal{S}$
7  **for** *each set* $x \in \mathcal{F}$ **do**
8  **if** *$x$ shrinks when we remove $\mathcal{S}$* **then**
9  Remove all UE $i \in x \cap \mathcal{S}$ from $x$ and update $c(x)$
10  Remove any set $h$ in $\mathcal{C}$ if $h \subset \mathcal{S}$
11  Add set $\mathcal{S}$ to $\mathcal{C}$
12 **return** $\mathcal{C}$

cannot shrink, we do not remove any UE from $x$ even though a UE is already covered by $\mathcal{S}$, based on earlier discussions. Next we check if set $\mathcal{S}$ covers all UEs in a previously selected set $h$. If yes, we remove set $h$ from $\mathcal{C}$. Finally $\mathcal{S}$ is added to $\mathcal{C}$. The algorithm outputs a list of sets with the minimal total cost, while all UEs are covered by those sets. In summary, IASC tries to reduce the overall network cost while minimizing the set overlapping such that each UE is associated with one beam.

Set covering is an NP-hard problem. The IASC algorithm is a heuristic greedy algorithm to solve the weighted set covering problem. In each iteration, the time is mainly on updating the set cost on line 7, which requires $\mathcal{O}(|\mathcal{F}|)$ time. The worst case run time of IASC is $\mathcal{O}(|\mathcal{U}||\mathcal{F}|)$.

### 2.4.4 BEAMFORMING

We adopt the *linearly constrained minimum variance* (LCMV) beamforming scheme [56, p.513]. It is able to constrain the beamforming output to achieve a given gain in the directions of intended UE signals while minimizing the power response toward UE directions in other beams and BSs. To achieve a complex gain $g^*$ in the UE direction $\theta$, the beamforming weight vector is subject to $\mathbf{a}_\theta^H \mathbf{w} = g^*$, where $\mathbf{a}(\bullet)$ is given by (21). Let $\mathbf{A}_\theta = [\mathbf{a}_{\theta_1}, \ldots, \mathbf{a}_{\theta_k}]$ be the constraint matrix for total $k$ UEs of a BS, with AoAs $\theta_1, \ldots, \theta_k$, and $\mathbf{f}$ be the $k$-dimension single column response vector. The covariance of the $k$ UE signals $\mathbf{R} = \mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right] = \frac{L|\alpha|^2}{\rho}\mathbf{A}_\theta \mathbf{A}_\theta^H$, where $\mathbf{h}$ is the $k$ UE channel vectors given in (20), $L$ is the number of antenna elements in the BS, $\rho$ is the path loss and $\alpha$ is the complex gain of the signal. The LCMV beamforming to minimize transmit power is formulated as

$$\min\{\mathbf{w}^H \mathbf{R} \mathbf{w}\} \text{ such that } \mathbf{A}_\theta^H \mathbf{w} = \mathbf{f}. \tag{16}$$

The solution of (41) is obtained as follows using the Lagrange multiplier method

$$\mathbf{w} = \mathbf{R}^{-1}\mathbf{A}_\theta(\mathbf{A}_\theta^H \mathbf{R}^{-1} \mathbf{A}_\theta)^{-1}\mathbf{f}. \tag{17}$$

By clustering UEs into groups and choosing an appropriate beamforming weight vector for each group, it is possible to approximately eliminate or significantly reduce the inter-beam interference. The time complexity of the LCMV beamforming in (42) is $\Theta(\max(kL^2, L^{2.373}))$ assuming $L \geq k$, where $L^{2.373}$ is the matrix inverse time for an $L \times L$ matrix.

We use the channel vector $\mathbf{h}_{mn}$ between BS $m$ and UE $n$ in (20) to compute a beamforming weight vector $\mathbf{w}$ for each set in $\mathcal{C}$ output by IASC, to serve all UEs in this set. Let $\mathcal{U}_i$ denote the UEs contained in the $i$th set in $\mathcal{C}$. Also let $m_i$ denote the corresponding BS of set $\mathcal{C}_i$, i.e., the $i$th set in $\mathcal{C}$. For each UE set $\mathcal{U}_i$, we construct the constraint matrix $\mathbf{A}_\theta$, the corresponding covariance matrix $\mathbf{R}$, and the response vector $\mathbf{f}$. Then we use (42) to compute

**Table 2.** Simulation parameters in BOON

| **Parameter** | **Value** |
|---|---|
| System operating frequency (GHz) | 73 |
| Network dimension ($m^2$) | $[400 \times 200]$ |
| Number of antennas in ULA per BS | 16 |
| Adjacent antenna spacing $\Delta$ in ULA | $\lambda/2$ |
| Maximum number of beamformers $K$ per BS | 10 |
| System bandwidth (FDD duplex mode) (MHz) | 500 |
| LOS path loss exponent $\eta$ | 2.2 |
| Minimum required SINR $\gamma_o$ at UE (dB) | 1 to 15 |
| Noise power spectral density $N_o$ (dBm/Hz) | $-174$ |
| Noise figure NF at UE receiver (dB) | 6 |

$|\mathcal{C}|$ number of weight vectors $\mathbf{w}_1, \ldots, \mathbf{w}_{|\mathcal{C}|}$, where $\mathbf{w}_i$ is associated with set $\mathcal{C}_i$. Then using (27), the transmit power allocated to the signal for set $\mathcal{C}_i$ is computed as

$$p_i^t = \max\left(\frac{(I_n + \sigma^2)\gamma_o}{|\mathbf{h}_{m_i,n}^H \mathbf{w}_i|^2}\right), \quad \forall n \in \mathcal{U}_i, i \in \mathcal{C} \tag{18}$$

where $\mathbf{h}_{m_i,n}$ is the channel vector between BS $m_i$ and UE $n$, $\gamma_o$ is the required minimum SINR threshold for all UEs, and $\sigma_n^2$ is the noise power in the AWGN channel for UE $n$. Note that interference $I_n$ in (18) is close to zero because IASC finds the best sets of UEs to avoid interference. Furthermore, the beamformer constraint and response vector are applied to form null in the direction of UEs of other sets.

## 2.5 PERFORMANCE EVALUATION

We compare BOON with MUB, MCB based topology control, and a state-of-the-art

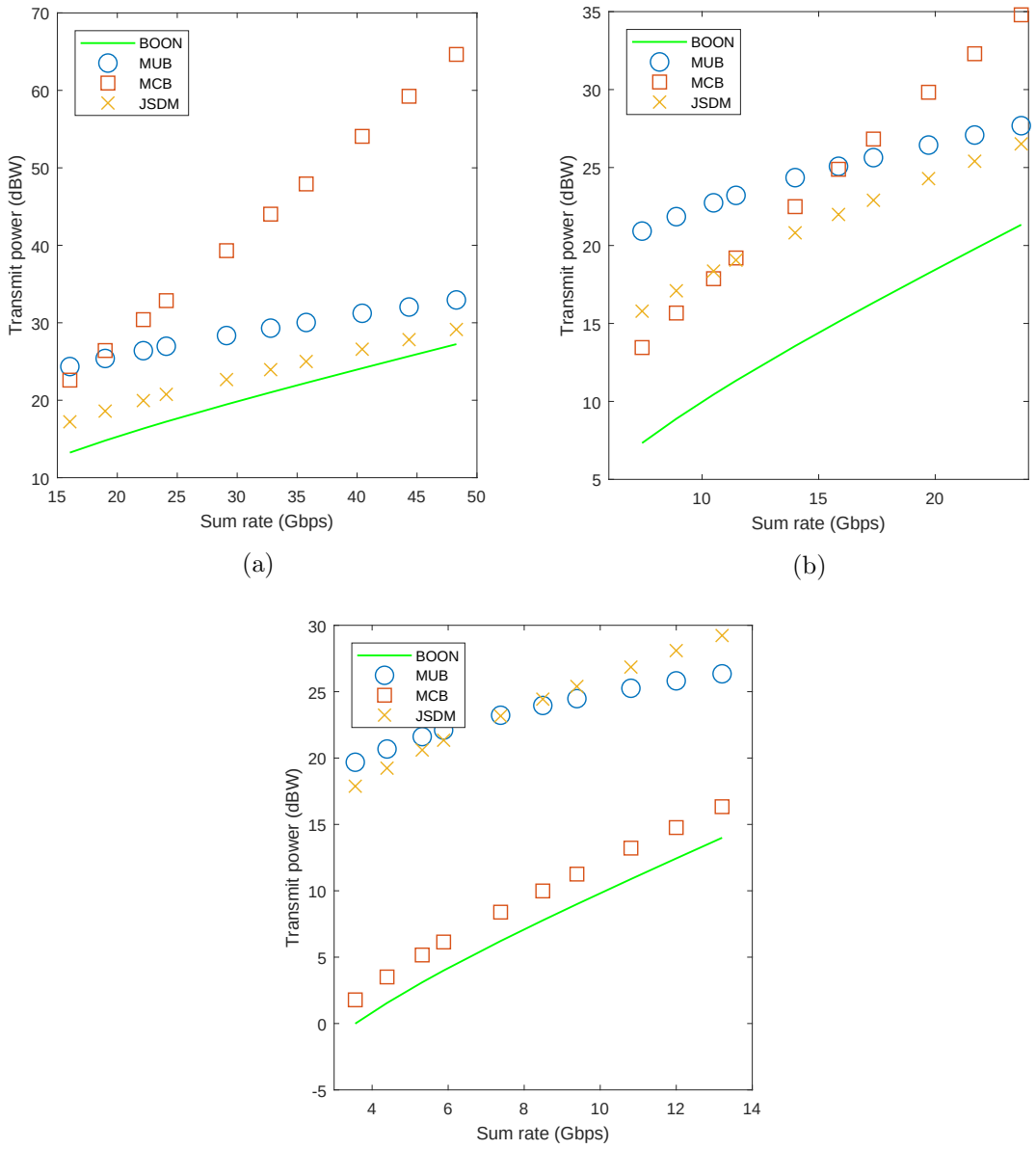



**Figure 10.** Average sum rate versus transmit power of 50 experiments: (a) spread UE distribution (b) grouped UE distribution, (c) dense UE distribution. $\gamma_0 = 1$ to 15 dB [1].

scheme JSDM [45]. All schemes use the same beamforming technique, LCMV. Note that both MUB [37,38,49,57] and MCB [42–44,58] are special cases of BOON. The former treats each UE as a set, forming one dedicated beam for each UE, while the latter treats all UEs in

one set, i.e., forming one beam to serve all UEs. In JSDM, UEs are partitioned into groups with approximately similar channel covariances. All UEs in a group are served simultaneously using one beamforming vector. We assume all BSs are configured with a ULA, while each UE is equipped with a single antenna. The OFDMA system is assumed for spectrum sharing among UEs in a beam. We consider a frequency division duplex (FDD) mode where 80% of the system bandwidth, i.e., 400 MHz, is used for downlink data signal transmission and the remaining 20% is used for control signal purposes. The system parameters used in simulations are summarized in Table 2. The performance metrics we use are the BS transmit power, sum rate of all UEs, SINR, and computation complexity. The sample networks are in a $[400 \times 200]$ $m^2$ outdoor area, with 6 BSs and 40 UEs. There are 10 blockages in four different shapes: rectangle, square, hexagon, and pentagon. The *reference signal received power* based UE association is used in MUB, MCB, and JSDM. The BSs are evenly distributed in the network area. UEs are distributed by three patterns; *spread*, *grouped*, and *dense*. In the spread UE distribution, UE locations are randomly distributed. In the grouped UE distribution, 40 UEs form into 10 UE groups, each with 4 UEs. The center of each group is randomly generated, and the 4 UE locations are randomly generated, but closely located to the center. In the dense UE distribution, UEs form into one to two groups around each BS. The group center and UE locations are generated similarly as in the grouped UE distribution.

We have generated 50 networks for each UE distribution pattern, respectively. We record the average transmit power and sum rate for each network. The required minimum SINR $\gamma_o$ varies in a range of 1 to 15 dB. In simulations, we first find the transmit power for BOON to meet a minimum SINR $\gamma_o$ at every UE and record the corresponding sum rate. Then we apply the same transmit power in MUB, MCB and JSDM to find their SINR and sum rate. Fig. 10 illustrates the *scatter plot* of the average sum rate versus transmit power for 50 networks/experiments in each UE distribution pattern, using different seeds for simulations. Each data point in the figure indicates the required transmit power for each scheme to achieve a given sum rate. The average error bar length (95% confidence interval) of BOON,

MUB, MCB, and JSDM is 2.3, 3.2, 3.5, and 4.8 dBW, respectively. They are not plotted in the figure as the error bars of different schemes overlap with each other significantly, which makes them hard to be viewed. Overall, BOON significantly outperforms other schemes, thanks to the effective BOC clustering, and the IASC algorithm which smartly associates UEs to BSs to avoid inter-BS interference. BOON not only uses less transmit power, but also has a smaller error bar. In the dense UE distribution, BOON performs similarly as MCB as it is best to form one or two beams in this case. They perform significantly better than MUB and JSDM, as the latter form multiple beams even for one group, and results in significant interference between UEs. *In summary, BOON smartly adapts to different UE distributions to achieve the best performance in all scenarios. Other schemes cannot adapt to different UE distributions, and thus perform well in certain scenarios, but perform poorly in other scenarios.*

**Table 3.** Average required transmit power of BOON (in %) with respect to MUB, MCB, and JSDM, to achieve the same average sum rate

| UE Pattern | Gbps | vs. MUB | vs. MCB | vs. JSDM |
|---|---|---|---|---|
| Spread | 31.2 | 15.08 % | 2.71 % | 50.37 % |
| Grouped | 15.05 | 10.87 % | 12.9 % | 20.6 % |
| Dense | 8.03 | 3.3 % | 80.4 % | 2.9 % |

In Table 3, we average the percentage transmit power BOON uses in all experiments, compared with MUB, MCB, and JSDM, to achieve the same sum rate. On average, BOON requires 10%, 32%, and 25% transmit power of MUB, MCB, and JSDM, respectively, to achieve an average 0.45 Gbps per UE rate. In other words, BOON achieves 90%, 68% and 75% gain in transmit power compared with MUB, MCB, and JSDM, respectively.

Fig. 10 only gives the average result among 50 experiments. Next we examine how spread the performance is across experiments. Fig. 11 plots the *probability density distribution* (PDF) of the sum rate obtained from the results of the 50 experiments for each UE distribution pattern, when the required minimum SINR $\gamma_o = 10$ dB. The PDF of the sum

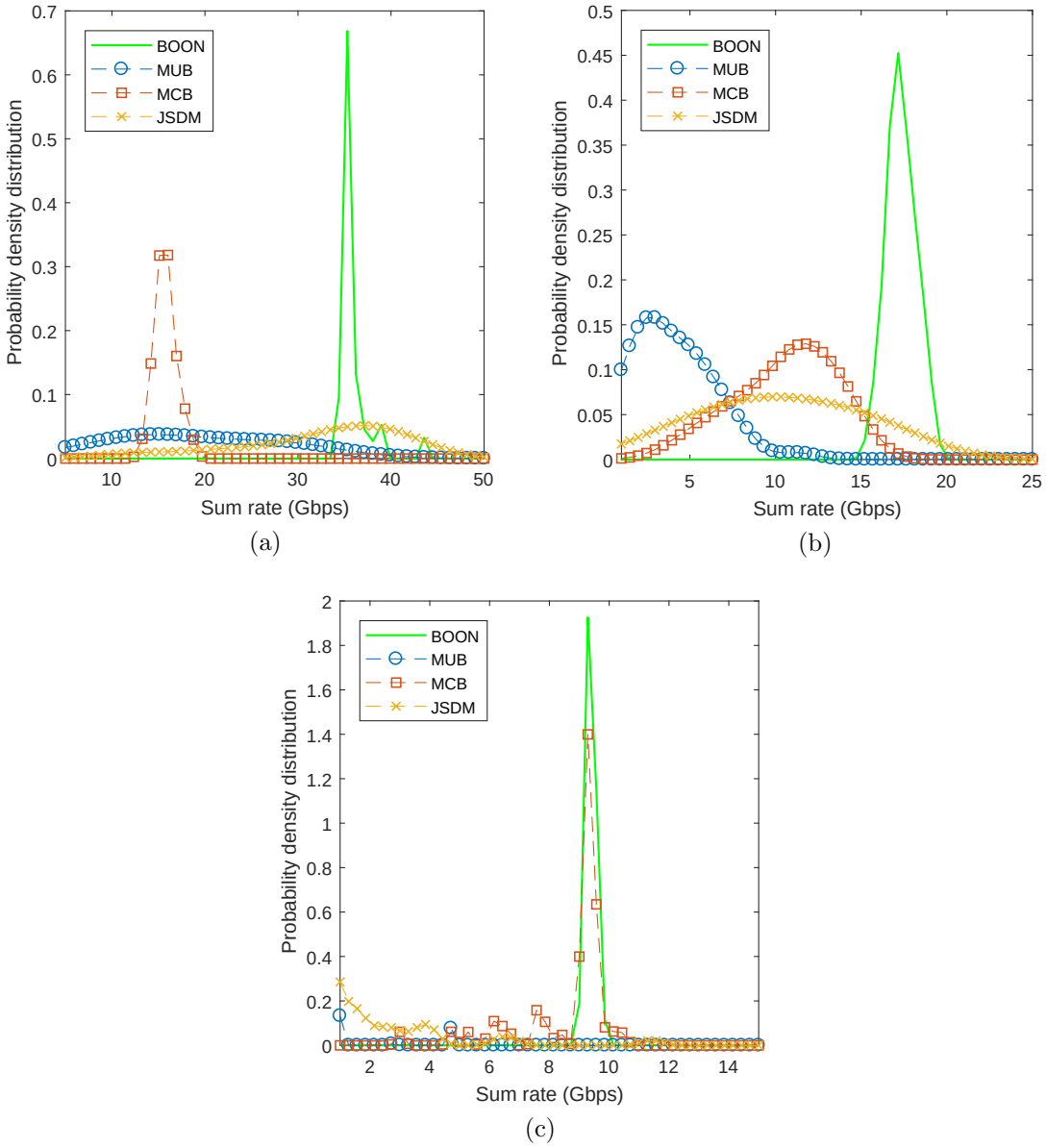



**Figure 11.** Probability density distribution of the sum rate, (a) spread UE distribution, (b) grouped UE distribution, and (c) dense UE distribution, $\gamma_0 = 10$ dB [1].

rates with different values of $\gamma_o$ have similar trends and are omitted due to space limitation. From Fig. 11, BOON not only achieves a higher average sum rate than MUB, MCB and JSDM, which can be verified by Fig. 10, but also has a much narrower spread on the sum rate overall. This indicates that for the 50 experiments, BOON can achieve most sum rates

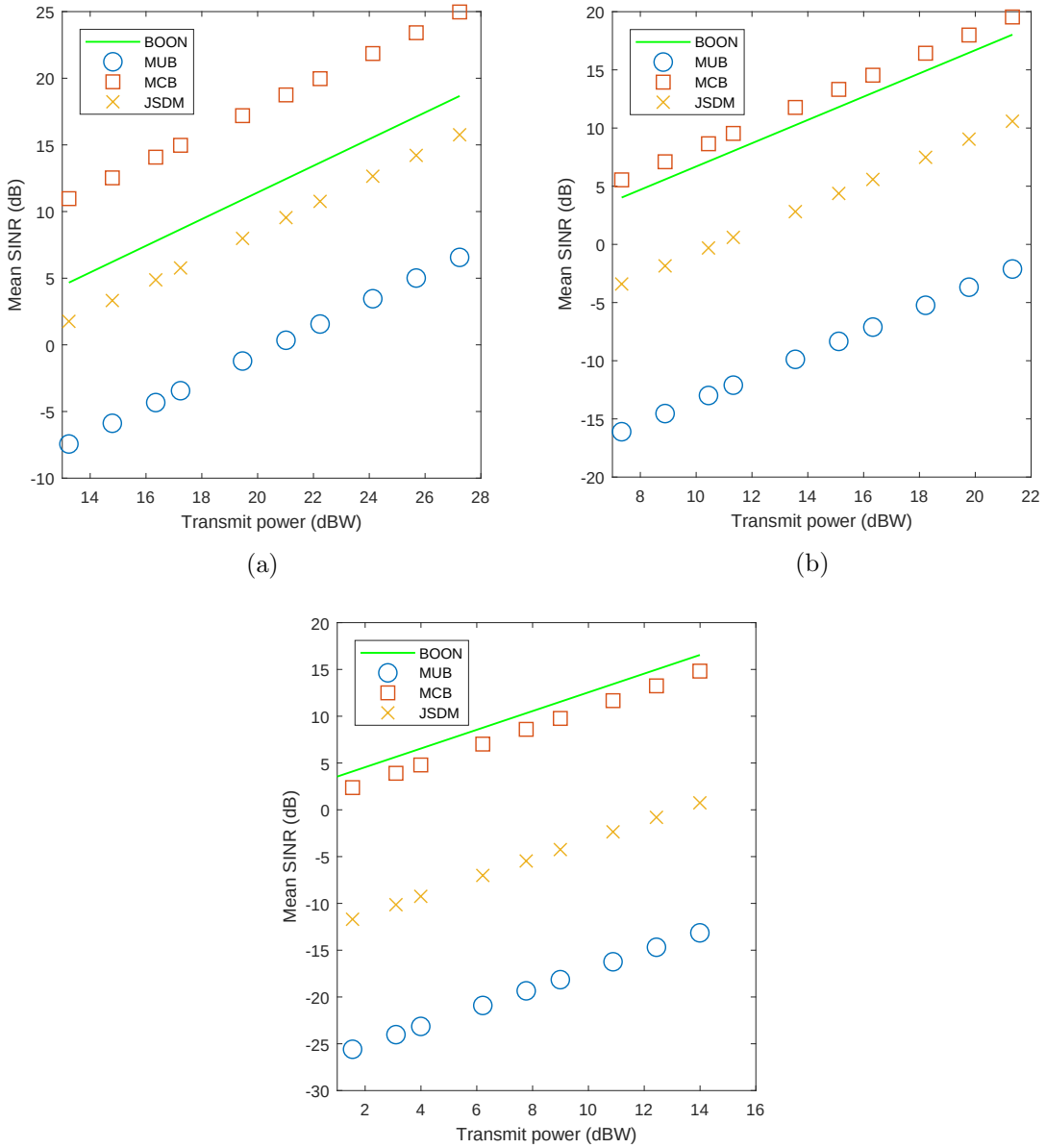



**Figure 12.** The mean SINR versus transmit power of 50 experiments, (a) spread UE distribution, (b) grouped UE distribution, and (c) dense UE distribution. $\gamma_0 = 1$ to 15 dB [1].

close to the average sum rate. In contrast, for JSDM and MUB, the average sum rate is not only smaller, but also the sum rates are much more spread out. This means that for some networks, their sum rates are very poor, significantly smaller than the average, while

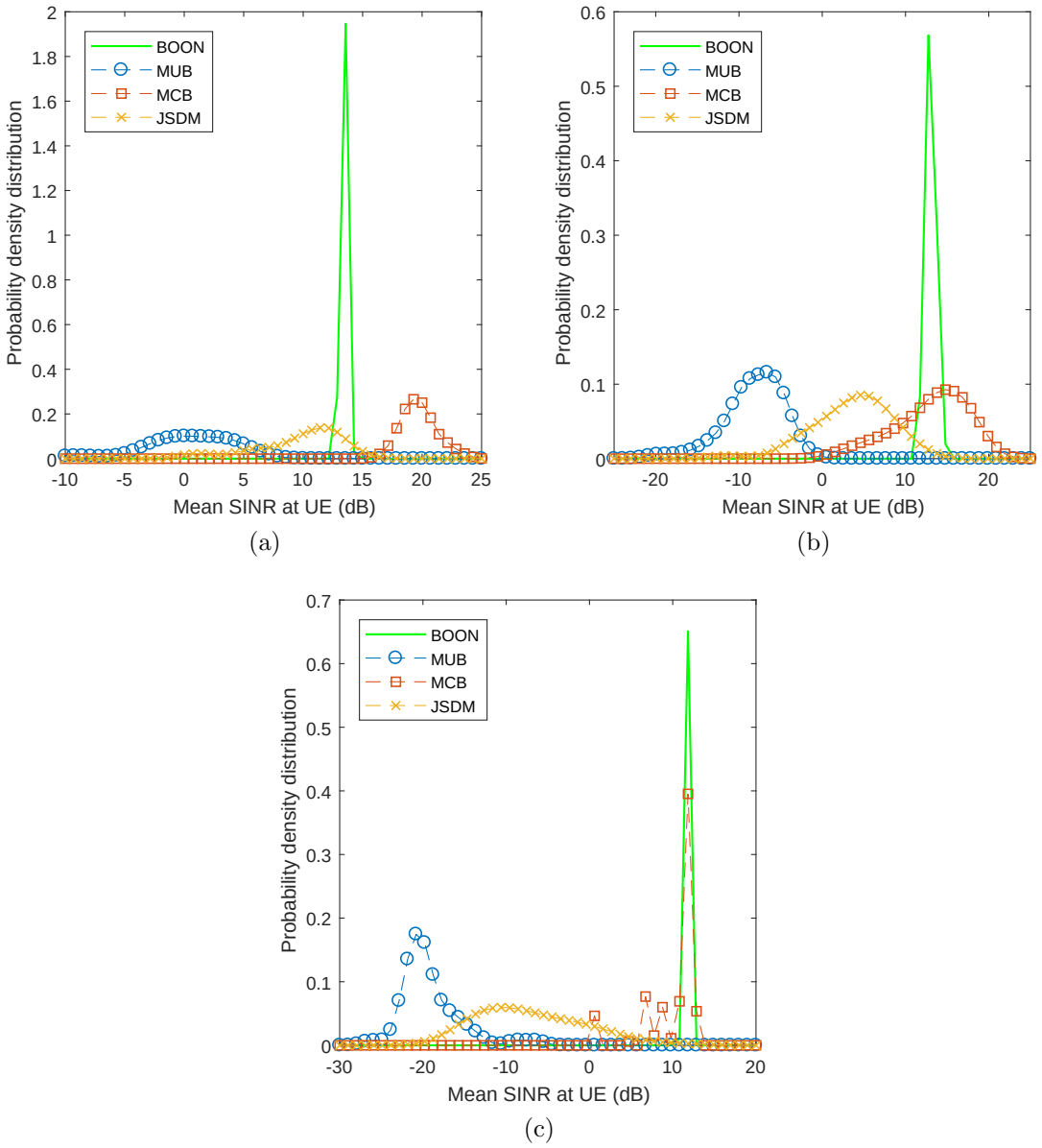



**Figure 13.** Probability density distribution of mean SINR, (a) spread UE distribution, (b) grouped UE distribution, and (c) dense UE distribution, $\gamma_0 = 10$ dB [1].

for some other networks, their sum rates are much higher than the average. In other words, their performance highly relies on the specific UE location distribution. *This is another demonstration that BOON smartly adapts to all network scenarios or UE location distributions, to achieve a good performance in all scenarios, while other schemes cannot adapt*

*to different network scenarios well.* Note that although MCB also has a relatively narrow spread, its sum rate is much lower than BOON, except in the dense UE distribution pattern.

Fig. 12 plots the average value of the mean SINR of the 50 experiments in each UE distribution pattern. BOON clearly outperforms JSDM and MUB. The average error bar length of BOON, MUB, MCB, and JSDM is 0.5, 3.8, 3.0, and 4.7 dB, respectively. They are not shown in the figure due to significant overlapping, which makes it hard to view them. Clearly BOON has a much smaller error bar than other schemes, i.e., it has a more stable performance for different scenarios. On average, the SINR gain of BOON is about 8 dB over JSDM, and about 20 dB over MUB. MCB performs better than BOON on the UE SINR in the spread UE distribution pattern. However, this is at the expense of a lower sum rate, which can be seen from Fig. 10(a). This is because all UEs of a BS share one beam. Although the SINR is high, the spectral share per UE is small, and thus the sum rate is not high. On the other hand, BOON smartly adapts to the UE distribution and determines the approximately optimal number of beams to be formed by each BS. Thus, even though the SINR is not as high as that of MCB, the spatial reuse of the spectral resource results in a higher sum rate for BOON compared with MCB.

The PDF of the mean SINR of all UEs for the 50 experiments is illustrated in Fig. 13, with $\gamma_o = 10$ dB. The PDF results of the mean SINR with different values of $\gamma_o$ also have similar trends and hence are omitted due to space limitations. Similar to Fig. 11, overall the mean SINR of BOON has a much smaller deviation from the average value than other schemes. MCB has a higher SINR than BOON in the spread UE distribution pattern, but it has a lower sum rate. This is because MCB forms one beam per BS and avoids the inter-beam interference in a BS. Thus it achieves a high SINR. Nevertheless, the sum rate is the smallest as all UEs share the same beam, and thus the spectral share for each UE in a beam is small. MUB has the worst performance on SINR because forming one beam per UE causes excessive interference among beams.

Table 4 compares the computation complexity of BOON, MUB, MCB, and JSDM, where

**Table 4.** Computation complexity of BOON, MUB, MCB, and JSDM [1]

| | BOON | MUB | MCB | JSDM |
|---|---|---|---|---|
| Time complexity | $\Theta(L^2 \sum_{1\leq i\leq \lvert B\rvert} q_i Q_i)$ | $\Theta(L^2 \sum_{1\leq i\leq \lvert B\rvert} k_i^2)$ | $\Theta(L^2 \lvert\mathcal{U}\rvert)$ | $\Theta(L^2 \sum_{1\leq i\leq \lvert B\rvert} k_i G_i)$ |

$L$ is the number of elements of the BS antenna, $|\mathcal{U}|$ is the total number of UEs, $|B|$ is the number of BSs, $k_i$ is the number of UEs for BS $i$ in MUB, MCB and JSDM, $G_i$ is the number of UE groups/beams for BS $i$ in JSDM, $q_i$ and $Q_i$ are the number of UEs and beams for BS $i$ in BOON, respectively. Note that generally $q_i \neq k_i$ and $Q_i \neq G_i$ since the UE association and clustering of BOON and JSDM are different. The $k_i$ is the same for MUB, MCB, and JSDM as they use the same user association scheme, the reference signal received power based UE association. The beamforming operation in (41–42) of Section 2.4.4 has to be conducted for every beam of each BS. Its time complexity is $\Theta(\max(kL^2, L^{2.373}))$, where $k$ is the total number of UEs for a BS. For MCB, the time complexity is $\Theta(|\mathcal{U}|L^2)$ assuming the number of UEs for each BS is larger than $L^{0.373}$, which is $2.8, 3.6, 4.7$ when $L = 16, 32, 64$, respectively. For MUB, the time complexity is $\Theta(L^2 \sum_{1\leq i\leq |B|} k_i^2)$. For JSDM, the UE grouping based on channel covariance takes $\Theta(\sum_{1\leq i\leq |B|} k_i^2)$ time and the beamforming takes $\Theta(\sum_{1\leq i\leq |B|} k_i L^2 G_i)$ time. In general, $k_i < L^2 G_i$. Hence the time complexity of JSDM is $\Theta(L^2 \sum_{1\leq i\leq |B|} k_i G_i)$. For BOON, the UE clustering takes $\mathcal{O}(\sum_{1\leq i\leq |B|} q_i^2)$ time. The set covering by the IASC algorithm has a time complexity $\mathcal{O}(|\mathcal{U}||\mathcal{F}|)$. The time for CDSC is small and dominated by the set covering time. The beamforming takes $\Theta(\sum_{1\leq i\leq |B|} q_i L^2 Q_i)$ time. Again, generally $q_i < L^2 Q_i$. Moreover, the set covering time $\mathcal{O}(|\mathcal{U}||\mathcal{F}|)$ is typically smaller than $\Theta(\sum_{1\leq i\leq |B|} q_i L^2 Q_i)$. Hence the BOON time complexity is $\Theta(L^2 \sum_{1\leq i\leq |B|} q_i Q_i)$. From Table 4, in general the computation complexity of MCB is the lowest, while the one of MUB is the highest. The computation complexities of BOON and JSDM are comparable and both are between the ones of MCB and MUB.

Fig. 14 illustrates the average execution time of the experiments in each UE distribution,

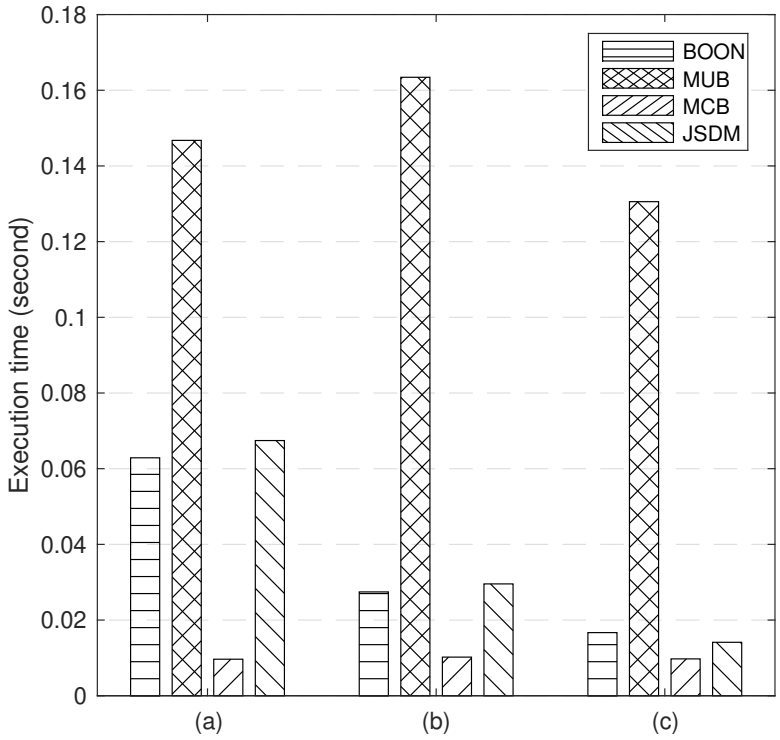



**Figure 14.** Average execution time of the experiments in each UE distribution: (a) spread UE, (b) grouped UE , (c) dense UE [1].

on a workstation with an Intel Core i7 CPU and 32 GB RAM. The execution time of BOON and JSDM are comparable and their time is between the ones of MCB and MUB. This is consistent with the time complexity comparison in Table 4. MUB uses much more time than BOON and JSDM in all scenarios. MCB uses less time than BOON and JSDM, but has much poorer performance on average.

## 2.6 CHAPTER SUMMARY

In this chapter, we discuss a topology control framework called *Beamforming Oriented tOpology coNtrol*(BOON), which has an objective to reduce the total transmit power of all BSs while forming beams to cover UEs. The BOON framework includes four components such as UE clustering, set construction, set covering, and beamforming. BOON has been

compared with the multi-user and multicast beamforming based topology control schemes, and a state-of-the-art scheme JSDM on transmit power, sum rate, SINR, and computation complexity. The results indicate that overall BOON significantly outperforms them. For instance, BOON uses only 10%, 32% and 25% transmit power on average, of the other three schemes, respectively, to achieve the same sum rate in the network.

# CHAPTER 3

# UE ASSOCIATION AND TRAFFIC SCHEDULING

In the previous chapter, we discussed the BOON framework to efficiently reduce the downlink transmit power in order to cover all UEs in the network. However, when a BS has a limitation on the transmit power and number of beamformers, covering all UEs simultaneously may not be possible, especially for dense mmWave networks. Therefore, BSs need to associate and schedule UEs over different time slots. This *UE association and scheduling* raises a great challenge due to the limited power budget and beamformers in BS, diverse user traffic loads, user quality of service requirement, etc.

In this chapter, we present a novel framework to reduce the downlink traffic transmission time in multi-BS mmWave networks, called *clustering Based dOwnlink UE assOciation, Scheduling, beamforming with power allocaTion* (BOOST). The objective of BOOST is to reduce the overall downlink transmission time, subject to the BS power budget, the number of beamformers at BS, UE traffic loads, and the minimum SINR requirement at each UE. BOOST consists of three principal components: UE clustering, UE-base station association, and beamforming and traffic scheduling. In this chapter, each component is discussed and three heuristic algorithms are developed to achieve the overall goal. At last, the performance of BOOST is evaluated and compared with three existing state-of-the-art user selection and scheduling schemes. The performance metrics are traffic transmission time, sum rate, and user fairness.

The rest of the chapter is structured as follows: Section 3.1 presents background and some related works, system model is discussed in Section 3.2, traffic scheduling problem is formulated in Section 3.3, the BOOST framework is broadly presented in Section 3.4, performance of BOOST is evaluated in Section 3.5 and Section 3.6 concludes the chapter.

## 3.1 BACKGROUND AND RELATED WORK

Unlike typical cellular networks, UE association in mmWave networks does not depend on the distance between BS and UE, but depends on if a UE is covered by a beam. This makes the UE association and beamforming a joint optimization problem, which is even more challenging as BS is power limited and cannot cover all its UEs simultaneously. Each BS usually schedules a set of UEs in each time slot and forms beams to transmit the UE traffic. The UE association and scheduling also has to consider various other factors including heterogenous UE traffic loads, the quality of service requirement for UEs, as well as the beam interference, power allocation, UE fairness, etc. More importantly, those factors are intermingled, which makes the UE association and scheduling a fundamentally challenging joint optimization problem. In this chapter, we develop a novel framework for UE association, scheduling, and beamforming in mmWave networks, called the *clustering Based dOwnlink UE assOciation, Scheduling, beamforming with power allocaTion* (BOOST). The objective is to reduce the transmission time for all UEs traffic, subject to the power budget of BSs, number of beamformers of BSs, UE traffic loads, and the required minimum SINR for each UE. Note that reducing traffic transmission time or latency is a key objective of 5G. Moreover, in the context of this work, reducing transmission time also contributes to increasing the network capacity, another key objective of 5G.

Most existing works on UE scheduling focus on choosing a set of UEs for a time slot to maximize the sum rate given the BS power budget. Several works chose a group of UEs with orthogonal channels to maximize the sum rate (e.g., see [39–41]). The *semi-orthogonal user selection* (SUS) chooses a UE with the largest channel gain at first, and then in each subsequent iteration, selects a UE such that its channel has the largest orthogonal component to the subspace spanned by the previously chosen UEs [59–61]. In [45], UEs are partitioned into groups, with each group having similar channel covariance. In [1], a UE clustering and set covering based scheme was developed to minimize the transmit power in multi-BS

networks. In [38], a scheme was presented to maximize the harmonic sum of UEs SINR in order to prioritize cell-edge UEs. The switched random beamforming has also been studied in the literature [48, 49], where a BS forms a beam in a random direction. The channel *chordal distance* (Chord-Dis) based scheme aims to maximize the sum rate, where UEs are chosen based on their channel chordal distances, which are a measure of orthogonality between UE channels (e.g., see [62–65]). The *UE partitioning and beamforming* (MUBFP) scheme maximizes the group-average sum rate for a single BS MU-MIMO system [37].

Most existing works focused on BS-UE channels for selecting UEs to maximize the sum rate. However, this often results in a large transmission time to deliver all UEs traffic, because to achieve a high sum rate, UEs with good channels (high SINR) would be allocated with much more network resources such as the transmit power and bandwidth. As a result, the transmission time for UEs with poor channels can be very large. Moreover, most of those works focused on one BS, while the UE association among multiple BSs assumed a simple scheme based on the received signal power. BOOST jointly considers UE association among multiple BS, UE scheduling, and beamforming altogether, to reduce the transmission time of the entire network.

## 3.2 SYSTEM MODEL

We assume BSs are connected to a cloud or a backend management system of the network operator, through either wired or wireless backhaul connections. The BSs information, including the location, number of antennas, etc., is known to the cloud which manages UE association and scheduling. We also assume all BSs and UEs are synchronized on both time and frequency for each time slot, in which a set of UEs are scheduled for traffic transmission. Table 5 lists major notations.

### 3.2.1 MMWAVE CHANNEL

Let $\mathcal{N}$, $\mathcal{M}$ and $\mathcal{K}$ be the set of UEs, BSs, and beams, respectively. Let $\mathcal{N}_m \subseteq \mathcal{N}$ be the

set of UEs that can be covered by BS $m$ under the maximum range, which is determined by a beam with the minimum beamwidth and the maximum BS transmit power. Throughout this chapter, we assume each BS is equipped with an antenna array with $L$ antenna elements, and each UE has a single antenna. Nevertheless, BOOST can be easily extended to accommodate UEs with multiple antennas. Let $K$ be the number of beamformers (or RF chains) for a BS. Each beamformer can form one beam. Note that the required processing power, design complexity and fabrication cost of a BS grows with the number of beamformers; hence a BS can only have a limited number of beamformers. Let $\mathcal{K}_m \subseteq \mathcal{K}$ denote the set of beams of BS $m$. For each beam $k \in \mathcal{K}_m$, let $\mathcal{N}_{mk} \subseteq \mathcal{N}_m$ denote the set of UEs that are inside beam $k$ of BS $m$, called beam $\langle m, k \rangle$.

Let $d_{mn}$ and $\mathrm{PL}_{mn}$ denote the distance and average LOS path loss, respectively, between BS $m$ and UE $n$. According to the 3GPP UMi-street canyon LOS model, the large scale path loss in dB is given as [66]

$$\mathrm{PL}_{mn}[dB] = 10\eta \log_{10}\left(\frac{d_{mn}}{d_0}\right) + 20\log_{10}\left(\frac{4\pi d_0 \times 10^9}{c}\right) + 20\log_{10}(f) + \mathcal{X}_{\sigma_{SF}}, \qquad (19)$$

where $\eta$ is the path loss exponent of the LOS path, $d_0$ is the close-in free space reference distance which is usually 1, $c$ is the speed of light, $f$ is the system frequency in GHz, and $\mathcal{X}_{\sigma_{SF}}$ is a zero-mean Gaussian random variable with a standard deviation $\sigma_{SF}$ in decibels. Let $\theta_{mn}$ be the *angle of arrival* (AoA) of UE $n$, i.e., the direction of the downlink signal to UE $n$ with regard to the BS $m$ antenna, and $\alpha_{mn}$ be the complex power gain of small-scale fading between BS $m$ and UE $n$, which is modeled using a complex Gaussian distribution, i.e., $\alpha_{mn} \sim \mathcal{CN}(0, 1)$. We assume a block fading channel between UEs and BSs, as in the existing cellular and WLAN standards [45]. The channel characteristics are about the same in the order of seconds. The channel vector between BS $m$ with $L$ antenna elements and UE $n$ ($n \in \mathcal{N}_{mk}$) with a single antenna is given as [48]

$$\mathbf{h}_{mn} = \sqrt{\frac{L}{\mathrm{PL}_{mn}}}\alpha_{mn}\mathbf{a}(\theta_{mn}), \tag{20}$$

where $\mathbf{a}(\theta)$ is the response vector of the BS antenna when the UE AoA is $\theta$. The response vector for a *uniform linear array* (ULA) at a BS is given as

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{L}}\left[1, \quad e^{-j2\pi\theta}, \ldots, e^{-j2\pi(L-1)\theta}\right]^T, \tag{21}$$

where $(\bullet)^T$ denotes the transpose of a vector. The concatenated channel matrix formed by all UEs covered by BS $m$ is therefore written as $\mathbf{H}_m = \left[\mathbf{h}_{m,1}, \ldots, \mathbf{h}_{m,|\mathcal{N}_m|}\right]$. The received signals at all UEs covered by BS $m$ are given as

$$\mathbf{Y}_m = \mathbf{H}_m^H \mathbf{W}_m \mathbf{x}_m + \mathbf{z}, \tag{22}$$

where $(\bullet)^H$ denotes the complex conjugate transpose, $\mathbf{x}_m \in \mathbb{C}^{|\mathcal{N}_m|\times 1}$ is the vector of transmitted signals to UEs of BS $m$, $\mathbf{z} \in \mathbb{C}^{|\mathcal{N}_m|\times 1}$ is the total noise, including the interference and the Additive White Gaussian Noise (AWGN) with zero mean and unit variance, $\mathbf{W}_m = \left[\mathbf{w}_{m,1}, \ldots, \mathbf{w}_{m,|\mathcal{K}_m|}\right] \in \mathbb{C}^{L\times|\mathcal{K}_m|}$ is the beamforming weight matrix of BS $m$, formed by concatenating individual weight vectors $\mathbf{w}$.

### 3.2.2 DOWNLINK TRANSMISSION TIME

In BOOST, each BS $m$ constructs a set of unit-power beams, denoted as $\{\mathbf{w}_{m,1}, \ldots, \mathbf{w}_{m,|\mathcal{K}_m|}\}$, with the power of each beam $||\mathbf{w}||^2 = 1$. As illustrated in Fig. 15(d), each beam serves a group of UEs in a time slot. Let $\hat{\theta}_{mk}$ denote the direction of the main lobe of beam $\langle m, k\rangle$, formed by $\mathbf{w}_{mk}$ to serve the UEs in $\mathcal{N}_{mk}$. For a beam covering multiple UEs, the beam direction $\hat{\theta}_{mk}$ may not be aligned exactly with the AoA of a UE. Let $\theta_{n,m,k} = |\hat{\theta}_{mk} - \theta_{mn}|$ denote the difference between the beam direction $\hat{\theta}_{mk}$ and the UE AoA $\theta_{mn}$. By the cosine antenna pattern of the ULA antenna, the effective channel gain of UE $n$

**Table 5.** Major notations in BOOST

| | |
|---|---|
| $m, n, k$ | index for BS, UE and beam, respectively |
| $\mathcal{M}, \mathcal{N}, \mathcal{K}$ | set of BSs, UEs, and beams, respectively |
| $\langle m, k \rangle$ | the $k$-th beam of the $m$-th BS |
| $\theta_{mn}, \mathbf{a}(\theta_{mn})$ | angle of arrival of UE $n$ signal from BS $m$, and the corresponding steering vector |
| $b_n$, $t_{n,m,k}$ | traffic load of UE $n$, and its transmission time when served by beam $\langle m, k \rangle$ |
| $\mathbf{h}_{mn}$ | channel gain vector between BS $m$ and UE $n$ |
| $\mathbf{w}_{mk}$ | beamforming weight vector of beam $\langle m, k \rangle$ |
| $p_{mk}$, $p_o$ | transmit power of beam $\langle m, k \rangle$ and power budget in each BS |
| $K$ | maximum number of beamformers at a BS |
| $I_{n,m,k}, \lambda_{n,m,k}, \gamma_{n,m,k}$ | interference, rate, and SINR at UE $n$ when it is served by beam $\langle m, k \rangle$ |
| $\gamma_o$ | minimum required SINR for UEs |

covered by beam $\langle m, k \rangle$ can be written as [67, 68]

$$|\mathbf{h}_{mn}^H \mathbf{w}_{mk}|^2 \approx \begin{cases} \frac{L|\alpha_{mn}|^2}{\mathrm{PL}_{mn}} \cos^2\left(\frac{L\pi\theta_{n,m,k}}{2}\right) & \text{if } \theta_{n,m,k} \leq \frac{1}{L}, \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

where $(\bullet)^H$ denotes the complex conjugate transpose, $\alpha_{mn}$ is the complex gain of the LOS path between BS $m$ and UE $n$, and $\mathrm{PL}_{mn}$ is the average LOS path loss between BS $m$ and UE $n$. In (23), the UE channel gain is at maximum when $\theta_{n,m,k} = 0$, i.e., when the beam direction perfectly matches the UE AoA. The cosine antenna pattern provides a good approximation for the main lobe gain [67, 68]. Note that the main impact of sidelobes is to cause interference to other beams, which can be efficiently suppressed if there is a sufficient

angular distance between two beams, e.g., using the *linearly constrained minimum variance* beamforming technique (to be discussed later). Therefore, in this chapter, we do not consider sidelobes.

Let $p_{mk}$ be the transmit power of beam $\langle m,k\rangle$. Let $I_{n,m,k}$ denote the total interference at UE $n$ ($n \in \mathcal{N}_{mk}$) from all beams. Let $\sigma^2$ be the thermal noise at a UE, which is modeled as $\sigma^2 = N_o + 10\log(v) + \text{NF}$, where $v$ is the system bandwidth, $N_o$ is the noise power spectral density and NF denotes the noise figure at the UE. The received SINR at UE $n$ from beam $\langle m,k\rangle$ is given as

$$\gamma_{n,m,k} = \frac{p_{mk}|\mathbf{h}_{mn}^H \mathbf{w}_{mk}|^2}{I_{n,m,k} + \sigma^2}, \quad \forall n \in \mathcal{N}_{mk}. \tag{24}$$

The total interference $I_{n,m,k}$ is the summation of intra-BS interference and inter-BS interference given as follows.

$$I_{n,m,k} = \underbrace{\sum_{i \in \mathcal{K}_m \setminus k} p_{mi}|\mathbf{h}_{mn}^H \mathbf{w}_{mi}|^2}_{\text{intra-BS interference}} + \underbrace{\sum_{j \in \mathcal{M}\setminus m} \sum_{l \in \mathcal{K}_j} p_{jl}|\mathbf{h}_{jn}^H \mathbf{w}_{jl}|^2}_{\text{inter-BS interference}}, \tag{25}$$

where $\mathcal{M}$ denotes the set of BSs and $\mathcal{K}_m$ denotes the set of beams of BS $m$.

To achieve a certain quality of service for a UE, the SINR at the UE has to be greater than or equal to a minimum SINR $\gamma_o$, i.e.,

$$\gamma_{n,m,k} \geq \gamma_o, \quad \forall n \in \mathcal{N}_{mk}. \tag{26}$$

In order to maintain the minimum $\gamma_o$ at all UEs in a beam, the required transmit power $p_{mk}$ is obtained using (24)-(26) as

$$p_{mk} = \max_{n \in \mathcal{N}_{mk}} \frac{(I_{n,m,k} + \sigma^2)\gamma_o}{|\mathbf{h}_{mn}^H \mathbf{w}_{mk}|^2}. \tag{27}$$

The transmit power vector for all beams of BS $m$ is given as

$$\mathbf{p}_m = \left[p_{m1}, \ldots, p_{m|\mathcal{K}_m|}\right]. \tag{28}$$

In each time slot, the transmit power from all beams of a BS must not be larger than the power budget $p_o$, i.e.,

$$\sum_{k \in \mathcal{K}_m} p_{mk} \leq p_o, \quad \forall m \in \mathcal{M}. \tag{29}$$

We assume each UE $n$ has a downlink traffic load $b_n$ that needs to be delivered with a minimum SINR $\gamma_o$. Let $\Delta_t$ be the length of each time slot. The minimum number of time slots needed to carry the traffic load $b_n$ to the UE is $\left\lceil \frac{b_n}{v_{n,m,k}\Delta_t \log_2(1+\gamma_{n,m,k})} \right\rceil$, where $v_{n,m,k}$ is the sum of spectrum from all downlink sub-carriers assigned to UE $n$ and $\gamma_{n,m,k}$ is the received SINR.

### 3.3 PROBLEM STATEMENT

As stated in the preceding section, we assume all BSs in the network are synchronized on time and frequency at each time slot. BSs schedule a set of UEs for downlink transmission in each slot. Our goal is to minimize the total number of time slots required to complete delivery of all UEs traffic loads, subject to the BS power budget, the minimum SINR requirement at UEs, and the number of beamformers of BSs. If there is more than one UE in a beam, the throughput capacity of the downlink channel is shared among all UEs. The UEs can use a multiplexing scheme to share resources, e.g., the *time division multiple access* (TDMA), *orthogonal frequency division multiple access* (OFDMA), or *non-orthogonal multiple access* (NOMA). In the ensuing discussions, we assume OFDMA is used. Nevertheless, BOOST can be extended for NOMA and TDMA.

If UE $n$ is scheduled in beam $\langle m, k \rangle$, the data rate for UE $n$ in slot $t$, denoted as $\lambda_{n,m,k}^{[t]}$, is given as

$$\lambda_{n,m,k}^{[t]} = v_{n,m,k}^{[t]} \log_2(1 + \gamma_{n,m,k}^{[t]}), \tag{30}$$

where $v_{n,m,k}^{[t]}$ is the sum of spectrum from all downlink sub-carriers assigned to UE $n$ in the beam, and $\gamma_{n,m,k}^{[t]}$ is the SINR at UE $n$ in time slot $t$. Let $x_{nmk}^{[t]} \in \{0, 1\}$ be a binary variable to indicate whether UE $n$ is covered by beam $\langle m, k \rangle$ in time slot $t$. If $x_{nmk}^{[t]} = 1$, then the remaining traffic of UE $n$ to be transported in the next slot $t + 1$ is given as

$$b_n^{[t+1]} = \max(0, b_n^{[t]} - \Delta_t \lambda_{n,m,k}^{[t]}). \tag{31}$$

To reduce the number of slots required to transmit traffic from all UEs, we formulate the objective function to minimize the remaining UE traffic in each slot given in (31). For the ease of description, we drop the time slot index $t$ in the formulation below. We want to find the values of $x_{nmk}$, $\mathbf{p}_m$ (transmit power) in (28), and $\mathbf{W}_m$ (beamforming weight matrix) in (22) for each BS. Let $\mathbf{X}$ denote the matrix $[x_{nmk}]$. The UE association and scheduling can be formulated as a nonlinear programming problem as below

$$\min_{\mathbf{X}, \mathbf{p}_m, \mathbf{W}_m} \sum_{n \in \mathcal{N}} [b_n - x_{nmk} v_{n,m,k} \Delta_t \log_2(1 + \gamma_{n,m,k})] \tag{32}$$

$$\begin{aligned}
\text{s. t.} \quad & \text{C1} : \forall n \in \mathcal{N}, \quad \sum_{m \in \mathcal{M}} \sum_{k \in \mathcal{K}_m} x_{nmk} \leq 1 \\
& \text{C2} : \forall n \in \mathcal{N}, m \in \mathcal{M}, k \in \mathcal{K}_m, \quad \gamma_{n,m,k} \geq x_{nmk} \gamma_o \\
& \text{C3} : \forall m \in \mathcal{M}, \quad \sum_{k \in \mathcal{K}_m} p_{mk} \leq p_o \\
& \text{C4} : \forall m \in \mathcal{M}, \quad |\mathcal{K}_m| \leq K \\
& \text{C5} : \forall m \in \mathcal{M}, k \in \mathcal{K}_m, \sum_{n \in \mathcal{N}_{mk}} x_{nmk} v_{n,m,k} \leq v \\
& \text{C6} : \forall n \in \mathcal{N}, m \in \mathcal{M}, k \in \mathcal{K}_m, \quad v_{n,m,k} \Delta_t \log_2(1 + \gamma_{n,m,k}) \leq x_{nmk} b_n.
\end{aligned}$$

Constraint C1 guarantees each UE is associated with only one BS and only one beam. C2

ensures the received SINR at each UE is greater than the threshold. C3 ensures the power allocated in all beams of a BS cannot surpass the power budget $p_o$. C4 ensures the number of beams in a BS is less than the number of beamformers $K$ in a BS. C5 ensures that the sum of allocated spectrum for all scheduled UEs in a beam must be not more than the available system bandwidth $v$. C6 guarantees that the allocated spectrum to a UE is no more than it needs to carry its traffic load in a time slot. The second term of (32) is the weighted sum rate maximization problem which has been proven NP-hard [37, 67]. Therefore, in this chapter, we develop a heuristic framework, BOOST, to solve the optimization problem (32).

## 3.4 BOOST FRAMEWORK

In the literature, the UE discovery usually can be conducted through two approaches: beam sweeping [50, 51] or coexisting macrocells [18]. With the first approach, each BS sweeps the whole angular space and transmits initial signals, to discover UEs. This approach may result in a large delay. With the second approach, the co-existing macrocells such as LTE towers are used to discover UEs and send the information to the cloud of the network operator, which can achieve a low delay. Given that major cellular operators all have deployed LTE systems, this is a practical approach.

Through UE discovery, the UEs information, including locations and AoAs, is known to the network operator. BOOST resides in the cloud of the operator to optimize UEs association and scheduling. Specifically, for given UE traffic loads, BOOST finds the optimal number of beams for each BS, assigns UEs to be covered by each beam, computes the beamforming vectors, and allocates spectrum for UEs and power for all beams in each slot, until the traffic of all UEs is delivered. The objective is to reduce the *network transmission time*, i.e., the total number of time slots required to deliver the UEs traffic, subject to the BS power budget, number of beamformers for BSs, and minimum UE SINR.

As illustrated in Fig. 15, BOOST contains three stages: 1) *UE clustering for interference control* (CIC), 2) *traffic balancing UE association* (TUA), and 3) *beamforming and UE traffic*

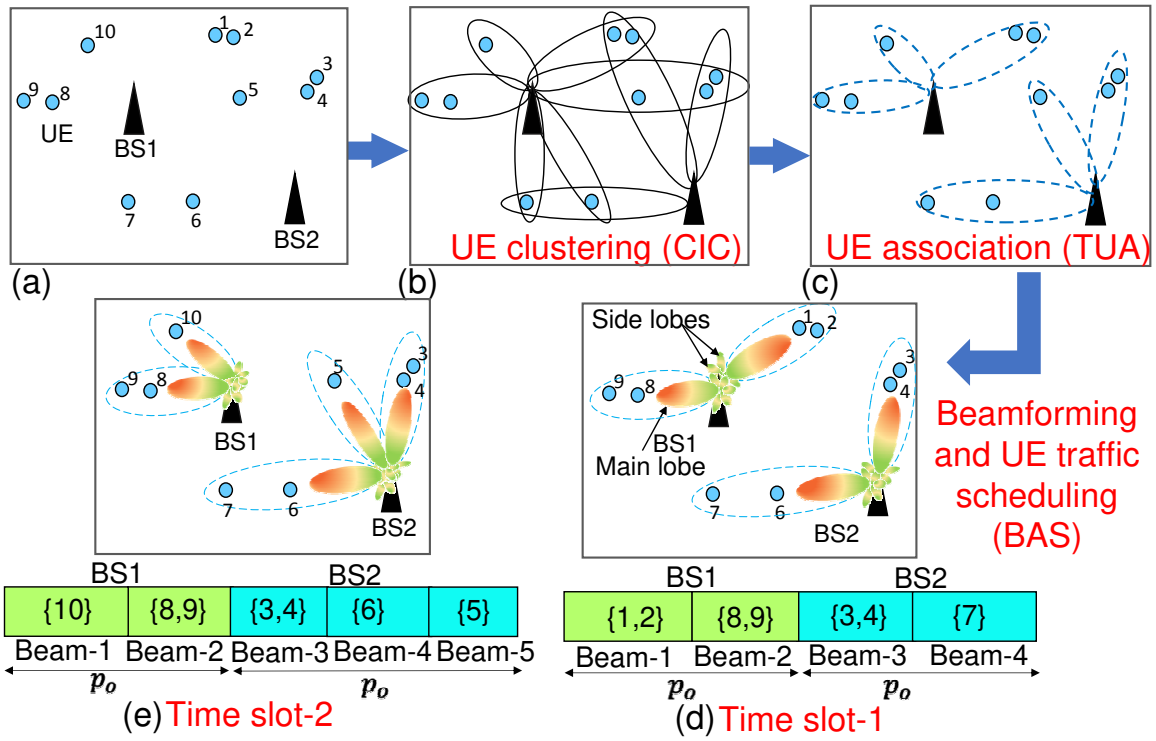



**Figure 15.** BOOST framework: (a) a mmWave network with 2 BSs and 10 UEs, (b) UE clustering to reduce intra-BS interference, (c) UE association to balance traffic and reduce inter-BS interference, (d-e) scheduling UE traffic into two slots and forming beams, constrained by the power budget and UE SINR requirement [6].

*scheduling* (BAS). In our experiments, we have observed that if a group of UEs are close to each other, then forming multiple beams to individual UEs creates significant interference between those beams. Therefore, for each BS, we first group its UEs into clusters with a CIC algorithm, based on their AoAs and spatial distances. The clusters have a radial shape with the BS as the center, as illustrated in Fig. 15(b). In the CIC stage, the UEs of a BS refer to the UEs that can be reached with the maximum range of the BS through beamforming. Many UEs may be reached by two or more BSs. Hence, different BSs can have *overlapping clusters* due to such UEs. For example, in Fig. 15(b), CIC forms nine clusters for a network with ten UEs and two BSs. The three clusters of BS2 are all overlapped with clusters of BS1. There are 7 *overlapping UEs* included in these overlapping clusters.

While the BS association for non-overlapping UEs is straightforward, since each of them can be reached by only one BS, it is a challenge to properly associate overlapping UEs such that the network transmission time is minimized while the inter-BS interference is avoided or significantly reduced. As illustrated in Fig. 15(c), TUA smartly associates overlapping UEs to the clusters based on the cluster transmission time, to achieve this objective. The CIC and TUA algorithms are re-run when there is a significant change to the channel. Typically the channel characteristics between a UE and a BS remain about the same in the order of seconds. Therefore, the CIC and TUA algorithms do not result in significant overhead.

Due to the power budget constraint, and the minimum SINR requirement, a BS often cannot form beams to cover all clusters simultaneously, with one beam to cover one cluster. Hence, BAS has to schedule the beams of a BS across a number of time slots. For example, the clusters in Fig. 15(c) have to be covered across two slots illustrated in Fig. 15(d) and (e). In each slot, BAS computes beamforming vectors for selected clusters, allocates the transmit power for each beam, and allocates spectrum for each UE in the beam to carry its traffic. At last, some beams may each have to be scheduled across multiple time slots, as the traffic transportation for the covered UE set cannot be finished in one slot. For example, in Fig. 15(d) and (e), the beams covering UE sets {3, 4}, {6, 7}, and {8, 9} are scheduled across two time slots while the beams covering sets {1, 2}, {5}, and {10} are scheduled in one slot. The objective of BAS is to minimize the network transmission time, i.e., the number of slots to finish traffic delivery for all beams/UEs. Next, we describe each component of BOOST in detail.

### 3.4.1 UE CLUSTERING FOR INTERFERENCE CONTROL (CIC)

Clustering is generally NP-hard [54]. Hence, in practice, heuristic algorithms are widely used. The challenge for clustering UEs is that unlike classic clustering that typically uses the spatial distance, we have to take beamforming into account. Furthermore, we also have to design the CIC algorithm to help achieve the objective of BOOST, reducing latency. In

this chapter, we design a heuristic clustering algorithm to reduce the interference between beams, and increase the average SINR at UEs in each cluster, both of which help to increase capacity and reduce latency. Moreover, UEs are grouped based on both AoAs (with regard to the BS) and the spatial distance.

Algorithm 3 describes the CIC algorithm. At the beginning each UE is treated as a cluster with the corresponding AoA for the beamforming direction. We use the equal power allocation for all clusters of a BS to compute the UE SINR. In each successive iteration, the CIC algorithm identifies two best candidate clusters such that merging them could increase the UE SINR (line 7–9). We select two candidate clusters with the minimum angular distance, which also needs to be not larger than $1/L$ (line 4), because the beamforming gain approaches zero by (23) if it is larger than $1/L$. This process continues until we cannot find such candidate clusters or all UEs are processed. The CIC algorithm output is the set of UE clusters in $\mathcal{K}$. Note that we do not limit the number of beams, nor consider the inter-BS interference between beams in the CIC stage, which will be taken care of by TUA and BAS.

Both CIC and TUA use the cosine pattern channel gain given in (23) to compute the UE interference and SINR by (25) and (24), respectively, while BAS computes the actual channel gain and SINR using the beamforming vector.

### 3.4.2 TRAFFIC BALANCING UE ASSOCIATION (TUA)

In BOOST, the CIC algorithm groups UEs into clusters for each BS independently, assuming the maximum range for the BS. Nevertheless, we should not directly form beams to cover those clusters, because clusters from different BSs can overlap with each other, resulting in significant inter-BS interference. TUA selects a unique BS (or the corresponding cluster) for each overlapping UE (belonging to more than one cluster), so that each overlapping UE goes to one cluster only and there is no overlapping anymore between clusters. The objective of TUA is to reduce transmission time, which is closely relevant to the traffic load. This is achieved through two techniques. First, to reduce the network transmission time, BOOST

**Algorithm 3:** UE Clustering for Interference Control [6]

**Input:** UE set $\mathcal{N}_m$ of BS $m$, AoAs $\hat{\theta}$, transmit power $p_o$

Let each UE be a cluster, with the set of clusters $\mathcal{K}_m = \{\{1\}, \ldots \{|\mathcal{N}_m|\}\}$, and their AoAs $\hat{\theta} = \{\hat{\theta}_1, \ldots, \hat{\theta}_{|\mathcal{N}_m|}\}$

$\mathcal{X} \leftarrow \emptyset$

**while** $\mathcal{N}_m \setminus \mathcal{X}$ *is not empty* **do**

  Select two clusters $(i, j) = \underset{(s,k)\in\mathcal{K}_m, s\neq k}{\operatorname{argmin}} |\hat{\theta}_s - \hat{\theta}_k|$ such that $|\hat{\theta}_s - \hat{\theta}_k| \leq 1/L$

  **if** $(i, j)$ *is empty* **then**

    break

  Find the UE SINR in the combined cluster $h = (i, j)$ and other clusters in $\mathcal{K}_m$ except $i$ and $j$, each with power $\frac{p_o}{|\mathcal{K}_m|-1}$ using (24)

  **if** *min. SINR for UEs in cluster* $h$ *is greater than the min. SINR in both clusters* $i$ *and* $j$ **then**

    Use cluster $h$ to replace clusters $i, j$ in $\mathcal{K}_m$, and accordingly update the AoAs $\hat{\theta}$.

  **else**

    Add all UEs in clusters $i$ and $j$ to $\mathcal{X}$

**Output:** $\mathcal{K}_m$ as the set of clusters of BS $m$

smartly balances UE traffic load between BSs. Second, to reduce the transmission time for a cluster, BOOST balances the transmission times of UEs in the cluster, i.e., avoids the situation where while some UEs in the cluster have completed traffic transmission, other UEs still have significant residual traffic to be transmitted. In other words, we would like the transmission time for every UE in the cluster to be the same. Hence we allocate spectrum for each UE in proportion to its traffic load and received SINR. Next, we find the transmission

time for a cluster.

Let $\mathcal{S}_{mk} \in \mathcal{K}_m$ denote the $k$-th UE cluster of BS $m$ obtained by the CIC algorithm. The UEs in cluster $\mathcal{S}_{mk}$ are covered by the corresponding beam $\langle m, k \rangle$ from BS $m$. Let $\gamma_{n,m,k}$ and $b_n$ denote the received SINR and traffic load of UE $n \in \mathcal{S}_{mk}$, respectively.

**Theorem 2.** *Let $v$ denote the total system bandwidth. If the system bandwidth is allocated to UEs of $\mathcal{S}_{mk}$ in proportion to their traffic loads and SINR, i.e., $\frac{b_n}{\log_2(1+\gamma_{n,m,k})}$, the transmission time for cluster $\mathcal{S}_{mk}$, denoted as $t(\mathcal{S}_{mk})$, is equal to the transmission time of every UE, and is given as*

$$t(\mathcal{S}_{mk}) = \sum_{i \in \mathcal{S}_{mk}} \frac{b_i}{v \log_2(1 + \gamma_{i,m,k})}. \tag{33}$$

*Proof.* Let $v_{n,m,k}$ denote the spectrum share for UE $n \in \mathcal{S}_{mk}$, the required transmission time for UE $n$, denoted as $t_{n,m,k}$, is given by

$$t_{n,m,k} = \frac{b_n}{v_{n,m,k} \log_2(1 + \gamma_{n,m,k})}. \tag{34}$$

If we allocate spectrum for each UE in proportion to its traffic load and received SINR, then the spectrum share for UE $n$ can be written as

$$v_{n,m,k} = \frac{\beta_n v}{\sum_{i \in \mathcal{S}_{mk}} \beta_i}, \tag{35}$$

where $\beta_n$ is the coefficient used for spectrum share of UE $n$, and is computed using its traffic load and SINR as below

$$\beta_n = \frac{b_n}{\log_2(1 + \gamma_{n,m,k})}. \tag{36}$$

After substituting $\beta_n$ into (35), the spectrum share for UE $n$ is derived as follows

$$v_{n,m,k} = \frac{v b_n \prod_{i \in \mathcal{S}_{mk}, i \neq n} \log_2(1+\gamma_{i,m,k})}{\sum_{i \in \mathcal{S}_{mk}} b_i \left[ \prod_{j \in \mathcal{S}_{mk}, j \neq i} \log_2(1+\gamma_{j,m,k}) \right]}. \tag{37}$$

After substituting $v_{n,m,k}$ into (34), the transmission time for UE $n$, $t_{n,m,k}$, becomes

$$t_{n,m,k} = \frac{b_n}{v_{n,m,k} \log_2(1+\gamma_{n,m,k})} = \sum_{i \in \mathcal{S}_{mk}} \frac{b_i}{v \log_2(1+\gamma_{i,m,k})}. \tag{38}$$

It is clear that by (38), the transmission time to complete traffic delivery to every UE in $\mathcal{S}_{mk}$ is the same, i.e., $t_{1,m,k} = \cdots = t_{|\mathcal{S}_{mk}|,m,k}$. Hence, (38) also denote the transmission time $t(\mathcal{S}_{mk})$ for cluster $\mathcal{S}_{mk}$. ■

Next, we use the cluster transmission time in (33) as the cost metric for UE association, to balance traffic amount and UEs between BSs, with the objective to reduce the network transmission time. For a BS, some of its clusters may be overlapped with clusters of other BSs, while others are not, as illustrated in Fig. 16. Note that the clusters of the same BS do not overlap with each other. Let $\mathcal{L}$ denote the set of UE clusters of all BSs which is obtained by the CIC algorithm. Let $\mathcal{R}$ denote the set of *non-overlapping* clusters in $\mathcal{L}$, i.e.,

$$\mathcal{R} = \{S_i \mid S_i \in \mathcal{L} \text{ and } \forall S_j \in \mathcal{L} \backslash S_i, j \neq i, S_i \cap S_j = \emptyset\}. \tag{39}$$

Let $\mathcal{R}_m \subseteq \mathcal{R}$ denote the set of non-overlapping clusters of BS $m$. We define the cost of BS $m$, denoted as $c_m$, as the transmission time of the BS, which is the maximum transmission time of its clusters. This is used in the TUA algorithm to determine which BS an overlapping UE should be associated to. However, the challenge is that we do not know the transmission times of overlapping clusters until the association of overlapping UEs is completed. Hence, during the execution of the TUA algorithm, we set the BS cost as the maximum transmission time of its clusters that have been processed so far. Initially, the BS cost is set as the maximum transmission time of the non-overlapping clusters, i.e.,

$$c_m = \max_{s \in \mathcal{R}_m}(t(s)), \text{ or } 0 \text{ if } \mathcal{R}_m = \emptyset. \tag{40}$$

It is updated whenever an overlapping cluster of the BS is processed in the TUA algorithm. That is, $c_m$ dynamically changes whenever a UE is associated to BS $m$.

TUA only needs to be applied on UEs in overlapping clusters. TUA associates an overlapping UE (belonging to more than one cluster) to the right cluster, such that the interference between BSs is reduced. *In the ensuing discussions, for the ease of description, we interchangeably refer to UE association to a cluster as to the BS of the cluster.* Algorithm 4 illustrates the TUA algorithm. The main idea is that for each overlapping UE $n$, we associate it with the cluster that results in the lowest BS cost. This in turn results in a smaller network transmission time. In the algorithm, $\mathcal{X}$ denotes the UEs that have been associated so far. In each iteration, TUA picks a cluster $\mathcal{S}^*$ that has the highest traffic load. Then it chooses an unprocessed UE $n^*$ in cluster $S^*$ with the highest channel gain. Next, TUA finds all other clusters that overlap due to UE $n^*$, denoted as $\{\bar{S}_1, \ldots, \bar{S}_j, \ldots, \bar{S}_J \mid n^* \in \bar{S}_j \in \bar{\mathcal{L}}\}$. UE $n^*$ is associated to the cluster that results in the lowest BS cost (lines 9–13). While associating $n^*$, we bundle it with the UEs with a higher channel gain as they are likely to be associated together. Hence, in Algorithm 4, $\{S_1, \ldots, S_j, \ldots, S_J\}$ are used instead of $\{\bar{S}_1, \ldots, \bar{S}_j, \ldots, \bar{S}_J\}$.

Fig. 16 illustrates the main idea of TUA, where $S_1$ and $S_2$ are overlapping clusters, with UE $b$ being an overlapping UE. Fig. 16(b) illustrates the BS costs before processing UE $b$, which are the transmission times of clusters $S_3$ and $S_4$, respectively. Fig. 16(c) and (d) illustrate the BS cost change if we associate $b$ to cluster $S_1$ or $S_2$. In option 1, the BS1 cost changes to the transmission time of cluster $S_1$, as it is larger than the transmission time of cluster $S_3$. However, this is still lower than the BS2 cost. Hence the network transmission time is still the BS2 cost, which slightly increases due to the power splitting to cover UE $c$. In option 2, UE $b$ is associated to cluster $S_2$. To cover both $b$ and $c$, BS2 needs to split

**Algorithm 4:** Traffic Balancing UE Association (TUA) [6]

**Input:** Clusters $\mathcal{L}$ from CIC, UE set $\mathcal{N}$

**Output:** Clusters $\bar{\mathcal{L}}$ with UEs uniquely associated

$\bar{\mathcal{L}} \leftarrow \mathcal{L} \setminus \mathcal{R}$. $\quad \mathcal{X} = \bigcup_{S \in \mathcal{R}} S$

Compute the initial cost vector $\mathbf{c}$ of $|\mathcal{M}|$ BSs by (40)

**while** $\mathcal{X} \neq \mathcal{N}$ **do**

  Select a cluster $S^* = \underset{S \in \bar{\mathcal{L}}}{\operatorname{argmax}} \sum_{n \in S} b_n$ and $S^* \setminus \mathcal{X} \neq \emptyset$

  Select a UE $n^* = \underset{n \in S^* \setminus \mathcal{X}}{\operatorname{argmax}} \|\mathbf{h}_n\|^2$

  Let $\{\bar{S}_1, \ldots, \bar{S}_j, \ldots, \bar{S}_J \mid n^* \in \bar{S}_j \in \bar{\mathcal{L}}\}$ denote the clusters in $\bar{\mathcal{L}}$ that include $n^*$

  For $1 \leq j \leq J$, $S_j = \{n^*\} \bigcup \{n | n \in \bar{S}_j, \|\mathbf{h}_n\|^2 > \|\mathbf{h}_{n^*}\|^2\}$

  **if** *$n^*$ is an overlapping UE, i.e. $J > 1$* **then**

    For $1 \leq j \leq J$, compute the temporary cost of the BS of $S_j$,

    $c'_{m(S_j)} = \max(c_{m(S_j)}, t(S_j))$

    $j^* = \underset{1 \leq j \leq J}{\operatorname{argmin}} c'_{m(S_j)}$

    For any $S \in \bar{\mathcal{L}}, S \neq S_{j^*}, S \leftarrow S \setminus S_{j^*}$

    Update BS $m(S_{j^*})$ cost $c_{m(S_{j^*})} = c'_{m(S_{j^*})}$

    $\mathcal{X} = \mathcal{X} \bigcup S_{j^*}$

  **else**

    If $t(S_1) > c_{m(S_1)}$, let $c_{m(S_1)} = t(S_1)$

    $\mathcal{X} = \mathcal{X} \bigcup S_1$

more power; hence, the transmission time of $S_4$ increases further, which results in a higher network transmission time compared with option 1. Hence TUA selects option 1.

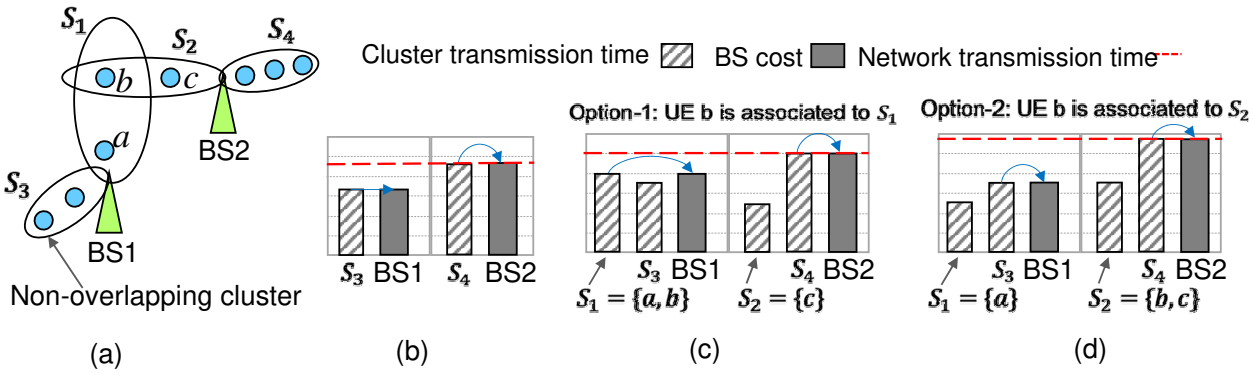



**Figure 16.** UE association by TUA, (a) a 2-BS network, (b) before processing UE $b$, (c) associating $b$ to S1, (d) associating $b$ to S2 [6].

### 3.4.3 BEAMFORMING AND UE TRAFFIC SCHEDULING (BAS)

Given the UE clusters obtained by TUA, the BAS algorithm schedules which clusters into a time slot. Then it computes the BS beamforming weight vectors, and allocates power for each beam. Through scheduling as much traffic as possible into every slot, BAS aims to reduce the number of time slots required to deliver traffic of all UEs subject to the BS power budget, UE SINR requirement, and UE traffic loads.

#### Beamforming Weight Vector

Similar to the previous chapter, we use the *linearly constrained minimum variance* (LCMV) scheme to compute beamforming weight vectors [56, p.513]. LCMV is able to suppress the power response toward the directions of other beams if there is a sufficient angular space from those beams. To achieve a complex gain $g^*$ in the UE direction $\theta$, the weight vector $\mathbf{w}$ needs to satisfy $\mathbf{a}(\theta)^H\mathbf{w} = g^*$, where $\mathbf{a}(\theta)$ is given in (21). Let $\mathbf{A}_\theta = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_k)]$ be the constraint matrix for total $k$ UEs of a BS, with AoAs $\theta = \{\theta_1, \ldots, \theta_k\}$. Let $\mathbf{f}$ be the $k$-dimension single column response vector of the $k$ UEs. The covariance matrix of the $k$ UE signals $\mathbf{R} = \mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right] = \frac{L|\alpha|^2}{\mathrm{PL}}\mathbf{A}_\theta\mathbf{A}_\theta^H$, where $\mathbf{h}$ is a matrix consisted of the $k$ UE channel vectors given in (20), $L$ is the number of antenna elements of a BS, PL is the path loss vector

of the $k$ UEs, and $\alpha$ is the complex gain vector of the $k$ UE signals. The LCMV beamforming to minimize the transmit power is formulated as

$$\min\{\mathbf{w}^H\mathbf{R}\mathbf{w}\} \text{ subject to } \mathbf{A}_\theta^H\mathbf{w} = \mathbf{f}. \tag{41}$$

The closed-form solution of (41) is obtained as follows using the Lagrange multiplier method

$$\mathbf{w} = \mathbf{R}^{-1}\mathbf{A}_\theta(\mathbf{A}_\theta^H\mathbf{R}^{-1}\mathbf{A}_\theta)^{-1}\mathbf{f}. \tag{42}$$

By choosing an appropriate weight vector for each UE cluster, it is possible to approximately eliminate or significantly reduce the interference between beams, as long as they are sufficiently separated in the angular space. The time complexity of the LCMV beamforming in (42) is $\Theta(\max(kL^2, L^{2.373}))$ assuming $L \geq k$, where $L^{2.373}$ is the matrix inverse time for an $L \times L$ matrix.

**Interference Controlled Power Allocation**

While BAS schedules a cluster into a time slot, it may select to cover only a subset of UEs in the cluster. Let $\mathcal{F}$ denote the list of UE sets to be scheduled in a slot. Let $\mathcal{F}_i$ denote the $i$th set in $\mathcal{F}$. Let $m_i$ denote the corresponding BS for set $\mathcal{F}_i$. For the UE sets $\mathcal{F}_i$ $(1 \leq i \leq |\mathcal{F}|)$, we construct the constraint matrix $\mathbf{A}_\theta$, the corresponding covariance matrix $\mathbf{R}$, and the response vector $\mathbf{f}$. Then we use (42) to compute $|\mathcal{F}|$ number of normalized weight vectors $\mathbf{w}_1, \ldots, \mathbf{w}_{|\mathcal{F}|}$, where $\mathbf{w}_i$ is for set $\mathcal{F}_i$. By (27), the transmit power allocated to the beam to cover set $\mathcal{F}_i$ is computed as

$$p_i = \max_{n \in \mathcal{F}_i}\left(\frac{(I_n + \sigma^2)\gamma_o}{|\mathbf{h}_{m_i,n}^H\mathbf{w}_i|^2}\right), \quad \forall\ 1 \leq i \leq |\mathcal{F}| \tag{43}$$

where $\mathbf{h}_{m_i,n}$ is the channel vector between BS $m_i$ and UE $n$, $\gamma_o$ is the required minimum SINR threshold for all UEs, and $\sigma^2$ is the noise power in the AWGN channel at the UE.

The power computed by (43) is the minimum power to ensure the SINR threshold for UEs. If a BS has more power than needed for its beams, the following theorem states that the residual power can be allocated to the beams in proportion to their minimum power, while achieving higher SINRs for all UEs.

**Theorem 3.** *Let $\mathbf{p}_m$ denote the power vector of the beams of BS $m$ computed by (43) and $p_{mi}$ denote the power of beam $< m, i >$. Let $\gamma_{n,m,i}(\mathbf{p}_m)$ denote the SINR of UE $n$ in beam $< m, i >$ under power allocation $\mathbf{p}_m$. If $\sum_i p_{mi} < p_o$, the power allocation $\mathbf{p}'_m = \frac{p_o}{\sum_i p_{mi}} \mathbf{p}_m$ meets the BS power constraint $p_o$, while the UE $n$ SINR $\gamma_{n,m,i}(\mathbf{p}'_m)$ is larger than the original SINR $\gamma_{n,m,i}(\mathbf{p}_m)$.*

*Proof.* The total transmit power of the BS under $\mathbf{p}'_m$ is $\sum_i \frac{p_o}{\sum_i p_{mi}} p_{mi} = \frac{p_o}{\sum_i p_{mi}} \sum_i p_{mi} = p_o$. Hence it meets the power budget requirement. With BS power vector $\mathbf{p}_m$, the SINR at UE $n$ is given by (24) as $\gamma_{n,m,i}(\mathbf{p}_m) = \frac{p_{mi}|\mathbf{h}_{mn}^H \mathbf{w}_{mi}|^2}{I_{n,m,i}+\sigma^2}$, where $\sigma^2$ is the noise power and $I_{n,m,i}$ is the sum of intra-BS and inter-BS interference at UE $n$ given by (25). As the CIC algorithm ensures that AoAs of UEs in different clusters of the same BS are spatially separated by at least $1/L$, where $L$ is the number of antennas at BS, the constraint in (41) guarantees negligible interference between UEs in different clusters of the same BS. That means the normalized intra-BS interference at UE $n$ from other beams $\langle m, j \rangle$, $j \neq i$, is approximately 0, i.e.,

$$\sum_{j \in \mathcal{K}_m \setminus i} |\mathbf{h}_{mn}^H \mathbf{w}_{mj}|^2 = \sum_{j \in \mathcal{K}_m \setminus i} \frac{L|\alpha_{mn}|^2}{\mathrm{PL}_{mn}} |\mathbf{a}^H(\theta_{mn}) \mathbf{w}_{mj}|^2 \approx 0,$$

where $\alpha_{mn}$ is the signal path complex gain, $\mathrm{PL}_{mn}$ is the path loss, and $\mathbf{a}^H(\theta_{mn})$ is the BS $m$ antenna response vector toward the direction of UE $n$. For clusters from different BSs, the inter-BS interference becomes very high if the angular spatial separation between those clusters is small. Line 12 of Algorithm 5 (to be discussed) effectively prevents such

significantly interfering beams of different BSs from being scheduled in the same slot. This technically reduces the inter-BS interference to close to zero. Thus, the normalized inter-BS interference at UE $n$ from beams $\langle k, l\rangle$, $k \neq m$, is

$$\sum_{k\in\mathcal{M}\setminus m}\sum_{l\in\mathcal{K}_k}\frac{L|\alpha_{kn}|^2}{\mathrm{PL}_{kn}}\underbrace{|\mathbf{a}^H(\theta_{kn})\mathbf{w}_{kl}|^2}_{\approx 0}\approx 0.$$

As a result, increasing power in one beam does not cause significant interference to another beam since both interference $I'_{n,m,i}$ with power $\mathbf{p}'_m$ and interference $I_{n,m,i}$ with power $\mathbf{p}_m$ are very small. If $\sum_i p_{mi} < p_o$, we have $p'_{mi} > p_{mi}$ by the allocation of residual power. Therefore, we have

$$\gamma_{n,m,i}(\mathbf{p}'_m)=\frac{p'_{mi}|\mathbf{h}_{mn}^H\mathbf{w}_{mi}|^2}{I'_{n,m,i}+\sigma^2}\approx\frac{p'_{mi}|\mathbf{h}_{mn}^H\mathbf{w}_{mi}|^2}{\sigma^2}>\frac{p_{mi}|\mathbf{h}_{mn}^H\mathbf{w}_{mi}|^2}{\sigma^2}\approx\frac{p_{mi}|\mathbf{h}_{mn}^H\mathbf{w}_{mi}|^2}{I_{n,m,i}+\sigma^2}=\gamma_{n,m,i}(\mathbf{p}_m).$$

■

**UE Scheduling**

The UE scheduling has two components. First, for each BS, it selects which UE clusters for transmission in the current slot, subject to the BS power budget and number of beamformers. We use a greedy scheme to select clusters with larger traffic loads. Second, for a selected cluster for the current slot, depending on its total traffic load, the UE scheduling may select all UEs in the cluster, or a subset of UEs, or even a single UE for transmission. We introduce two propositions for selecting UEs in a cluster.

**Proposition 1.** *Let $\Delta_b$ denote the traffic load that can be delivered with SINR $\gamma_o$ in a slot duration. If the total traffic load of a cluster $\mathcal{S}$ is less than $\Delta_b$, i.e., $\sum_{n\in\mathcal{S}} b_n \leq \Delta_b$, the transmission time for cluster $\mathcal{S}$ can be reduced by $(|\mathcal{S}|-1)$ slots if a single beam is formed to cover all UEs in $\mathcal{S}$ rather than one beam per UE (multi-user beamforming).*

*Proof.* If *one beam per UE* is formed for all UEs simultaneously, the interference between

beams is typically very high as those UEs are very close to each other. Hence, the minimum SINR cannot be met. To avoid interference, each UE has to be scheduled into a different slot. Note that even though the transmission of the traffic of one UE needs only a fraction of slot duration, it is usually not possible to form a beam for another UE in the middle of a slot [69]. This results in total $|\mathcal{S}|$ slots. In contrast, if one beam is formed to cover all UEs in the cluster, the total traffic can be transmitted in one slot, reducing the transmission time by $(|\mathcal{S}| - 1)$ slots. ■

**Proposition 2.** *For a cluster $\mathcal{S}$, if $\max_{n\in\mathcal{S}} b_n < \Delta_b$ and $\sum_{n\in\mathcal{S}} b_n > \Delta_b$, i.e., the traffic load of any single UE is smaller than $\Delta_b$ but the total load is larger than $\Delta_b$, then a subset of UEs $\mathcal{S}' = \underset{s\subseteq\mathcal{S}}{\operatorname{argmin}} \left([\Delta_b - \sum_{n\in s} b_n] \geq 0\right)$ can be scheduled into the current slot. This helps to reduce the cluster $\mathcal{S}$ transmission time.*

*Proof.* The total UE traffic load of $\mathcal{S}'$ is not larger than $\Delta_b$. Hence by Proposition 1, the traffic transmission to all UEs in $\mathcal{S}'$ can be completed in one slot, by forming one beam to cover all UEs. Moreover, scheduling $\mathcal{S}'$ into the current slot minimizes the remaining traffic of cluster $\mathcal{S}$ to be transmitted in future slots. If such scheduling is repeatedly applied on cluster $\mathcal{S}$, the total cluster transmission time is effectively reduced. ■

Algorithm 5 illustrates the BAS algorithm. In each iteration, BAS chooses a cluster $s^*$ with the maximum traffic load (line 3). It then schedules UEs of $s^*$ to the current slot as follows. *First*, if the maximum-load UE $n^*$ has a traffic load larger than $\Delta_b$, then schedule only UE $n^*$ in the current slot (line 7). *Otherwise*, if the total traffic load of all UEs in $s^*$ is larger than $\Delta_b$, then by Proposition 2, schedule a subset $s' \subset s^*$ to this slot (line 9). Lastly, if the total traffic load of all UEs in $s^*$ is smaller than $\Delta_b$, then by Proposition 1, schedule the entire $s^*$ to this slot. The beamforming weight vectors and transmit power vectors for all clusters in $\mathcal{F}$ are then computed (line 11). Next, every BS is checked and enforced for the power budget constraint (line 12). This is a critical step to prevent significant interference between beams from different BSs to transmit in the same slot, as such interference likely

**Algorithm 5:** Beamforming and UE Traffic Scheduling [6]

**Input:** UE clusters $\bar{\mathcal{L}}$, UE traffic $\mathbf{b}^{[t]}$ at slot $t$

**Output:** Remaining UE traffic $\mathbf{b}^{[t+1]}$ for slot $t+1$

$\mathcal{F} \leftarrow \{\emptyset\}$. $\quad \mathcal{L} = \bar{\mathcal{L}} \cup R$

**while** $\mathcal{L} \neq \emptyset$ **do**

  $s^* = \underset{s}{\operatorname{argmax}} \left( \sum_{n \in s, s \in \mathcal{L}} b_n^{[t]} \right)$

  **if** *# of clusters in* $\mathcal{F}$ *for the BS of* $s^* < K$ **then**

    Let $s' = s^*$ and select a UE in $s'$ with max. traffic, i.e., $n^* = \underset{n \in s'}{\operatorname{argmax}} \left( b_n^{[t]} \right)$

    **if** $b_{n^*}^{[t]} \geq \Delta_b$ **then**

      $s' = \{n^*\}$

    **else if** $\sum_{n \in s^*} b_n^{[t]} > \Delta_b$ *and* $b_{n^*}^{[t]} < \Delta_b$ **then**

      $s' = \underset{s \subseteq s^*}{\operatorname{argmin}} \left( [\Delta_b - \sum_{n \in s} b_n^{[t]}] \geq 0 \right)$

    $\mathcal{F} \leftarrow \mathcal{F} \cup \{s'\}$ `// Add` $s'$ `to` $\mathcal{F}$

    Compute $\mathbf{w}_1, \ldots, \mathbf{w}_{|\mathcal{F}|}$ and $\mathbf{p} = \left[p_1, \ldots, p_{|\mathcal{F}|}\right]$ by (42) and (43)

    **if** *sum of power of all beams of any BS is greater than* $p_o$ **then**

      $\mathcal{F} \leftarrow \mathcal{F} \setminus s'$

  $\mathcal{L} \leftarrow \mathcal{L} \setminus s^*$

For all $m \in \mathcal{M}$, update $\mathbf{p}_m$ as $\mathbf{p}_m = \frac{p_o}{\sum_i p_{mi}} \mathbf{p}_m$ by Theorem 3

Compute the SINR $\gamma_n^{[t]}$, bandwidth $w_n^{[t]}$, and remaining traffic $b_n^{[t+1]}$ for each UE in the clusters in $\mathcal{F}$ by (24), (37), and (31)

Remove a UE from $\mathcal{N}$ and all clusters in $\bar{\mathcal{L}} \cup R$ if its traffic delivery can be finished in this slot

**Table 6.** System parameters in BOOST

| **Parameter** | **Value** |
|---|---|
| System operating frequency | 73 GHz |
| Number of BSs | 4 |
| Number of UEs | 100 |
| Network dimension | $[300 \times 300]$ m$^2$ |
| Number of antennas in ULA | 64 |
| Adjacent antenna spacing $\Delta_d$ in ULA | $\lambda/2$ |
| Number of beamformers $K$ per BS | 10 |
| System bandwidth (FDD duplex mode) | 400 MHz |
| Time slot duration $\Delta_t$ | 0.125 ms |
| LOS path loss exponent $\eta$ | 2.1 |
| Minimum required SINR $\gamma_o$ at UE | 15 dB |
| Noise power spectral density $N_o$ | $-174$ dBm/Hz |
| Noise figure NF at UE receiver | 6 dB |

results in violation of the BS power constraint. The power allocation in line 15 follows Theorem 3. The BAS algorithm continues for the next time slot, until the traffic of all UEs is scheduled.

## 3.5 PERFORMANCE EVALUATION

We evaluate the performance of BOOST comparatively with three state-of-the-art UE scheduling schemes in the literature: MUBFP [37], Chord-Dis [62], and SUS [60]. Those three schemes use *multiuser beamforming*, where one beam is formed for each UE. Table 2 lists system parameters used in simulations. We assume BSs have a ULA antenna with 64 elements. UEs are equipped with a single antenna. The sample networks are in a $[300 \times 300]$

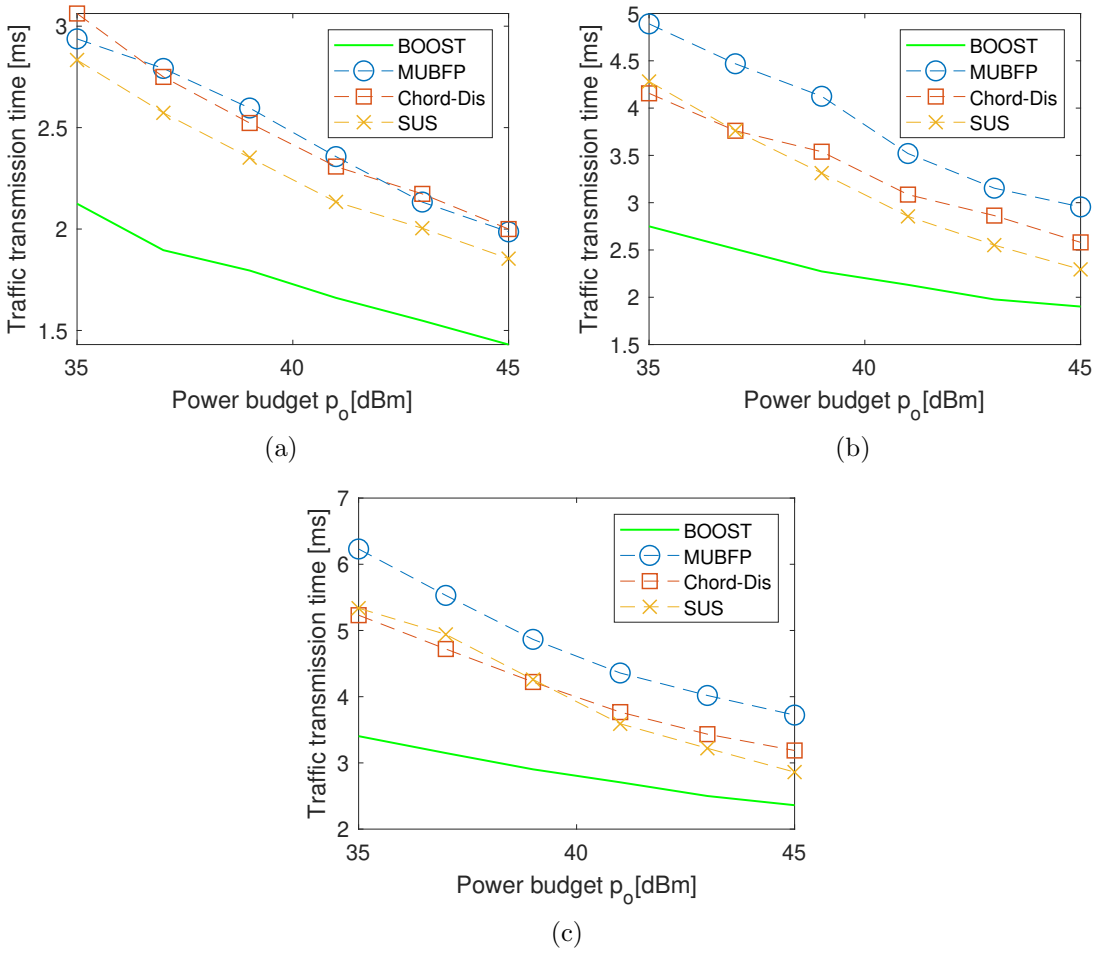



**Figure 17.** Transmission time vs. BS power budget under 3 UE distributions: (a) spread, (b) grouped, and (c) dense [6]

$m^2$ outdoor area, with 4 BSs and 100 active UEs. The BS positions are evenly distributed in the network. The UEs are distributed in three patterns: *spread*, *grouped*, and *dense*. With the spread UE distribution, UE locations are randomly generated in the network area. With the grouped UE distribution, 20 UE groups are created first and then each of the 100 UEs randomly selects a group. Each BS is assigned with 5 groups. Note that this assignment of groups to BSs is for the sole purpose of generating UE locations, irrelevant to the actual UE association. For each group, we randomly generate a 20 or 25-degree sector with regard to its BS. For each UE, we generate a random angle within the sector of its group, and

**Table 7.** Transmission time improvement of BOOST

| UE Pattern | vs. MUBFP | vs. Chord-Dis | vs. SUS |
|---|---|---|---|
| Spread | 29% | 29% | 24% |
| Grouped | 41% | 32% | 28% |
| Dense | 40% | 30% | 28% |

a random distance between 10 m and 106 m from the BS of its group. The dense UE distribution is similar to the grouped UE distribution, but with 3 UE groups for each BS. In the experiments, the LCMV beamforming scheme is used. The *reference signal received power* based UE association is used in MUBFP, Chord-Dis, and SUS. The OFDMA scheme is assumed for spectrum sharing among UEs in a beam. The time slot duration is assumed to be 0.125 ms, which is one of the slot durations defined in 3GPP [70]. The required minimum SINR $\gamma_o$ is set to 15 dB which is needed to use high rate modulation schemes [69]. The traffic load of each UE is randomly generated between 12.5 KB and 250 KB, following a uniform distribution. Note that this is not the UE data rate, but the traffic amount of the UE. As the UE location from the UE discovery may be inaccurate, we use a Gaussian distribution with a zero mean and 0.5 m standard deviation to model the error on the UE X and Y coordinates from the UE discovery [71]. The location error of UEs, i.e., the distance between the actual UE location and the reported location by the UE discovery, ranges from 0.1 m to 1.6 m. The experiment results using larger standard deviation values have similar trends and are omitted due to space limitations.

Fig. 17 illustrates the average network transmission time as a function of the BS power budget $p_o$, for 50 experiments using different seeds for UE location generation. Each data point indicates the network transmission time to complete delivery of the traffic loads of all UEs. Overall, BOOST significantly outperforms other schemes, thanks to its effective UE clustering, association, and scheduling, which altogether make it possible to form beams with significantly reduced interference. Moreover, BOOST not only achieves lower transmission

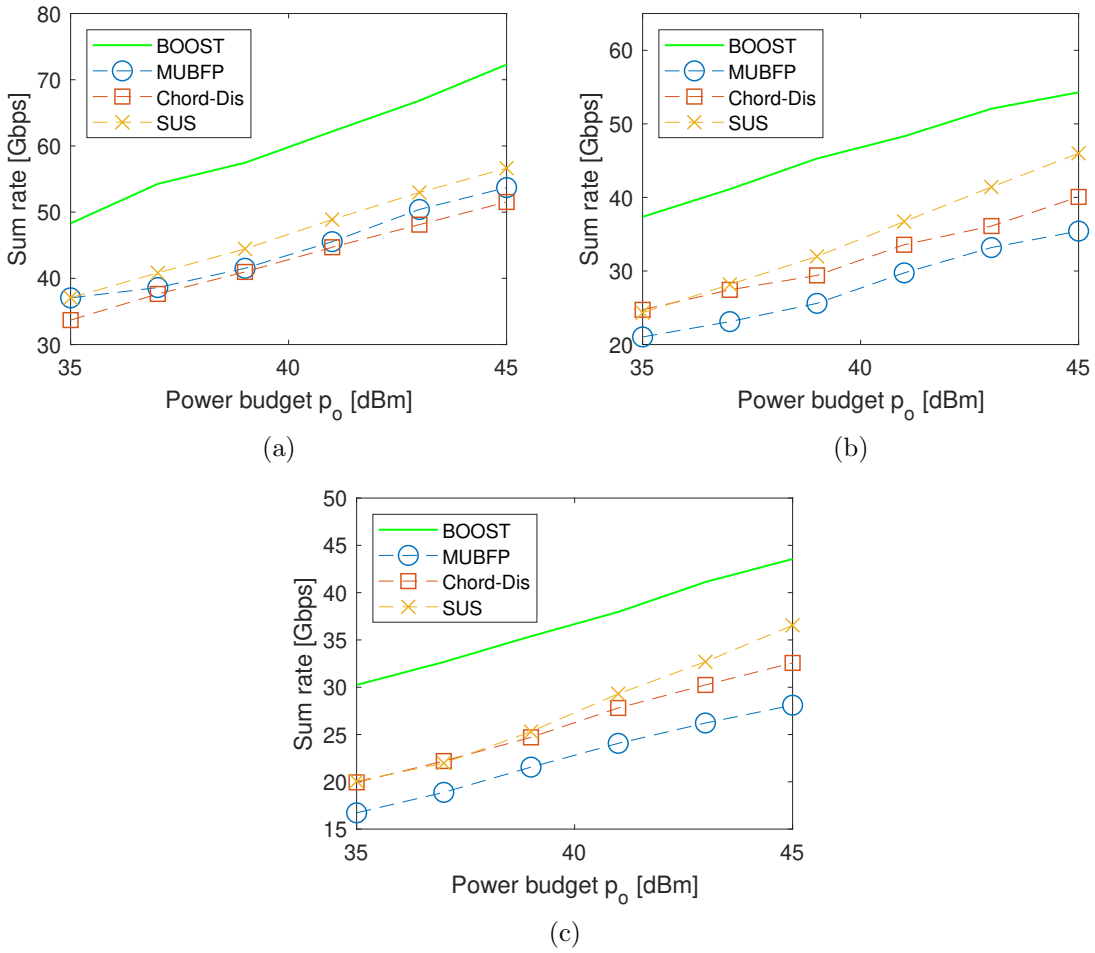



**Figure 18.** Sum rate vs. BS power budget under 3 UE distributions: (a) spread, (b) grouped, and (c) dense [6]

time but also a better confidence interval. The 95% confidence interval of the transmission time from the 50 experiments is 0.15, 0.54, 0.33, and 0.53 ms for BOOST, MUBFP, Chord-Dis, and SUS, respectively. From the 'spread' to the 'dense' UE distribution, the transmission time increases. This is because the interference between beams increases. Thus a smaller number of beams can be scheduled in each slot to maintain the minimum SINR requirement at UEs. Table 7 illustrates the percentage decrease of the average transmission time of BOOST, with respect to MUBFP, Chord-Dis, and SUS. Averaging over the three UE distributions, BOOST reduces transmission time by 37%, 30%, and 26%, respectively.

**Table 8.** Sum rate increase of BOOST

| UE Pattern | vs. MUBFP | vs. Chord-Dis | vs. SUS |
|---|---|---|---|
| Spread | 36% | 41% | 29% |
| Grouped | 67% | 46% | 36% |
| Dense | 65% | 41% | 36% |

Next we evaluate the sum rate of UEs with regard to the BS power budget, as plotted in Fig. 18. BOOST clearly outperforms other schemes on the sum rate as well. Table 8 illustrates the percentage increase of the average sum rate of BOOST compared with other schemes. On average, BOOST achieves 56%, 43%, and 34% gain, respectively. From the 'spread' to the 'dense' UE distribution, the sum rate decreases in all schemes, due to the higher interference as discussed above.

The above discussions are for the average performance among 50 experiments under each UE distribution. Next we examine how dispersed the results of individual experiments are. Figs. 19 and 20 plot the *probability density distribution* (PDF) of the network transmission time and UE SINR, respectively, obtained from 50 experiments under each UE distribution, given the BS power budget $p_o = 35$ dBm. The PDF results with different values of simulation parameters such as the power budget, minimum SINR, and slot duration exhibit similar trends and are omitted. As shown in Fig. 19, BOOST has a smaller spread on both the transmission time and SINR. This indicates that BOOST can smartly adapt to different networks to achieve good performance in all scenarios. In contrast, other schemes cannot adapt to different networks well, i.e., their performance is good for some networks, but poor for others. In particular, the spread of UE SINR for BOOST is much smaller than other schemes. This demonstrates the advantage of BOOST–it effectively guarantees a required SINR for each UE while wisely providing just enough SINR to all UEs, through smartly clustering and associating UEs to different clusters/beams. In contrast, other schemes result in highly dispersed SINR for different UEs.

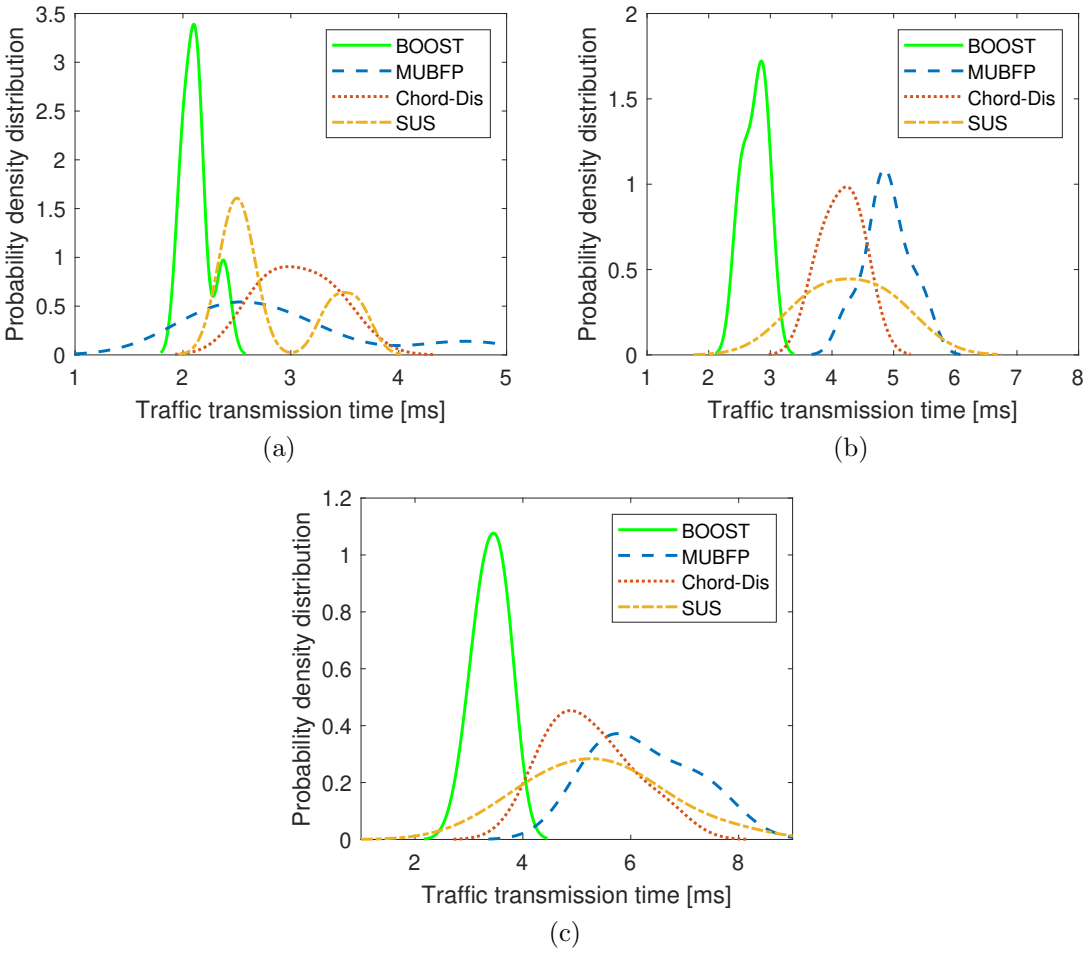



**Figure 19.** PDF of transmission time under (a) spread, (b) grouped, and (c) dense UE distribution, with $p_o = 35$ dBm [6]

Lastly, we evaluate the UE fairness on the transmission time using the Jain index, which ranges from 0 (worst fairness) to 1 (best fairness). It is at the maximum when the transmission time for all UEs is the same regardless of their traffic loads and locations in the network. Fig. 21 plots the mean and 95% confidence interval of the Jain index from all experiments. BOOST clearly outperforms other schemes, on both the Jain index value and the confidence interval. Furthermore, the fairness of BOOST is similar among all three UE distributions. Together with the much better confidence interval in all three cases, again, this indicates that BOOST can smartly adapt to different networks and achieves similar fairness. On the

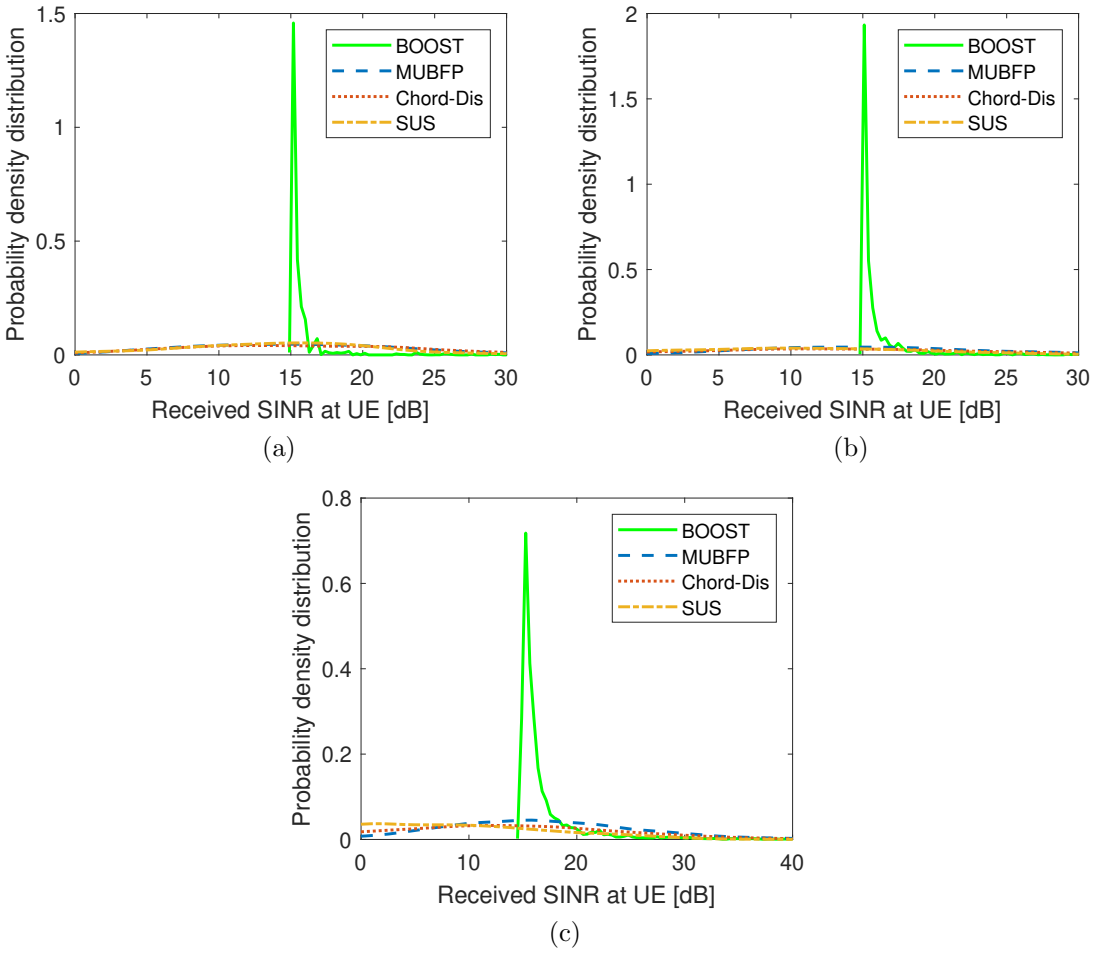



**Figure 20.** PDF of UE SINR under (a) spread, (b) grouped, and (c) dense UE distribution, with power budget $p_o = 35$ dBm [6]

other hand, other schemes have significantly worse confidence intervals as well as dispersed fairness across three UE distributions, which means the fairness has a large variance across different networks.

## 3.6 CHAPTER SUMMARY

In this chapter, we have discussed a novel *clustering Based dOwnlink UE assOciation, Scheduling, beamforming with power allocaTion* (BOOST) framework for mmWave networks, with the objective to reduce the network transmission time for given UE traffic loads, subject

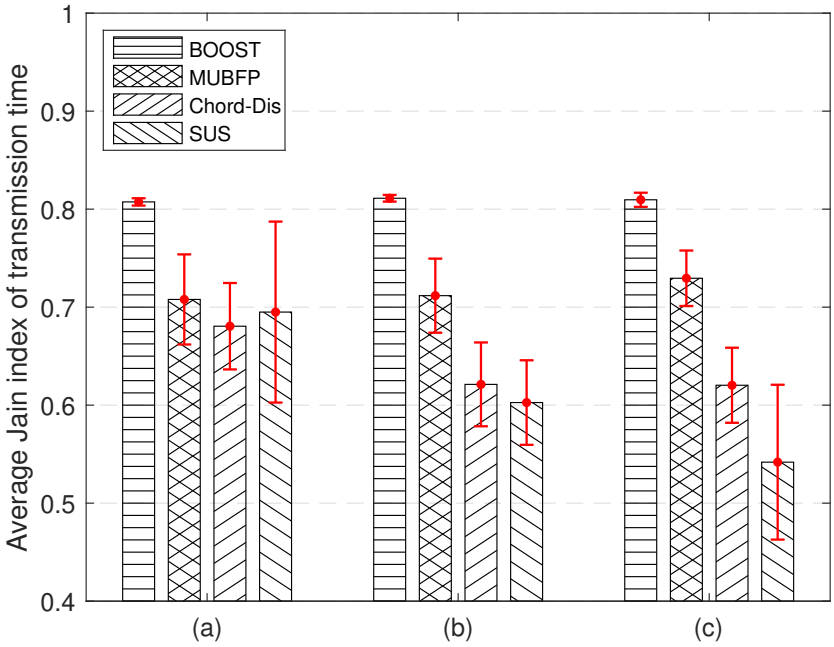



**Figure 21.** Average Jain index of transmission times with 95% confidence interval under (a) spread, (b) grouped, and (c) dense UE distribution, with power budget $p_o = 35$ dBm [6]

to the BS power budget, minimum UE SINR requirement, and number of beamformers. The BOOST framework includes three components, UE clustering, UE association, and beamforming, power allocation and UE scheduling. We have compared BOOST with three state-of-the-art schemes on the transmission time, sum rate, SINR, and UE fairness. The results indicate that overall BOOST significantly outperforms them. On average, BOOST reduces transmission time by 37%, 30%, and 26%, and achieves a sum rate gain of 56%, 43%, and 34%, respectively, compared with the other three schemes.

# CHAPTER 4

# SUBDIVIDED BAND SPECTRUM SENSING

mmWave communications and dynamic spectrum sharing have widely been recognized as two important features to realize the full benefits of envisioned 5G and beyond. Spectrum sensing is the foundation of a dynamic spectrum sharing system that is usually implemented in the cognitive radio domain. In cognitive radio, SUs constantly sense the spectrum environment and dynamically adapt the transreception parameters to access the partially or completely unused spectrum without causing harmful interference to PU.

In this chapter, we first briefly discuss some major spectrum sensing methods to detect PU signals on a single band. These spectrum sensing methods typically are difficult to use for sub-divided band spectrum sensing, which aims to identify all idle sub-bands in a sub-divided band. Hence, we further introduce sub-divided band spectrum sensing methods, which are mostly based on wavelet transform, such as the *wavelet transform multi-scale summation* (WTMS), *wavelet transform multi-scale product* (WTMP), *wavelet transform modulus maxima* (WTMM), etc. The objective of sub-divided band spectrum sensing is to find the *edges* of the signals, i.e., the start and the end frequencies of each signal, on the frequency domain. Hence, wideband spectrum sensing is also called *edge detection*. The existing wideband spectrum sensing methods often have difficulty to accurately detect both the presence of the signals and the locations of the signal edges, in particular for real world signals. As such, as the last part of this chapter, we present a new wavelet based wideband spectrum sensing algorithm, called the *exponentially moving average based multi-scale summation* (EMAMS), to effectively address this issue.

## 4.1 BACKGROUND AND RELATED WORKS

Spectrum sensing plays an important role in cognitive radio techniques which is vital in 5G to enable spectrum sharing. There are two main objectives for spectrum sensing. First, SUs perform spectrum sensing to identify the unused PU frequency bands for opportunistic spectrum access for SU communications. Second, SUs need to detect the presence of PU signals on a frequency band so that the PU transmission is protected from the interference of SU communications. For a sub-divided licensed band, such as an unlicensed WiFi band, existing works face a challenge to detect the presence of PU. To detect which sub-bands are active or signals with unpredictable start and end frequencies, usually, the received signal in the time domain is transformed into the frequency domain to get the *power spectral density* (PSD) of the signal, e.g., through the (Fast Fourier Transform) FFT. The PSD signal in the frequency domain gives information about the start and end frequencies for each signal, which are called the edges of a signal. However, the PSD of sub-band signals can be highly irregular and unpredictable in the real world. Hence, detecting the edges of such signals in real time requires a precise and efficient edge detection algorithm. The WTMP is a widely used edge detection algorithm. It utilizes the adjacent scale multiplication to detect the edges in PSD signals. When the wavelet transform scale increases, the edges of the PSD signal get more visible as the noise is reduced. However, the edge locations also shift. Therefore, WTMP often cannot accurately detect the true start and end frequency of a signal.

Here we devise an *exponential moving average based multi-scale sum* (EMAMS) spectrum sensing approach to decompose the PSD signal and extract the edges of signals, by assigning exponentially changing weights to the different scales of the wavelet transform. Compared with WTMP, the EMAMS approach effectively avoids the edge shifting issue, and successfully detects the active sub-bands in a licensed band so that SUs can utilize the unused sub-bands. This would dramatically reduce the band switching of SUs, since SUs do not have to evacuate from a licensed band when there is a PU signal on the band. Only if all sub-bands are used

by PUs, SUs need to switch to a different band.

In recent literature, several spectrum sensing works based on wavelet transform can be found. In [72], the authors discussed the WTMS based signal edge detection method for wide band or sub-divided band. It detects the sharp variation points, called *singularities*, of synthetic cosine shaped PSD signal in the entire wide band to classify the sub-bands as occupied or vacant. The authors in [73] studied the DWPT based energy detection algorithm under uncertain noise. The algorithm estimates the received signal noise level by performing DWPT and then uses energy detection over the wide band to detect the sub-band signal. They claimed that DWPT is better than wavelet transform (WT) for estimating the uncertain noise power, because it can decompose both low and high frequency components simultaneously in every scale. In [74], the authors studied a wavelet transform based wideband spectrum sensing method using the Hilbert transform. They examined the rectangular shape PSD signal with additive noise for edge detection. In [75], the authors discussed an edge detection based spectrum sensing technique by using the phase-field segmentation. They claimed that their sensing performance results can be improved by properly setting the phase-field segmentation parameters.

The existing works detect the edges of synthetic PSD signals with regular shapes, such as cosine or rectangle, by using the wavelet transform. However, in practice, the real world signals have a highly irregular shape in the frequency domain. Moreover, real world noise is also unpredictable. Hence, it requires a robust and precise edge detection scheme to accurately detect edges of the signals in the frequency domain. Our objective is to detect the sub-channel signals within a licensed band using edge detection. As such, we develop the exponential moving average based multiscale sum approach.

## 4.2 PROBLEM STATEMENT AND SYSTEM MODEL

Fig.22 illustrates the PSD power of a received signal along the frequency domain at a specific time. It can be seen that there are three signals on this band. Note that there may

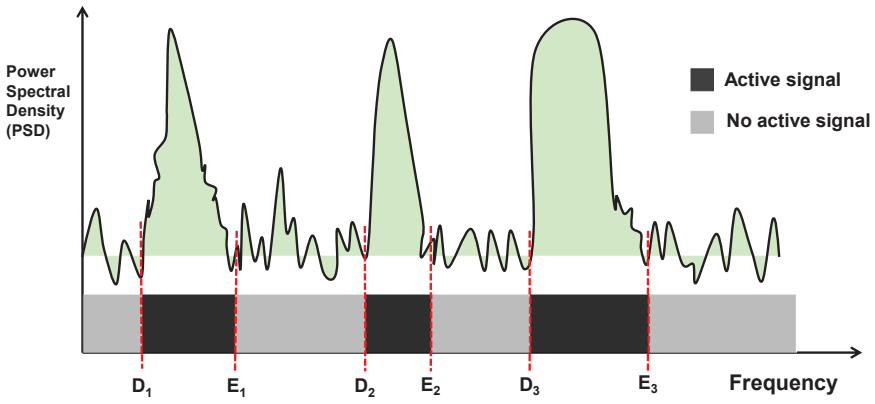



**Figure 22.** Illustration of three active PSD signals in a licensed band: Three active signals present in the licensed band at any specific time. Edges are the start and end frequencies of each active signal represented by $D$ and $E$, respectively [7].

be multiple signals sitting on several consecutive sub-bands, so that those signals may appear like one single signal. Each signal has a start frequency and an end frequency, which are called *edges* of the signal. The *objective of edge detection* of spectrum sensing is to detect the start and the end frequency of each signal. The PSD shape of a real world signal is significantly different from a regular cosine or rectangular form that has been assumed in most existing studies of edge detection. Therefore, precise localization of the edges of a signal is a great challenge in designing an edge detection scheme. The DWT signal decomposition has a good localization property at the low scale, but performance degrades when the scale increases. This is because the edges of a signal are dislocated in the frequency axis at higher DWT scales. On the other hand, there may be many false edges due to random noise at lower DWT scales. In this dissertation, we present a reliable and robust signal edge detection algorithm which is able to preserve start and end frequency edges structures of the PSD signal even at higher DWT scales; hence, performance of signal detection improves with good localization accuracy.

The DWT will be used in our work for multi-scale time-frequency analysis to extract the start and end frequencies of the PSD signal in a specific band. The wavelet transform

can effectively separate the *detail* and the *approximation* of a signal. The detail is the high frequency component of a signal, while the approximation is the low frequency component. The edge features of the PSD signal are hidden in detail or high frequency component parts of the DWT decomposed PSD signal. As mentioned earlier, PSD signal edges are not clearly distinguishable from noise components at lower DWT scale detail parts. In DWT successive scale decomposition, noise dilutes, but at the same time, the high frequency component, i.e., edges of the PSD signal becomes imprecise. In the literature, there are some edge detection techniques including WTMP, WTMS, wavelet transform multi-scale modulus maxima (WTMM), etc. The performance of signal detection in real-world PSD signals with these existing techniques is not satisfactory. More specifically, these edge detectors are prone to miss the narrow width PSD signal and/or dislocate the corresponding edges frequency at higher DWT scale decomposition.

Directly summing up or multiplying the transformed signals at adjacent scales, as adopted in most existing works, results in unwanted noise in the edge features. Moreover, the edges are shifted in the frequency domain, which leads to dislocated edges of a signal. The EMAMS approach can effectively address this issue and detect the edges of signals. The basic idea is that the recent DWT scale results obtain strict priority over the previous scale; hence, noise is suppressed with a low weight at a low scale. The edges are detected in two stages. In the first stage, the wavelet transform results are compared with a power threshold to detect signals. In the second stage, some spurious edges are discarded by a bandwidth threshold.

## 4.3 EXPONENTIALLY AVERAGED MULTI-SCALE SUM

In this section, we first describe the main idea of wavelet transform. Then we describe the EMAMS approach. Lastly, we discuss how to use an adaptive threshold to achieve excellent performance in edge detection.

### 4.3.1 WAVELET TRANSFORM

A function $\Psi(x)$ is called a wavelet if its average is zero and has the orthogonal property as shown in (1) and (2).

$$\int_{-\infty}^{\infty} \Psi(x)dx = 0, \tag{44}$$

$$\langle \Psi_l(x), \Psi_m(x) \rangle = \delta_{l,m} = \begin{cases} 1, & \text{if } l = m \\ 0, & \text{if } l \neq m \end{cases} \tag{45}$$

where $l, m$ are integers which represent two translated versions $\Psi_l(x)$ and $\Psi_m(x)$ of wavelet function $\Psi(x)$, and $\delta_{l,m}$ is the Kronecker delta. For signal edge detection, it is highly desirable that the wavelet basis function has the singularity property. We use the Daubechies wavelet basis function with the 4th vanishing moment which has a good singularity property. Vanishing moment is a criterion about how a function decays toward infinity. The wavelet function $\Psi(x)$ has $v$ vanishing moments if

$$\int_{-\infty}^{\infty} x^h \Psi(x)dx = 0 \quad \text{for} \quad 0 \leq h \leq v. \tag{46}$$

Without loss of generality, we consider signal edge detection in one licensed band that is subdivided into $N_c$ sub-bands by PUs. SUs do not need to have the subdivision knowledge of the licensed band. Specifically, SUs do not need to know $N_c$. Let $F[n]$ $(1 \leq n \leq M)$ denote the PSD of the combination of all signals on the licensed band at a specific time. Here, we consider $F$ as a discrete function, for efficient computation. That is, the PSD of the licensed band is sampled with a frequency resolution $\nabla$, with total $M$ number of samples. $F[1]$ is the PSD at the first sample, and $F[M]$ is the PSD of the last sample. A continuous PSD function over the licensed band can also be used, but it usually consumes significantly longer computation time.

For a given function $f(t)$, the wavelet transform for high frequency decomposition of this function means to apply a certain wavelet function to this function through a convolution, denoted as $W_s f(t)$, where $s$ denotes the scale of the wavelet transform. Specifically, $W_s f(t)$ indicates the strength of a wavelet scaled by $s$ at time $t$. The wavelet extends for only a short period, so its effect is limited to the area immediately surrounding $t$. The wavelet transform gives information for the strength of a signal in the frequency domain at a given time [76]. However, the wavelet function is achieved from the scaling function which is used for approximation decomposition of a signal.

The scaling and wavelet functions, denoted as $\Phi(t)$ and $\Psi(t)$, are given as [77]

$$\Phi(t) = \sum_{i=-\infty}^{\infty} h_\phi[i]\sqrt{2}\Phi(2t-i), \tag{47}$$

$$\Psi(t) = \sum_{i=-\infty}^{\infty} h_\psi[i]\sqrt{2}\Phi(2t-i). \tag{48}$$

The $h_\phi[i]$ and $h_\psi[i]$ are scaling and wavelet vectors, respectively. For instance, the discrete time-domain representation of the scaling vector or low pass filter for the Daubechies wavelet with 2nd vanishing moment is written as

$$h_\phi[i] = \frac{\sqrt{2}+\sqrt{6}}{8}\delta[i] + \frac{3\sqrt{2}+\sqrt{6}}{8}\delta[i-1] + \frac{3\sqrt{2}-\sqrt{6}}{8}\delta[i-2] + \frac{\sqrt{2}-\sqrt{6}}{8}\delta[i-3]. \tag{49}$$

Note that the minimum filter size is two times the vanishing moment. The corresponding wavelet vector or high frequency filter $h_\psi[i]$ is just a mirror filter of $h_\phi[i]$. The high frequency filter $h_\psi[i]$ is written as

$$h_\psi[i] = (-1)^i h_\phi[1-i].$$

Let $\Psi_j[n]$ denote the dyadic dilation of Daubechies wavelet basis function $\Psi[n]$ in the

discrete frequency domain. Then $\Psi_j[n]$ is given as

$$\Psi_j[n] = 2^{-2j}\Psi[2^{-j}n] \tag{50}$$

The wavelet basis function $\Psi[n]$ gets stretched with a factor of $2^j$ in every successive DWT signal decomposition. Since we are only interested in the detail part of the PSD signal in the decomposition, the detail coefficient of the DWT decomposition for $F[n]$ at dyadic scale $2^j$ is defined as:

$$W_jF[n] = F * \Psi_j[n], \tag{51}$$

where $*$ denotes the convolution.

### 4.3.2 EDGE DETECTION BY EXPONENTIAL MOVING AVERAGE BASED MULTI-SCALE SUMMATION (EMAMS)

Most existing work on signal edge detection uses the multi-scale product which multiplies the wavelet transform results at two adjacent scales, $W_jF$, and $W_{j+1}F$. Specifically, the multi-scale product $P_j[n]$ is computed as follows.

$$P_j[n] = W_jF[n]W_{j+1}F[n] \tag{52}$$

Then $P_j[n]$ is used to detect the edges of signals (see [76] for more details). The multi-scale product technique exploits the adjacent DWT scale correlation. However, this approach has difficulty detecting a signal if its bandwidth is relatively narrow. For example, in our experiments, we found that if the bandwidth of the signal is less than 5 MHz, then this approach has difficulty detecting the signal, i.e., it cannot find the edges of the signal.

An alternative way of extracting the edge information is to sum up the DWT scale details. Nevertheless, this approach also has a problem as it accumulates noise from the low scale to the high scale, i.e., the summation of the DWT results at multiple scales contains a high level

of noise mixed with the signal edge information. To address this issue, we need to develop an approach to preserve the signal edge information while suppressing the noise from low DWT scales.

Our developed approach utilizes an exponential moving average of DWT scale details. As discussed earlier, on the detail of the higher scale, the noise is low; hence, it is easier to detect the signal edges. Nevertheless, the location of the edges are also shifted in the frequency domain. In other words, with the detail at the higher scale, we can detect the existence of signals, but cannot find the accurate start and end frequencies of the signals, as they are shifted. On the other hand, on the detail of the lower scale, the noise is high; hence, it is often difficult to detect the existence of the signal. Nevertheless, once the existence of a signal is detected, the edges of the signal can be often precisely detected, as the locations of the signal edges are not shifted on the detail of the lower scale. Therefore, we devise the exponential moving average based multi-scale summation approach to intelligently combine the details at both lower and higher scales so that both the signal edge information is preserved while the noise is low. The exponential moving average based multi-scale summation for the wavelet transform of function $F[n]$ is written as:

$$S_{j+1}[n] = (1 - a)S_j[n] + aW_{j+1}F[n], \tag{53}$$

where $a$ is a weighting factor. Eq. (53) represents the EMAMS results after $j + 1$ scales. Specifically, the EMAMS results are as follows.

$$
\begin{aligned}
S_0[n] &= 0 \\
S_1[n] &= aW_1F[n] \\
S_2[n] &= a(1-a)W_1F[n] + aW_2F[n] \\
S_3[n] &= a(1-a)^2W_1F[n] + a(1-a)W_2F[n] + aW_3F[n] \\
&\vdots \\
S_j[n] &= a(1-a)^{j-1}W_1F[n] + a(1-a)^{j-2}W_2F[n] + \cdots + aW_jF[n]
\end{aligned}
$$

Here, $W_1F[n], W_2F[n], \cdots, W_jF[n]$ are high frequency components of the PSD signal decomposed by DWT. If the weighting factor $a$ is large, the higher scale detail weighs more than the lower scale detail. This exponential moving average based multi-scale summation can mitigate the noise effect at lower scales while it assigns a larger weight to the higher scale details. To further remove noise and the false signal edges, we design adaptive thresholds for our approach to detect signal edges and be adaptable to the noise variance and bandwidth.

### 4.3.3 ADAPTIVE THRESHOLDS

The EMAMS approach uses two thresholds: the power threshold $T_P$, and the bandwidth threshold $T_B$. The power threshold is used to detect if there is a signal on a frequency. The bandwidth threshold is used to detect the edges of a signal after observing the existence of a signal. We assume that the noise is additive white Gaussian noise (AWGN) with zero mean and variance $\sigma^2$. Let $N_0$ denote the single sided power spectral density of AWGN and $B$ denote the total bandwidth of the interested frequency band in the unit of Hz. The noise variance is then given as $\sigma^2 = \frac{N_0 B}{2}$. The power threshold for the EMAMS detector is calculated as $T_P = \frac{k\sqrt{(\sigma^2)}}{max(S_{j+1}[n])}$ where $k$ is a constant with a typical value 20.

In general, there is a tradeoff to setting the value for $T_P$. If $T_P$ is too high, some signals

**Algorithm 6:** EMAMS Edge Detection [7]

**Require:** $F[n]$ $(1 \leq n \leq M)$, $j$, $T_P$, $T_B$
**Ensure:** $K, \alpha(k), \beta(k)$ $(1 \leq k \leq K)$

```
K = 0
Compute S_j from F[n] using Eq. (53)
for n = 1 to M do
    if S_j[n] ≥ T_P then
        E[n] ← 1
    else
        E[n] ← 0
    end if
end for
{Add two auxiliary values to facilitate edge detection}
Let G[0] = 0 and G[M + 1] = 0
n = 1
while n ≤ M do
    if G[n] = 1 and G[n − 1] = 0 then
        {Find a segment of consecutive 1's in G }
        Find m such that n ≤ m ≤ M, G[n], G[n + 1], ..., G[m] are all equal to 1, and
        G[m + 1] = 0
        if m − n + 1 ≥ T_B then
            K = K + 1
            α(K) = n
            β(K) = m
            n = m + 1 {continue to scan from G[m + 1]}
        end if
    end if
    n = n + 1
end while
```

may be missed. However, with a lower $T_P$, some noise impulses may appear as signals. As a matter of fact, we have conducted extensive analysis for spectrum measurement data and found that many noise impulses have even higher PSD magnitude than signals. Thus, those

noise impulses will be treated as signals. Fortunately, from our observation of the spectrum measurement data, the bandwidth of the noise impulses is very narrow, typically in the range of a few kHz. This is much lower than the bandwidth of a signal. This observation motivates us to introduce another threshold, the bandwidth threshold, to distinguish between signals and noise impulses. Let $\lambda_B$ denote the minimum bandwidth for any signal. We compute the bandwidth threshold $T_B$ as follows.

$$T_B = \left\lceil \frac{\lambda_B}{\nabla} \right\rceil \tag{54}$$

where $\nabla$ is the frequency resolution of the discrete PSD signal $F[n]$. For example, if $F[n]$ is the values sampled at every 30 Hz on the continuous PSD of a signal, then $\nabla = 30$.

### 4.3.4 EMAMS ALGORITHM

Algorithm 6 describes how to detect the start and end frequencies of signals using the EMAMS method. The input is the PSD function $F[n]$, the number of scales $j$, the power and bandwidth thresholds $T_P$ and $T_B$. The output is the number of signals $K$, and the start and end frequency indices $\alpha(k)$ and $\beta(k)$ for each signal $k$. At first, it computes the multiscale summation by the exponentially moving average, $S_j$. Then the algorithm uses the power threshold $T_P$ to detect the existence of signals at each frequency, which is recorded in vector $G[n]$. Next, the bandwidth threshold is utilized to filter out the spurious noise impulses, and the start and end frequencies of real signals are recorded in $\alpha(k)$ and $\beta(k)$.

EMAMS is a reliable and robust signal edge detection algorithm which is able to preserve the start and end frequency edge structure of the PSD function even at higher DWT scales. Hence, the performance of signal detection improves. Compared with WTMP, EMAMS effectively avoids the edge shifting issue, and more accurately detects the active sub-bands in a licensed band.

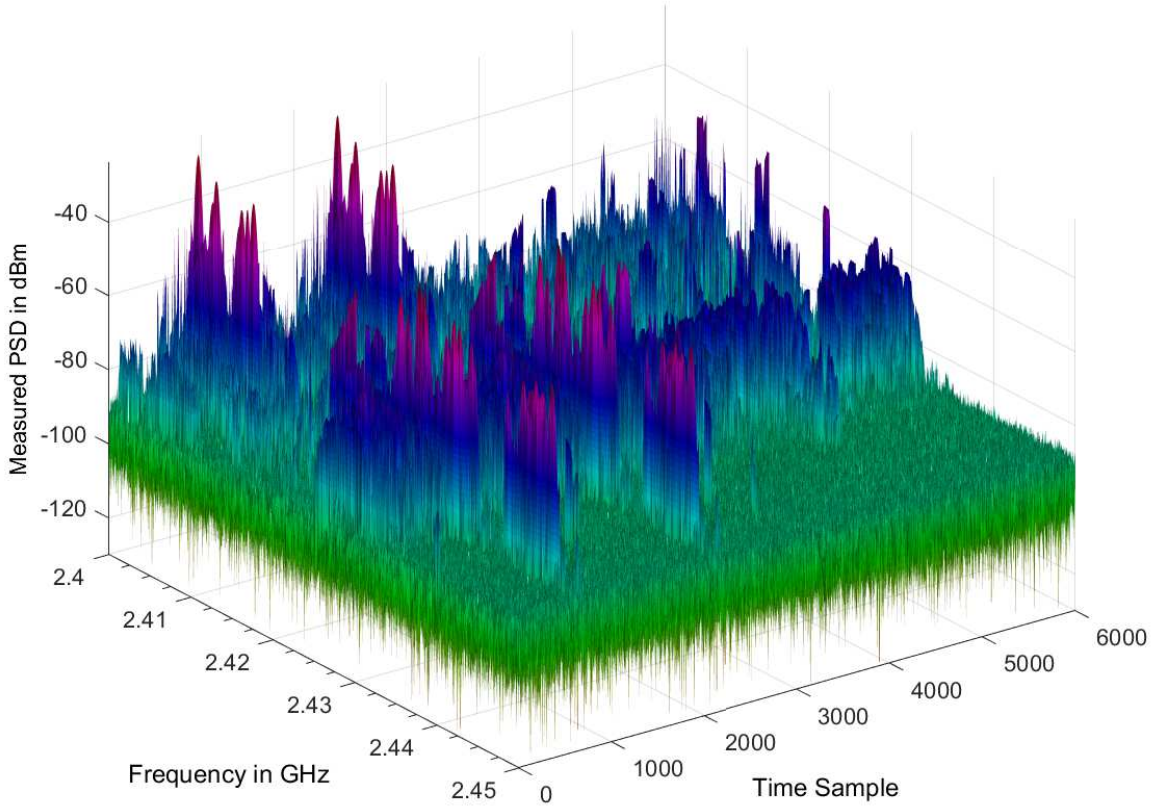



**Figure 23.** The PSD signal measured at ODU Kaufman Building in the 2.4 GHz ISM band [7]

## 4.4 PERFORMANCE EVALUATION

In this section, we evaluate the performance of the EMAMS signal detection approach. We compare the performance of EMAMS with the performance of the widely used WTMP edge detection (e.g., see [76]). Two performance metrics, false alarm and true detection, will be used for the performance evaluation.

In the first experiment, we measure the PSD signal of 2.4 GHz *industrial, scientific, and medical radio* (ISM) band (2.4 GHz to 2.5 GHz) using a spectrum analyzer in the Kaufman Engineering building at Old Dominion University (ODU). A total of 6000 time samples

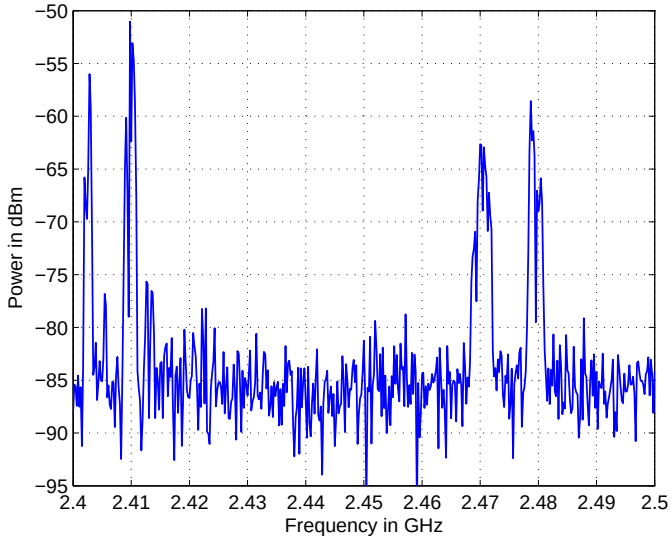



**Figure 24.** One sample of the measured PSD signal at a specific time [7]

are measured with a 1 second interval. Fig. 23 shows the measured PSD signal in three dimensions, frequency, time, and PSD, which indicates the spectrum usage in the WiFi band that varies in both time and frequency domains. The received signal attenuation is 10 dB. The PSD signal reference is $-50$ dBm. The total measured bandwidth $B$ is 100 MHz, from 2.4 GHz to 2.5 GHz.

Fig. 24 illustrates one sample from the measured PSD signal at a specific time. It can be seen that there are four signals and those signals do not have the ideal sinusoidal or rectangular shapes. In fact, signals usually have large fluctuations in the PSD magnitude. There are also noise impulses that have high a PSD magnitude but narrow bandwidth.

Next we decompose the PSD sample in Fig. 24 to get the details at multiple scales using Eq. (51), as illustrated in Fig. 25. We have shown the details of the first 4 scales from the decomposition. In the first scale decomposition, the features of signals are not clear since there is a high level of noise. When the scale increases, the details of the signal become apparent and the noise power remarkably attenuates. However, it is difficult to accurately detect the edges of PSD signals from the higher detail decomposition since the edges shift

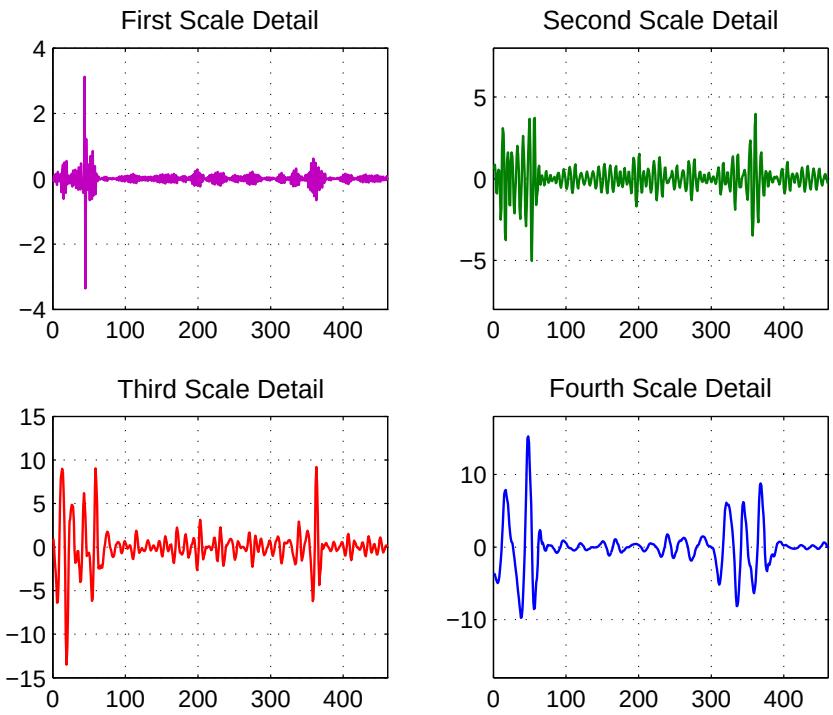
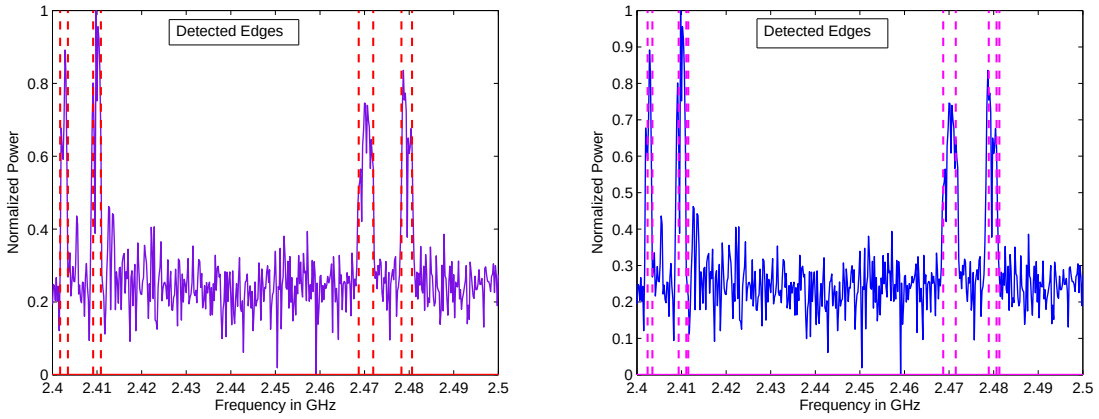



**Figure 25.** The detail at multiple scales resulted from the decomposition of the PSD sample in Fig. 24 using Eq. 51 [7]

in the frequency axis. In fact, in the 4th scale detail, it seems that there are 5 signals.

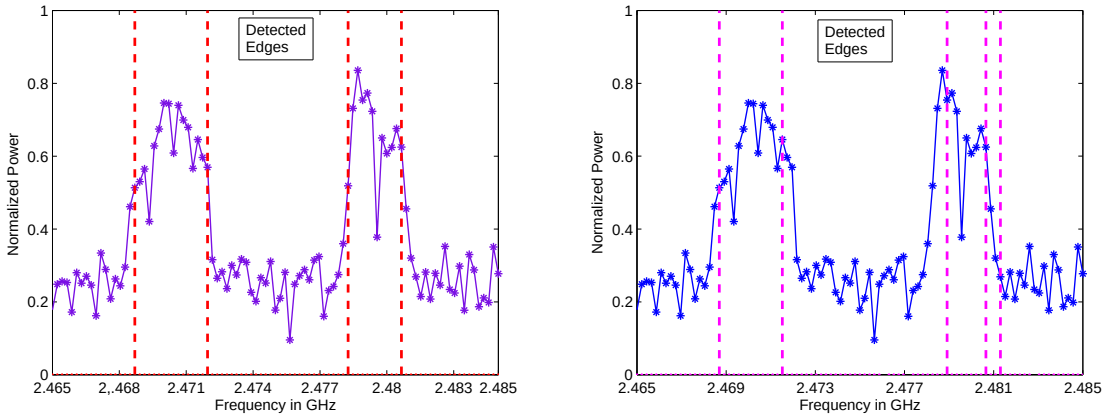
Now we evaluate the EMAMS edge detection and compare it with the widely used WTMP edge detection. Fig. 26 depicts the signal edge detection results using 5 scales, by the EMAMS and WTMP approaches, respectively.

The single sided noise PSD $N_0$ is $-90$ dBm. The threshold constant $k$ is 20, and the weight constant $a$ is 0.5. The measured PSD signal frequency resolution $\nabla$ is 217 kHz. The $\lambda_B$ is set as 868 kHz. There are four signals in this sample PSD signal. Both EMAMS and WTMP have detected these four signals and the edges of signals are shown by vertical solid lines. However, WTMP often cannot give accurate locations for the signal edges. Fig. 27 is the zoomed-in version of Fig. 26 in the 2.465 GHz to 2.485 GHz frequency range. There are 2 signals in this frequency range. However, the edges detected by WTMP have a frequency




(a) (b)

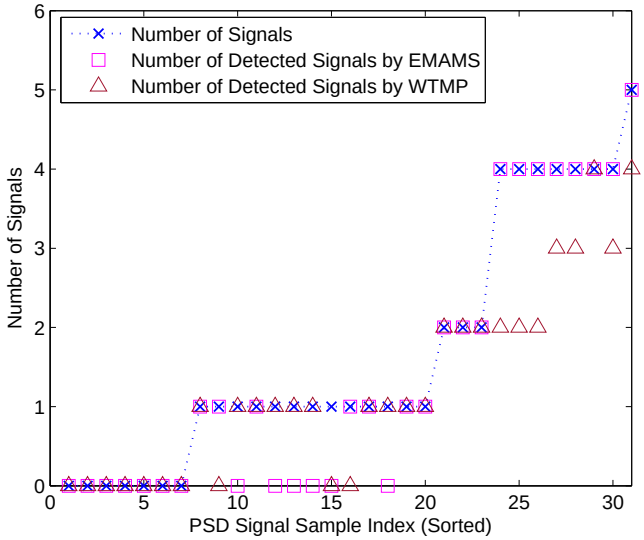
**Figure 26.** Signal edge detection for the PSD sample in Fig. 24 by EMAMS and WTMP: (a) signal edges detected by EMAMS after 5th scale DWT decomposition; (b) signal edges detected by WTMP after 5th scale decomposition [7].



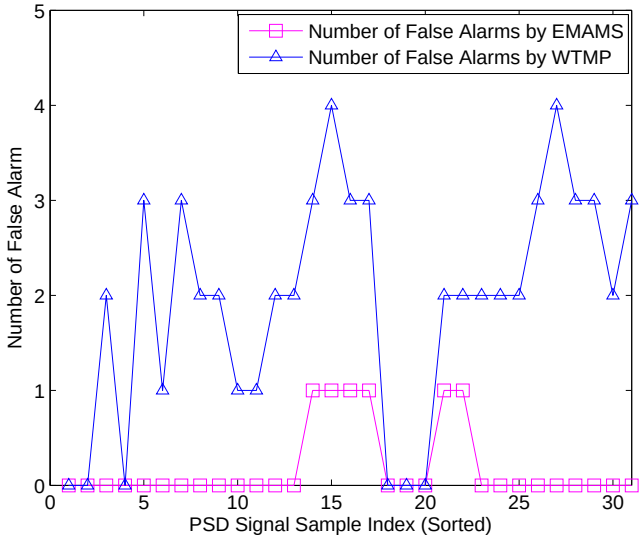

(a) (b)

**Figure 27.** Zoomed-in version of Fig. 26 in the 2.465 GHz to 2.485 GHz frequency range: (a) signal edges detected by EMAMS, and (b) signal edges detected by WTMP [7].

localization problem, i.e., WTMP fails to extract the correct frequency location of signal. Specifically, the right edge detected by WTMP for the first signal in Fig. 27(b) is not at the correct location. It actually sits on the left of the true right edge. In addition, WTMP found three edges for the second signal which makes it difficult to determine the correct left




(a)

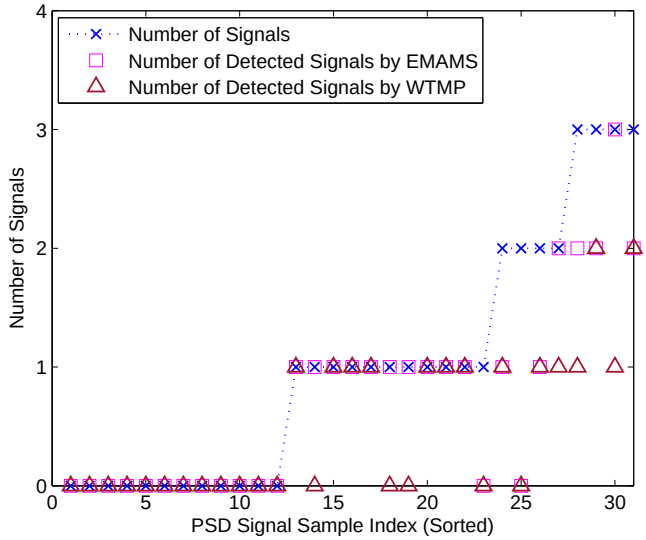



(b)

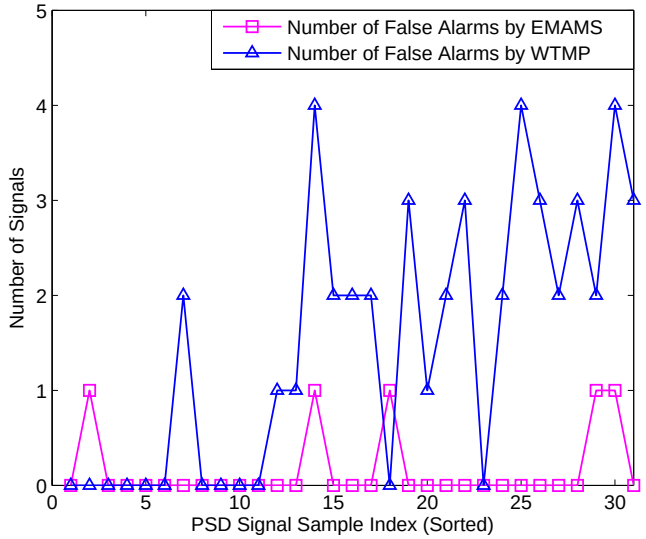
**Figure 28.** (a) Number of signals in 31 PSD signal samples versus the number of detected signals by EMAMS and WTMP, (b) False detection for the PSD samples in (a) [7]

and right edges, as any of them could be a false edge.

Next we examine the number of true detections and false alarms. The number of true detections denotes the number of signals that has not been correctly detected, while the number of false alarms denotes the number of 'false' signals that has been detected, but are




(a)




(b)

**Figure 29.** (a) Number of signals in the second data set, versus the number of detected signals, (b) False detection for the PSD samples in (a) [7]

actually not true signals. Fig. 28(a) illustrates the number of signals of 31 PSD signal samples, and the corresponding number of detected signals by EMAMS and WTMP, respectively. To present the results better, we have sorted the sample indices in ascending order of the number of occupied signals in each PSD signal sample. We can see that the EMAMS edge

detection approach outperforms the WTMP approach in existing works. EMAMS misses 6 signals while WTMP misses 13 signals.

Fig. 28(b) plots the number of false alarms for the same PSD signal samples in Fig. 28(a). The figure shows that with regard to the false alarm, EMAMS significantly outperforms WTMP. There are 60 false alarms with WTMP, while there are only 6 false alarms with EMAMS. Figs. 29(a) and 29(b) plot the number of true detections and false alarms for a different set of 31 samples measured at a different date. For this PSD sample set, EMAMS misses 8 signals while WTMP misses 14 signals. With regard to false alarms, there are 46 false alarms with WTMP, while EMAMS has only 5 false alarms. In conclusion, the EMAMS edge detection approach significantly outperforms the WTMP approach.

## 4.5 CHAPTER SUMMARY

In this chapter, we first outline the spectrum sensing problem on a wide band or a sub-divided licensed band to identify all idle sub-bands, which offers greater potential for SUs to exploit the spectrum. Then we discuss the major techniques for wideband spectrum sensing, the wavelet transform based spectrum sensing methods. The wideband spectrum sensing is reduced to detecting edges of signals on the frequency domains, i.e., finding the start and end frequencies of each signal. Hence, the wavelet transform based wideband spectrum sensing is also called edge detection. Lastly, we present a new edge detection algorithm based on exponentially moving averaged multi-scale summation, called EMAMS. The designed algorithm intelligently summates the details of multiple scales after the wavelet transform of the power spectral density measure. This preserves the edge information of signals while suppressing the noise so that both the existence of the signals and corresponding edges can be accurately detected. The simulation results indicate that EMAMS outperforms the existing wavelet transform based edge detection algorithms.

# CHAPTER 5

# CONCLUSIONS AND FUTURE WORK

In previous chapters, we have studied beamforming based topology control and scheduling in mmWave networks and spectrum sensing for dynamic spectrum sharing in 5G. In this chapter, we will present a summary of the contributions of the dissertation. We will also discuss future research directions and ideas.

## 5.1 CONCLUSIONS

The envisioned 5G network will address the phenomenal traffic demand from emerging intelligent wireless devices and a large number of internet applications. The key objective in 5G is to utilize the massive spectrum in mmWave frequency to drastically increase system capacity and UE data rate. The path loss in the mmWave is significant which can be fully compensated for by using an antenna array beamforming technique. However, topology control, UE association and scheduling are challenging with beamforming. This dissertation developed several novel techniques and algorithms to solve these challenges. We also devised a novel spectrum sensing technique for subdivided licensed band which can be used for spectrum sharing in 5G. The main contributions discussed in the dissertation are listed below.

- A novel topology control framework has been designed for multi-BSs mmWave networks, called *Beamforming Oriented tOpology coNtrol* (BOON). The objective of BOON is to reduce the total transmit power of all BSs while forming beams to cover UEs with a minimum SINR. BOON includes four major components: UE clustering, set construction, set covering, and beamforming. Here, we have developed a beamforming oriented UE clustering algorithm to group UEs, with the objective to reduce transmit

power and interference between beams. We have devised a set construction scheme and an interference aware set covering algorithm that addresses the unique challenges in beamforming to smartly associate UEs to beams and BSs, with the objective to reduce interference between beams. We have designed an approach to form beams for all BSs with the objective to reduce the total transmit power and inter-BS interference, subject to coverage of all UEs and meeting a minimum quality of service. At last, we have compared BOON with the multi-user and multicast beamforming based topology control schemes, and a state-of-the-art scheme JSDM on transmit power, sum rate, SINR, and computation complexity. The results indicate that overall BOON significantly outperforms them. For instance, BOON uses only 10%, 32% and 25% transmit power on average, of the other three schemes, respectively, to achieve the same sum rate in the network.

- A novel framework, called *clustering Based dOwnlink UE assOciation, Scheduling, beamforming with power allocaTion* (BOOST), has been discussed for UE association and scheduling. The main objective of BOOST is to reduce the downlink network transmission time, subject to the BS power budget, number of beamformers per BS, user traffic loads, and the quality of service requirement at UEs. The BOOST framework has the following three unique schemes: (1) a UE clustering scheme that capitalizes on unique features of beamforming to group UEs into clusters to reduce interference for beamforming, (2) a novel UE association scheme that effectively reduces interference and balances UE traffic loads between BSs to decrease the transmission time, and (3) a scheme for joint beamforming, power allocation, and UE scheduling to reduce transmission time, subject to the BS power budget, UEs traffic loads, and the minimum SINR requirement at UEs. We compare BOOST with three state-of-the-art user scheduling schemes. On average, BOOST reduces the transmission time by 37%, 30%, and 26%, and achieves a sum rate gain of 56%, 43%, and 34%, respectively.

- For sub-divided band spectrum sensing, a novel technique has been developed through wavelet transform, called *exponentially moving average based multi-scale summation* (EMAMS). EMAMS can be used in 5G as a potential technique to detect unused subbands in the PU spectrum. EMAMS does not depend on the number of used sub-bands and adaptively detects active signals in a spectrum band measured by a spectrum analyzer. Here, we have developed an exponentially averaged multi-scale sum algorithm to effectively detect the active sub-bands in a licensed band so that SUs can utilize the residual spectrum in the licensed band to significantly increase spectrum utilization. Further, instead of using synthetic PU signals such as raised cosine, rectangular pulse, sinusoidal, we have used the measured real world signals by a spectrum analyzer for performance evaluation. The simulation results indicate that EMAMS outperforms the existing wavelet transform based edge detection algorithms.

## 5.2 FUTURE WORKS

In this section, future research directions will be discussed. We have identified several research areas that can be explored in the future to improve the results presented in this dissertation. The following lists some possible directions for future directions.

- In the BOON and BOOST framework discussed in Chapter 2 and Chapter 3, we assumed line of sight (LOS) signal propagation and formulated the channel between multi-antenna BS and single-antenna UE. One of the research directions is to explore the possibility of incorporating (non-line of sight) NLOS paths and consider the multi-antenna system at both BS and UE. In this case, a UE can also form beam(s) to increase received signal gain and cancel interference from other beams by directing nulls toward them. Moreover, it is possible to extend the study on how to efficiently adjust the UE clustering, set covering, and beamforming in BOON to avoid re-running the entire process for a small number of UEs moving out of the original beams. Such

improvements will strengthen BOON to efficiently address the UE mobility.

- In the BOOST framework presented in Chapter 3, it is assumed that each UE has fixed downlink traffic in the queue of BS. Another possible research direction could be the case when UE traffic is a continued arrival process, instead of fixed size traffic. In this case, we need to study how to efficiently adjust the UE association, beamforming, and scheduling when UE's new traffic needs to be considered for scheduling if it finishes in the current slot. It is also possible to extend the study considering the case when each UE forms multiple beams to receive coordinated spatial streams from different BSs simultaneously, similar to the carrier aggregation technique in LTE-advance. Such improvements will strengthen BOOST.

- Both BOON and BOOST frameworks are designed for low UE mobility use case scenarios where channel characteristics remain unchanged for several seconds (coherence time). It is still an open challenge to design topology control and UE scheduling techniques for high UE mobility mmWave networks. In this case, coherence time will be very small; thus, we need to re-run the algorithms frequently which results in a large system overload. A robust and fast beam tracking technique will be required to adjust the UE association and scheduling. Such improvements will strengthen our frameworks and improve the results presented in the dissertation.

# BIBLIOGRAPHY

[1] P. Paul, H. Wu, C. Xin, and M. Song, "Beamforming oriented topology control for mmwave networks," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2019.

[2] Eiman Mohyeldin, ITU-R Workshop on IMT-2020 terrestrial radio interfaces, "Minimum technical performance requirements for imt-2020 radio interface(s)," 2016. [Online]. Available: https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Documents/S01-1_Requirements%20for%20IMT-2020_Rev.pdf

[3] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications mmWave for 5G: opportunities and challenges," *Wireless networks*, vol. 21, no. 8, pp. 2657–2676, 2015.

[4] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, and G. Chen, "Millimeter wave communication: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1616–1653, 2018.

[5] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, 2014.

[6] P. Paul, H. Wu, and C. Xin, "BOOST: A user association and scheduling framework for beamforming mmwave networks," *IEEE Transactions on Mobile Computing*, 2020 (Accepted).

[7] P. Paul, C. Xin, M. Song, and Y. Zhao, "Spectrum sensing for a subdivided band in cognitive radio networks," in *Proc. in IEEE International Conference on Computer Communication and Networks (ICCCN)*, 2015.

[8] Ericsson Report, "Internet of things forecast," 2018. [Online]. Available: https://www.ericsson.com/en/mobility-report/internet-of-things-forecast

[9] IHS Markit, "The 5g economy," 2019. [Online]. Available: https://www.qualcomm.com/media/documents/files/ihs-5g-economic-impact-study-2019.pdf

[10] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[11] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. of the IEEE*, vol. 102, no. 3, pp. 366–385, 2014.

[12] C. R. Anderson and T. S. Rappaport, "In-building wideband partition loss measurements at 2.5 and 60 GHz," *IEEE Trans. Wireless Commun.*, vol. 3, no. 3, pp. 922–928, 2004.

[13] F. Khan and Z. Pi, "mmWave mobile broadband (MMB): Unleashing the 3–300 GHz spectrum," in *Proc. IEEE Sarnoff Symposium*, 2011.

[14] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, 2011.

[15] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[16] Z. He, S. Mao, S. Kompella, and A. Swami, "On link scheduling in dual-hop 60 GHz mmWave networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 11 180–11 192, Dec. 2017.

[17] Y. Wang, S. Mao, and T. Rappaport, "On directional neighbor discovery in mmwave networks," in *Proc. IEEE ICDCS 2017*, Atlanta, GA, June 2017, pp. 1704–1713.

[18] H. Shokri-Ghadikolaei, C. Fischione, G. Fodor, P. Popovski, and M. Zorzi, "Millimeter wave cellular networks: A MAC layer perspective," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3437–3458, 2015.

[19] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[20] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka *et al.*, "Scenarios for 5G mobile and wireless communications: the vision of the METIS project," *IEEE commun. mag.*, vol. 52, no. 5, pp. 26–35, 2014.

[21] C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE commun. mag.*, vol. 52, no. 2, pp. 122–130, 2014.

[22] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE journal of selected topics in signal processing*, vol. 8, no. 5, pp. 742–758, 2014.

[23] S. Sun, T. S. Rappaport, R. W. Heath, A. Nix, and S. Rangan, "MIMO for millimeter-wave wireless communications: Beamforming, spatial multiplexing, or both?" *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 110–121, 2014.

[24] I. Ahmed, H. Khammari, A. Shahid, A. Musa, K. S. Kim, E. De Poorter, and I. Moerman, "A survey on hybrid beamforming techniques in 5G: Architecture and system model perspectives," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3060–3097, 2018.

[25] L. Chen, Y. Yang, X. Chen, and W. Wang, "Multi-stage beamforming codebook for 60GHz WPAN," in *Proc. IEEE International ICST Conference on Communications and Networking in China*, 2011.

[26] Y. M. Tsang, A. S. Poon, and S. Addepalli, "Coding the beams: Improving beamforming training in mmwave communication system," in *Proc. IEEE Global Telecommunications Conference-GLOBECOM*, 2011.

[27] V. Venkateswaran and A.-J. van der Veen, "Analog beamforming in MIMO communications with phase shift networks and online channel estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4131–4143, 2010.

[28] S. Han, I. Chih-Lin, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 186–194, 2015.

[29] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 501–513, 2016.

[30] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, "Subspace estimation and decomposition for large millimeter-wave MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 528–542, 2016.

[31] J. Ma, G. Y. Li, and B. H. Juang, "Signal processing in cognitive radio," *Proc. of the IEEE*, 2009.

[32] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *Communications Surveys & Tutorials, IEEE*, vol. 11, no. 1, pp. 116–130, 2009.

[33] Y. Zhao, P. Paul, C. Xin, and M. Song, "Performance analysis of spectrum sensing with mobile sus in cognitive radio networks," in *Proc. IEEE International Conference on Communications (ICC)*, 2014.

[34] Y. Zhao, M. Song, and C. Xin, "A weighted cooperative spectrum sensing framework for infrastructure-based cognitive radio networks," *Computer communications*, vol. 34, no. 12, pp. 1510–1517, 2011.

[35] T. Yucek and H. Arslan, "Spectrum characterization for opportunistic cognitive radio systems," in *Proc. IEEE Military Communications Conference*, 2006.

[36] A. Mathew, N. Chandrababu, K. Elleithy, and S. Rizvi, "IEEE 802.11 & bluetooth interference: simulation and coexistence," in *Proc. IEEE Communication Networks and Services Research Conference*, 2009.

[37] B. Hu, C. Hua, C. Chen, X. Ma, and X. Guan, "MUBFP: Multiuser beamforming and partitioning for sum capacity maximization in MIMO systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 233–245, 2017.

[38] D. Lee, G. Y. Li, X.-L. Zhu, and Y. Fu, "Multistream multiuser coordinated beamforming for cellular networks with multiple receive antennas," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3072–3085, 2016.

[39] W.-L. Shen, K. C.-J. Lin, M.-S. Chen, and K. Tan, "SIEVE: Scalable user grouping for large MU-MIMO systems," in *Proc. IEEE Conference on Computer Communications (INFOCOM)*, 2015.

[40] X. Xia, S. Fang, G. Wu, and S. Li, "Joint user pairing and precoding in MU-MIMO broadcast channel with limited feedback," *IEEE Communications Letters*, vol. 14, no. 11, pp. 1032–1034, 2010.

[41] S. Huang, H. Yin, J. Wu, and V. C. Leung, "User selection for multiuser MIMO downlink with zero-forcing beamforming," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 7, pp. 3084–3097, 2013.

[42] L.-N. Tran, M. F. Hanif, and M. Juntti, "A conic quadratic programming approach to physical layer multicasting for large-scale antenna arrays," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 114–117, 2014.

[43] J. Choi, "Iterative methods for physical-layer multicast beamforming," *IEEE Transactions on Wireless Communications*, vol. 14, no. 9, pp. 5185–5196, 2015.

[44] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2239–2251, 2006.

[45] A. Adhikary, E. Al Safadi, M. K. Samimi, R. Wang, G. Caire, T. S. Rappaport, and A. F. Molisch, "Joint spatial division and multiplexing for mm-wave channels," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1239–1255, 2014.

[46] J. Yoon, K. Sundaresan, M. A. A. Khojastepour, S. Rangarajan, and S. Banerjee, "Joint multicell beamforming and client association in OFDMA small-cell networks," *IEEE Transactions on Mobile Computing*, vol. 15, no. 9, pp. 2260–2274, 2016.

[47] G. Dimic and N. D. Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm," *IEEE Transactions on Signal processing*, vol. 53, no. 10, pp. 3857–3868, 2005.

[48] Z. Ding, P. Fan, and H. V. Poor, "Random beamforming in millimeter-wave NOMA networks," *IEEE access*, vol. 5, pp. 7667–7681, 2017.

[49] G. Lee, Y. Sung, and J. Seo, "Randomly-directional beamforming in millimeter-wave multiuser MISO downlink," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1086–1100, 2016.

[50] C. N. Barati, S. A. Hosseini, S. Rangan, P. Liu, T. Korakis, S. S. Panwar, and T. S. Rappaport, "Directional cell discovery in millimeter wave cellular networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6664–6678, 2015.

[51] Y. Li, J. G. Andrews, F. Baccelli, T. D. Novlan, and C. J. Zhang, "Design and analysis of initial access in millimeter wave cellular networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6409–6425, 2017.

[52] T. S. Rappaport, E. Ben-Dor, J. N. Murdock, and Y. Qiao, "38 GHz and 60 GHz angle-dependent propagation for cellular & peer-to-peer wireless communications," in *Proc. IEEE International Conference on Communications (ICC)*, 2012.

[53] Y. Cho and J. Kim, "Line-of-sight MIMO channel in millimeter-wave beamforming system: Modeling and prototype results," in *Proc. IEEE Vehicular Technology Conference*, 2015.

[54] M. Mahajan, P. Nimbhorkar, and K. Varadarajan, "The planar k-means problem is np-hard," *Theoretical Computer Science*, vol. 442, pp. 13 – 21, 2012.

[55] C. A. Balanis, *Antenna Theory: Analysis and Design, 3rd ed.* Wiley-Interscience, 2005.

[56] H. L. Van Trees, *Optimum array processing: Part IV of detection, estimation, and modulation theory.* John Wiley & Sons, 2004.

[57] E. Bjornson, M. Bengtsson, and B. Ottersten, "Optimal multiuser transmit beamforming: A difficult problem with a simple solution structure [lecture notes]," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 142–148, 2014.

[58] A. B. Gershman, N. D. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 62–75, 2010.

[59] M. Min, Y.-S. Jeon, and G.-H. Im, "On achievable rate of user selection for MIMO broadcast channels with limited feedback," *IEEE Transactions on Communications*, vol. 65, no. 1, pp. 122–135, 2017.

[60] G. Lee and Y. Sung, "A new approach to user scheduling in massive multi-user mimo broadcast channels," *IEEE Trans. Commun.*, vol. 66, no. 4, pp. 1481–1495, Apr. 2018.

[61] J. Mao, J. Gao, Y. Liu, and G. Xie, "Simplified semi-orthogonal user selection for MU-MIMO systems with ZFBF," *IEEE Wireless Communications Letters*, vol. 1, no. 1, pp. 42–45, 2012.

[62] J. Choi, G. Lee, and B. L. Evans, "User scheduling for millimeter wave hybrid beamforming systems with low-resolution adcs," *IEEE Transaction on Wireless Communications*, vol. 18, no. 4, pp. 2401–2414, Apr. 2019.

[63] K. Ko and J. Lee, "Multiuser MIMO user selection based on chordal distance," *IEEE Transactions on Communications*, vol. 60, no. 3, pp. 649–654, 2012.

[64] M. Taniguchi, H. Murata, S. Yoshida, K. Yamamoto, D. Umehara, S. Denno, and M. Morikura, "Indoor experiment of multi-user mimo user selection algorithm based on chordal distance," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2013.

[65] B. Zhou, B. Bai, Y. Li, D. Gu, and Y. Luo, "Chordal distance-based user selection algorithm for the multiuser MIMO downlink with perfect or partial CSIT," in *Proc. IEEE International Conference on Advanced Information Networking and Applications*, 2011.

[66] 3GPP TR 38.901 version 14.3.0 Release 14, "Study on channel model for frequencies from 0.5 to 100 GHz," 2018. [Online]. Available: https://www.3gpp.org/release-14

[67] M. F. Hanif, L.-N. Tran, A. Tölli, M. Juntti, and S. Glisic, "Efficient solutions for weighted sum rate maximization in multicellular networks with channel uncertainties," *IEEE Transaction on Signal Processing*, vol. 61, no. 22, pp. 5659–5674, Nov. 2013.

[68] N. Deng and M. Haenggi, "A novel approximate antenna pattern for directional antenna arrays," *IEEE Transaction on Wireless Communications*, vol. 7, no. 5, pp. 832–835, Oct. 2018.

[69] 3GPP TS 36.101 version 14.5.0 Release 14, "Evolved universal terrestrial radio access (E-UTRA); user equipment (ue) radio transmission and reception," 2018. [Online]. Available: https://www.3gpp.org/release-14

[70] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3GPP NR at mmWave frequencies," vol. 21, no. 1, pp. 173–196, 2018.

[71] M. Koivisto, A. Hakkarainen, M. Costa, P. Kela, K. Leppanen, and M. Valkama, "High-efficiency device positioning and location-aware communications in dense 5G networks," vol. 55, no. 8, pp. 188–195, Aug. 2017.

[72] S. E. El-Khamy, M. El-Mahallawy, and E. Youssef, "Improved wideband spectrum sensing techniques using wavelet-based edge detection for cognitive radio," in *Proc. IEEE International Conference on Computing, Networking and Communications (ICNC)*, 2013.

[73] Z. Shi-bing and Q. Jin-jing, "Energy detection algorithm based on wavelet packet transform under uncertain noise for spectrum sensing," in *Proc. IEEE International Conference on Wireless Communications Networking and Mobile Computing*, 2010.

[74] S.-S. Jeng, J.-M. Chen, H.-Z. Lin, and C.-W. Tsung, "Wavelet-based spectrum sensing for cognitive radios using hilbert transform," *World Academy of Science, Engineering and Technology*, pp. 596–600, 2011.

[75] M. Eslami and S.-S. Sadough, "Wideband spectrum sensing for cognitive radio via phase-field segmentation," in *Proc. IEEE Conference on Wireless Advanced*, 2010.

[76] L. Zhang and P. Bao, "Edge detection by scale multiplication in wavelet domain," *Pattern Recognition Letters*, vol. 23, no. 14, pp. 1771–1784, 2002.

[77] S. Mallat, *A wavelet tour of signal processing*. Academic press, 1999.

# VITA

Prosanta Paul
Department of Electrical & Computer Engineering
Old Dominion University, Norfolk, VA 23529

Education

- Ph.D. Electrical and Computer Engineering, May 2020, Old Dominion University
- M.Sc. Electrical and Computer Engineering, August 2014, South Dakota School of Mines & Technology
- B.Sc. Electrical and Electronic Engineering, March 2009, Khulna University of Engineering & Technology

Publications

- Paul, Prosanta and Xin, ChunSheng and Wu, Hongyi and Song, Min, "*Beamforming Oriented Topology Control for mmWave Networks*," in Trans. on Mobile Computing, April 2019.
- P. Paul, H. Wu, and C. Xin, "BOOST: A User Association and Scheduling Framework for Beamforming mmWave Networks," in Trans. on Mobile Computing, (Accepted).
- Xin, ChunSheng and Paul, Prosanta and Song, Min,"*On Dynamic Spectrum Allocation in Geo-location Spectrum Sharing Systems*," in Trans. on Mobile Computing, June 2018.
- Paul, Prosanta, and Xin, ChunSheng and Zhao, Yanxiao and Song, Min, "*Spectrum Sensing for a Subdivided Band in Cognitive Radio Networks*," in Proc. ICCCN, Las Vegas, USA, August 2015.
- Zhao, Yanxiao and Paul, Prosanta and Xin, ChunSheng and Song, Min, "*Performance Analysis of Spectrum Sensing with Mobile SUs in Cognitive Radio Network*," in Proc. ICC, Sydney, Australia, June 2014.
- Paul, Prosanta, and Xin, ChunSheng and Zhao, Yanxiao and Song, Min, Chapter 9, "*Wideband Spectrum Sensing in Cognitive Radio Networks*," in Computer Communications and Networks: From Green, Mobile, Pervasive Networking to Big Data Computing, ISBN: 9788793379879, River Publishers, February 9, 2017.

Typeset using LaTeX.